

# Data-Driven Collective Variables in Meta Dynamics (Review)

Anna Denisenko<sup>1, 2</sup> and Barak Hirshberg<sup>1</sup>

<sup>1</sup>School of Chemistry, Center for the Physics & Chemistry of Living Systems, Tel Aviv University, 6997801 Tel Aviv, Israel

<sup>2</sup>Raymond and Beverly Sackler School of Physics & Astronomy,  
 Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel

(Dated: April 3, 2022)

Molecular dynamics simulations are a great tool, but suffer from a timescale problem when dealing with rare events. One solution is to use Meta Dynamics enhanced sampling method, with a collective variable (CV) as a bias, to progress faster toward a molecular transition. When dealing with complex molecules, it's a struggle to define a good CV, which will allow us to extract the free energy surface of the molecule and its thermodynamic properties. These days, different data-driven machine learning methods are used to extract CVs. In this paper we address 3 linear dimensionality reduction methods - PCA, LDA and HLDA, while trying to assess whether these methods can find us a good CV for a molecule of medium complexity, on which we already know almost everything - alanine dipeptide. PCA produced very poor results (3 transitions in 20 ns), while LDA and HLDA produced impressive results (44 and 39 transitions respectively in 20 ns), with a slight advantage to LDA.

## I. INTRODUCTION

### A. Molecular Dynamics Simulations

Since the first simulation of a protein 45 years ago, molecular dynamics (MD) simulations have progressed towards an ever more realistic representation of the solvent and surroundings of molecules, and towards larger systems and longer equilibration times. These models can be parametrized using hundreds of nanoseconds of sampling time, allowing accurate tuning to reproduce average equilibrium properties, such as free energy surfaces (FES). For systems under the ergodic hypothesis [1], the FES and other averages of a MD simulation can represent the thermodynamic properties of a system at the macroscopic scale, like protein folding for example. Complicated systems can be simulated, such as fully solvated membrane protein complexes. Thus, MD simulations can be applied to attack problems like protein folding in different solvents, explaining spectroscopic data and designing novel ligands. All of these were driven by the improvement over the years in simulation methodology, increasing accuracy of force fields and the increasing power of computers [2]. MD simulations are even more powerful these days, when every researcher and student can run them on their private computer, and are extensively used in fields as diverse as chemistry, biology, and materials science.

In MD simulations, the time is divided into discrete time steps, no more than a few femtoseconds each, as the accuracy of such calculations would depend on the length of the time interval. At each time step the forces acting on each atom are computed, using a molecular mechanics force field and then the atoms are moved a little bit, the position and velocity of each atom are updated accordingly, using Newton's laws of motion, assuming the atoms are not quantum particles and the connections between atoms act like springs with specific force constant. One can also determine the system size, boundary conditions,

and more features of the simulations [3].

#### 1. The Time Scale Problem

So far we can conclude that MD simulations are very powerful, but they still suffer from a major problem - the time scale problem. When studying systems that exhibit complex free energy landscapes in which long-lived metastable states separated by kinetic bottlenecks, this problem surfaces quickly. The kinetic bottlenecks hinder transition between metastable states and restrict the time scale that can be explored [4]. Special purpose machines[5] have been constructed to push this limit. However, in today's science there are still quite a few problems with time scales that puts them out of the reach of direct simulations, not to mention the fact that such machines can be accessed only by a restricted number of researchers. This has motivated a vast community of modelers to develop enhanced sampling methods that allow investigating the properties of complex systems in an affordable computational time, overcoming kinetic bottlenecks, and exploring different metastable states.

#### 2. Meta Dynamics to the Rescue

One enhanced sampling method is called Meta Dynamics (MetaD), which relies on identifying Collective Variables (CVs), a small subset of the degrees of freedom that describes the slowest modes of the process. MetaD is based on iteratively 'filling' the potential energy of the system by a sum of Gaussians centred along the trajectory followed by a suitably chosen set of CVs, thereby forcing the system to migrate from one minimum to the next [6].

Collective variables (also called Reaction Coordinates), are building blocks in our understanding of molecular potential energy surfaces. In a similar way to generalized coordinates in analytical mechanics [7], a good CV is one

that can differentiate between two (or more) states of the system, but also needs to work well as a bias for enhanced sampling.

But CVs are not always easy to find and calculate. Data driven methods allow us to choose our CVs wisely, and thus better understand free energy surfaces of complex molecular systems. Using biased simulations in MetaD can also help us retrieve free energy surfaces in much smaller time scales.

In this paper we will test the performance of several dimensionality reduction algorithms to locate optimal CVs and answer the question: whether these data driven methods can supply us with good CVs, when a small amount of prior knowledge is used.

### B. Alanine Dipeptide

Alanine dipeptide is the canonical system of molecular dynamics. It has two main conformers, one (C7eq) is more stable than the other (C7ax), which can be differentiated by the dihedral angles  $\Phi$  and  $\Psi$ . It is already known that the dihedral angle  $\Phi$  is a good CV while  $\Psi$  is not. Another dihedral angle, which is of less importance but still used in some papers is  $\Theta$  [8] and will also play a role in this paper.

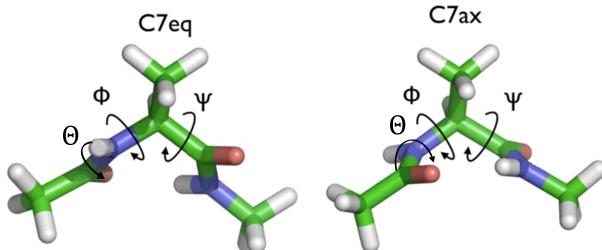


FIG. 1: The two main conformers of alanine dipeptide - C7eq and C7ax. Those can be differentiated by the dihedral angles -  $\Phi$  and  $\Psi$ , accompanied by the torsional angle  $\Theta$ .

## II. METHODS

The data for the dimensionality reduction algorithms was obtained from 2 ns unbiased simulations of each conformer, separately, using the Amber99-SB force field and performed with GROMACS 2018.6 [9], with an integration time step of 0.5 fs. Trajectories were generated in the NVT ensemble with the temperature maintained at 300 K using the stochastic velocity re-scaling thermostat, according a Maxwell distribution in a well tempered MetaD.

Meta Dynamics calculations were performed within the PLUMED2 plugin with Gaussian hills deposited every 500 integration steps with an initial hill height of 1.2  $kJ/mol$ , a width of 0.03 and a bias factor  $\gamma = 6$ .

All the files used to create the .tpr files for GROMACS can be found [here](#).

