Machine learning molecular kinetics and coarse-grained molecular dynamics forcefields

ML4MS: Young Researcher's Workshop on Machine Learning for Materials Science

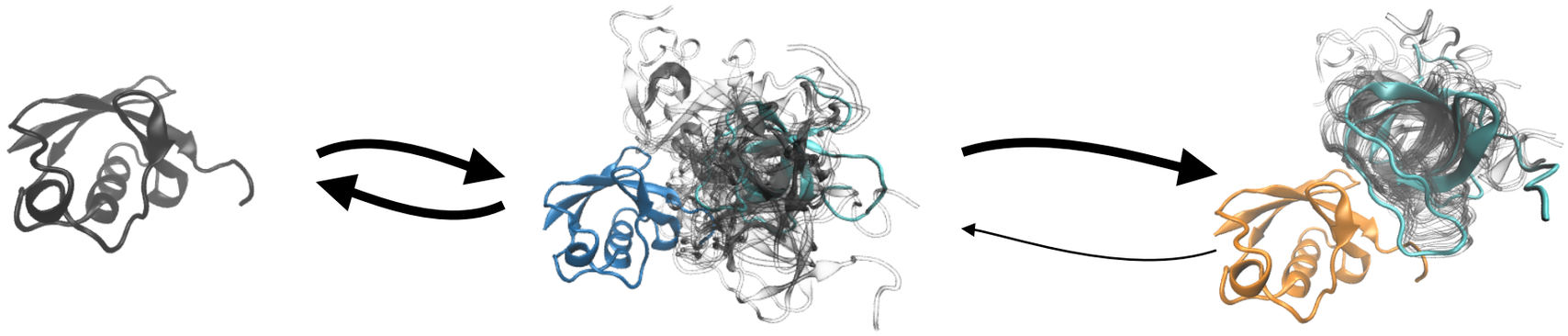
Simon Olsson
FU Berlin

Aalto University, Helsinki, Finland
6.-10. May 2018

@smnlssn
simon.olsson@fu-berlin.de

The role of conformational change in biology: a mechanistic view

Protein-protein and protein-ligand binding



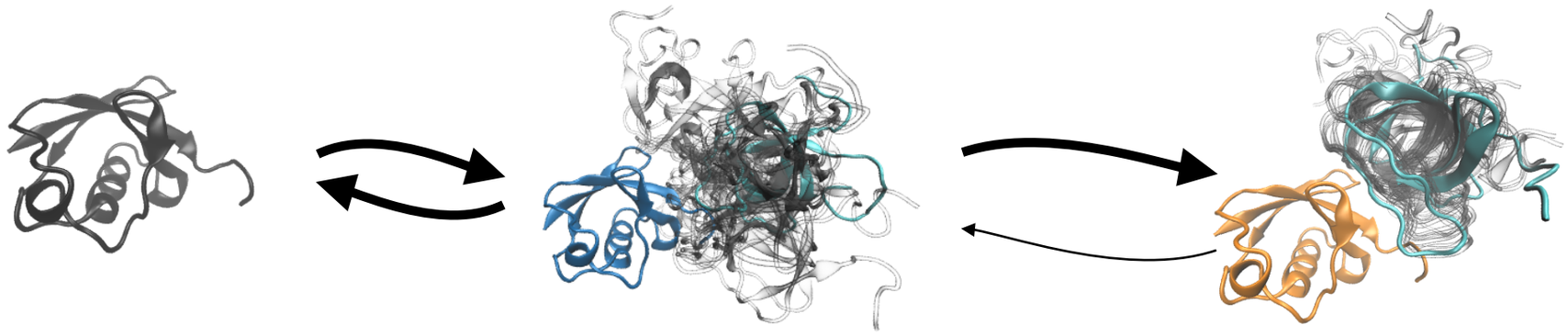
Quantities of interest:

$\Delta G_{\text{bind}}, k_{\text{on}}$ and k_{off}

With Chakrabarty … Weikl, Noé and Griesinger

The role of conformational change in biology: a mechanistic view

Protein-protein and protein-ligand binding

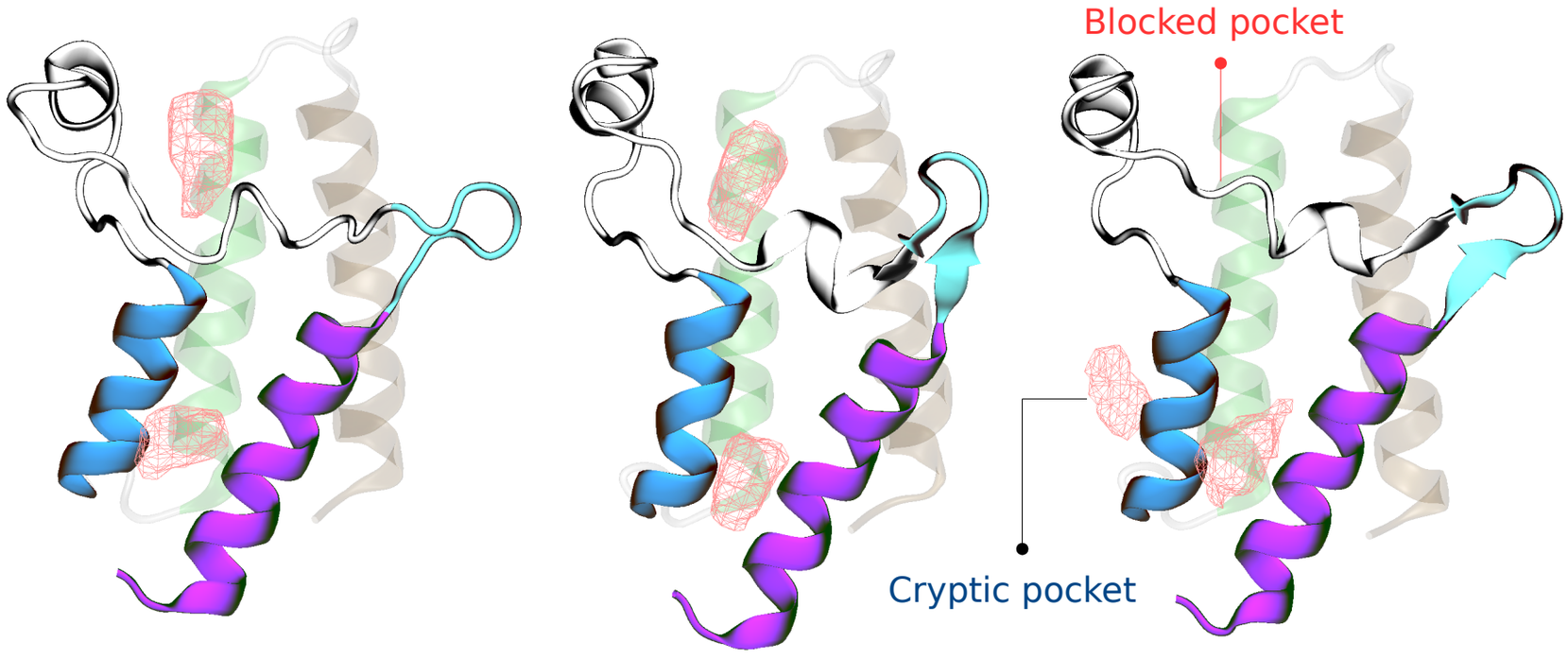


Quantities of interest:

ΔG_{bind}, k_{on} and k_{off}

With Chakrabarty … Weikl, Noé and Griesinger

Allosteric regulation — Cryptic pockets



Quantities of interest:

ΔG_{pocket}, τ_{open} and k_{open}

With Raich

Molecular simulations as a tool to study conformational change

$$\underbrace{U(x)}_{\text{Model of physical interactions}} \xrightarrow{\text{Sampling}} \{x_0, x_t, \ldots, x_T\}$$

Molecular simulations as a tool to study conformational change

$$\underbrace{U(x)}_{\text{Model of physical interactions}} \xrightarrow{\text{Sampling}} \{x_0, x_t, \ldots, x_T\}$$

Important conformational states are often inter-connected by rare events

$\longrightarrow \{x_0, x_t, \ldots, x_T\}$ are all from the same meta-stable state

Molecular simulations as a tool to study conformational change

$$\underbrace{U(x)}_{\text{Model of physical interactions}} \xrightarrow{\text{Sampling}} \{x_0, x_t, \ldots, x_T\}$$

Important conformational states are often inter-connected by rare events

$\longrightarrow \{x_0, x_t, \ldots, x_T\}$ are all from the same meta-stable state

Naive solution: increase T

Molecular simulations as a tool to study conformational change

$$\underbrace{U(x)}_{\text{Model of physical interactions}} \xrightarrow{\text{Sampling}} \{x_0, x_t, \ldots, x_T\}$$

Important conformational states are often inter-connected by rare events

$\longrightarrow \{x_0, x_t, \ldots, x_T\}$ are all from the same meta-stable state

Naive solution: increase T $\longrightarrow$ Expensive and slow!

Molecular simulations as a tool to study conformational change

$$\underbrace{U(x)}_{\text{Model of physical interactions}} \xrightarrow{\text{Sampling}} \{x_0, x_t, \ldots, x_T\}$$

Important conformational states are often inter-connected by rare events

$\longrightarrow \{x_0, x_t, \ldots, x_T\}$ are all from the same meta-stable state

Naive solution: increase T $\longrightarrow$ Expensive and slow!

