# Big Data Summer, 11 Sept 2019

## Variational autoencoders for dimensionality reduction and clustering of molecular dynamics data

**Jose**

Max Planck Institute fo

**+ Automated detection of many-particle solvation states using hidden Markov models**

# Big Data Summer, 11 Sept 2019

## Variational autoencoders for dimensionality reduction and clustering of molecular dynamics data

**Joseph F. Rudzinski**

Max Planck Institute for Polymer Research, Mainz, Germany

Dr. Tristan Bereau

Prof. Kurt Kremer

Dr. Marc Radu

Svenja Wörner

Yasemin Bozkurt Varolgunes

Dr. Yani Zhao

**Funding**

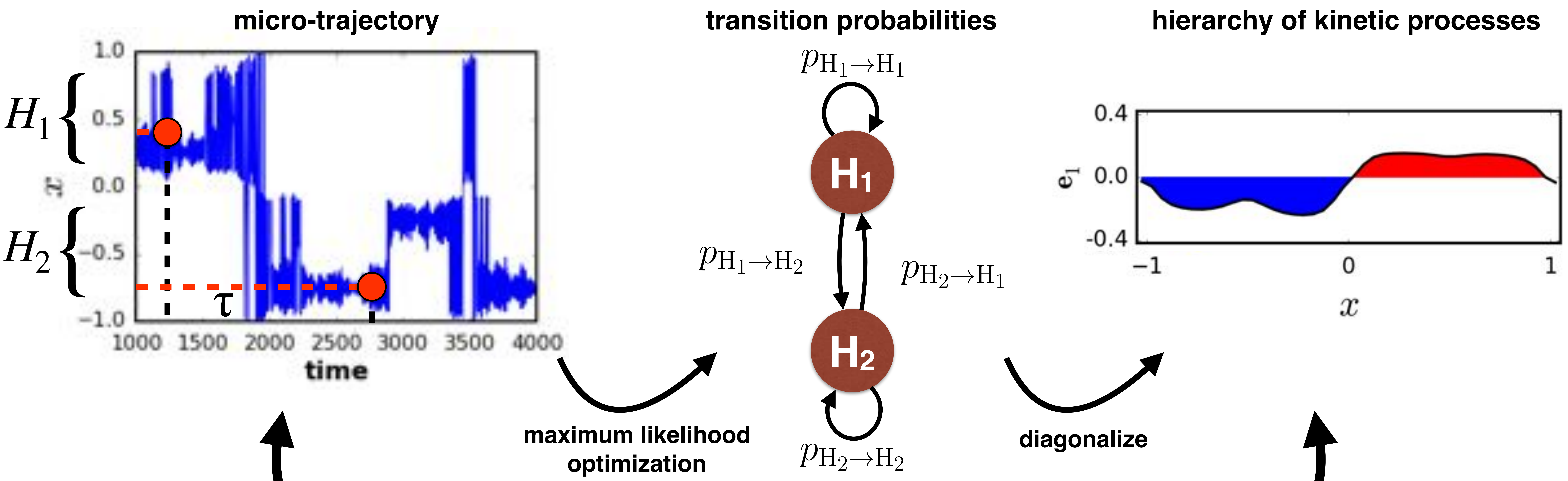
**Humboldt postdoctoral fellowship**

Alexander von Humboldt
Stiftung/Foundation

DFG Deutsche Forschungsgemeinschaft

TRR 146

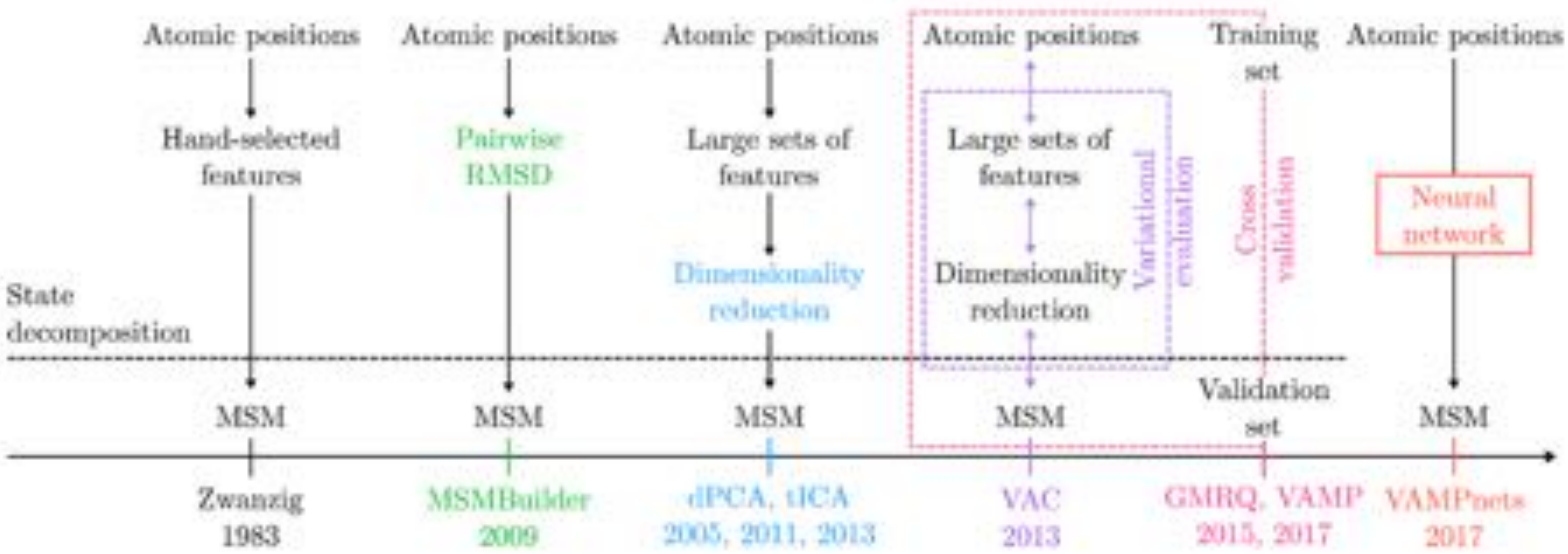
# Markov State Models




MSMs link microtrajectories with long timescale kinetic processes

Bowman *et al.* “An introduction to Markov state models…” *Springer* (2014)

# Analysis: Markov state models (MSMs)

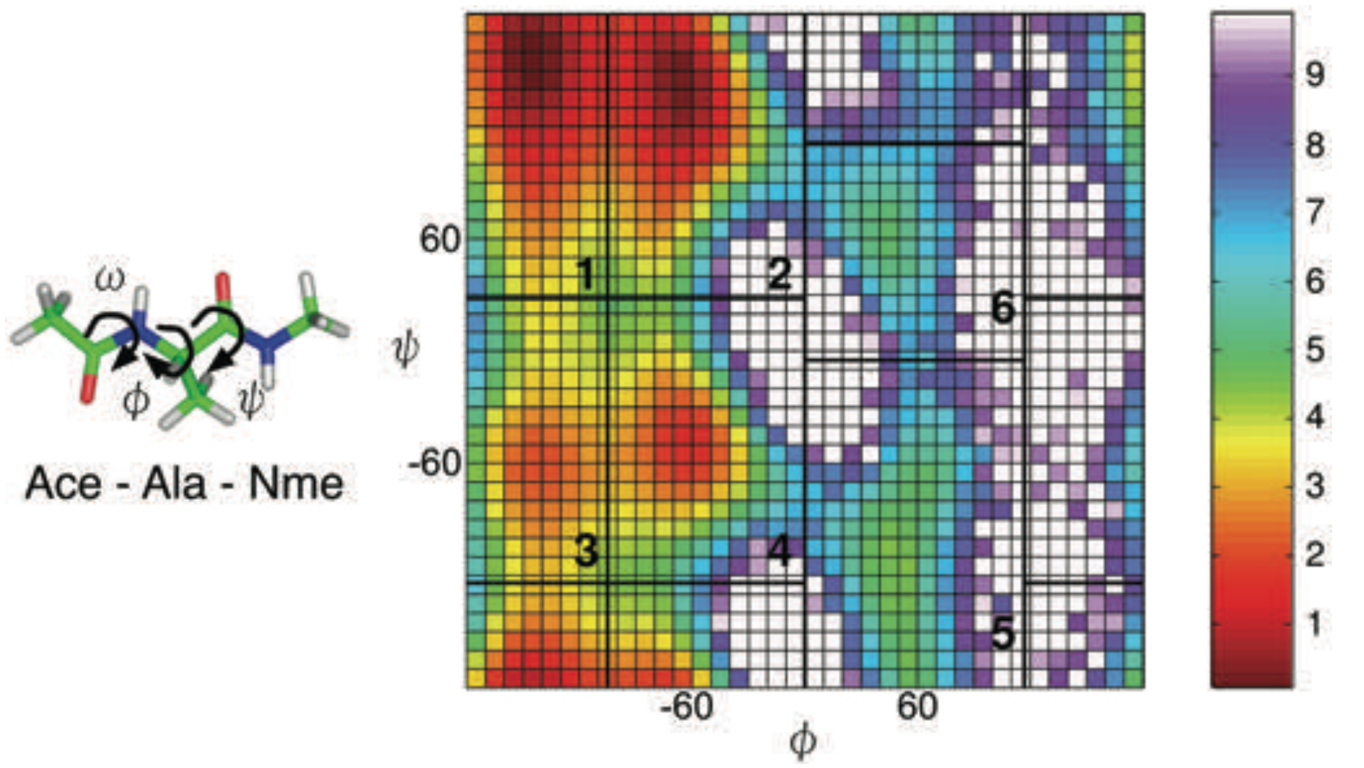



Husic, Pande *JACS* ***(2018)*** "Markov State Models: From an Art to a Science"

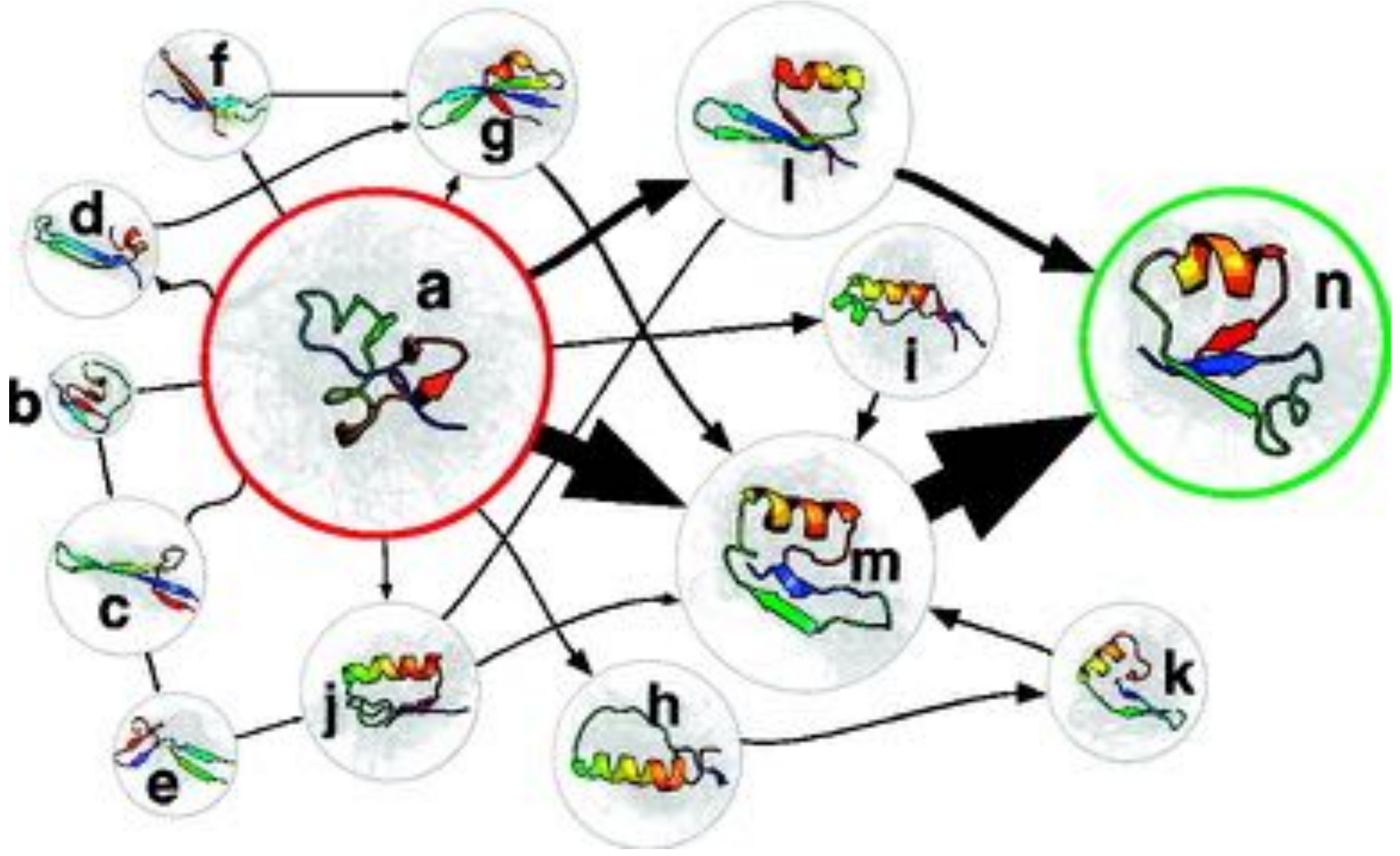
# Analysis: Markov state models (MSMs)

# Success of MSMs for protein systems

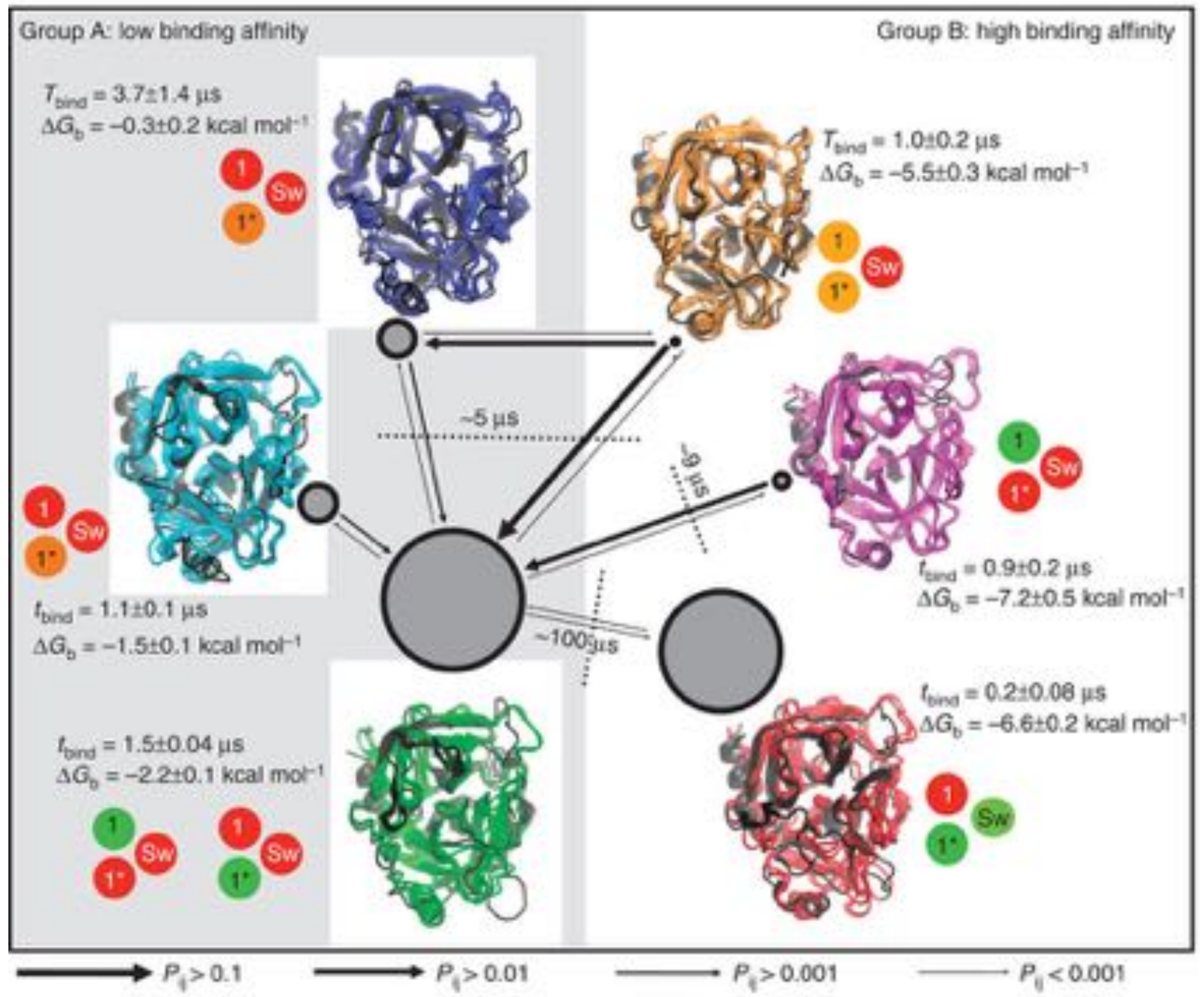
Well-characterized, low-dimensional features provide a great starting point for building kinetic models.




