

# (DEEP?) ENCODING OF MOLECULAR DYNAMICS

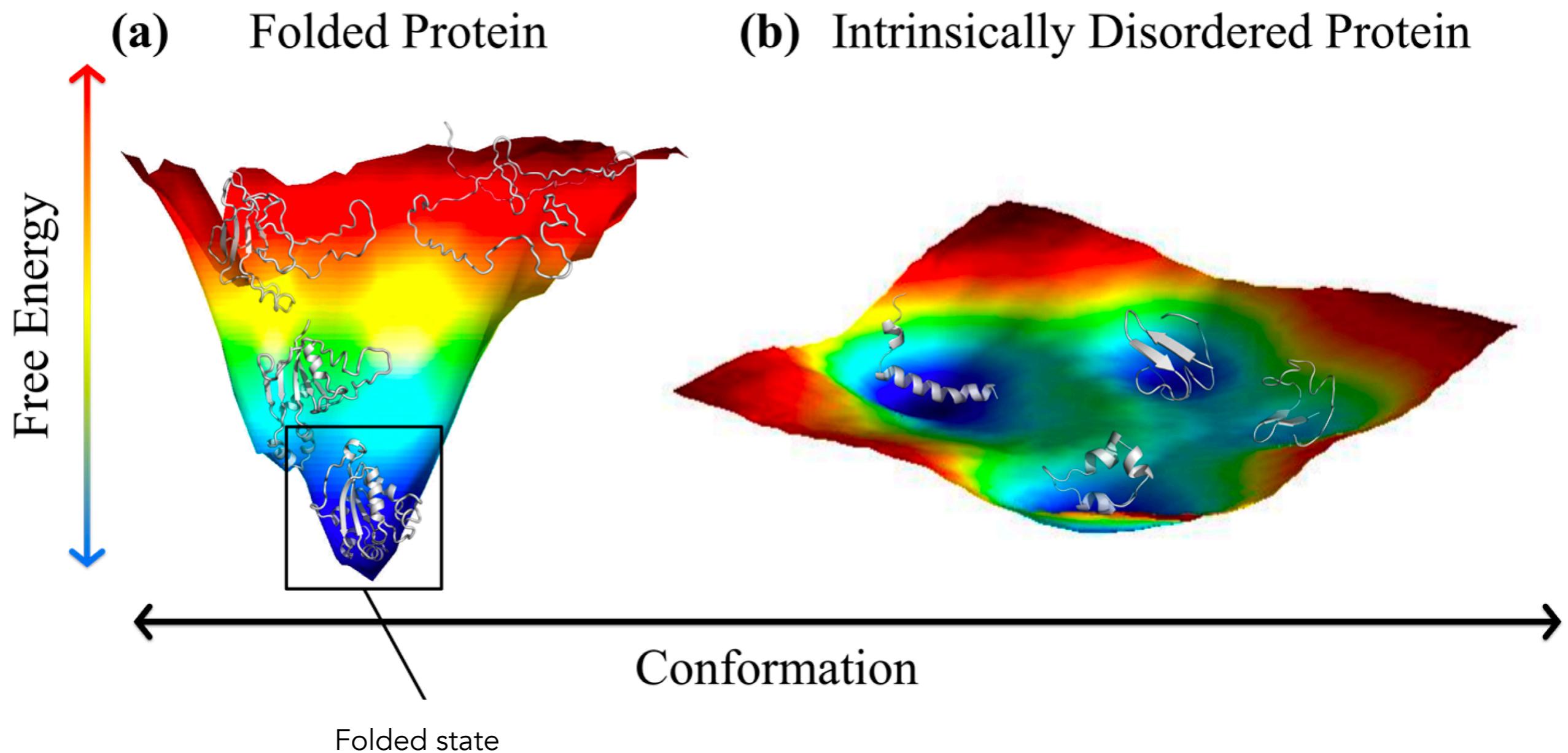
Eugene Klyshko

January 31, 2020

# Outline

1. Proteins
2. Molecular Dynamics Simulations
3. Markov State Models (MSMs) as encoding of dynamics
4. Variational Approach to Conformational Dynamics
5. MSMs revisited
6. VDE - Variational Dynamics Encoder NN

# Proteins



# Molecular Dynamics simulations – “computational microscope”

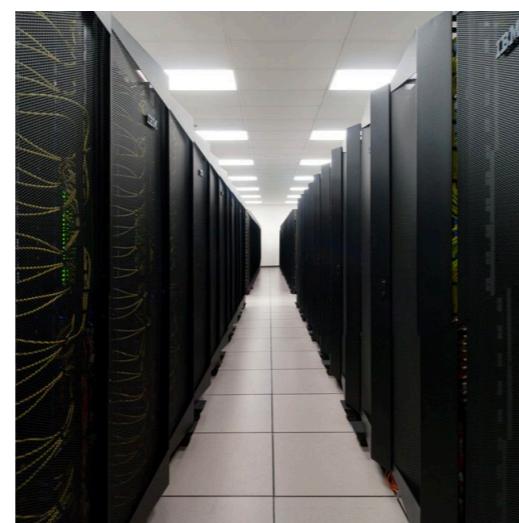
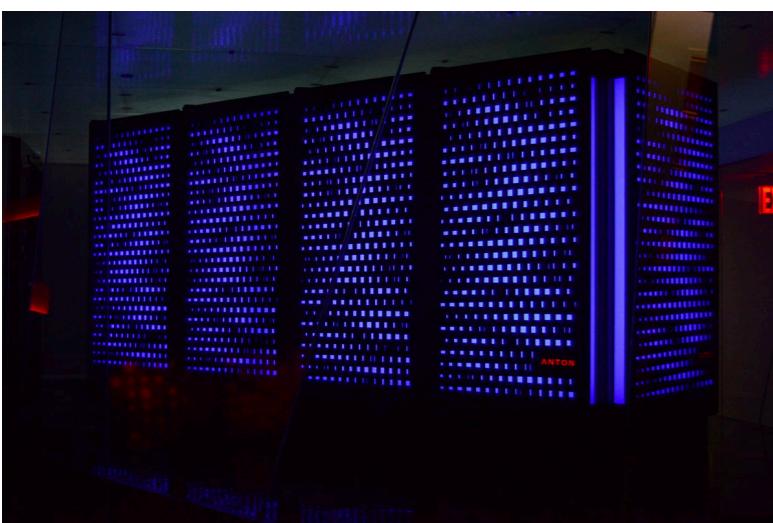
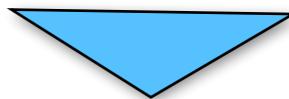
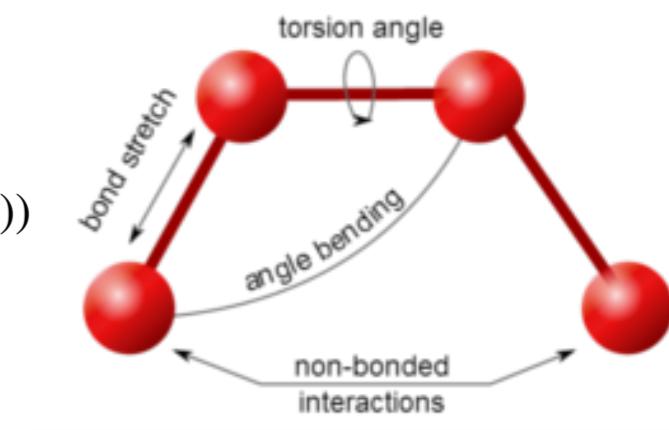
Numerical integration of equations of motion for every atom in the system (protein + solvent + ions):

$$m_i \frac{d^2 r_i(t)}{dt^2} = \sum_j F_{ij}(t) = - \sum_j \nabla_i V(r_{ij}(t))$$

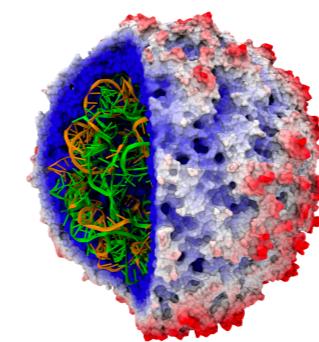
based on empirical potential energy function:  $V = V_b + V_{nb}$

bonded term:  $V_b = \sum_{bonds} \frac{1}{2} K_b (b - b_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{torsions} K_\phi (1 - \cos(n\phi - \phi_0))$

non-bonded term:  $V_{nb} = \sum_{i,j} \left( \frac{q_i q_j}{4\pi\epsilon_0 \epsilon r_{ij}} + \epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \right)$



