

# Umbrella Sampling Tutorial

Practical aspects and application to Alanine di-peptide

Naushad AL Velgy

January 26, 2018

# Outline

- 1 Introduction
- 2 Summary of Theory
- 3 Test System
- 4 Umbrella Sampling in Practice
- 5 WHAM and 1D PMFs

# Outline

1 Introduction

2 Summary of Theory

3 Test System

4 Umbrella Sampling in Practice

5 WHAM and 1D PMFs

# Prerequisites

## GROMACS

- Some understanding of GROMACS is required.
- A working version of GROMACS patched with Plumed (GMX5.1.2 with PLUMED2 in this tutorial)

## Alanine di-peptide topology

- Provided (with CHARMM27), but can be recompiled with different forcefields if needed

## WHAM

- A working version of WHAM, written by Alan Grossfield (version 2.0.9 used here)
- <http://membrane.urmc.rochester.edu/content/wham>

# Goals

## Goals

- Perform Umbrella sampling on  $\phi$  dihedral of alanine dipeptide
- Use output and WHAM to calculate PMF
- Perform Umbrella sampling on both  $\phi$  and  $\psi$  dihedrals
- Use output and WHAM-2D to plot free energy surface

## Inspiration

- This tutorial is modelled after the Belfast Tutorial, and some files and images are from the website.
- <https://plumed.github.io/doc-v2.4/user-doc/html/belfast-4.html>
- Credit will be given to each file and image used this way

# Outline

- 1 Introduction
- 2 Summary of Theory
- 3 Test System
- 4 Umbrella Sampling in Practice
- 5 WHAM and 1D PMFs

# Enhanced Sampling Methods

## Brief theoretical background

- It is possible that a given system has several minima, difficult to capture *in crystallo*.
- These might be separated by regions of high energetic barriers, and while one might want to sample the transitions between the minima, it might not be possible using unbiased MD.
- Not knowing the location of these minima can be an issue, particularly when trying to extrapolate meaningful free energy surfaces.

# Enhanced Sampling Methods

## Collective Variables

- Systems often contain too many atoms to describe the free energy surface with.
- We can condense the system into a set of descriptors to facilitate any conformational transitions.
- These descriptors, known as 'Collective Variables' (CVs), can be anything from distances between 2 atoms, to dihedral angle changes of particular residues.
- While any measurement can be used, it is important to use a descriptor that reflects the slowest change in the system.

# Enhanced Sampling Methods

Which enhanced sampling method to use?

- PLUMED offers multiple enhanced sampling methods to explore the free energy surface of a system.
- These include:
  - ▶ Steered MD
  - ▶ Metadynamics
  - ▶ Replica Exchange MD
  - ▶ Umbrella Sampling
- Using one method over another depends on the system and the collective variable being used. In this tutorial, we will be using Umbrella Sampling.

# System Potential

"The greatest barrier to someone achieving their potential is their denial of it." - S Travaglia

- Let's say you have an idea of where the major energetic barrier of your system is.
- You can try to add energy to your system to cross that barrier and measure the energy required to do so.

## Bias added to the system

$$V(s) = \frac{\kappa}{2}(s - s_0)^2 \quad (1)$$

- We refer to this as 'opening an umbrella'.
- Larger values of  $\kappa$  will encourage the system to explore regions close to  $s_0$ 
  - ▶ Too large, and the system won't deviate much from  $s_0$ ;
  - ▶ Too small, and the system might not be able to cross the barrier

- Now, you can open an umbrella at specific values of the collective variable  $s$ .
- It is important to allow sequential umbrellas to overlap while fluctuating around certain central values.
- Let's see an example in motion!

# Outline

1 Introduction

2 Summary of Theory

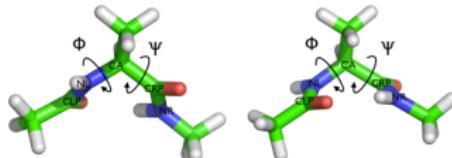
3 Test System

4 Umbrella Sampling in Practice

5 WHAM and 1D PMFs

# Alanine Dipeptide

- Our model system is the widely used alanine dipeptide.
- We will use PLUMED to find the free energy profile of di-alanine while varying both  $\phi$  and  $\psi$  dihedral angles.
- The image below show the dihedrals in questions (adapted from <https://plumed.github.io/doc-v2.4/user-doc/html/belfast-2.html>)



# Discretisation of Dihedral angles

- When performing umbrella sampling calculations, you must discretise the collective variable.
- How discretised the system becomes in our analysis can be critical in the convergence of calculations.
- If one has too few umbrellas, the free energy will not converge; too many, one would waste computational resources.