As the paper focuses on whether a data driven method can lead us to a good CV, we should first define one. A good CV will be considered by the number of transitions performed between two conformers, during a biased MetaD simulation. When looking at a CV vs time plots, it seemed like the CV is very fluctuating, and for CVs with bad performance, calculating the number of transitions using the CV was nearly impossible (the value of the CV for both conformers can be too close to differentiate). Considering these facts, it was decided that the number of transitions between each conformer will be determined by the angle  $\Phi$  which is (unsurprisingly) differentiating very well between the two conformers. Each 6 timesteps were averaged (to reduce noise, caused by the periodic boundaries) and the conformers were differentiated by the following rule:

- C7ax - where  $\Phi > 0.9$  rad
- C7eq - where  $\Phi < -2.2$  rad

Another way to evaluate the robustness of the methods, is to use references. This paper will rely on three of them: 1D FES, 2D FES and  $\Delta F$ .

The 1D FES show the 1 dimensional free energy surface for alanine dipeptide per each angle ( $\Phi$  and  $\Psi$ ). For each method the FES will be compared to the one created when running MetaD with  $\Phi$  or  $\Psi$  as a bias, respectively. The FES is calculated using the PLUMED2 method of umbrella-sampling-like re-weighting [10]. The metadynamics bias potential  $V(s)$  is calculated using all the Gaussians deposited along the entire trajectory and the weight  $w$  of each frame can be obtained by using the expression

$$w \propto \exp \frac{V(s)}{k_B T} \quad (1)$$

The weights are being used to calculate a histogram every 50 steps which is printed every 400k steps, using PLUMED2 functionality. The histograms  $h(s)$  are converted into free energies by using the formula

$$F(s) = -k_B T \cdot \log(h(s)) \quad (2)$$

In a similar way, the 2D FES show the 2 dimensional free energy surface for alanine dipeptide for both angles ( $\Phi$  and  $\Psi$ ). The FES will be compared to the one created when running MetaD with  $\Phi$  and  $\Psi$  together as a bias, as it shows the broadest view of the FES. The 2D FES was calculated by using a 2D histogram with the same weights as in EQ. (1) leading to the FES function in the same way as in EQ. (2).

$\Delta F$  will be calculated using the 1D FES functions, as it symbolises the convergence of the alanine dipeptide free energy difference. The free energy difference between the basins of the 1D FES is as follows:

$$\Delta F = \frac{1}{\beta} \log \frac{\int_A e^{-\beta F(\Phi)} d\Phi}{\int_B e^{-\beta F(\Phi)} d\Phi} \quad (3)$$

Where  $\Phi$  is the dihedral angle and  $F(\Phi)$  is the reweighted free energy. The two integrals are computed over the regions of phase space corresponding to metastable state A (C7eq) and B (c7ax), in this case  $\Phi < 0$  and  $\Phi > 0$ . The update of  $\Delta F$  is performed every 0.02 ns. We will consider a fine convergence as one that stays within the 0.5  $k_B T$  uncertainty. We expect the  $\Delta F$  value to be around 10  $kJ/mol$ . [11].

All of the comparisons above were calculated after running 10 different simulations (with 10 different seeds), for each case, using the mean and the standard deviation of all sizes. All of the analysis and running scripts can be found [here](#) for future reference.

### A. PCA

Principal Component Analysis (PCA) is used in exploratory data analysis and for making predictive models as well as for dimensionality reduction by projecting each data point onto only the first few principal components to obtain lower-dimensional data while preserving as much of the data's variation as possible. The first principal component can equivalently be defined as a direction that maximizes the variance of the projected data. The  $i$ -th principal component can be taken as a direction orthogonal to the first  $i - 1$  principal components that maximizes the variance of the projected data [12].

More mathematically speaking, consider an  $n \times p$  data matrix  $\mathbf{X}$ , with column-wise zero empirical mean (which we can assume relates to our case), where each row presents a different timestamp in the simulation, for both conformers concatenated, and each column is a feature (a dihedral angle in our case). The transformation is defined by a set of size  $l$  (this paper discusses only the first principle component, so  $l = 1$ ) of  $p$ -dimensional vector of weights  $\mathbf{w}$  that map each row vector  $\mathbf{x}_i$  of  $\mathbf{X}$  to a new vector of principal component score  $\mathbf{t}_i$ , given by

$$\mathbf{t}_i = \mathbf{x}_i \cdot \mathbf{W} \quad (4)$$

in such a way that the individual variables  $t_i$  of  $\mathbf{t}$  considered over the data set, successively inherit the maximum possible variance from  $\mathbf{X}$ , with the coefficient vector  $\mathbf{w}$  constrained to be a unit vector.

In order to find the principal component and maximize the variance, the weight component should satisfy

$$\mathbf{W} = \arg \max \left( \sum_i (\mathbf{x}_i \cdot \mathbf{W})^2 \right) \quad (5)$$

or, equivalently

$$\mathbf{W} = \arg \max (\mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W}) \quad (6)$$

Since  $\mathbf{w}$  has been defined to be a unit vector, it equivalently satisfies

$$\mathbf{W} = \arg \max \left( \frac{\mathbf{W}^T \mathbf{X}^T \mathbf{X} \mathbf{W}}{\mathbf{W}^T \mathbf{W}} \right) \quad (7)$$

The quantity that has to be maximized (the one inside the brackets) can be recognized as a Rayleigh quotient. A standard result for a positive semi-definite matrix such as  $\mathbf{X}^T \mathbf{X}$  is that the quotient's maximum possible value is the largest eigenvalue of the matrix, which occurs when  $\mathbf{w}$  is the corresponding eigenvector.

### B. LDA and HLDA

The following methods were done accordingly to Dan Mendels et al. paper [13].

Linear Discriminant Analysis (LDA) is used to find a linear combination of features that characterizes or separates two or more classes of objects or events. The resulting combination may be used as a linear classifier, or, more commonly, for dimensionality reduction before later classification.

While trapped in the metastable states, the  $N_d$ ,  $d$  dimensional descriptors (dihedral angles) of alanine dipeptide, will have an expectation value  $\mu_{A,B}$  and a multivariate variance  $\Sigma_{A,B}$ . These quantities can be evaluated in short unbiased runs. The data generated in such runs will be separated in the 3 dimensional space of the dihedral angles -  $\Phi$ ,  $\Psi$  and  $\Theta$ . Our first goal is to find a one-dimensional projection of these two sets of data along which they still do not overlap, otherwise it will be very hard to separate them. If we take a generic projection  $x = \mathbf{W}^T \mathbf{d}$ , where  $W$  is a vector in the  $N_d$  space, there is no guarantee that they will not overlap. LDA aims at finding the direction  $\mathbf{W}$  along which the projected data is best separated. To do this one needs a measure of their degree of separation. Following Fisher [14], this is measure is given by the ratio between the so-called between class  $\mathbf{S}_b$  and the within class  $\mathbf{S}_w$  scatter matrices. The former is measured by the square of the distance between the projected means

$$\tilde{\mu}_A - \tilde{\mu}_B = \mathbf{W}^T (\mu_A - \mu_B) \quad (8)$$

which can be written as

$$\mathbf{W}^T \mathbf{S}_b \mathbf{W} \quad (9)$$

where

$$\mathbf{S}_b = (\mu_A - \mu_B)(\mu_A - \mu_B)^T \quad (10)$$

The within spread matrix is estimated from the sum of the two spreads, leading to the expression

$$\mathbf{W}^T \mathbf{S}_w \mathbf{W} \quad (11)$$

with

$$\mathbf{S}_w = \boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_B \quad (12)$$

