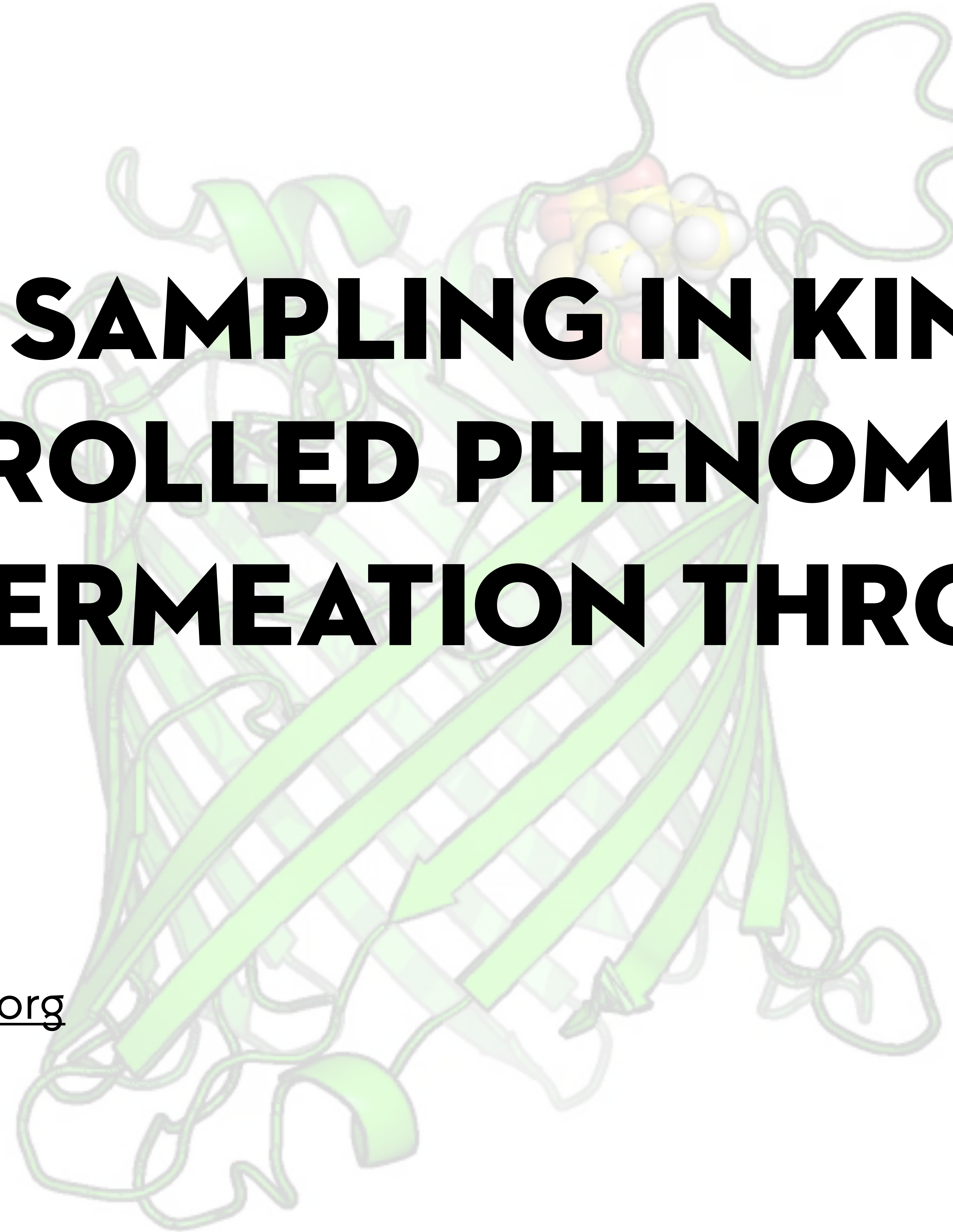
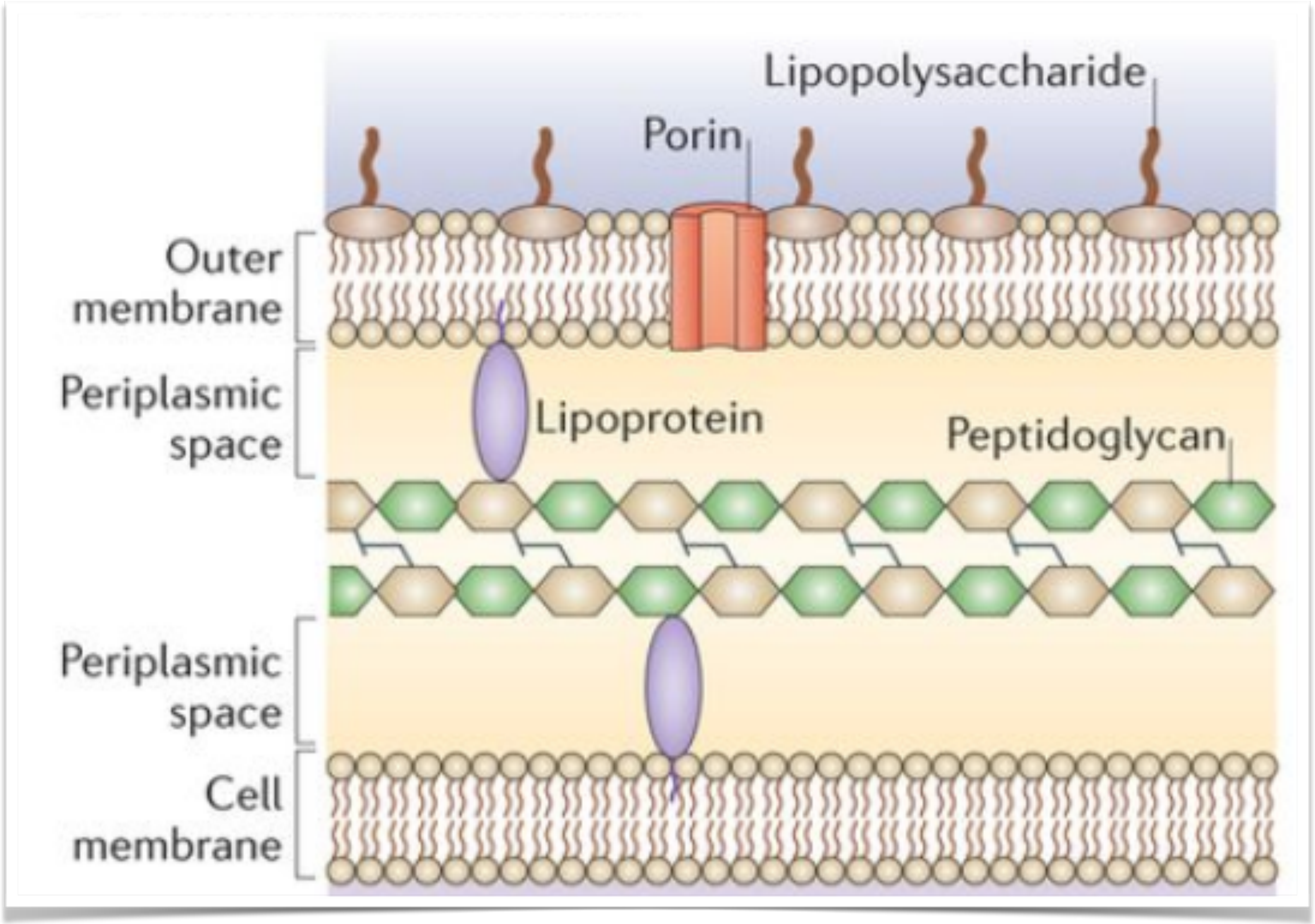


ENHANCED SAMPLING IN KINETICALLY CONTROLLED PHENOMENA: THE CASE OF PERMEATION THROUGH PORINS

Ana Silveira
Sloan Kettering Institute
http://www.choderalab.org

6 Dec 2018

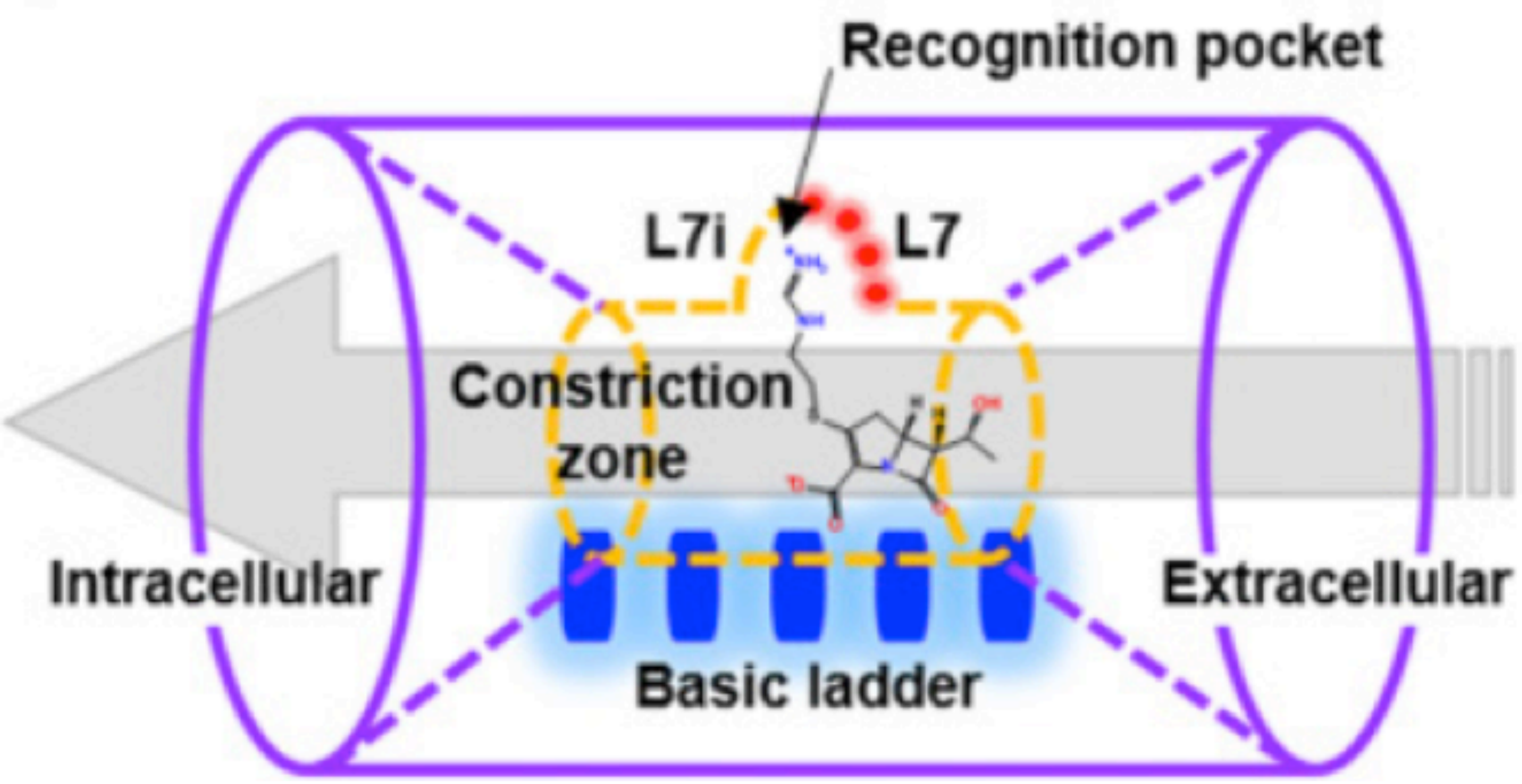
The outer membrane is the first line of defense for Gram-negative bacteria against toxic compounds