Adaptive strategy

$U(x_a), U(x_b) \ldots U(x_z)$

Multiple initial conditions simulated in parallel

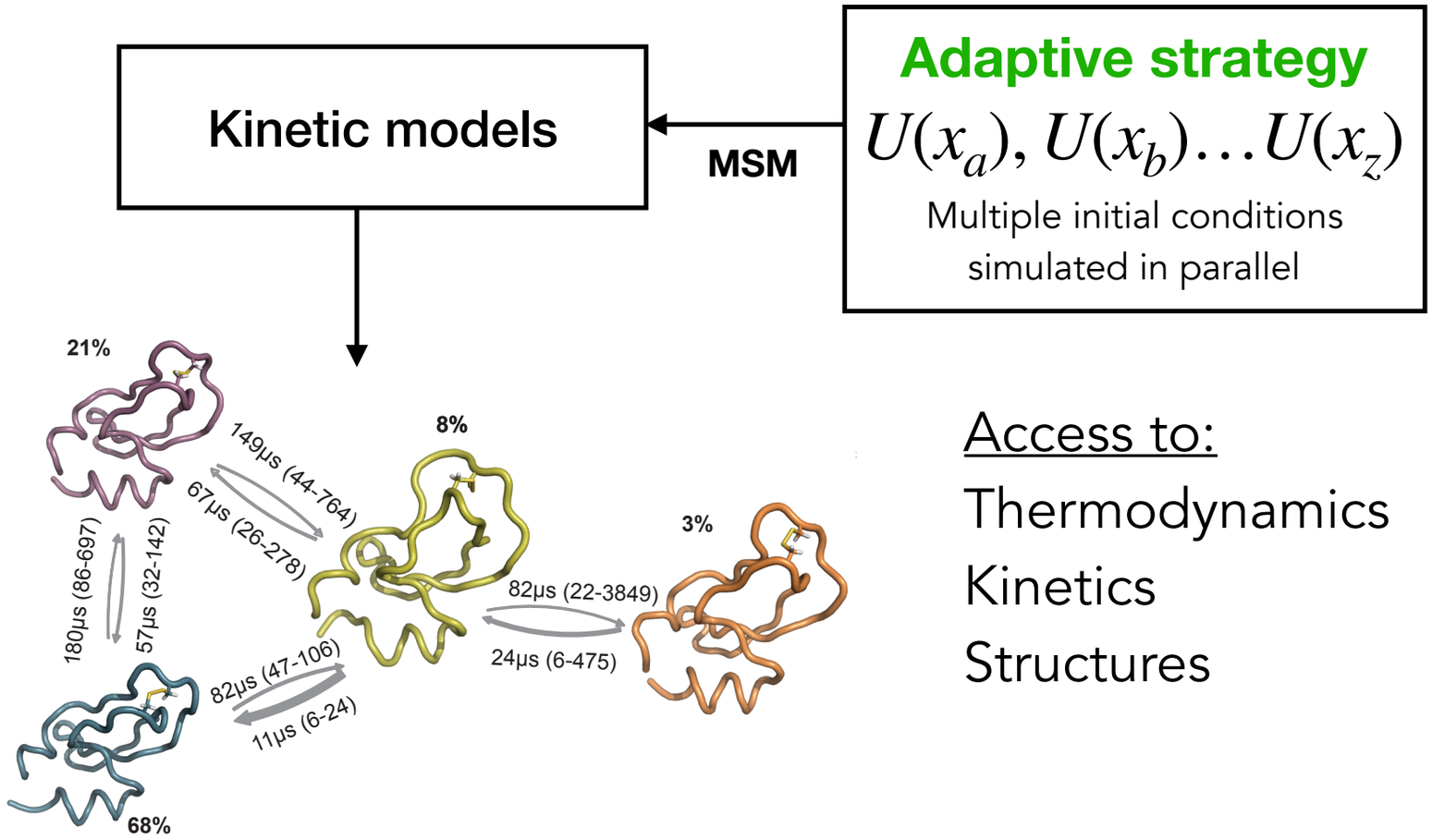
Molecular simulations as a tool to study conformational change

$$\underbrace{U(x)}_{\text{Model of physical interactions}} \xrightarrow{\text{Sampling}} \{x_0, x_t, \ldots, x_T\}$$

Important conformational states are often inter-connected by rare events

$\longrightarrow \{x_0, x_t, \ldots, x_T\}$ are all from the same meta-stable state

Naive solution: increase T $\longrightarrow$ Expensive and slow!



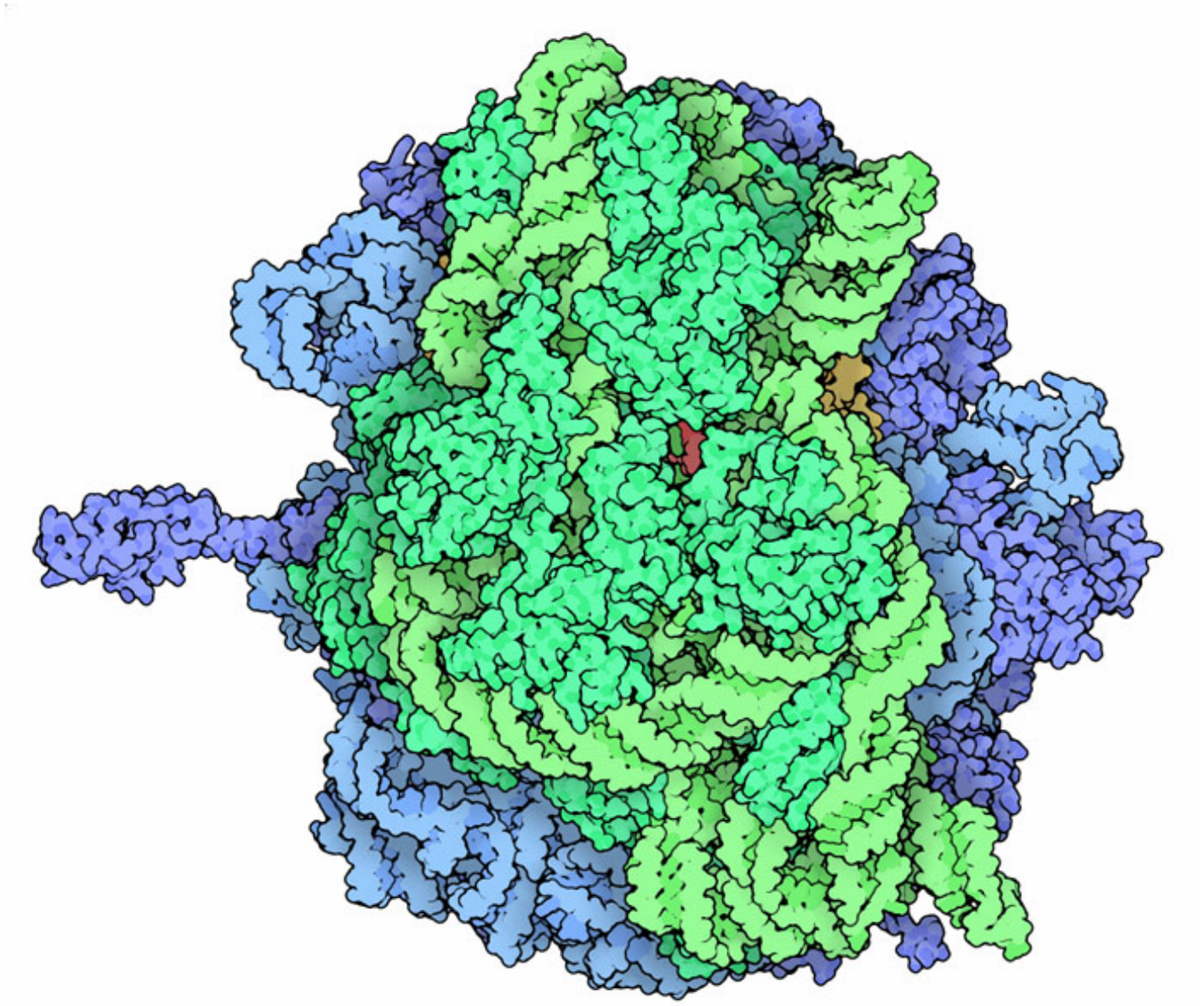
Access to:

