Umbrella sampling in GROMACS

BioExcel Riga Workshop 2018

Paul Bauer <paul.bauer@scilifelab.se>

SciLifeLab

KTH

26/10/2018

Outline

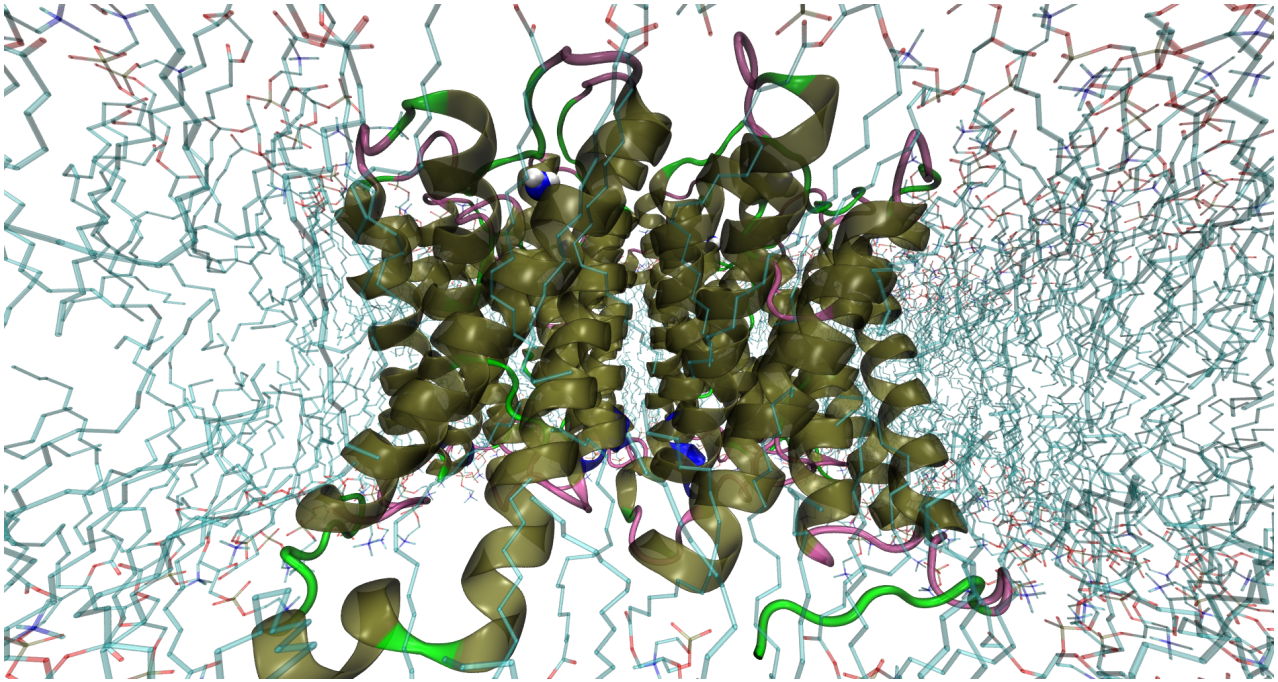
1. Basics of Umbrella Sampling (US)
2. Umbrella sampling in GROMACS
3. Data analysis with gmx wham

Sampling problems in Molecular Dynamics

Basics of Umbrella Sampling (US)

Umbrella sampling in GROMACS

Data analysis with gmx wham



Example problem being studied by MD: Transport of a solute through a membrane channel.

Covering of complete phase space is unrealistic

Basics of Umbrella Sampling (US)

Umbrella sampling in GROMACS

Data analysis with gmx wham

- Systems explore phase space according to boundary conditions given by the ensemble
- Larger differences in energy between states make crossing less likely
- Interesting changes often involve large changes in free energy and large free energy barriers
- System needs to be forced to either cross barriers or ignore them

Biasing potentials help crossing of free energy barriers

Basics of Umbrella Sampling (US)

Umbrella sampling in GROMACS

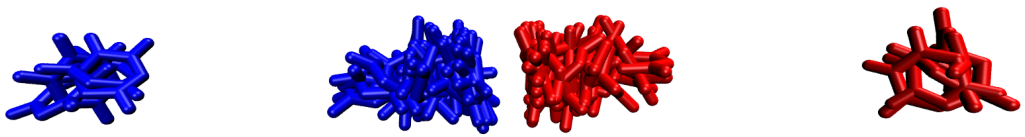
Data analysis with gmx wham

Basic math behind use of Umbrella Sampling

$$\begin{aligned} U_{\text{window}}(\mathbf{r}) &= U(\mathbf{r}) + W(\mathbf{r}_{\text{rest}}) \\ W(\mathbf{r}_{\text{rest}}) &= k(\mathbf{r} - \mathbf{r}_{ref})^2 \end{aligned}$$