Brown et al. *<u>Nature Reviews Microbiology</u>* *(2015)*

Porins

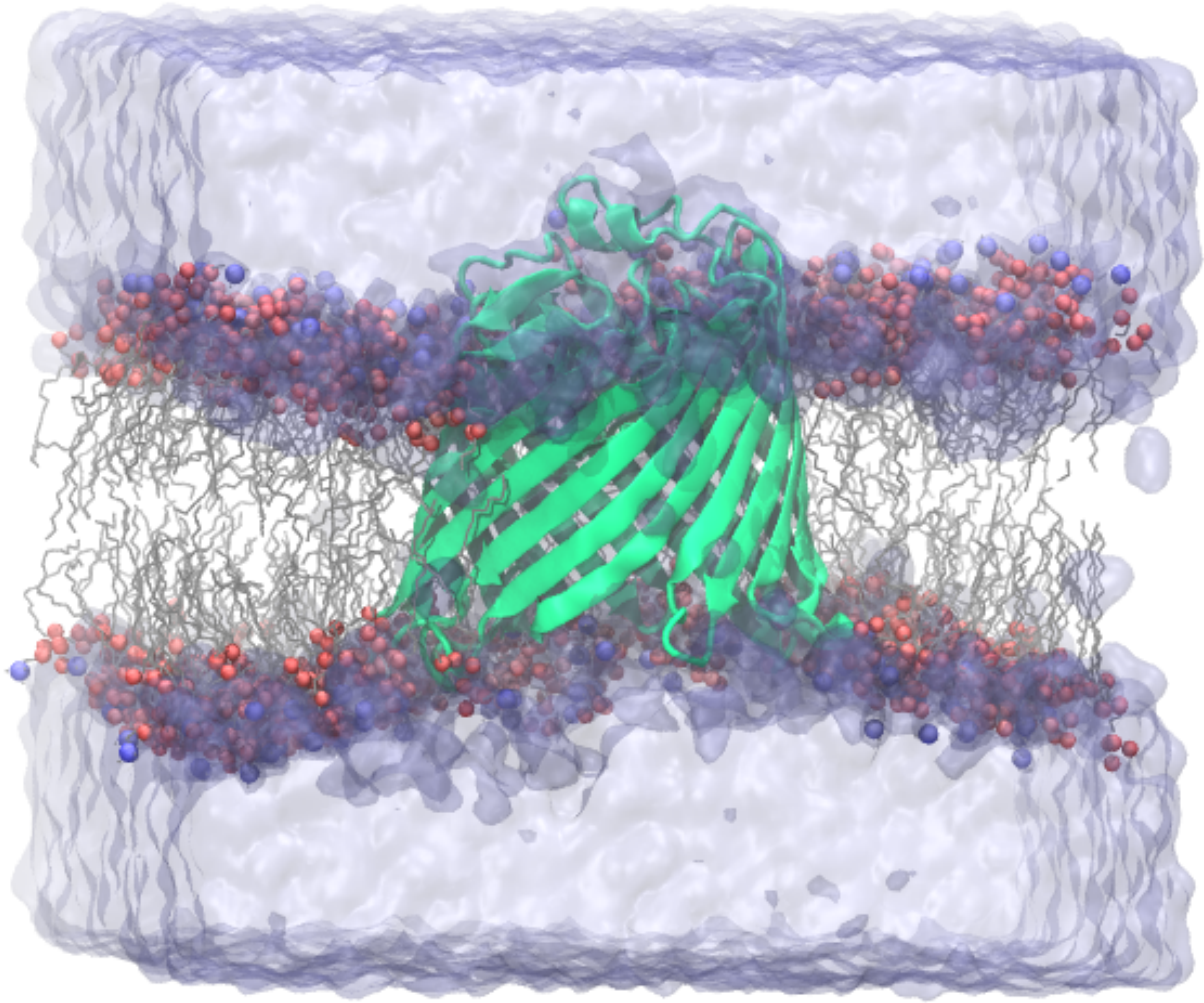
channels where nutrients, ions, substrates, and water pass through from the extracellular space



also antibiotics!

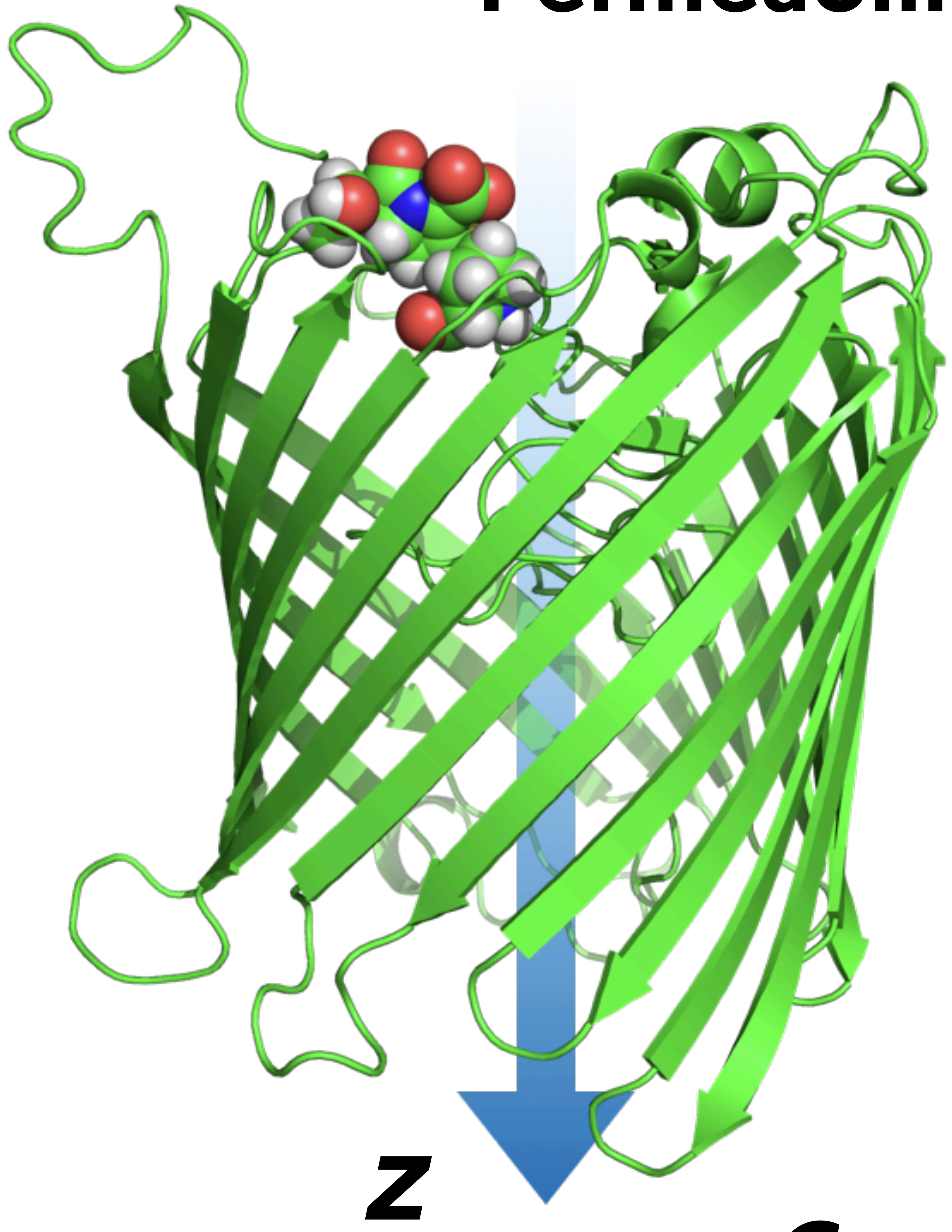
Isabella *et al*. Chem.Biol. (2015)

OprD porin embedded in a lipid bilayer retrieved from MemProtMD data base



http://memprotmd.bioch.ox.ac.uk

Permeability coefficient (P_m) from Molecular Simulation



Inhomogeneous solubility-diffusion model

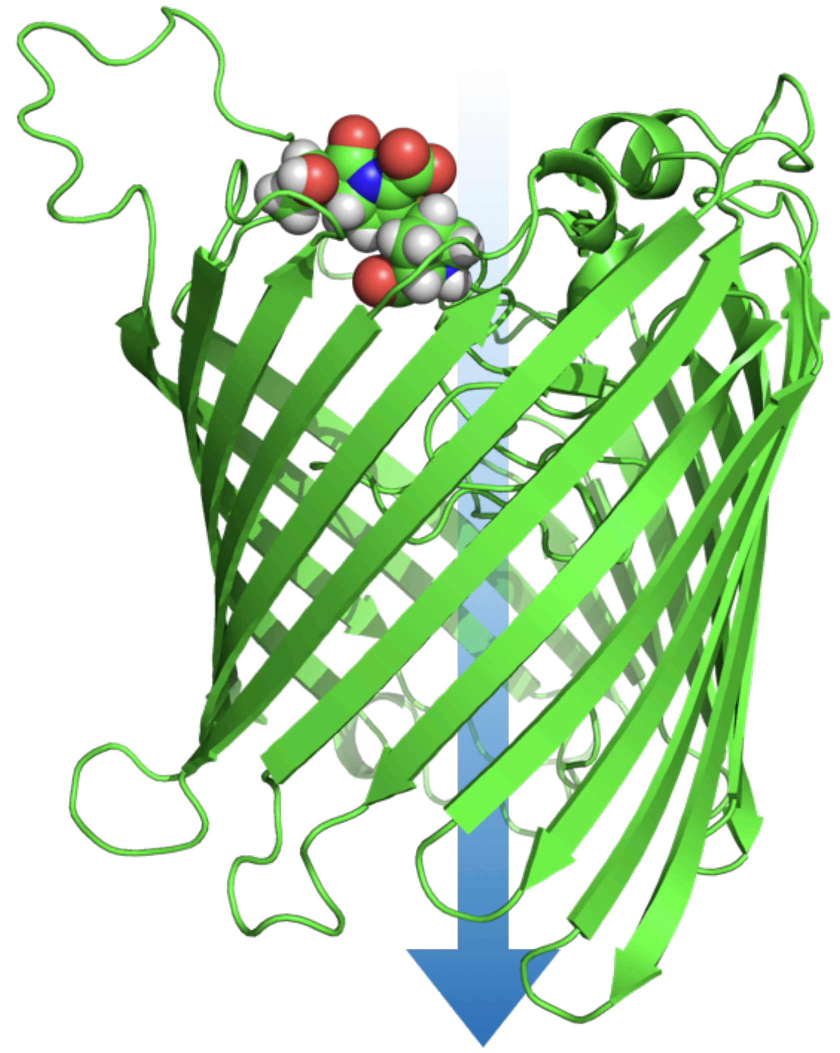
$$\frac{1}{P_m} = \int_{z_{int}}^{z_{ext}} \frac{\exp[w(z)/RT]}{D(z)} dz$$

$w(z)$: Potential of mean force

$D(z)$: Diffusion coefficient

Can the potential of mean force reveal the rate/mechanism of transport?

Potential of mean force



Only possible with Biased Simulations

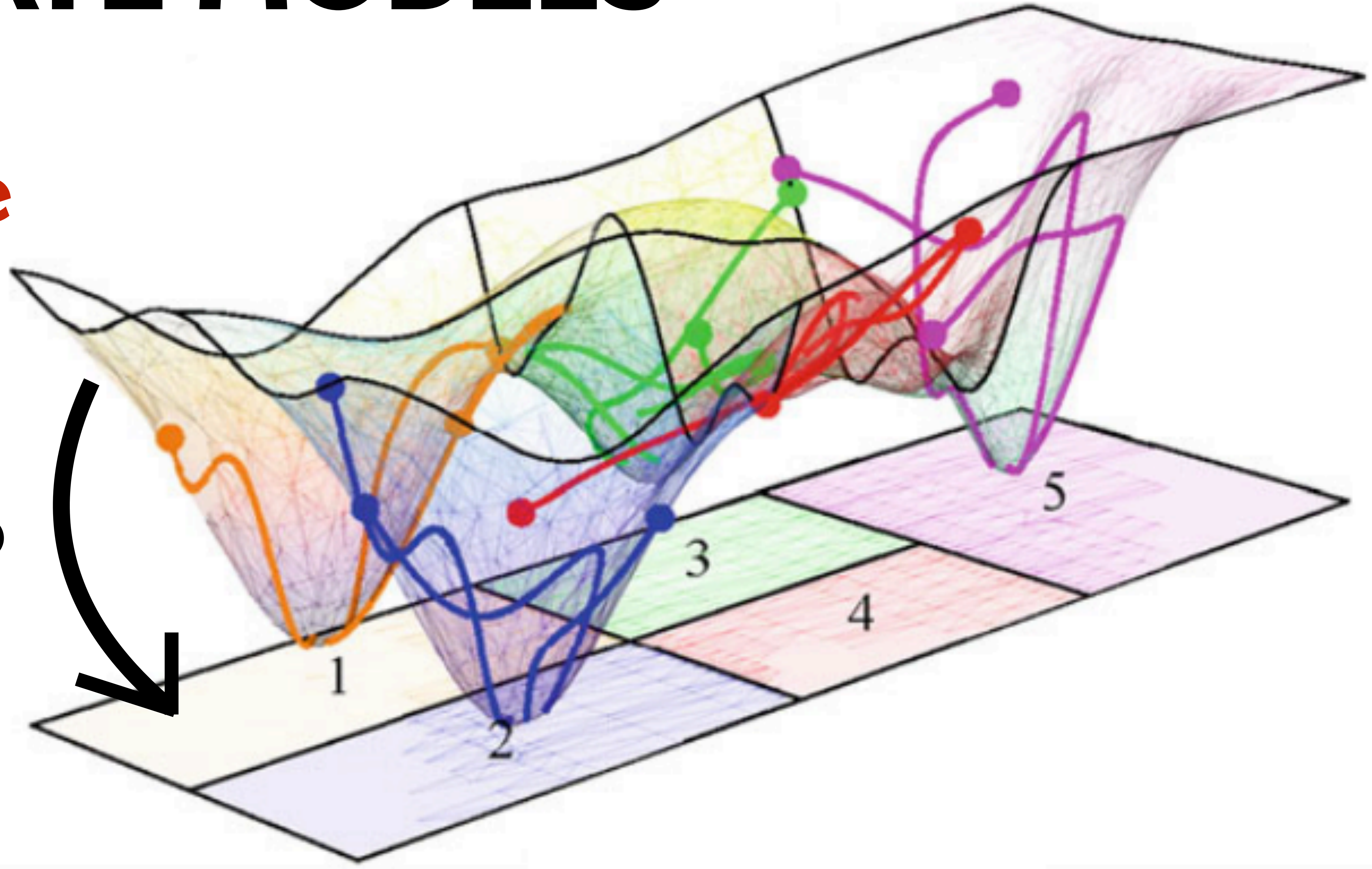
- Umbrella sampling, simulated tempering, parallel tempering....