Thermodynamics

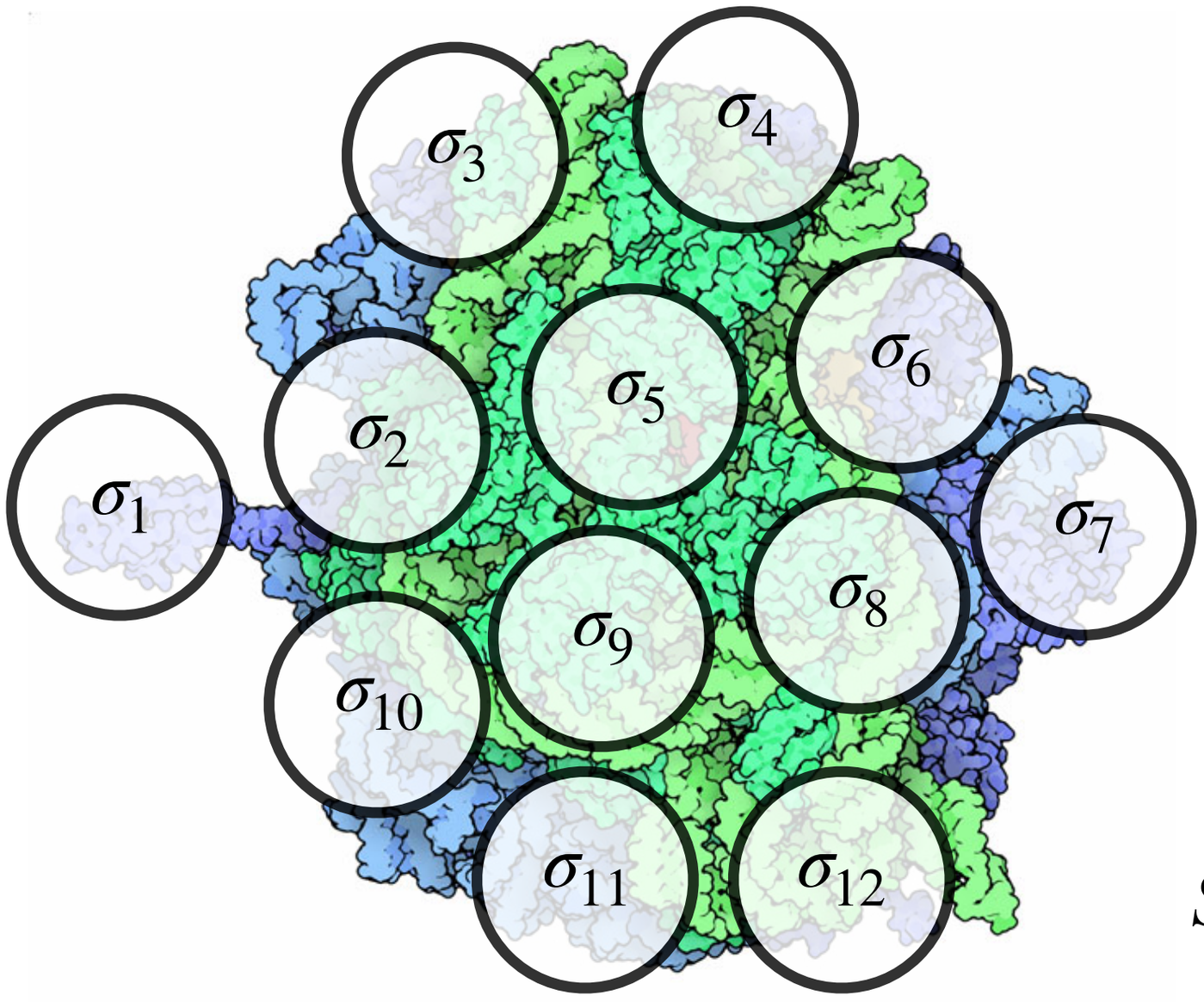
Kinetics

Structures

Large molecular systems may adopt an intractable number of states

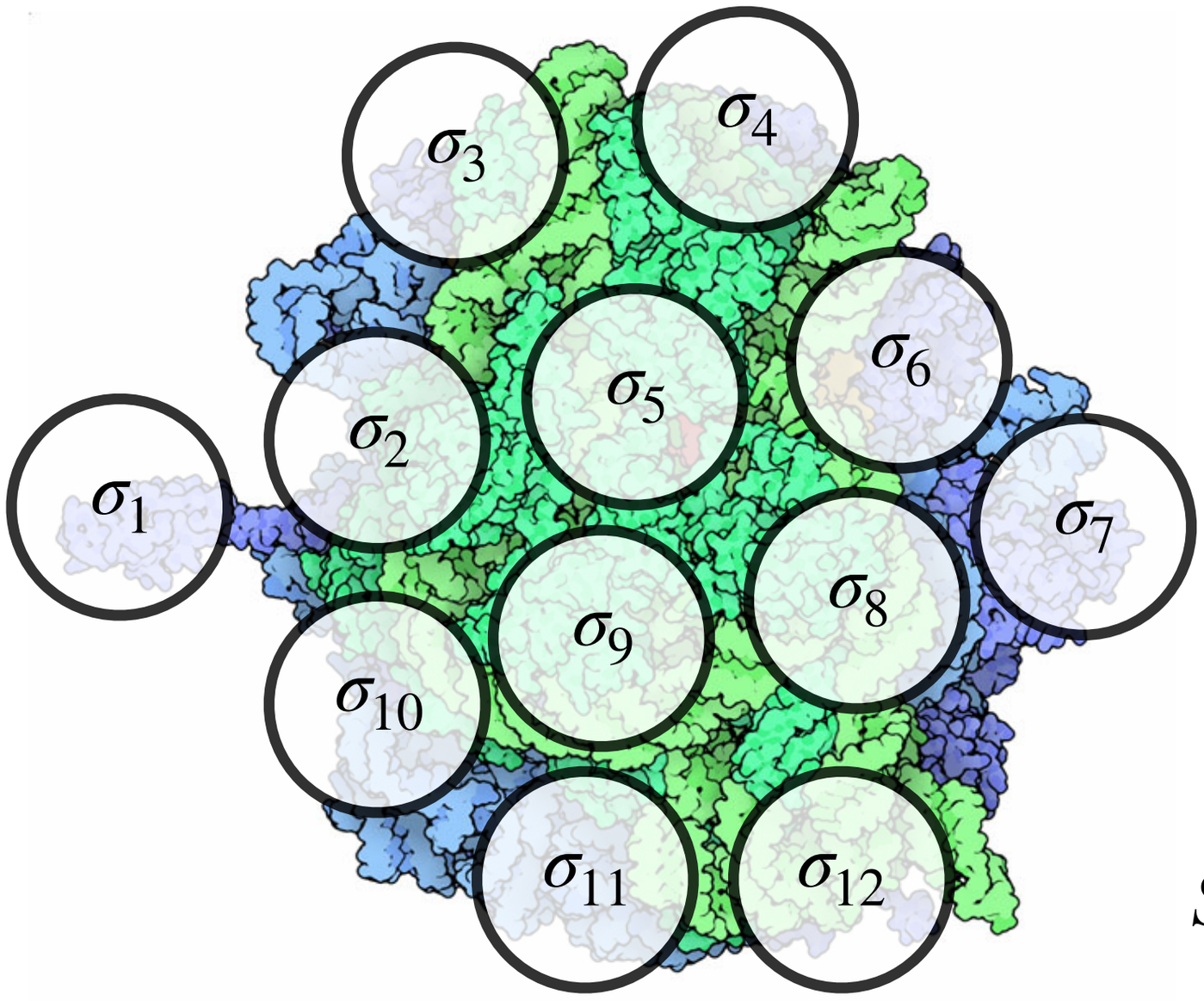


Large molecular systems may adopt an intractable number of states



$$S = \{\sigma_1, \ldots, \sigma_{12}\}$$

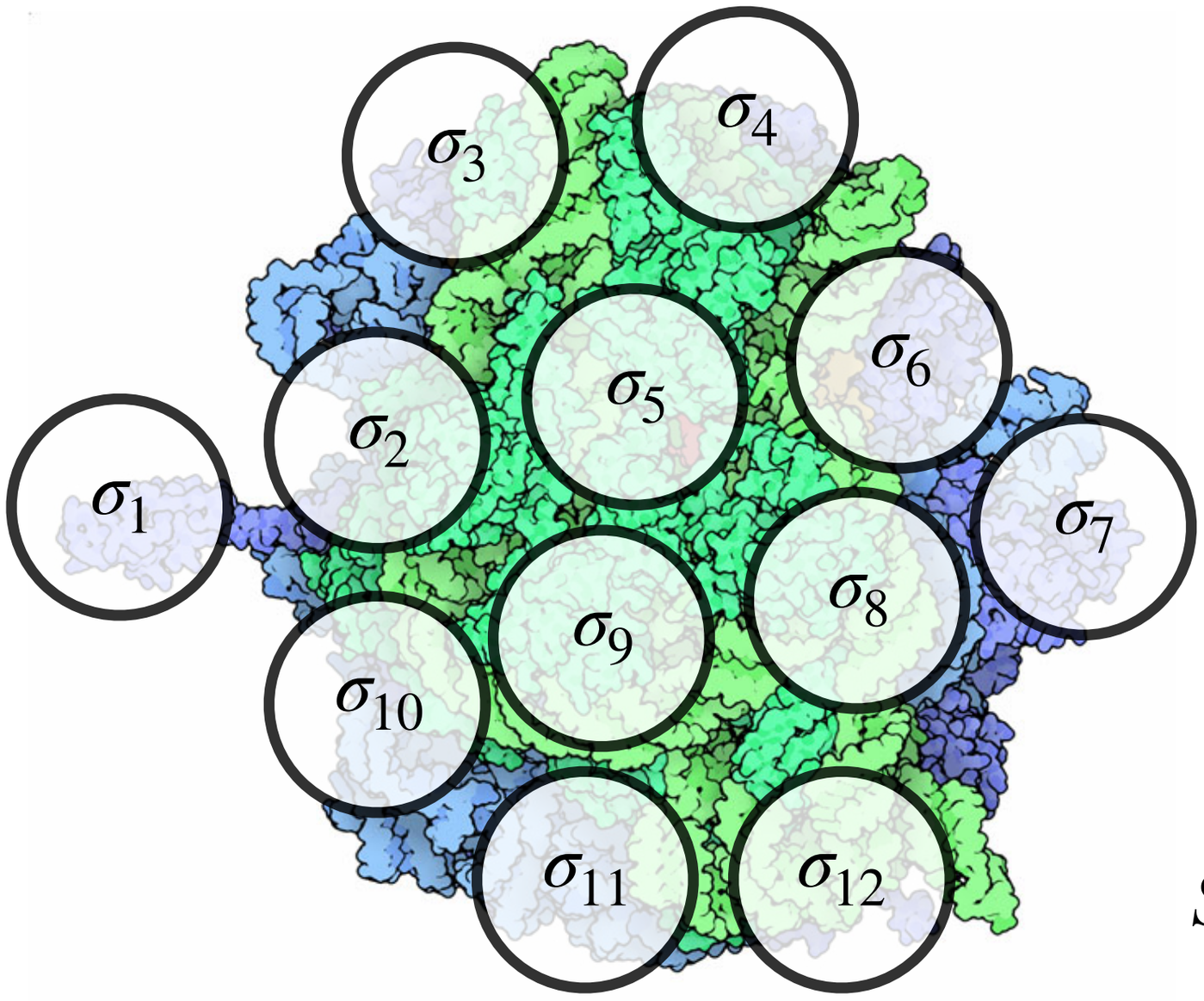
Large molecular systems may adopt an intractable number of states