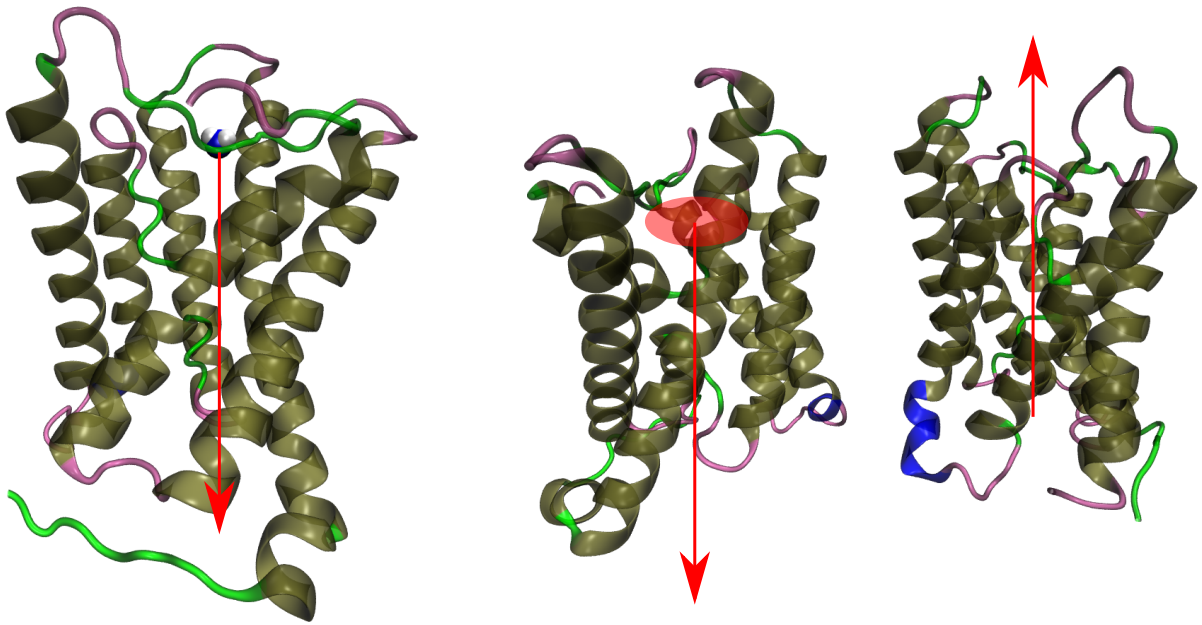
System is moved through several windows with different references coordinates.

Example visualization of US windows



Moving two pyrimidine molecules through different US windows.

Ways to define US window coordinates



Different ways to define US window coordinates.

Minimum requirements for converging simulation

- Each window needs to represent physical configuration
- Windows need to have sufficient sampling overlap
- No large changes between configurations that can't be sampled

How to generate initial configurations?

- Manual placement of initial configurations
- Molecular dynamics with constant potential
- Simulation with restraints
- QM calculations

Some application examples

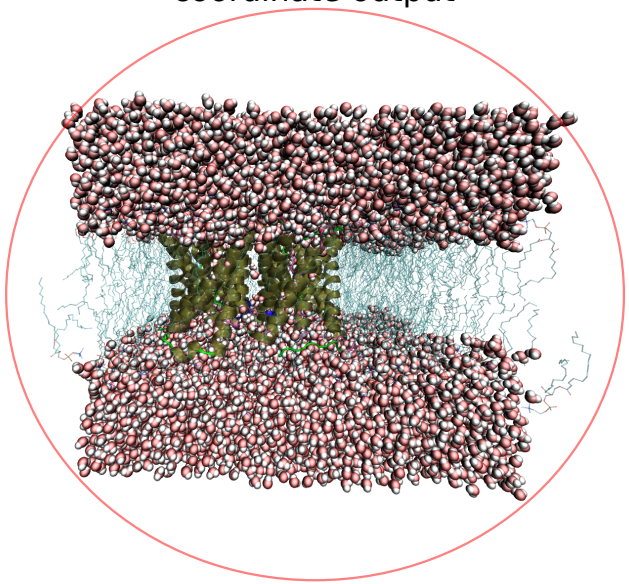
- Calculation of dimerization energies
- Clearly defined change within simulated structure
- QM/MM for chemical reactions in enzyme systems
- What can you think of?