Degrees of freedom orthogonal to the collective variable can be highly autocorrelated

- **Markov State Models? (include experimental data)**

MARKOV STATE MODELS

The problem of sampling **metastable** ("long-lived") states



map

MSM
Transition Matrix
Spectral Decomposition

↓

Implied time scales $t_i = \frac{-\tau}{\ln|\lambda_i(\tau)|}$

Da et al. Adv Exp Med Biol. 2014

Dynamics can be cast as the action of the Transfer Operator

$$\rho_\tau = T_r(\tau)\,\rho$$

Metastability

Eigenvalues $\sigma(T_\tau) \subset [0,r] \cup \{\lambda_M, \ldots, \lambda_2, 1\}$

Eigenfunctions weighted by the stationary distribution $\pi(x)$ $\phi_m(x) = \pi(x)\,\psi_m(x)$

Finite basis $f_m(x) = \sum_{k=1}^{n} b_{mk}\chi_k(x)$

Which is the "optimal" numerical approximation for ϕ, λ ?

$$\sum_{m=1}^{M} \lambda_m = \sup_{\mathbf{b}} \sum_{m=1}^{M} \langle T_\tau f_m, f_m \rangle_\mu$$

Rayleigh trace

The optimization problem reduces to an eigenvalue problem

Finite basis

$$f_m(x) = \sum_{k=1}^{n} b_{mk}\chi_k(x)$$

$$C^\tau b_m = \hat{\lambda}_m C^0 b_m$$

$$C^\tau(i,j) = \langle T_\tau f_i, f_j \rangle_\mu$$

Time-lagged

$$C^0(i,j) = \langle f_i, f_j \rangle_\mu$$

Instantaneous

C^τ, C^0 **Correlation matrix of basis functions**

Special (important) cases:

1. **State space partition** $\{S_1, \ldots, S_n\}$

$$f_i = \chi S_i$$

$$T^\tau b_m = \hat{\lambda}_m b_m$$

MARKOV STATE MODEL

$$T^\tau(i,j) = p^\tau(S_i, S_j)$$

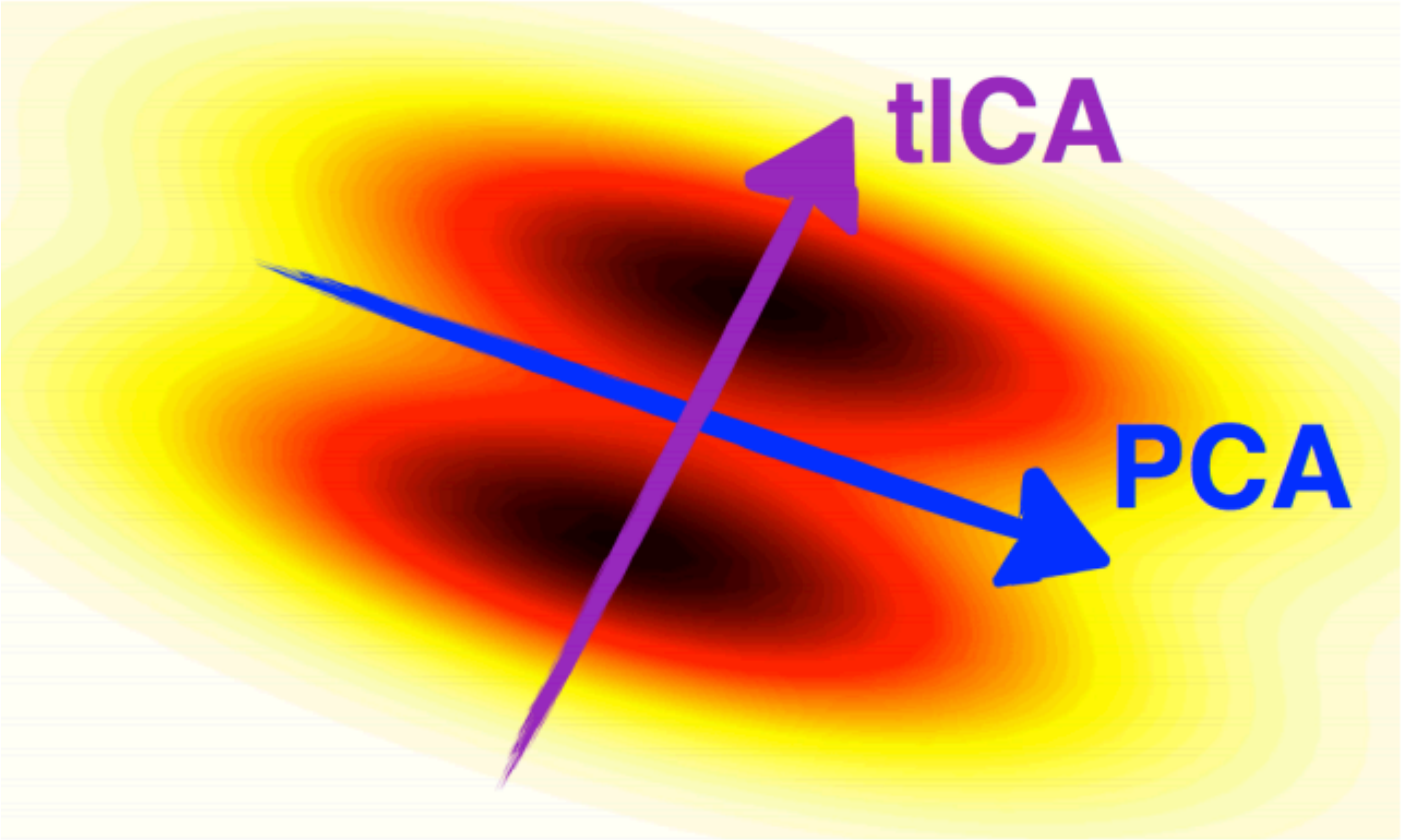
Transition Matrix

2. Mean-free order parameters $f_i = x_i - \langle x_i \rangle_\mu$

$$\langle f_2(x)_t \, f_2(x)_{t+\tau} \rangle_t \le e^{-\frac{\tau}{t_2}}$$

$$t_2^{\ddagger} \le t_2 \quad \left(t_i^{\ddagger} \le t_i \right)$$

Time-lagged independent component analysis (TICA)



TICA finds the most autocorrelated directions versus PCA yields the directions of most variance

LARGE-AMPLITUDE MOTIONS ARE NOT NECESSARILY ASSOCIATED WITH SLOW TRANSITIONS

MSM builder

VAMP (Variational approach for Markov Processes) allows to define "scores" for model selection

Approximation of Koopman operator

Wu and Noé (2017) https://arxiv.org/abs/1707.04659

VAMP-r Scores

Koopman operator

$$\mathcal{K}_\tau g(\mathbf{x}) = \mathbb{E}[g(\mathbf{x}_{t+\tau}) \,|\, \mathbf{x}_t = \mathbf{x}]$$

Linear transformation

$$\mathbb{E}[g(\mathbf{x}_{t+\tau})] = \mathbf{K}^T \mathbb{E}[f(\mathbf{x}_t)]$$

$$\mathbb{E}[\phi_i(\mathbf{x}_{t+\tau})] = \sigma_i \mathbb{E}[\psi_i(\mathbf{x}_t)]$$

$\mathbf{f} = (\psi_1, ..., \psi_m)^T$ **right-singular functions**

$\mathbf{g} = (\phi_1, ..., \phi_m)^T$ **left-singular functions**

$\mathbf{K} = diag(\lambda_1, ..., \lambda_m)$ **largest singular values of** $\mathcal{K}_\tau$

Wu and Noé (2017) https://arxiv.org/abs/1707.04659

The k dominant singular components of a Koopman operator are the solution of:

$$\sum_{m=1}^{M} \lambda_m^r = \sup_{\mathbf{f},\mathbf{g}} \sum_{m=1}^{M} \langle \mathcal{K}_\tau g_m, f_m \rangle_\mu^r$$

$$\mathcal{R}_r[\mathbf{f}, \mathbf{g}] = \sum_{m=1}^{M} \langle \mathcal{K}_\tau g_m, f_m \rangle_\mu^r$$

VAMP-r score

VAMP-1: maximizes the Rayleigh trace

VAMP-2: maximizes the kinetic variance $\sum_{m=2}^{M} = \lambda_m^2(\tau)$

Wu and Noé (2017) https://arxiv.org/abs/1707.04659

For known basis functions (χ_1, χ_0)

$$\mathbf{f} = \mathbf{U}^T \chi_0$$
$$\mathbf{g} = \mathbf{V}^T \chi_1$$

Find **U, V** →

$$\sup_{\mathbf{U},\mathbf{V}} \mathcal{R}_r[\mathbf{U}, \mathbf{V}]$$

$$\mathcal{R}_r(\mathbf{U}, V) = \sum_{i=1}^{k} \left(\mathbf{u}_i^T \mathbf{C}_{01} \mathbf{v}_r\right)^r$$

Algorithm

1. Compute covariance matrices
2. Perform the truncated SVD $\overline{\mathbf{K}} = \mathbf{C}_{00}^{-\frac{1}{2}} \mathbf{C}_{01} \mathbf{C}_{11}^{-\frac{1}{2}} \approx \mathbf{U}'\mathbf{S}\mathbf{V}'^T$
3. Compute $\mathbf{U} = \mathbf{C}_{00}^{-\frac{1}{2}} \mathbf{U}'$

   $\mathbf{V} = \mathbf{C}_{11}^{-\frac{1}{2}} \mathbf{V}'$

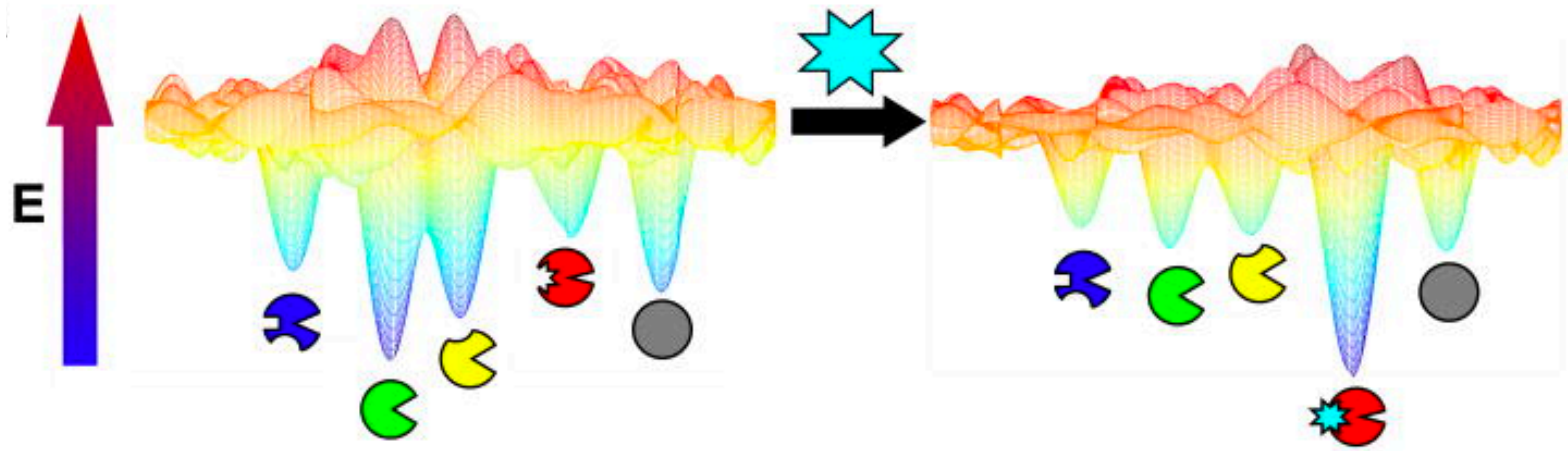
$$\mathcal{R}_r(\mathbf{U}, V) = \sum_{i=1}^{k} \left(\mathbf{u}_i^T \mathbf{C}_{01} \mathbf{v}_r\right)^r$$

VAMP-r score

The Dynamic Conformational Landscapes of the Protein Methyltransferase SETD8

- cancer invasiveness and metastasis
- pediatric leukemia

Covalent ligands can trap hidden conformations of SETD8 to allow visualization via x-ray crystallography

