ESCOLA GAÚCHA DE BIOINFORMÁTICA

Molecular Dynamics Simulations:

Theory and Applications

DSc. Conrado Pedebos
DSc. Pablo Ricardo Arantes

Porto Alegre, 14-18 de Julho de 2025

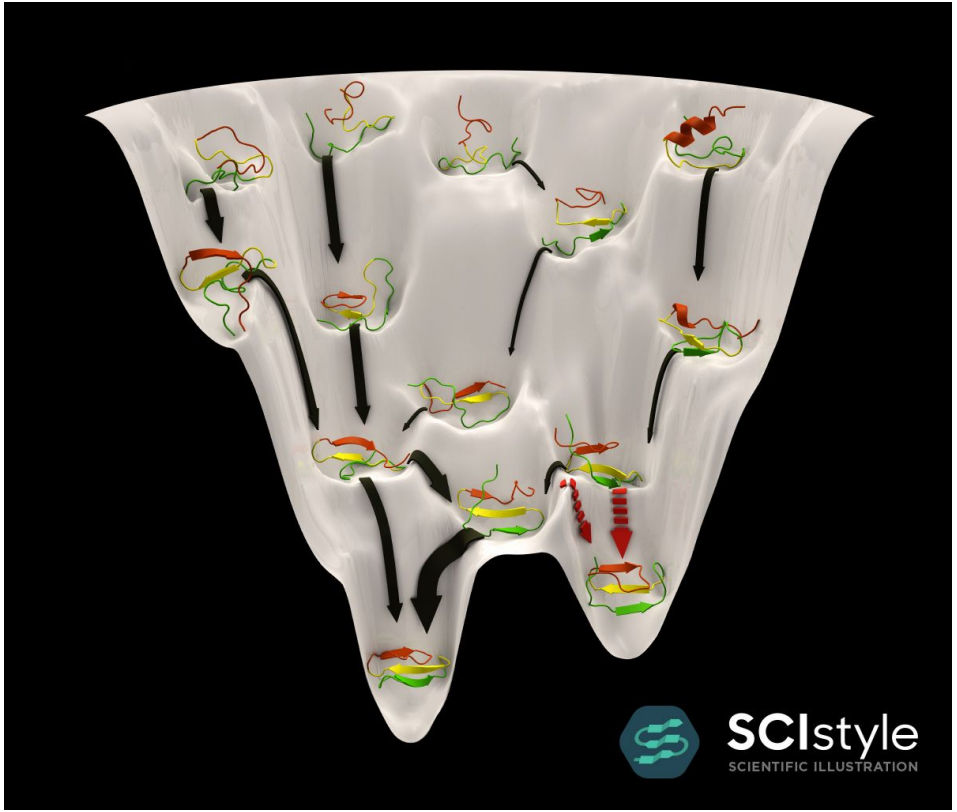
CBiot Centro de Biotecnologia UFRGS · PPGBCM CBiot/UFRGS · InsCer Instituto do Cérebro · UFRGS Universidade Federal do Rio Grande do Sul · UFCSPA · PUCRS · PPGBio UFCSPA · PPGBM Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS

What is Sampling?

Sampling, in a mathematical context, means visiting regions of phase space representative of a statistical ensemble

- In MD, we aim to sample the configuration space over time
  - Key challenge: Can short trajectories represent long-time behavior?
  - Phase space = coordinates + momenta (p) of all atoms

$$p = m*v$$

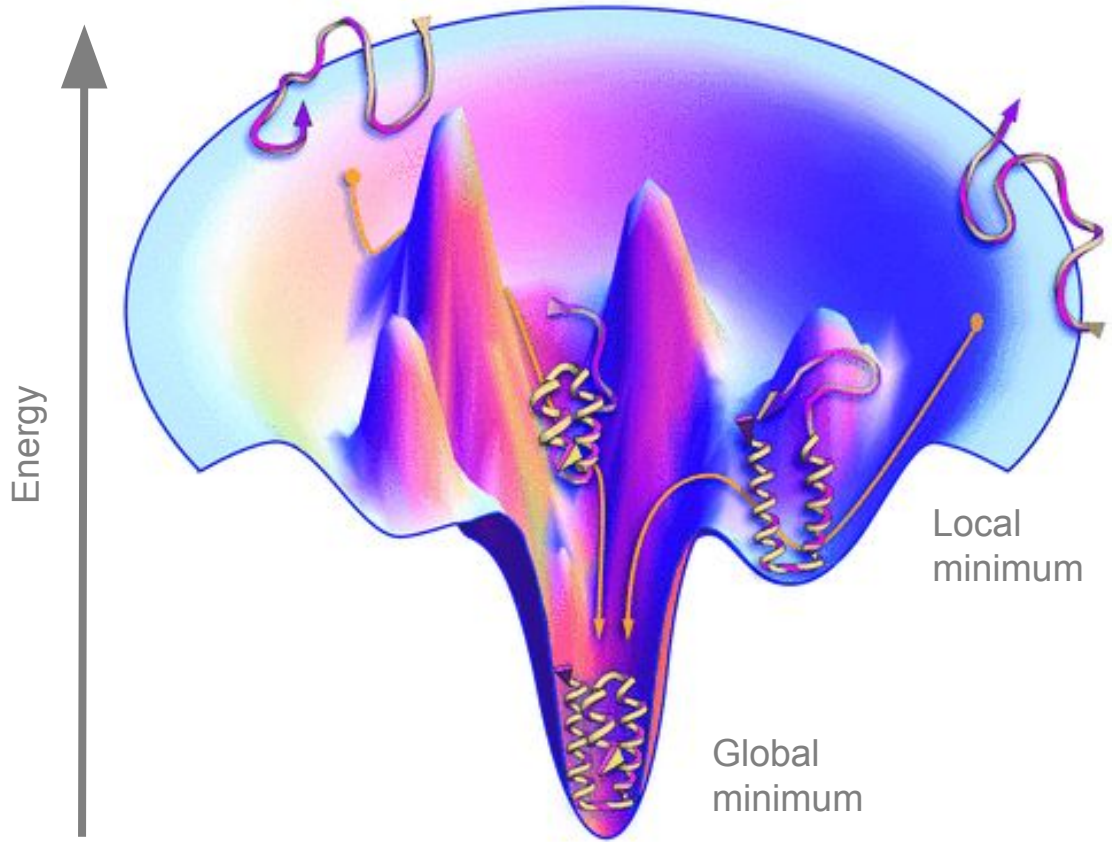


source: https://www.scistyle.com/portfolio/protein-folding-funnel/

A Sampling Problem...