Chodera *et al. JCP* (2007)




Voelz *et al. JACS* (2010)




Plattner *et al. Nat Comm* (2015)

Pande and coworkers
Noé and coworkers
many more…

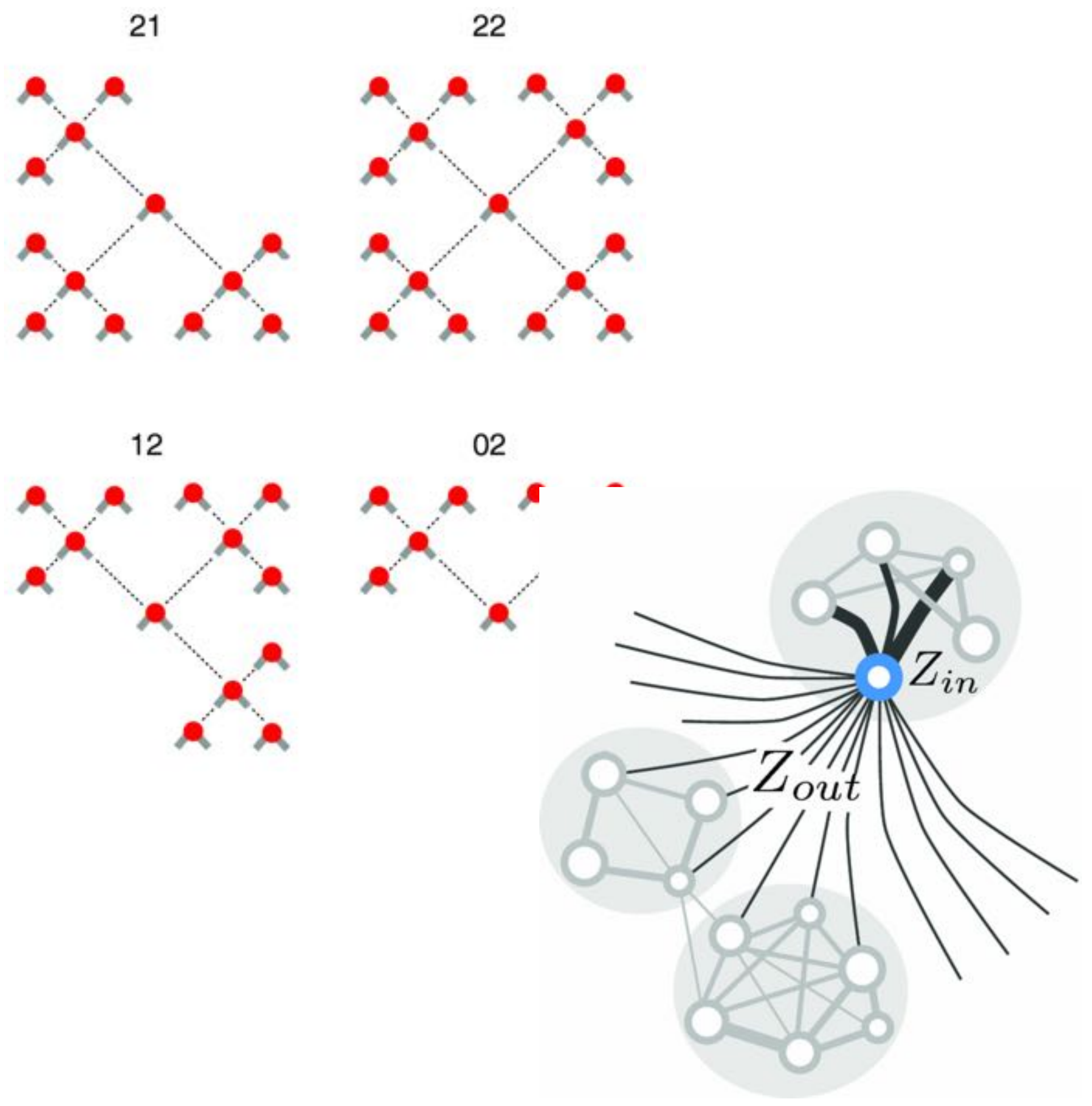
Recent Review - Husic and Pande “Markov state models: from an art to a science” *JACS* (2018)

# MSMs for many-particle systems

Specialized, system-dependent input features

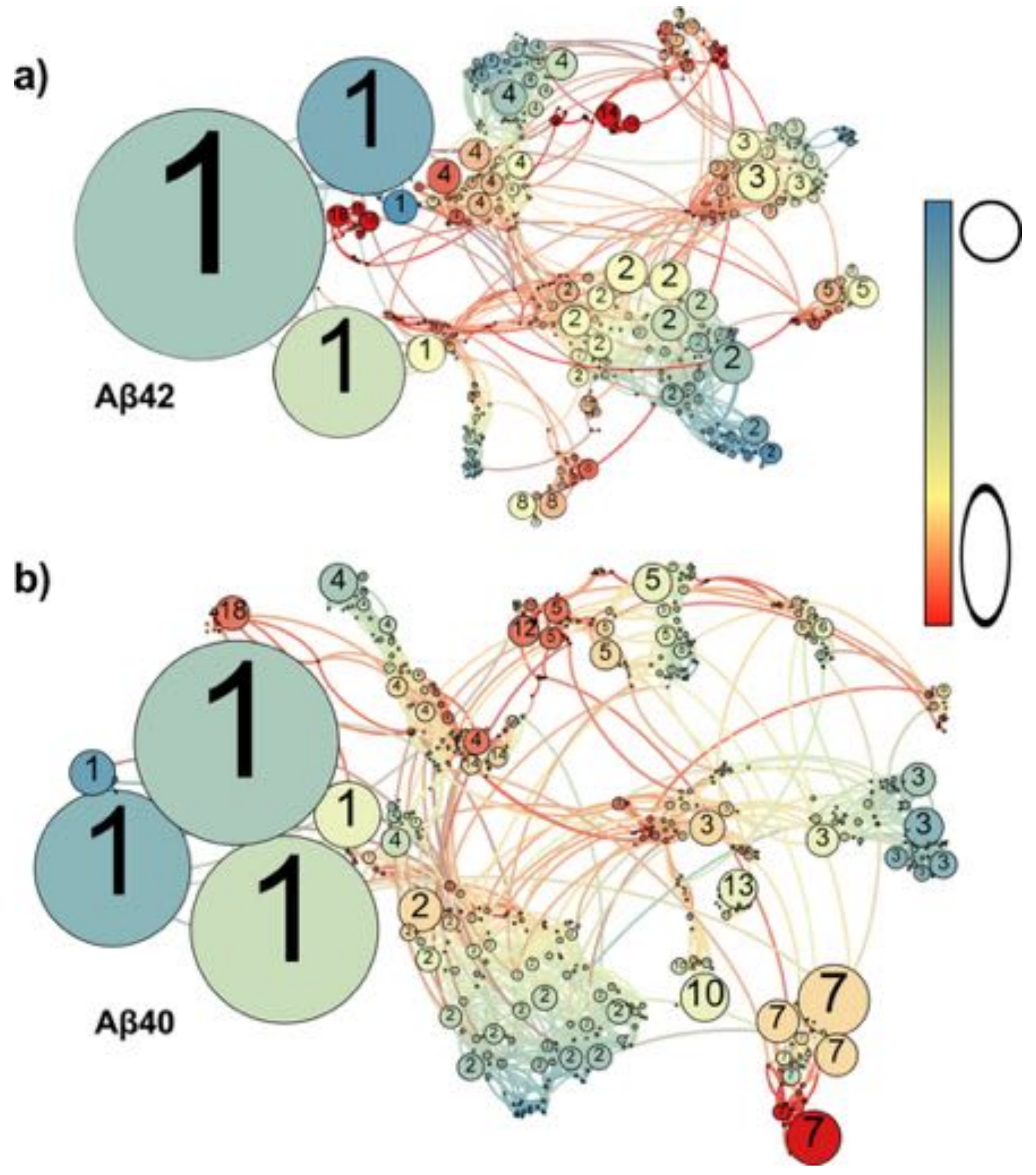
Hidden-Markov-model-based identification of states

**Water hydrogen bonding dynamics**



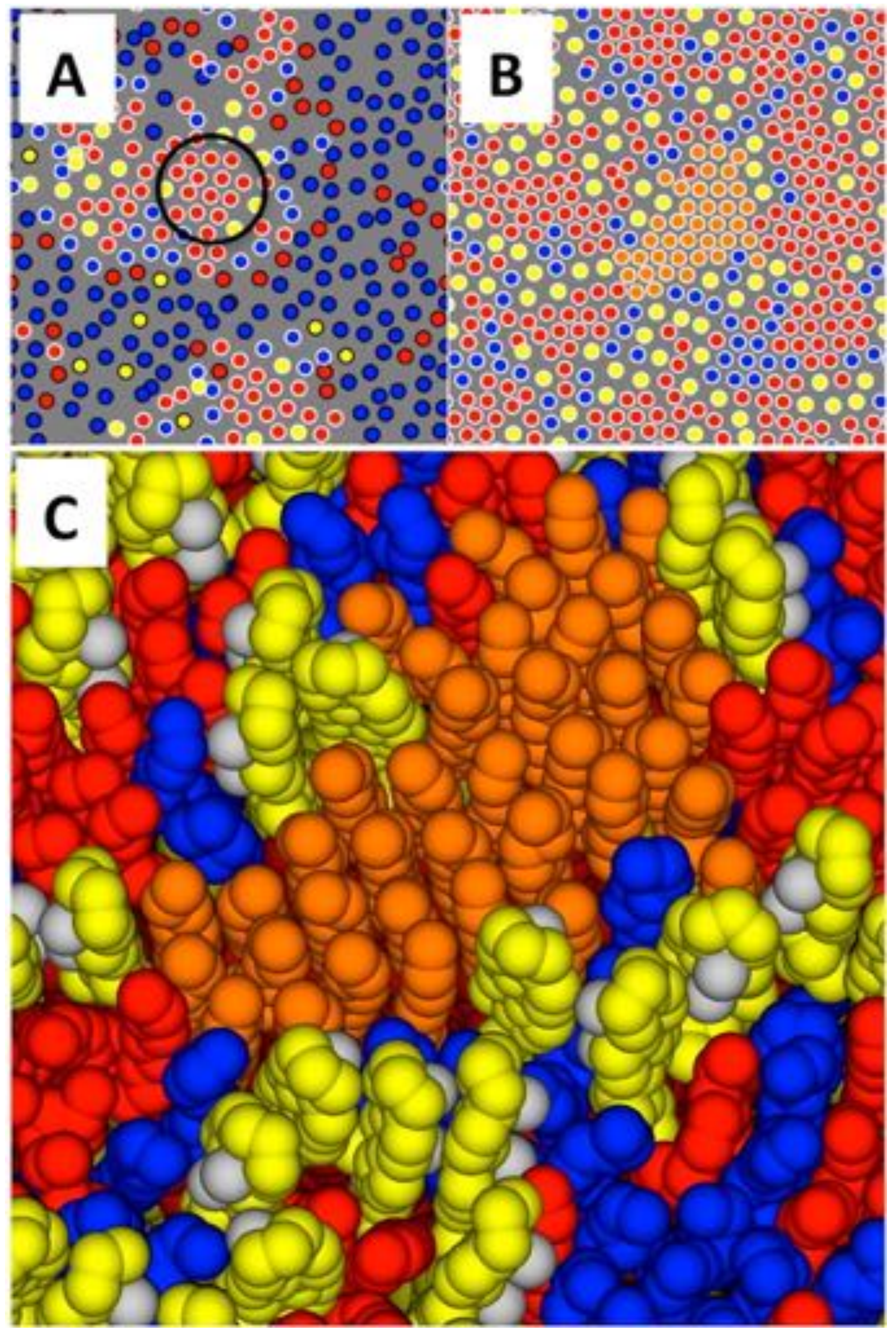

Prada-Garcia *et al. JCP* (2012)

**Amyloid aggregation kinetics**



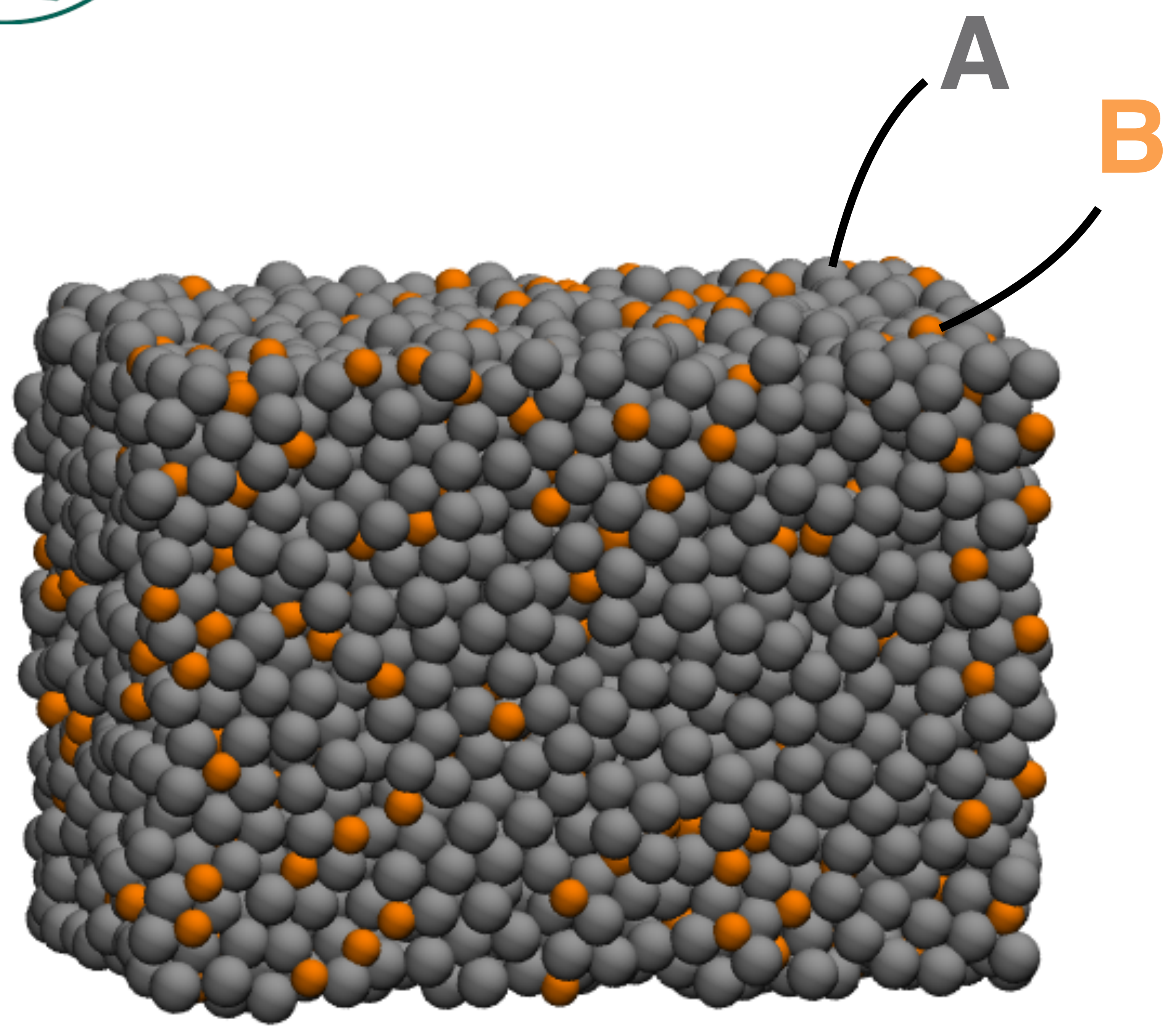

Barz, Liao, Strodel *JACS* (2018)