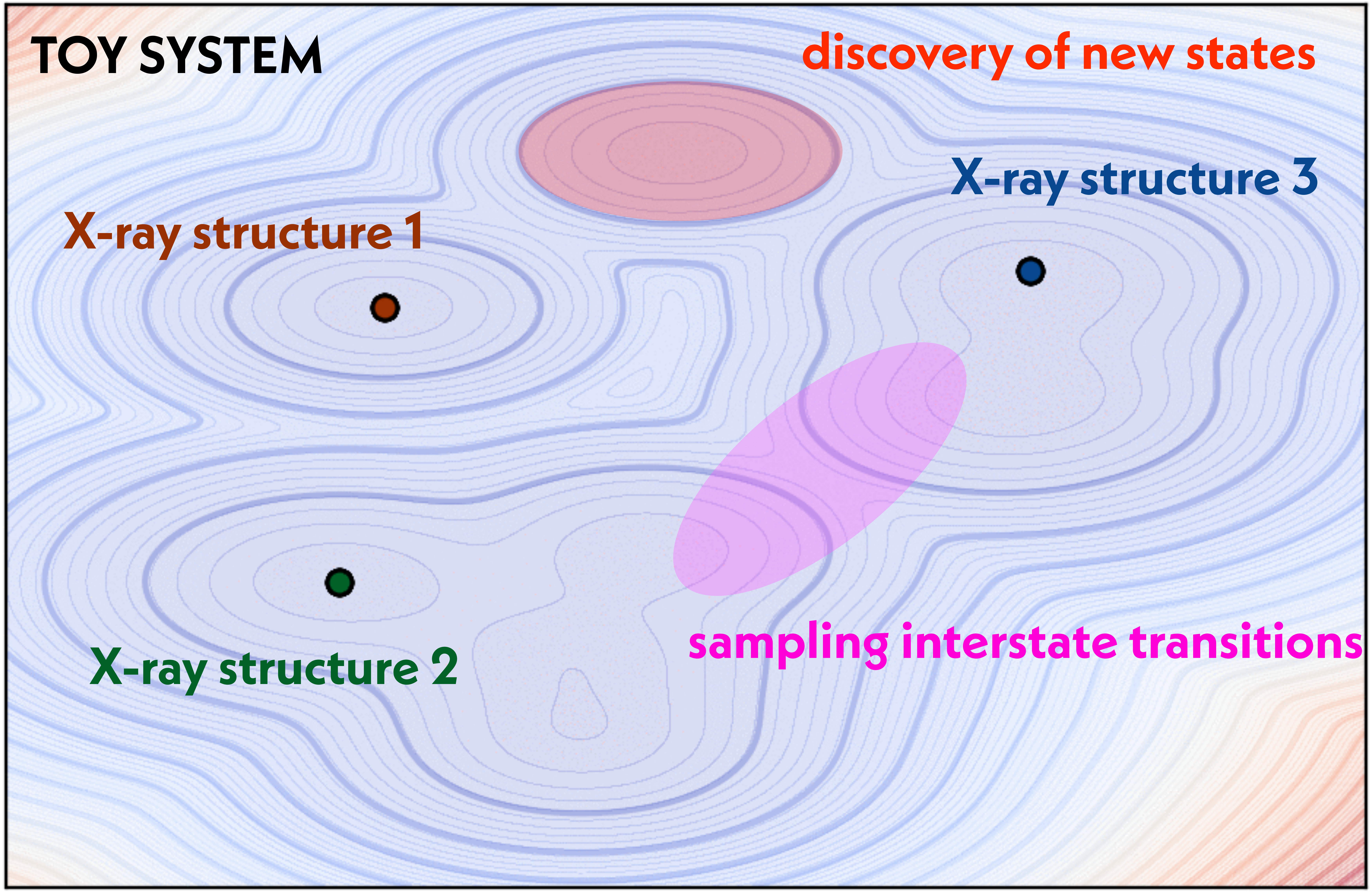


Lee and Craik. Science 324:213,

Rafal Wiewiora

bioRxiv (2018) https://doi.org/10.1101/438994

Simulations seeded from x-ray structures can identify hidden conformations and characterize functional dynamics



Methods

- **~ 5000 independent MD simulations (1 μs each) on Folding@home (~ 5 milliseconds)**

- **Observables**:

(a) residue-residue distances (heavy atoms) separated in sequence by at least two neighboring residues

(b) logistic distance

(c) backbone (phi, psi) and side chains (chi 1) dihedral angles

- **tICA : at lag times of either 5 or 50 ns, with either kinetic or commute mapping**

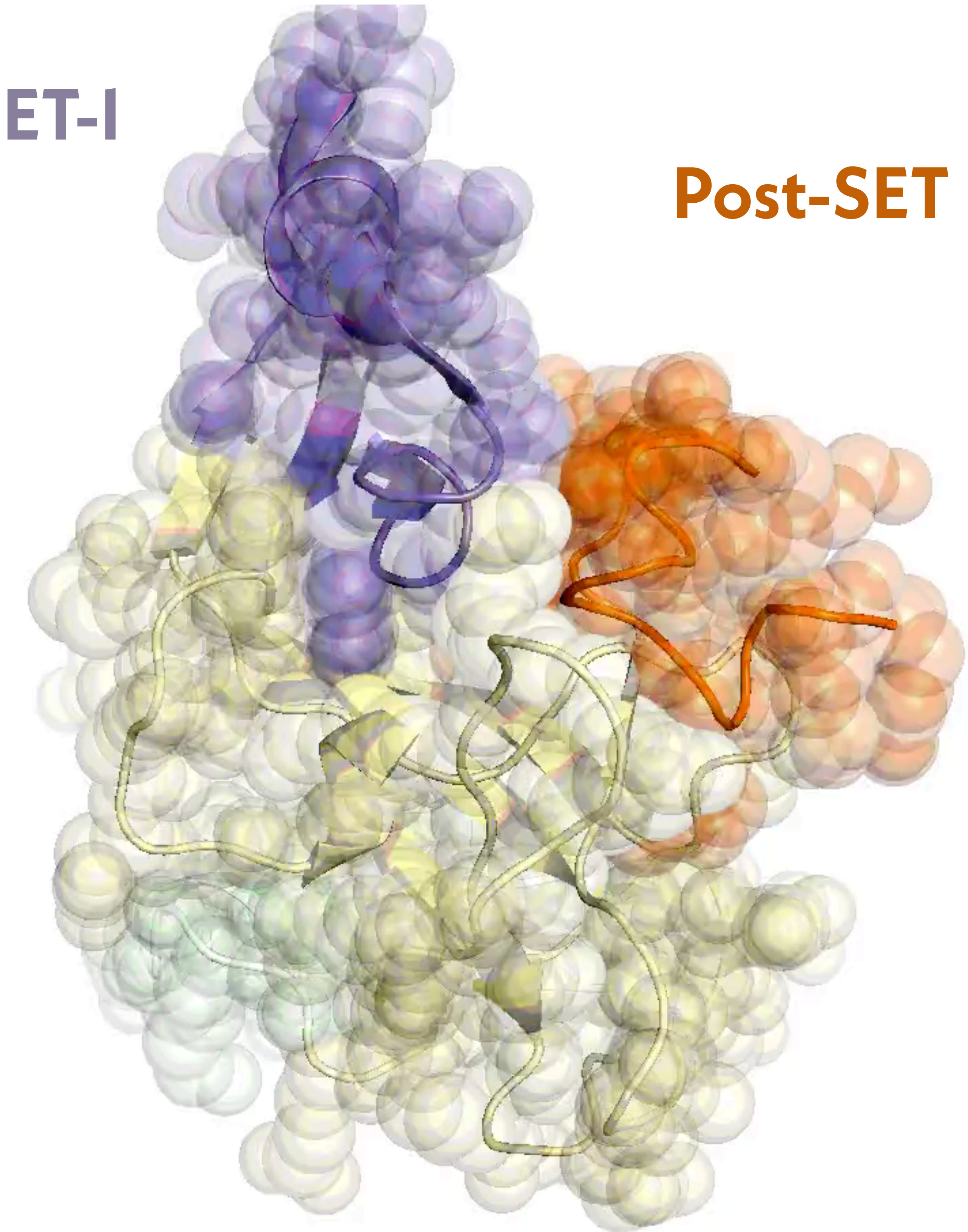
- **k-means clustering** into 50, 100, 500, or 1000 microstate clusters

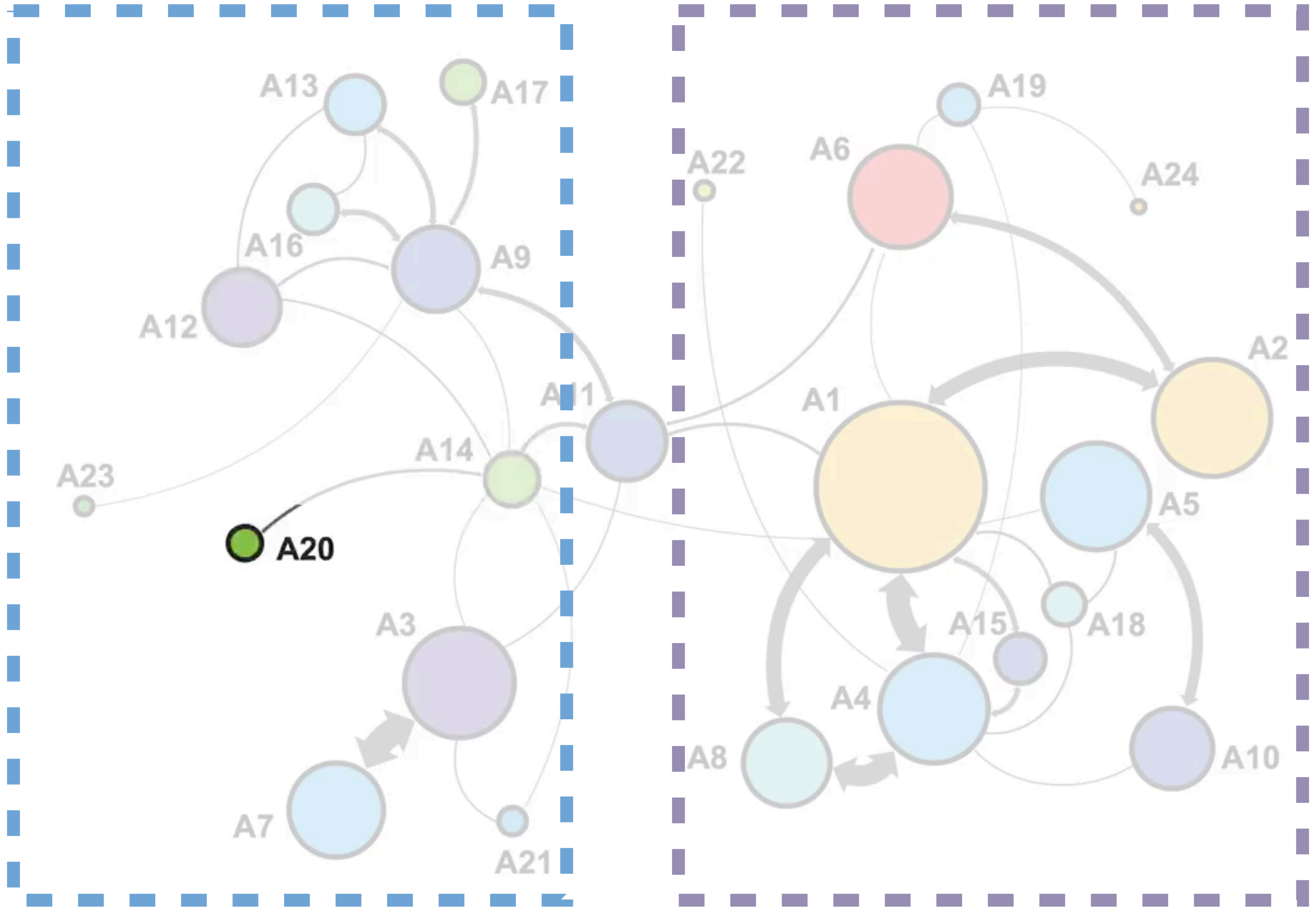
- Cross validation: **training** + **test** sets (50:50 shuffle-split)