Where  $\Sigma$  is the co-variance matrix. The Fisher's object function then reads like a Rayleigh ration, in the same way as in PCA

$$W = \arg \max \frac{\mathbf{W}^T \mathbf{S}_b \mathbf{W}}{\mathbf{W}^T \mathbf{S}_w \mathbf{W}} \quad (13)$$

that maximized by

$$\mathbf{W}^* = \mathbf{S}_W^{-1} (\boldsymbol{\mu}_A + \boldsymbol{\mu}_B) \quad (14)$$

This would suggest that a useful one-dimensional CV is

$$s_{LDA}(\mathbf{R}) = (\boldsymbol{\mu}_A - \boldsymbol{\mu}_B)^T (\boldsymbol{\Sigma}_A + \boldsymbol{\Sigma}_B)^{-1} d(\mathbf{R}) \quad (15)$$

where  $\mathbf{R}$  are the atomic coordinates.

LDA is closely related to PCA as they both look for linear combinations of variables which best explain the data. LDA explicitly attempts to model the difference between the classes of data. PCA, in contrast, does not take into account any difference in class. People are used to think that LDA is generally better than PCA, but one conclusion is that when the training data set is small, PCA can outperform LDA and, also, that PCA is less sensitive to different training data sets. [15].

In Dan Mendels et al. paper, the performance of an LDA CV seemed to be mediocre. In their paper, a Diels-Alder reaction was examined, and the free energy surface was of course different than alanine dipeptide's, so the explanation given was that when taking the arithmetic averages of the covariances, it is the larger one that carries more weight. From a data analysis point of view, this seems a bit counter intuitive because the data with smaller variance is better defined. Also, from the rare event point, it would seem more appropriate to give more weight to the state with the smaller fluctuations, which is more difficult to get into and escape from, as the covariance and variance are also strongly related to the free energy plot width. For these reasons, they propose a different measure of the scatter, and rather than using the arithmetic average, they base the measure of the within scatter matrix on the harmonic average as follows

$$\mathbf{S}_w = \left( \frac{1}{\boldsymbol{\Sigma}_A} + \frac{1}{\boldsymbol{\Sigma}_B} \right)^{-1} \quad (16)$$

Which leads to a different expression for the collective variable

$$s_{HLDA}(\mathbf{R}) = (\boldsymbol{\mu}_A - \boldsymbol{\mu}_B)^T \left( \frac{1}{\boldsymbol{\Sigma}_A} + \frac{1}{\boldsymbol{\Sigma}_B} \right) d(\mathbf{R}) \quad (17)$$

Opposing Dan Mendels et al. paper [13], as described in the SI, it was decided not to use the angles themselves, considering the periodicity of the angles. While the dihedral angles are periodic (from  $\pi$  to  $-\pi$ ), and the bias on one angle or more is still periodic in the same way, a linear combination of the angles might not have the same periodicity. It is necessary for the linear combination to act in the same way, to portray the angles transformations in the most accurate way, therefore trigonometric functions, which are periodic by definition, were used instead.

### C. Determining the Phase Factor $\Theta_0$

When dealing with trigonometric function, the phase should be discussed. This is an odd discussion, considering that we are dealing with quantum mechanical systems, in which the phase can't even be measured, but we can assume that while transitioning our system to work accordingly to Newton mechanics (as we should in MD simulations), the phase does play a role. This idea came to us while looking at J. McCarty et al. paper [16], in which the reference angle  $\Theta_0$  was chosen to be 1.2 rad, for a CV of a form

$$0.5 + 0.5 \cos(\Theta - \Theta_0) \quad (18)$$

The value  $\Theta_0$  was determined in P. Tiwarya and B. J. Berne's paper [17], while looking at a comparison of spectral gap.

Choosing  $\Theta_0$  that is different than zero, creates a change in the phase between the conformers, and following our need in any of the methods above to separate between the two conformers, it is safe to assume that there might be an optimal value for  $\Theta_0$  (which will of course be of periodic nature).

To test  $\Theta_0$ , it is of course possible to use the methods described above (number of transitions, 1D FES, etc) and indeed we compared the number of transition for each  $\Theta_0$ , but MetaD simulations are much heavier than unbiased ones, so it was decided to explore what are the different methods we have that might help us choose the best  $\Theta_0$  before even running one MetaD simulation. The methods we used are the distribution width of each conformer CV, and the distance between histograms.

The distribution of each collective variable per conformer was determined by calculating the histogram of the CV over the unbiased simulation time of 2 ns, as can be shown in Fig. 2, as well as the power of  $\Theta_0$  on the distribution.