**GROMACS**  
**FAST. FLEXIBLE. FREE.**



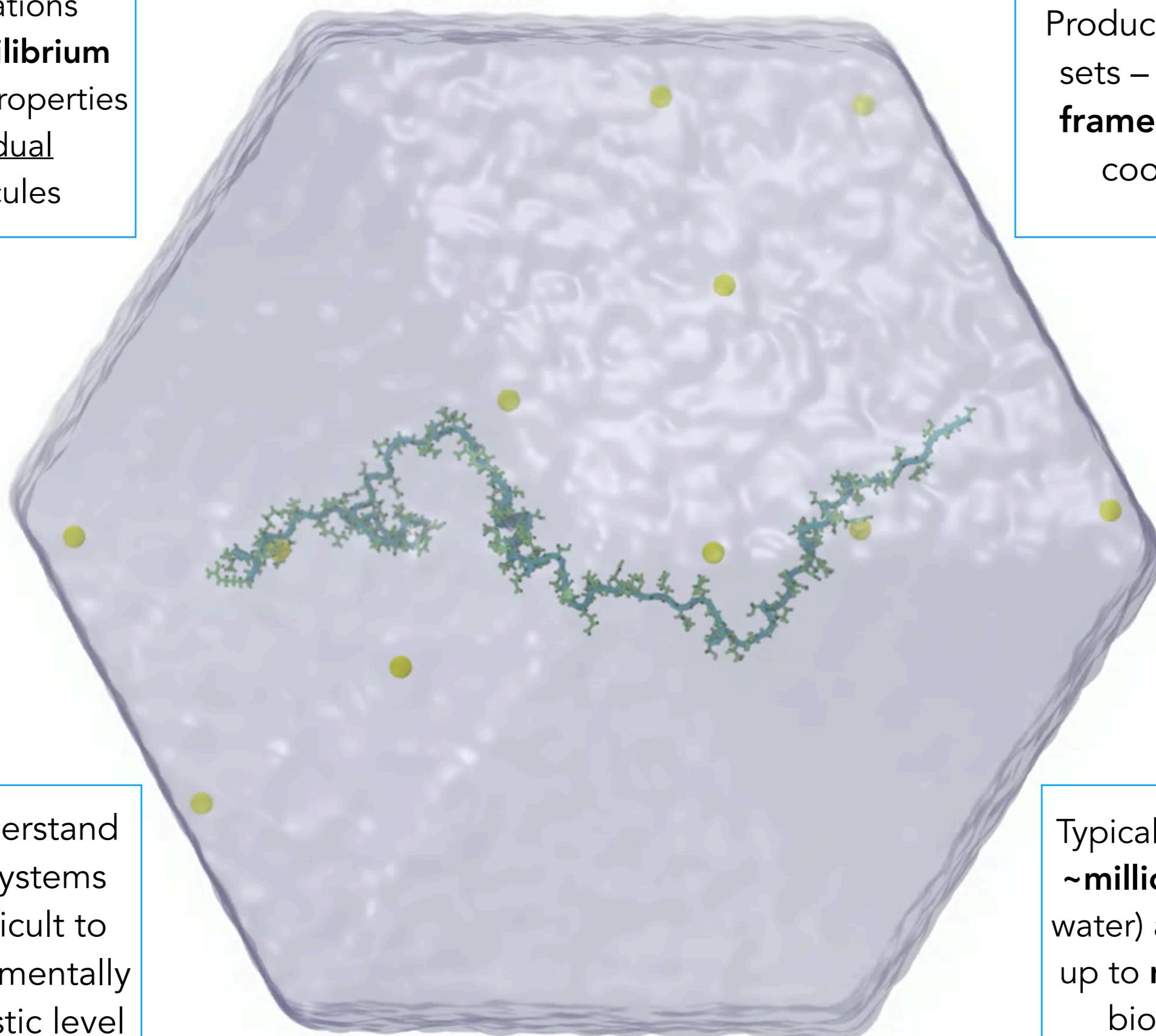
# MD simulations

MD simulations capture **equilibrium** and **kinetic** properties of individual biomolecules

Produce large data sets – **millions of frames** of atomic coordinates

Used to understand molecular systems that are difficult to probe experimentally at the atomistic level

Typically systems with ~**million atoms** (90% water) and can sample up to **milliseconds** of biological time



20 ns of the MD of Sic1

# Vanilla Markov State Models (MSMs)

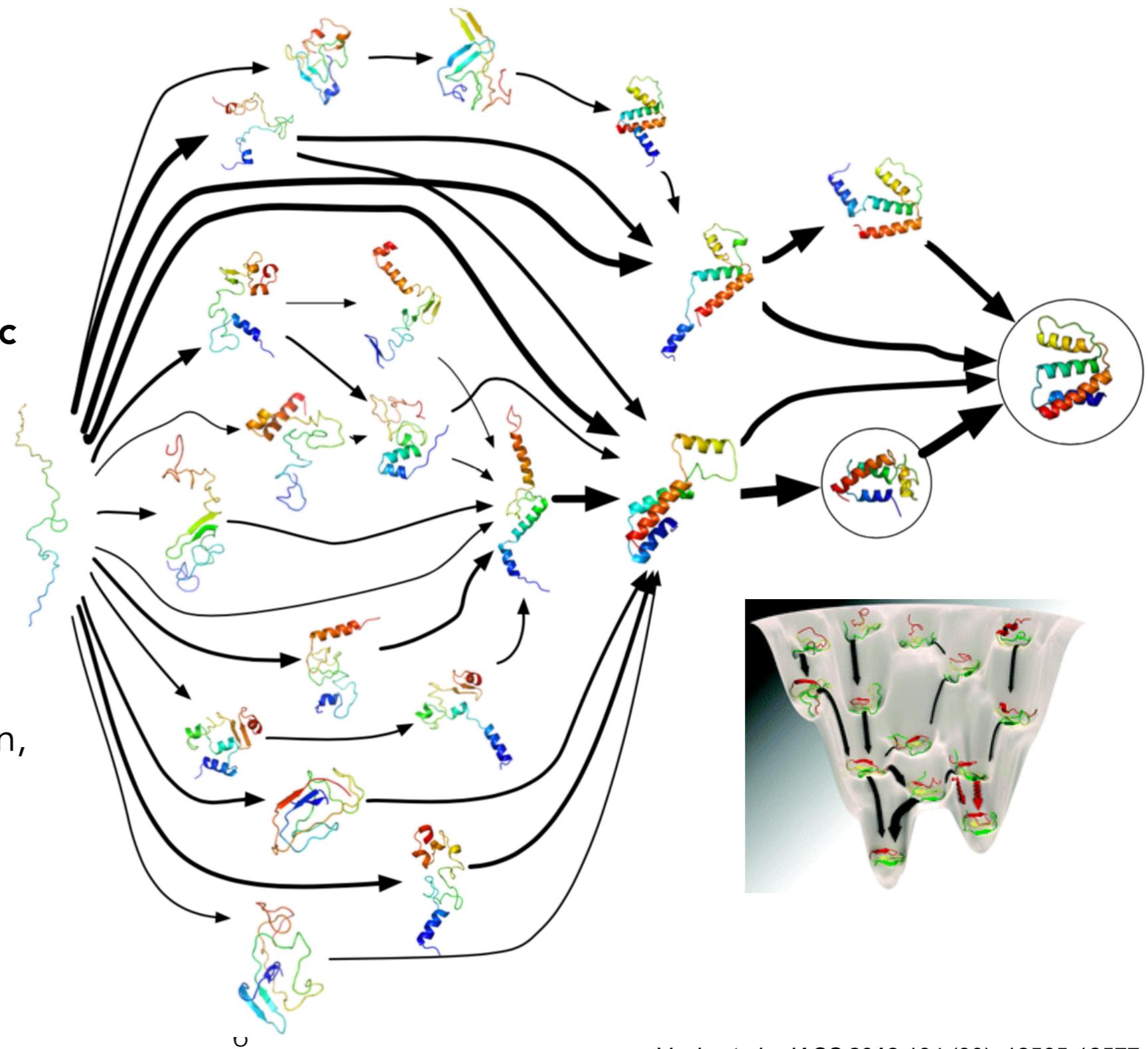
“States and rates” approach is more *intuitive* than the large number of 3D-structures generated by MD simulations

- ▶ Master equation approach to describe conformational dynamics
- ▶ Network of **conformational states** and a **transition probability** matrix
- ▶ An MSM contain all **thermodynamic** and **kinetic** information

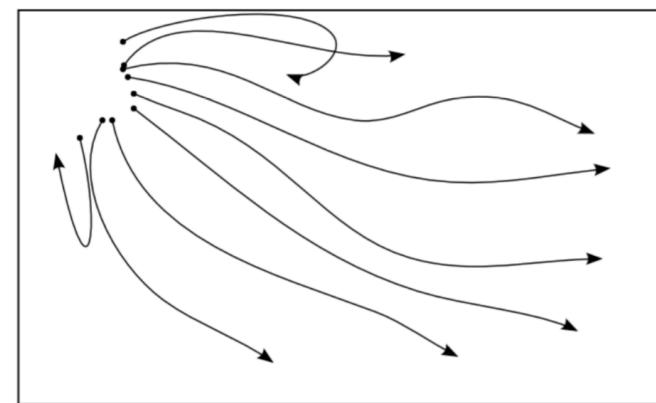
**Markov State Model  
of protein folding**