Certain regions of the configurational space have higher probability of being visited more frequently than others

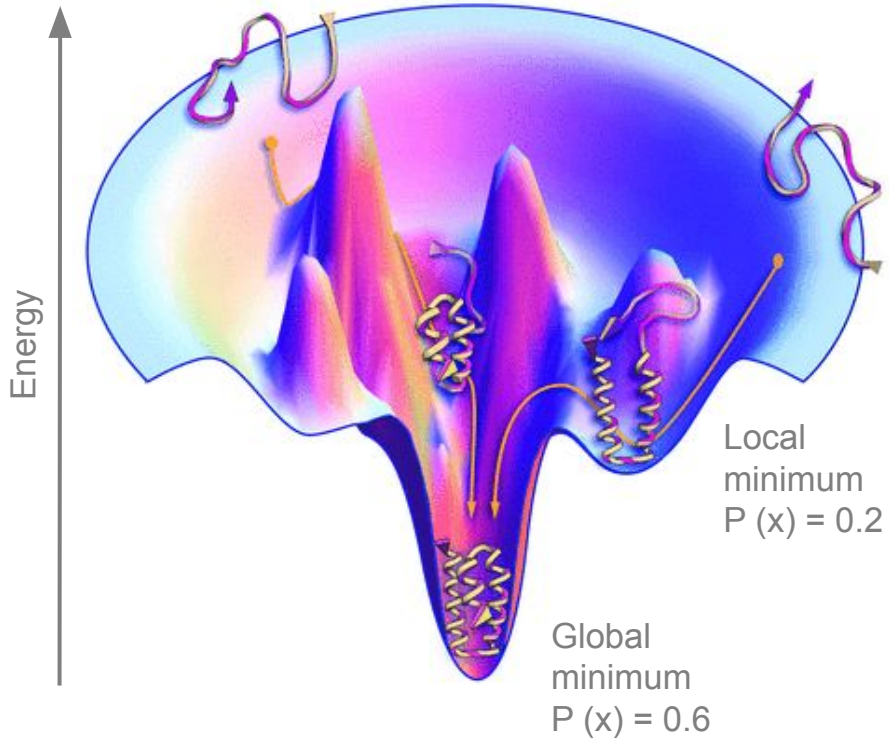
Statistical Mechanics → Boltzmann Distribution



A Sampling Problem...

- **Statistical mechanics**: the link between observable **macroscopic properties** such as temperature and pressure and the **behaviour of individual particles** that make up the system (**microscopic states**).
- Probability distribution = probability that a system will be in a certain state as a function of that state's energy and the temperature of the system
- **Boltzmann distribution**:

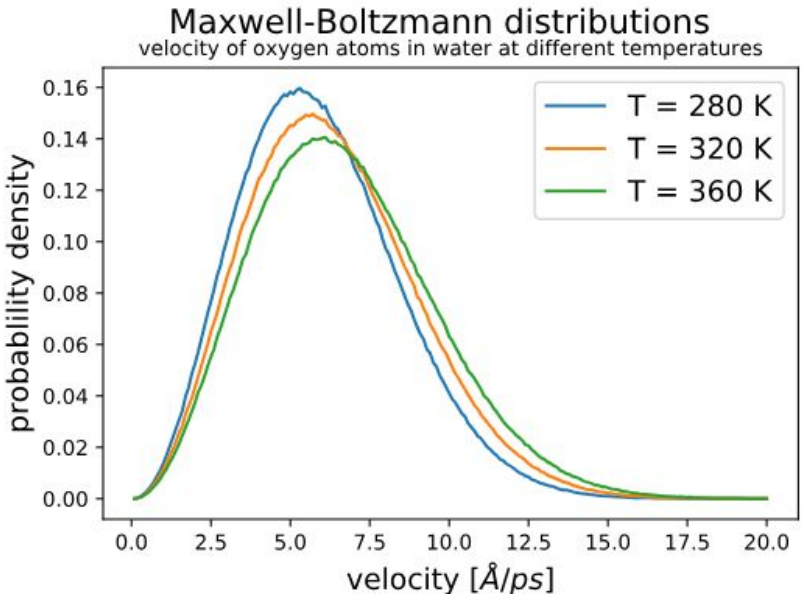
$$P(x) \propto e^{-E(x)/kBT}$$



A Sampling Problem...

- **Statistical mechanics**: the link between observable **macroscopic properties** such as temperature and pressure and the **behaviour of individual particles** that make up the system (**microscopic states**).
- Probability distribution = probability that a system will be in a certain state as a function of that state's energy and the temperature of the system
- **Boltzmann distribution**:

$$P(x) \propto e^{-E(x)/kBT}$$

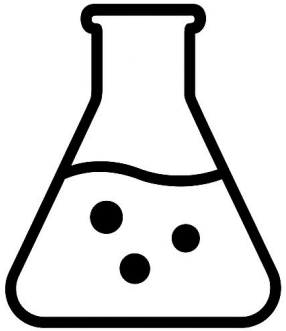


https://computecanada.github.io/molmodsim-md-theory-lesson-novice/07-thermostats/index.html

A Sampling Problem...

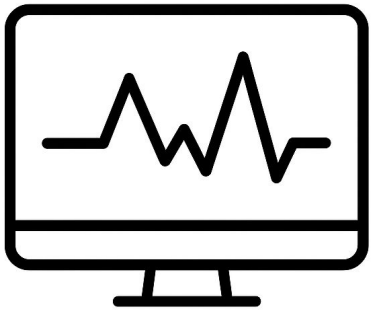
- Over a long time period the **microstates sampled by MD** will match the microsta the statistical ensemble; so we can take average of a property over the time of th trajectory as the true ensemble average (**ergodic hypothesis**)
- **Experimental methods** usually provide measure of **ensemble averages**
- Sampling the right regions of phase space = getting correct probabilities for P(x).