$$S = \{\sigma_1, \ldots, \sigma_{12}\}$$

$$\sigma_a = \{0,1\} \qquad \#S = 2^{12} = 4096$$

Large molecular systems may adopt an intractable number of states

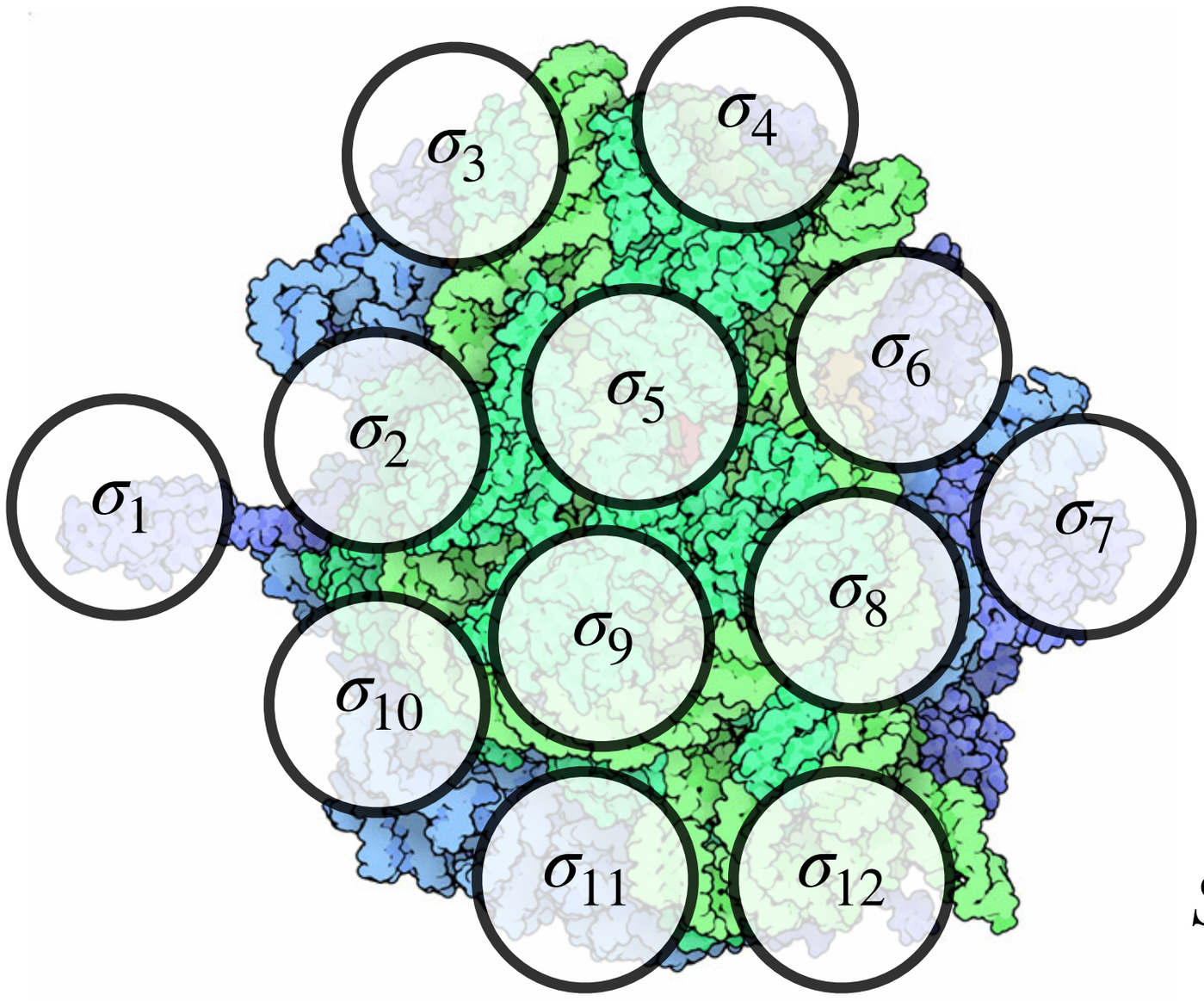


$$S = \{\sigma_1, \ldots, \sigma_{12}\}$$

$\sigma_a = \{0,1\}$ $\quad \#S = 2^{12} = 4096$

$\sigma_a = \{0,1,2\}$ $\quad \#S = 3^{12} = 531441$

Large molecular systems may adopt an intractable number of states



$$S = \{\sigma_1, \ldots, \sigma_{12}\}$$

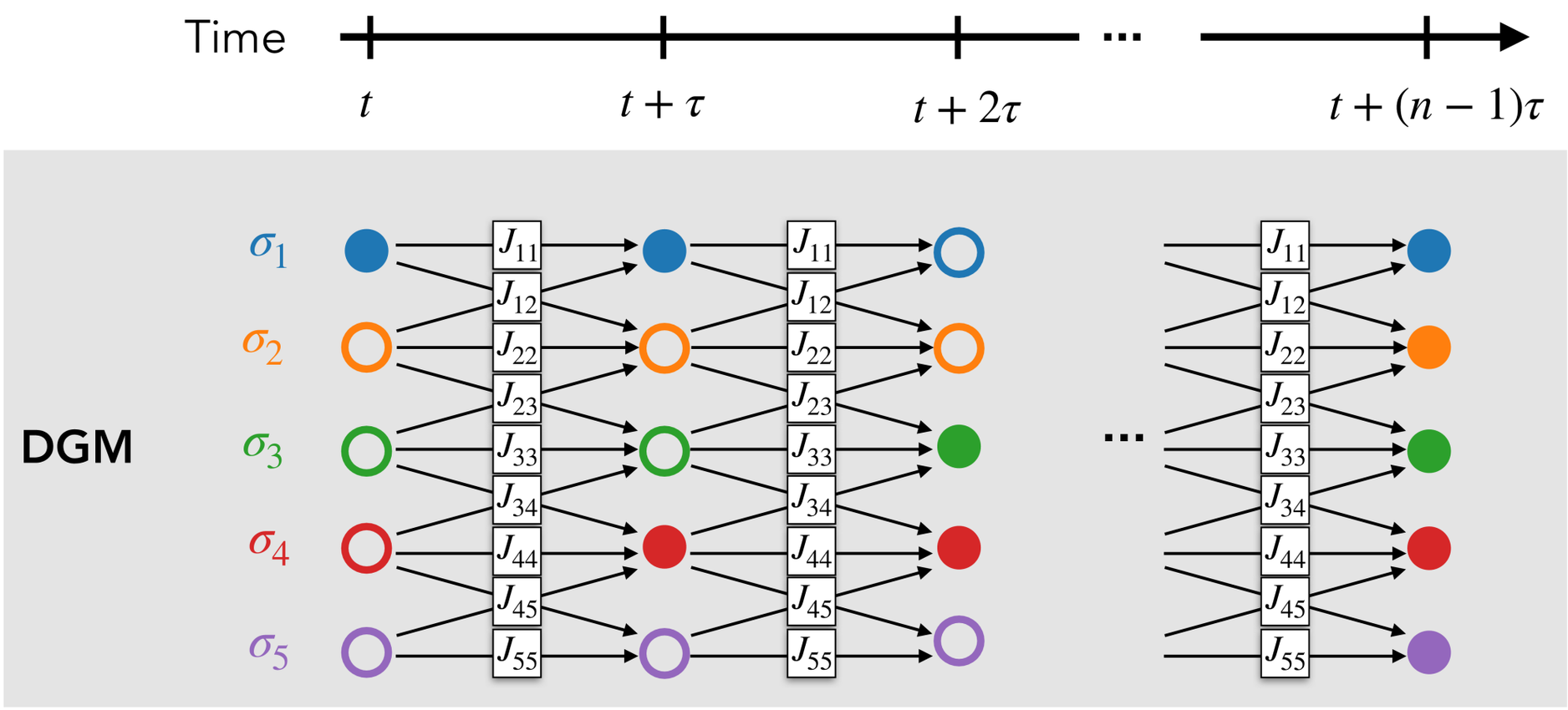
$\sigma_a = \{0,1\}$ $\quad \#S = 2^{12} = 4096$

$\sigma_a = \{0,1,2\}$ $\quad \#S = 3^{12} = 531441$

$\sigma_a = \{0,1,2,3\}$ $\quad \#S = 4^{12} = 16777216$

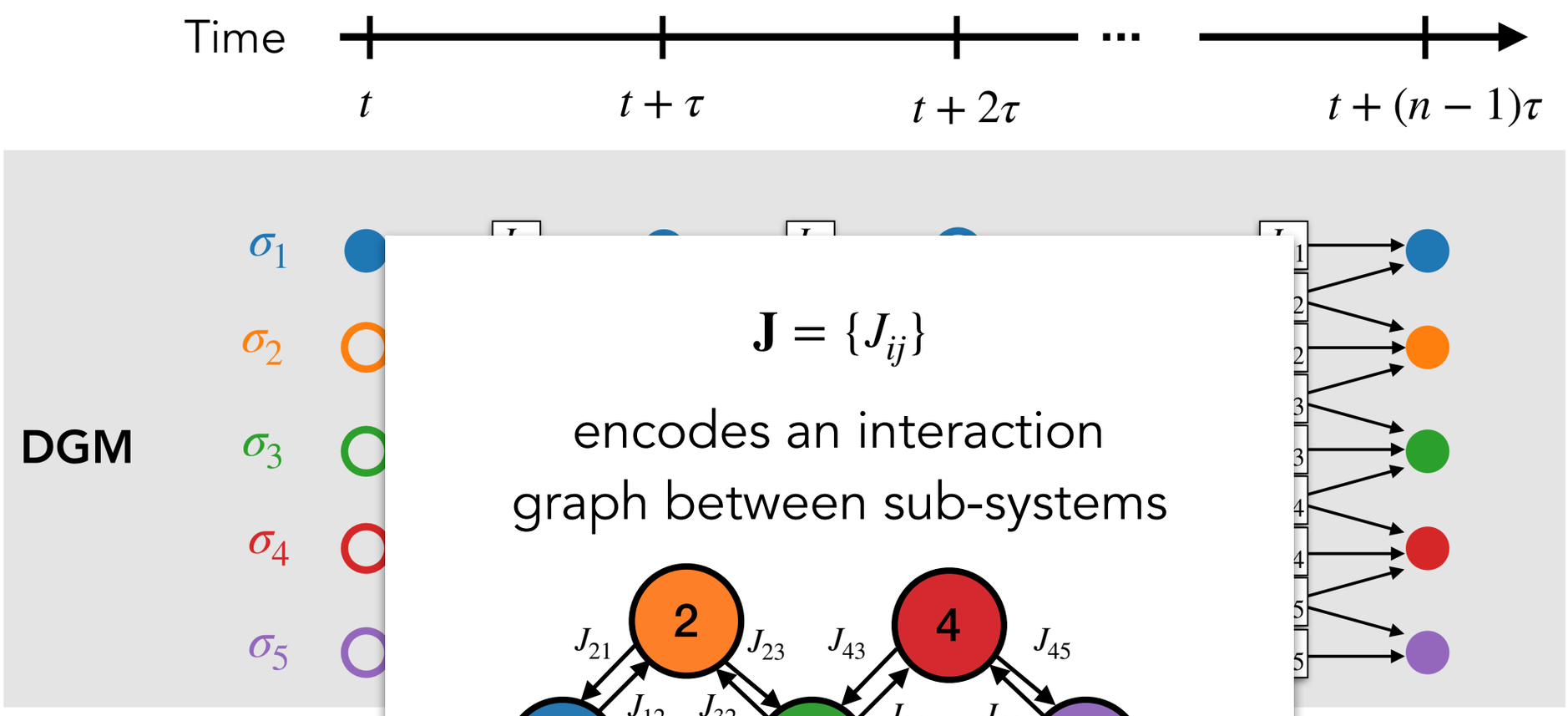
$\sigma_a = \{0,\ldots,N\}$ $\quad \#S = N^{12}$

