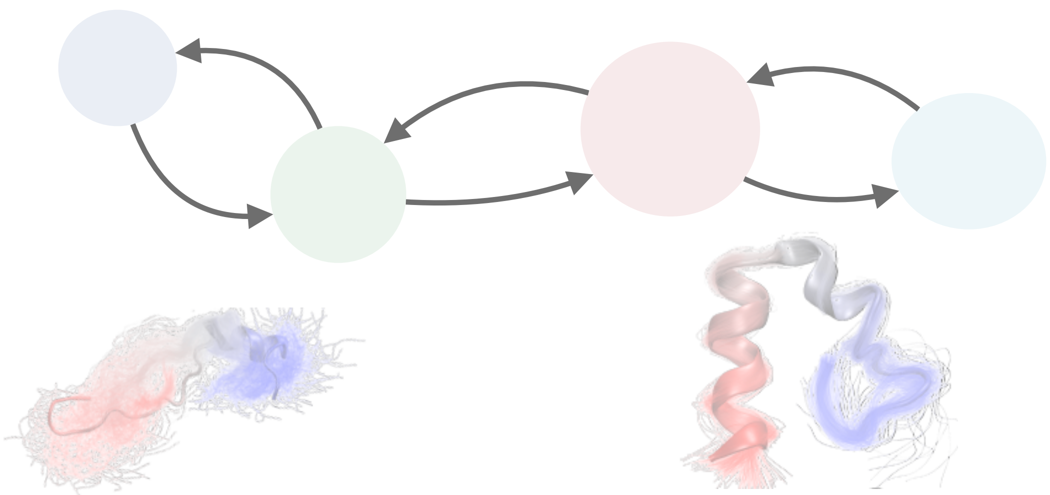
Introduction to Markov State Modelling

Theory, estimation and validation



Antonia Mey — antonia.mey@ed.ac.uk

School of Chemistry

University of Edinburgh, UK

What is a Markov State Model?

$$P = \begin{bmatrix} P_{1,1} & P_{1,2} & \dots & P_{1,j} & \dots & P_{1,S} \\ P_{2,1} & P_{2,2} & \dots & P_{2,j} & \dots & P_{2,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{i,1} & P_{i,2} & \dots & P_{i,j} & \dots & P_{i,S} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ P_{S,1} & P_{S,2} & \dots & P_{S,j} & \dots & P_{S,S} \end{bmatrix}.$$

Where to start with Markov Models?

https://www.livecomsjournal.org

A LiveCoMS Tutorial

Introduction to Markov state modeling with the PyEMMA software [Article v1.0]

Christoph Wehmeyer[1†*], Martin K. Scherer[1†], Tim Hempel[1†], Brooke E. Husic[1,2], Simon Olsson[1], Frank Noe[1,3*]

https://github.com/markovmodel/pyemma_tutorials

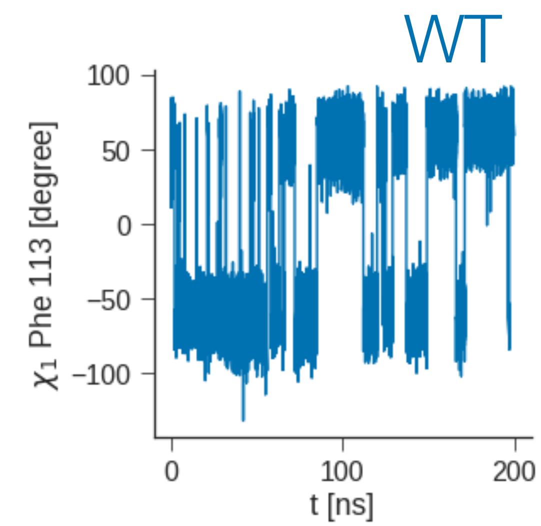
markovmodel / pyemma_tutorials

<> Code | Issues 1 | Pull requests 0 | Projects 0 | Wiki | Insights

How to analyze molecular dynamics data with PyEMMA

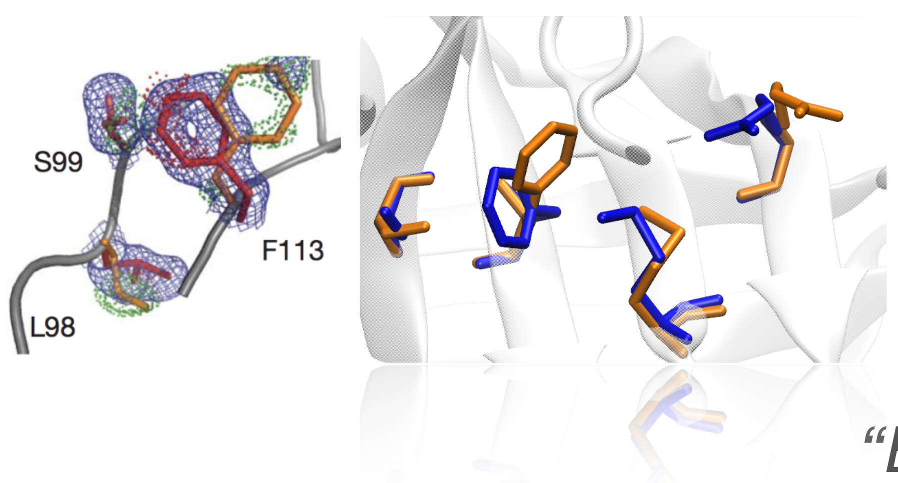
Motivation

experiments /
simulations

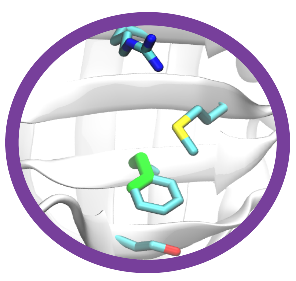


question

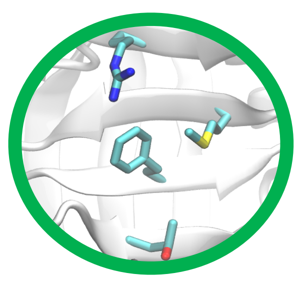
estimation



prediction



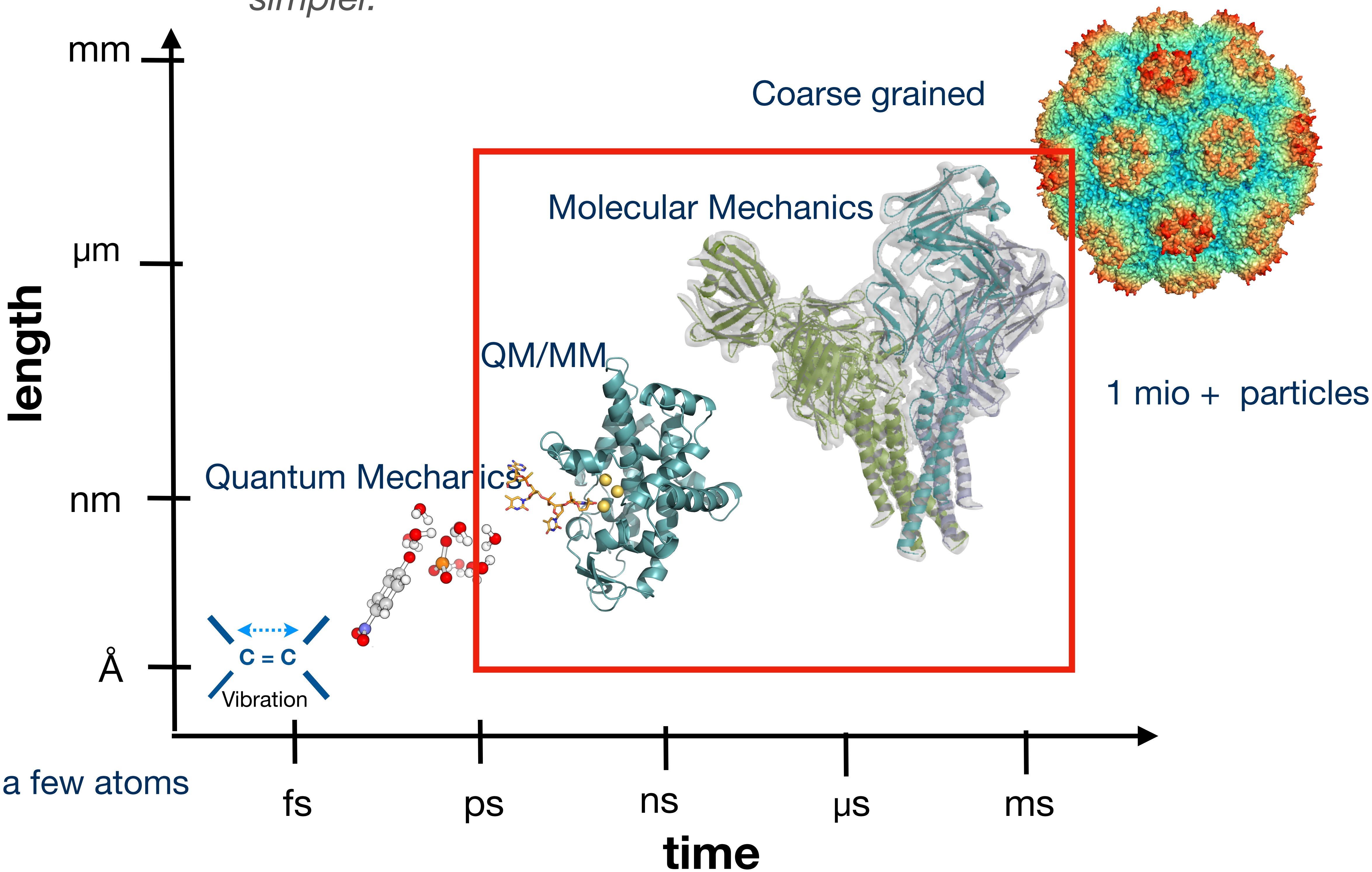
Fast
Local



"Everything should be made as simple as possible, but not simpler."

Timescales and lengthscales

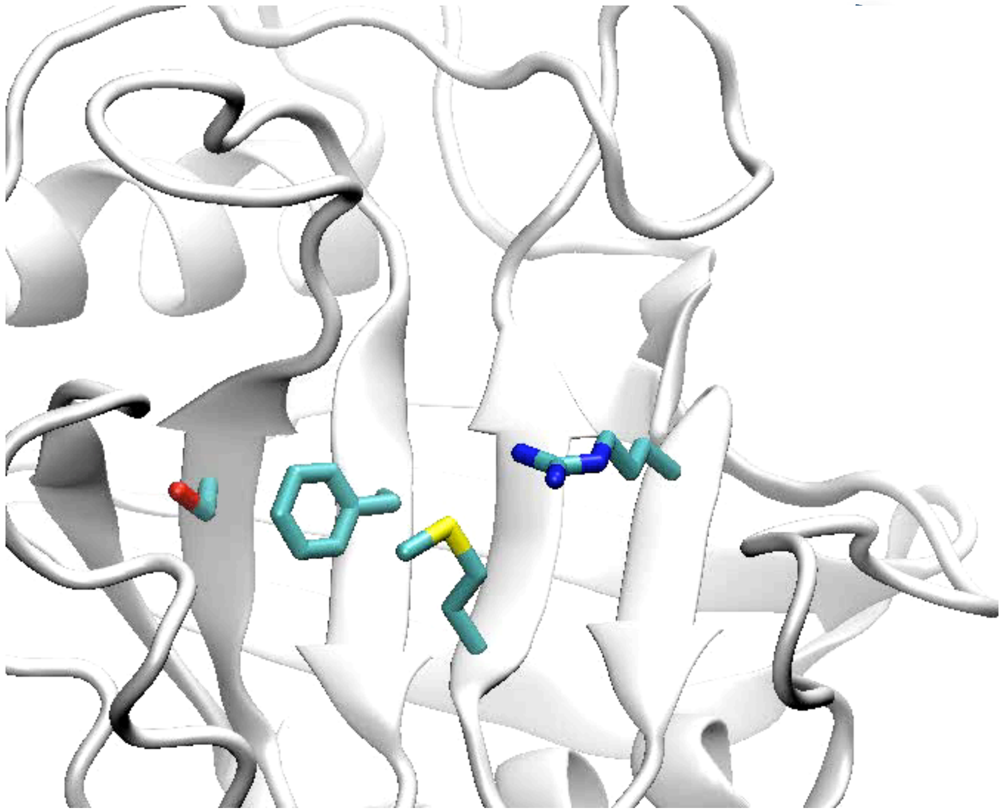
"Everything should be made as simple as possible, but not simpler."



https://quoteinvestigator.com/2011/05/13/einstein-simple/

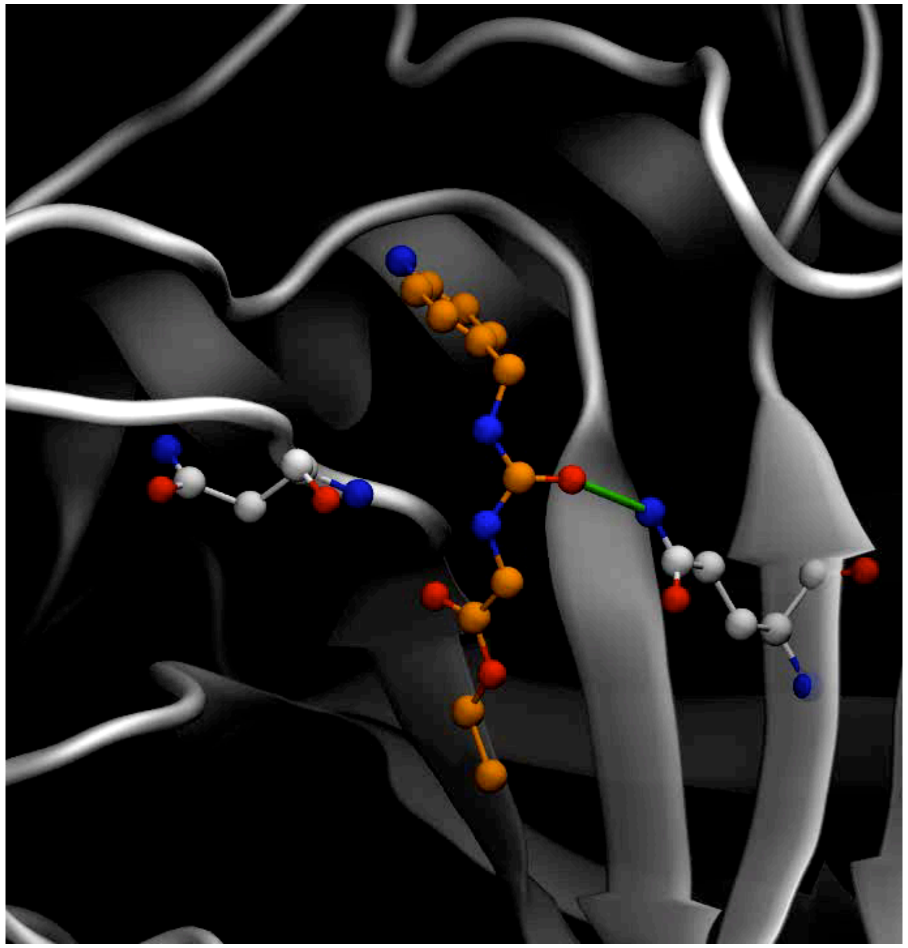
Molecular dynamics

Conformational dynamics



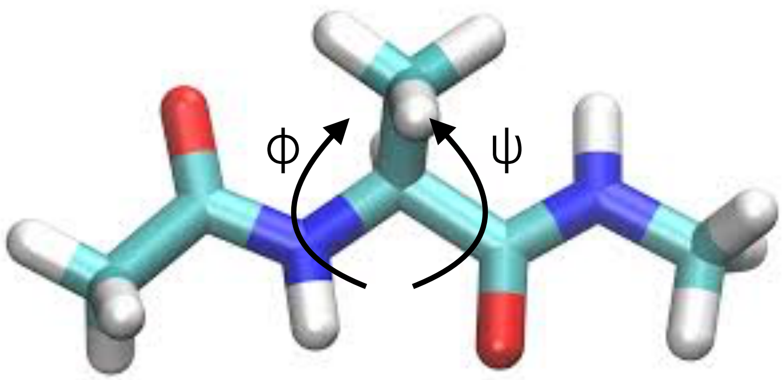
- Local Motions (0.01 - 5 Å, 10^{-15} - 10^{-1} s)
  - atomic fluctuations
  - sidechain motions
  - loop motions