# Outline

- 1 Introduction
- 2 Summary of Theory
- 3 Test System
- 4 Umbrella Sampling in Practice
- 5 WHAM and 1D PMFs

# Towards a 1D PMF

- In this tutorial, we'll use a variety of umbrellas to demonstrate the effect of appropriate discretisation
- Then, using those results, we'll get a 2D free energy profile.
- Our CV changes from  $-3\pi\text{rad}$  to  $3\pi\text{rad}$  (-180 to 180 degrees, as it is an angle), we will create restraints at discretised points along the CV.

# Basic PLUMED input

We need to generate a plumed.dat file as input for GROMACS. The file we will be generating looks like:

## plumed.dat file

```
phi: TORSION ATOMS=5,7,9,15
```

```
psi: TORSION ATOMS=7,9,15,17
```

```
restraint-phi: RESTRAINT ARG=phi KAPPA=50.0 AT=-3.0
```

```
PRINT ARG=phi,psi,restraint-phi.bias FILE=COLVAR_-3.0
```

# Parameter explanations

- The 'phi' and 'psi' lines tell plumed to define 'psi' and 'psi' as dihedrals (TORSION) using the ATOMS indexed.
- 'restraint-phi' tells plumed to apply a RESTRAINT on phi (ARG), using a harmonic potential (KAPPA) of 50kJ/mol, at the value specified in AT.
- Then, plumed will PRINT 'phi', 'psi', and 'restraint-phi.bias' (ARG) into a FILE called 'COLVAR\_-3.0'.

# Your first plumed run

When your plumed.dat file is ready, run

## GROMACS run

```
$ mdrun -s topol.tpr -x traj.xtc -plumed plumed.dat -nsteps 250000
```

After this run (which shouldn't take more than a few seconds!), you will have at least two new files: 'traj.xtc' and 'COLVAR\_-3.0'. Keep both, as both are very important.

## More frames

- So that we get a better idea of the energetics of changing the  $\phi$  dihedral, we should run a restraint on every  $\phi$  value between -3 and 3.
- We can automate the process of generating a plumed.dat file and running the GROMACS simulation using a simple Bash script (see an example in umbrellas.sh)

## Example Bulk MD run

```
for AT in $(seq -3 1 3): # changes the value of AT from -3 to 3 in steps  
of 1  
do  
cat > plumed.dat << EOF  
phi: TORSION ATOMS=5,7,9,15  
psi: TORSION ATOMS=7,9,15,17  
# changes 'AT' in the following 2 lines  
restraint-phi: RESTRAINT ARG=phi KAPPA=100.0 AT=$AT  
PRINT ARG=phi,psi,restraint-phi.bias FILE=COLVAR_$AT  
EOF  
# And finally changes 'AT' here  
mdrun -s topol.tpr -plumed plumed.dat -nsteps 250000 -x traj_$AT.xtc  
done
```

# Output

Now you should have a variety of new files:

- 7 'traj.xtc' files, each with a number on them (e.g. 'traj\_-3.0.xtc')
- 7 'COLVAR' file, each with a number on them (e.g. 'COLVAR\_-3.0')

Copy these files into a new directory, so we can do analysis on them without losing the originals (though they should not be hard to re-generate).

# Preparation of files for WHAM

- In reality, we only really need the change in  $\phi$ ; even the time column is irrelevant.
- We can use some bash magic to pass only the relevant information to WHAM.

# Outline

- 1 Introduction
- 2 Summary of Theory
- 3 Test System
- 4 Umbrella Sampling in Practice
- 5 WHAM and 1D PMFs

# Towards a 1D PMF

Contrary to the remainder of the Belfast tutorial, we will now be using WHAM2.0.9, developed by Alan Grossfield (<http://membrane.urmc.rochester.edu/content/wham>).