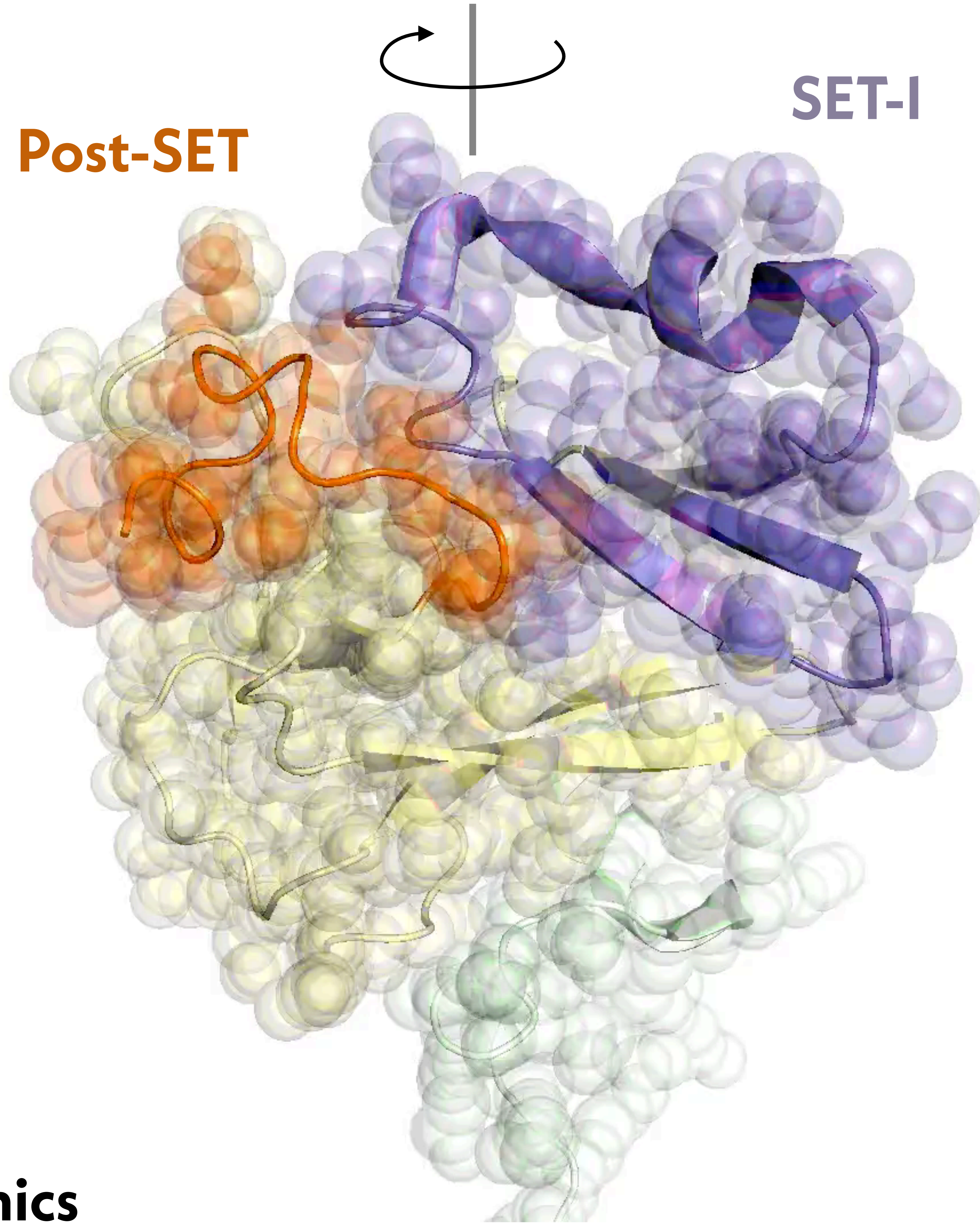
- **Scoring**: Rank-10 **VAMP-2**

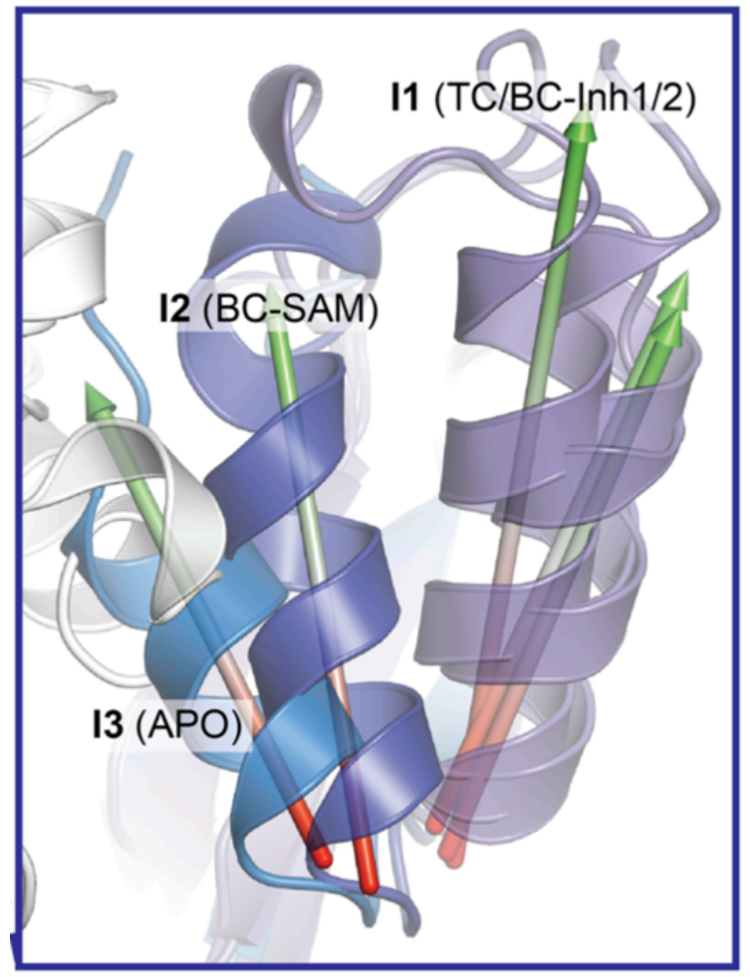
VAMP-2 scores for the different features

The Markov State Model summarizes slow kinetically distinct conformational states







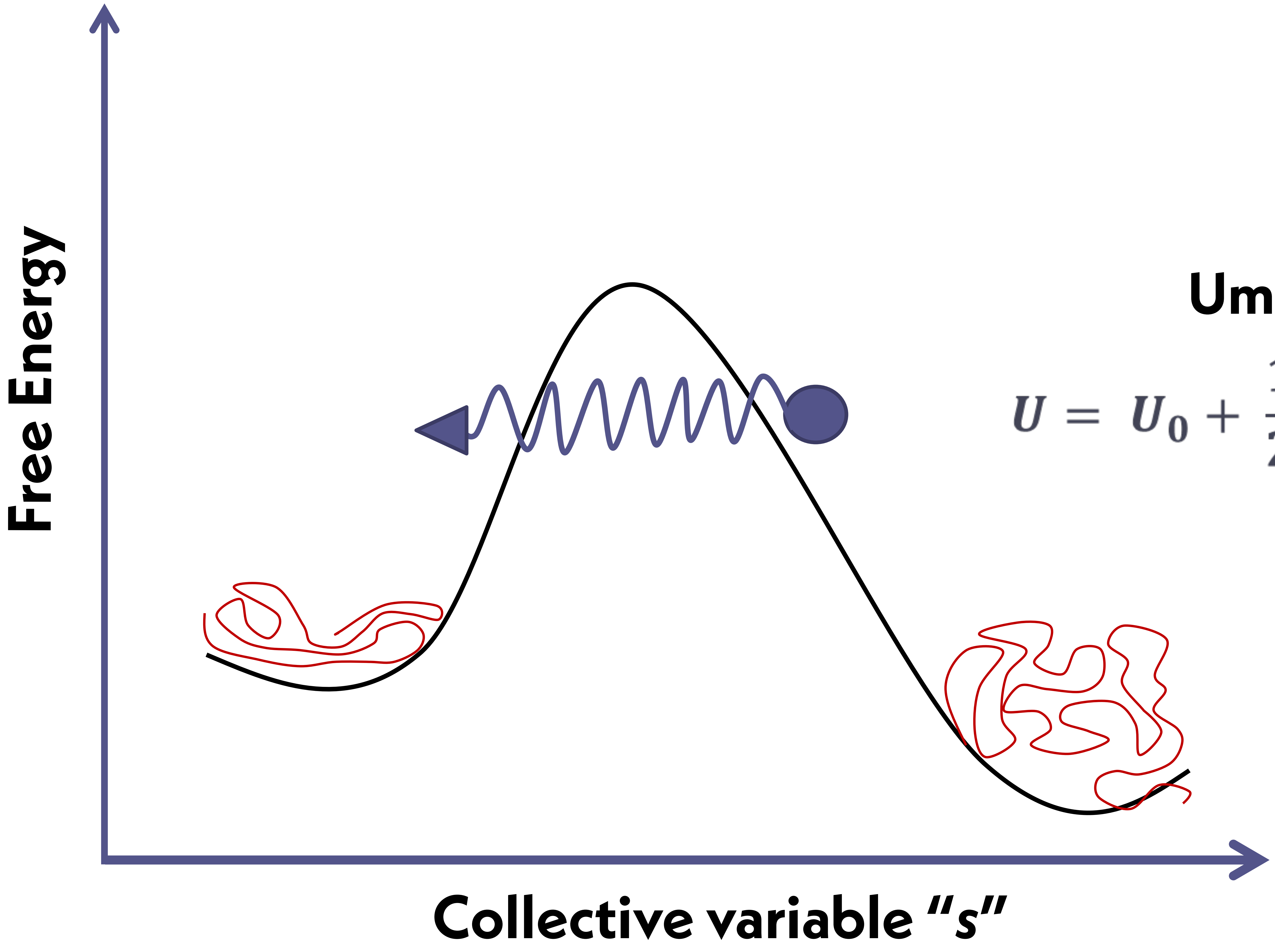


Slow SET-I dynamics

divides the network into TWO LOBES

bioRxiv: https://doi.org/10.1101/438994

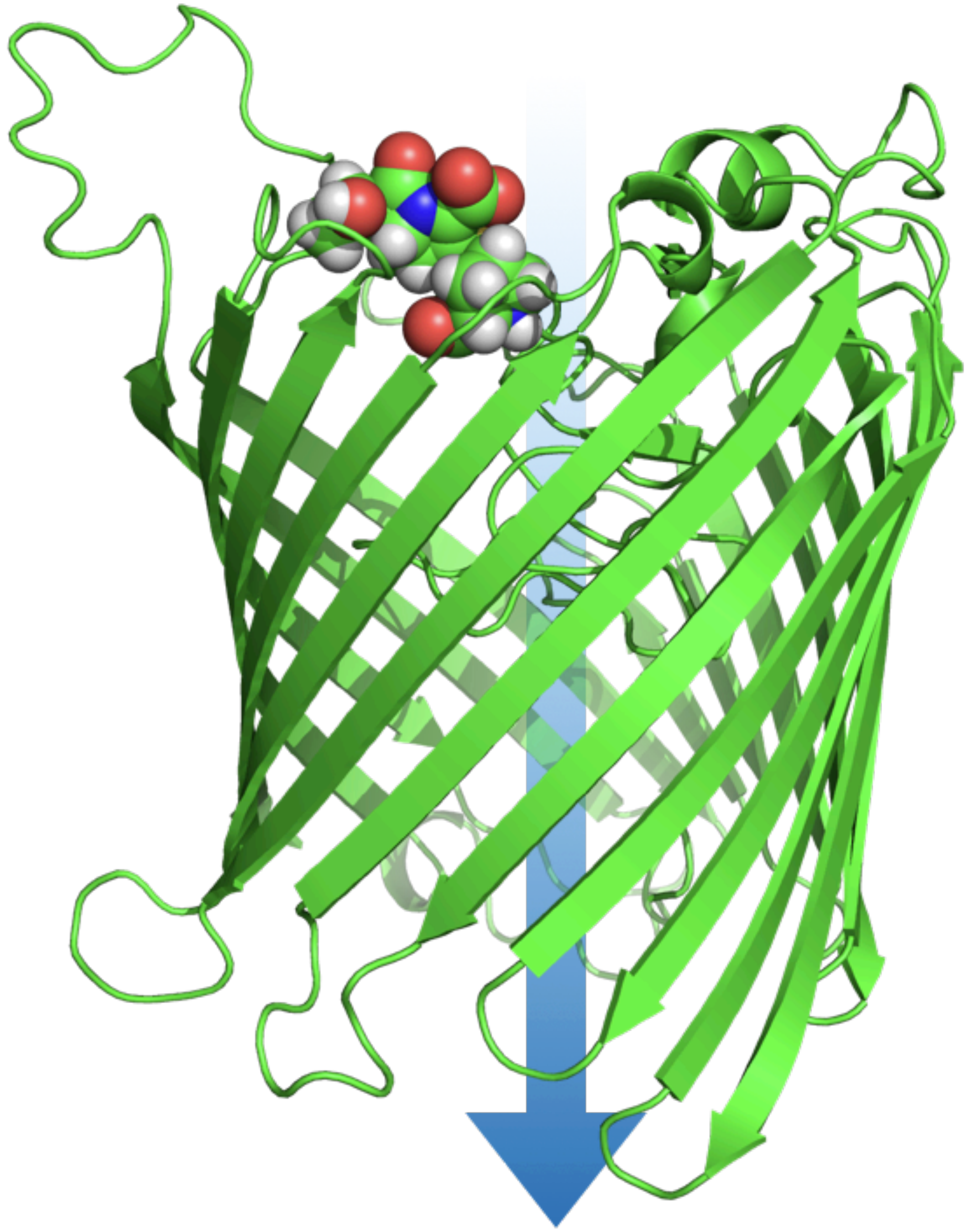
The permeation through porins is a "rare event"



Umbrella sampling

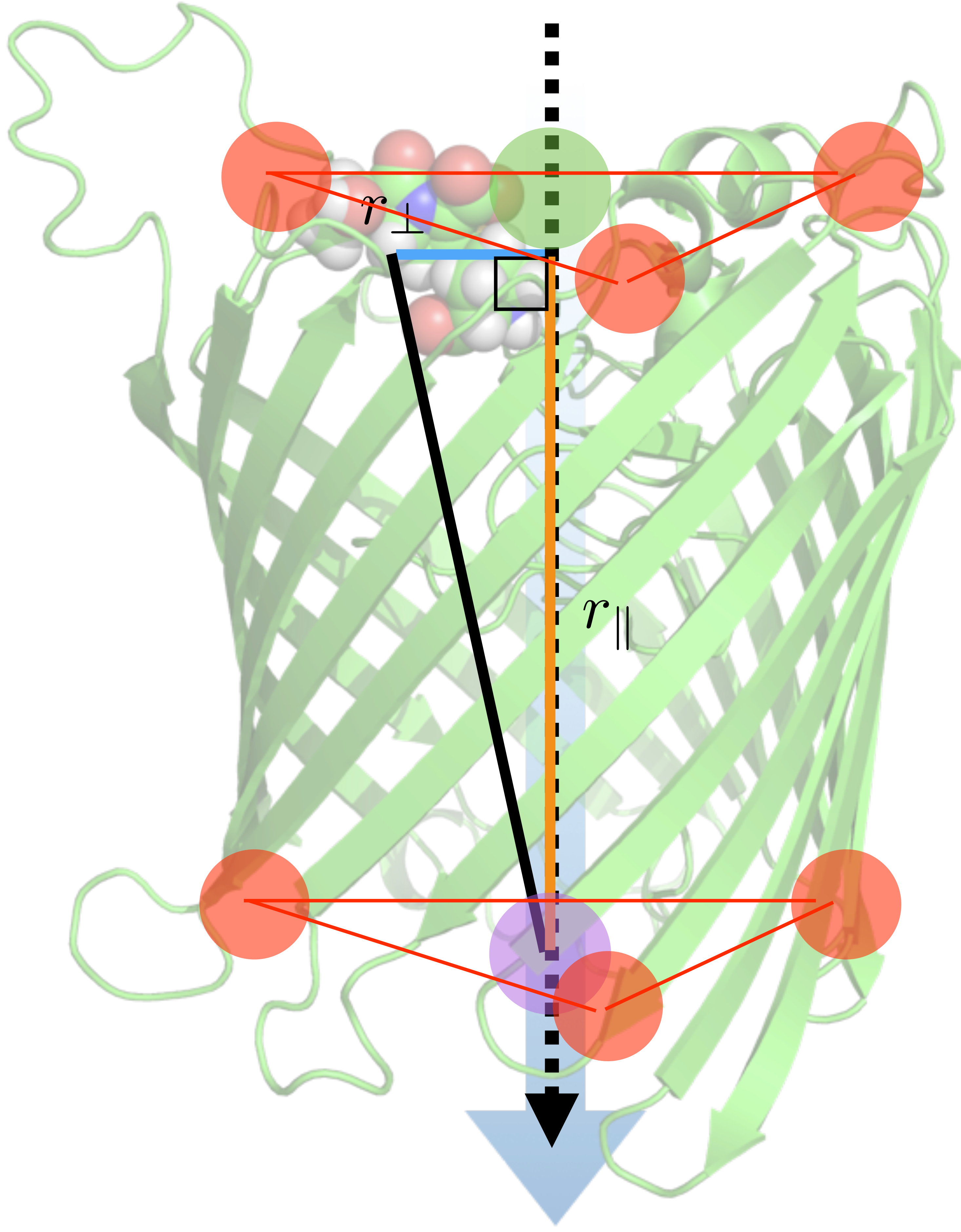
$$U = U_0 + \frac{1}{2}k[f(x_1,..,x_N) - s]^2$$