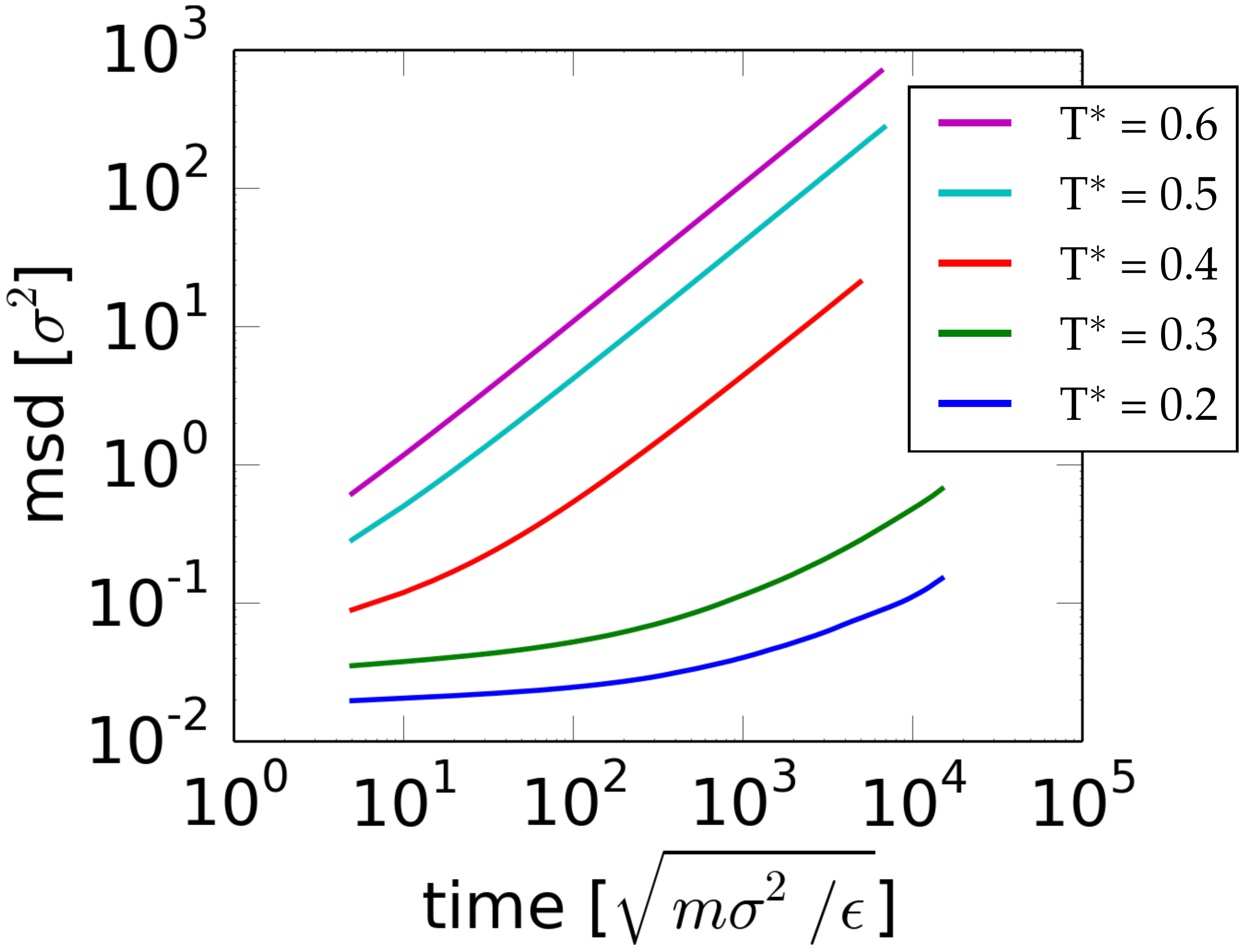
**Lipid ordering in bilayers**

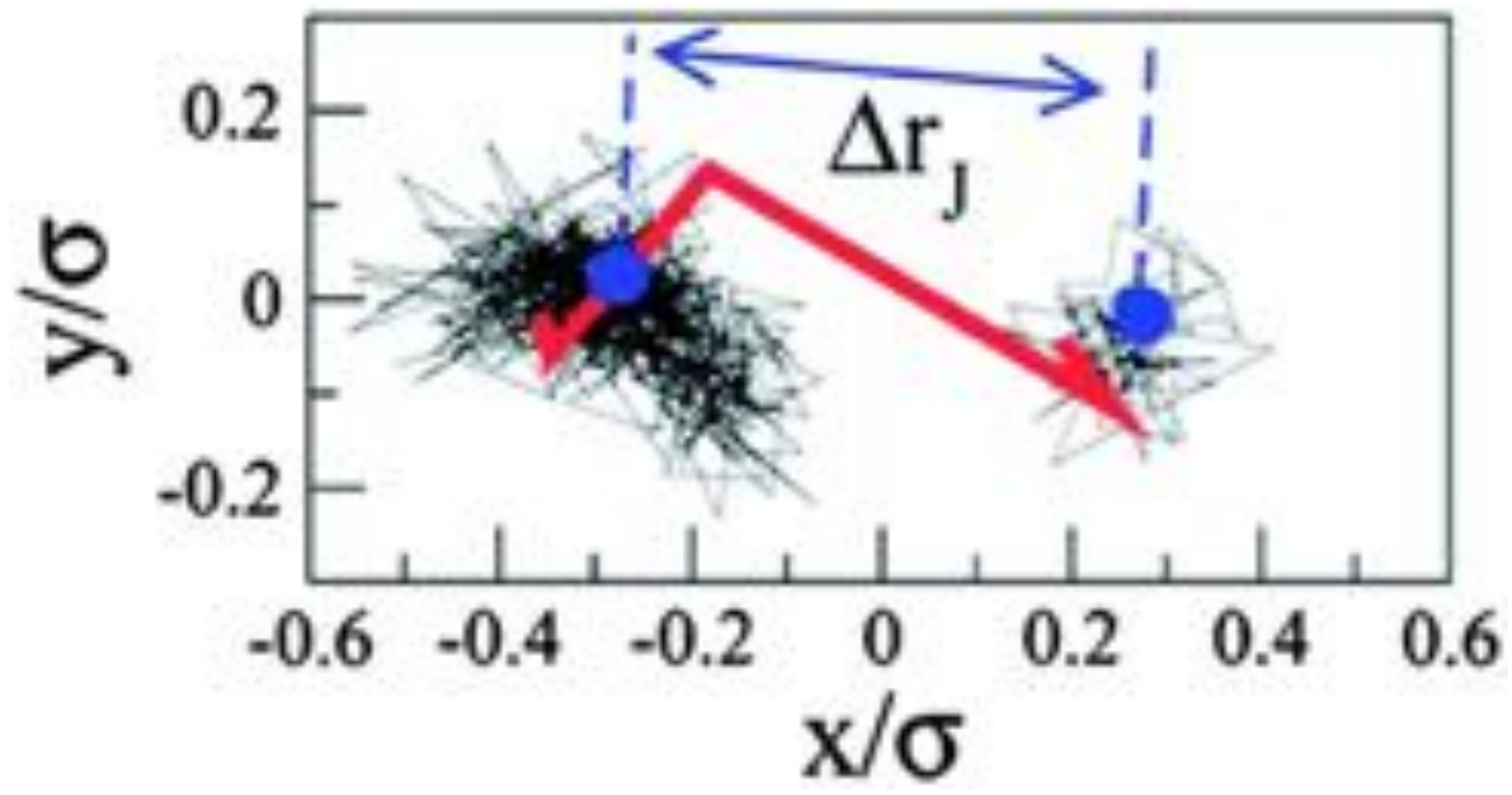



Sodt *et al. JACS* (2014)

# Kob-Andersen model for glassy liquids



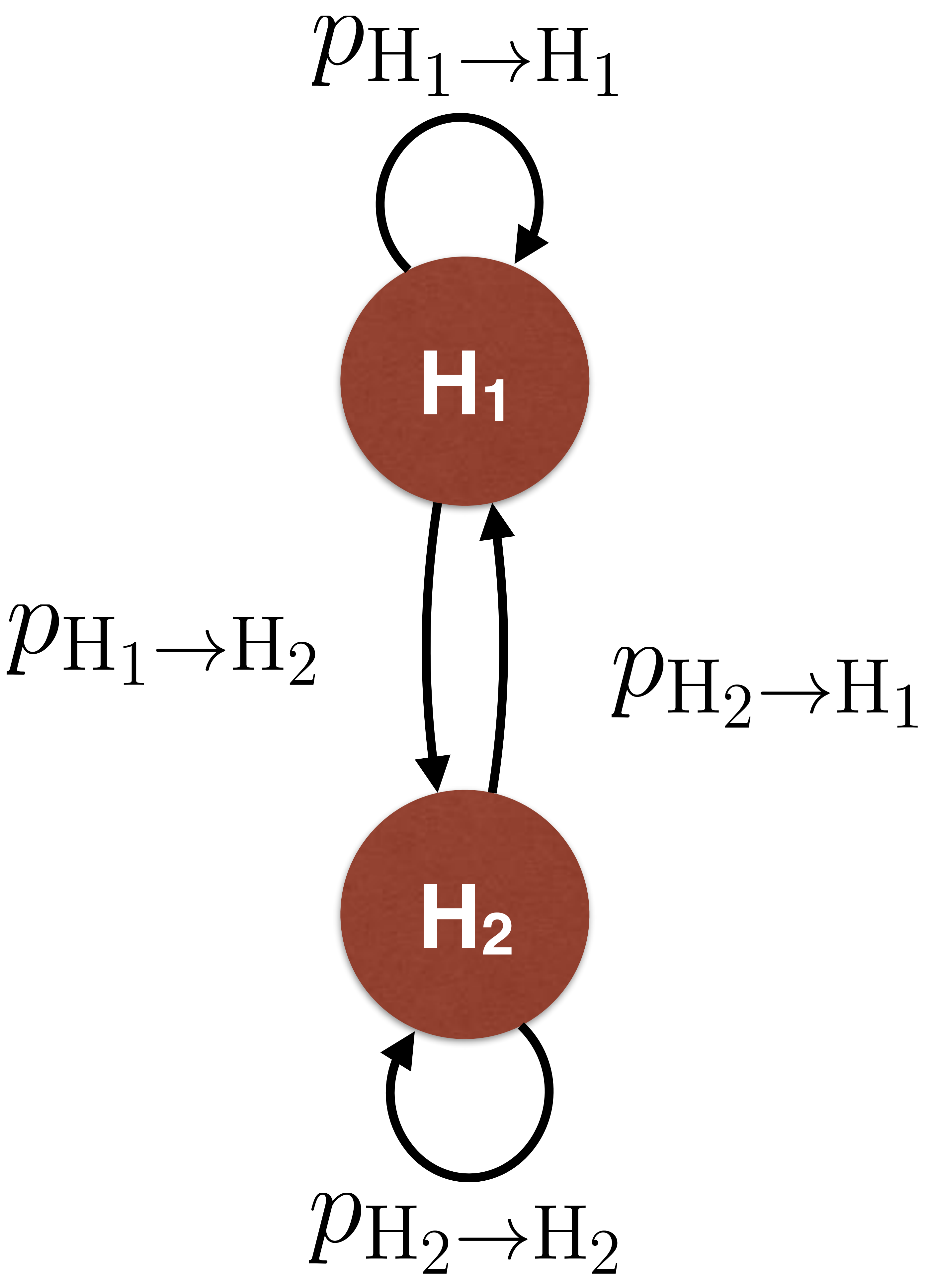



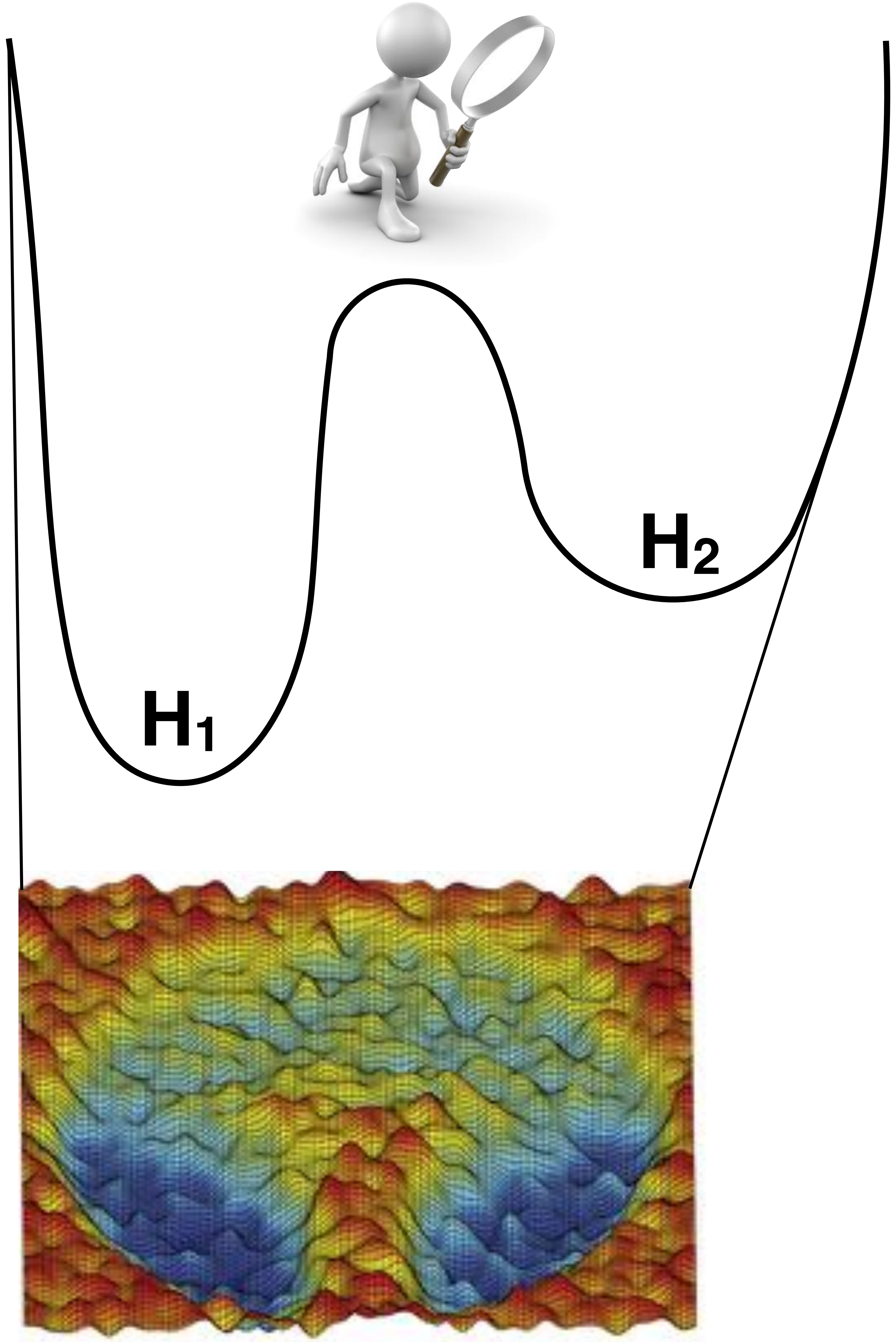




JFR, Radu, Bereau *J Chem Phys* (2019)

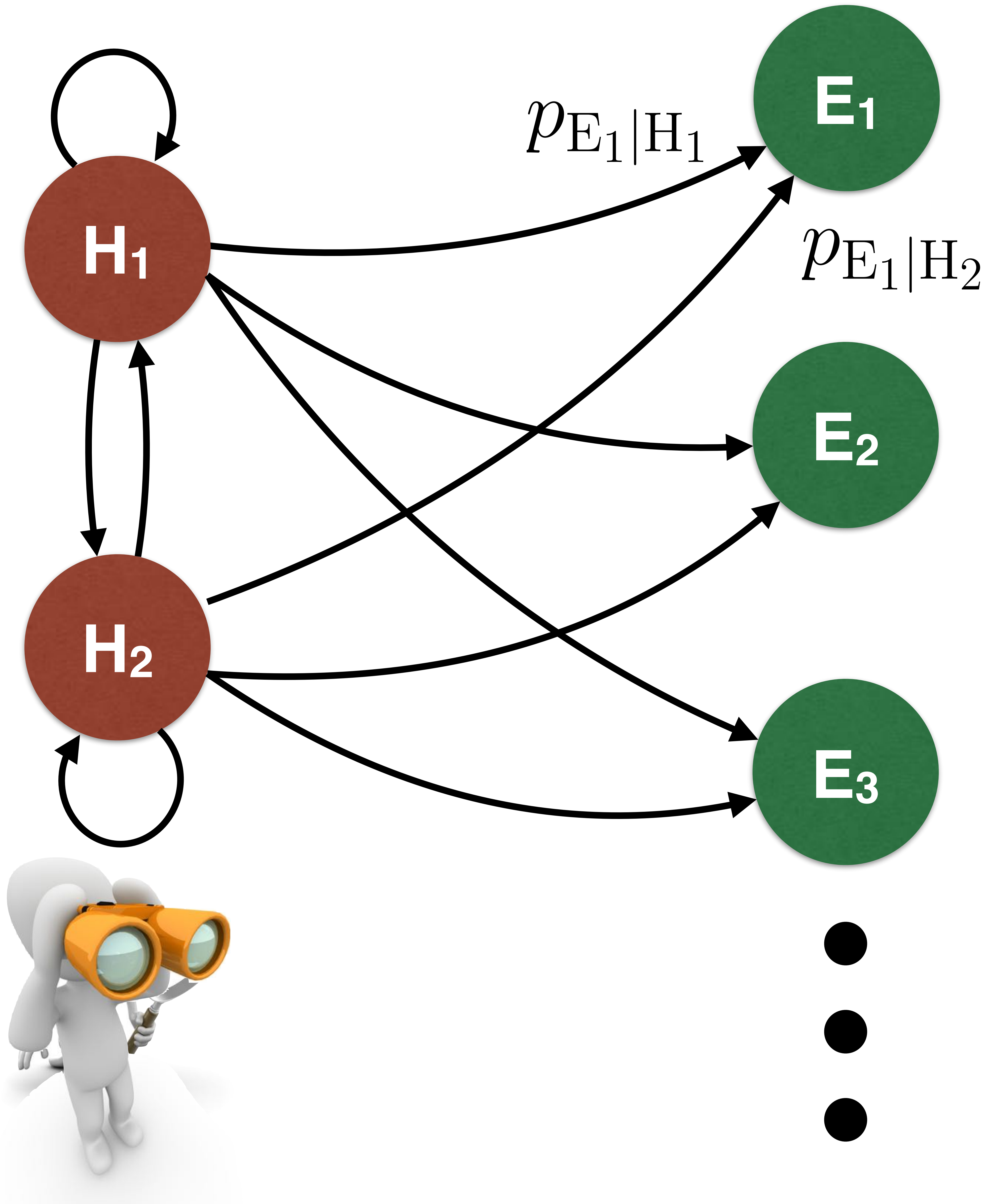
Ciamarra, Pastore, Coniglio "Particle jumps in structural glasses" *SM* 2016