From the documentation of WHAM, the command line use is:

## WHAM usage

```
$ wham [P|Ppi|Pval] hist min hist max num bins tol temperature  
numpad metadatafile freefile
```

# Explanation of WHAM arguments I

It is very important to understand these individual input arguments, so we get the correct PMF out.

- 'P|Ppi|Pval' refers to the periodicity of our CV, and it is optional
  - ▶ 'P' tells WHAM the CV has a periodicity of 360, used for angles
  - ▶ 'Ppi' tells WHAM the CV has a periodicity of  $2\pi$ , for angles measured in radians
  - ▶ 'Pval' allows you to specify the periodicity. For example, 'P180.0' tells WHAM the CV has a periodicity of 180 degrees.
  - ▶ Leaving this argument empty tells WHAM there is no periodicity in the system (for distances and the like).
- Since our CV is both an angle and is measured in radians, we will use 'Ppi' during the input.

## Explanation of WHAM arguments II

- 'hist min' and 'hist max' are the boundaries of our PMF. It's very simple in our case, since our dihedral starts at -3 and ends at 3. These should also be floating point numbers (i.e. -3.0 and 3.0).
- 'num bins' specifies the number of bins. The result doesn't seem to change with the number of bins, so just set it to the number of windows. It should be an integer. However, we will plot this PMF while changing the number of bins, just so you see the effect.
- 'tol' is the convergence for the WHAM calculations. See the documentation for more, but set it to '0.001'.
- 'temperature' should be a floating point number detailing the temperature in Kelvin. It doesn't have to be the same one used for the simulation. Set it to '303.15'.

# Explanation of WHAM arguments III

- 'numpad' should be set to 0 for aperiodic CVs. Set it to '1'.
- 'freefile' is the name of the output file. Set it to 'pmf\_7\_windows'.
- 'metadatafile' is the path to the metadata file. The format of this file is explained in the next slide.

# Metadatafile file format

- The metadatafile is a file with at least 3 columns (for a 1D PMF).
- The first column is the path to the timeseries file (generating these files is not complicated; see `truncate_colvars.sh` for an examples).
- The second column should specify the AT value for that timeseries.
- The third column should specify the KAPPA value used, **divided by 2**.
  - ▶ This is very important, as WHAM does not account for this scaling (see equation 1).

# Timeseries file

- Each timeseries file should consist of only 2 columns: a column of time values and one with the change in CV.
- In reality, WHAM only needs the 2nd column, so it doesn't particularly matter what you have on the first column (provided, of course, they are floating point numbers).
- `truncate_colvars.sh` will take the first and second columns of 'COLVAR' files and remove the headers (which start with a '#') and the last two columns. The output of `truncate_colvars.sh` is the name of the colvar file with '\_trunc' at the end. Feel free to modify it as you wish.

# Metadata file

- Now that we have a set of timeseries files, all we need to do is generate a metadatafile.
- gen\_metadata.sh is provided for that purpose. Feel free to change any relevant parts of the script.
- Make sure that the 'spring' is appropriately changed.

# Running WHAM

Now that you have a metadatafile, it's time to run WHAM!

## Run WHAM

```
$ wham Ppi -3.0 3.0 $(ls *trunc | wc -w) 0.001 303.15 1 metadata  
pmf_7_windows
```

And plot 'pmf\_7\_windows' using your favourite plotting software!

# Plot of 7 windows

The simplest way to plot is using gnuplot.