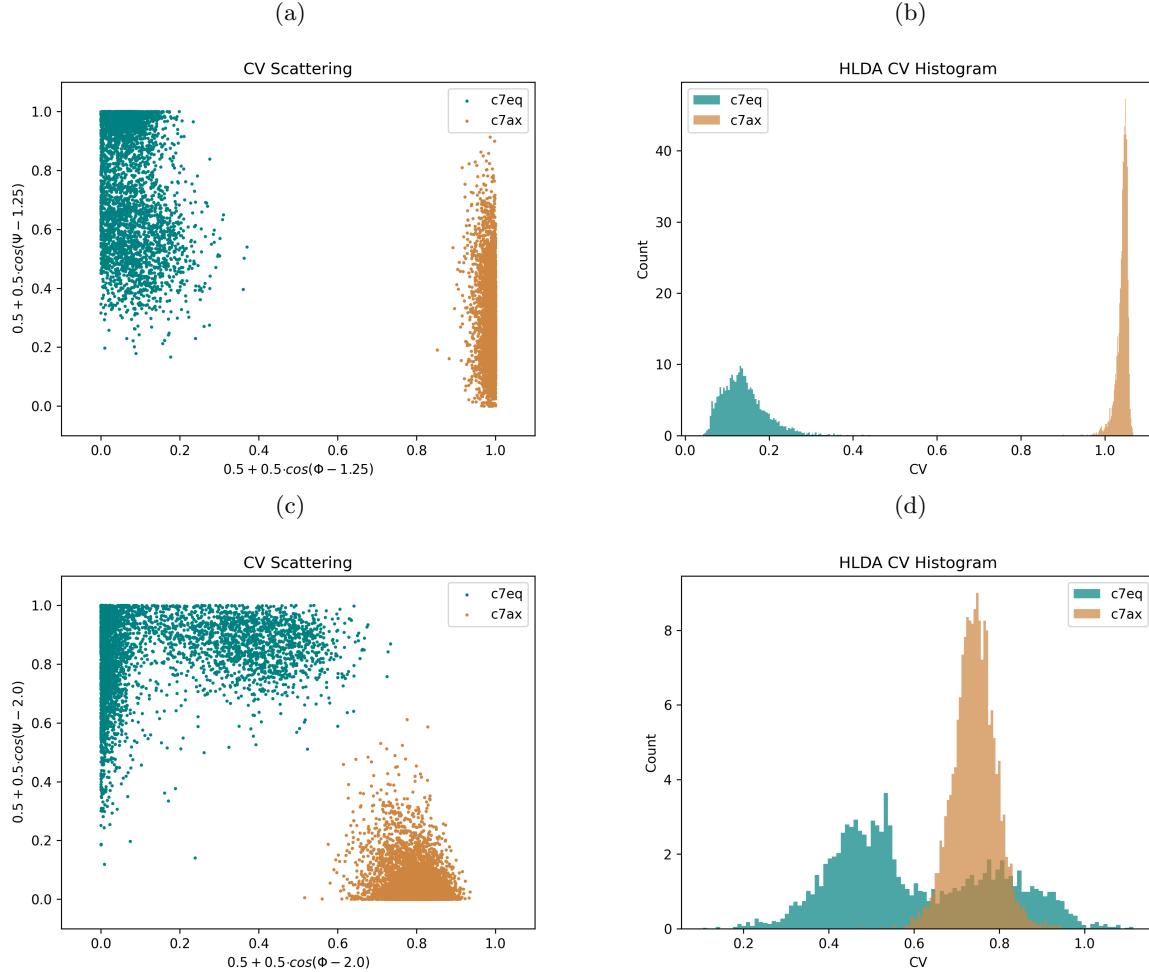


FIG. 2: Collective variable distribution for HLDA method, using  $\Theta_0 = 1.25$  rad (a+b), and  $\Theta_0 = 2.0$  rad (c+d) in comparison, showing the more stable conformer, C7eq, on the left, and C7ax on the right.

The distance between each distribution was also calculated using

$$D = \frac{|\mu_A - \mu_B|}{\sqrt{\sigma_A^2 + \sigma_B^2}} \quad (19)$$

where  $\mu_A$  and  $\mu_B$  are the means of each conformer distribution and  $\sigma_A$  and  $\sigma_B$  are the standard deviations respectively.

### III. RESULTS

Let's convince ourselves first that the timescale problem does exist. As can be seen in FIG. 4, when running a 20 ns long simulation (took 2 hours, 0.098 hour/ns), only the unstable conformer C7ax fliped once. If we define one transition as a flip forward and then back, we don't see even one full transition even when running an altogether 40 ns simulation [18].

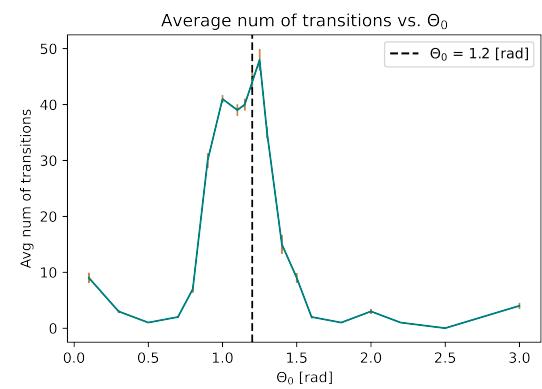


FIG. 3: Averaged for 10 simulations of 5 ns, the number of transitions from one conformer to the other, as a function of  $\Theta_0$ , with  $\Theta_0 = 1.2$  as a reference.

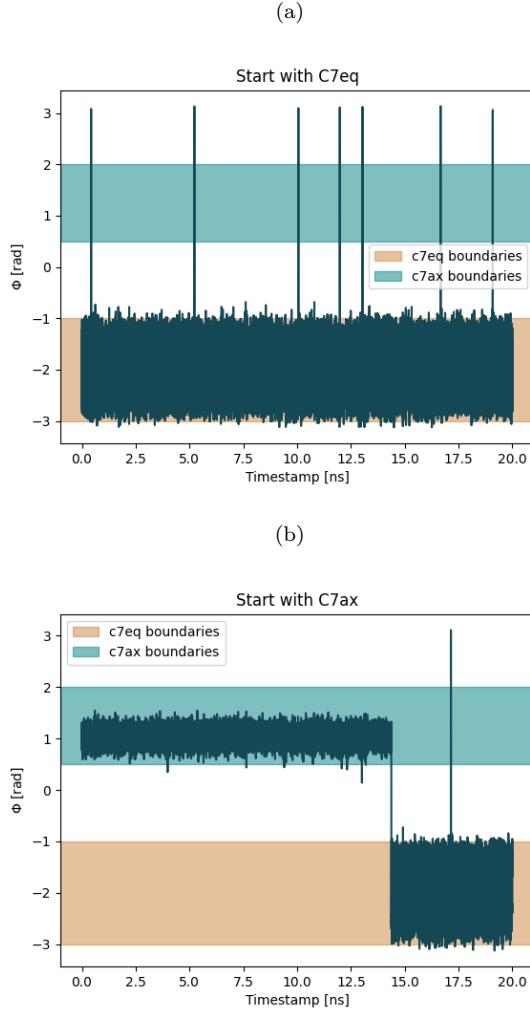


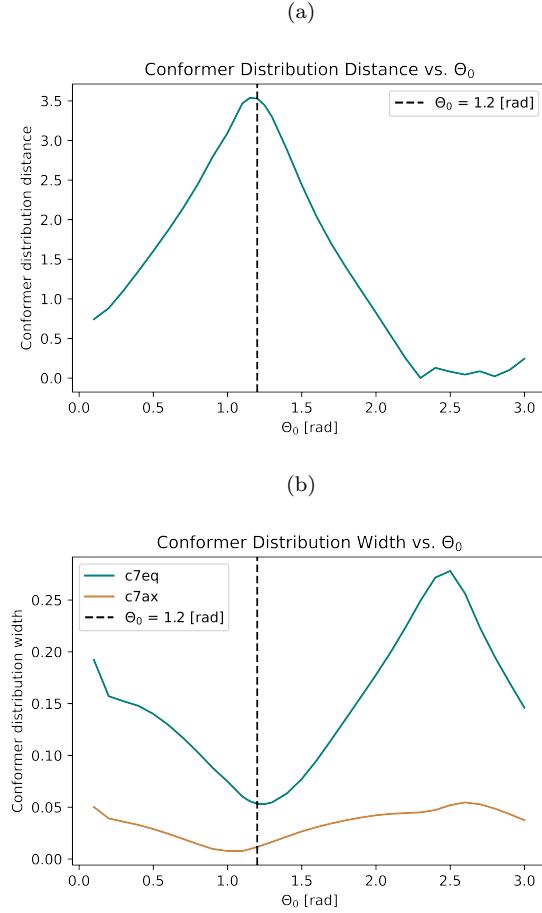
FIG. 4: Transitions in unbiased simulations of 20 ns, (a) starting with C7eq, the more stable conformer, (b) starting with C7ax. The extreme jumps in value are caused by the periodic boundaries.