Experiment



Ensemble averages

Simulation



Time averages

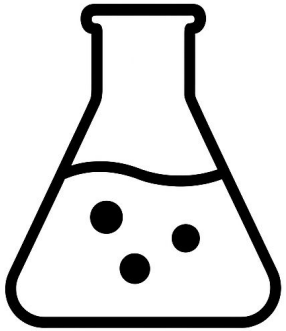
Main ensembles in statistical mechanics:

- Microcanonical: fixed **N, V, E**
- Canonical: fixed **N, V, T**
- Isobaric-isothermic: fixed **N, P, T**

Time Averages vs Ensemble Averages

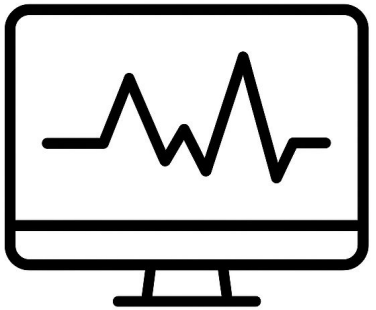
- MD produces time averages:
- $\langle A \rangle_{time} = (1/T) \int A(t)\, dt$
- Ensemble average:
- $\langle A \rangle_{ensemble} = \int A(x)\, P(x)\, dx$
  - Are these equal?
  - Only if system is ergodic: all accessible states visited with correct probability

Experiment



Ensemble averages

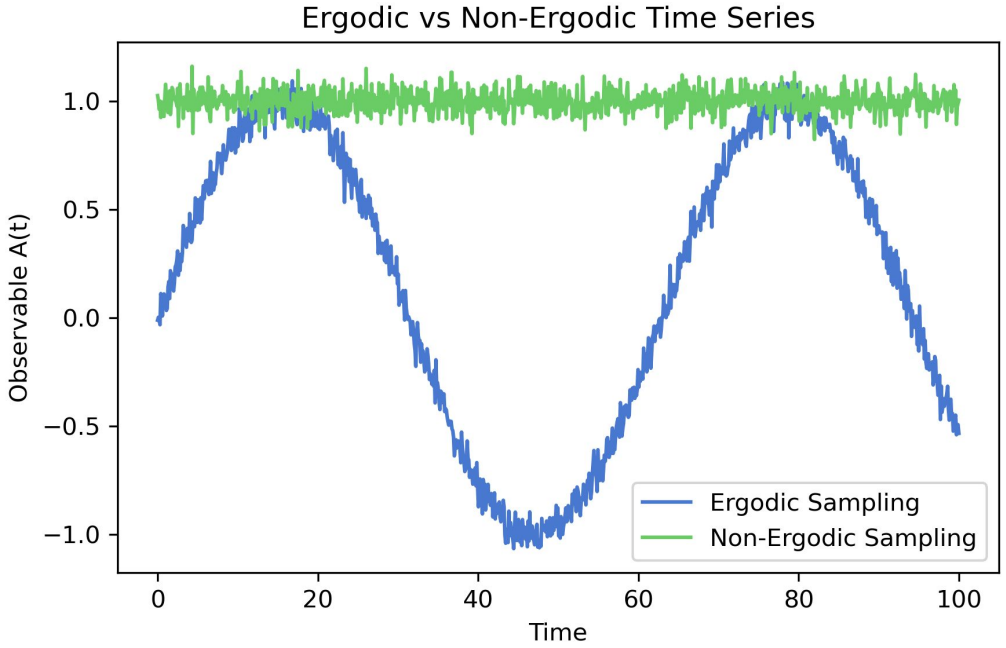
Simulation



Time averages

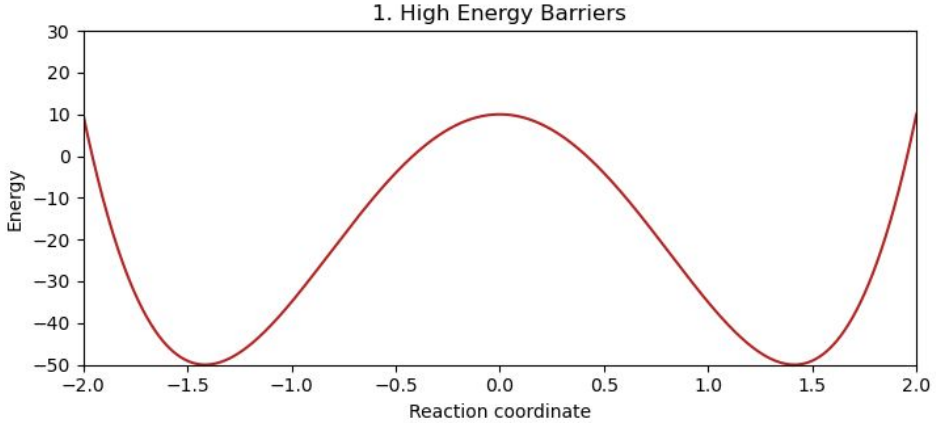
The Ergodic Hypothesis

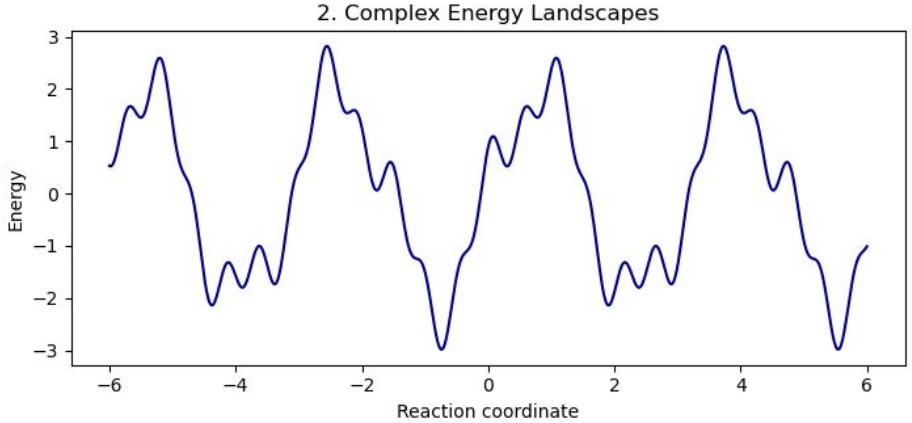
- Assumes equivalence between time and ensemble averages
- Works if simulation explores the full phase space
  - Real systems often exhibit kinetic barriers
  - Non-ergodicity in finite time is common



Why Ergodicity Fails?