Binding dynamics

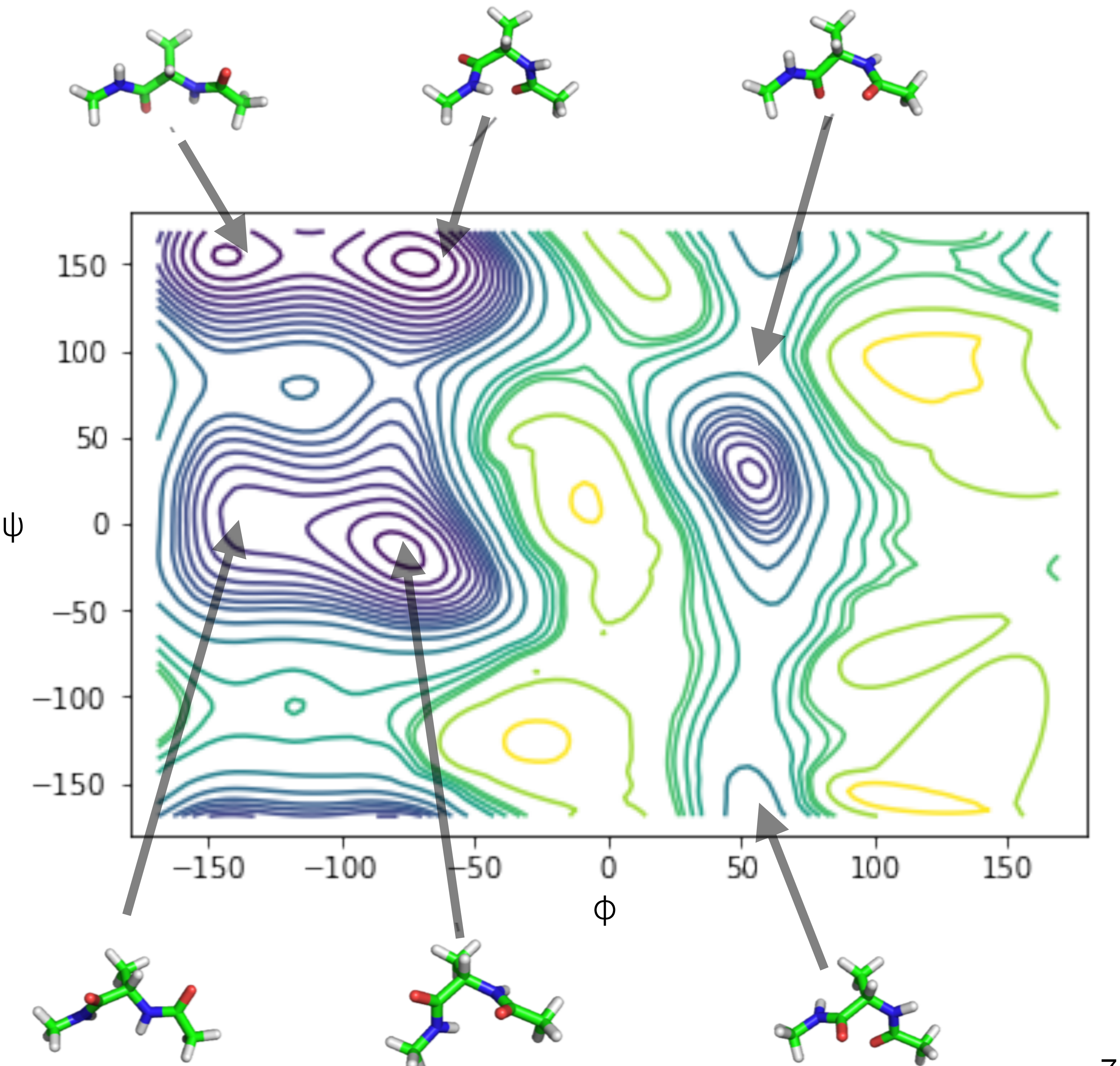


- Large-Scale Motions (> 5Å, 10^{-7} to 10^{4} s)
  - helix coil transitions
  - dissociation/association
  - folding and unfolding

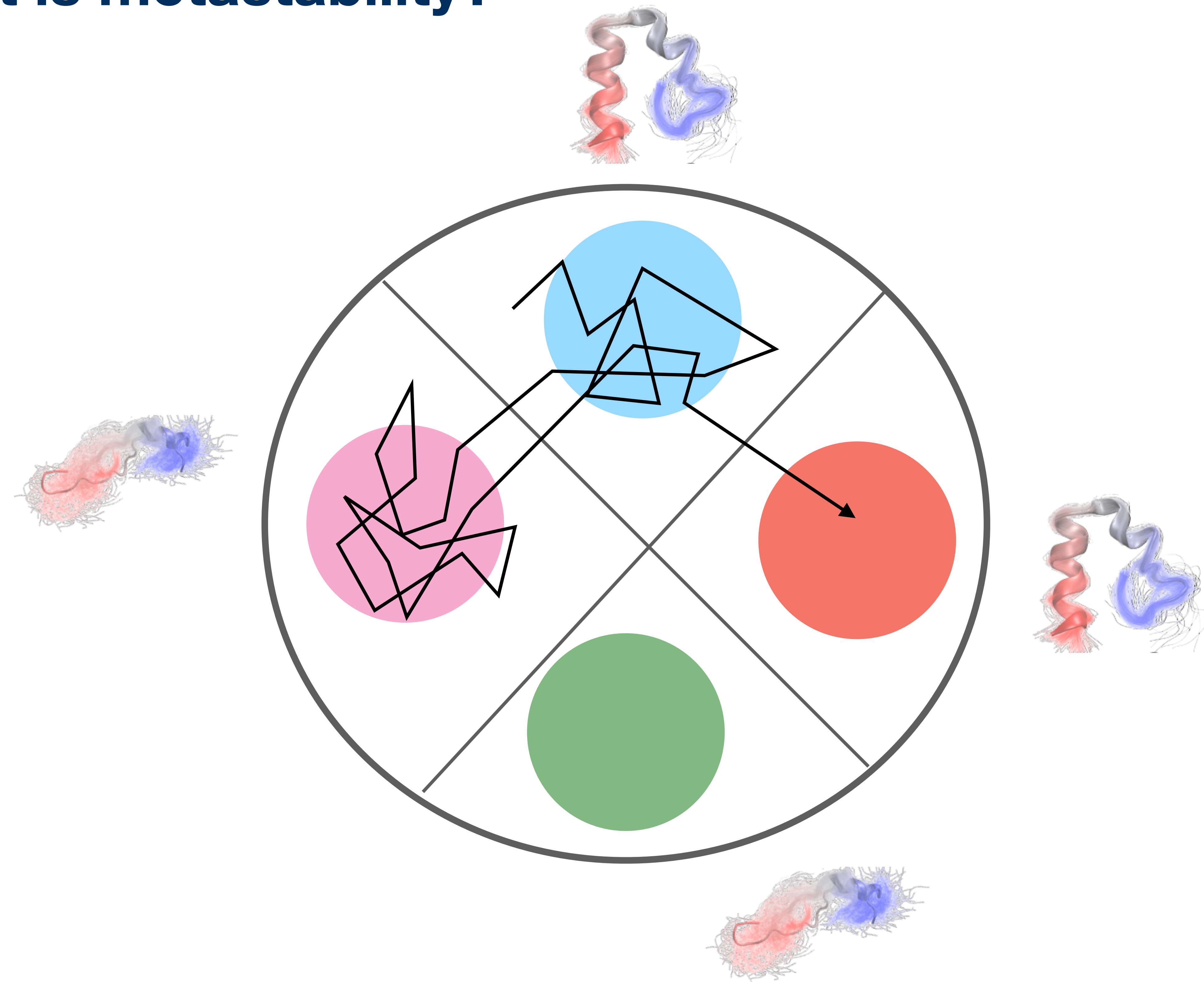
Conformational dynamics



Alanine dipeptide



What is metastability?



Ω = space of all possible configurations

What is Ergodicity

No segments of the space Ω are dynamically disconnected

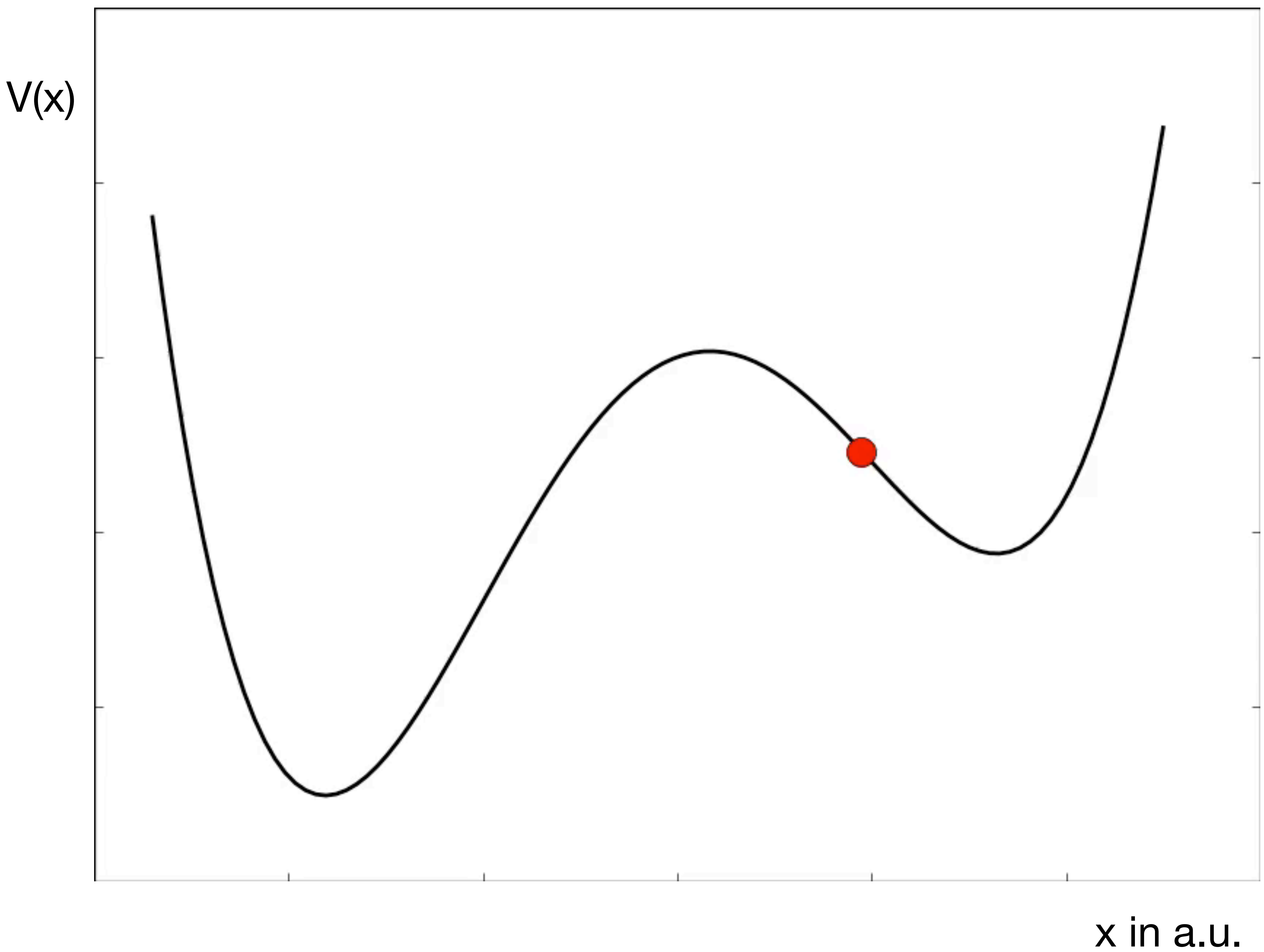
and

an infinitely long simulation will have visited every state $\mathbf{x}$ in Ω infinitely many times.

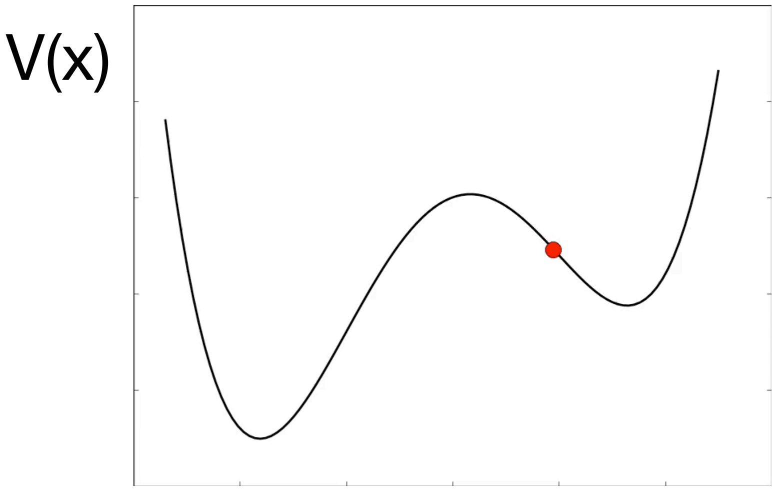
$$\lim_{t \to \infty} \hat{A}_t = \mathbb{E}_\mu(A)$$

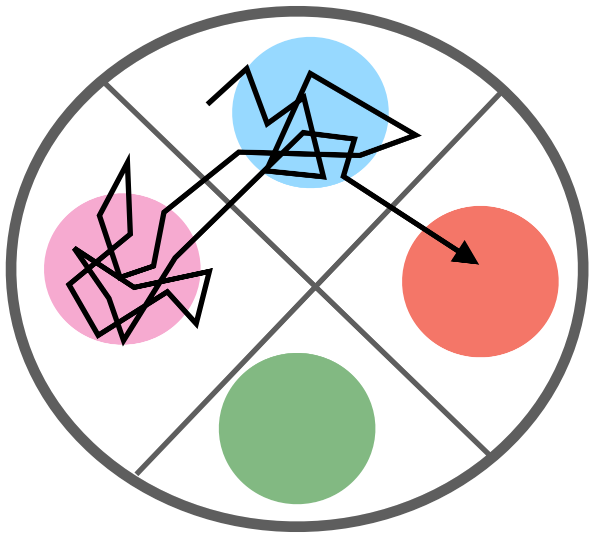
We can use time averages to observe conformational averages!

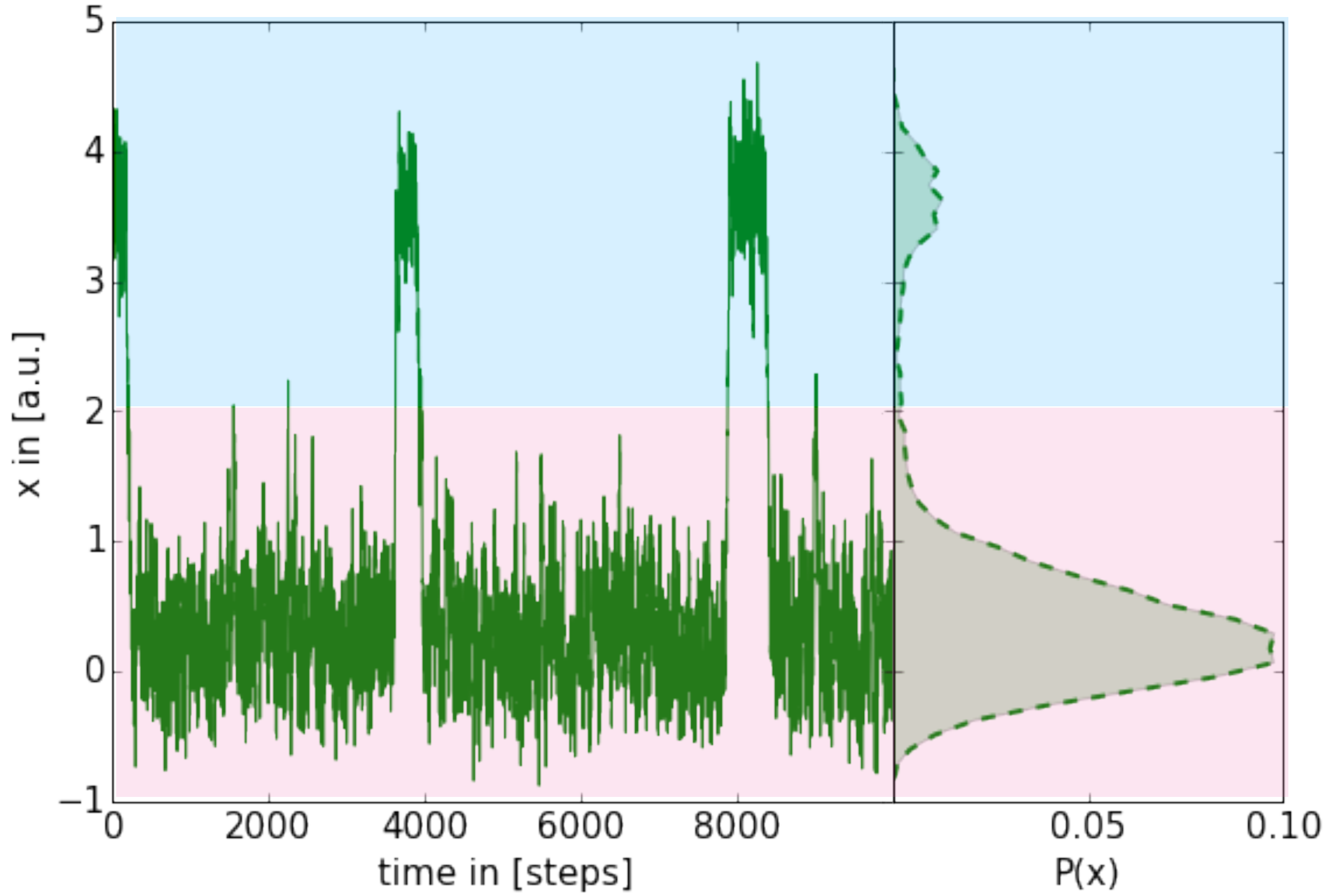
Langevin dynamics



Langevin dynamics







What is Reversibility?