### A. But First, $\Theta_0$

Now that we are convinced that the timescale problem is real, and transition between the conformers is a rare event, we shall proceed to finding the best CV there is to make this event a little bit less rare. But first the optimal  $\Theta_0$  was determined.

One of the methods (HLDA) was chosen to test the optimal angle using MetaD. As described in EQ. 18,  $\Theta_0$  was chosen from 0.1 till 3.0 rad with a step of 0.1 rad. The average number of transitions per each value of the angle can be seen in FIG. 3. We can see that the maximum value of transitions is received when  $\Theta_0 = 1.25$ . This is the angle that was chosen for the rest of the analysis.

It should also be noted that MetaD simulations using  $\cos(\Theta)$  and  $\sin(\Theta)$  were run, and they had the same effect (the sine function worked better than the cosine) as the sine and cosine functions also have a phase between them,



but none of them has the optimal one.

We were interested to see if the influence of  $\Theta_0$  is similar to the influence of choosing a low, inappropriate, coefficient for the cosine functions of  $\Phi$  or  $\Psi$ , these results are shown in the appendix, FIG. A.

It is hard to compare both settings because of the different scale, but it can be said that the influence is very similar. Which means the phase is playing a relatively big role, but probably still less than the coefficients. As we can see in Appendix A mainly in (b) but also in (a) the relative amount of good coefficients (over 30 transitions) is much smaller than in FIG. 3.

In retrospect, as already noted, it also makes sense to look at unbiased simulations and try to see whether we can figure out the optimal  $\Theta_0$  just by looking at 2 ns unbiased simulations. The CV histogram width and distance between conformers as a function of  $\Theta_0$  is shown in FIG. 5.

The conformer distribution width in FIG. 5(b) demonstrates periodic effect, as expected when dealing with trigonometric functions and their phase.

There is a correlation between the conformer distri-

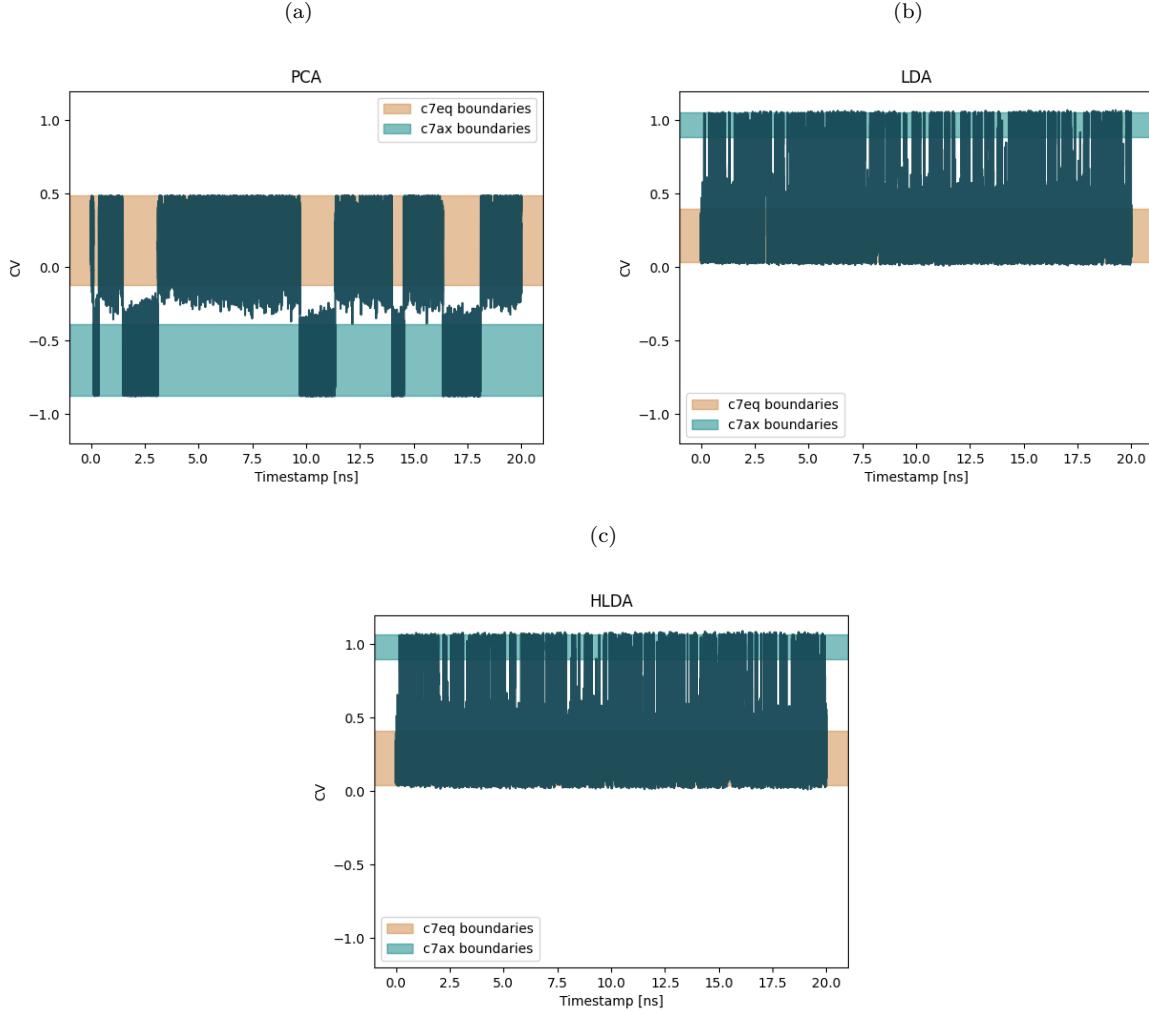


FIG. 6: The CV, as a function of time in a 20 ns biased simulation, per each method.

bution width (FIG. 5(b)) and the distance between the conformers (FIG. 5(a)). The narrower the distributions are, the farther they are. The reference angle  $\Theta_0$  is also correlating very well with the results. Therefore, we can learn a lot about the value of a CV from unbiased simulations, even before running one MetaD simulation.

### B. Comparing the Methods

All of the bias simulations were on CVs based on EQ. 18, with the optimal value of  $\Theta_0$ , 1.25 rad, with the 3 linear methods - PCA, LDA, HLDA.