US basics and requirements

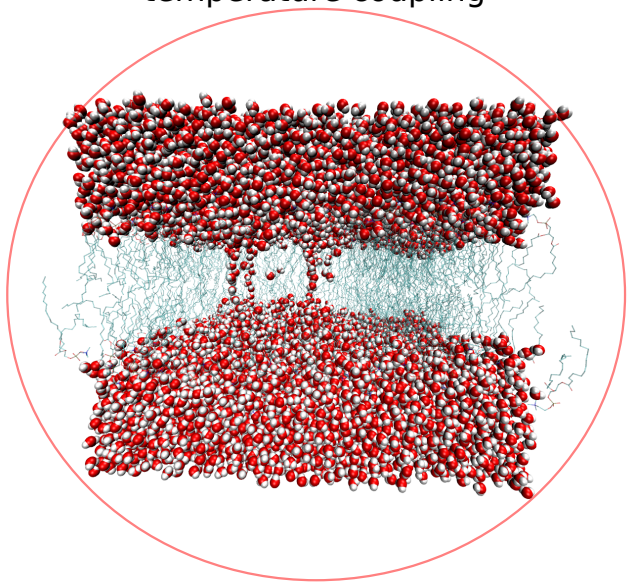
- All here based on reasonable recent version of GROMACS, tried on v2018
- Examples and later tutorials based on `https://barnett.science/tutorials/5_umbrella/`
- Need to understand pull code, index groups, (partially) WHAM
- Topology setup and definition not covered

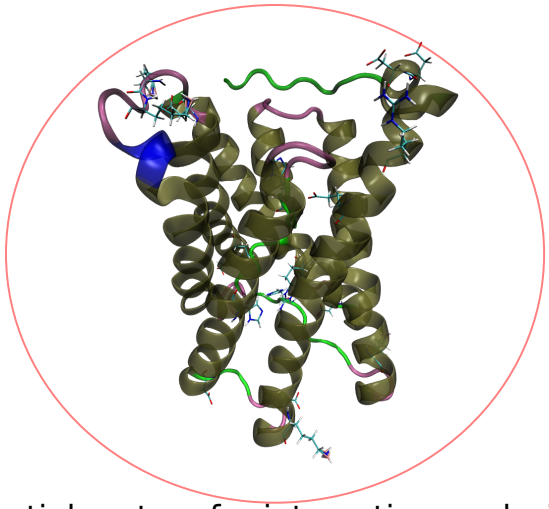
Introduction to index groups

Complete system for coordinate output

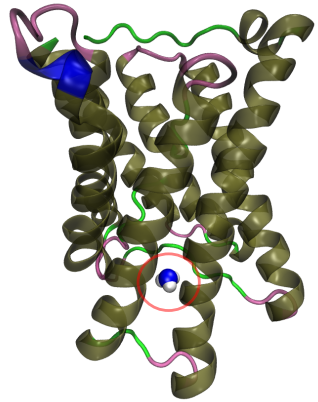


Partial system for e.g. temperature coupling





Partial system for interaction analysis



Single molecule for e.g. Umbrella sampling

Index group basics

- List of atoms being grouped together according to some requirement
- Used to specify important parts of simulation for running/analysis/data processing
- Defaults generate by all GROMACS tools internally
- User defined groups available through either gmx make_ndx or gmx select

Basics of Umbrella Sampling (US)
Umbrella sampling in GROMACS
Data analysis with gmx wham

Index group generation demo with gmx make_ndx

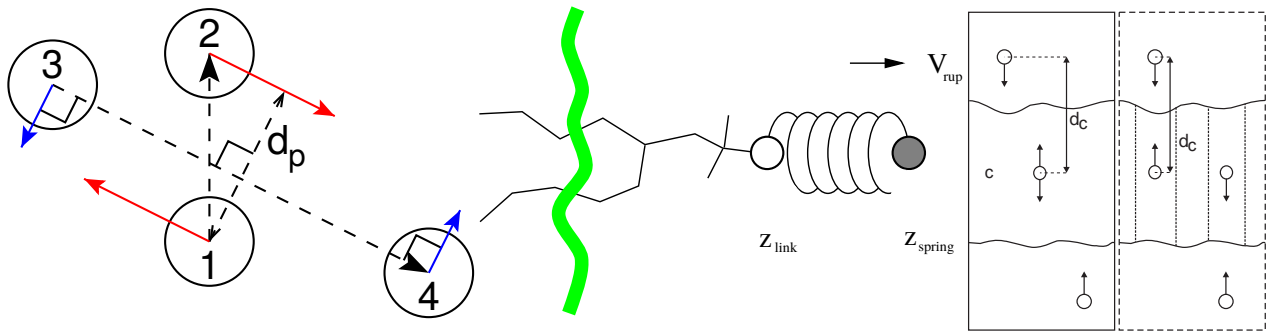
Demo time!