Defining the permeation coordinate



Defining the permeation coordinate

collective variables

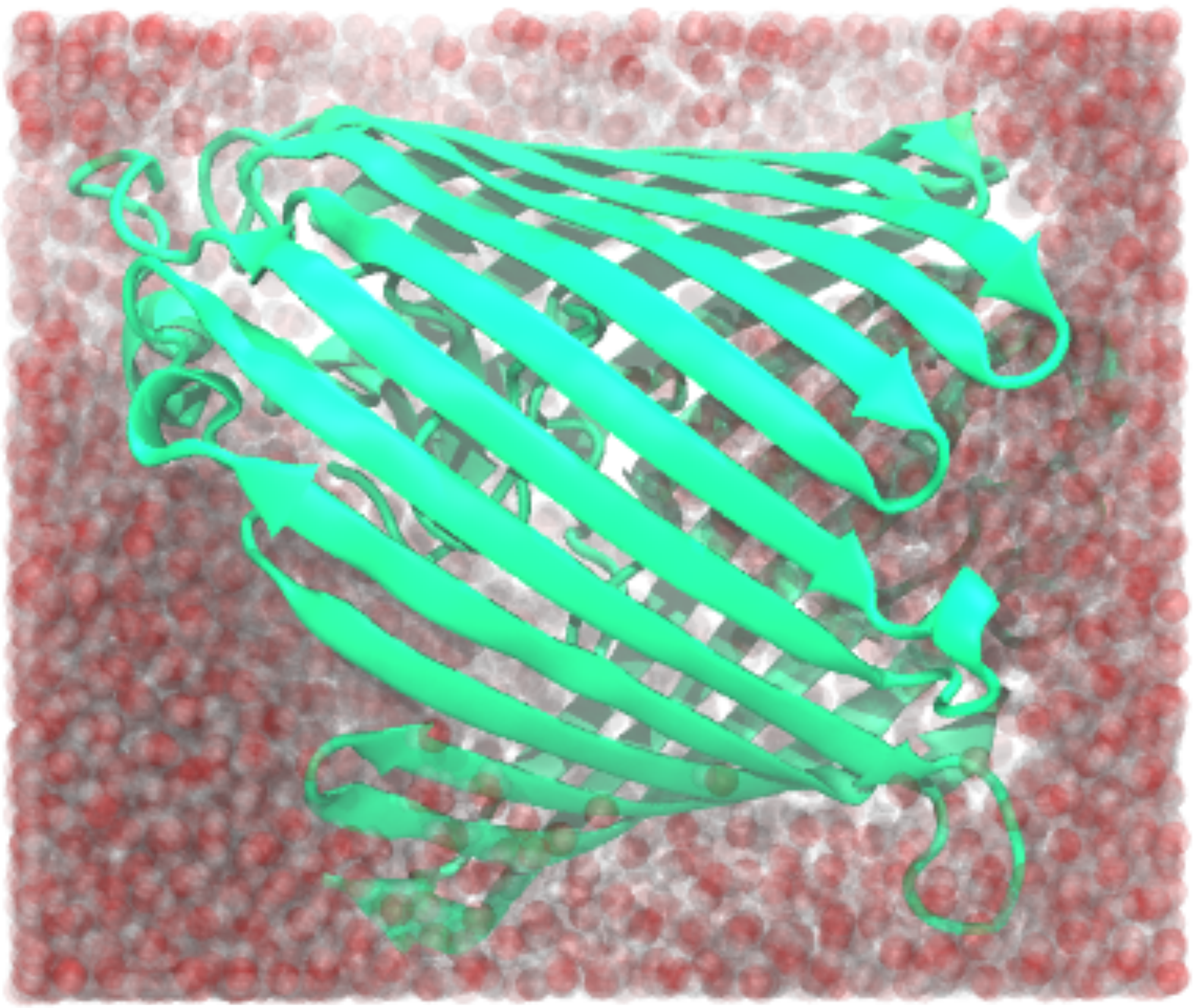


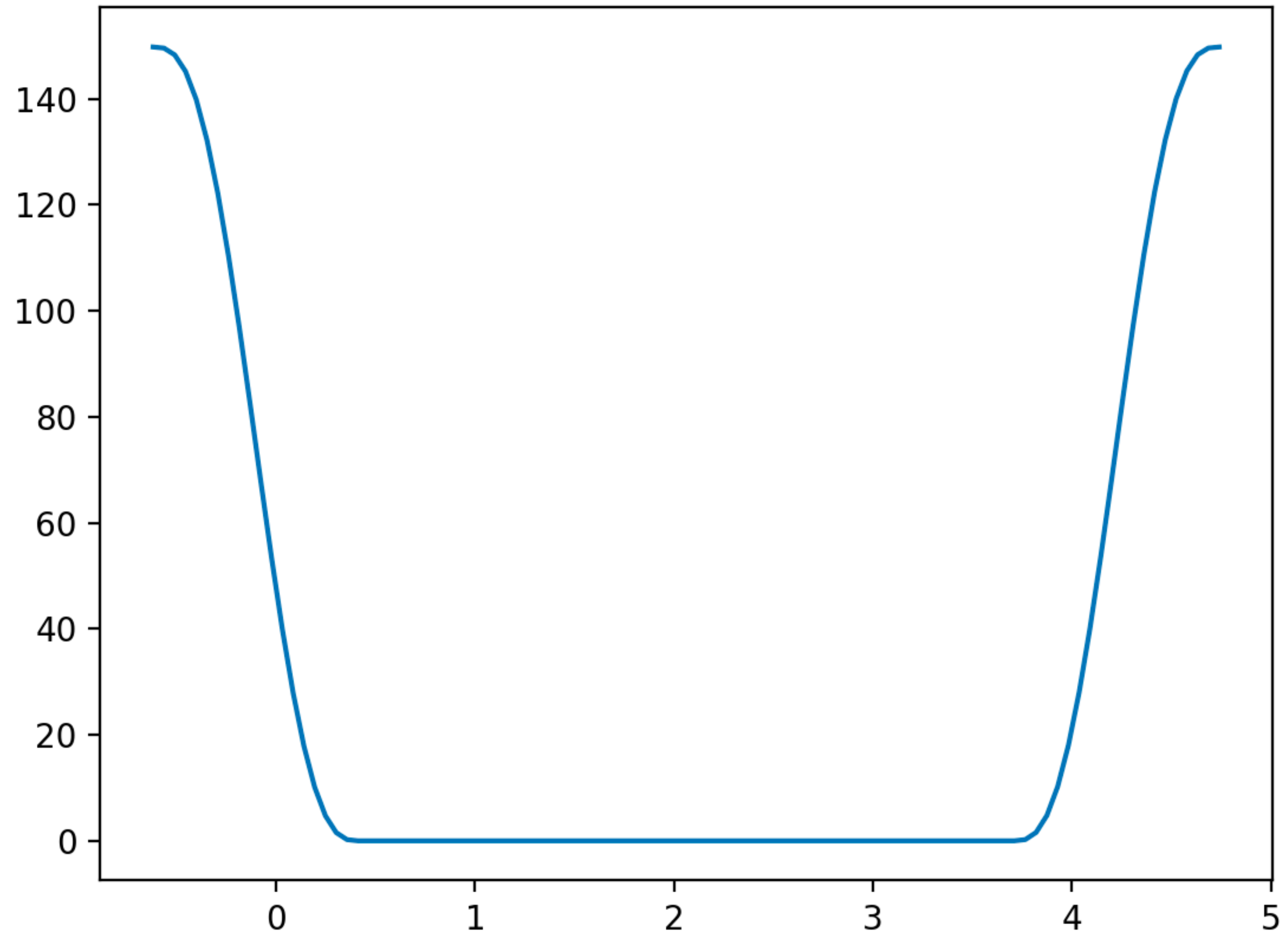
biasing potential

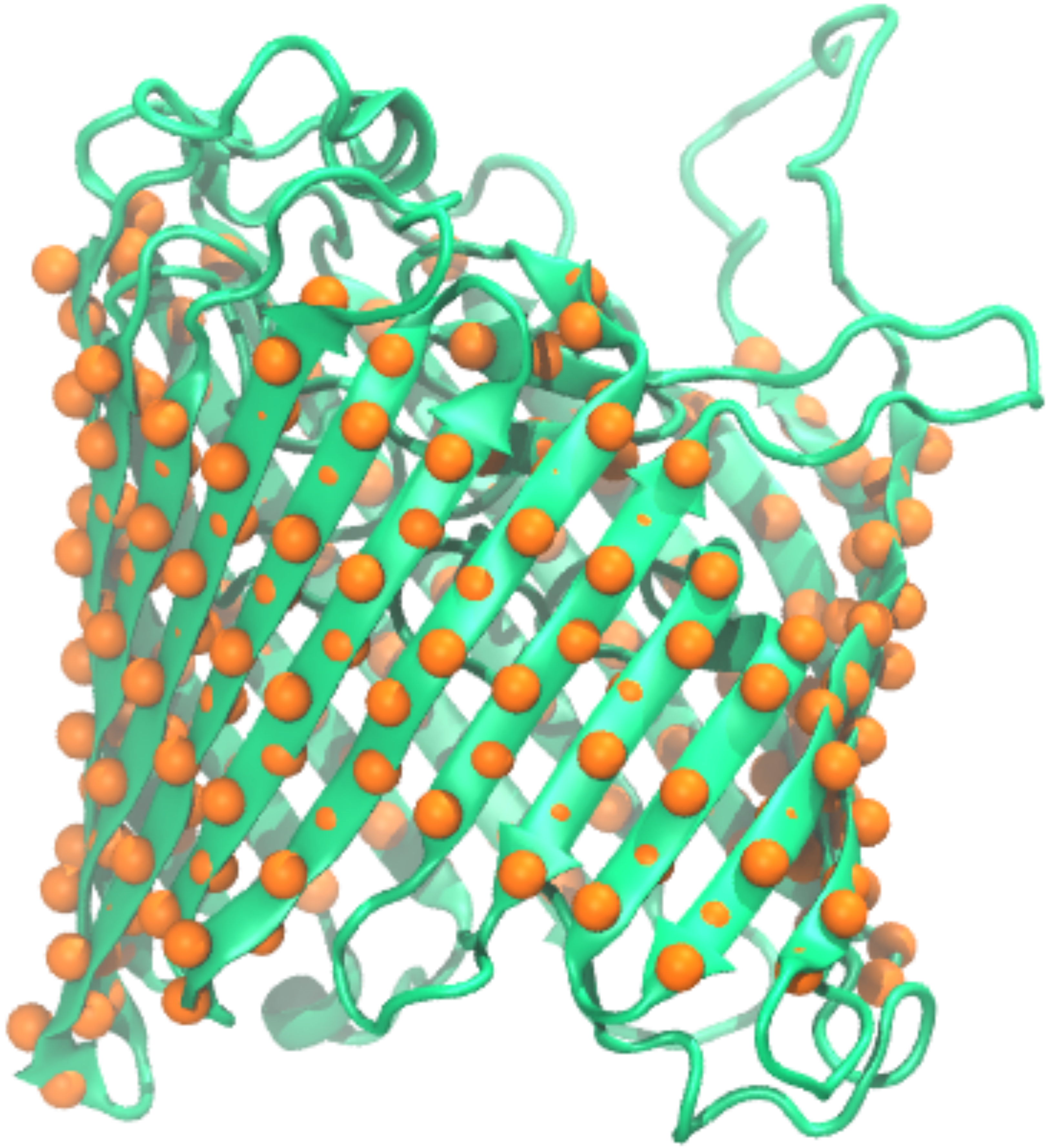
along-axis off-axis

$$U(x) = \frac{K_\parallel}{2}[r_\parallel - r_0(\lambda)]^2 + \frac{K_\perp}{2} r_\perp^2$$