Simulations fulfil the detailed-balance condition.

$$\mu(\mathbf{x})p(\mathbf{x},\mathbf{y};\tau) = \mu(\mathbf{y})p(\mathbf{y},\mathbf{x};\tau)$$

$$\mu(\mathbf{x}) = \frac{\exp(-\beta V(\mathbf{x}))}{Z(\beta)} \qquad Z = \int_\Omega \exp(-\beta V(x))dx$$

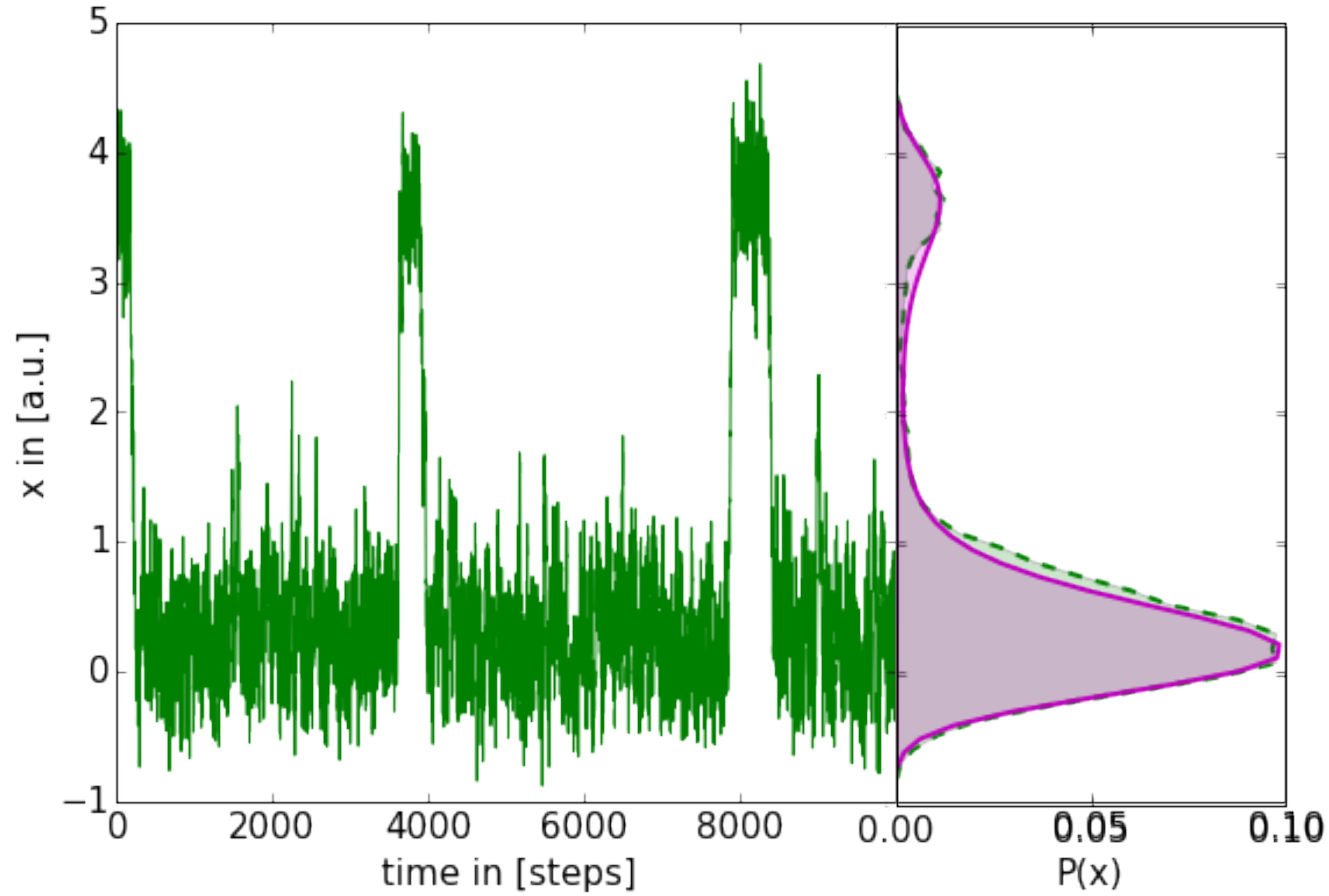
At equilibrium the probability of jumping from **x** to any **y**, is the same as jumping from **y** to **x**.

Typical questions?

What is the *transition rate* between the metastable states?

What is the *free energy difference* between the metastable states?

$$G = -k_B T \ln(\mu(x))$$

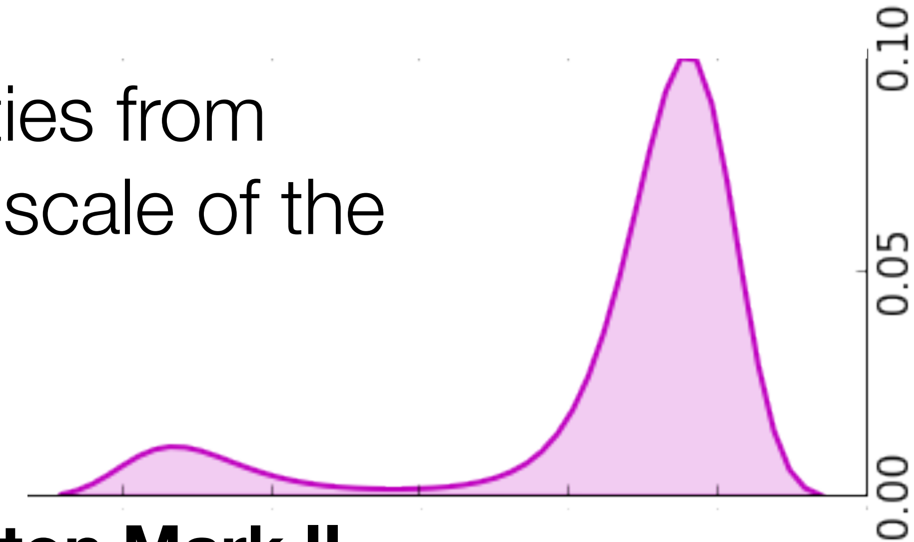


$$\mu(\mathbf{x}) = \frac{\exp(-\beta V(\mathbf{x}))}{Z(\beta)}$$

$$Z = \int_{\Omega} \exp(-\beta V(x)) dx$$

Simulating 50k atoms on a….

How can we reliably estimate equilibrium properties from simulations that are shorter than that of the time scale of the equilibrium process?



NVIDIA GTX1080

~350 ns/d

Anton Mark II

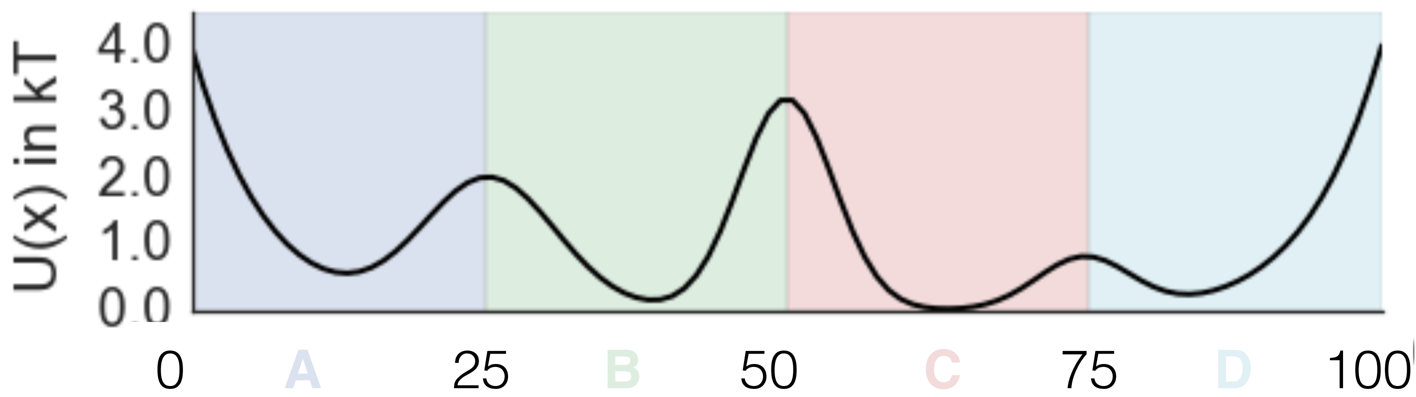
~70 μs/d

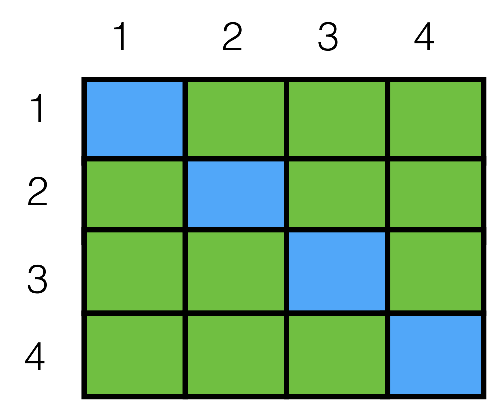
→ **Model molecular simulations as a Markov jump process between metastable states.**

MSM construction overview

1. Generate trajectories e.g. through MD

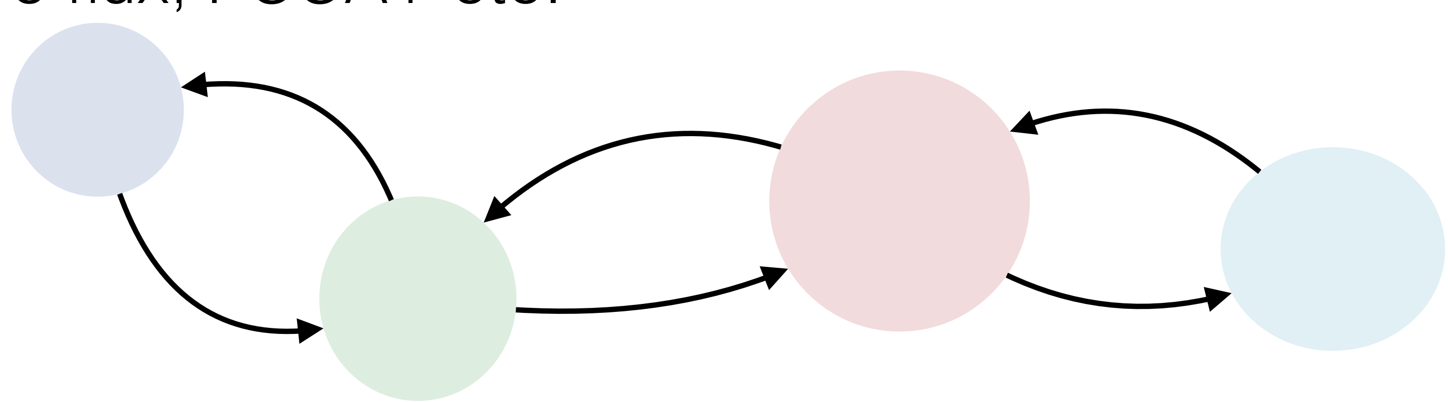
2. Discretise trajectories (two steps (a) dimensionality reduction (b) clustering