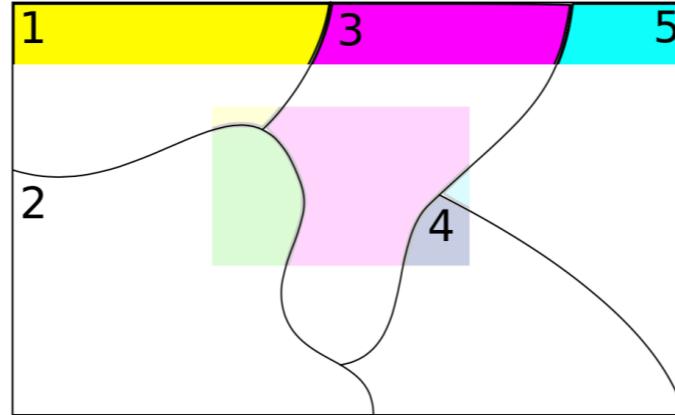
- ▶ Assumption: all states are Markovian, i.e. “memoryless”
- ▶ Reflects the free energy landscape (states are energy minima)



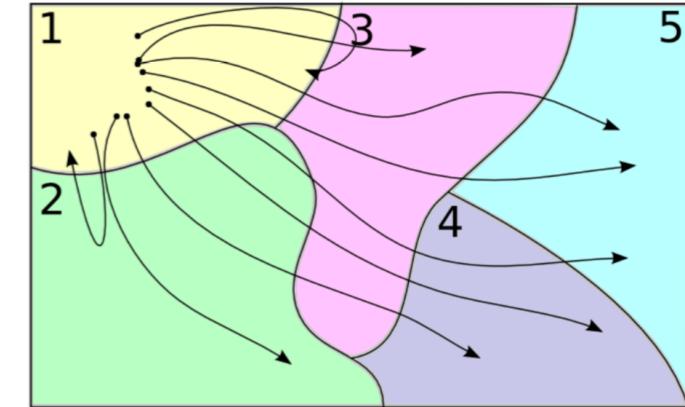
# Markov State Models - the pipeline



## 1. Obtain MD trajectories

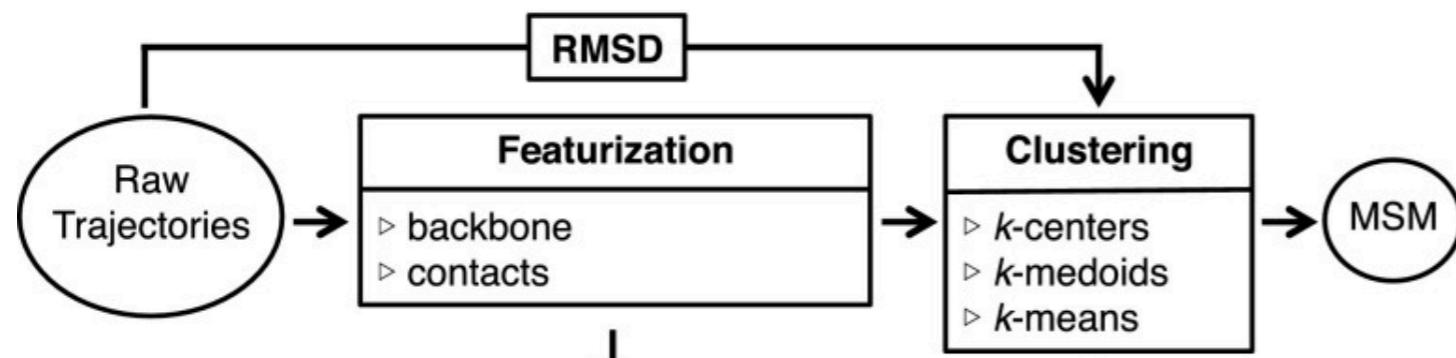


## 2. State decomposition



## 3. Transition matrix estimation

Suppose we slice our trajectories every  $\Delta t$  (lagtime) and count the observed transitions:



$$C_{ij} = C_{i \rightarrow j}$$

To get the transition probabilities, we simply “normalize” the counts:

$$T_{ij} = T_{i \rightarrow j} = \frac{C_{ij}}{\sum_k C_{ik}}$$

# MSM transition matrix $\mathbf{T}$

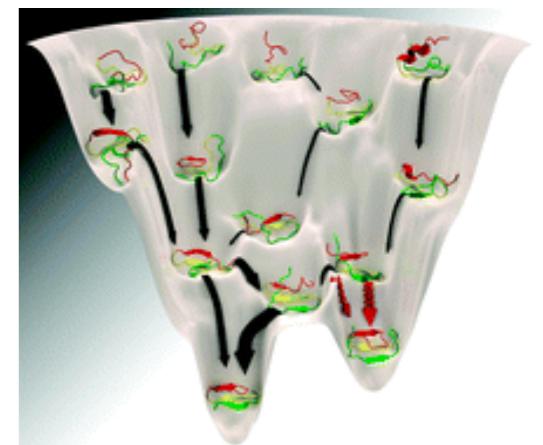
$$T\nu = \lambda\nu$$

Setting  $\lambda = 1$  gives the equilibrium populations:

$$T\pi = 1\pi = \pi$$

At long times system approaches equilibrium population:

$$x(t) \rightarrow \pi$$

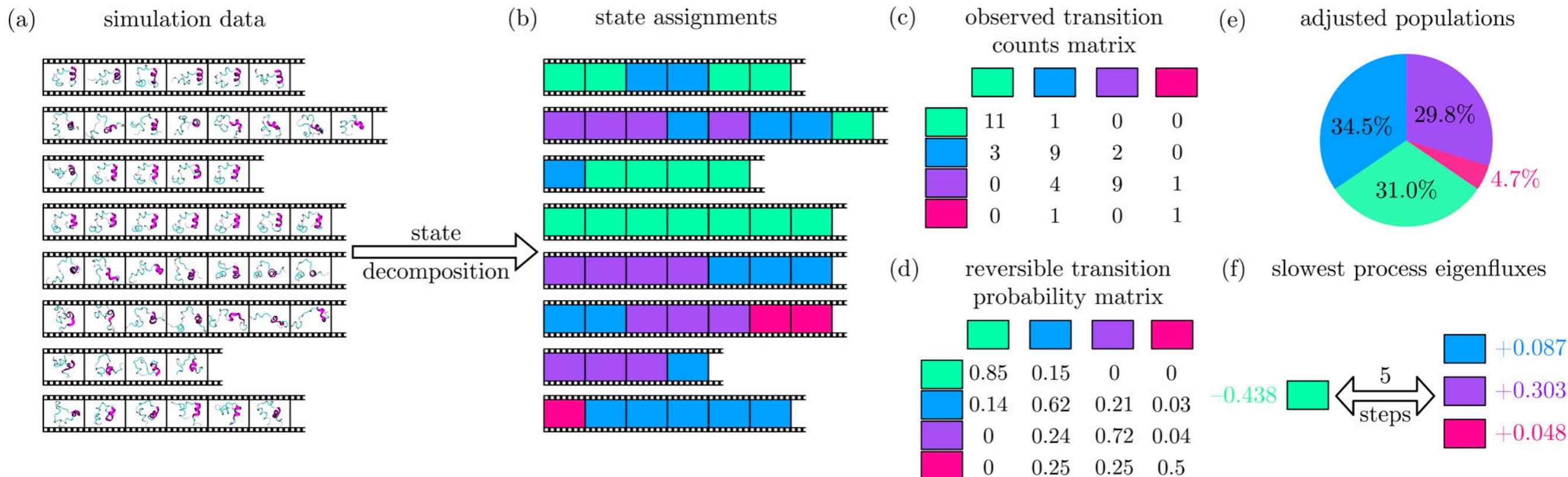


For the remaining eigenvalues,  $\lambda_i < 1$

These eigenvalues correspond to **characteristic timescales** at which different populations approach equilibrium (the slowest processes in the system)