Dynamic Graphical models: Going local to scale globally



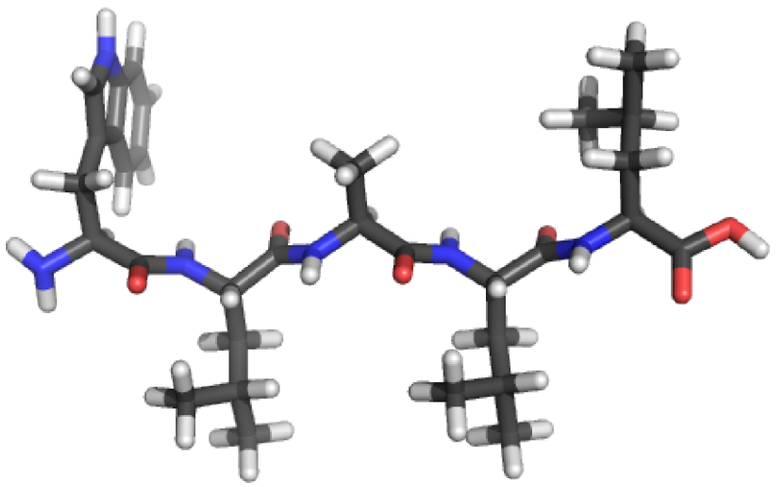
Olsson & Noé (bioRxiv:467050)

Dynamic Graphical models: Going local to scale globally



Olsson & Noé (bioRxiv:467050)

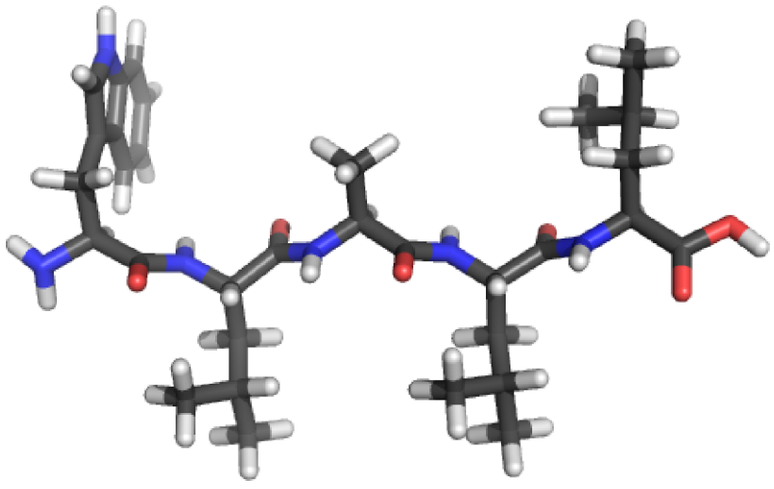
Consistent predictions for small molecular systems



DGM

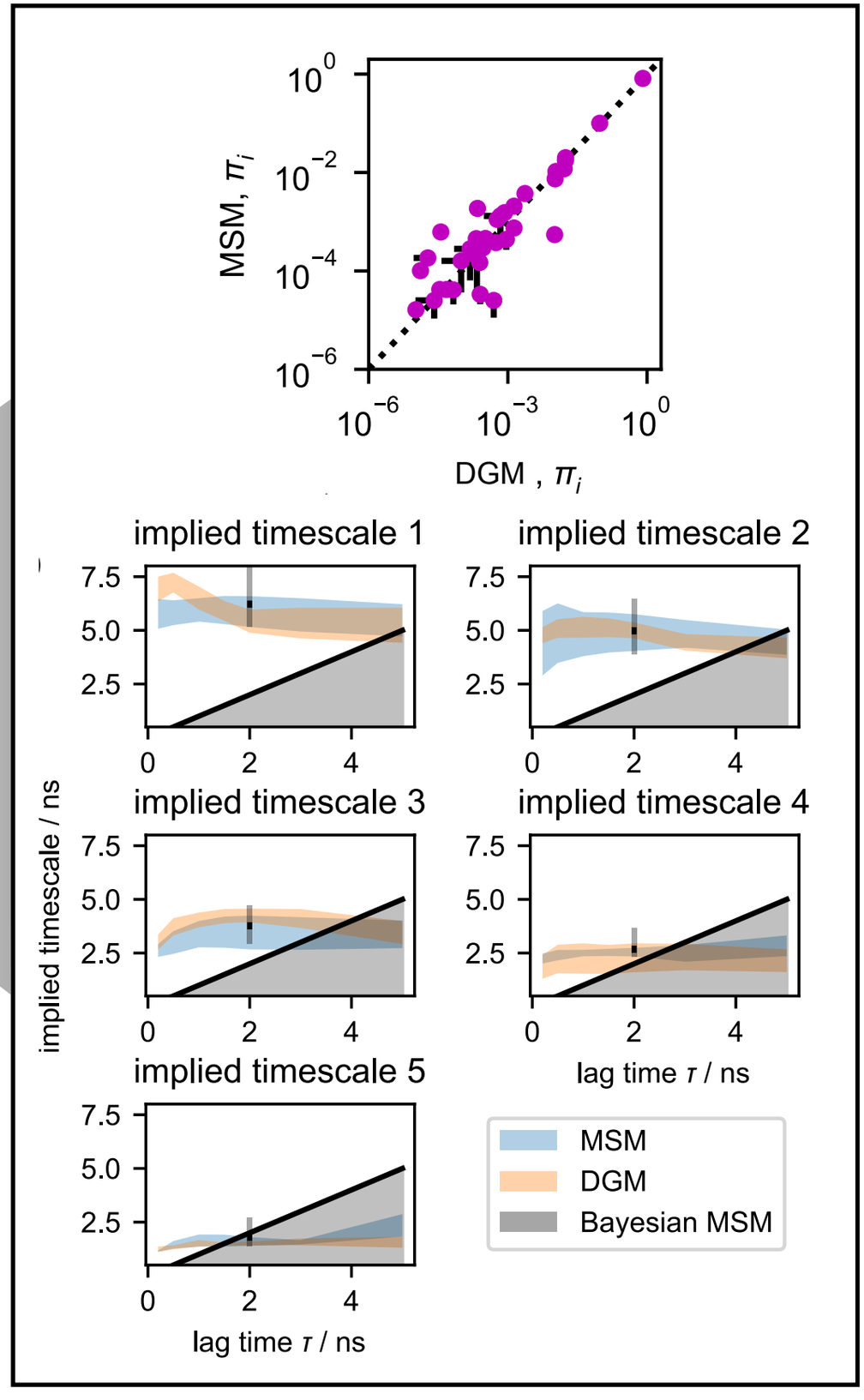
MSM

Consistent predictions for small molecular systems



DGM

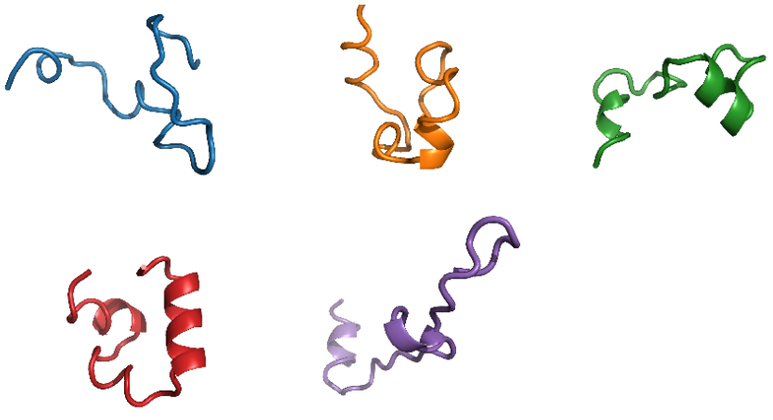
MSM



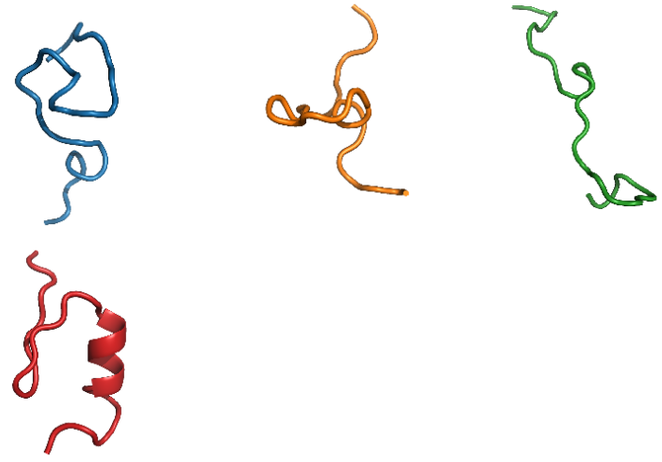
Predicting beyond the observed domain

We consider two fast folding proteins which was previously studied by Lindorff-Larsen et al.