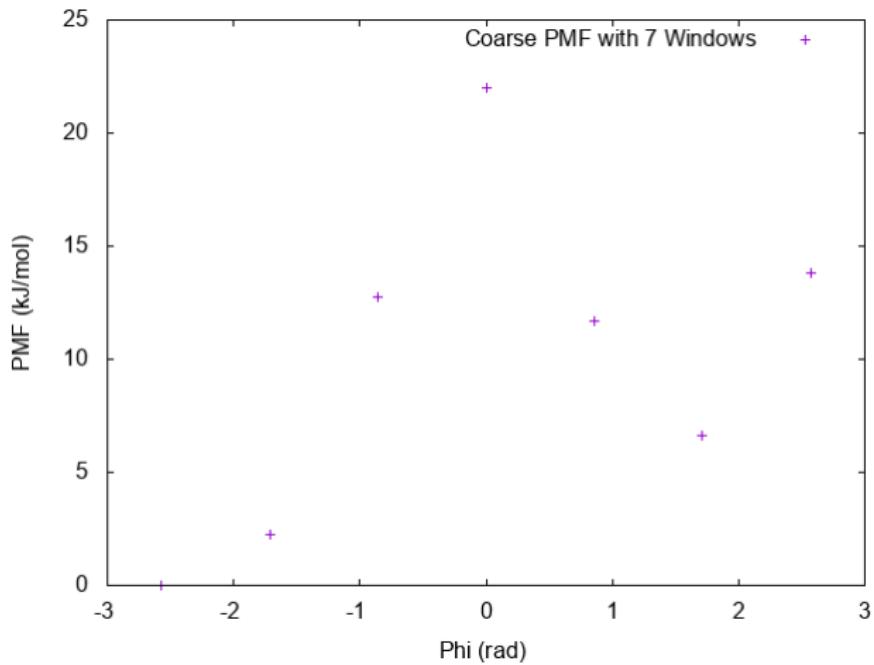
In gnuplot, type

```
> plot 'pmf_7_windows'
```

and a plot should come up almost immediately.

# A coarse PMF

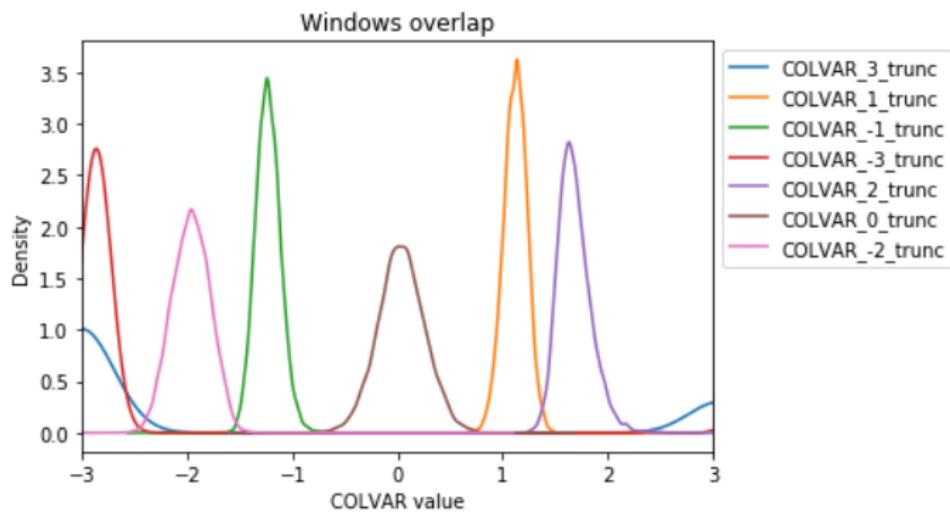
This should be the outplot of gnuplot (you'll have to set the labels yourself):



# Windows overlap

Now we should check whether the windows overlap during the simulations.

You can check using a simple python script (see the example set in `overlap.py`).