- High energy barriers prevent transitions between states
- Complex energy landscapes with multiple minima
- Limited simulation time (ns-μs) vs biological timescales (μs-ms)
- Systems with rare events: binding, folding, allostery

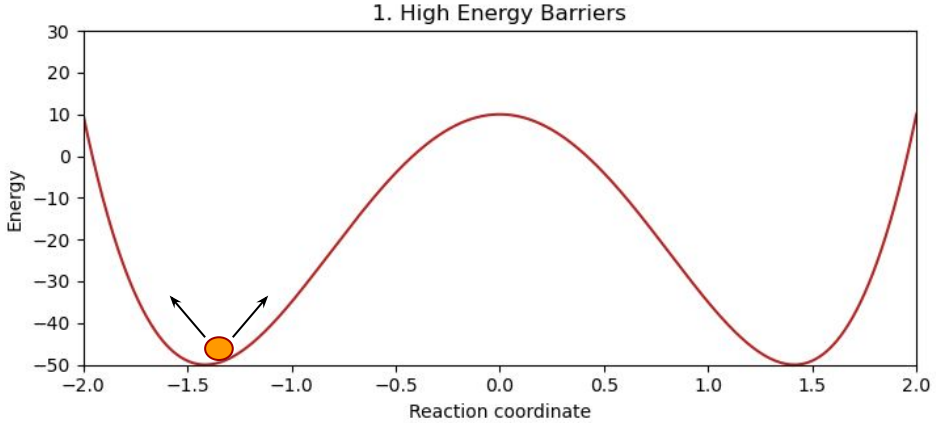




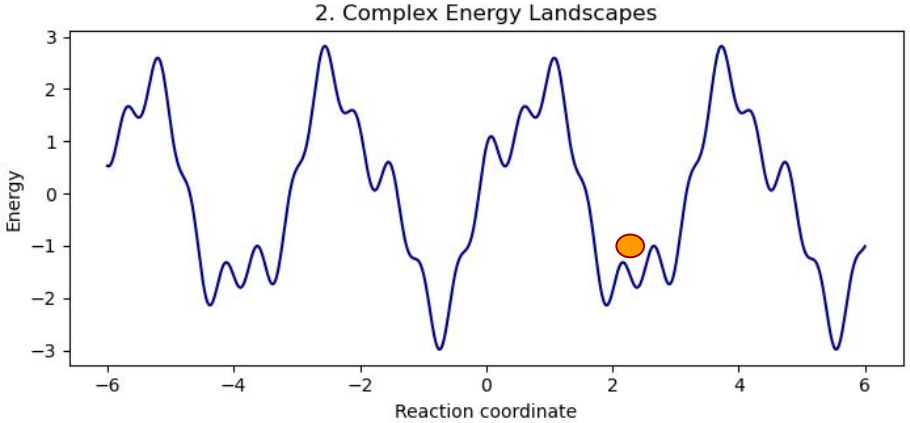
Why Ergodicity Fails?

- High energy barriers prevent transitions between states
- Complex energy landscapes with multiple minima

Molecule started in one well may not cross to the other during MD



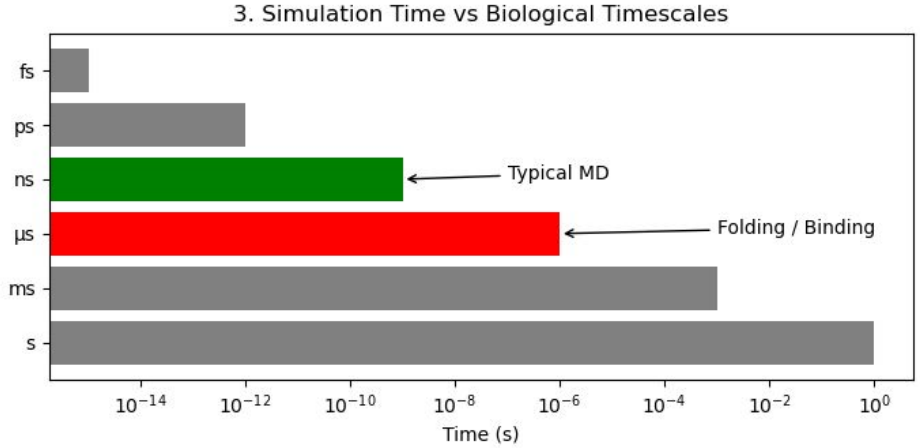
Molecule is trapped in local minimum state



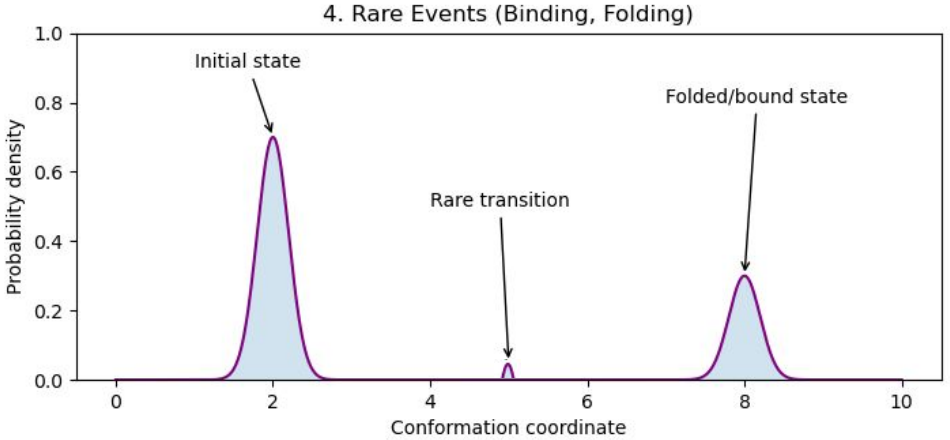
Why Ergodicity Fails?