Villin — 5 Meta-stable states



BBA — 4 Meta-stable states

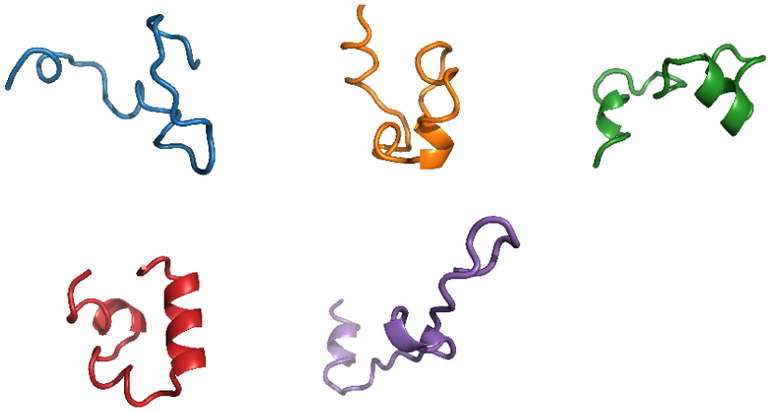


Olsson & Noé (bioRxiv:467050)

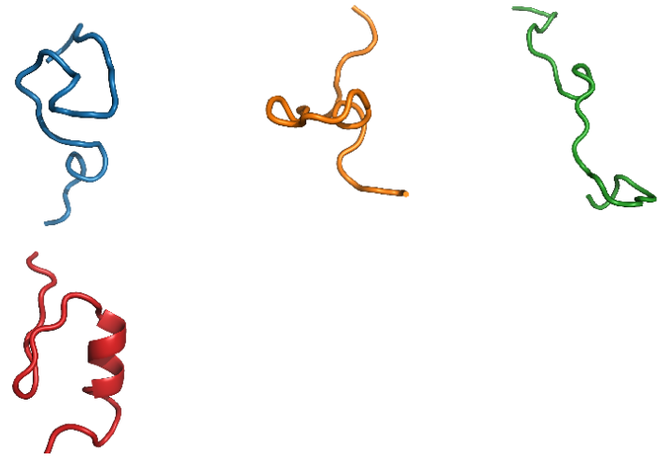
Predicting beyond the observed domain

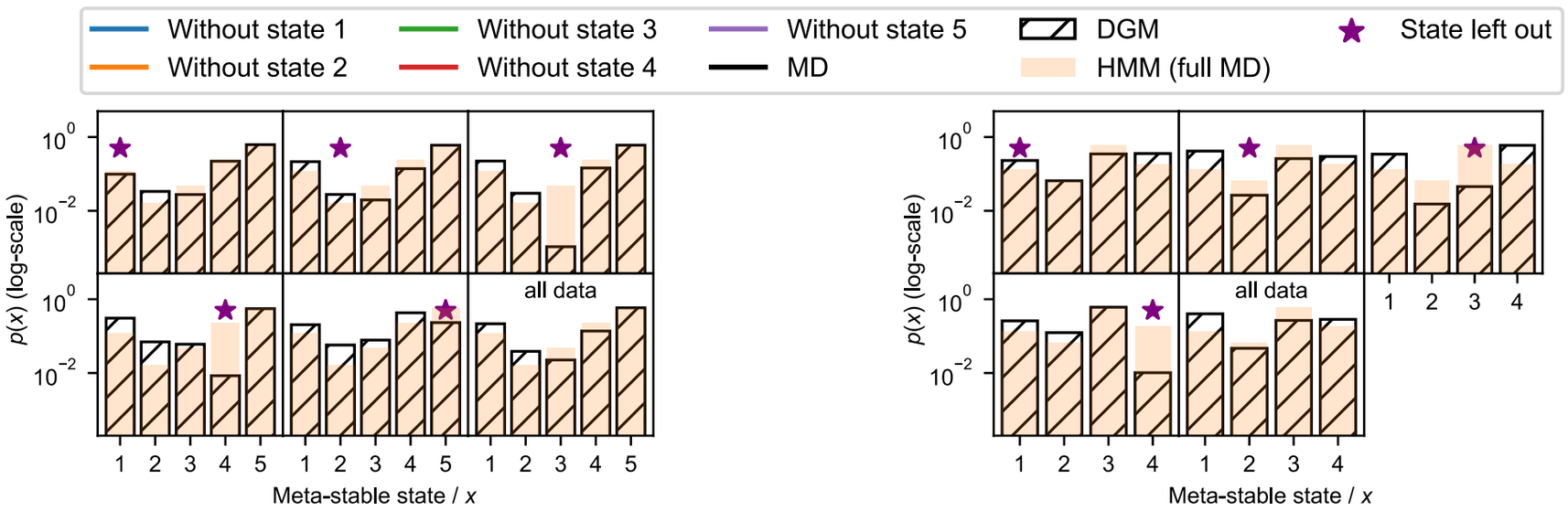
We consider two fast folding proteins which was previously studied by Lindorff-Larsen et al.

Villin — 5 Meta-stable states



BBA — 4 Meta-stable states



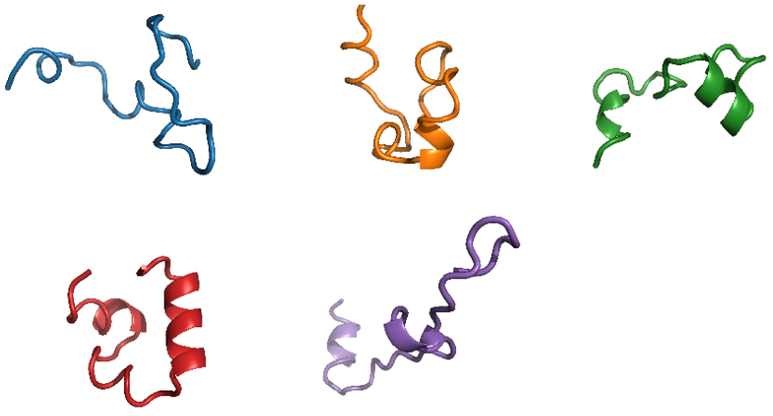


Olsson & Noé (bioRxiv:467050)

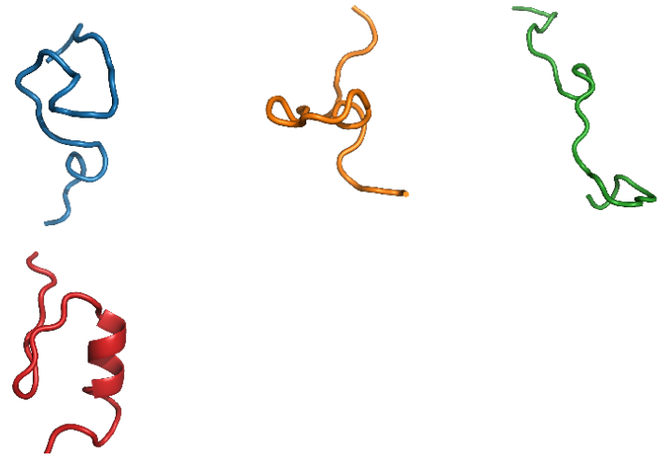
Predicting beyond the observed domain

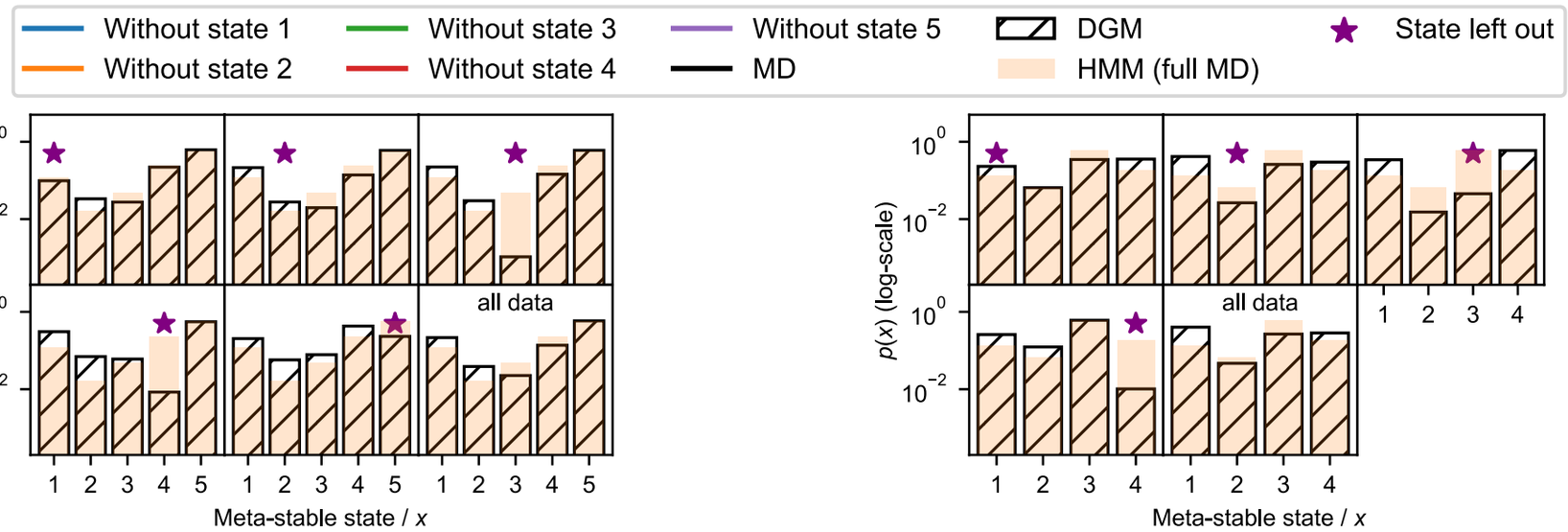
We consider two fast folding proteins which was previously studied by Lindorff-Larsen et al.