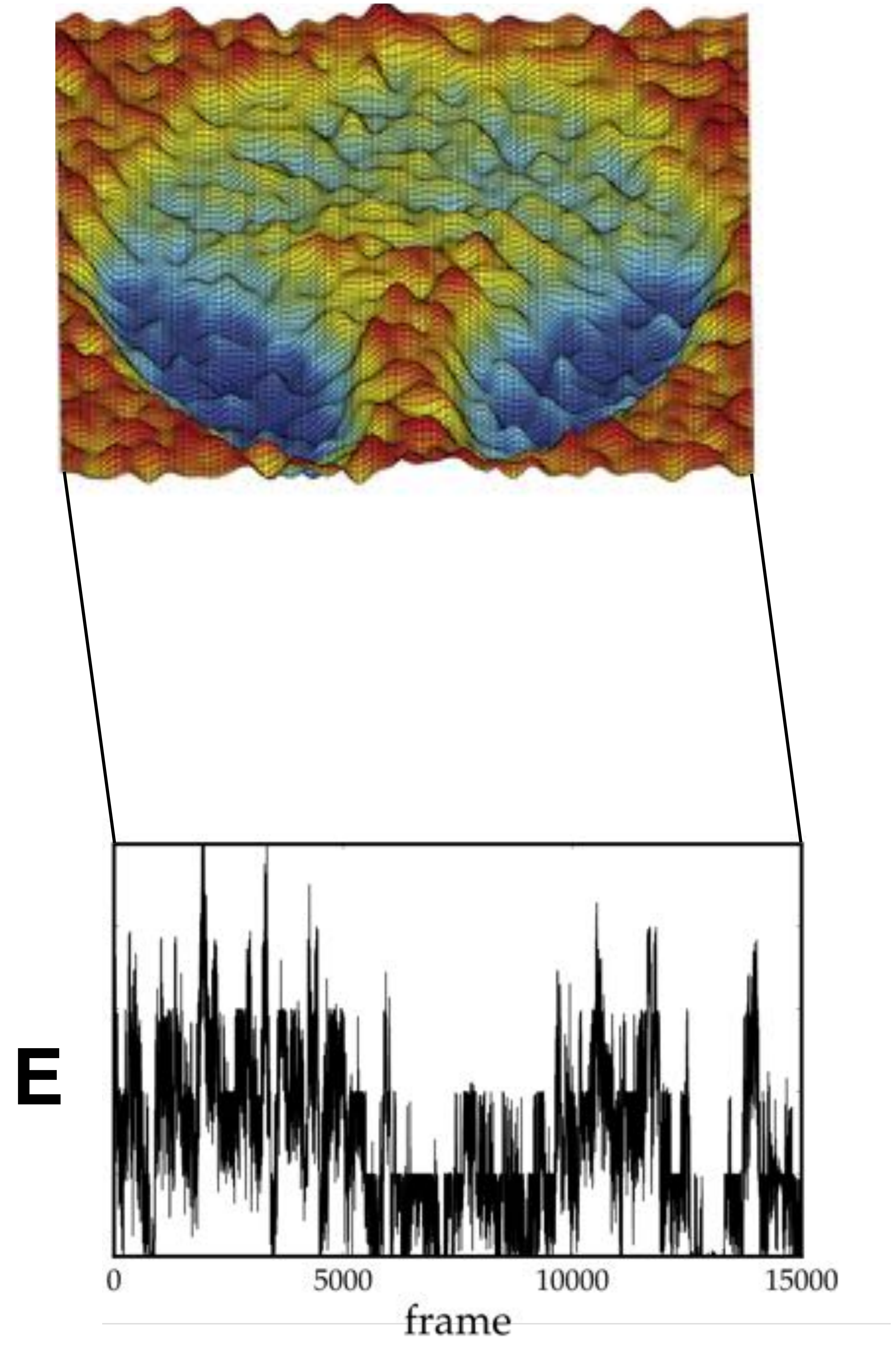
# MSMs on observable states



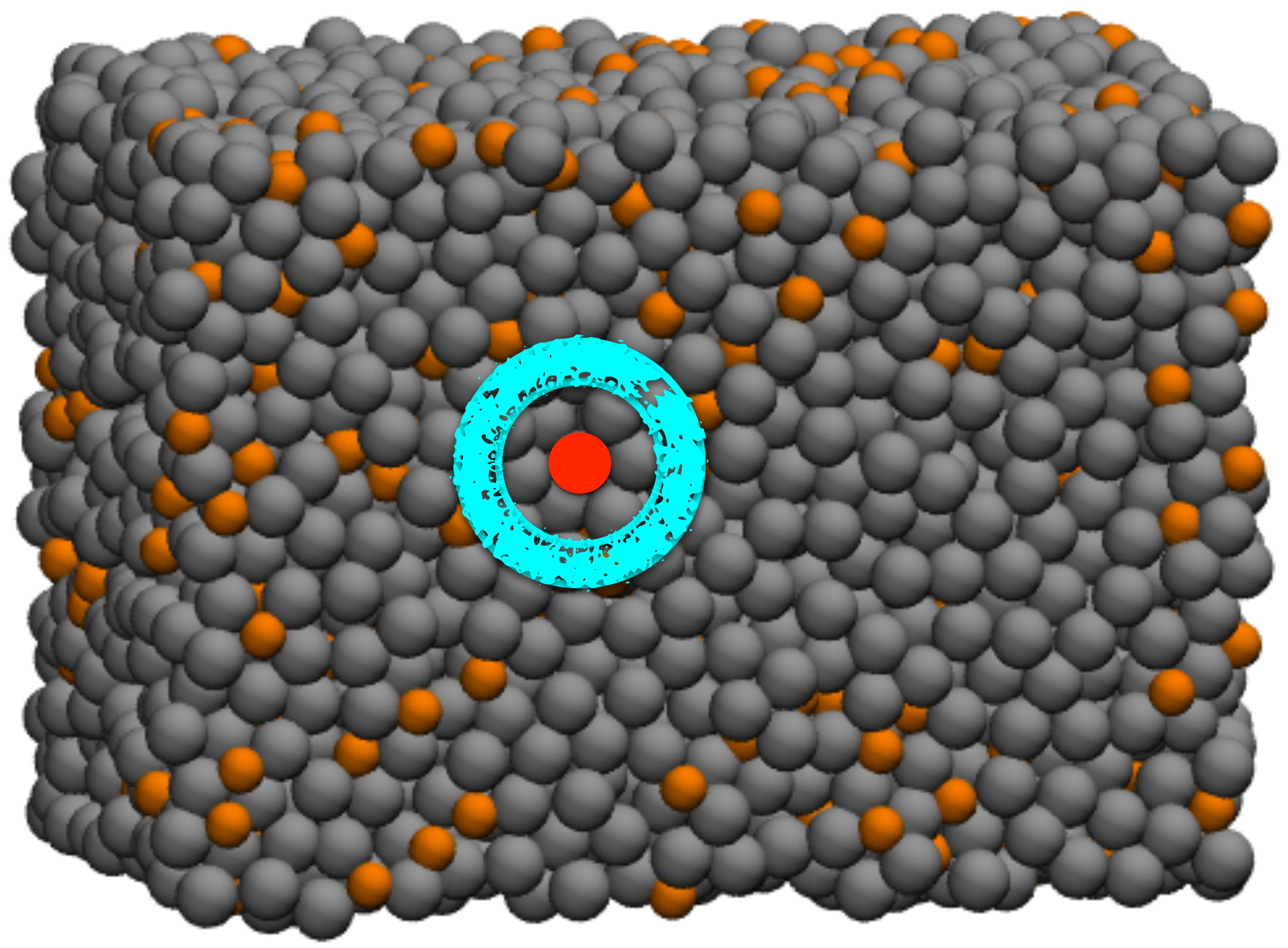
# MSMs on hidden states

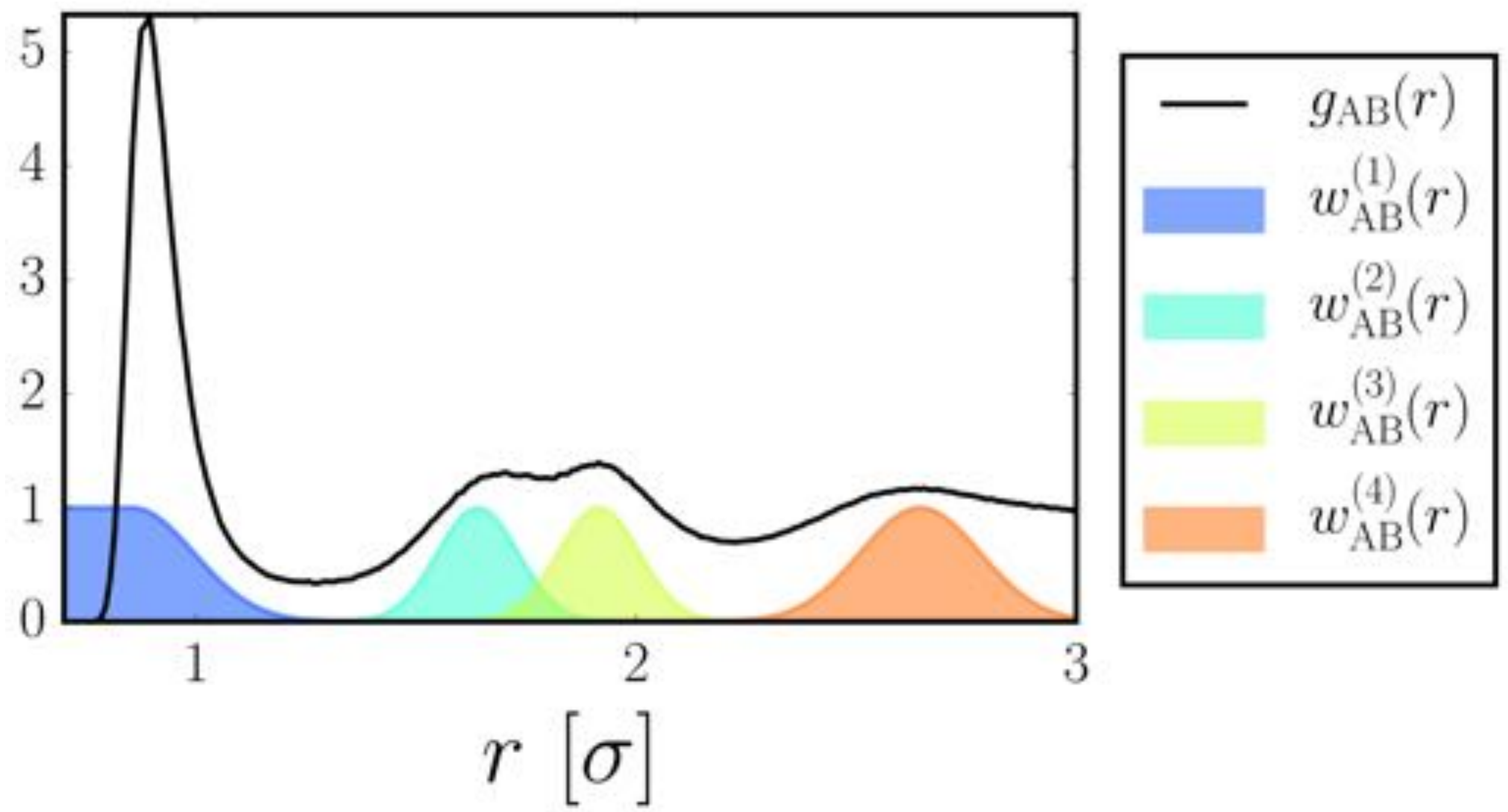


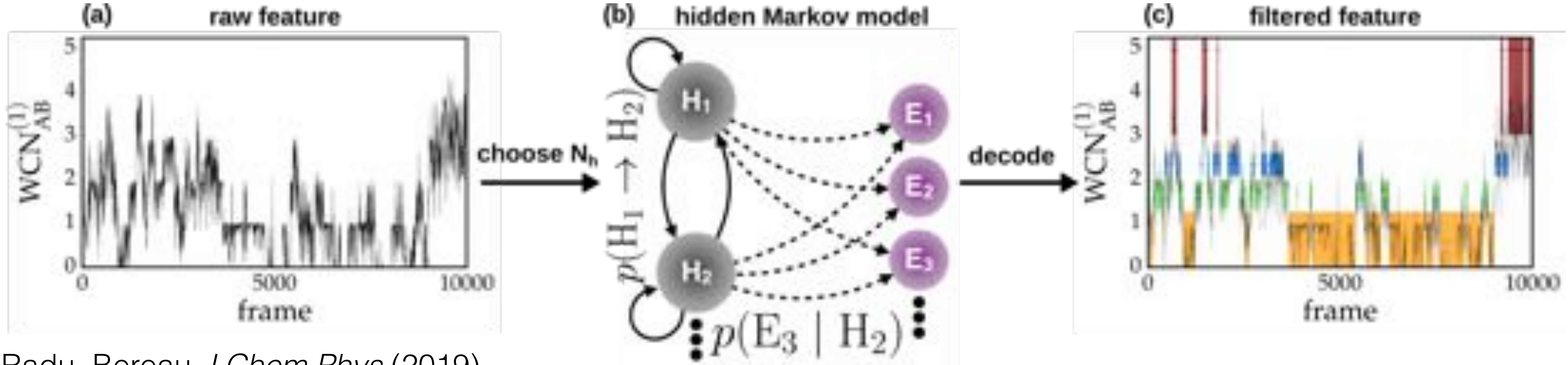



# Automated detection of many-particle solvation states

**Coordination numbers as input features**

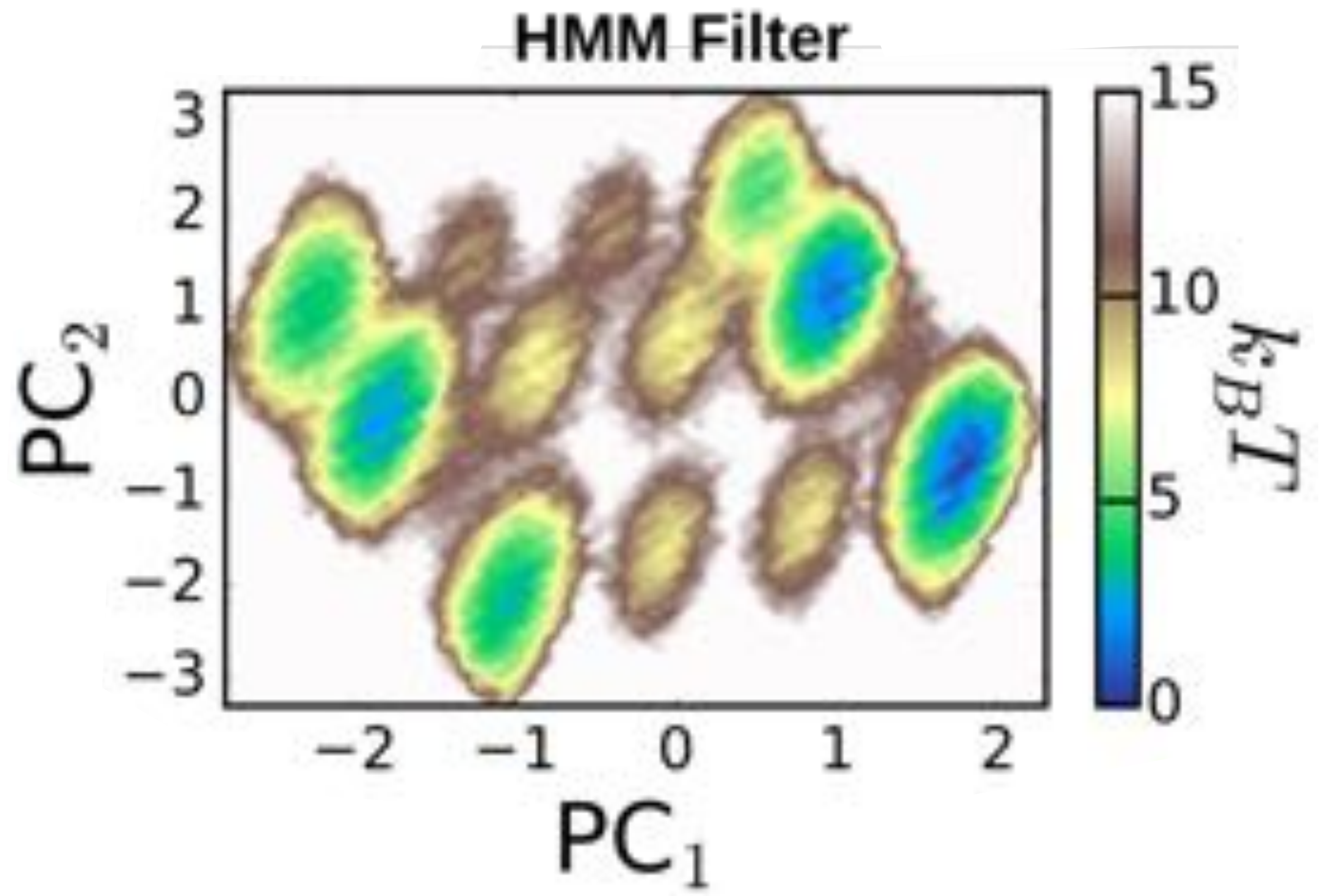




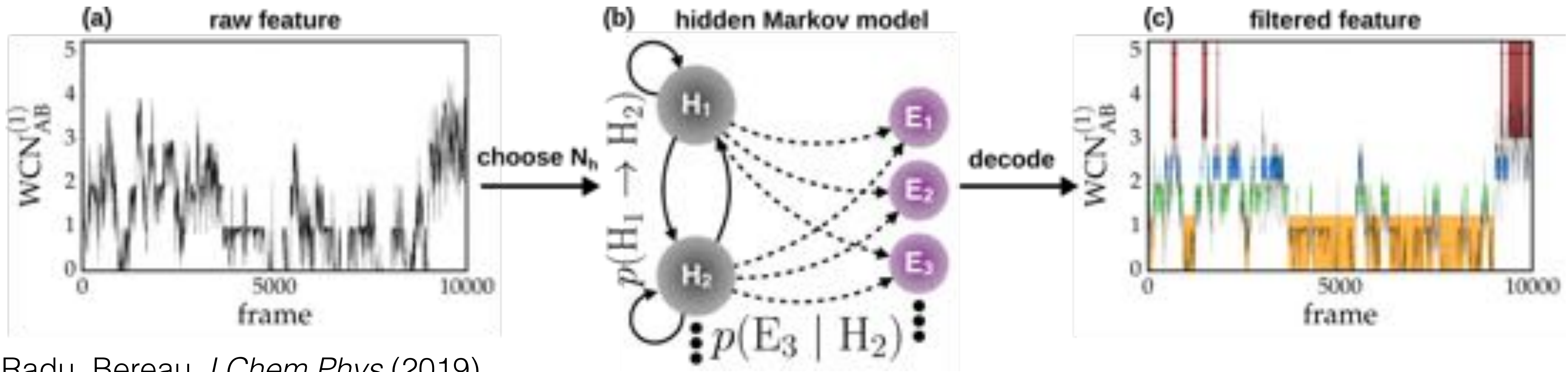




JFR, Radu, Bereau *J Chem Phys* (2019)