$$\tau_i = -\frac{\Delta t}{\ln(\lambda_i)}$$

# Markov State Model framework

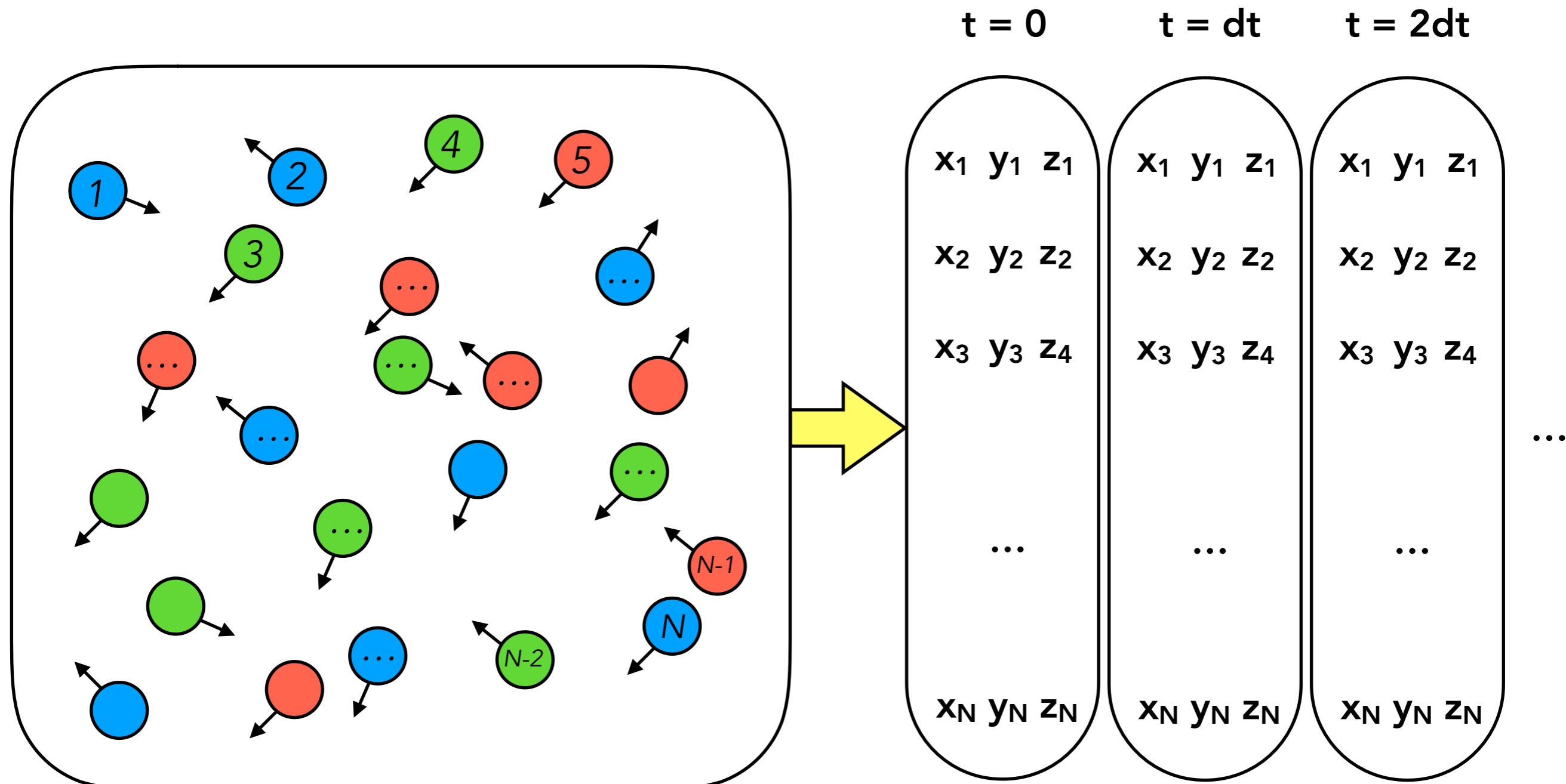


# Variational Approach to Conformational Dynamics formalizes MSM

In 2013, the derivation of a variational approach to conformational dynamics (VAC) showed that **estimates of MSM eigenvalues cannot exceed their true values**. Thus, the choice of MSM states can be optimized according to this variational principle.

In fact, the VAC applies in a more general case than MSMs: the eigenfunction approximations need not come from the discrete state decomposition that characterizes a MSM, but rather can come from other features; in the case of protein dynamics, these features might represent torsional angles or pairwise distances between amino acids.

# Data: coordinates of interacting particles



# Dynamical propagator theory

- Conformational space  $\mathbf{X}$  of a molecule consisting of  $\mathbf{N}$  atoms: the  $3\mathbf{N}$ -dimensional space
- The conformational dynamics of the molecule in this space can be represented by a dynamical process  $\{\mathbf{x}_t\}$ , which samples at a given time  $t$  a particular point  $\mathbf{x}_t \in \mathbf{X}$ .
- $\mathbf{x}_t$  is called a trajectory
- This process is governed by the equations of motion (simulated using MD).
- MD ensures that  $\mathbf{x}_t$  is time-homogeneous, Markovian, ergodic, and reversible with respect to a unique stationary density (usually the Boltzmann distribution).
- An infinite ensemble of molecules of the same type, distributed in the conformational space according to some initial probability density  $|p_0(x)>$ .

# Dynamical propagator theory

This initial probability density evolves in time in a definite manner that is determined by the equations of motion for the individual molecules.

We can define a transition density that is independent of time:

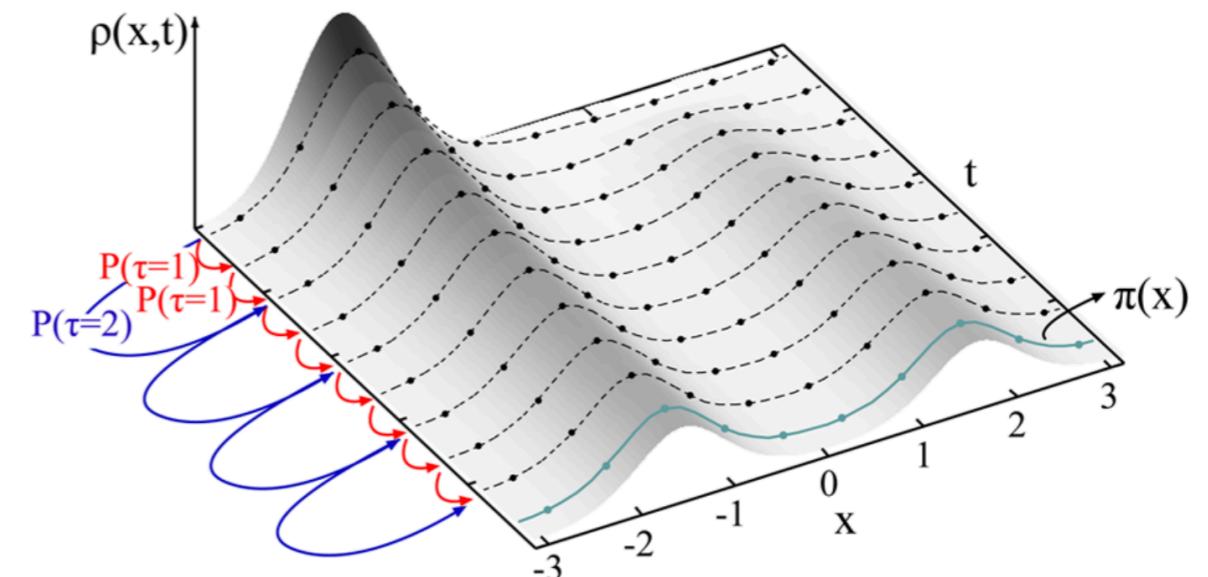
$$p(x, y; \tau) dy = Prob(x_{t+\tau} \in y dy | x_t = x) = Prob(x_\tau \in y dy | x_0 = x)$$

The unconditional probability density of finding a molecule in conformation  $\mathbf{y}$  at time  $t + \tau$  is obtained by integrating over all starting conformations  $\mathbf{x}$ :

$$\rho_{t+\tau}(y) = \int_X p(x, y; \tau) \rho_t(x) dx$$

This equation, in fact, defines an operator  $P(\tau)$  that propagates the probability density by a finite time step  $\tau$  - called lag time. Operator is called propagator:

$$|\rho_{t+\tau}(x) \rangle = P(\tau) |\rho_t(x) \rangle$$
$$|\rho_{t+n\tau}(x) \rangle = P(\tau)^n |\rho_t(x) \rangle$$



# Eigenspectrum

The way the propagator acts on the density can be understood in terms of its eigenfunctions  $\{|l_\alpha(x)\rangle\}$  and associated eigenvalues  $\{\lambda_\alpha\}$ , which are defined by the following eigenvalue equation:

$$P(\tau) |l_\alpha(x)\rangle = \lambda_\alpha |l_\alpha(x)\rangle$$

The eigenfunctions form a complete set of  $3N$ -dim space. Hence, any probability density (in fact any function) in this space can be expressed as a linear combination of  $\{|l_\alpha(x)\rangle\}$ .

$$|\rho_{t+n\tau}(x)\rangle = c_\alpha \lambda_\alpha^n |l_\alpha(x)\rangle = c_\alpha e^{-n\tau/t_\alpha} |l_\alpha(x)\rangle$$

The eigenfunctions can be interpreted as kinetic processes that transport probability density from one part of the conformational space to another and thus modulate the shape of the overall probability density.