*CV vs Time* - When comparing the CV versus time in FIG. 6 it gives us a good intuition about the number of transitions, and it is clearly seen that PCA is far behind LDA and HLDA in performance. While we can easily count the number of transitions in 20 ns in PCA CV, the transitions in LDA and HLDA CV are too fast to count by eye.

*$\Delta F$  Convergence* - Looking at the convergence of  $\Delta F$  versus time, we can see that  $\Phi$  and  $\Psi$  as a CV are not converging fast and both have relatively big errors, compared to the other methods. PCA is not even converging at all in the 20 ns time limit. LDA and HLDA are much faster to converge, with smaller errors. HLDA overtook LDA by negligible time, but had a worse start with larger standard deviation. Both methods converge even faster than  $\Phi$  as a bias.

*2D FES* - When comparing two 2D free energy surfaces, one from  $\Phi$  and  $\Psi$  as a bias, and one from LDA CV as a bias, in Appendix B, we can see that at a first glance the  $\Phi$  and  $\Psi$  one is more informative, but actually the LDA one shows us more of the mount between the two valleys - the conformers. Though both of the 2D FES were extracted from 20 ns simulations, the  $\Phi$  and  $\Psi$  is much more computationally heavy and took 6.5 hours (0.332 hour/ns) while the one took half of the time (0.175 hour/ns).

*1D FES* - In the 1D FES comparison in Appendix C we

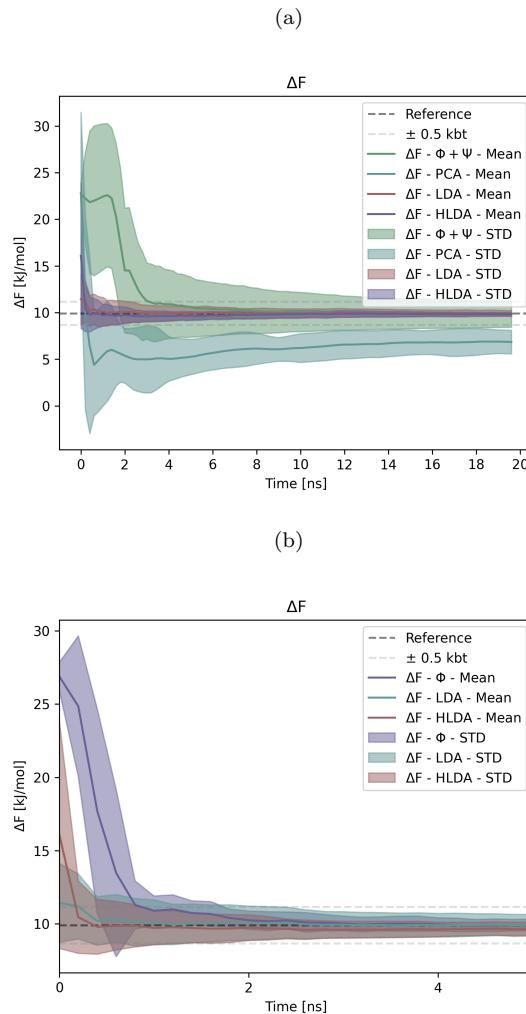


FIG. 7:  $\Delta F$  over time as calculated from EQ. 3 compared between a few methods, (a) best results in PCA, LDA and HLDA compared to  $\Phi$  and  $\Psi$  as a bias, (b) comparing the best three methods, showing that LDA and HLDA are leading in convergence.

can see very small differences between LDA and HLDA. It is easy to see that the PCA method is doing fine, but struggling more than the others as the standard deviation is bigger. Using  $\Phi$  and  $\Psi$  as a bias produces the poorest FES, but the big advantage of this CV is that the peak of the C7ax conformer is also reachable (while in the others there is not enough information), and is not far from the reference.

#### IV. SUMMARY

TABLE I summarizes all the different measurements taken per each CV. Through all the measurements and plots, we can see that  $\Phi$  is a good CV as expected, but LDA and HLDA CVs are even better, both with a very similar performance. PCA CV and  $\Phi$  and  $\Psi$  as a CV

showed poor results as CVs, though  $\Phi$  and  $\Psi$  was a little bit better.  $\Psi$  as a CV, as expected, showed the worst results.

It is worth noting that the width and distance for  $\Phi$  as a CV have worse stats than for LDA CV or HLDA CV, and this is caused of course by the unit difference.  $\Phi$  is measured in radians and goes between  $-\pi$  and  $\pi$  while all of the CVs for the three dimensionality reduction methods are with no units, and go between -1 and 1. So it makes sense that all of the measurements for the linear combination CVs will be relatively smaller.

#### V. DISCUSSION AND CONCLUSIONS

It is safe to say that dimensionality reduction techniques are able to learn a good CV from short (only 2 ns!) unbiased simulations of each state.

In addition, we can conclude that PCA is not a suitable method for identifying CVs, compared to LDA and HLDA. This is caused by the fact that to find the best CV, we need the most distinct separation between the conformers. All three methods are looking for linear combinations of variables which best explain the data, but while LDA and HLDA explicitly attempt to model the difference between the classes, PCA does not take into account any difference in class, which makes it usually an inferior method of separation.

No distinct difference was observed between LDA and HLDA, as in our case the molecule is too simple, and it's obvious that there is only one angle which is a good CV -  $\Phi$ . As discussed before, with alanine dipeptide the covariance of the two classes is approximately the same, as the two potential wells are of the same width. Checking the results on a molecule or process with higher complexity might give more insights to the difference between the two methods.

It was interesting to see that  $\Phi$  and  $\Psi$  together as a bias are a good method for exploring the 2D and 1D FES, but as can be seen from the  $\Delta F$  plots in FIG. 7 and from the measurements in TABLE I, it is not a good CV. This summarizes well our discussion about CVs in the beginning - a good CV is one that can differentiate between two (or more) states of the system, but they also need to work well as a bias for enhanced sampling.

As for  $\Theta_0$ , we learned that adding a degree of freedom can improve results, but as always, we should be cautious about adding too many degrees of freedom that might lead us to a wrong model. In this case, the phase  $\Theta_0$  was used to improve the statistical separation between the two conformers, but didn't hold any physical meaning.

*Future Work* - The observed mechanism is applicable to other more complex systems, providing new research directions, while checking all of the measurements above (number of transitions, distribution width and distance, as well as  $\Delta F$  convergence - which can be used even when the reference  $\Delta F$  value isn't known, as for at least two converging methods it's easy to see the expected value).