# Automated detection of many-particle solvation states

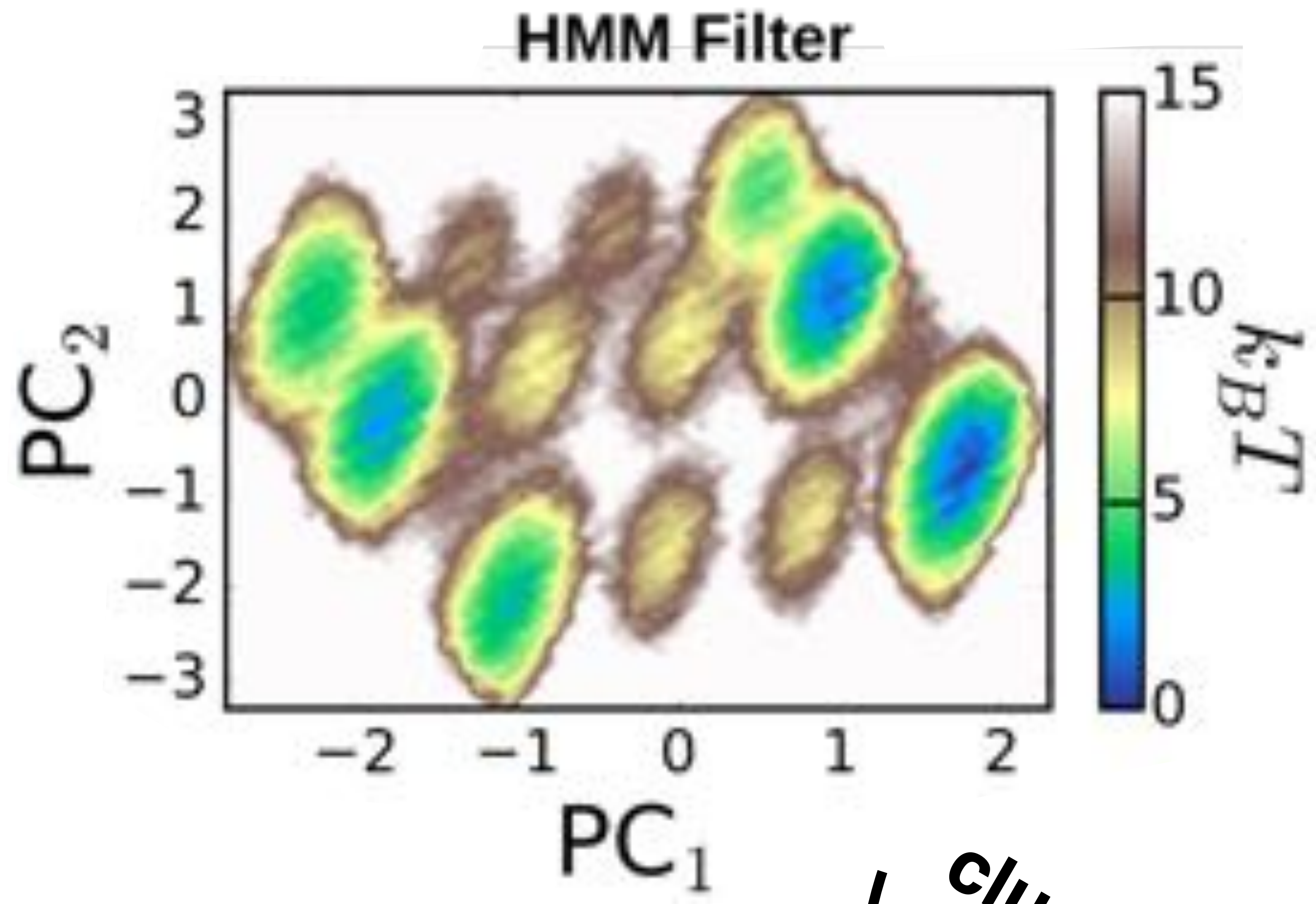


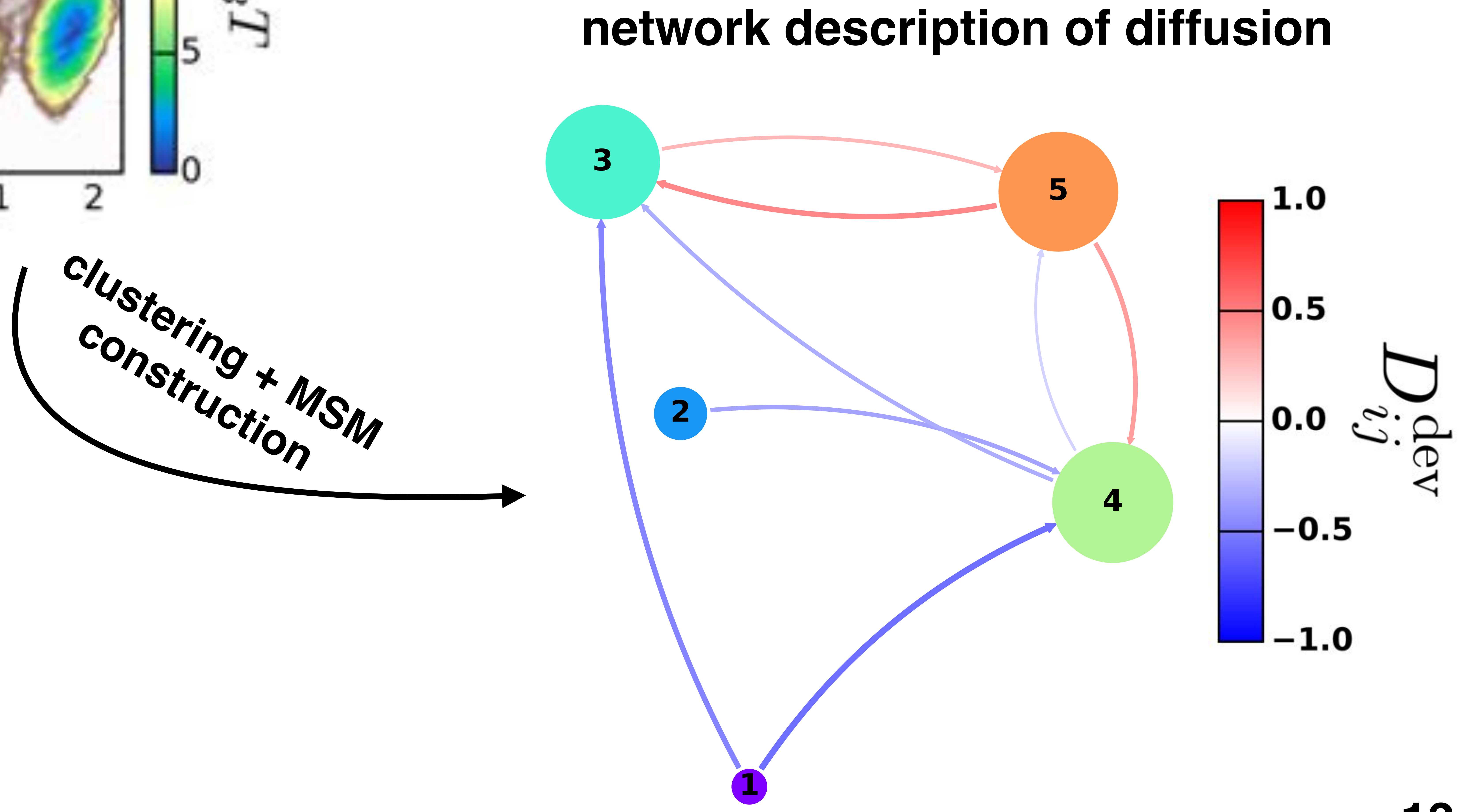

**Structured free-energy landscape describing distinct solvation states**




JFR, Radu, Bereau *J Chem Phys* (2019)

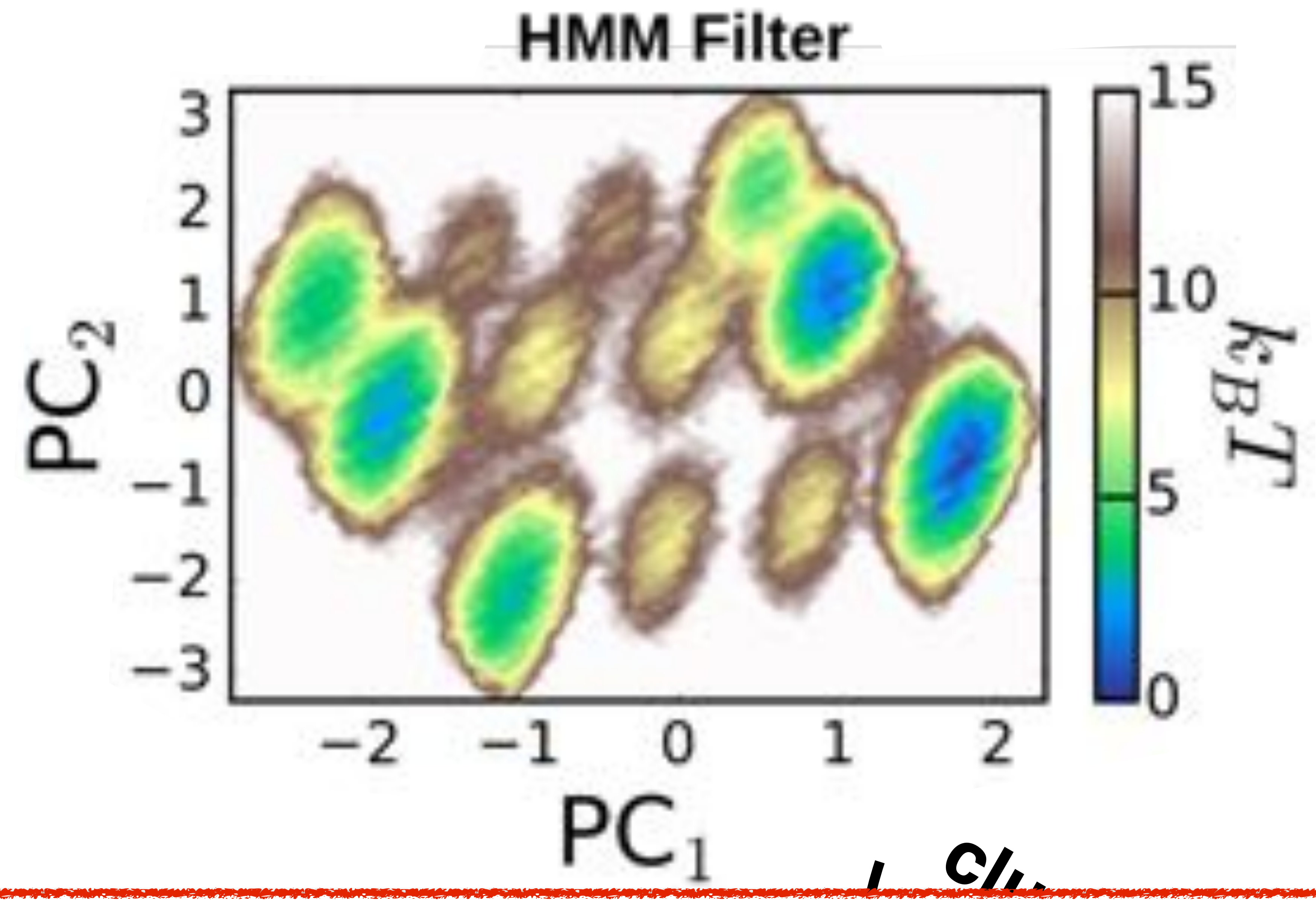
# Automated detection of many-particle solvation states



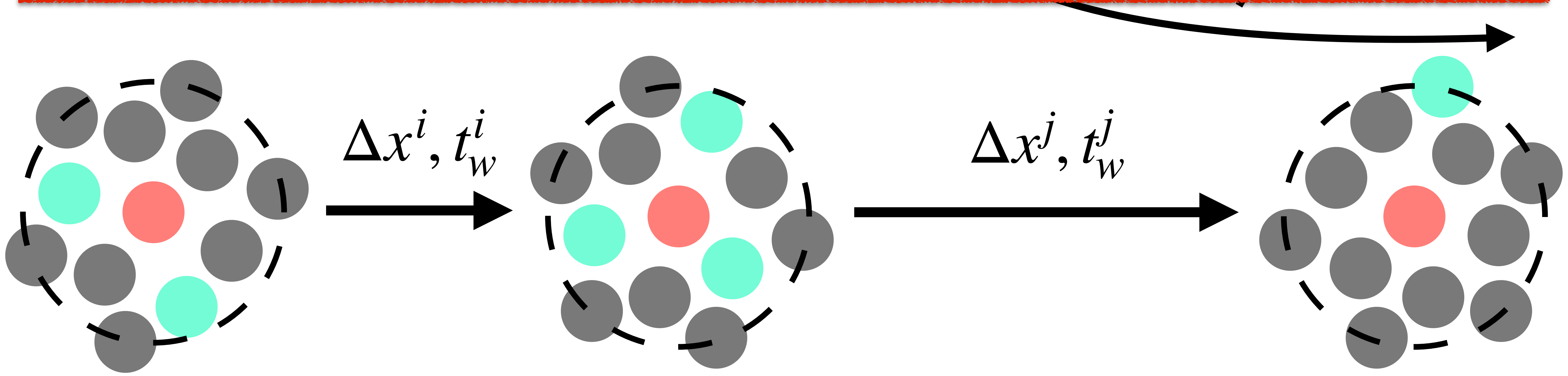




JFR, Radu, Bereau *J Chem Phys* (2019)

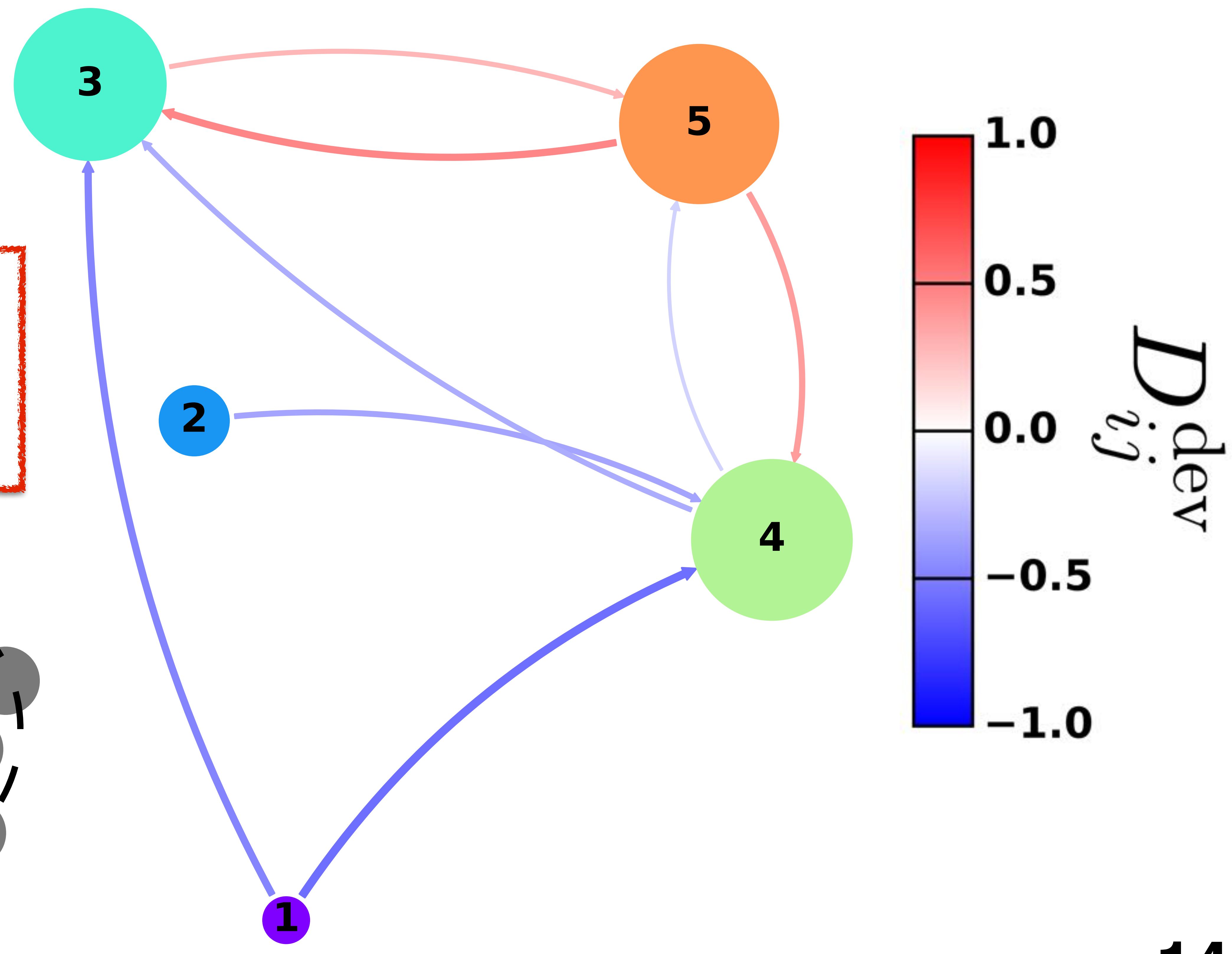
# Automated detection of many-particle solvation states




- **Mechanistic insight into solvation shell dynamics**
- **Quantification of dynamic heterogeneity**