- Limited simulation time (ns-μs) vs biological timescales (μs-ms)
- Systems with rare events: binding, folding, allostery

Events in biology have long timescales (and it's not guaranteed that MD will see them even if we reach those)



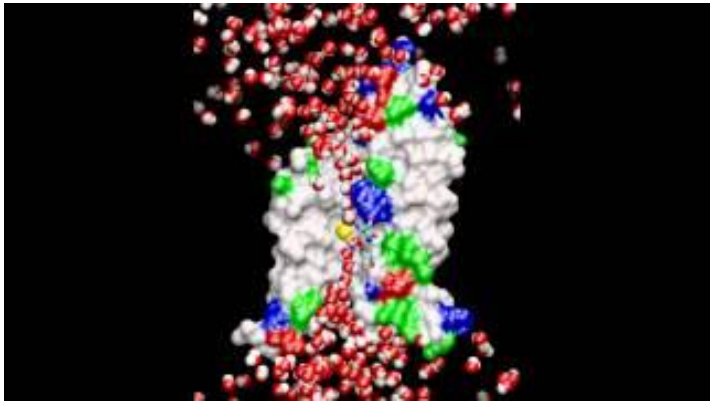
True probabilities might be hindered by difficult to observe rare events



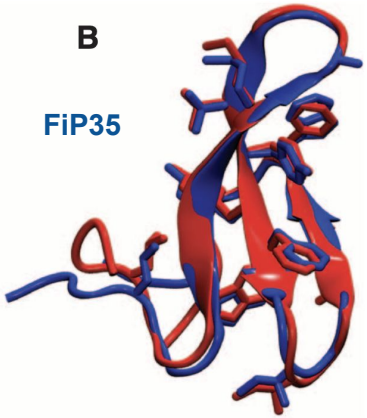
So… is MD any good?

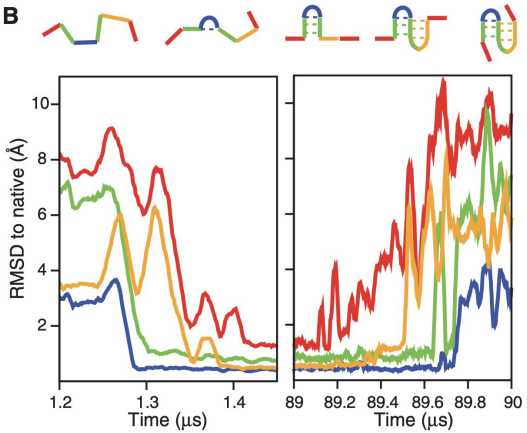
Trust the assumption: sample enough parts of the phase space around the native structure → reliable estimates of our property of interest.

- Did we miss important areas of the phase space? Hard to know…
- Understand/guess what are the slow motions → helps to estimate simulation time
- REPEATS!!! As many as you can get, ideally
- Some properties have different sensitivities to sampling → might equilibrate faster!
- Evaluate them during a simulation: if the mean and the SD are small and stable, it's a good sign



Aquaporins

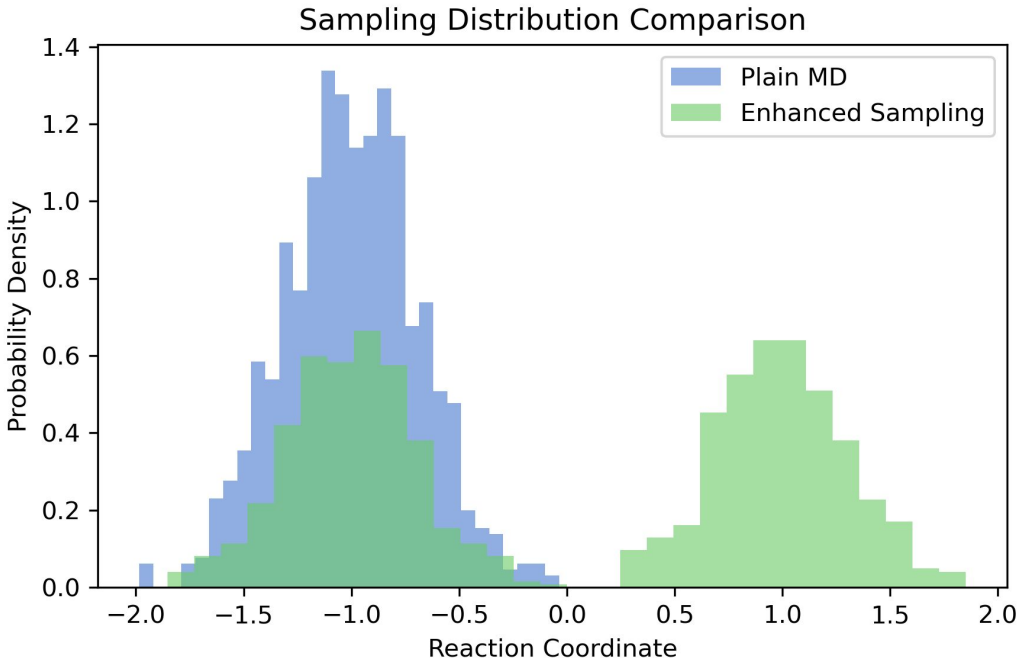




Shaw et al. (2010) Science

Strategies to Improve Sampling

- Replica Exchange MD (REMD)
- Metadynamics: add bias along collective variables
- (Gaussian) Accelerated MD: reduce energy barriers
- Steered MD + Umbrella Sampling: force sampling along reaction coordinates



Machine Learning-Aided Sampling

- Define better collective variables using dimensionality reduction
- Reinforcement learning to explore under-sampled regions
- Deep learning for bias potential generation

Biomolecular Emulator (BioEmu)

RAVE - Reweighted autoencoded variational Bayes for enhanced sampling

RAVE is a machine learning-based framework that learns the reaction coordinates, which can be used in biased MD simulation to enhance the sampling. Please read and cite these manuscripts when using RAVE:
https://aip.scitation.org/doi/abs/10.1063/1.5025487 https://www.nature.com/articles/s41467-019-11405-4
https://arxiv.org/abs/2002.06099