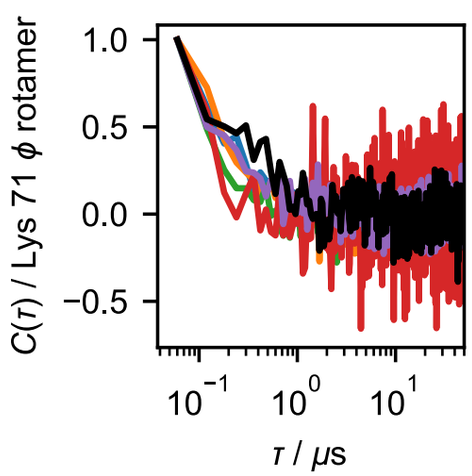
Villin — 5 Meta-stable states

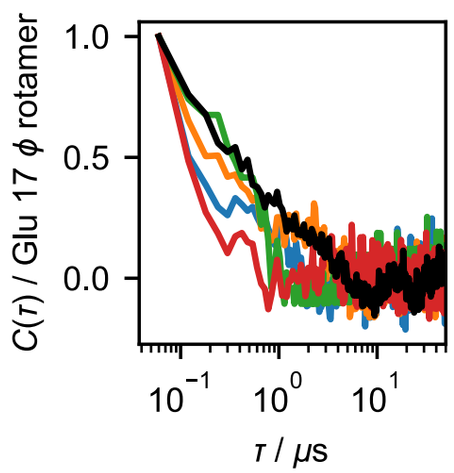


BBA — 4 Meta-stable states









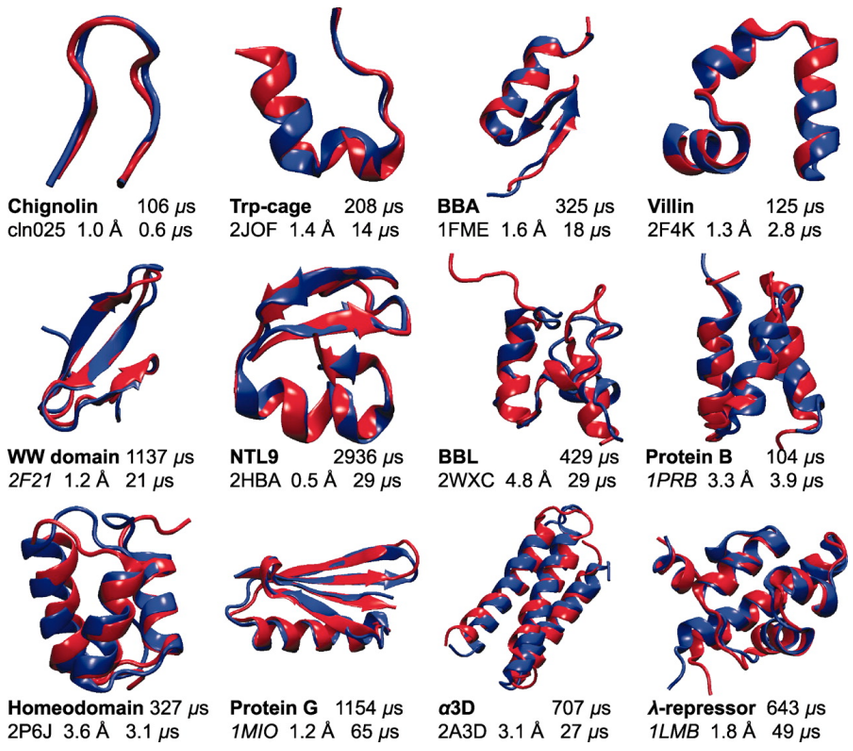
Olsson & Noé (bioRxiv:467050)

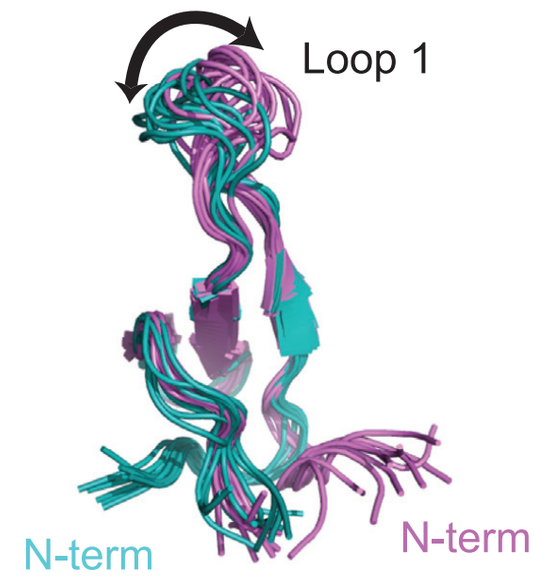
Machine Learning of Coarse-Grained Molecular Dynamics Force Fields

Wang L, Olsson S, Wehmeyer C, Perez A, Charron N, De Fabritiis G, Noé F, Clementi C. ACS Cent Sci (2019).

Simulating large molecular systems is challenging

- Atomistic molecular simulations have seen many successes including
  - Folding of small protein domains
  - Identifying transient structural states in proteins





- However, computational effort grows rapidly with the number of atoms:
  - Shorter simulations => less confident predictions about slow time-scale events

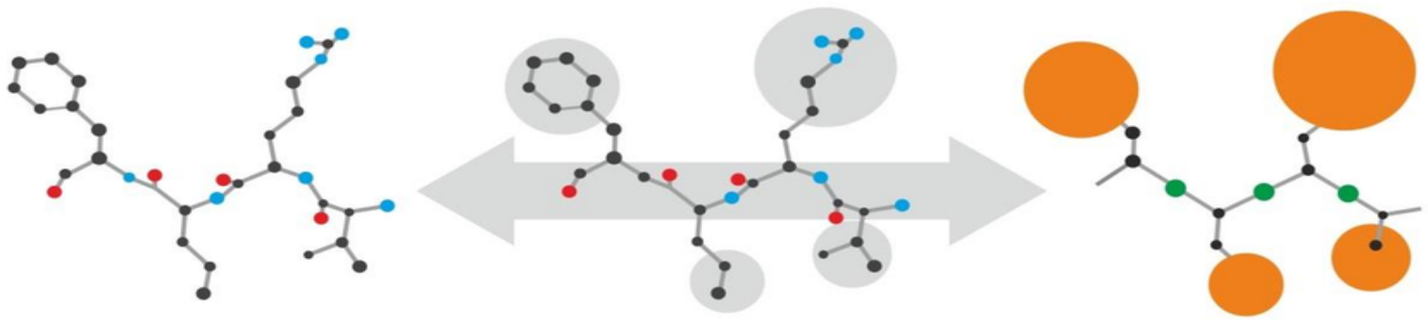
<u>Non-atomistic (coarse-grained) models may help us close the gap!</u>

Lindorff-Larsen et al. *Science* 2011 Olsson et al *Structure* 2016

Coarse-graining

Fine grained *Embedding function* *Coarse grained*

M



Representation $\mathbf{r}$ $\mathbf{R}$

(free) energy model $u(\mathbf{r})$ $U(\mathbf{R})$

Figure courtesy of A Kolinski

Coarse-graining with thermodynamic consistency

- We want to build a coarse-grained free energy model which matches the fine-grained model as closely as possible.

$$p(\mathbf{r}) = \exp(-u(\mathbf{r})/k_b T)$$ Boltzmann weights

Coarse-graining with thermodynamic consistency

- We want to build a coarse-grained free energy model which matches the fine-grained model as closely as possible.

$$p(\mathbf{r}) = \exp(-u(\mathbf{r})/k_b T)$$

Boltzmann weights

$$p(\mathbf{R}) = \frac{\int p(\mathbf{r})\delta(\mathbf{R} - \xi(\mathbf{r}))\mathrm{d}\mathbf{r}}{\int p(\mathbf{r})\mathrm{d}\mathbf{r}}$$

Coarse-grained distribution

uniquely defined, albeit intractable integral! — origin of pesky multi-body terms

Coarse-graining with thermodynamic consistency

- We want to build a coarse-grained free energy model which matches the fine-grained model as closely as possible.

$$p(\mathbf{r}) = \exp(-u(\mathbf{r})/k_b T)$$ Boltzmann weights

$$p(\mathbf{R}) = \frac{\int p(\mathbf{r})\delta(\mathbf{R} - \xi(\mathbf{r}))\mathrm{d}\mathbf{r}}{\int p(\mathbf{r})\mathrm{d}\mathbf{r}}$$ Coarse-grained distribution

$$U(\mathbf{R}) = -k_B T \log p(\mathbf{R}) + C$$ CG free energy potential

uniquely defined, albeit intractable integral! — origin of pesky multi-body terms

Can we approximate this Free energy potential with a neural network?

If so how?

Estimating CG Free energy potential model — a ML problem

- Matching instantaneous forces between CG and atomistic models

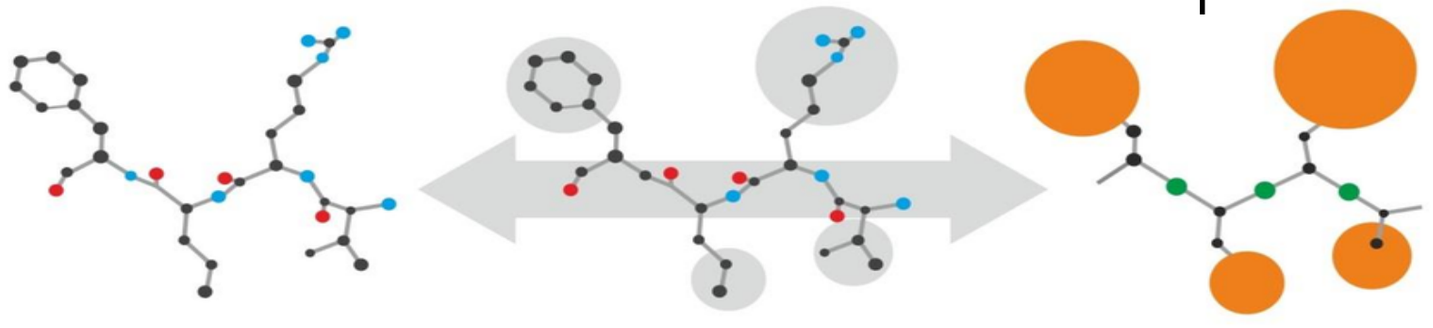
$$\chi^2(\boldsymbol{\theta}) = \langle \| \xi(\mathbf{F}(\mathbf{r})) + \nabla U_{\boldsymbol{\theta}}(\xi(\mathbf{r})) \|_2 \rangle$$