The eigenvalues are linked to the time scales  $t_\alpha$  on which the associated kinetic processes take place by

$$t_\alpha = -\frac{\tau}{\ln(\lambda_\alpha)}$$

# Properties of the propagator

1. The propagator has a unique stationary distribution  $\pi(x)$ , corresponding to eigenvalue 1

$$P(\tau) |\pi(x)\rangle = |\pi(x)\rangle$$

2. Its eigenvalue spectrum is bounded from above by  $\lambda_1 = 1$ . Also,  $\lambda_1$  is the only eigenvalue of absolute value equal to one

$$\lambda_\alpha < \lambda_1 = 1$$

3. Operator  $P(\tau)$  is self-adjoint with respect to the weighted scalar product

$$\langle f | g \rangle_{\pi^{-1}} = \int_X f(x)g(x)\pi^{-1}(x)dx$$

Consequently, its eigenfunctions  $|l_\alpha(x)\rangle$  form an orthonormal basis of the Hilbert space of square-integrable functions with respect to this scalar product. Its eigenvalues are real and can be numbered in descending order:

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \dots$$

# Variational Principle in 3 steps

A variational principle can be derived for any operator whose eigenvalue spectrum is bound (either from above or from below) and whose eigenvectors form a complete basis set and are orthonormal with respect to a given scalar product

**Step 1.** For the exact eigenvector holds (analogous to QM expectation value of the Hamiltonian - has an interpretation of the time-lagged autocorrelation):

$$\langle l_\alpha | P(\tau) | l_\alpha \rangle_{\pi^{-1}} = \lambda_\alpha(\tau) = e^{-\tau/t_\alpha}$$

**Step 2.** For any trial function  $f$  that is normalized holds true. Equal sign is if and only if  $f(x) = l_1(x)$

$$\langle f | P(\tau) | f \rangle_{\pi^{-1}} = \int_X f(x) \pi^{-1}(x) P(\tau) f(x) dx \leq \lambda_1 = 1 \quad (*)$$

**Step 3.** Same applies to all other eigenvectors, one by one. The trial function  $|f\rangle$  is linearly expanded using a basis of  $n$  basis functions  $\{|l_i\rangle\}_{i=1}^n$

$$\langle f | P(\tau) | f \rangle_{\pi^{-1}} \leq \lambda_\alpha$$

$$\langle f | l_\beta \rangle_{\pi^{-1}} = 0 \quad \forall \beta = 1, \dots, \alpha - 1$$

$$|f\rangle = \sum_{i=1}^n a_i |l_i\rangle \quad \begin{array}{l} \text{Vary real numbers } a_i \\ \text{using Lagrange} \\ \text{multipliers, maximizing} \\ (*) \end{array}$$

# Variational Principle continues

$$\{|\varphi_i\rangle\}_{i=1}^n$$

$$|f\rangle = \sum_{i=1}^n a_i |\varphi_i\rangle$$

Vary real numbers  $a_i$   
using Lagrange  
multipliers

$$1 \geq \left\langle \sum_{i=1}^n a_i \varphi_i | \mathcal{P} | \sum_{j=1}^n a_j \varphi_j \right\rangle_{\pi^{-1}}$$

$$= \sum_{i,j=1}^n a_i a_j \langle \varphi_i | \mathcal{P} | \varphi_j \rangle_{\pi^{-1}}$$

$$= \sum_{i,j=1}^n a_i a_j C_{ij}$$

Maximizing this

Subject to the constraints

$$\mathcal{L} = \sum_{ij=1}^n a_i a_j \langle \varphi_i | \mathcal{P} | \varphi_j \rangle_{\pi^{-1}}$$

$$- \lambda \left[ \sum_{ij=1}^n a_i a_j \langle \varphi_i | \varphi_j \rangle_{\pi^{-1}} - 1 \right]$$

$$= \sum_{ij=1}^n a_i a_j C_{ij} - \lambda \left[ \sum_{ij=1}^n a_i a_j S_{ij} - 1 \right]$$

$$\langle f | f \rangle_{\pi^{-1}} = \sum_{ij=1}^n a_i a_j \langle \varphi_i | \varphi_j \rangle_{\pi^{-1}} = \sum_{ij=1}^n a_i a_j S_{ij} = 1$$

$$C_{ii} = \langle \varphi_i | \mathcal{P} | \varphi_i \rangle_{\pi^{-1}}$$

$$S_{ij} = \langle \varphi_i | \varphi_j \rangle_{\pi^{-1}} = \langle \varphi_j | \varphi_i \rangle_{\pi^{-1}}$$

# Lagrange multiplier solution

$$\{|\varphi_i\rangle\}_{i=1}^n$$

$$|f\rangle = \sum_{i=1}^n a_i |\varphi_i\rangle$$

$$\begin{aligned}\frac{1}{2} \frac{\partial}{\partial a_k} \mathcal{L} &= \frac{1}{2} \sum_{j=1}^n a_j C_{ij} + \frac{1}{2} \sum_{i=1}^n a_i C_{ij} \\ &\quad - \frac{1}{2} \lambda \left[ \sum_{j=1}^n a_j S_{ij} + \sum_{i=1}^n a_i S_{ij} \right] \\ &= \sum_{i=1}^n a_i C_{ij} - \lambda \sum_{i=1}^n a_i S_{ij} \\ &= 0 \\ \forall k &= 1, 2, \dots, n\end{aligned}$$

Matrix equation, which is a generalized eigenvalue problem

$$\mathbf{C}\mathbf{a} = \lambda \mathbf{S}\mathbf{a}$$

By solving this, we can obtain  $\mathbf{a}_i$ , hence all eigenvectors  $|l_\alpha(x)\rangle$  and corresponding eigenvalues. **How to estimate  $\mathbf{C}$  and  $\mathbf{S}$ ?**

# Estimating matrix elements C and S

For this, we introduce a basis set  $\{\chi_i\}$  consisting of the n cofunctions of the original basis set  $\{\varphi_i\}$  by weighting the original functions with  $\pi^{-1}$

$$\chi_i(x) = \pi^{-1}(x)\varphi_i(x) \Leftrightarrow \varphi_i(x) = \pi(x)\chi_i(x)$$

$$\begin{aligned} C_{ij} &= \langle \varphi_i | \mathcal{P}(\tau) | \varphi_j \rangle_{\pi^{-1}} \\ &= \langle \chi_i \pi | \mathcal{P}(\tau) | \pi \chi_j \rangle_{\pi^{-1}} \\ &= \int_X \int_X \chi_i(z) p(y, z, \tau) \pi(y) \chi_j(y) dy dz \end{aligned}$$

It is the time-lagged cross-correlation between the functions  $\chi_i$  and  $\chi_j$

$$\begin{aligned} \text{cor}(\chi_i, \chi_j, \tau) &:= \int_X \int_X \chi_i(z) \mathbb{P}(x_{t+\tau} = z | x_t = y) \\ &\quad \times \chi_j(y) \mathbb{P}(x_t = y) dy dz \end{aligned}$$

which can be estimated from a time-continuous time series  $\mathbf{x}_t$  of length T (or time discretized)

$$\widehat{\text{cor}}_T(\chi_i, \chi_j, \tau) = \frac{1}{T - \tau} \int_0^{T-\tau} \chi_j(x_t) \chi_i(x_{t+\tau}) dt \quad \widehat{\text{cor}}_T(\chi_i, \chi_j, \tau) = \frac{1}{N_T - n_\tau} \sum_{t=1}^{N_T - n_\tau} \chi_j(x_t) \chi_i(x_{t+n_\tau})$$

# Transfer operator

$$\mathcal{J}(\tau)|f(z)\rangle = \frac{1}{\pi(z)} \int_X p(y, z, \tau) \pi(y) f(y) dy$$

$$\begin{aligned} C_{ij}(\tau) &= \langle \chi_i \pi | \mathcal{P}(\tau) | \pi \chi_j \rangle_{\pi^{-1}} \\ &= \langle \chi_i | \mathcal{J}(\tau) | \chi_j \rangle_{\pi} \end{aligned}$$

T has the same eigenvalues as the propagator P and its eigenfunctions are cofunctions of the propagator eigenfunction

$$r_\alpha(x) = \pi^{-1}(x) l_\alpha(x)$$

# Summary

1. Choose a basis set  $\{|\chi_i\rangle\}$ .
2. Estimate the matrix elements of the correlation matrix  $\mathbf{C}$  and of the overlap matrix  $\mathbf{S}$  using eq 25 with lag times  $\tau$  and 0, respectively.
3. Solve the generalized eigenvalue problem in eq 17. This yields the  $\alpha$ -th eigenvalue  $\lambda_\alpha$  of the propagator (and the transfer operator) and the expansion coefficients  $a_i^\alpha$  of the associated eigenvector.
4. The eigenvectors of the transfer operator are obtained directly from the expansion coefficients  $a_i^\alpha$  via

$$|r_\alpha\rangle = \sum_{i=1}^n a_i^\alpha |\chi_i\rangle \quad (48)$$

5. If an estimate of the stationary density  $\pi$  is available, the eigenvectors of the propagator  $\mathcal{P}(\tau)$  are obtained from

$$|l_\alpha\rangle = \sum_{i=1}^n a_i^\alpha |\varphi_i\rangle = \sum_{i=1}^n a_i^\alpha |\pi\chi_i\rangle \quad (49)$$

$$\widehat{\text{cor}}_T(\chi_i, \chi_j, \tau) = \frac{1}{N_T - n_\tau} \sum_{t=1}^{N_T - n_\tau} \chi_j(x_t) \chi_i(x_{t+n_\tau})$$

$$\mathbf{Ca} = \lambda \mathbf{Sa}$$

Now ... MSM is just an example of a propagator with indicator basis function

$$|\chi_i^{\text{MSM}}(x)\rangle = \begin{cases} 1 & \text{if } x \in S_i \\ 0 & \text{otherwise} \end{cases}$$