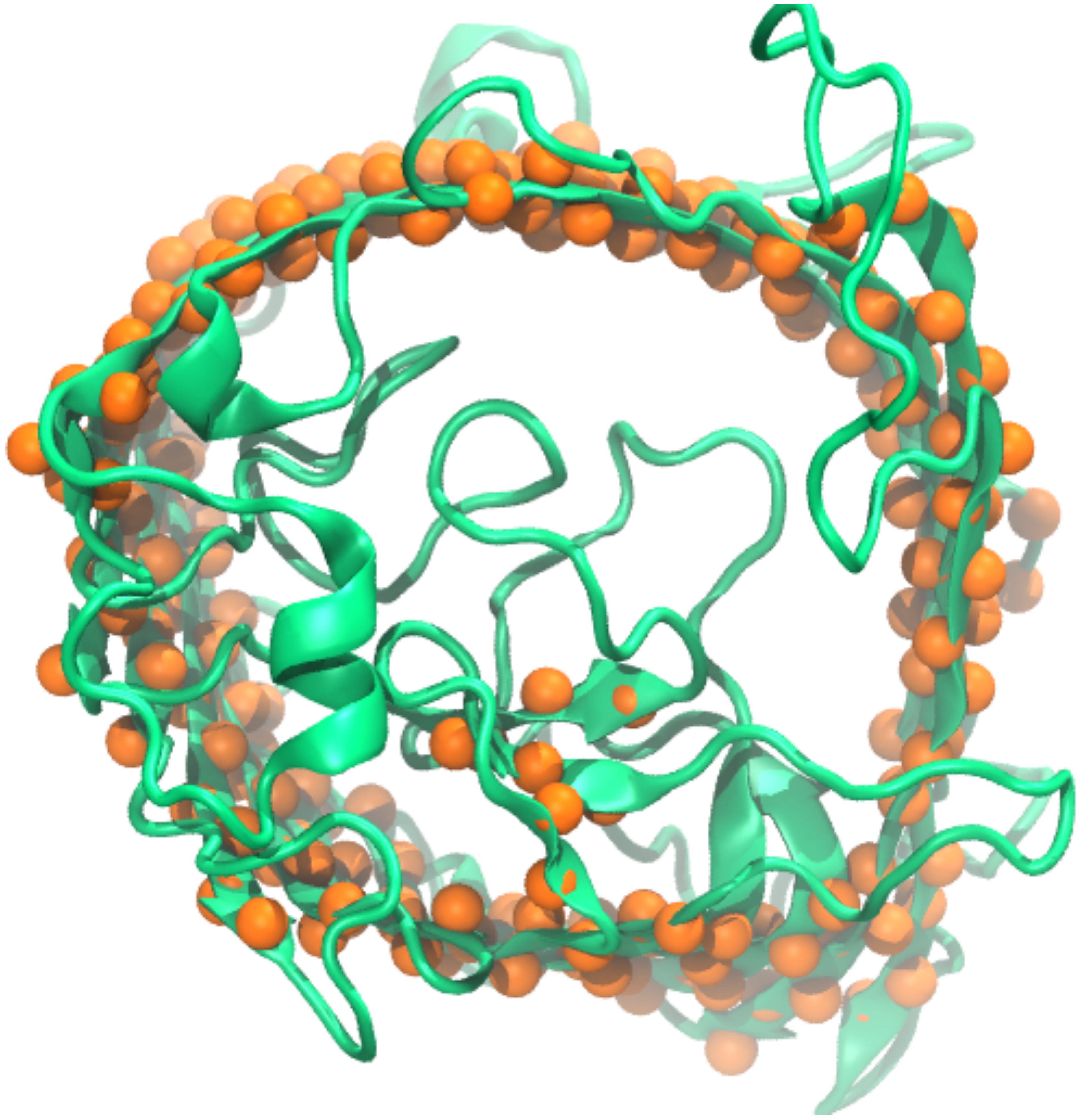
K orthogonal along the permeation coordinate





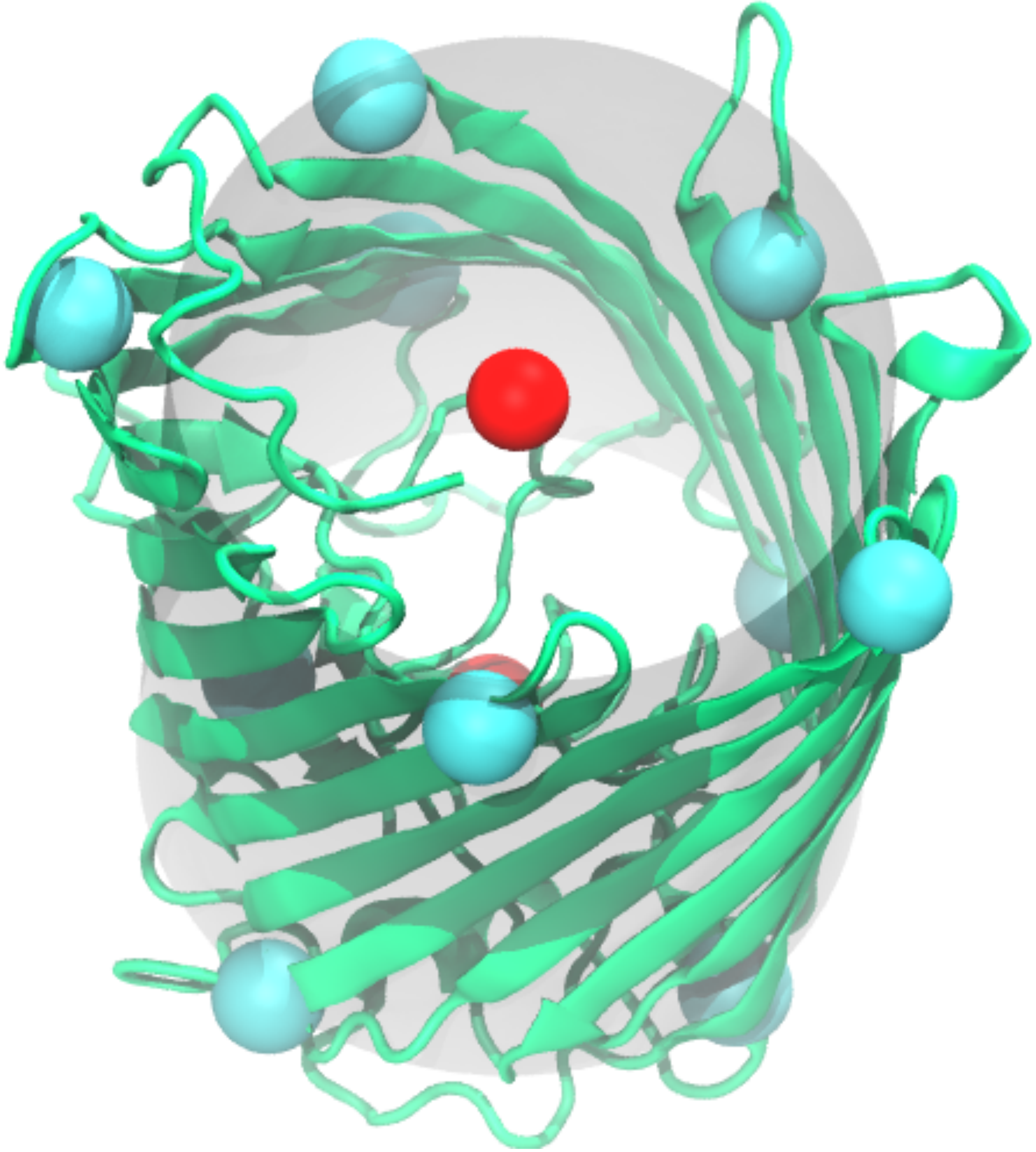


Side view

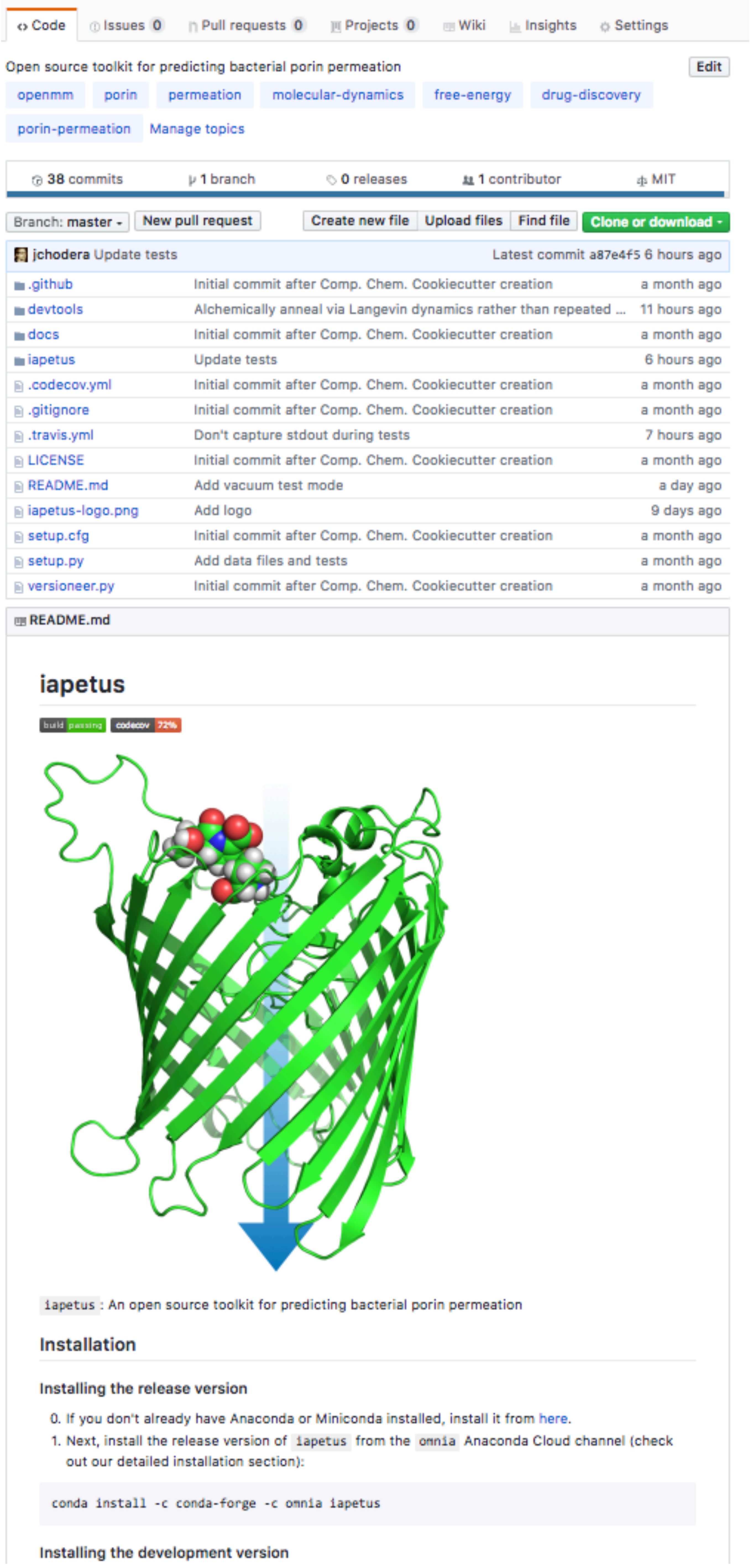


Top view

CYLINDER FITTING



- **alpha carbons** of **beta sheets**
- **Best cylinder: least squares minimization**
- Automatically select atoms in the bottom and top of the porin: geometrical center **closest** to the cylinder center

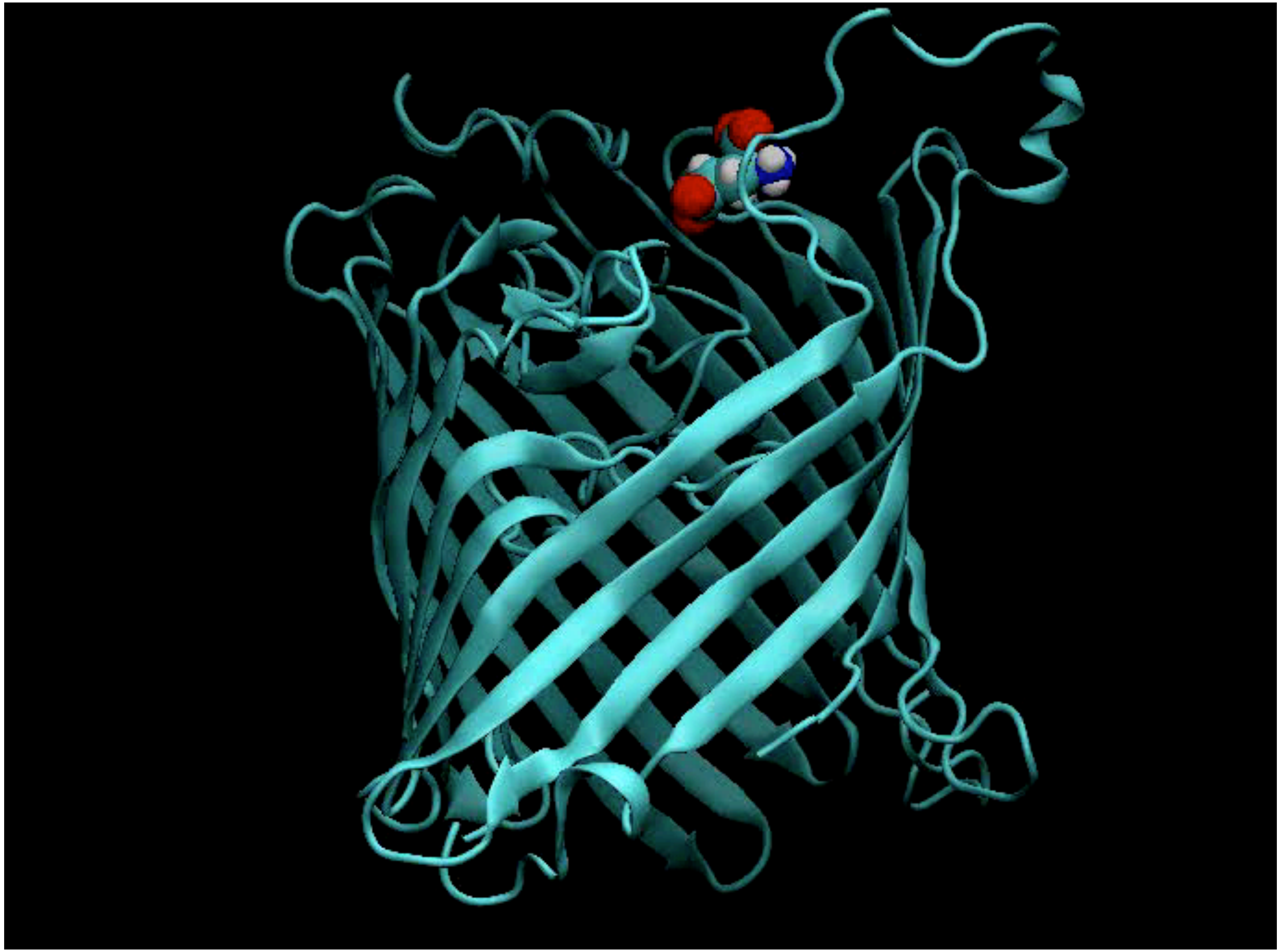


IAPETUS ("THE PIERCER")