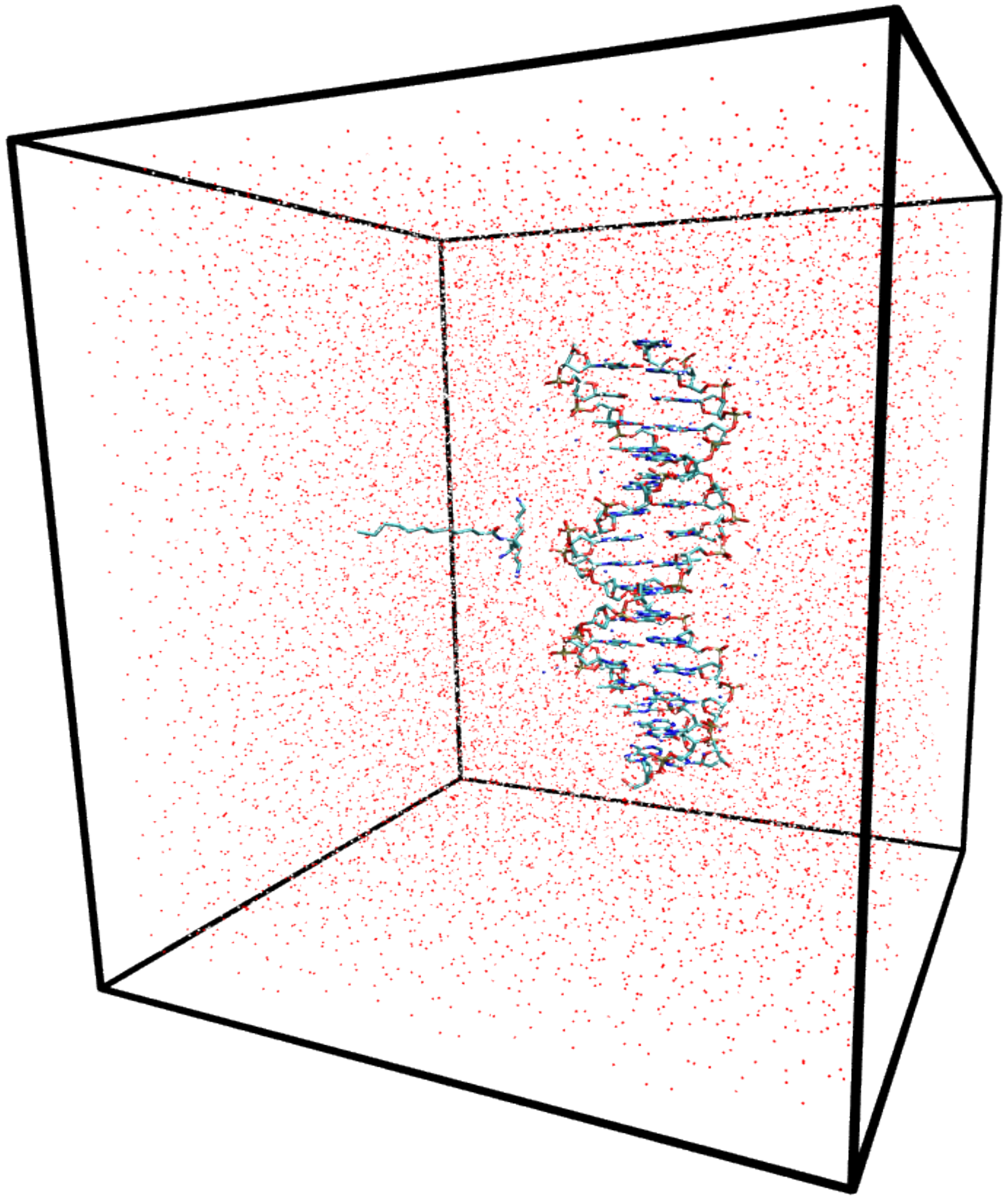


3. Estimate transition matrix (MSM) e.g. Baysian MSM, HMM

4. Analyse transition matrix (MSM) e.g. stationary properties, timescales, reactive flux, PCCA+ etc.



Trajectory generation



$$V = \sum_{\text{bonds}} K(\mathbf{r} - \mathbf{r_{eq}})^2$$

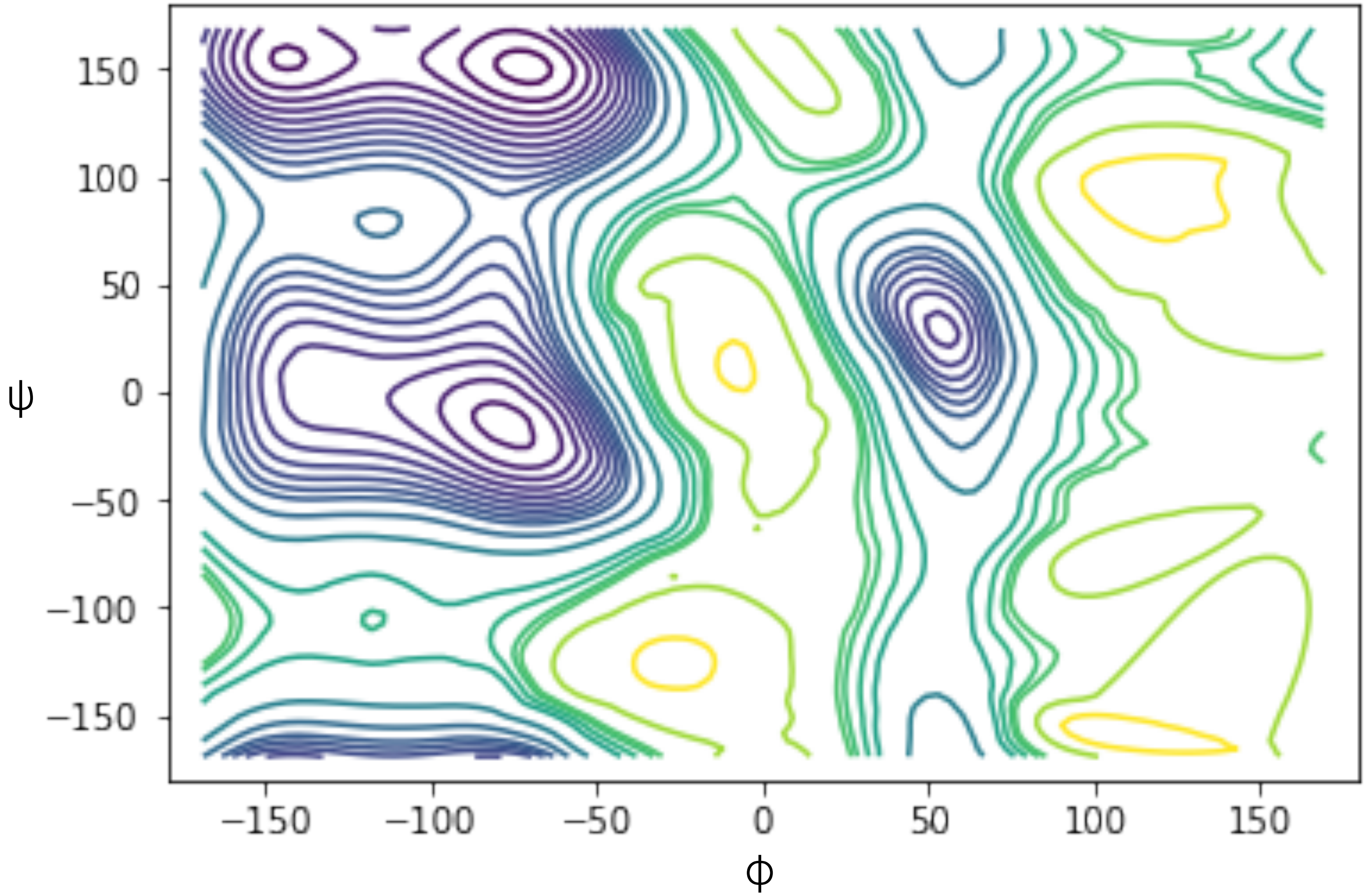
$$+ \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2$$

$$+ \sum_{\text{dihedral}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)]$$

$$+ \sum_{\text{non-bonded}} [\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}]$$

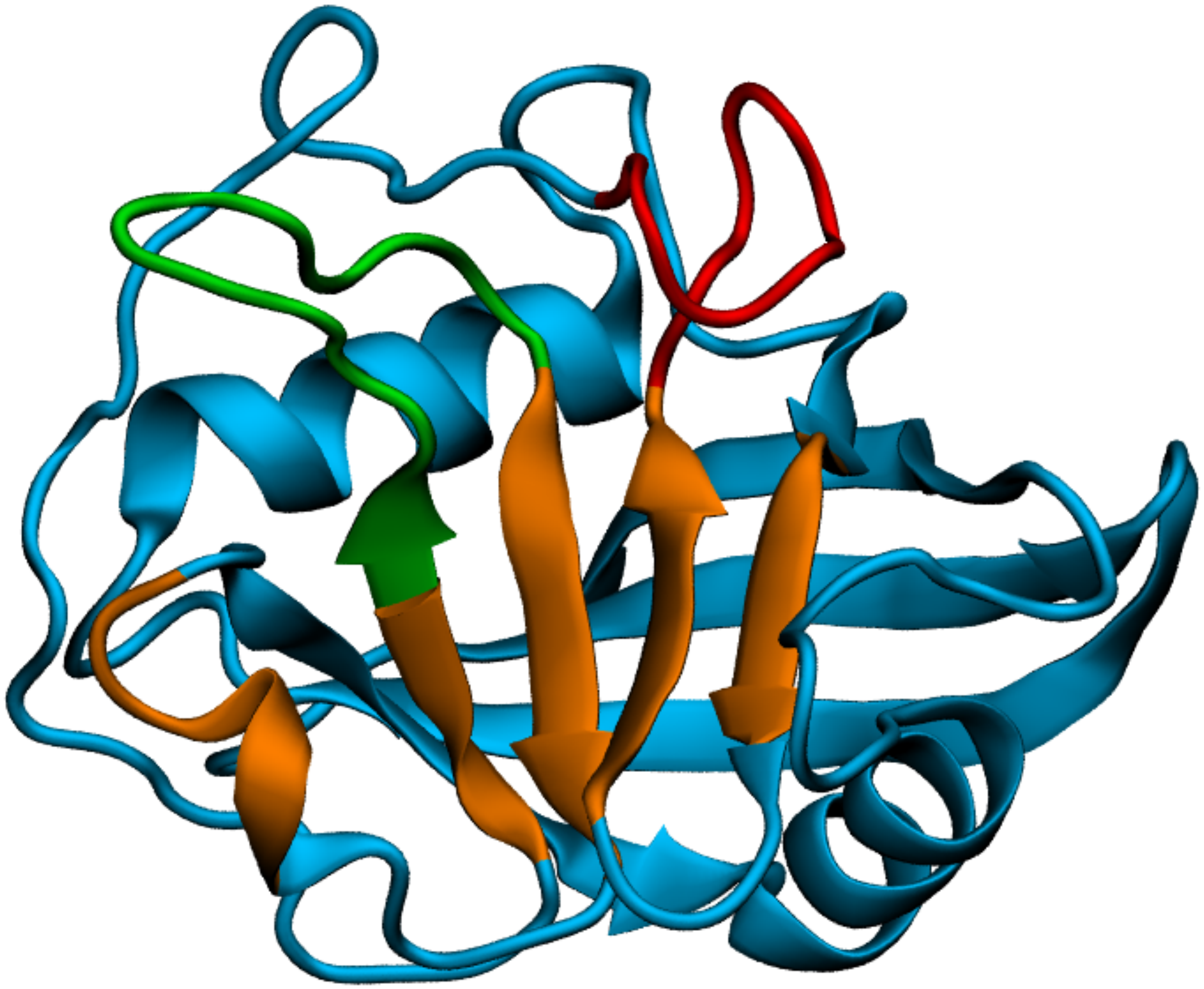
Dimensionality reduction — choosing features

```
['PHI 0 ALA 2', 'PSI 0 ALA 2']
```



2D features -> we can go straight to clustering

Dimensionality reduction — choosing features

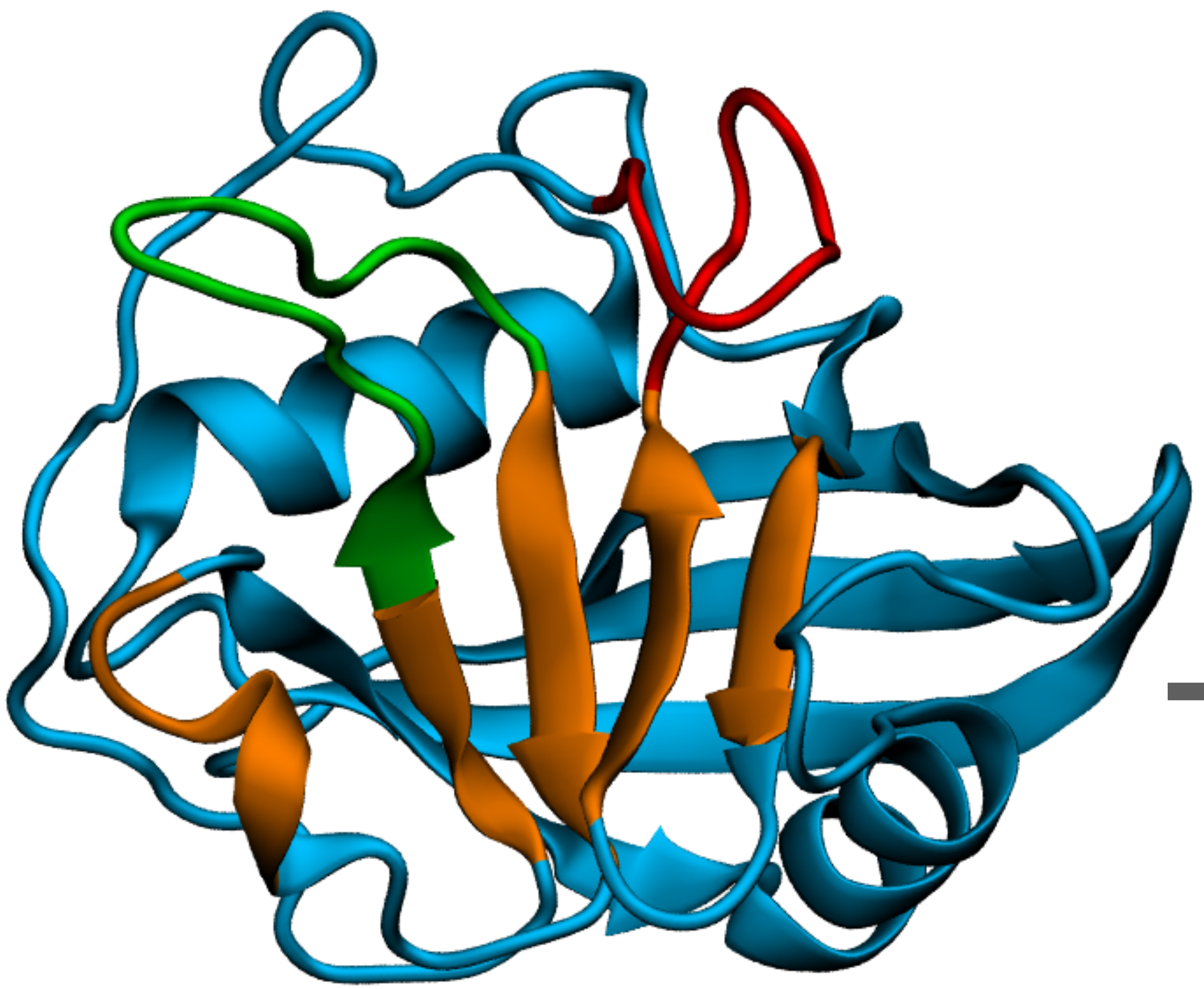


```
['ATOM:ACE 1 CH3 1 x',
 'ATOM:ACE 1 CH3 1 y',
 'ATOM:ACE 1 CH3 1 z',
 'ATOM:ACE 1 C 4 x',
 'ATOM:ACE 1 C 4 y',
 'ATOM:ACE 1 C 4 z',
 'ATOM:ACE 1 O 5 x',
 'ATOM:ACE 1 O 5 y',
 'ATOM:ACE 1 O 5 z',
 'ATOM:ALA 2 N 6 x',
 'ATOM:ALA 2 N 6 y',
 'ATOM:ALA 2 N 6 z',
 'ATOM:ALA 2 CA 8 x',
 'ATOM:ALA 2 CA 8 y',
 'ATOM:ALA 2 CA 8 z',
 'ATOM:ALA 2 CB 10 x',
 'ATOM:ALA 2 CB 10 y',
 'ATOM:ALA 2 CB 10 z',
 'ATOM:ALA 2 C 14 x',
```