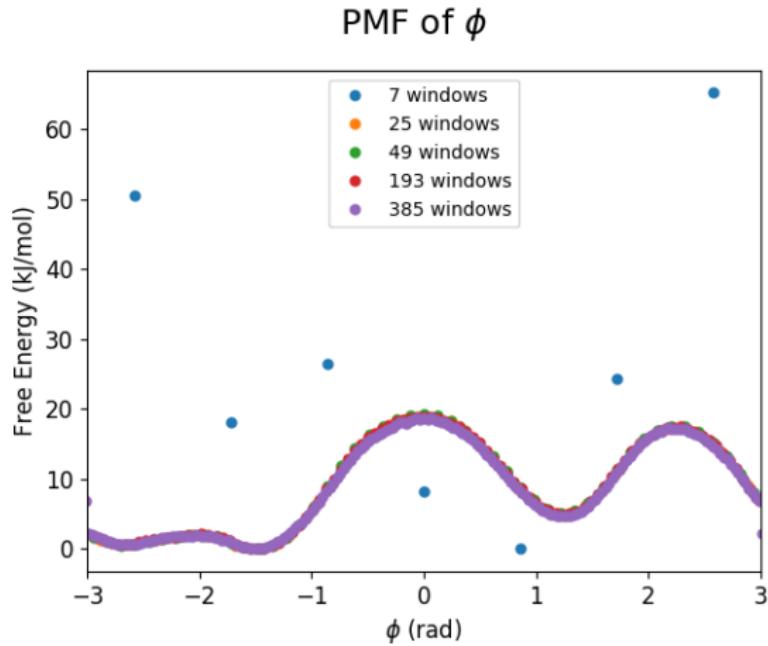


## A finer PMF

- The PMF plot, while giving us some information, is lacking
- It is clear from the density plot that the umbrellas do not overlap
- Spend the next moments getting a similar profile using more windows (use a 100ps simulation time for each window).
  - ▶ Start by changing the step between windows to 0.25 (i.e. seq -3 0.25 3)
  - ▶ Then a step of 0.125
  - ▶ And so on to a step size of 1/64
- Before running multiple simulations, run 'export GMX\_MAXBACKUP=-1' on the terminal. This prevents GROMACS from making backups, and thus crashing from making too many of them.

# Better PMFs

After running all the simulations, plot the PMF using any one gnuplot, matplotlib, matlab, octave... It should look a little something like this:



# Towards a 2D FE surface

- Now that we have an idea of the changes in free energy of one dihedral, why not do the same for both?
- We can start by setting a grid of simulations where we vary  $\phi$  and  $\psi$ . Maybe something that looks like this:

# $\phi$ and $\psi$ grid

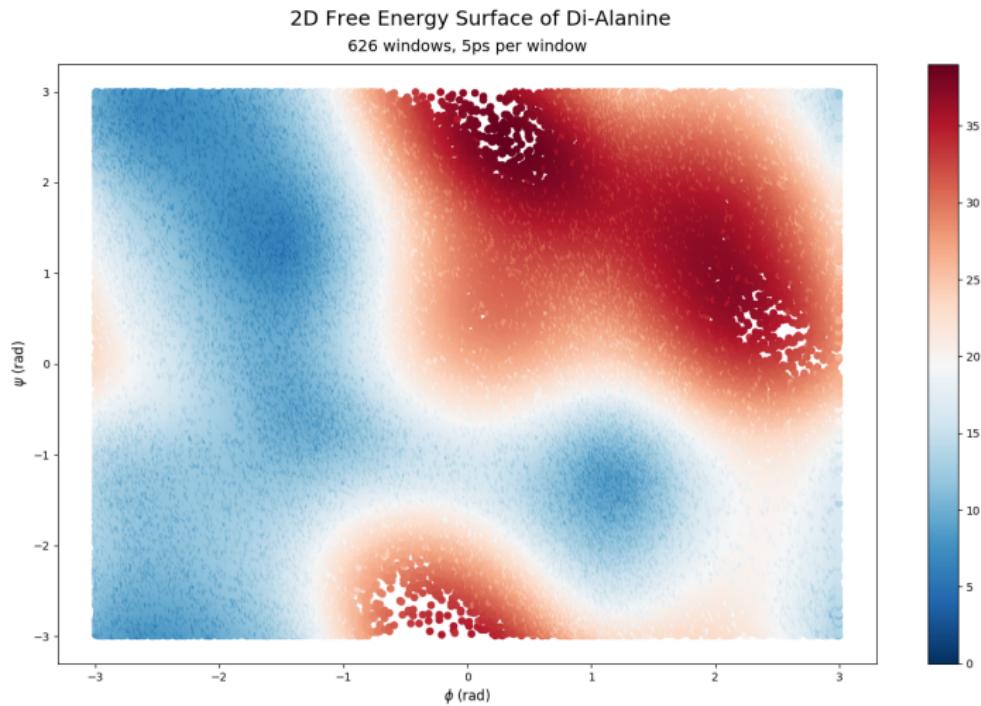
$\phi$ psi	-3	-2.75	-2.5	...	2.5	2.75	3
-3	5ps	5ps	5ps	...	5ps	5ps	5ps
-2.75	5ps	5ps	5ps	...	5ps	5ps	5ps
-2.5	5ps	5ps	5ps	...	5ps	5ps	5ps
...	...	...	...	...	...	...	...
2.5	5ps	5ps	5ps	...	5ps	5ps	5ps
2.75	5ps	5ps	5ps	...	5ps	5ps	5ps
3	5ps	5ps	5ps	...	5ps	5ps	5ps

Run these simulations using the same scripts as before, again making sure you run 'export GMX\_MAXBACKUP=-1' in the terminal.