TorchMD

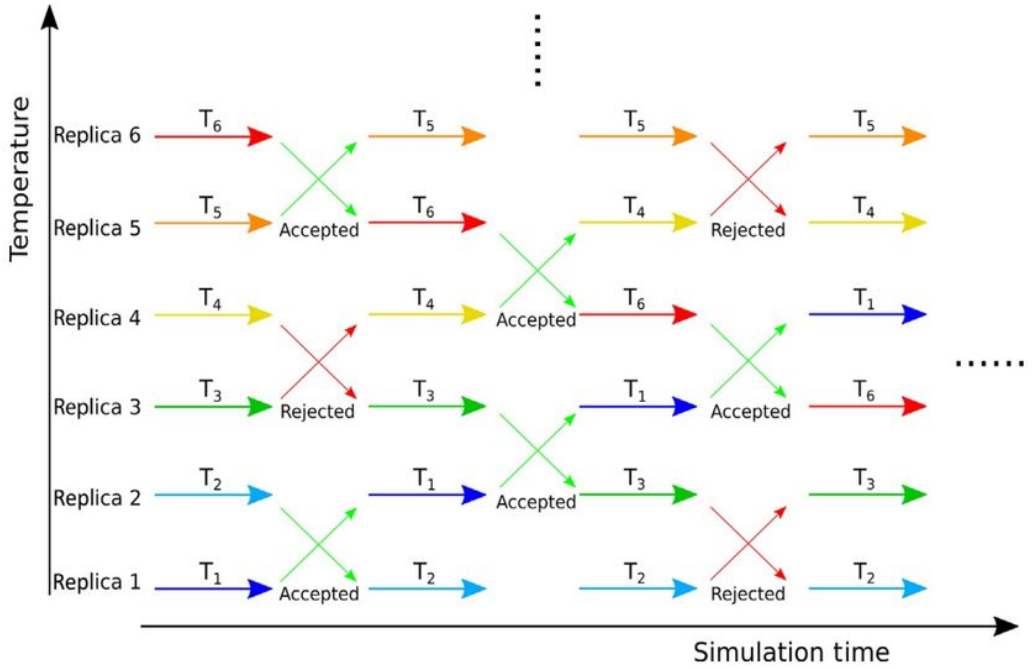
Replica Exchange

- Replica-Exchange Molecular Dynamics (REMD)

Increasing temperatures for each replica

Exchange of conformations at every X step

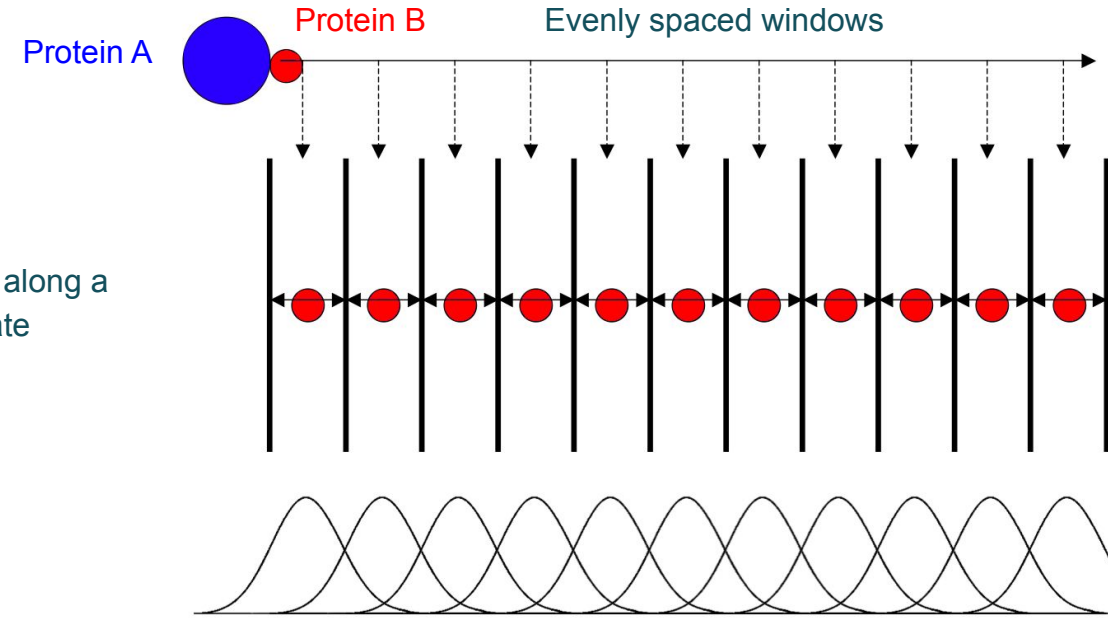
Probability of acceptance is important (0.25 is the optimal value)



http://dx.doi.org/10.13140/RG.2.2.28745.93288

Umbrella Sampling

- Steered Molecular Dynamics (Steered MD)