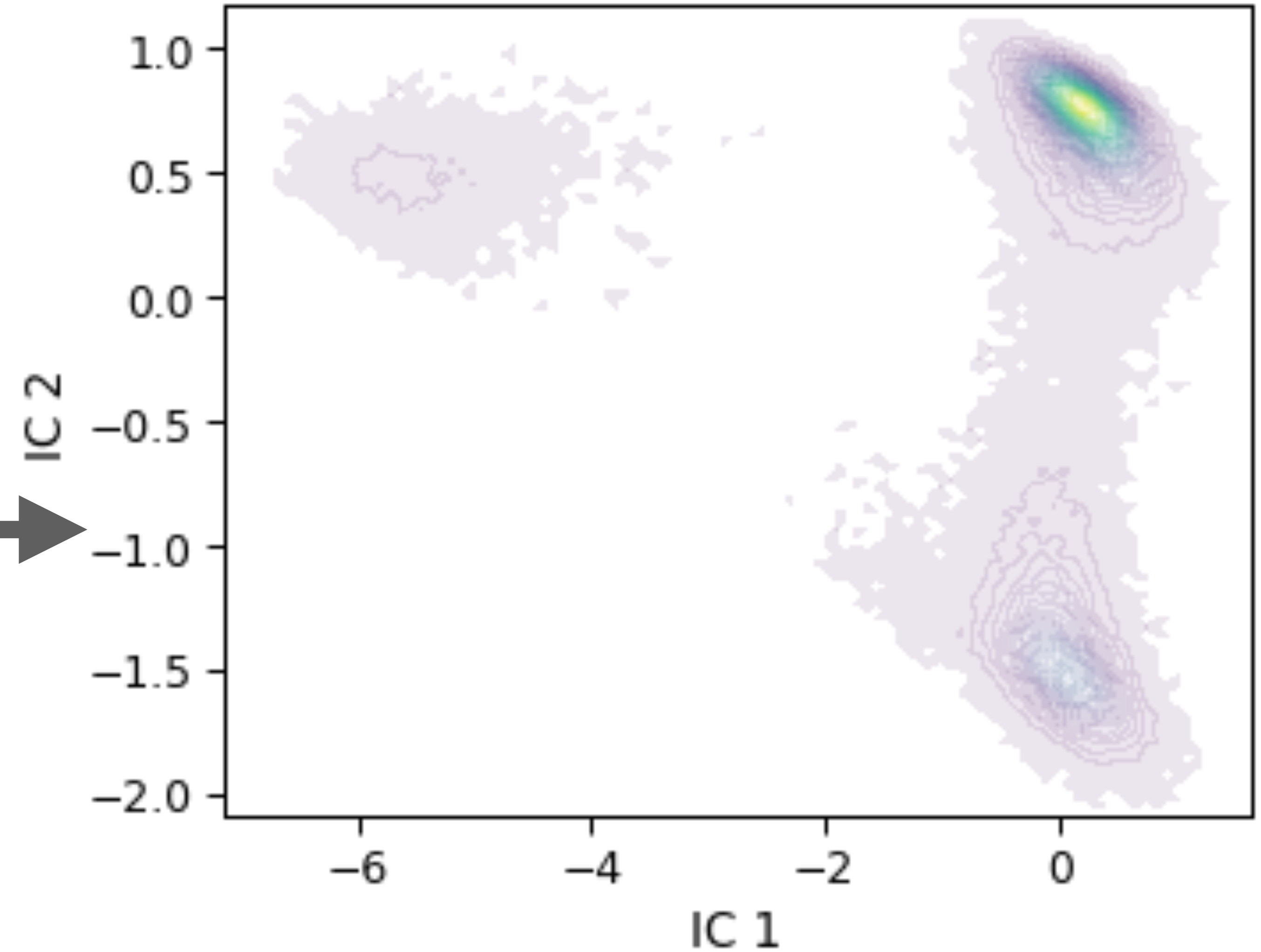
```
['PHI 0 ALA 2', 'PSI 0 ALA 2']
```

dimensionality of features >4? -> dimensionality reduction

Dimensionality reduction — TICA - PCA - VAMP



Project features onto low dimensional subspace



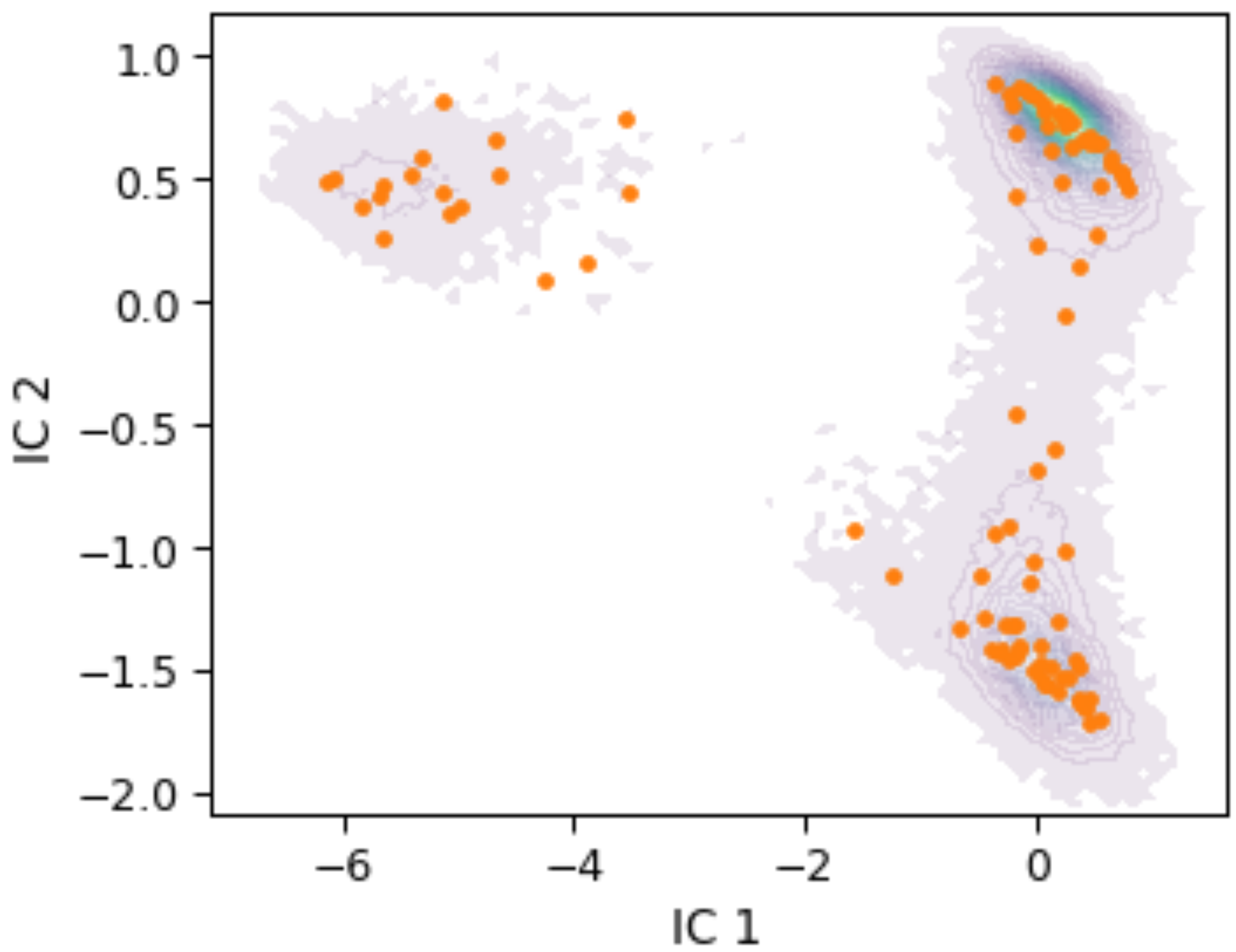
PCA: Linear combination of input features maximising the variance

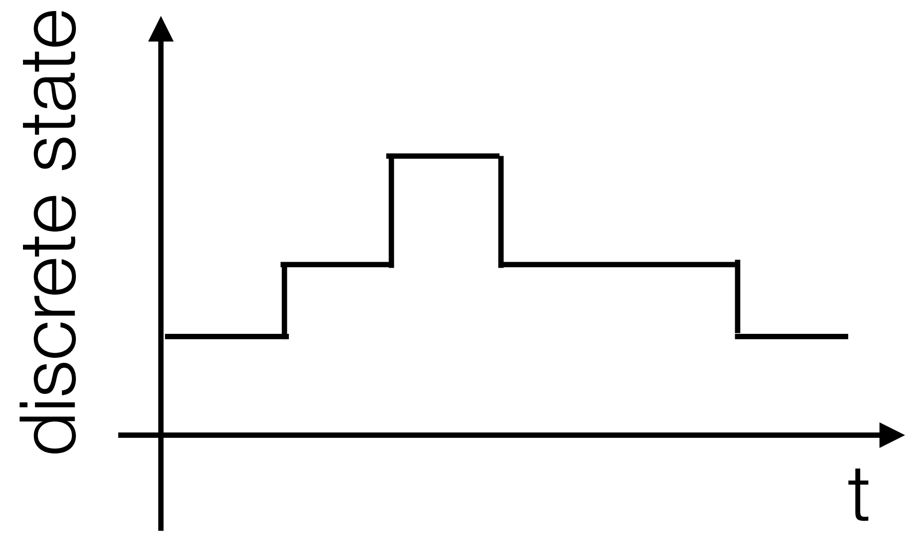
TICA: Linear combination of input features maximising time autocorrelation

VAMP: Variational approach for Markov Process, true for non-equilibrium data

TICA: J. Chem. Phys. **139**, 015102 (2013);
VAMP: J. Chem. Phys. **150**, 194108 (2019)

Discretisation





k-means

regular spatial

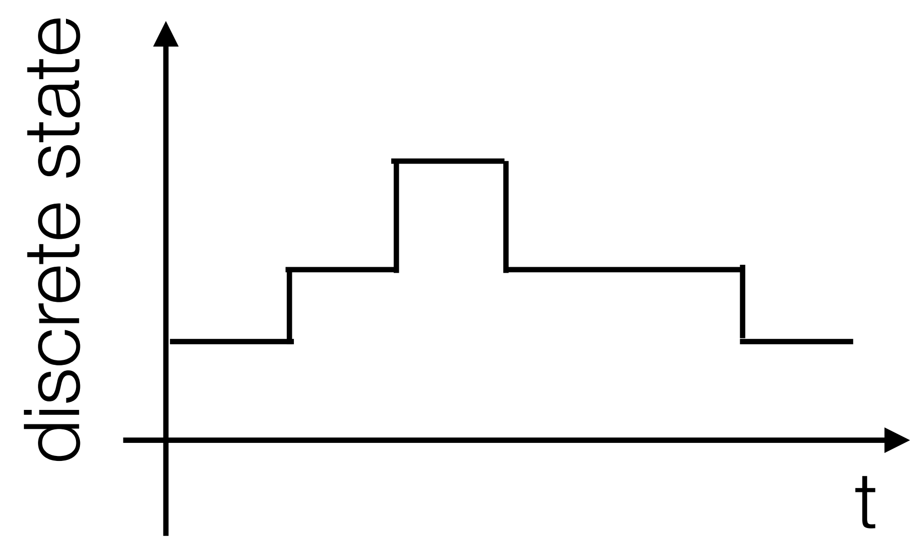
etc

Dimensionality reduction and discretisation require a lot of parameter optimisation. It is necessary to spend a good amount of time on hyper parameter optimisation.

The count matrix

$$\begin{bmatrix} \sum_t s_i \rightarrow s_i & \sum_t s_i \rightarrow s_j \\ \sum_t s_j \rightarrow s_i & \sum_t s_j \rightarrow s_j \end{bmatrix} = \mathbf{C}$$

The countmatrix contains the number of times a transition from state *i* to state *j* is observed.



Transition matrix estimation

$$\begin{bmatrix} \frac{c_{ii}}{\sum_i c_{ij}} & \frac{c_{ij}}{\sum_i c_{ij}} \\ & \end{bmatrix} = \mathbf{T}$$

The transition matrix contains conditional probabilities, of going from state i to state j.
Usually, a reversible estimation is used to ensure detailed balance.

$$\mu_i P_{ij} = \mu_j P_{ji}$$

Writing the transition matrix as a Markov jump process:

$$T_{ij}(\tau) = \mathbb{P}[\mathbf{x}(t+\tau) \in S_j | \mathbf{x}(t) \in S_i]$$

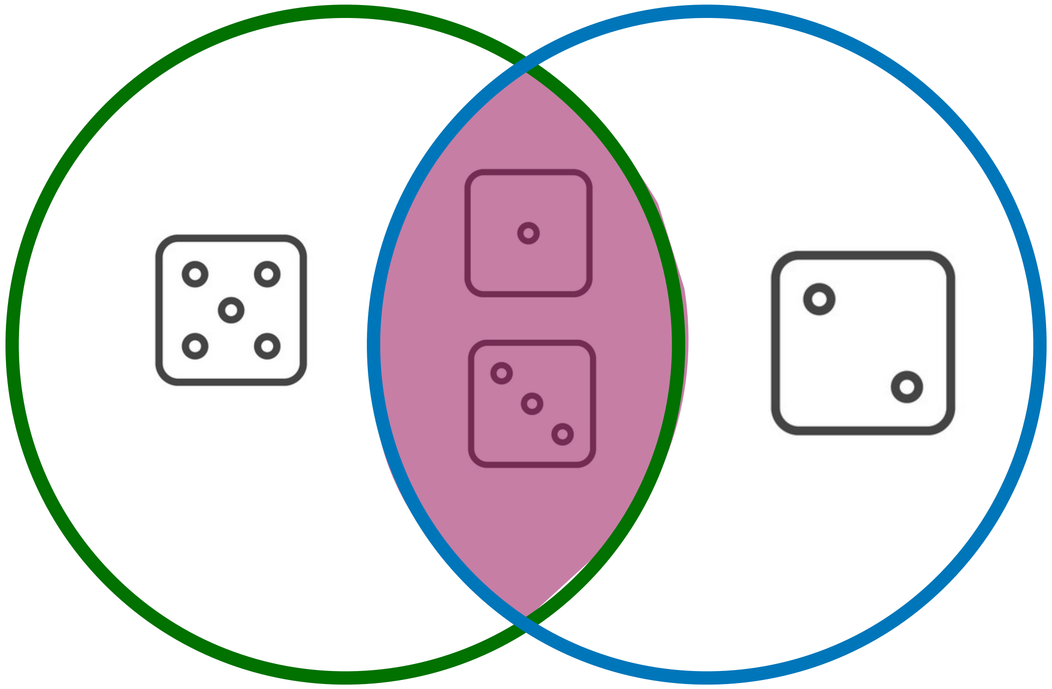
So what is a conditional probability?

Conditional probabilities

What is the probability of rolling
a die

and its value is less than 4,

$$P(A \mid B) = \frac{P(A \cap B)}{P(A)}$$



knowing (given) that the number is odd.

$$\boxed{P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}}$$

evidence

Bayes' theorem

hypothesis

Reversible Transition matrix estimation from counts

Objective: find the most likely **reversible transition** matrix, based on the **observed counts** using Bayes

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

evidence → Observed counts

hypothesis → Transition matrix

Likelihood

$$\mathbb{P}(\mathbf{T}|\mathbf{C}) = \prod_{i,j} t_{ij}^{c_{ij}}$$

We use log-likelihoods instead: $Q = \log \mathbb{P}(\mathbf{T}|\mathbf{C}) = \sum_{i,j} c_{ij} \log t_{ij}$

Maximise the log-likelihood, by taking its derivative and using the constraint, that detailed balance must hold, i.e. $\frac{\partial Q}{\partial x_{ij}} = 0$

$$\frac{\partial Q}{\partial x_{ji}} = \frac{c_{ij} + c_{ji}}{x_{ji}} - \frac{c_i}{x_i} - \frac{c_j}{x_j}$$

Prinz, J.-H. et al (2011) Markov models of molecular kinetics: Generation and Validation. J. Chem. Phys., 134 . p. 174105.

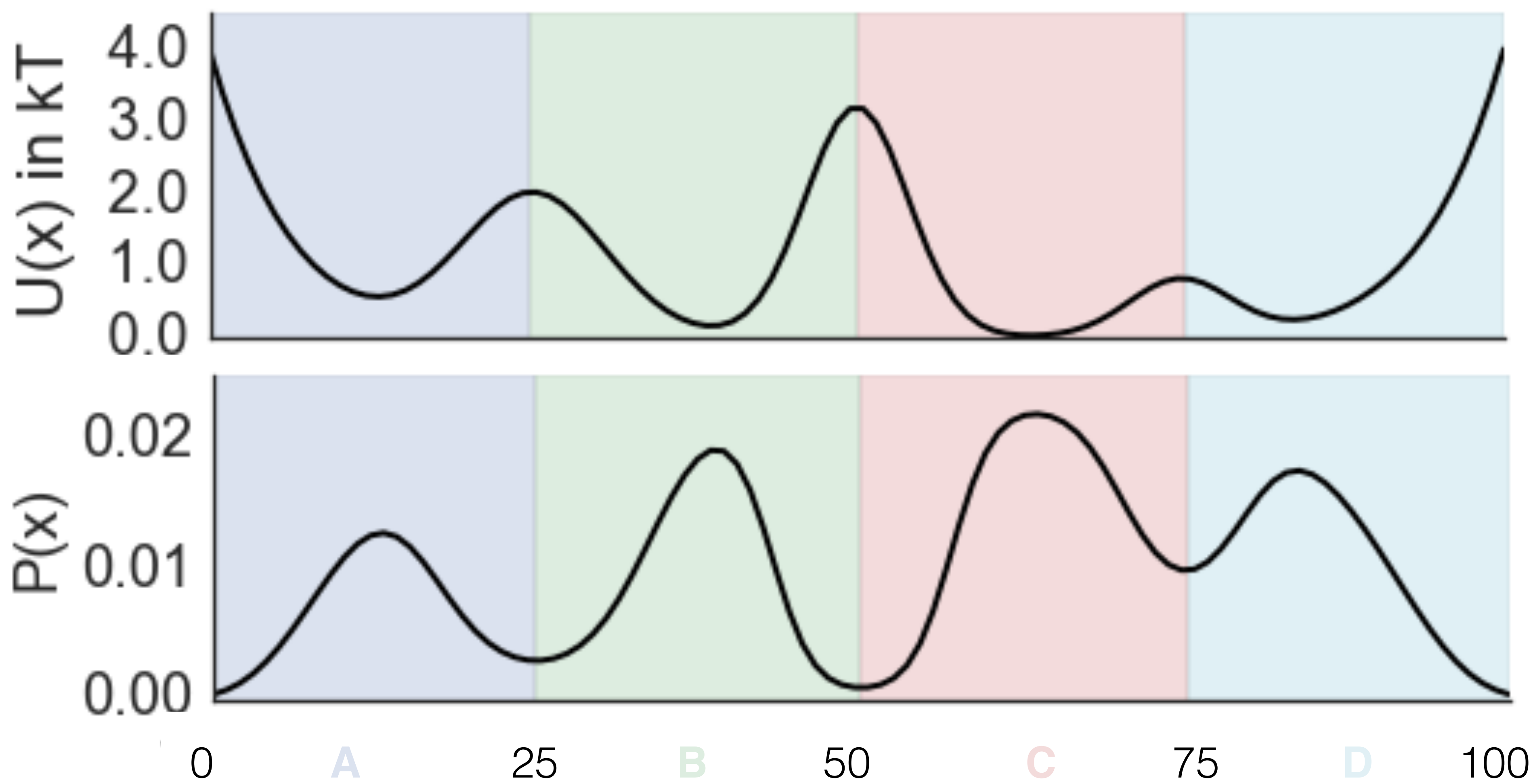
Transition matrix estimation

$$\begin{bmatrix} \frac{c_{ii}}{\sum_i c_{ij}} & \frac{c_{ij}}{\sum_i c_{ij}} \\ & \end{bmatrix} = \mathbf{T}$$

$$T_{ij}(\tau) = \mathbb{P}[\mathbf{x}(t+\tau) \in S_j | \mathbf{x}(t) \in S_i]$$

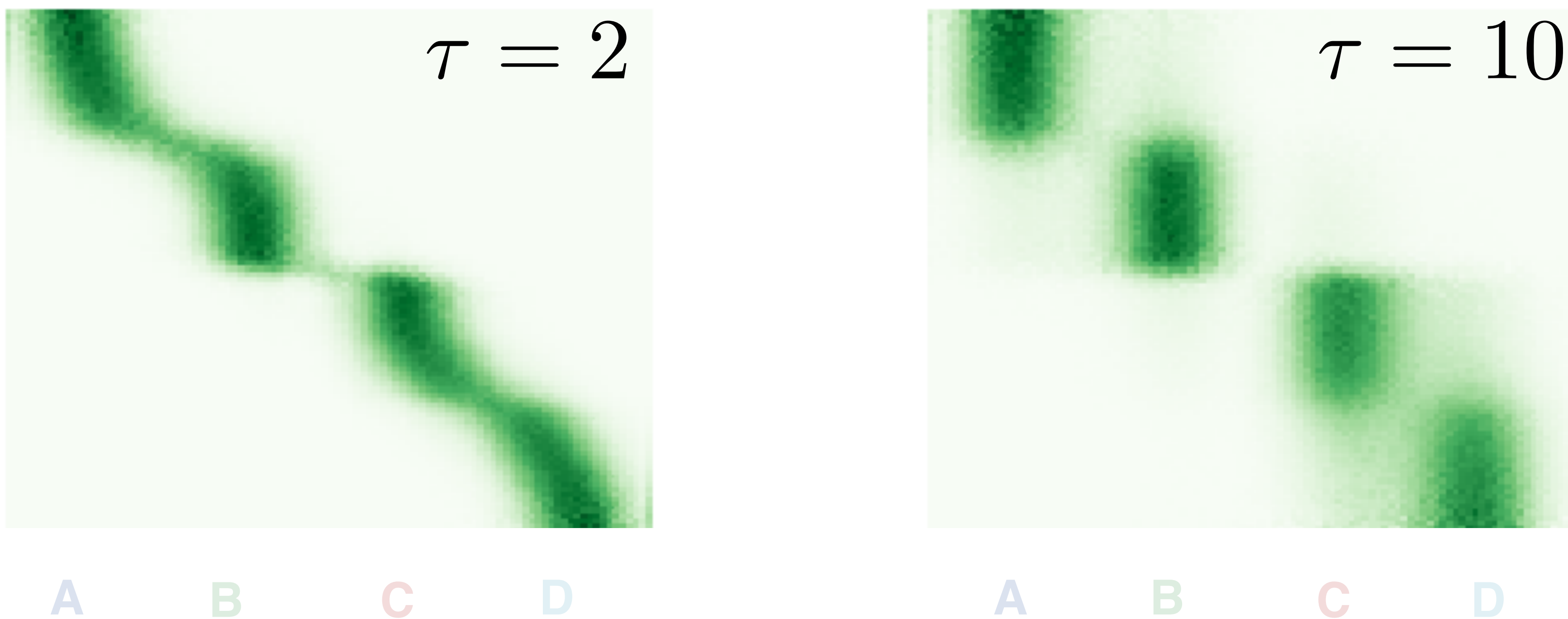
Obtaining error estimates on transition matrices and observables taken from them will be covered in more detail in the afternoon.