CV	Relative weight of $\Phi$	Number of Transitions	C7eq Width	C7ax Width	D	$\Delta F$ [kJ/mol]
$\Phi$	1	42	0.555128 rad	0.133758 rad	5.144912 rad	9.921239
$\Psi$	0	1	1.049933 rad	0.346322 rad	2.319498 rad	2.846318
$\Phi + \Psi$	0.5	9				9.842234
PCA	-0.868940	3	0.125775	0.076967	6.900959	6.845382
LDA	0.998280	44	0.052955	0.013992	16.422705	9.903184
HLDA	0.996697	39	0.053165	0.013710	16.399725	9.836306

TABLE I: Every value is summarized over 10 simulations of 20 ns. *Relative weight of  $\Phi$*  - the relative weight of dihedral angle  $\Phi$  or its trigonometric function in the CV, *Number of transitions* - calculated as explained in the methods section, *C7eq and C7ax width* - the width of each CV histogram per conformer, as shown in FIG. 2(b+d), *D* - the distance between the two histograms,  $\Delta F$  - last value of  $\Delta F$  after 20 ns, whether converged or not.

Applying the method on different molecules with diverse complexity is a key to a consistency test for the methods, when the results are reconstructed in different settings. But walking this way can also open a door to new results. For example, as discussed above, LDA and HLDA aren't showing distinct difference in the case of alanine dipeptide, but when running on a more complex molecule with different width for each potential well per each conformer, the harmonic part in HLDA plays a big role.

It is also important to note, that in this paper prior knowledge was used, as we know the dihedral angle  $\Phi$  is a good CV, which allowed us to concentrate only on the angles. But how we can start researching a molecule

which we know almost nothing about?

The linear dimensionality reduction methods are a good start with prior knowledge, but deep learning non-linear methods can be much more useful when unable to make head or tail. Such methods might yield information about the important degrees of freedom of each molecule and give us a head start on our knowledge on the molecule. For example, blindly using distances between atoms as a CV given to a DeepLDA neural network [11], might surface some important distances that can be directly related to the dihedral angle  $\Phi$ .

The research of data-driven collective variables is alive and kicking, there are many more sophisticated methods possible like deep learning and multiple class dimensionality cases that might lead to further improvement.

- 
- [1] L. Boltzmann, *Vorlesungen über Gastheorie: Th. Theorie van der Waals'; Gase mit zusammengesetzten Molekülen; Gasdissociation; Schlussbemerkungen*, Vorlesungen über Gastheorie (J. A. Barth, 1898).
- [2] T. Hansson, C. Oostenbrink, and W. van Gunsteren, Current Opinion in Structural Biology **12**, 190 (2002).
- [3] B. J. Alder and T. E. Wainwright, The Journal of Chemical Physics **31**, 459 (1959), <https://doi.org/10.1063/1.1730376>.
- [4] B. Peters, in *Reaction Rate Theory and Rare Events Simulations*, edited by B. Peters (Elsevier, Amsterdam, 2017) pp. 227–271.
- [5] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, Science **334**, 517 (2011), <https://www.science.org/doi/pdf/10.1126/science.1208351>.
- [6] G. Bussi and A. Laio, Nature Reviews Physics **2**, 200 (2020).
- [7] L. Hand and J. Finch, *Analytical Mechanics* (Cambridge University Press, 1998).
- [8] P. G. Bolhuis, C. Dellago, and D. Chandler, Proceedings of the National Academy of Sciences **97**, 5877 (2000), <https://www.pnas.org/doi/pdf/10.1073/pnas.100127697>.
- [9] P. Bauer, B. Hess, and E. Lindahl, Gromacs 2022 manual (2022).
- [10] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, Computer Physics Communications **185**, 604 (2014).
- [11] L. Bonati, V. Rizzi, and M. Parrinello, The Journal of Physical Chemistry Letters **11**, 2998 (2020), pMID: 32239945, <https://doi.org/10.1021/acs.jpclett.0c00535>.
- [12] Principal component analysis for special types of data, in *Principal Component Analysis* (Springer New York, New York, NY, 2002) pp. 338–372.
- [13] D. Mendels, G. Piccini, and M. Parrinello, The Journal of Physical Chemistry Letters **9**, 2776 (2018), pMID: 29733652, <https://doi.org/10.1021/acs.jpclett.8b00733>.
- [14] J. Cohen, P. Cohen, S. West, and L. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (Taylor & Francis, 2013).
- [15] A. Martinez and A. Kak, IEEE Transactions on Pattern Analysis and Machine Intelligence **23**, 228 (2001).
- [16] J. McCarty and M. Parrinello, The Journal of Chemical Physics **147**, 204109 (2017), <https://doi.org/10.1063/1.4998598>.
- [17] P. Tiwary and B. J. Berne, Proceedings of the National Academy of Sciences **113**, 2839 (2016), <https://www.pnas.org/doi/pdf/10.1073/pnas.1600917113>.
- [18] When running on LAMMPS unbiased, with what was apparently a broken model, in a simulation of 400 ns the number of transitions was, starting with C7ax - 0.5 and starting with C7eq - 1.

### Appendix A: Average Number of Transitions vs. $\Phi$ Coefficient

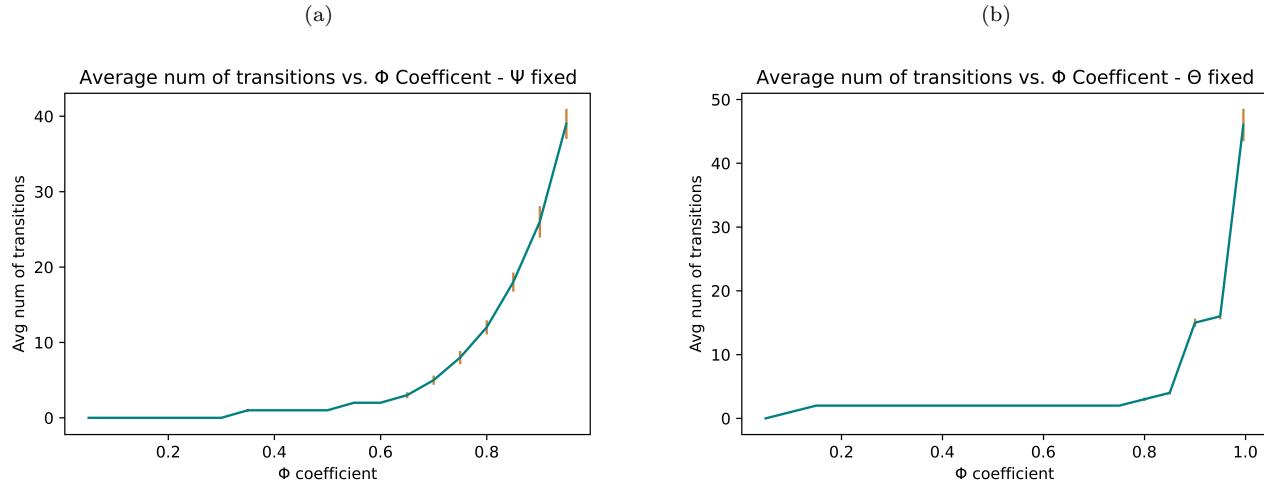


FIG. 8: Average number of transitions as a function of the relative weight of  $\Phi$ , (a) when the weight of  $\Psi$  is fixed on its optimal value, (a) when weight of  $\Theta$  is fixed on its optimal value. The coefficients vector is normalized to 1.