Pulling Protein B along a reaction coordinate

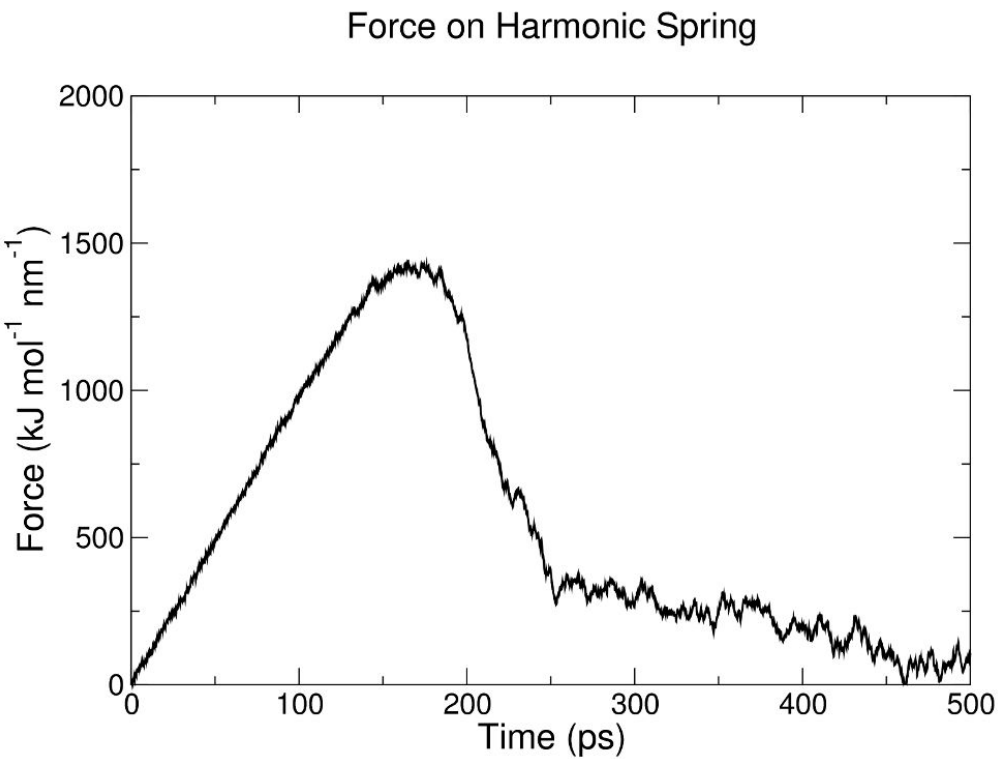
Harmonic potential (Umbrella) at every window

Histogram where each conformation overlaps with the next one

http://www.mdtutorials.com/gmx/umbrella/

Umbrella Sampling

Force builds up until dissociation occurs!

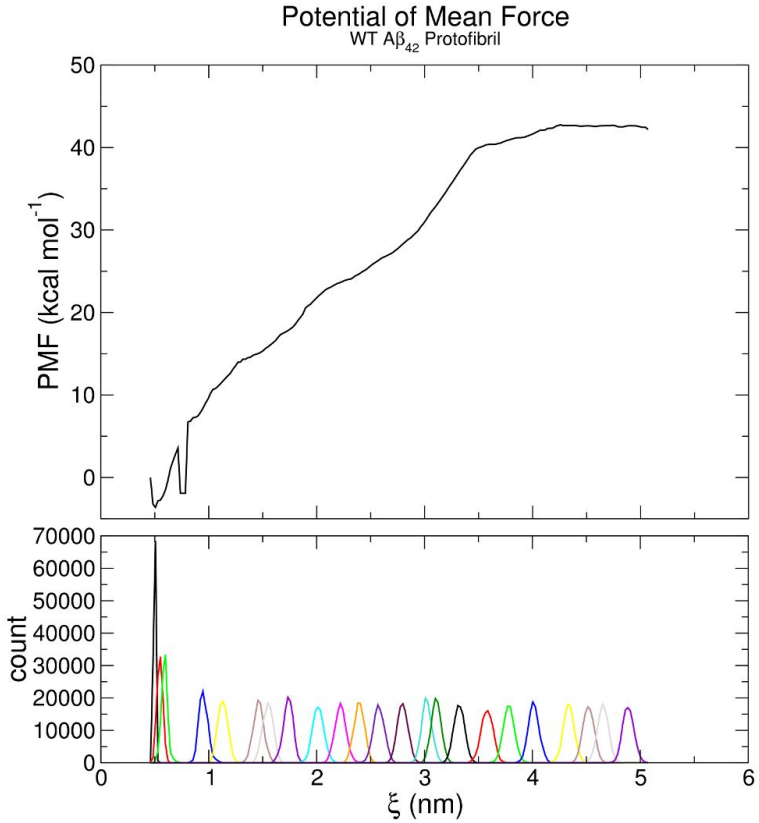


http://www.mdtutorials.com/gmx/umbrella/

Umbrella Sampling

Final PMF → can derive the free energy of the process using the lowest and the highest energy regions

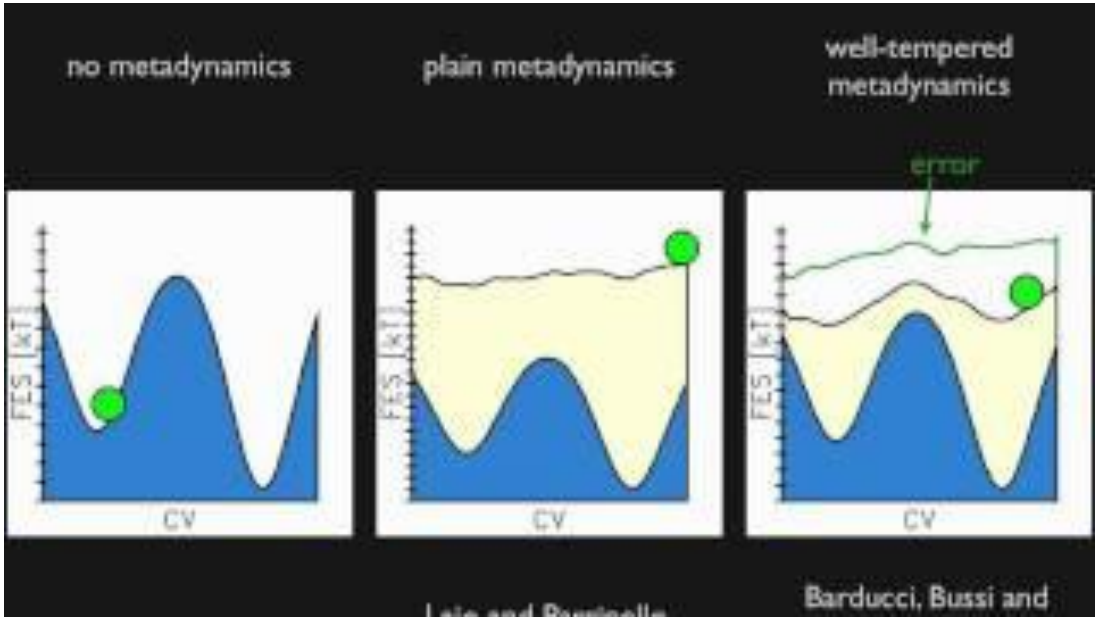
Windows overlap reasonably well (some issues might be corrected)