The Prinz potential



Varying the lagtime

$$T_{ij}(\tau) = \mathbb{P}[\mathbf{x}(t+\tau) \in S_j | \mathbf{x}(t) \in S_i]$$

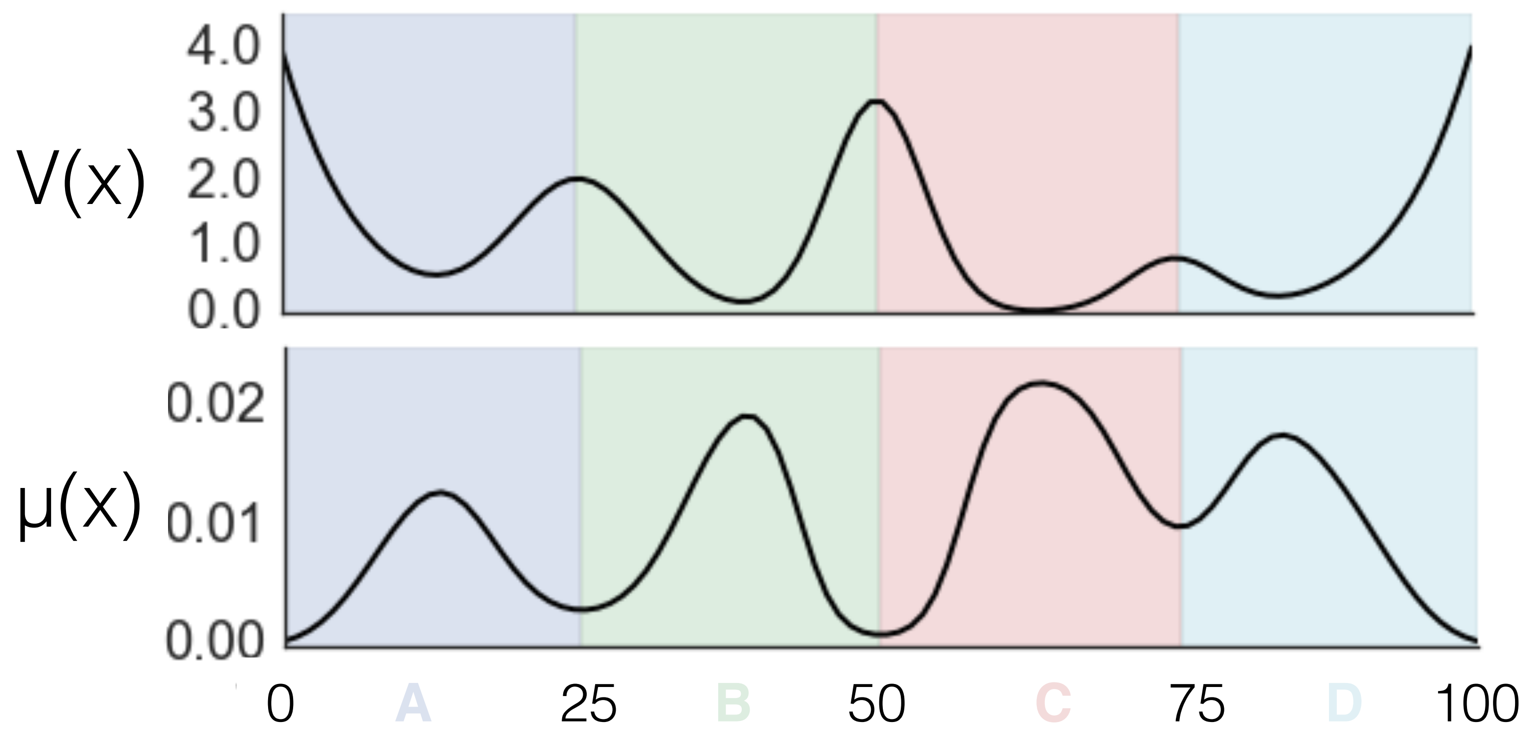


varying the lagtime

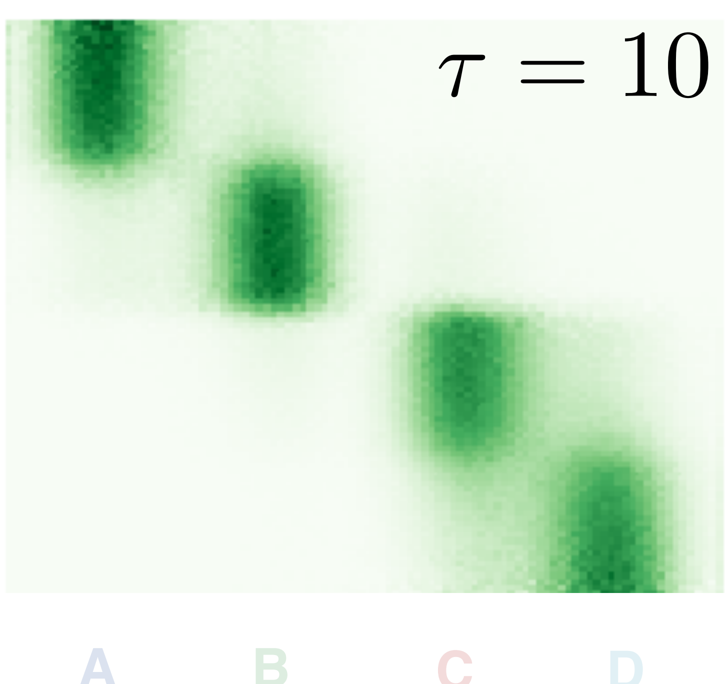
The transition matrix is:

- a stochastic matrix (rows sum to 1)
- has interesting properties that let us understand stationary and dynamic behaviour of the system

The transition matrix has lots of interesting properties



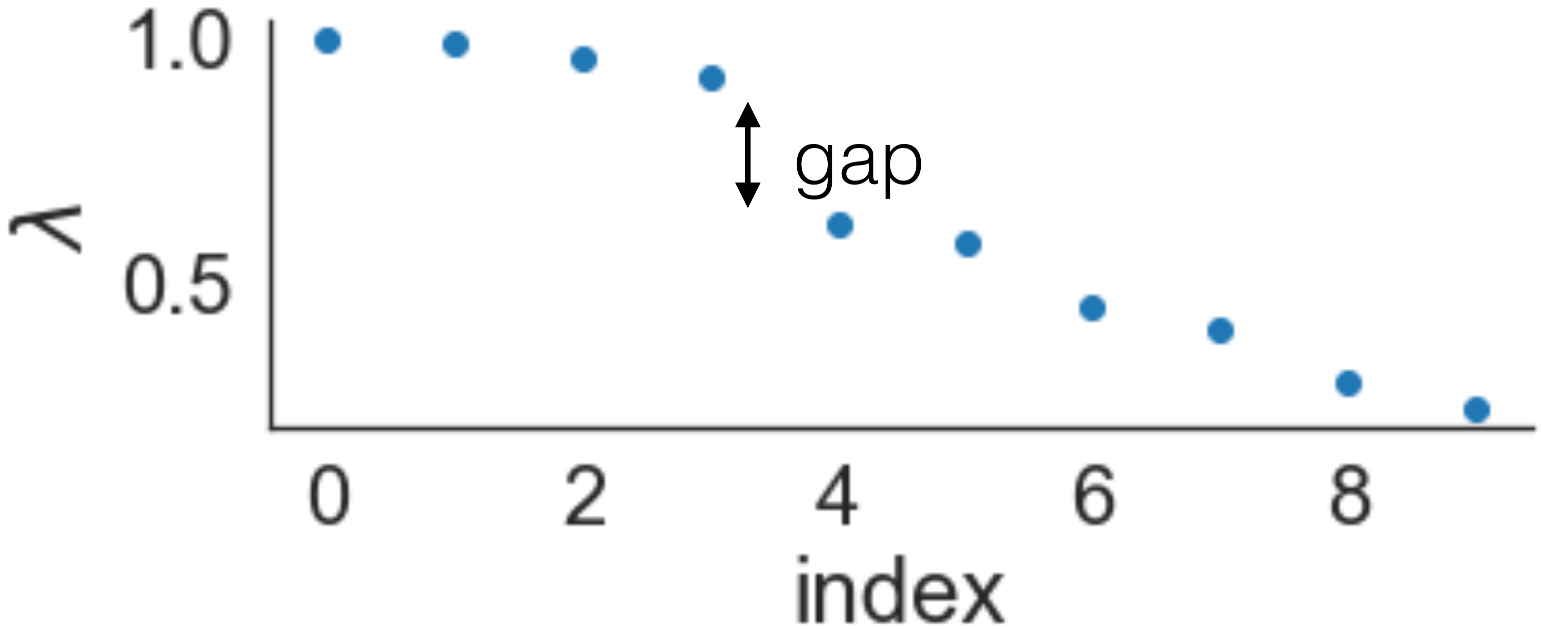
$$\lambda_1 = \boxed{1} > \boxed{\lambda_2 > \lambda_3, \ldots, > \lambda_n}$$



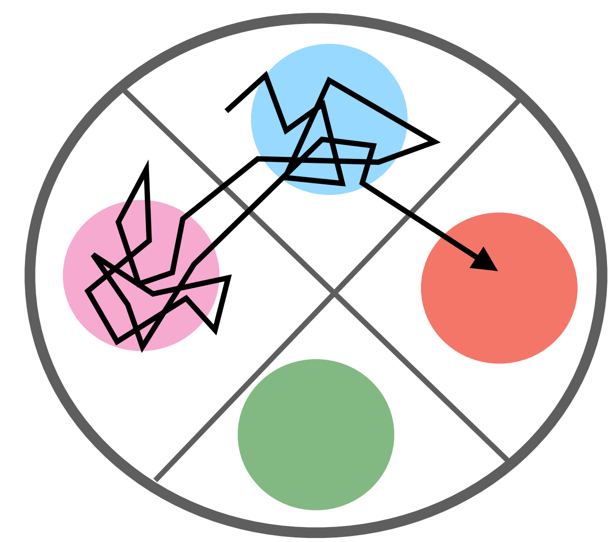
$T \; v = \lambda \; v$ right eigenvector

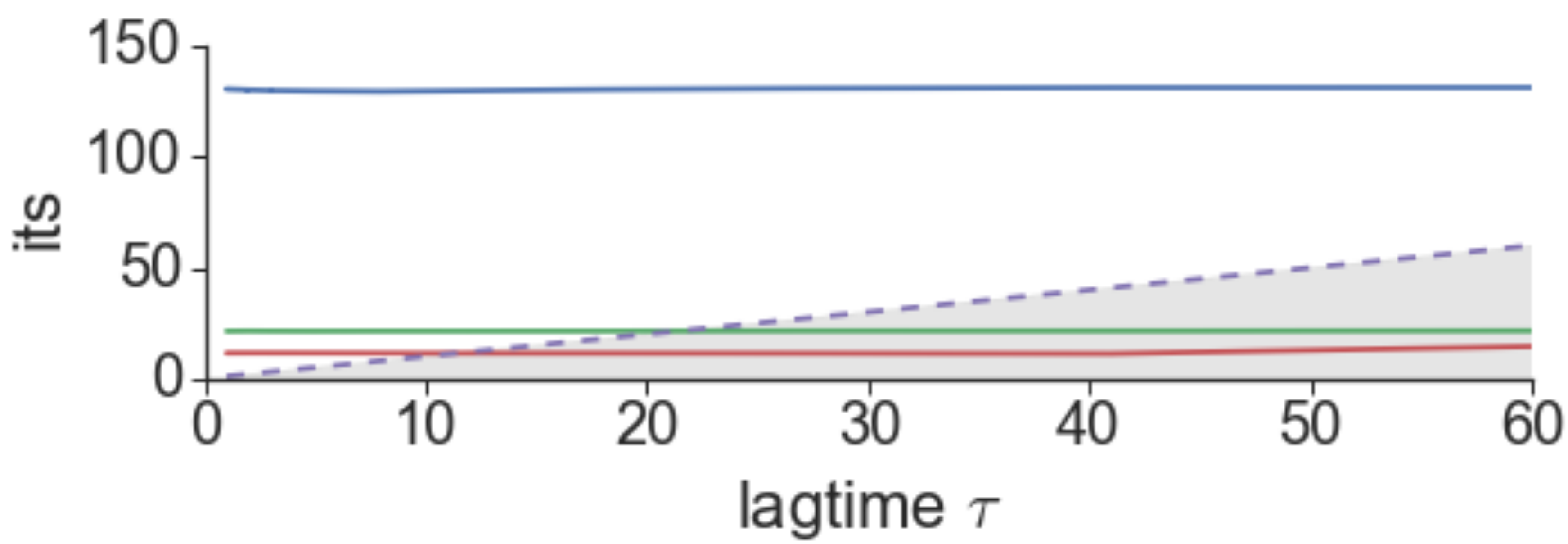
$T^T \; v = \lambda \; v$ left eigenvector