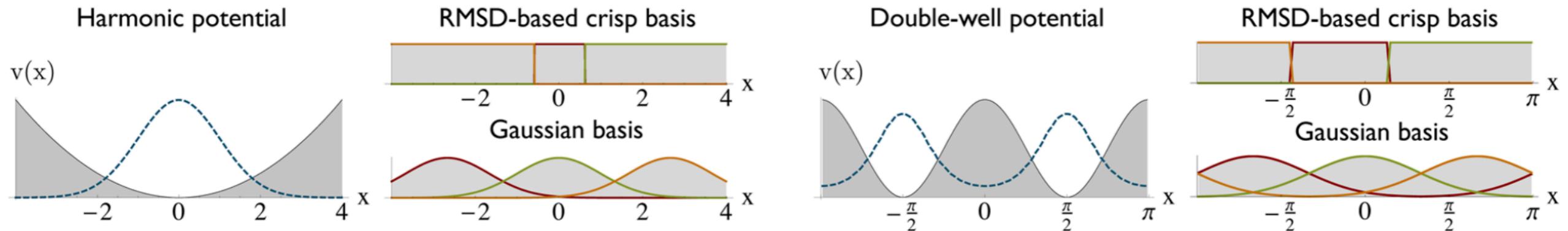
Estimating matrix C is just counting the observed transitions  $z_{ij}$  between sets  $S_i$  and  $S_j$

$$\begin{aligned} C_{ij} &= \frac{1}{N_T - n_\tau} \sum_{t=1}^{N_T - n_\tau} \chi_j^{\text{MSM}}(x_t) \chi_i^{\text{MSM}}(x_{t+n_\tau}) \\ &= \frac{z_{ij}}{N_T - n_\tau} \end{aligned}$$

# Crisp-basis (MSM) vs Gaussian Basis

**A**

## ID-Potentials and basis sets



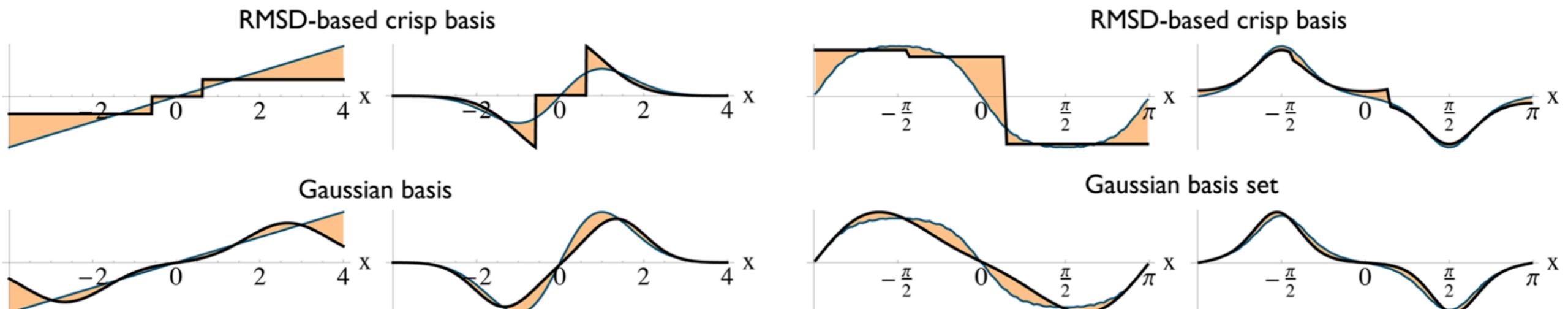
**B**

## Eigenvectors



**C**

## Basis set approximation of the eigenvectors



**Figure 2.** Illustration of the method with two one-dimensional potentials, the harmonic potential in the left half and a periodic double-well potential in the right half of the figure. (A) Potential  $v$  together with its invariant distribution  $\pi$  (shaded) next to two possible choices of basis functions: a three-element crisp basis and a set of three Gaussian functions. (B) Exact right and left second eigenfunctions,  $|r_2\rangle$  and  $|l_2\rangle$ . (C) Approximation results for these second eigenfunctions obtained from the basis sets shown.

# Methods for approximating the propagator

- time-structure-based independent component analysis (tICA) and extensions (kernel tICA, sparse tICA),
- variational approach for Markov processes (VAMP)nets,
- soft-max MSMs
- diffusion maps

These methods make different assumptions about the underlying eigenfunctions of the system; for example, approximating the eigenfunctions of the propagator under the constraint of linear combinations of features produces tICA, which can be further enhanced as a nonlinear approximation via the kernel trick.

	Less Interpretable	Interpretable
Non- Linear	Kernel tICA MSM TAE VAMPnet	VDE + Saliency
Linear	PCA	Sparse tICA tICA

# Methods for capturing slow collective variables

- Linear Dimensionality reduction: PCA +tICA
- Autoencoder
- EncoderMap = Autoencoder + MDS
- Time-lagged Autoencoder
- Variational Autoencoder
- Variational time-lagged Autoencoder (aka VDE)
- VAMPNet

# Variational encoding of complex dynamics

Carlos X. Hernández,<sup>1</sup> Hannah K. Wayment-Steele,<sup>2</sup> Mohammad M. Sultan,<sup>2</sup> Brooke E. Husic,<sup>2</sup> and Vijay S. Pande<sup>1,2,\*</sup>

<sup>1</sup>*Biophysics Program, Stanford University, Stanford, California, USA*

<sup>2</sup>*Chemistry Department, Stanford University, Stanford, California, USA*

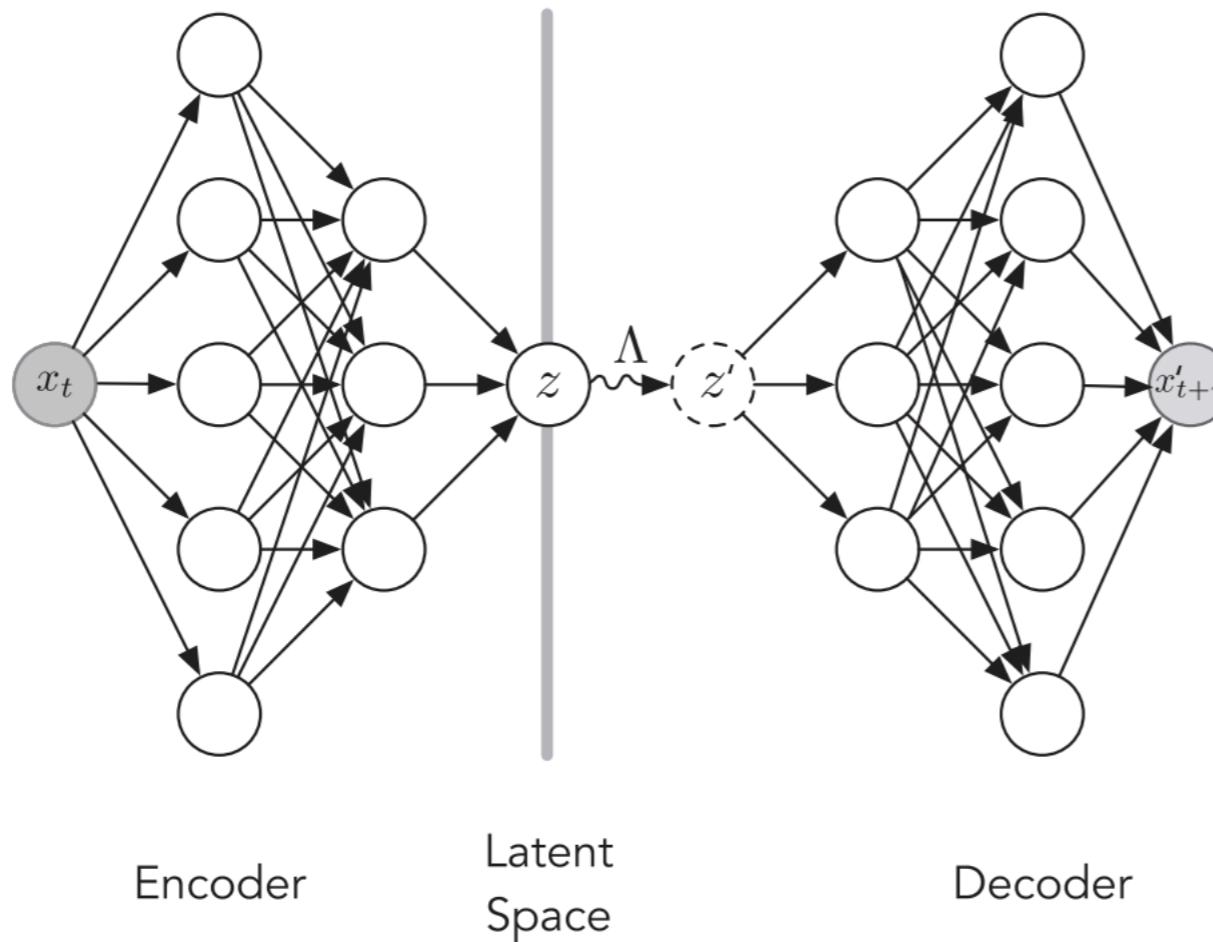


(Received 1 December 2017; published 18 June 2018)