## 2D WHAM

- After running the simulations, you should have a variety of small 'traj.xtc' and 'COLVAR' files, hopefully each indicating the 'AT' position of each dihedral (e.g. 'traj\_-3\_-3.xtc' and 'COLVAR\_-3\_-3' for a simulation with both  $\phi$  and  $\psi$  at -3 ( $\pi$ rad)).
- After that, isolate the 'time', 'phi', and 'psi' columns of each 'COLVAR' file, which we will feed into the wham-2d executable.
- Remove from the resulting 'freefile' the 4th column (as you won't need it), and every row for which the 3rd column has a value of '9999999.000000'. This file is now ready to plot!

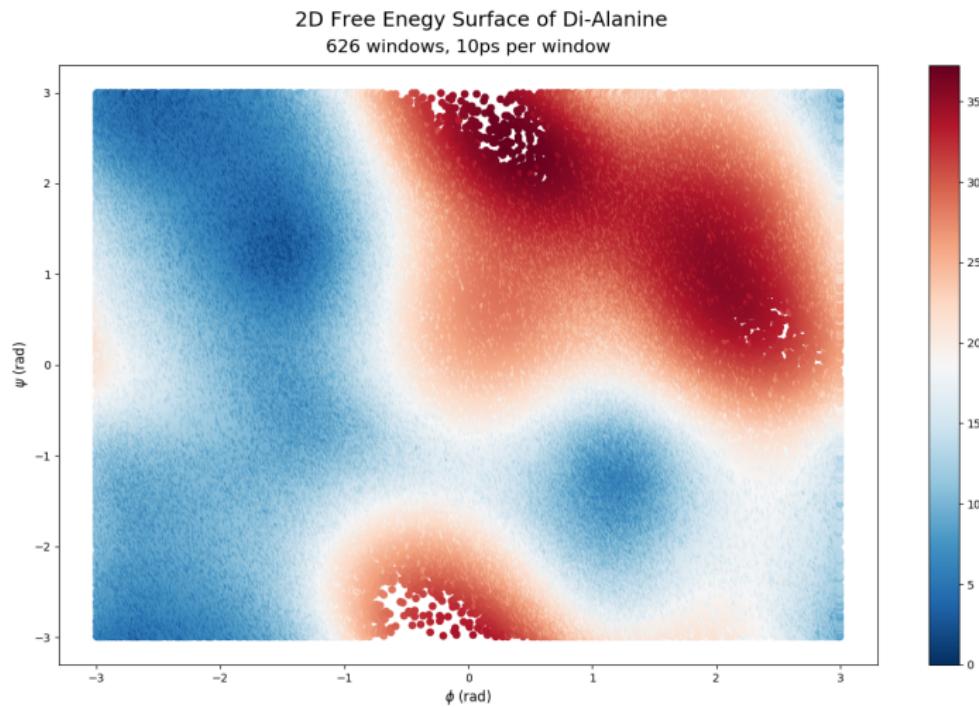
# A 2D surface



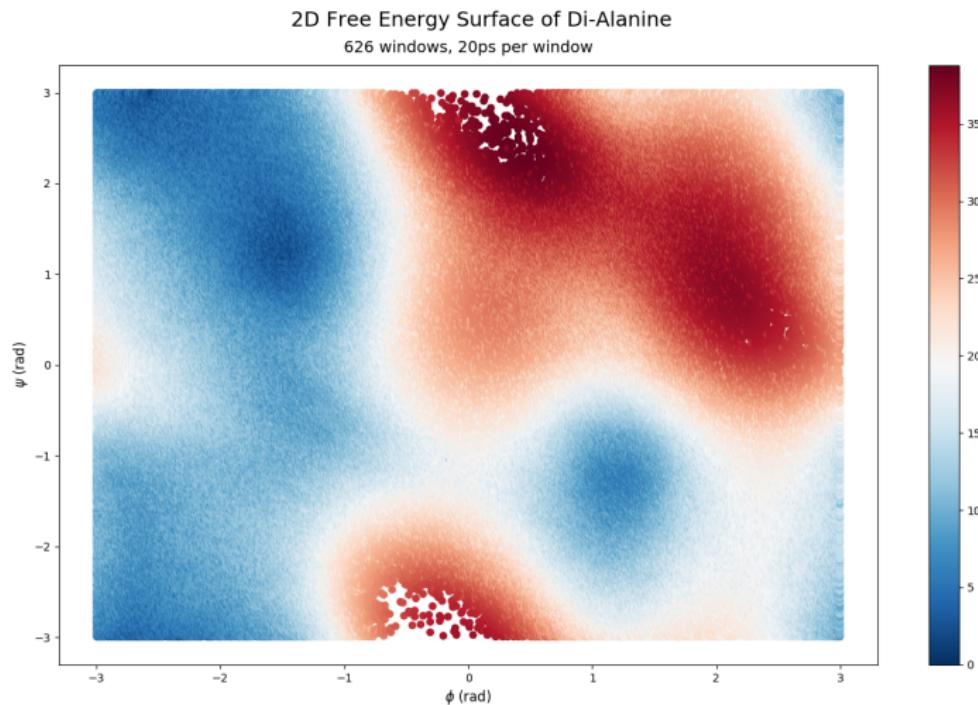
# Better 2D surfaces?