Obtaining timescales from the MSM



$$\lambda_1 = 1 > \boxed{\lambda_2 > \lambda_3, \ldots, > \lambda_n}$$

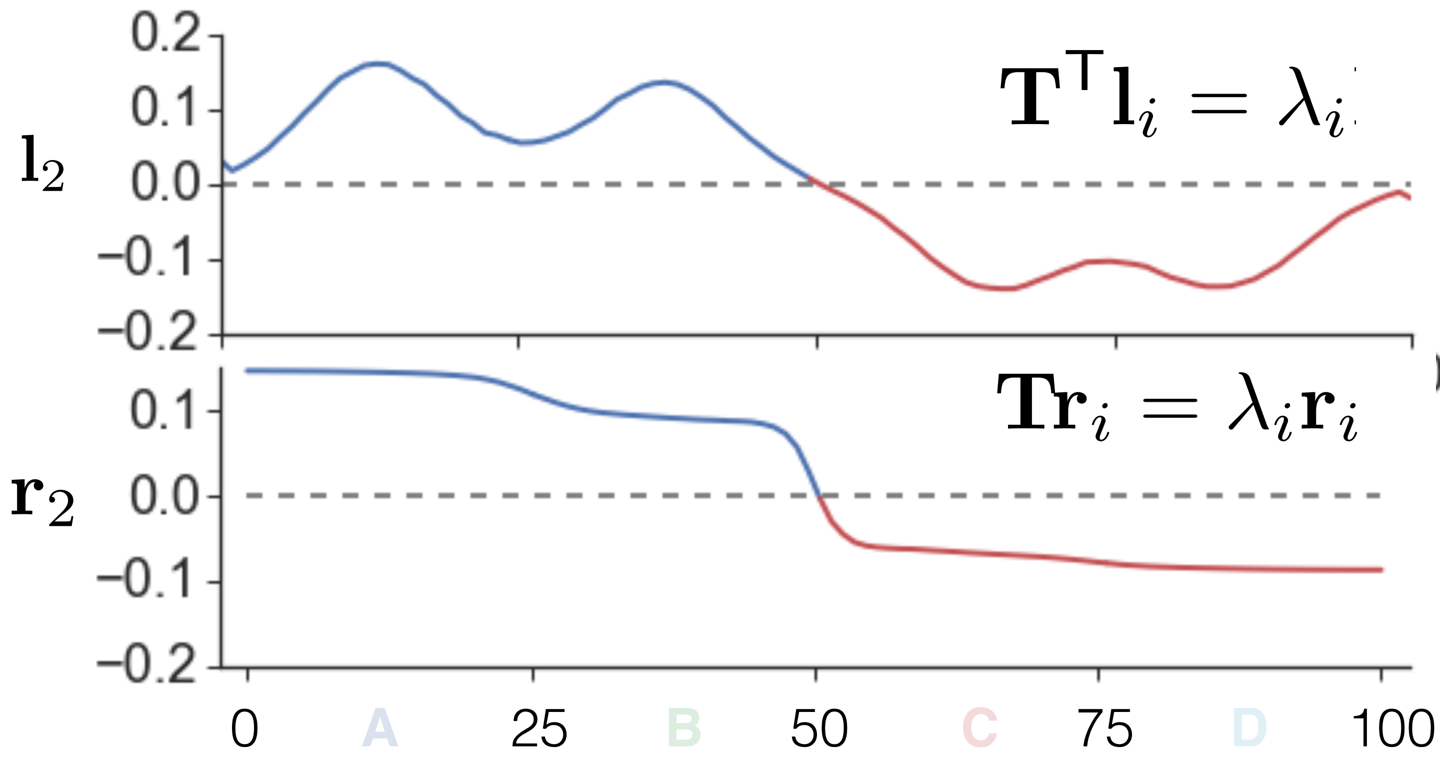


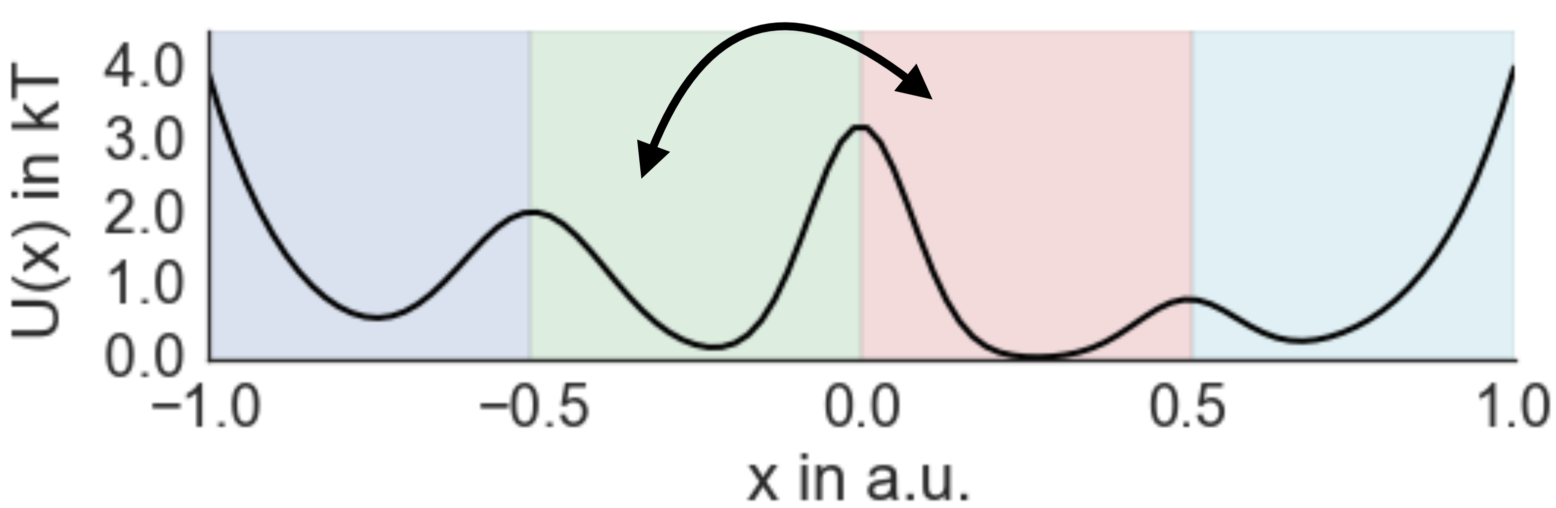


$$\boxed{t_i = -\frac{\tau}{\ln \lambda_i}}$$

Prinz, J.-H. et al (2011) Markov models of molecular kinetics: Generation and Validation. J. Chem. Phys., 134 . p. 174105.

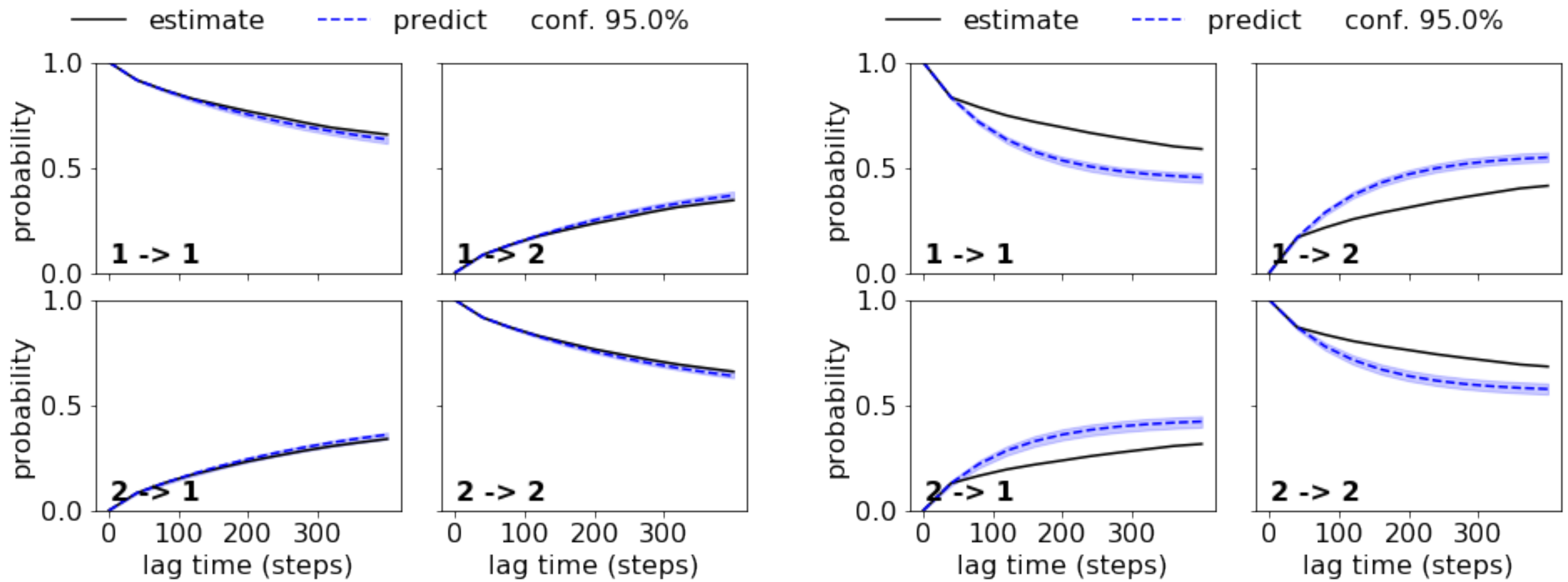
Identifying dynamic properties from the MSM





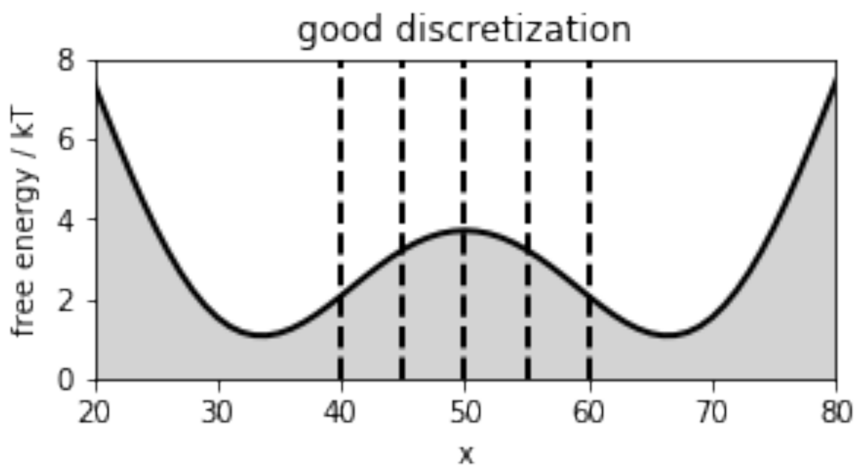
Validation with the Chapman Kolmogorov test

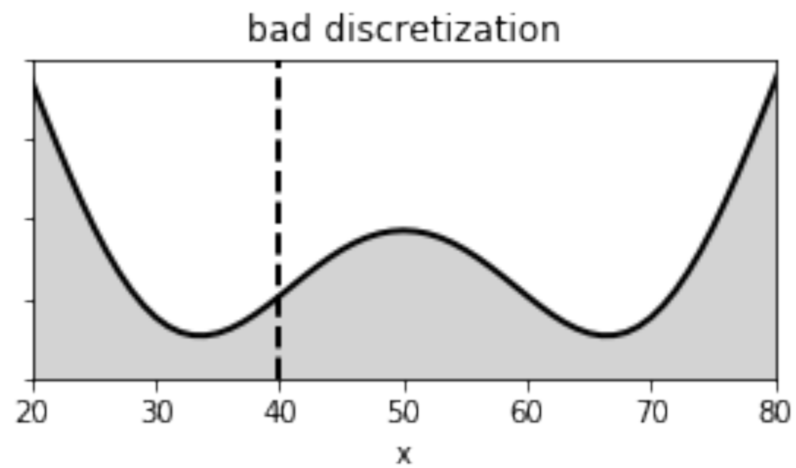
$$T(k\tau) = T(\tau)^k$$



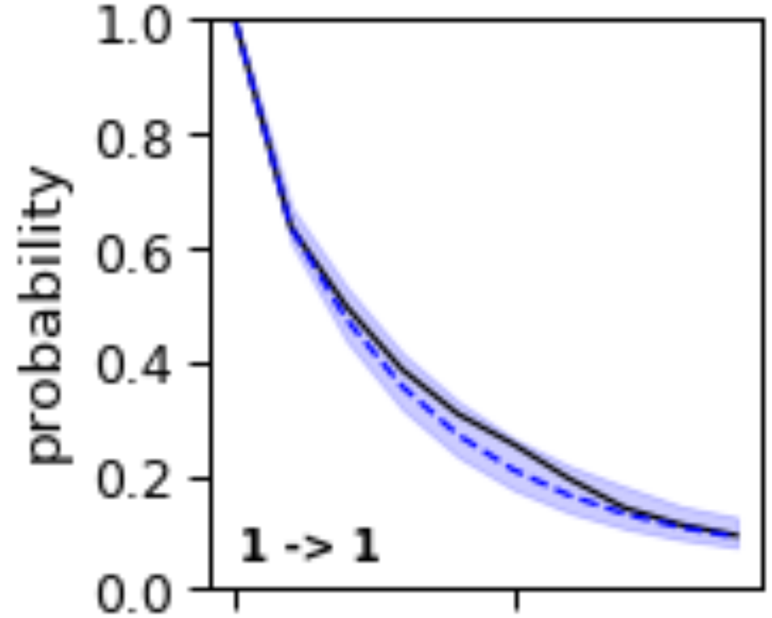
More validation

Build more than 1 MSM!



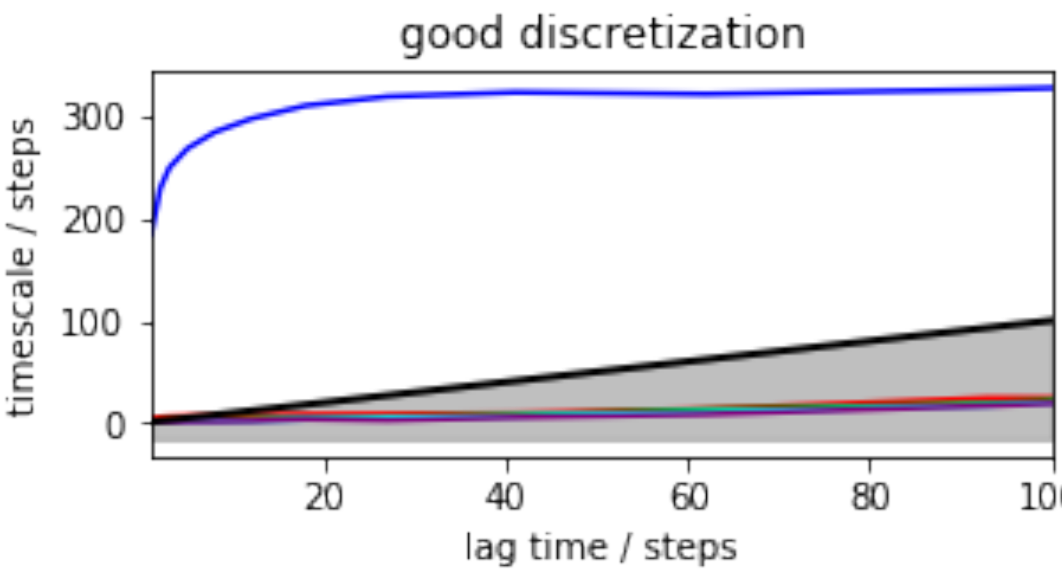


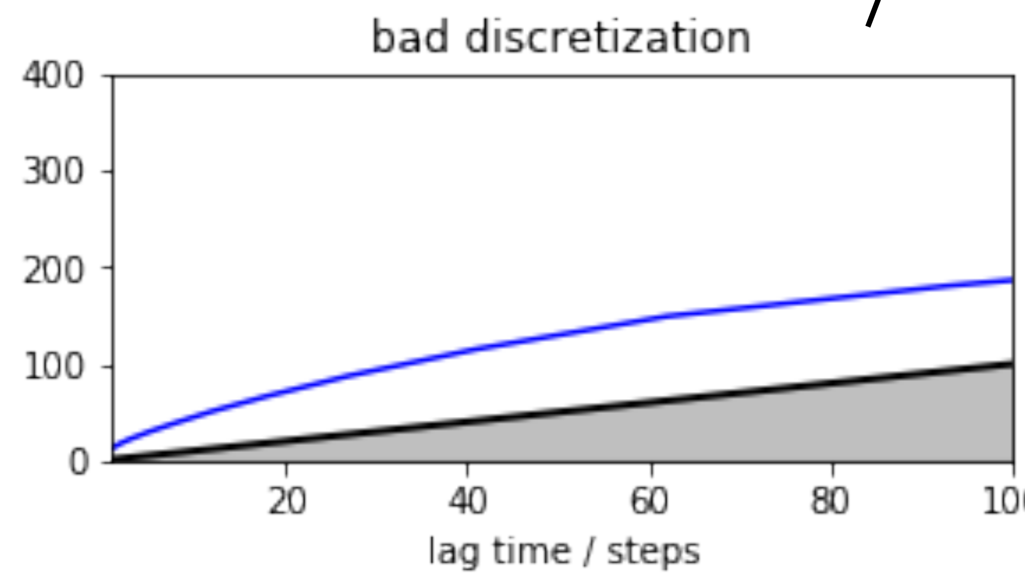
CK test!