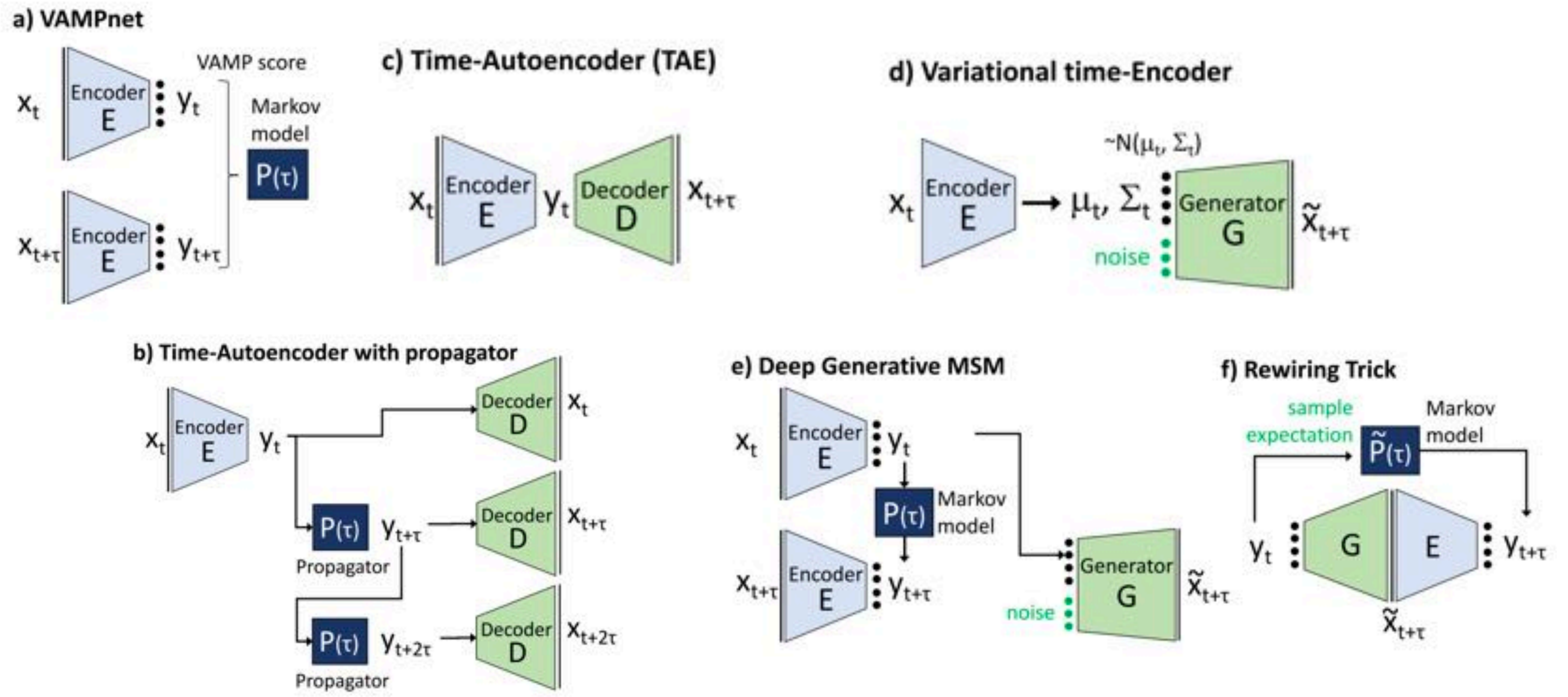



**network description of diffusion**



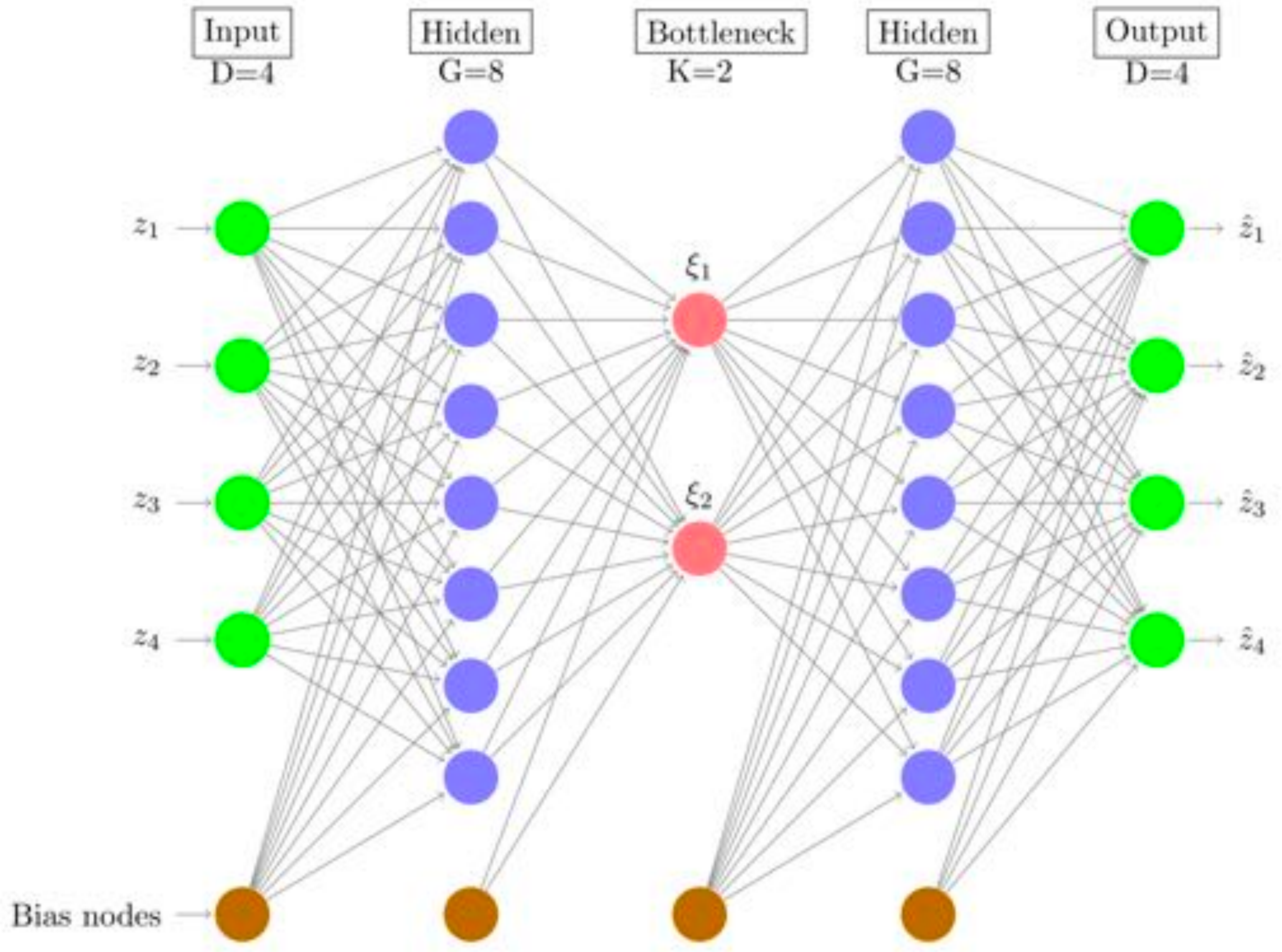

JFR, Radu, Bereau *J Chem Phys* (2019)

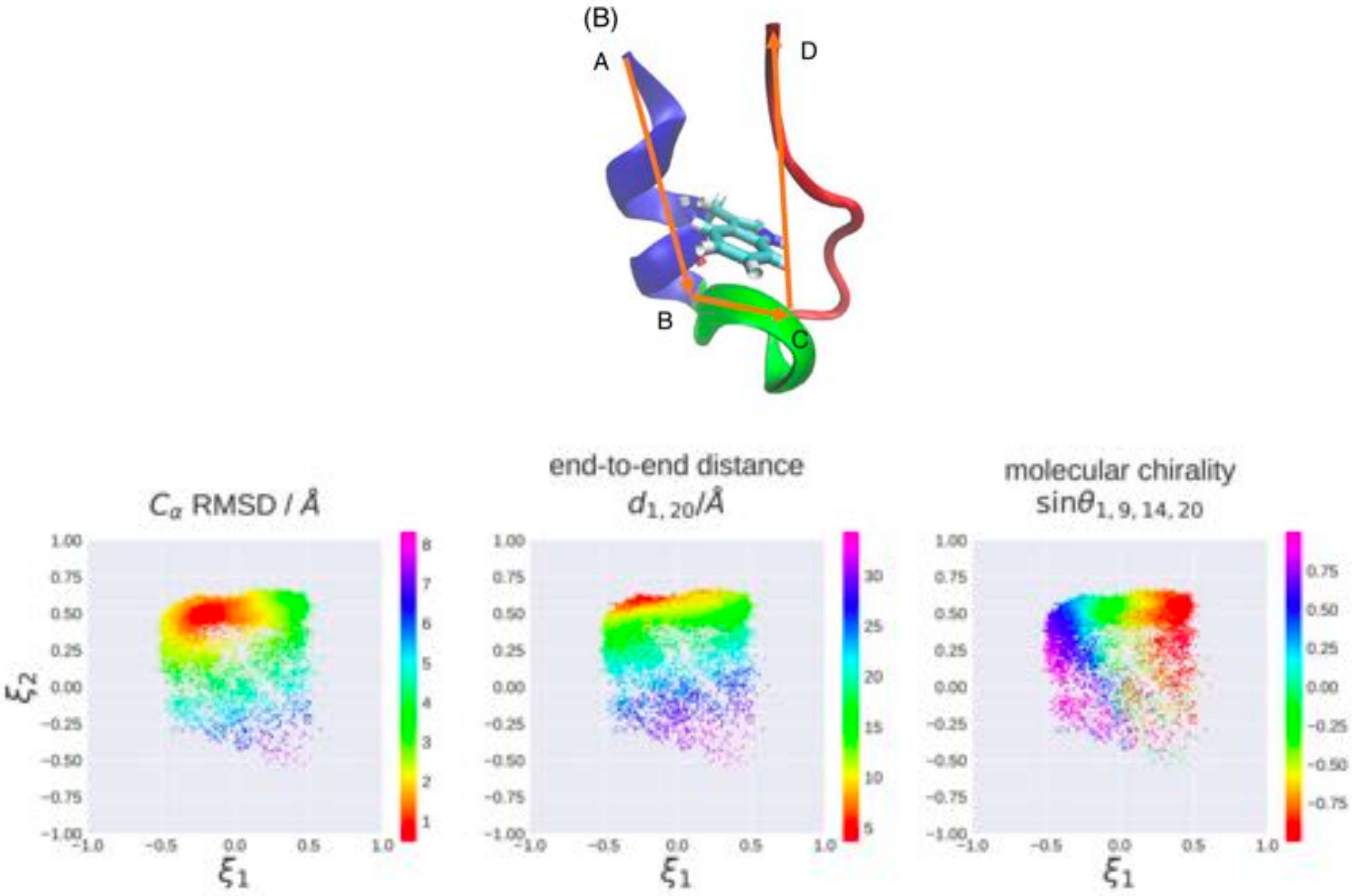
# VAEs for MD analysis




Noe *ArXiv* **(2019)** “Machine Learning for Molecular Dynamics on Long Timescales”

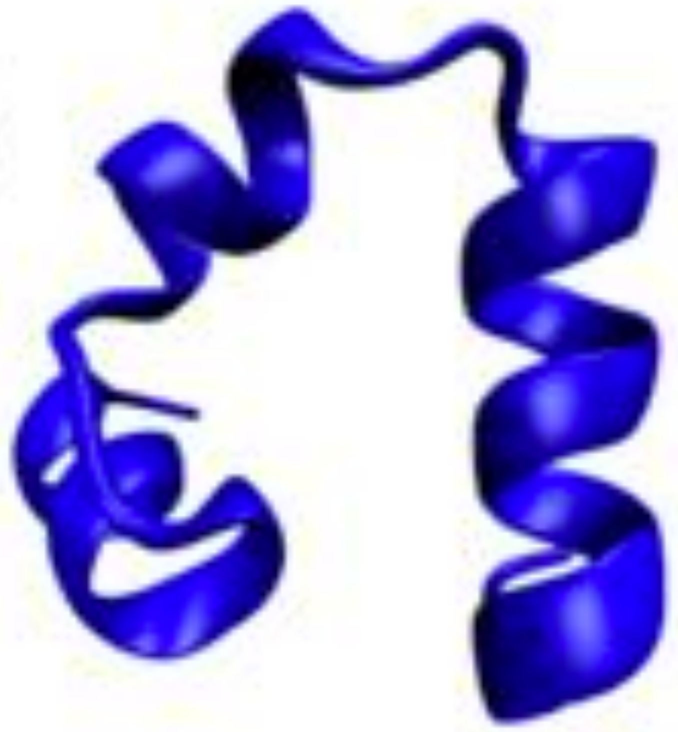
# VAEs for collective variable discovery