http://www.mdtutorials.com/gmx/umbrella/

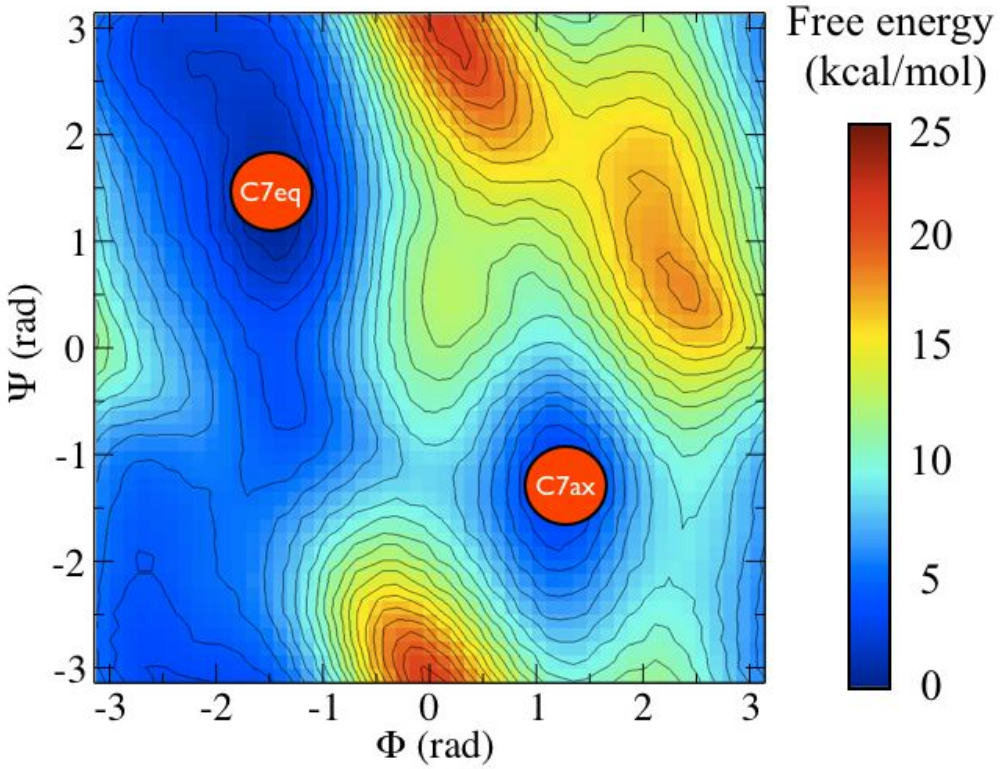
Metadynamics

- Metadynamics → time dependent bias designed so as to avoid re-exploring the already visited regions
- The bias is grown at a constant rate by adding "computational sand" at the position where the system actually is.
- Standard algorithm → fills the system constantly, might explore high energy areas
- Well-tempered → bias deposited decreases as the simulation runs → more accurate



Metadynamics

- Metadynamics → time dependent bias designed so as to avoid re-exploring the already visited regions
- The bias is grown at a constant rate by adding "computational sand" at the position where the system actually is.
- Standard algorithm → fills the system constantly, might explore high energy areas
- Well-tempered → bias deposited decreases as the simulation runs → more accurate



https://www.plumed.org/doc-v2.9/user-doc/html/marvel-2.html

Take-Home Messages

- Always evaluate sampling quality in MD
- Ergodicity is assumed but rarely achieved in short simulations
- Enhanced sampling methods are essential for complex systems
- Use visual, statistical, and physical intuition to assess convergence

Further studies

Prof. Erik Lindahl

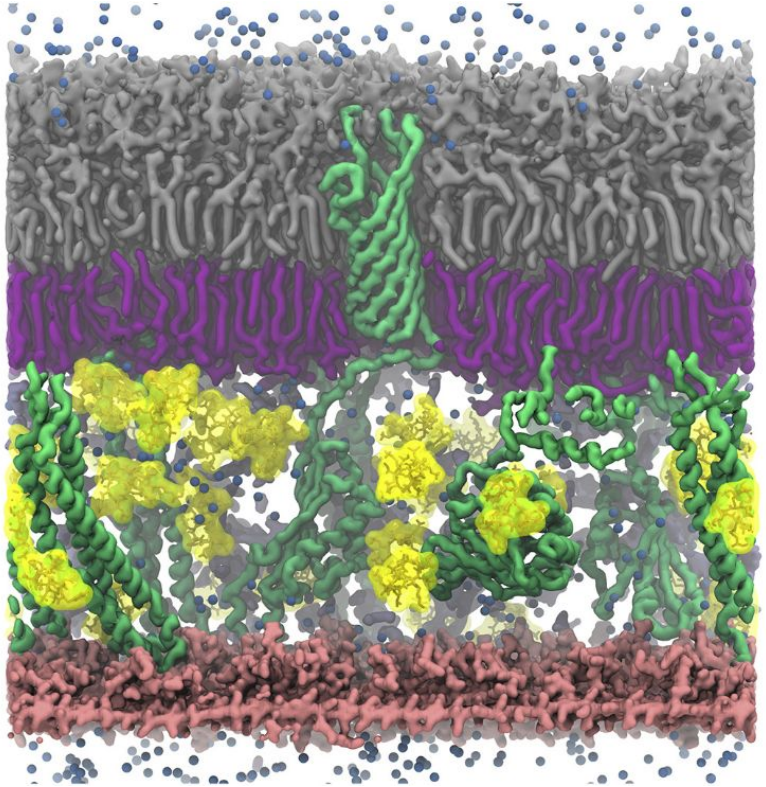
https://www.youtube.com/watch?v=TDzzvKoDOuQ&list=PLuIpgNT2hMwRQKFy4okoNQKiJwM8li3Sz

THANK YOU!

github.com/cpedebos

conrado.pedebos@ufcspa.edu.br

linkedin.com/in/conrado-pedebos



LAMASU

UFCSPA

Universidade Federal de Ciências da Saúde de Porto Alegre

Laboratório de Modelagem Molecular Aplicada à Saúde Única