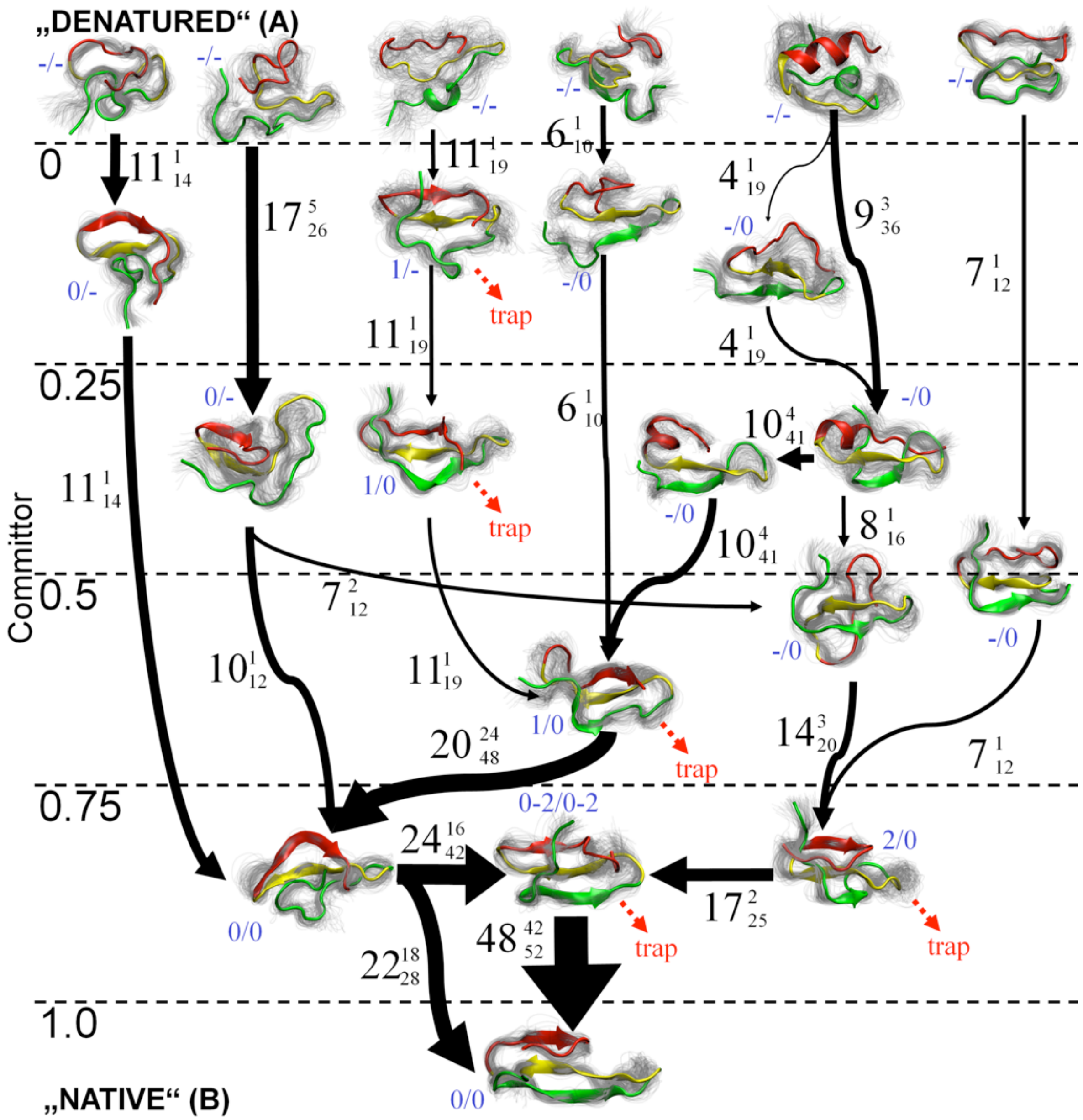
try and maximise your timescale!

?



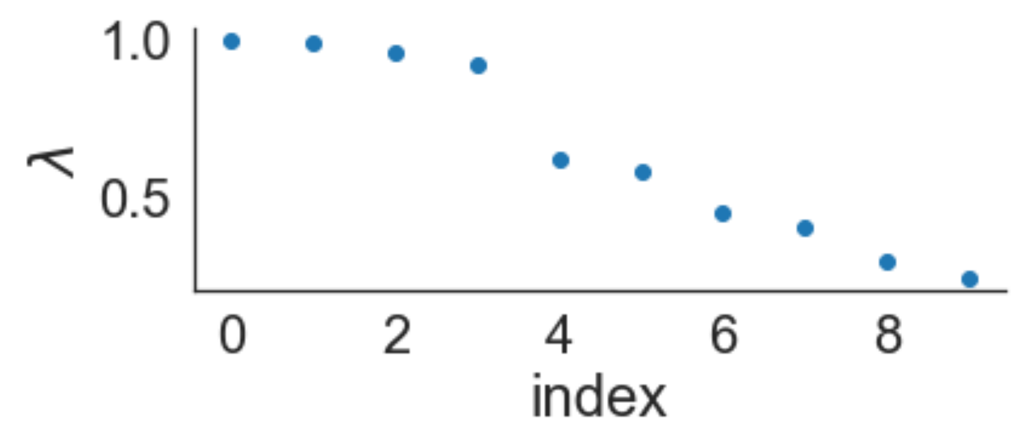


Coarse graining with PCCA



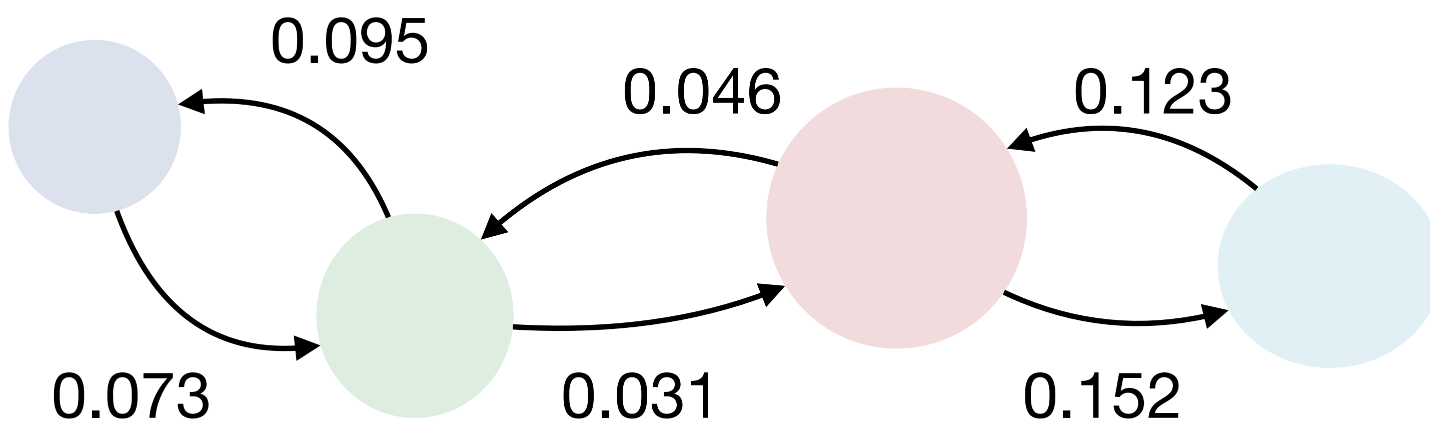
Noé et al. PNAS, 106, 19011 (2009)

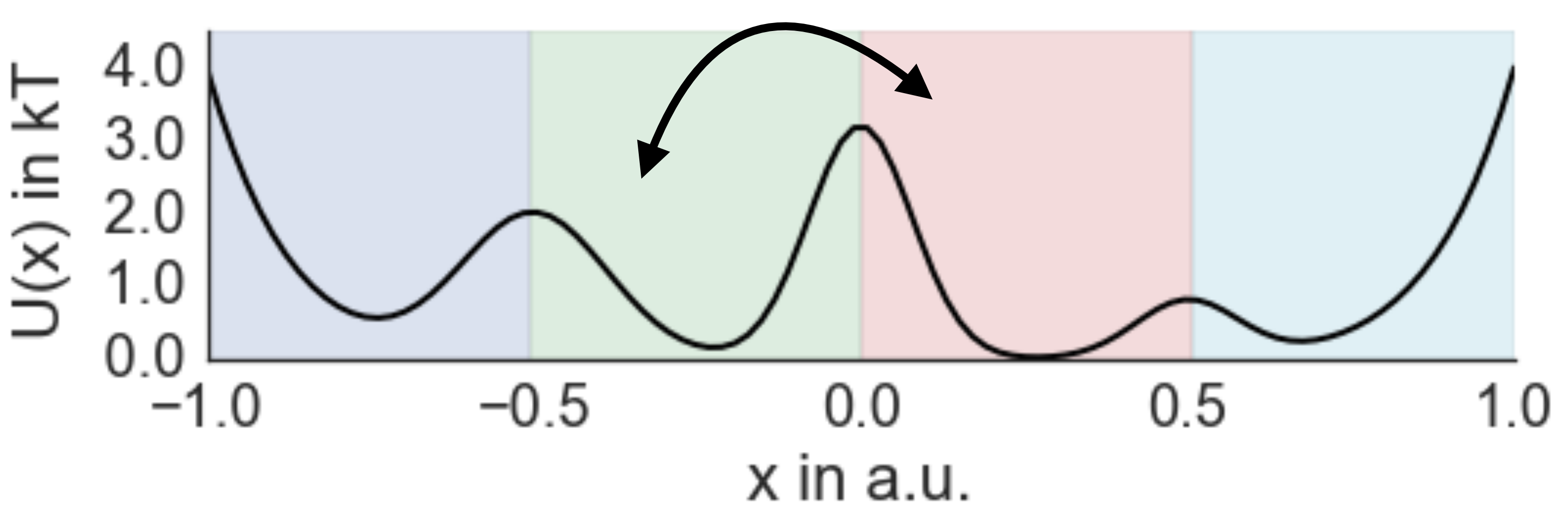
PCCA++



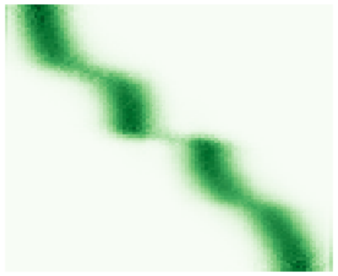
Or hidden Markov Models can be used for coarse graining

Transition paths and MFPT





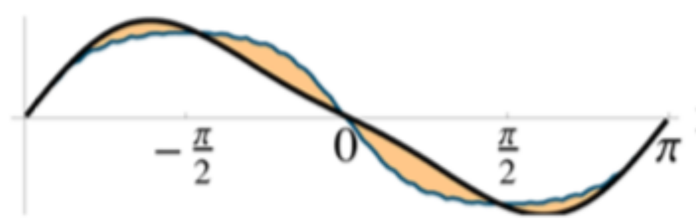
MSM what you should also know….

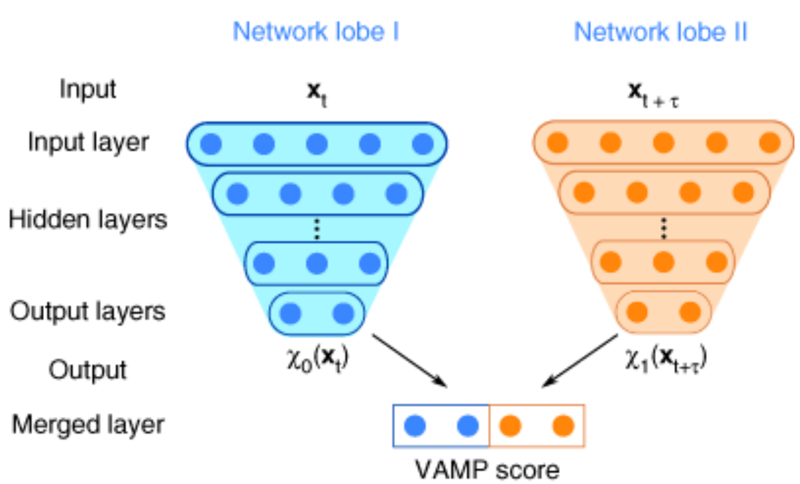


Reversible transition matrix estimator is one of many estimators that have been developed for estimating transition matrices, there are different/better approaches available

Variational approaches have been used to try and directly approximate the eigenfunctions of the propagator Q

Nüske, Mey, JCTC 10 (4), 1739-1752

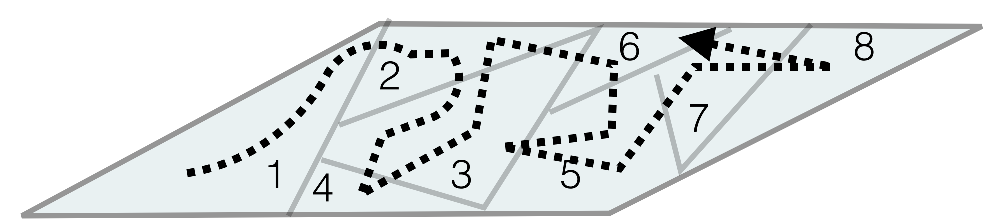




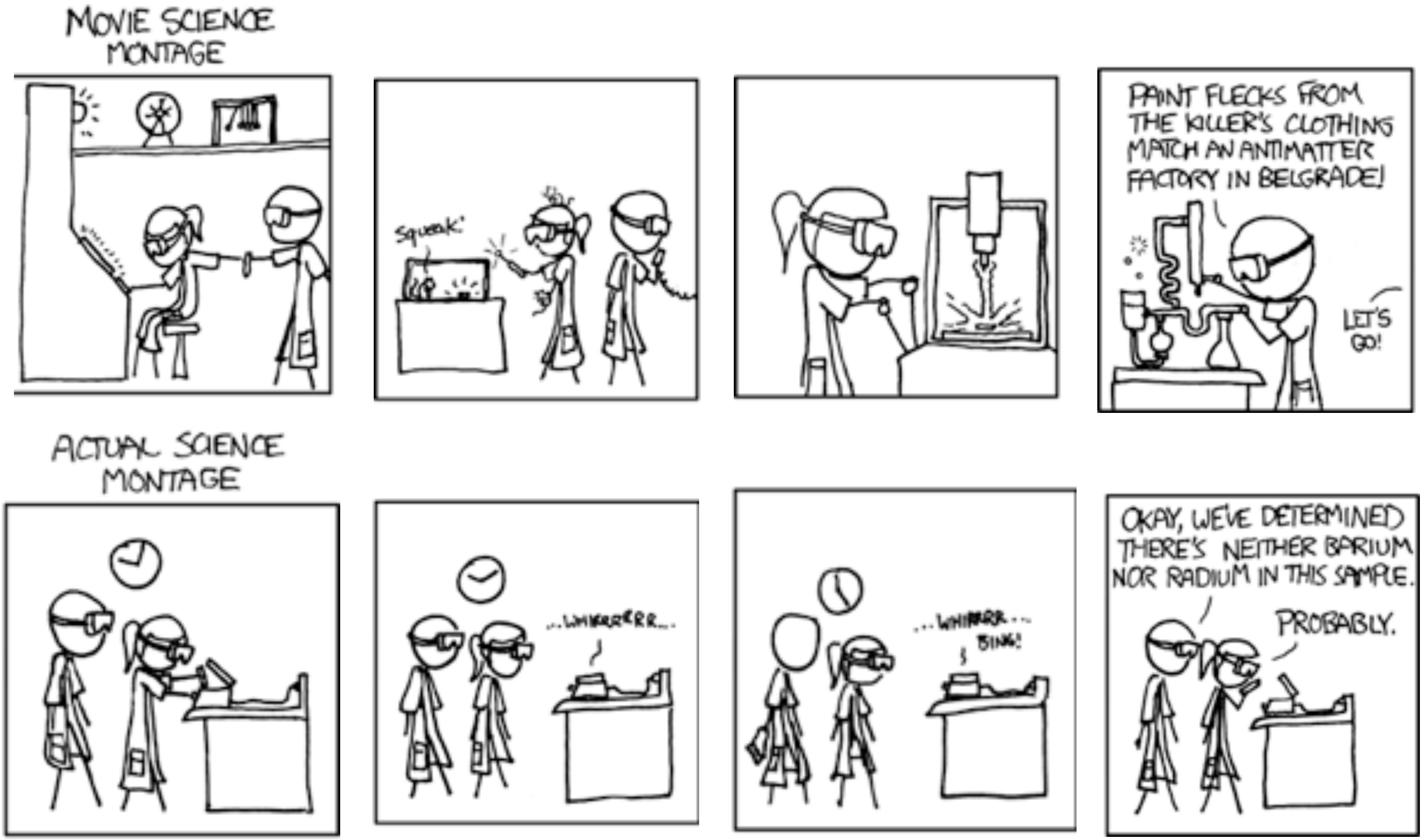
Neural networks can be used to learn and optimise MSMs

Mardt et al., Nature Communications ***9****, (2018)*

Dimensionality reduction and clustering of relevant data is still an open research problem.



Thank you


MOVIE SCIENCE MONTAGE
SQUEAK!
PAINT FLECKS FROM THE KILLER'S CLOTHING MATCH AN ANTIMATTER FACTORY IN BELGRADE!
LET'S GO!
ACTUAL SCIENCE MONTAGE
...WHIRRRRRR...
...WHIRRRR...
BING!
OKAY, WE'VE DETERMINED THERE'S NEITHER BARIUM NOR RADIUM IN THIS SAMPLE.
PROBABLY.