Often the analysis of time-dependent chemical and biophysical systems produces high-dimensional time-series data for which it can be difficult to interpret which individual features are most salient. While recent work from our group and others has demonstrated the utility of time-lagged covariate models to study such systems, linearity assumptions can limit the compression of inherently nonlinear dynamics into just a few characteristic components. Recent work in the field of deep learning has led to the development of the variational autoencoder (VAE), which is able to compress complex datasets into simpler manifolds. We present the use of a time-lagged VAE, or variational dynamics encoder (VDE), to reduce complex, nonlinear processes to a single embedding with high fidelity to the underlying dynamics. We demonstrate how the VDE is able to capture nontrivial dynamics in a variety of examples, including Brownian dynamics and atomistic protein folding. Additionally, we demonstrate a method for analyzing the VDE model, inspired by saliency mapping, to determine what features are selected by the VDE model to describe dynamics. The VDE presents an important step in applying techniques from deep learning to more accurately model and interpret complex biophysics.

DOI: [10.1103/PhysRevE.97.062412](https://doi.org/10.1103/PhysRevE.97.062412)

# VDE - variational dynamics encoder



- dimensionality reduction
- generative model
- approximation of a propagator
- captures slow collective variables

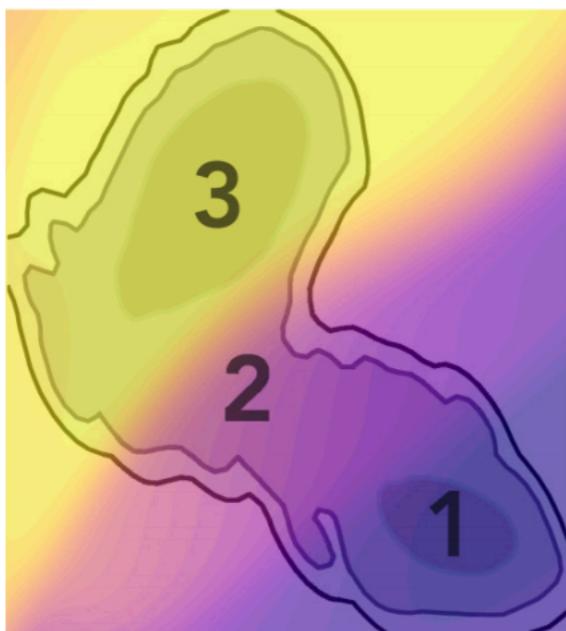
$$\mathcal{L}_{\text{VDE}} = \boxed{\mathcal{L}_R} + \boxed{\mathcal{L}_{\text{KL}}} + \boxed{\mathcal{L}_{\text{AC}}}.$$

$$\mathcal{L}_R = \mathbb{E}[\|x'_{t+\tau} - x_{t+\tau}\|].$$

$$\mathcal{L}_{\text{KL}} = \mathbb{E} \left[ \frac{1 + \log \sigma(z_t)^2 - \mu(z_t)^2 - \sigma(z_t)^2}{2} \right],$$

$$\mathcal{L}_{\text{AC}} = -\rho_{z_t, z_{t+\tau}} = -\frac{\mathbb{E}[(z_t - \bar{z}_t)(z_{t+\tau} - \bar{z}_{t+\tau})]}{s_{z_t} s_{z_{t+\tau}}}$$