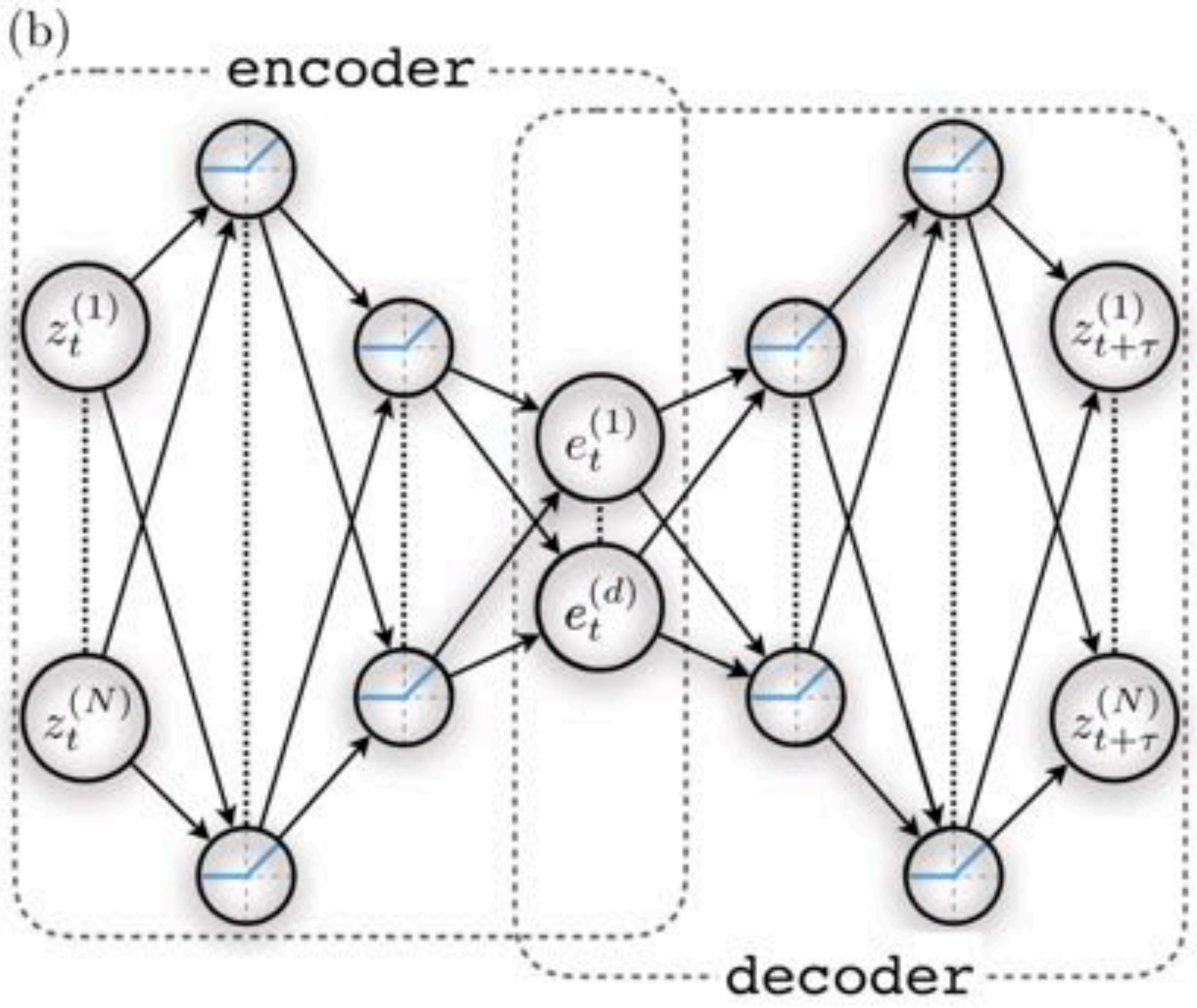


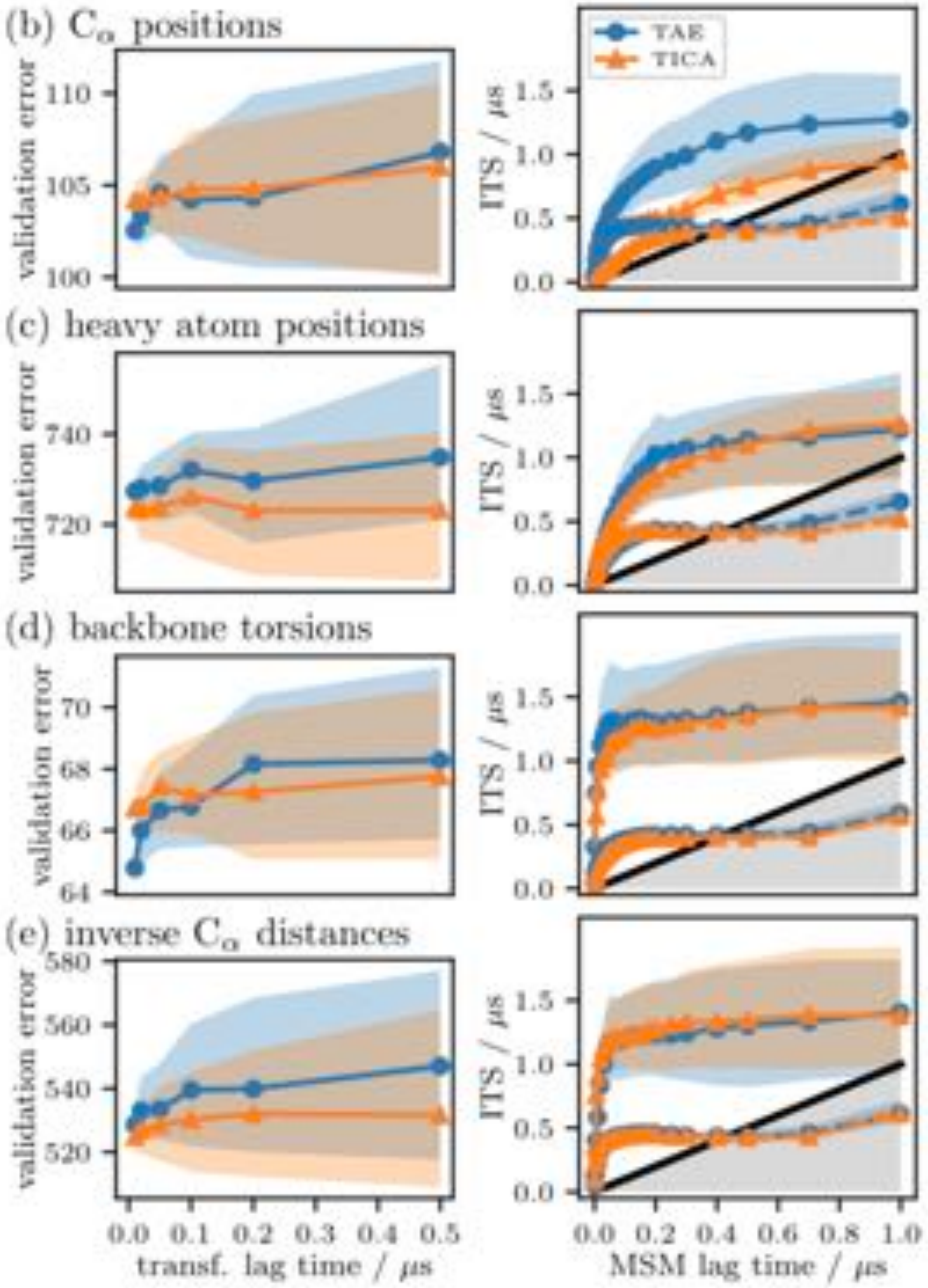



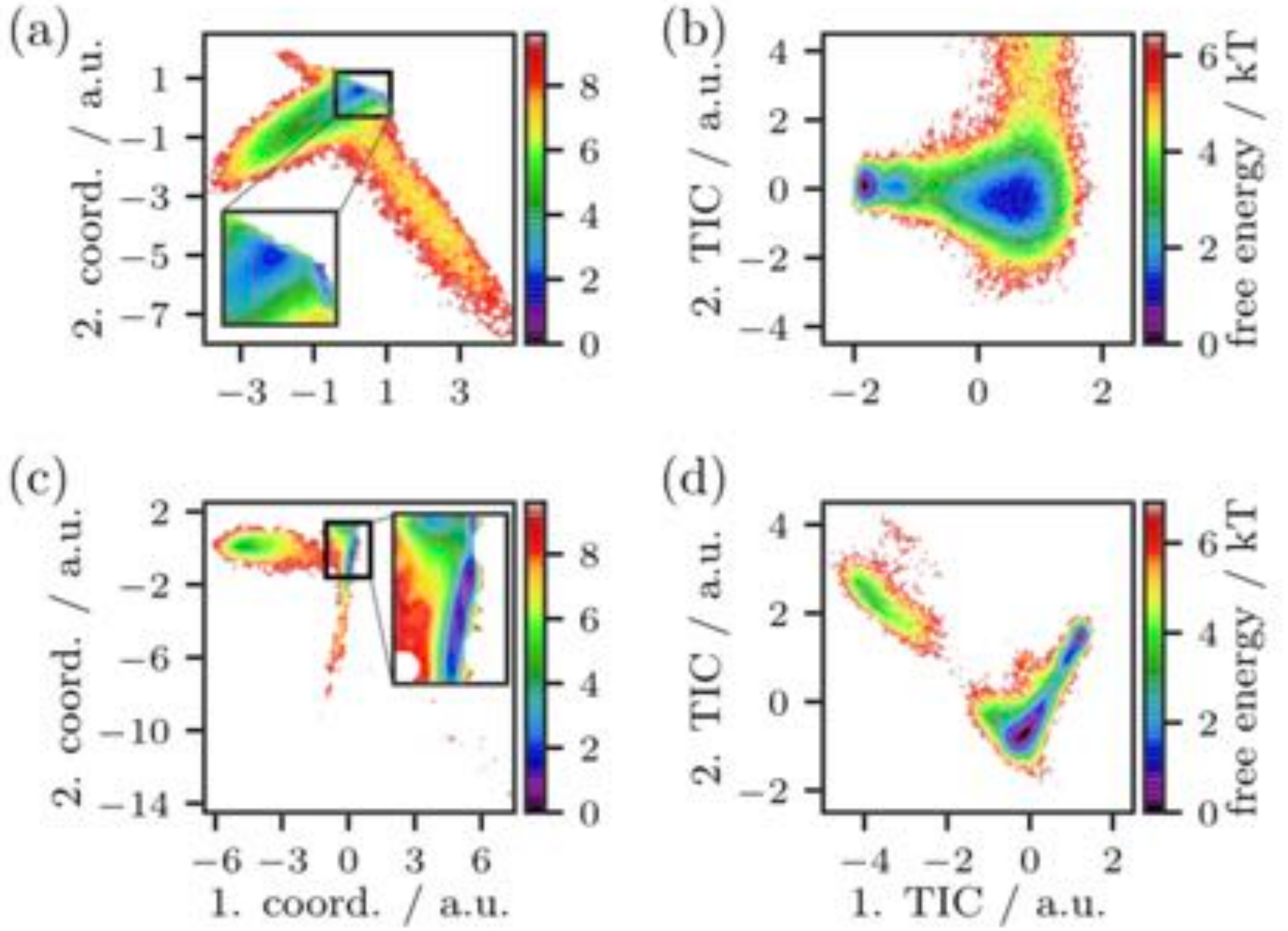

Chen, Ferguson *J Comp Chem* **(2018)** “Molecular enhanced sampling with autoencoders: On-the-fly collective variable discovery and accelerated free energy landscape exploration”

# Time-lagged autoencoders





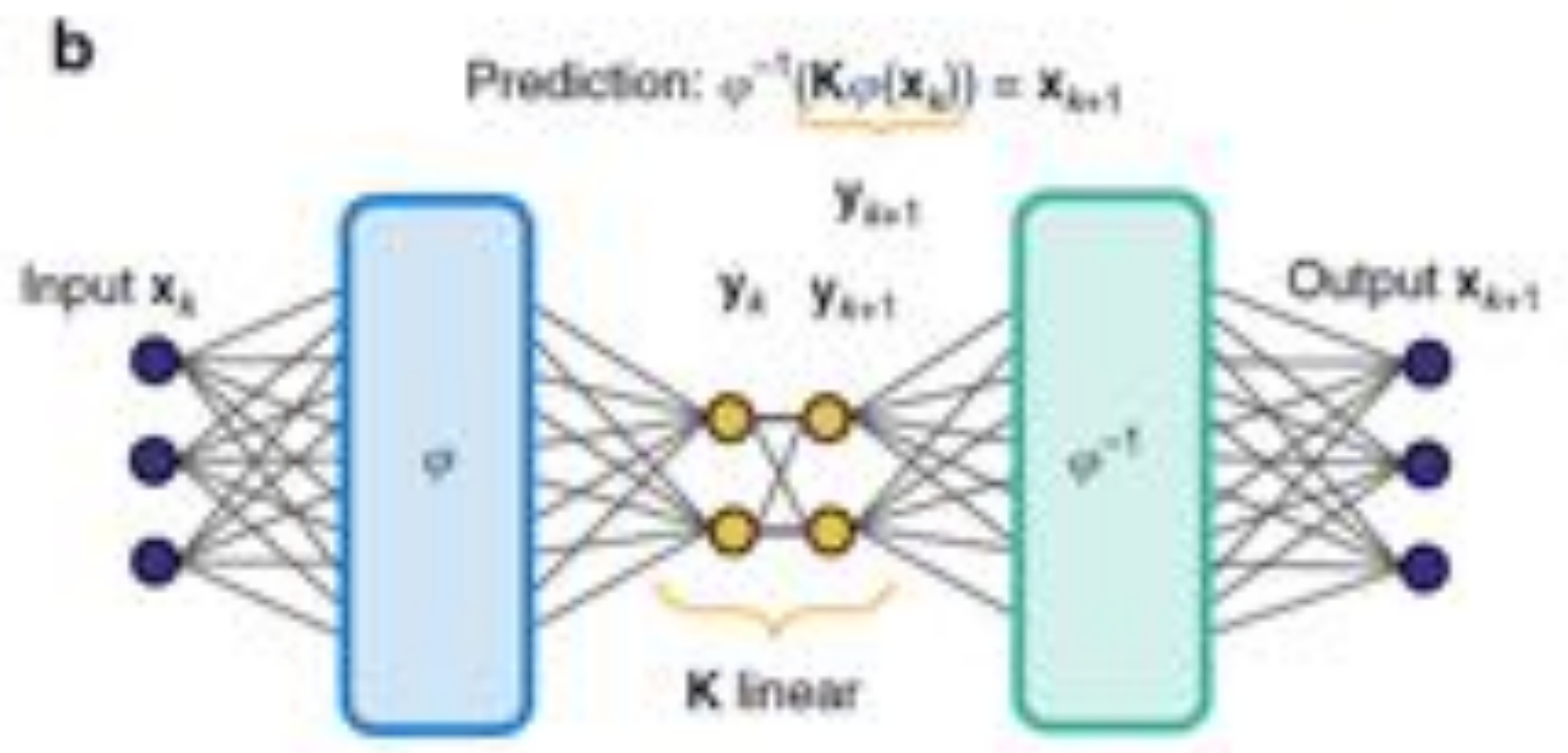




Christoph Wehmeyer and Frank Noe *JCP* **(2018)** “Time-lagged autoencoders: Deep learning of slow collective variables for molecular kinetics”

# Data driven approach to Koopman Theory

## Loss functions




$$\mathcal{L} = \alpha_1(\mathcal{L}_{\text{recon}} + \mathcal{L}_{\text{pred}}) + \mathcal{L}_{\text{lin}} + \alpha_2 \mathcal{L}_{\infty} + \alpha_3 \|\mathbf{W}\|_2^2$$

$$\mathcal{L}_{\text{recon}} = \|\mathbf{x}_1 - \varphi^{-1}(\varphi(\mathbf{x}_1))\|_{\text{MSE}}$$

$$\mathcal{L}_{\text{pred}} = \frac{1}{S_p}\sum_{m=1}^{S_p}\|\mathbf{x}_{m+1} - \varphi^{-1}(K^m\varphi(\mathbf{x}_1))\|_{\text{MSE}}$$

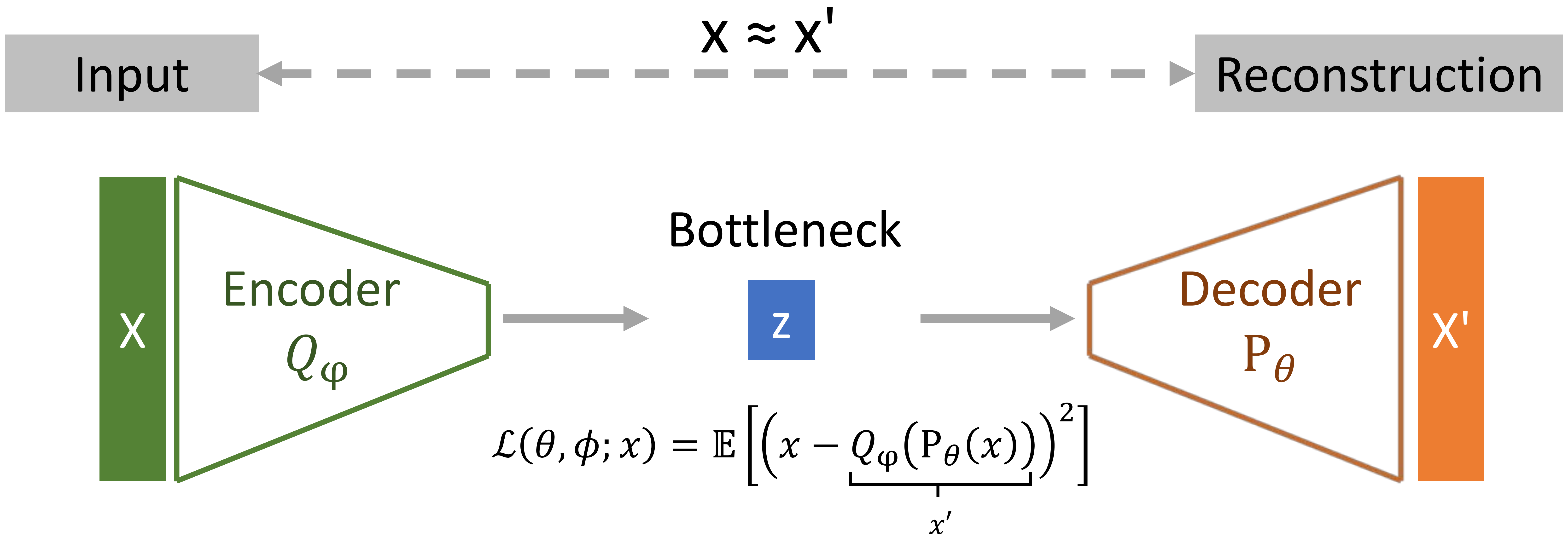
$$\mathcal{L}_{\text{lin}} = \frac{1}{T-1}\sum_{m=1}^{T-1}\|\varphi(\mathbf{x}_{m+1}) - K^m\varphi(\mathbf{x}_1)\|_{\text{MSE}}$$

$$\mathcal{L}_{\infty} = \|\mathbf{x}_1 - \varphi^{-1}(\varphi(\mathbf{x}_1))\|_{\infty} + \|\mathbf{x}_2 - \varphi^{-1}(K\varphi(\mathbf{x}_1))\|_{\infty}$$

Lusch *et al. Nat Comm* **(2018)** “Deep learning for universal linear embeddings of nonlinear dynamics”

# Interpretable embeddings for molecular kinetics using deep learning

## Standard Autoencoder

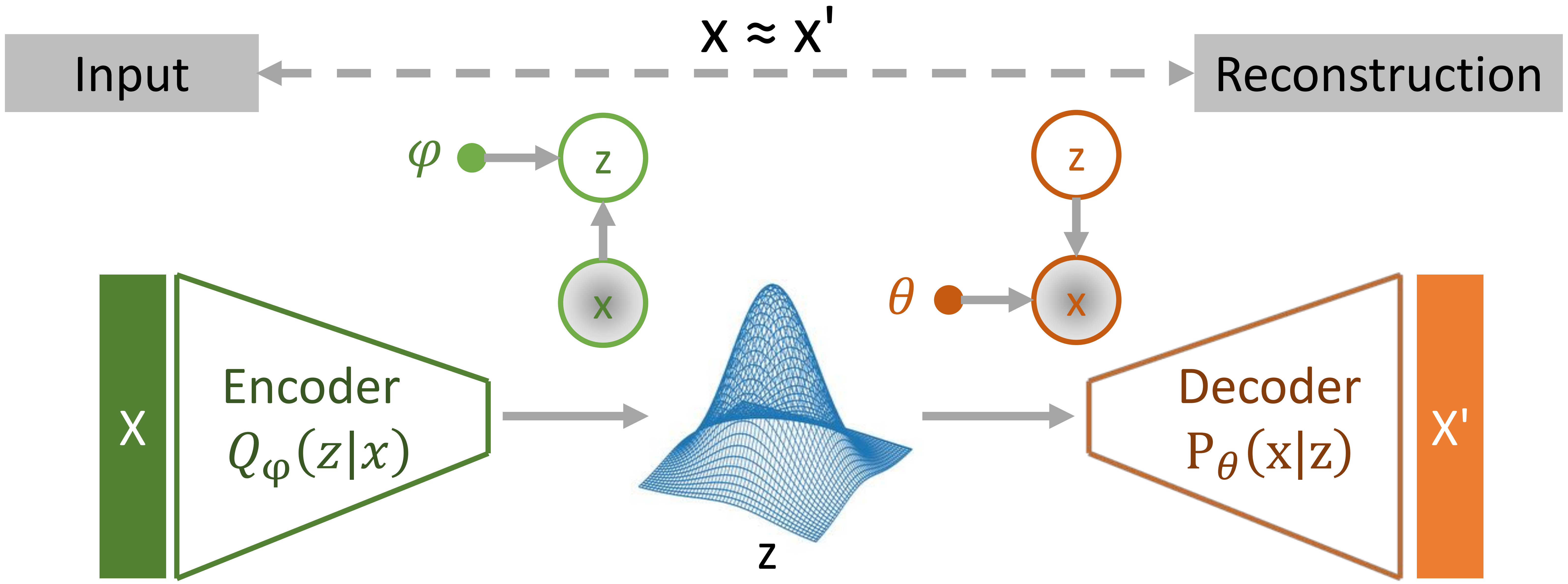



Yasemin Bozkurt Varolgunes

Bozkurt Varolgunes, Bereau, JFR *in progress*

# Interpretable embeddings for molecular kinetics using deep learning

## Variational Autoencoder with Unimodal Gaussian Prior




$$\mathcal{L}(\theta,\phi;x) = \mathbb{E}_{Q_\phi(x,z)}[\log P_\theta(x,z) - \log Q_\phi(z|x)]$$

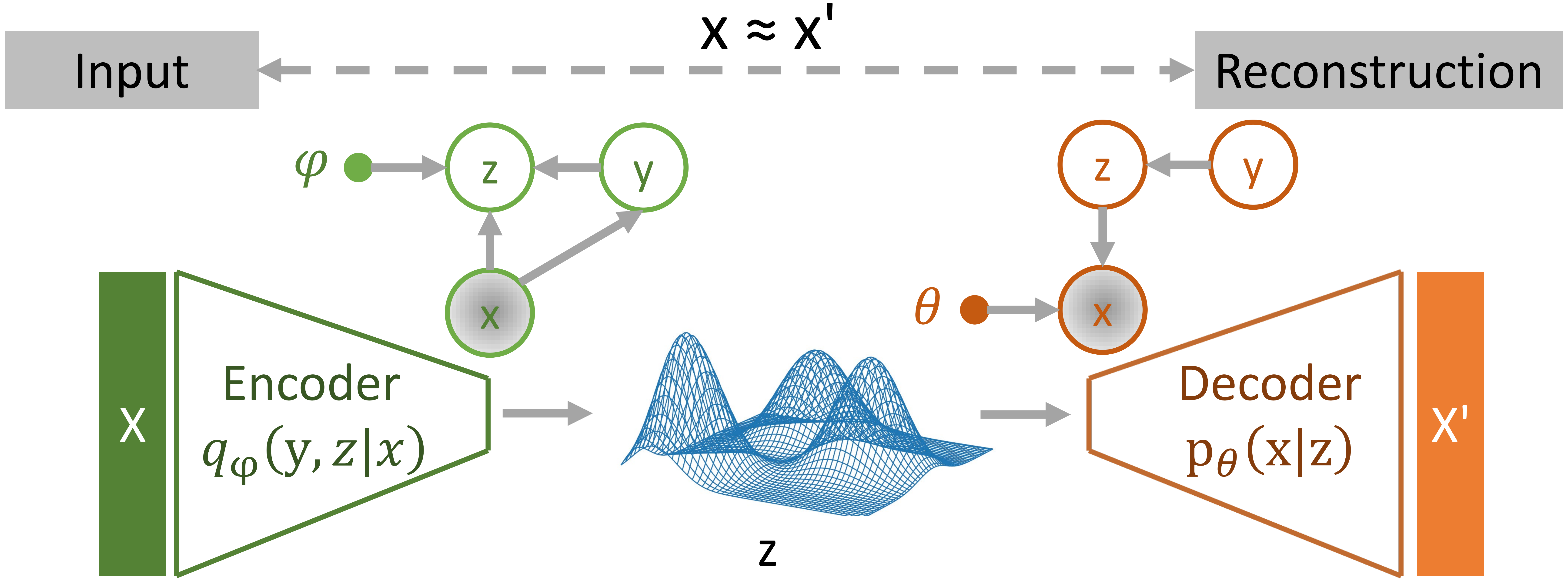
$$\mathcal{L}(\theta,\phi;x) = \underbrace{\boxed{\mathbb{E}_{Q_\phi(z|x)}[\log P_\theta(x|z)]}}_{\text{Reconstruction}} - \underbrace{\boxed{KL\left(Q_\phi(z|x)||P(z)\right)}}_{\text{Regularization}}$$