Setting up the simulation

- Define groups that should be restrainted by US windows
- Define coordinates to be used for restraint
- Set window spacing or generate initial configurations
- Set US window restraint force constant

Different kinds of coordinates and geometries

- Linear distance
- Directional or based on reference
- Cylinder pulling



Examples for pull or US coordinates in GROMACS from reference manual

Basics of US histogram analysis

- All information here based on `http://doi.org/10.1002/jcc.540130812`
- Full disclosure: I'm not an expert in WHAM
- More intendet as opening point to look for further information

Math behind WHAM

$$P_{\{\lambda\}j,\beta_j}(\{V\},\zeta) = \frac{\sum_{k=1}^{R} N_k(\{V\},\zeta)\exp\left(-\beta\sum_{j=0}^{L}\lambda_j V_j\right)}{\sum_{m=1}^{R} n_m \exp\left(f_m - \beta_m\sum_{j=0}^{L}\lambda_{j,m}V_j\right)}$$

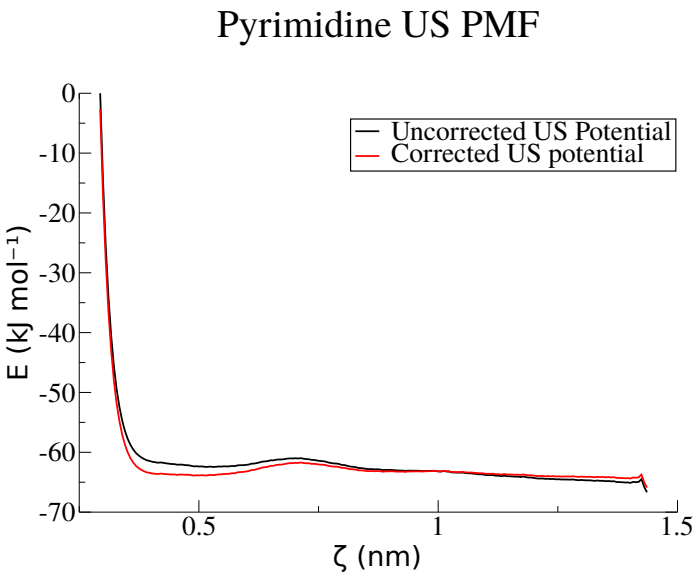
$$\exp(-f_i) = \sum_{\{V\},\zeta} P_{\{\lambda\}j,\beta_j}(\{V\},\zeta)$$

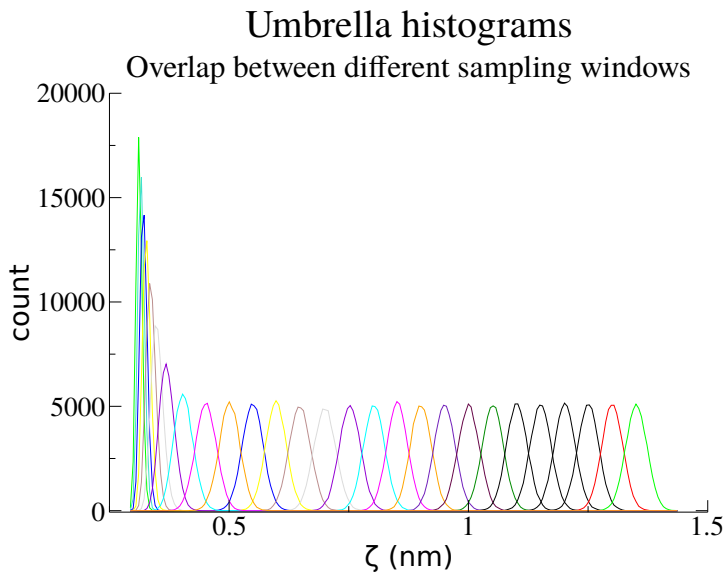
$$\exp(-f_i) = \sum_{k=1}^{R}\sum_{t=1}^{n_k} \frac{\exp\left[-\beta_i\sum_{j=0}^{L}\lambda_{j,i}V_{j,t}^{(k)}\right]}{\sum_{m=1}^{R} n_m \exp\left[f_m - \beta_m\sum_{j=0}^{L}\lambda_{j,m}V_{j,t}^{k}\right]}$$

Analysing GROMACS US simulations

- Needs at least run input file and pullx or pullf file for each window
- Will by default calculate both individual histograms and PMF
- Able to also obtain autocorrelation times and error estimates

Example analysis from simple US simulation





Well coverged and overlapping US simulation

Common pitfalls

- Insufficient sampling overlap
- Bad choice of reaction coordinate
- Bad choice of restraint parameters
- Insufficient equilibration

Thanks for your attention/patience

- Questions?
- Comments?
- Suggestions?