Atomistic forces projected onto CG particles

Predicted average CG particle force on mapped configuration

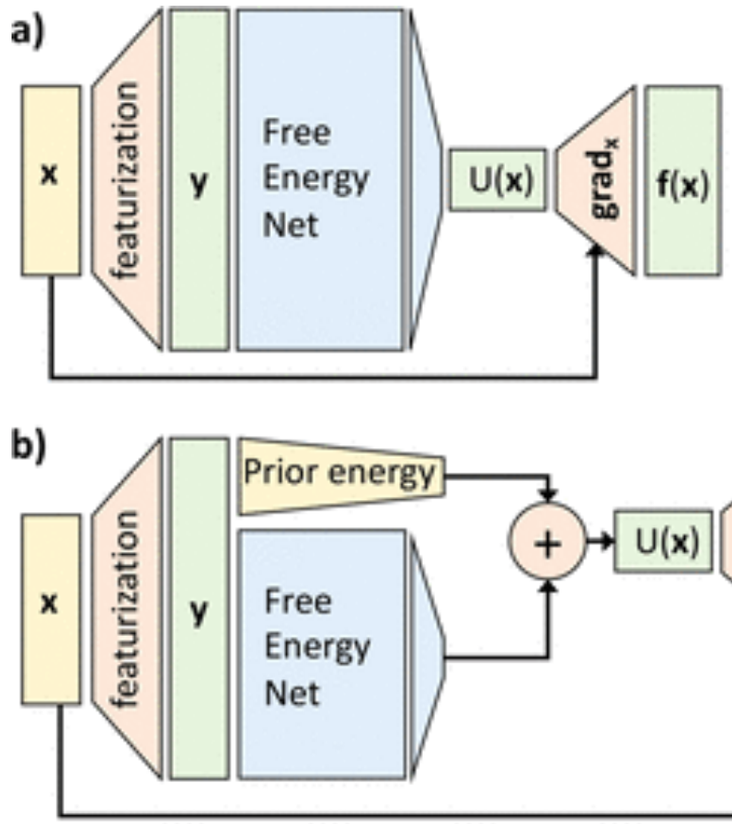
$$\xi(\mathbf{F}(\mathbf{r})) = \xi(-\nabla u(\mathbf{r})) \quad \overset{\textit{Compare}}{\longleftrightarrow} \quad -\nabla U_{\boldsymbol{\theta}}(\xi(\mathbf{r}))$$

$u(\mathbf{r})$ $\quad \xi(\mathbf{r}) \quad$ $U_{\boldsymbol{\theta}}(\mathbf{R})$



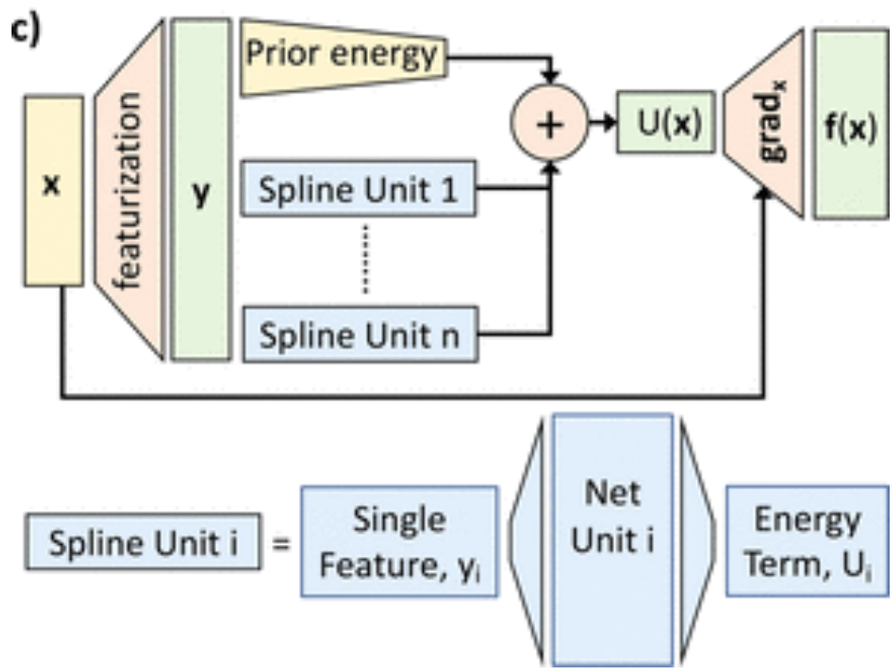
CGNet: A Deep neural network approach

- We train several models to approximate $U_{\theta}(\mathbf{R})$



CG Free energy Net

Regularized CG Free energy Net

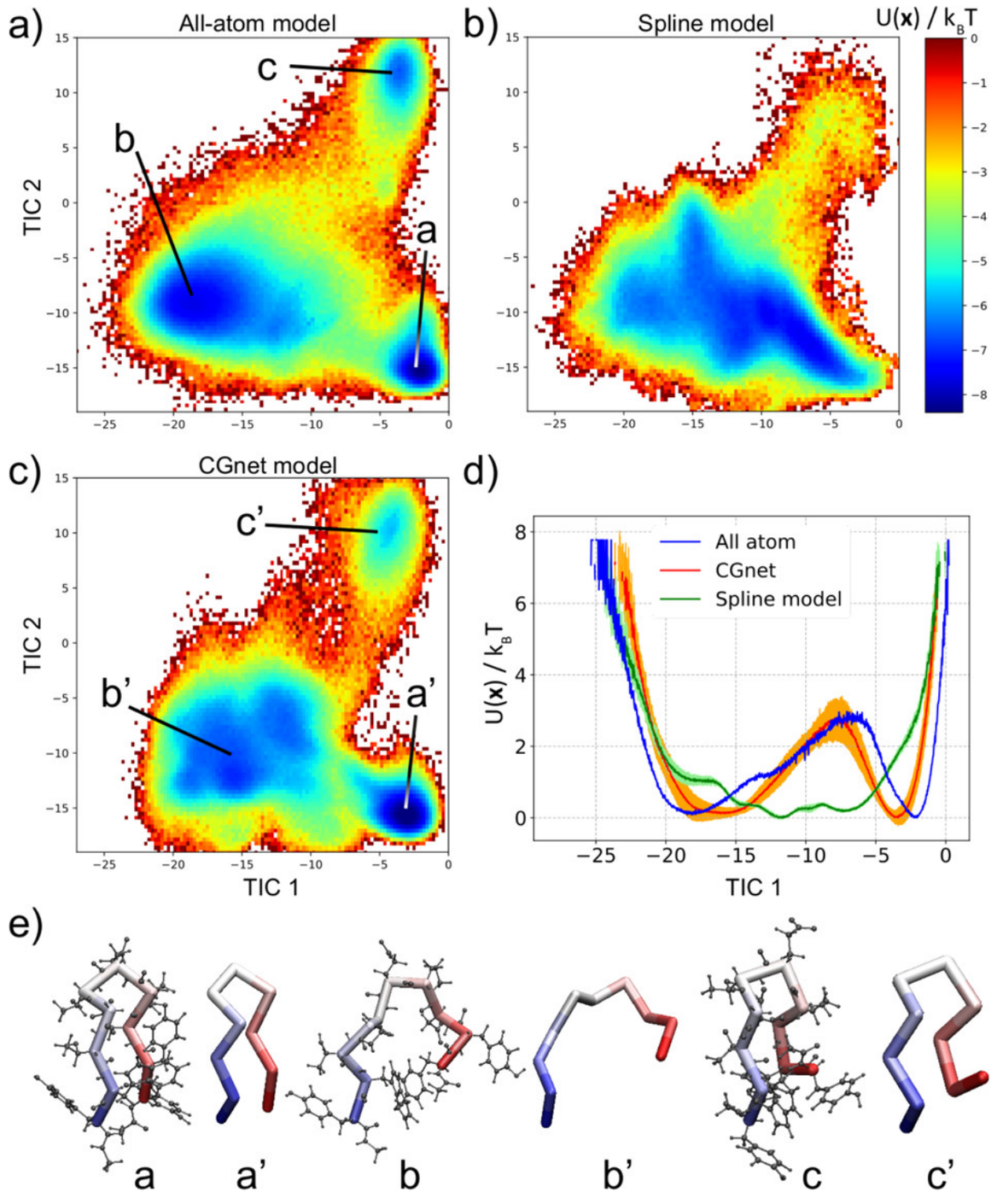


Reference mode

Wang, Olsson et al . ACS Cent. Sci 2019

CGNet: A Deep neural network approach - application to Chignolin

- Fast-folding micro protein
- All-atom simulation
  - Forces and coordinates stored
- Regularized CGNet captures three free energy minima well. Samples similar structures
- Reference model fails to capture free energy minima.



Wang, Olsson et al . ACS Cent. Sci 2019

CGNet: A Deep neural network approach - outstanding problems

- Transferability
  - CGnet currently only predicts properties of molecule for which training data is available.
    - Use parameter-sharing
    - Generate data-set of multiple molecular systems for training.

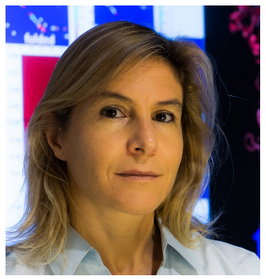
- Computational overhead associated with current CGnet architecture may become substantial.
  - Alternative network architectures <=> parameter sharing.

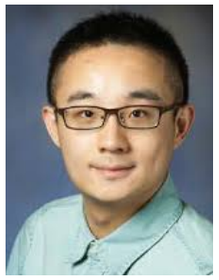
Wang, Olsson et al . ACS Cent. Sci 2019

Acknowledgements

FU Berlin

Rice University





Universitat Pompeu Fabra

Save the date:

MolKin2019

Molecular Kinetics

Sampling, Design and Machine Learning

June 19-21, 2019, Berlin Germany

Only a few spots left!

DAHLEM RESEARCH SCHOOL | Center for Junior Researchers

DANMARKS FRIE FORSKNINGSFOND

Unterstützt von / Supported by

Alexander von Humboldt Stiftung/Foundation