Yasemin Bozkurt Varolgunes

Bozkurt Varolgunes, Bereau, JFR *in progress*

# Interpretable embeddings for molecular kinetics using deep learning

## Gaussian Mixture Variational Autoencoder




$$\mathcal{L}(\theta, \phi; x) = \mathbb{E}_{q_\phi(y,z|x)}\left[\log p_\theta(x,y,z) - \log q_\phi(y,z|x)\right]$$

$$\mathcal{L}(\theta, \phi; x) = \mathbb{E}_{q_\phi(y,z|x)}\left[\underbrace{\log \frac{p_\theta(y)}{q_\phi(y|x)}}_{\textbf{clustering}} + \underbrace{\log \frac{p_\theta(z|y)}{q_\phi(z|x,y)}}_{\textbf{regularization}} + \underbrace{\log p_\theta(x|z)}_{\textbf{reconstruction}}\right]$$

- Dimensionality reduction + Clustering
- Synthetic trajectory generation

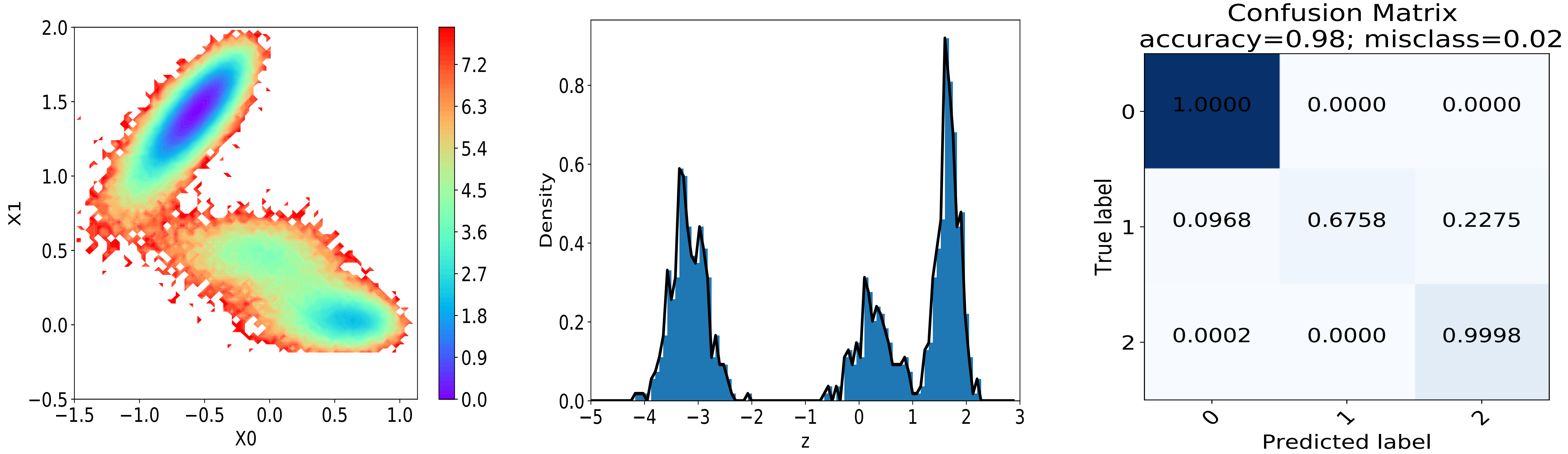
Yasemin Bozkurt Varolgunes

Bozkurt Varolgunes, Bereau, JFR *in progress*

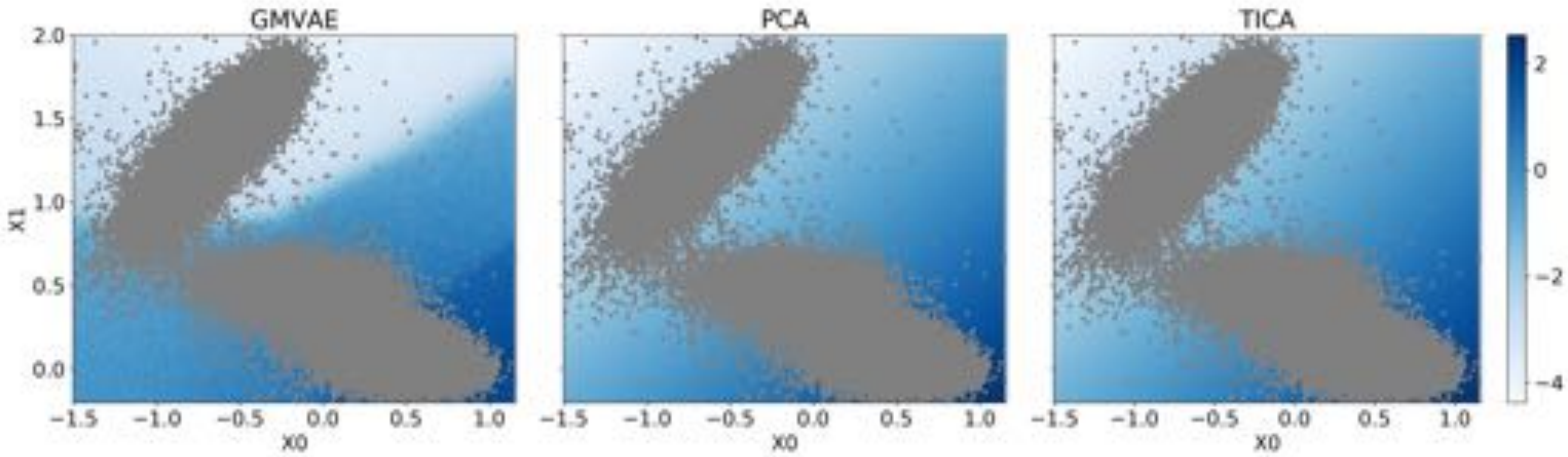
Interpretable embeddings for molecular kinetics using deep learning

## Two dimensional Muller-Brown potential

$$V(x) = \sum_{j=1}^{4} A_j \exp[a_j(x_1 - X_j)^2 + b_j(x_1 - X_j)(x_2 - Y_j) + c_j(x_2 - Y_j)^2]$$



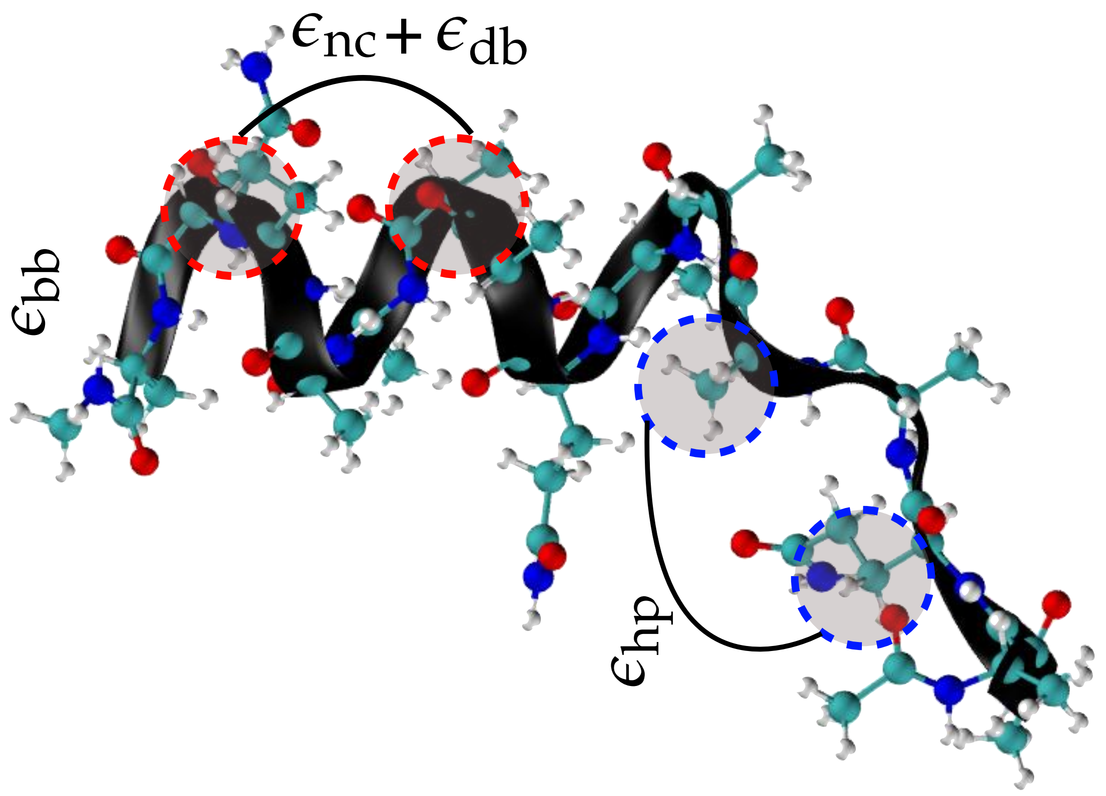

| | 0 | 1 | 2 |
|---|---|---|---|
| 0 | 1.0000 | 0.0000 | 0.0000 |
| 1 | 0.0968 | 0.6758 | 0.2275 |
| 2 | 0.0002 | 0.0000 | 0.9998 |

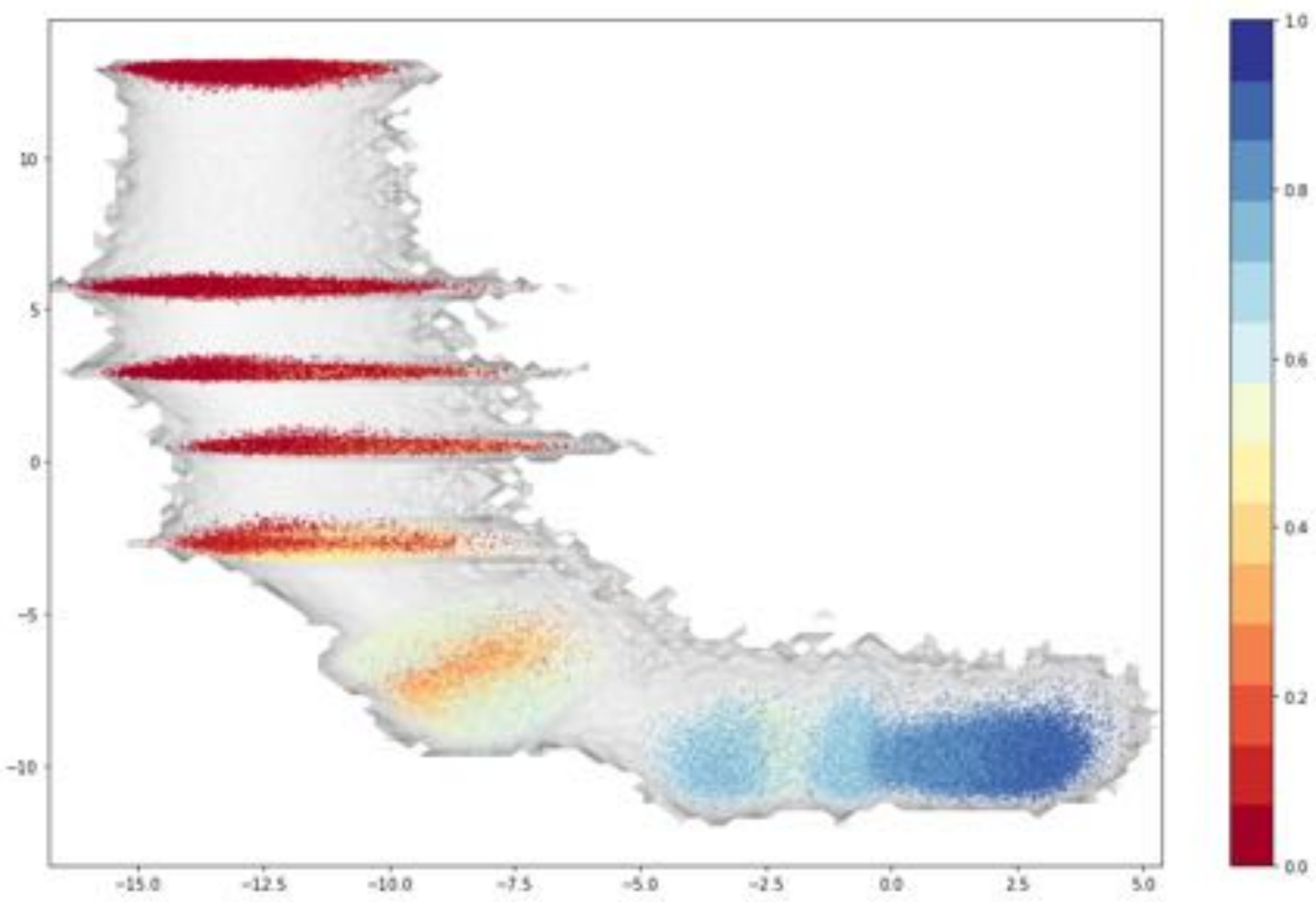


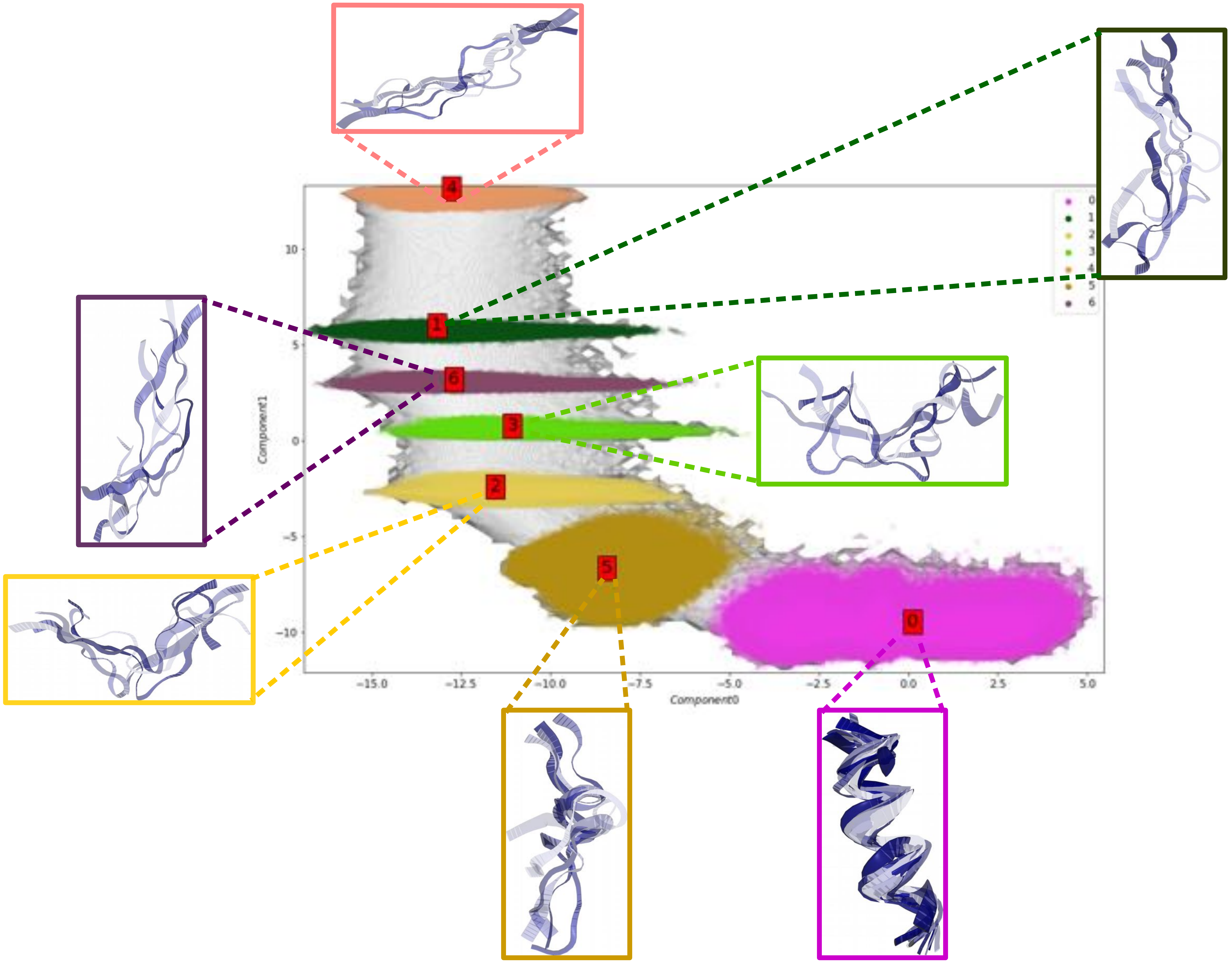

Bozkurt Varolgunes, Bereau, JFR *in progress*

# Interpretable embeddings for molecular kinetics using deep learning

**(AAQAA)$_3$**






Bozkurt Varolgunes, Bereau, JFR *in progress*

# Thank you for your attention!

Contact, papers, and current updates on my research @ RudzinskiResearch.com