### Appendix B: 2D Free Energy Surface Comparison

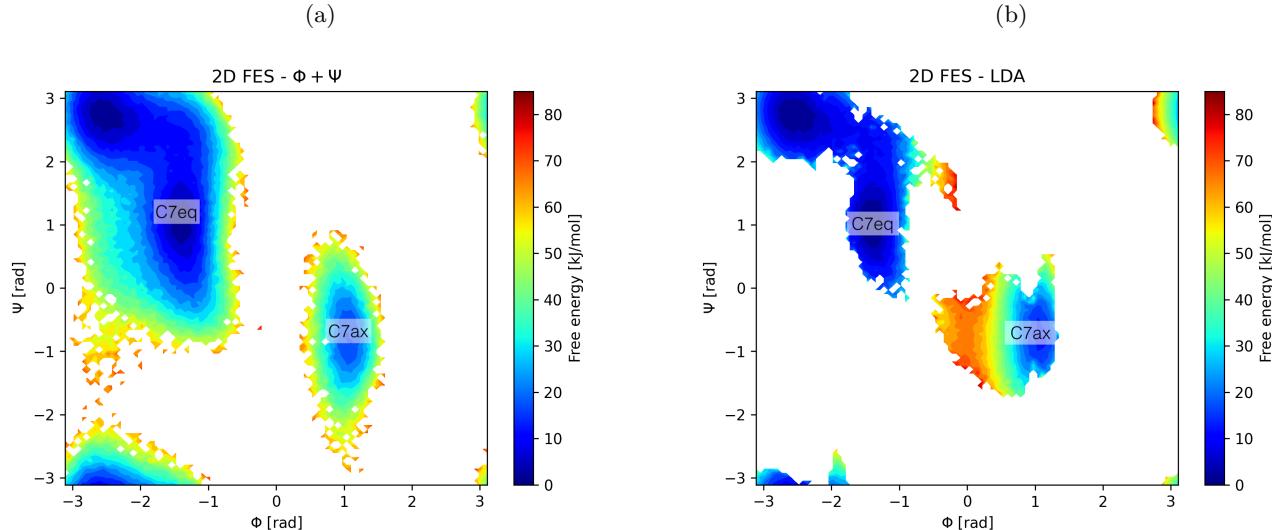


FIG. 9: 2D free energy surface comparison between (a) both angles  $\Phi$  and  $\Psi$  as a cv, (b) LDA CV, from a 20 ns simulation.

### Appendix C: 1D Free Energy Surface Comparison for $\Phi$

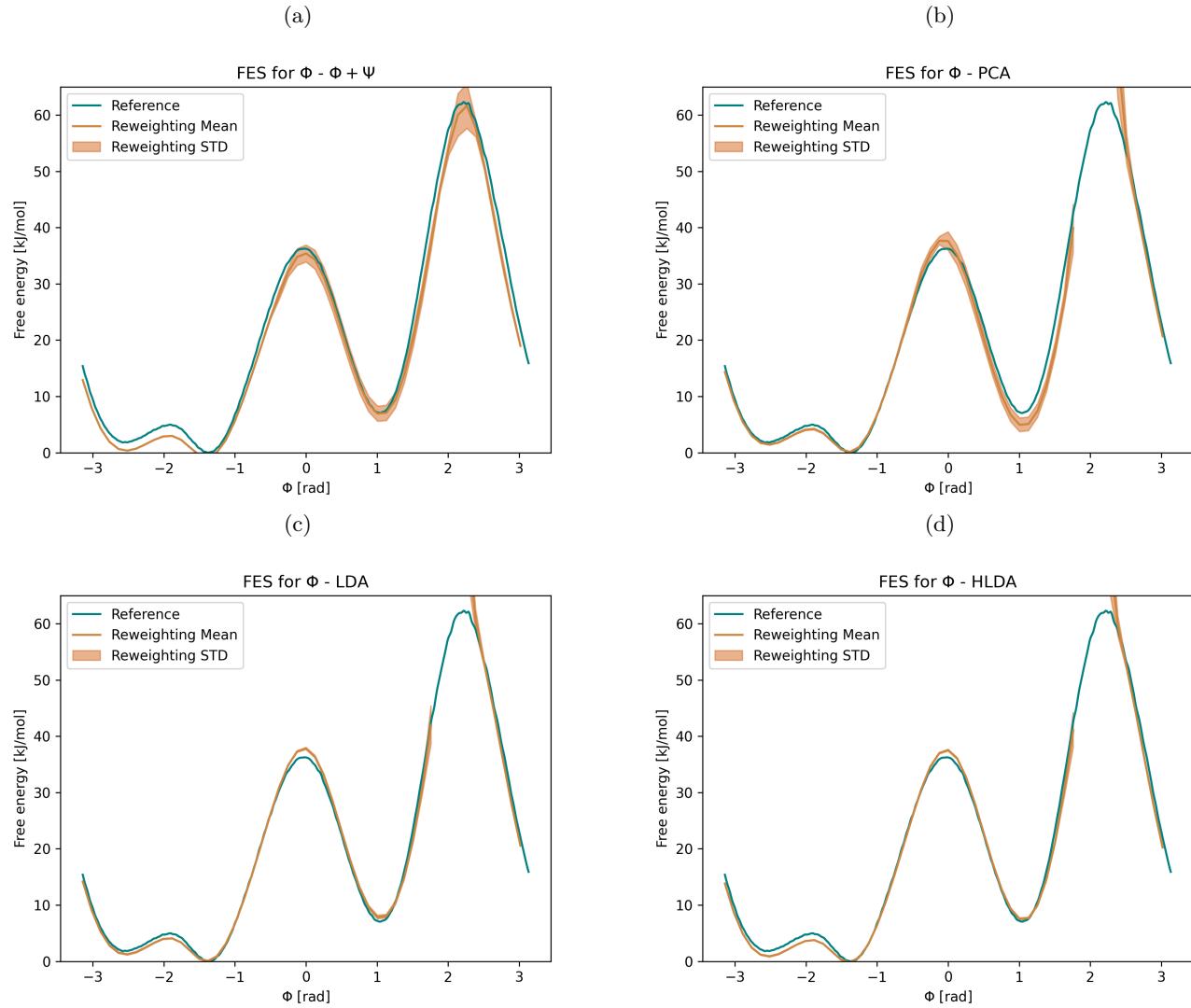


FIG. 10: 1D free energy surface for the angle  $\Phi$ , comparison between four different CV. The reference in each plot is the FES as extracted from a 200 ns simulation with  $\Phi$  as a bias. The re-weighting functions were calculated from EQ. 2