- So, what happens when you increase the sampling tile of each simulation?
- Or the number of windows?
- The following 3 plots are the result of increasing the simulation time of each window to 10ps, 20ps, and 50ps, respectively.

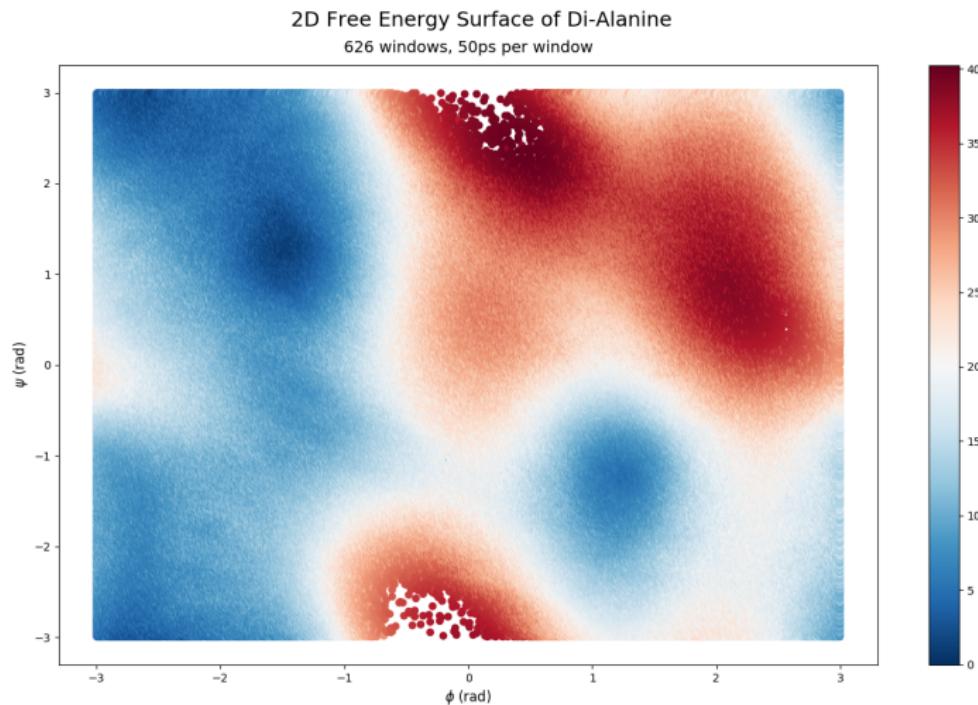
# 10ps Windows



# 20ps Windows



# 50ps Windows



# Precautions & further comments

- Measuring dihedrals is relatively simple; other collective variables may not be as simple!
- In cases where your collective variable represents a large conformational change, it might be worth performing some steered MD first
  - ▶ This way, you can get any number of windows you need.
  - ▶ Restarting the simulation from a frame of your steered MD run might not be an issue.
- Getting a 2D free energy surface might not be necessary to your research question.

Thank you for reading this tutorial. If you have any questions or comments, please feel free to email me at [naushad.velgy@dtc.ox.ac.uk](mailto:naushad.velgy@dtc.ox.ac.uk)