# Example 1: Brownian Dynamics



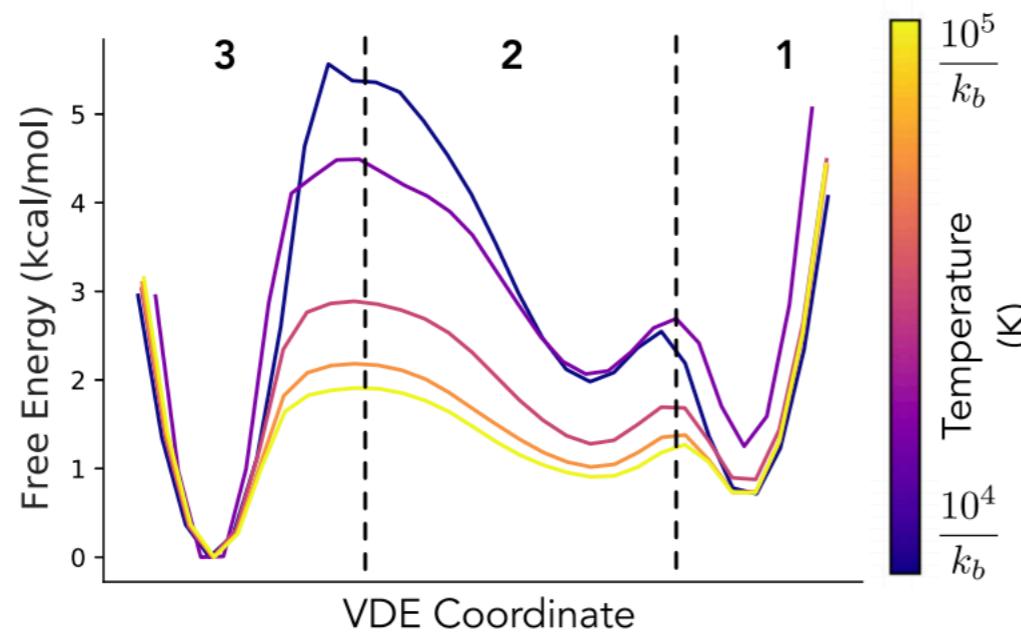
(a) VDE



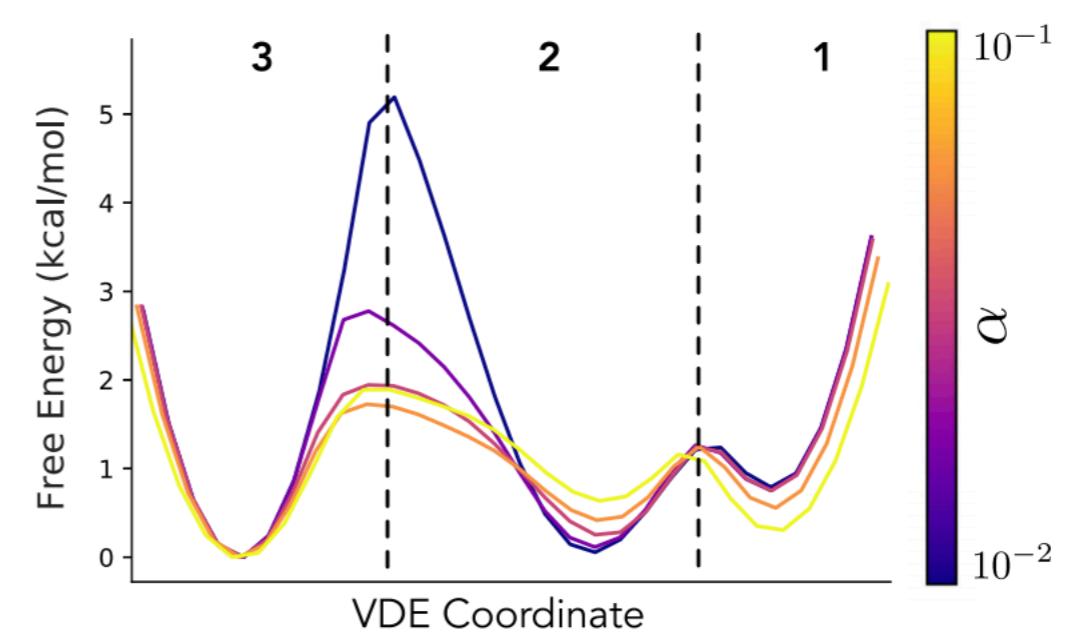
(b) tICA



(c) PCA

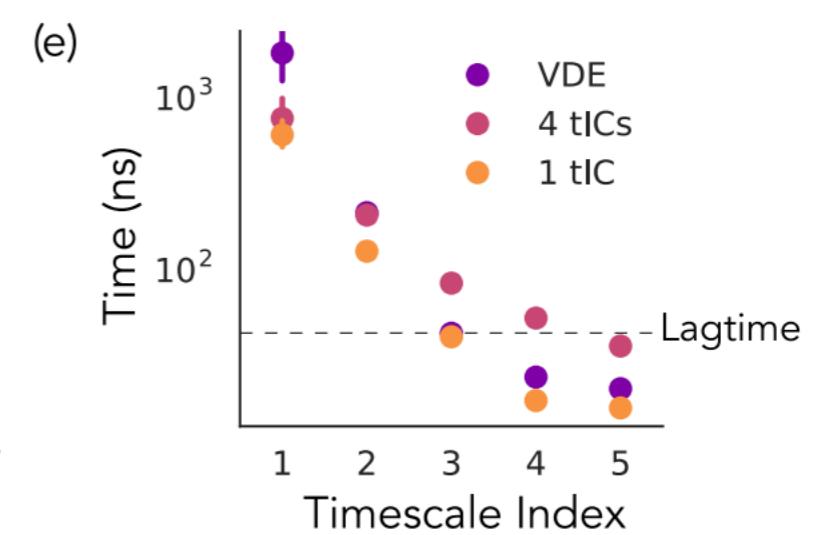
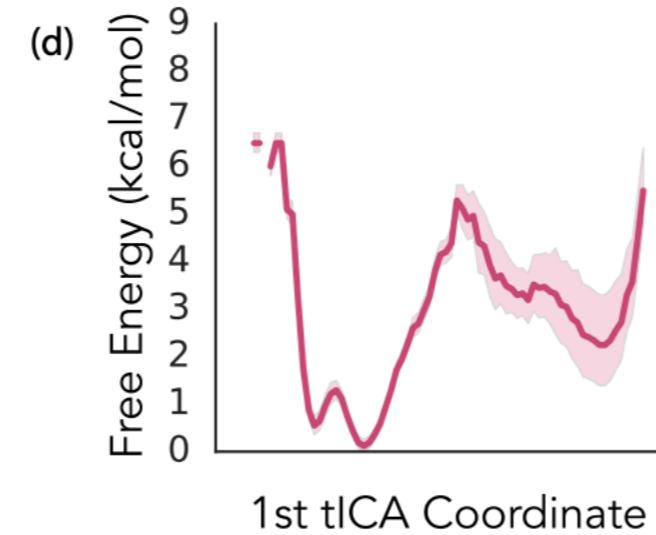
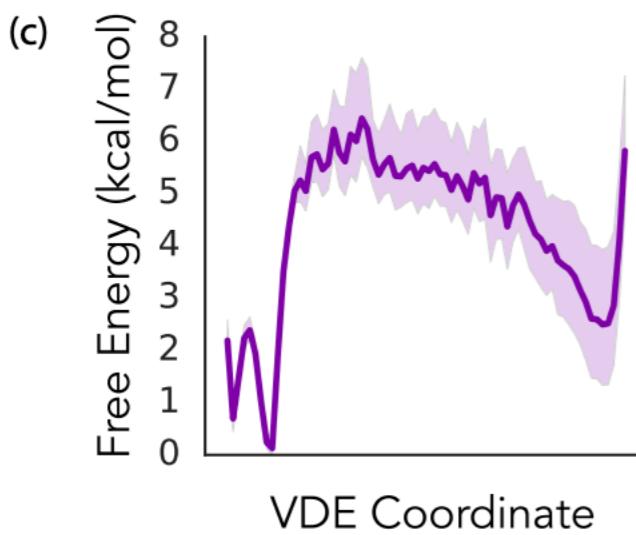
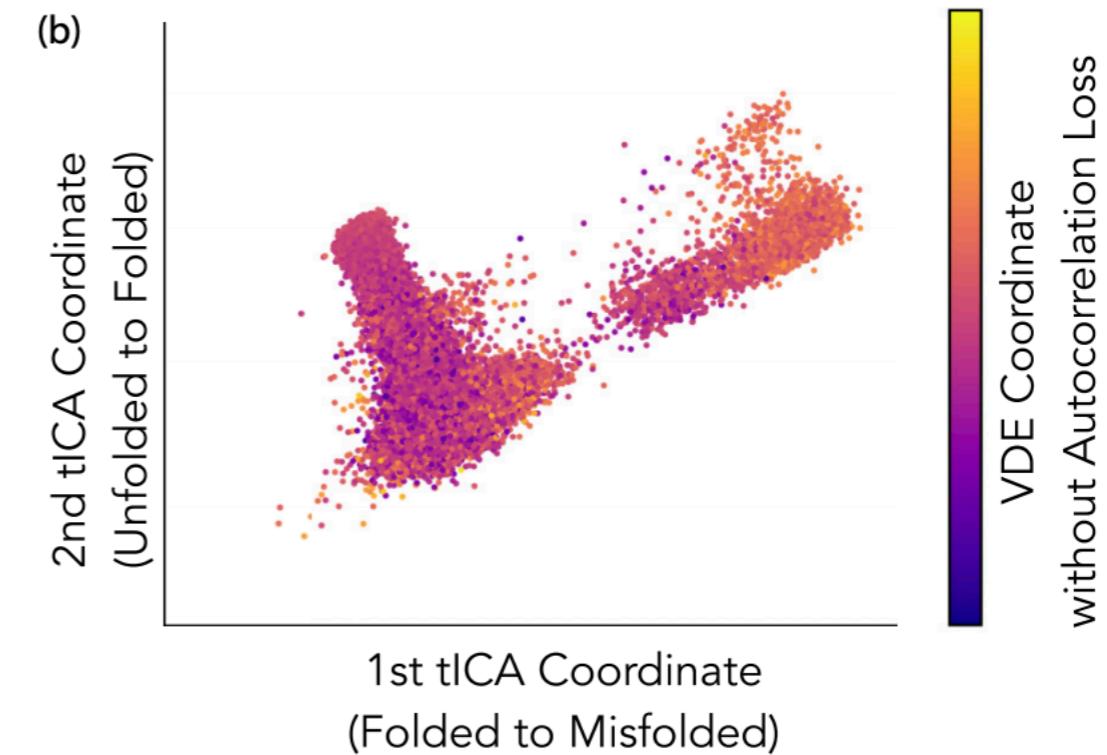
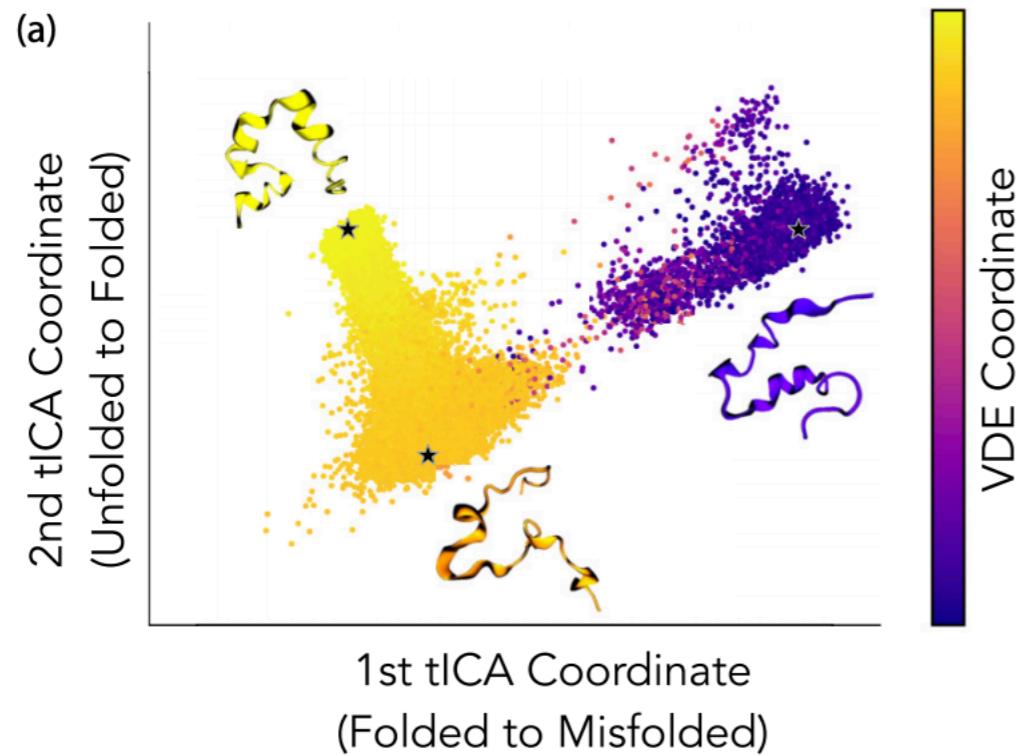


(a) Brownian thermodynamics at many temperatures



(b) VDE thermodynamics at many scaling values

# Example 2: protein folding



# Take home points

- Variational approach for conformational dynamics (VAC) formalizes MSMs
- All existing methods for capturing collective variables are (implicitly) motivated by VAC
- You can create your own propagator by utilizing NN by training on big data: what matters most is a loss function

# References:

- F Nüske et al. “Variational Approach to Molecular Kinetics.” J of Chem Theory and Comput. 2014 10 (4), 1739-1752
- CX Hernández et al, “Variational encoding of complex dynamics” Phys. Rev. E 97, 062412 (2018)