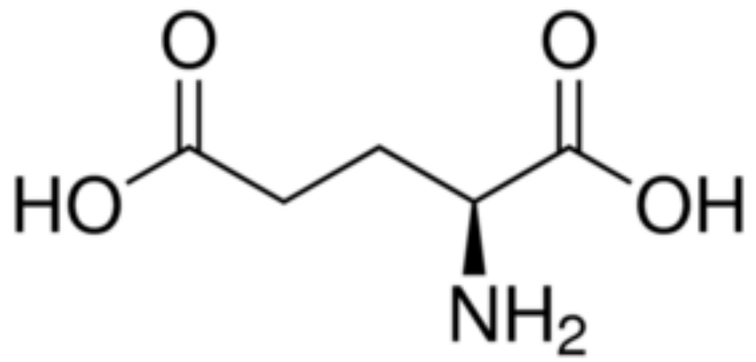
source code: https://github.com/choderalab/iapetus

examples: https://github.com/choderalab/iapetus-examples

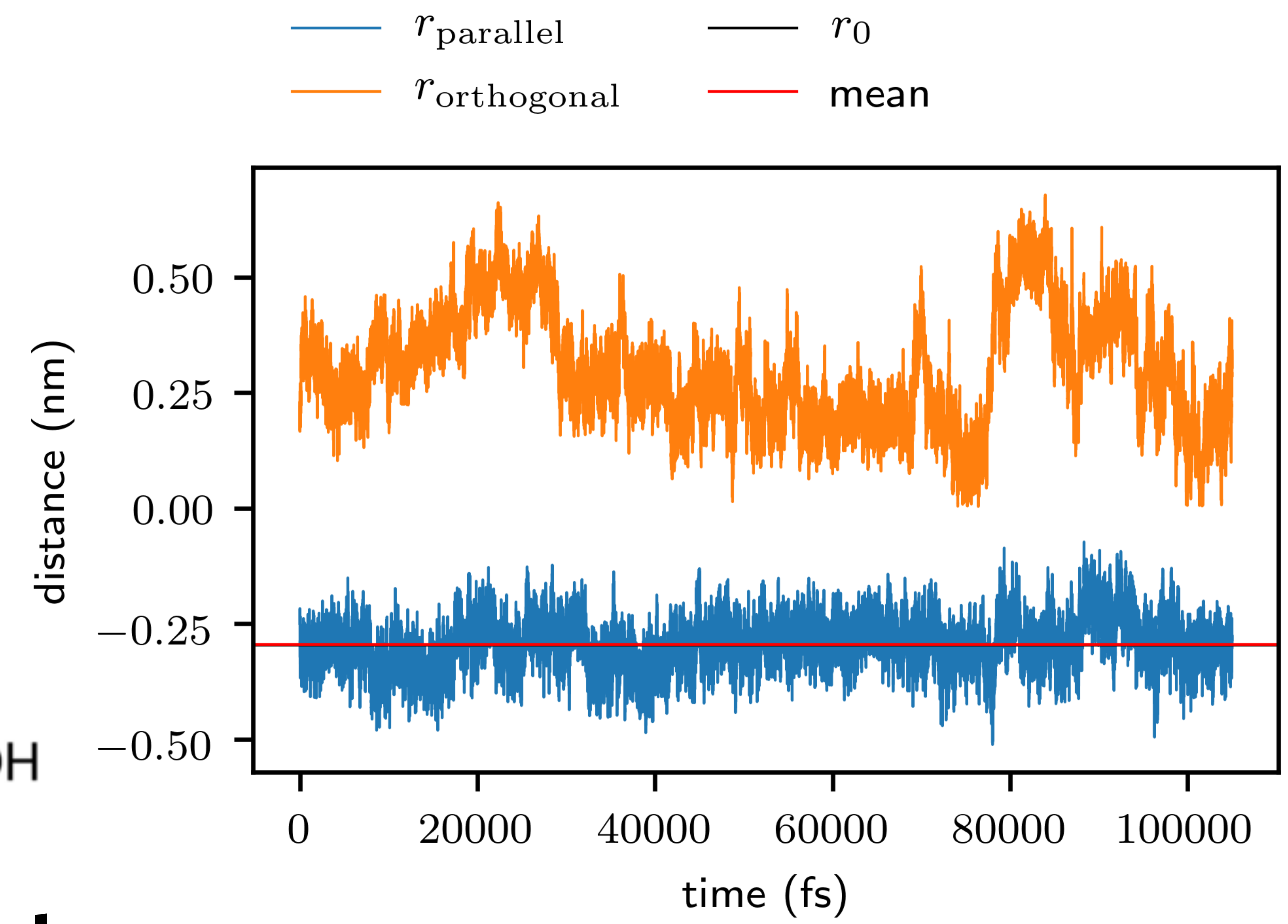
license: MIT (OSI-approved permissive license)

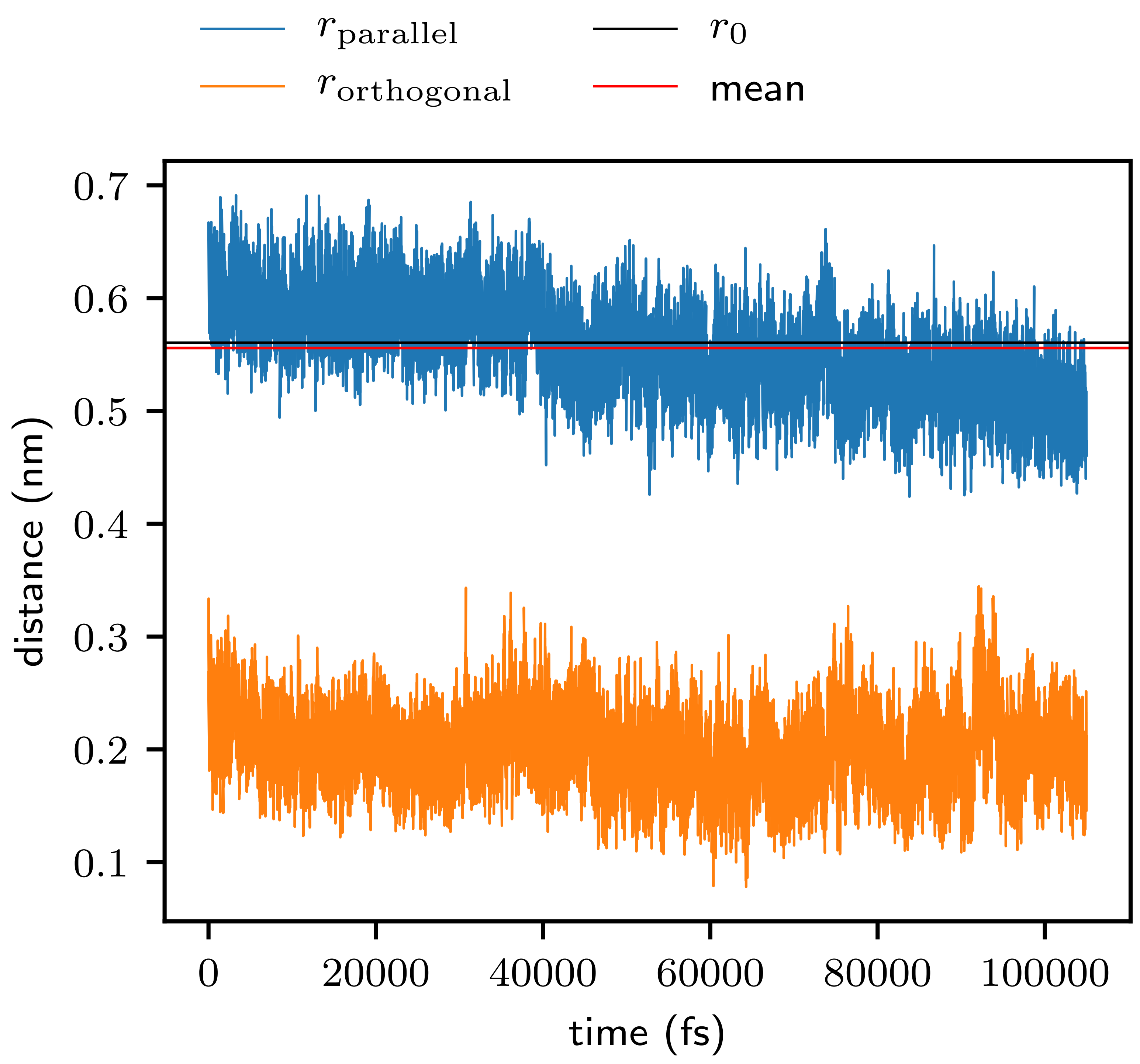


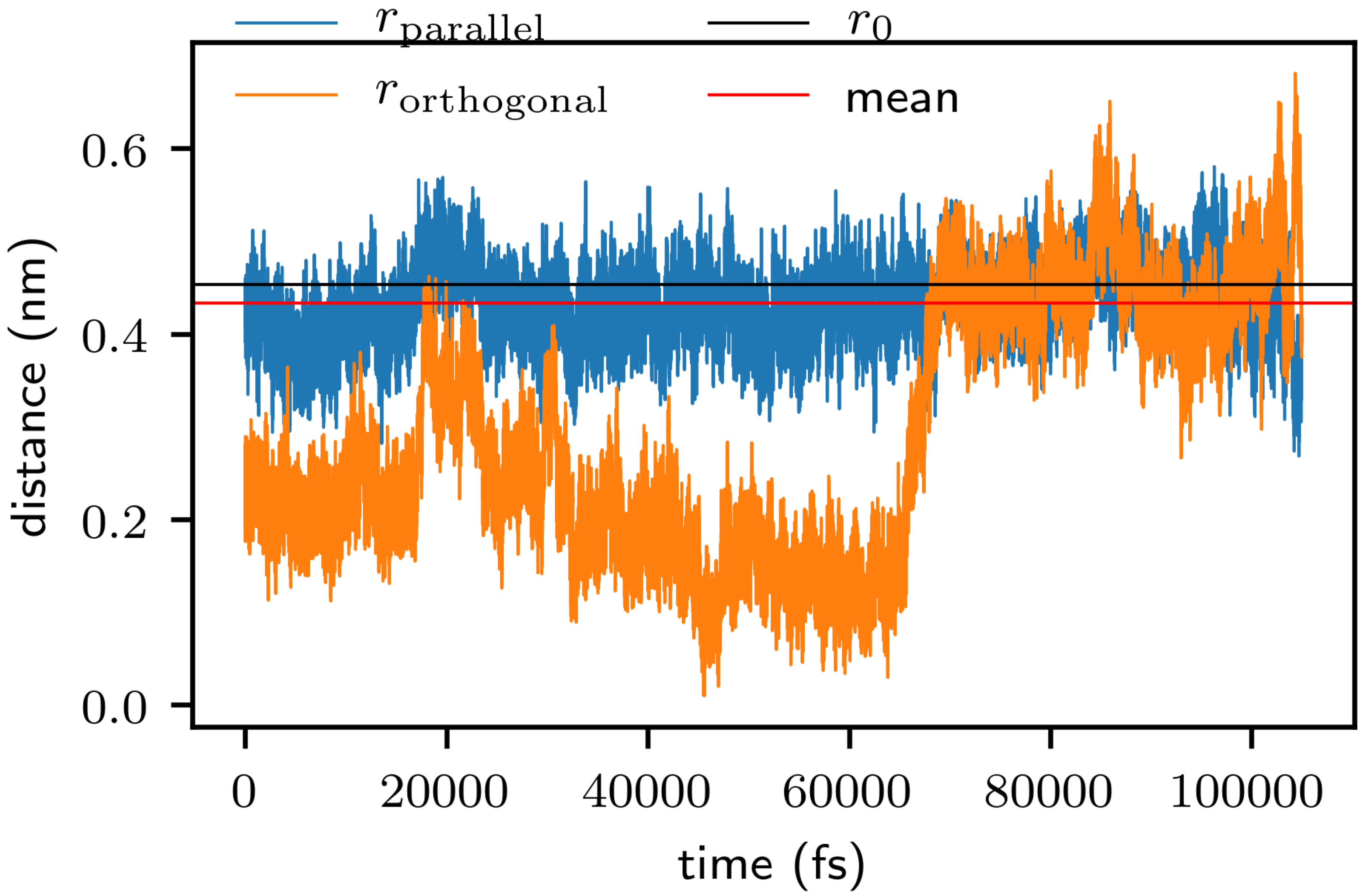
Umbrella Sampling

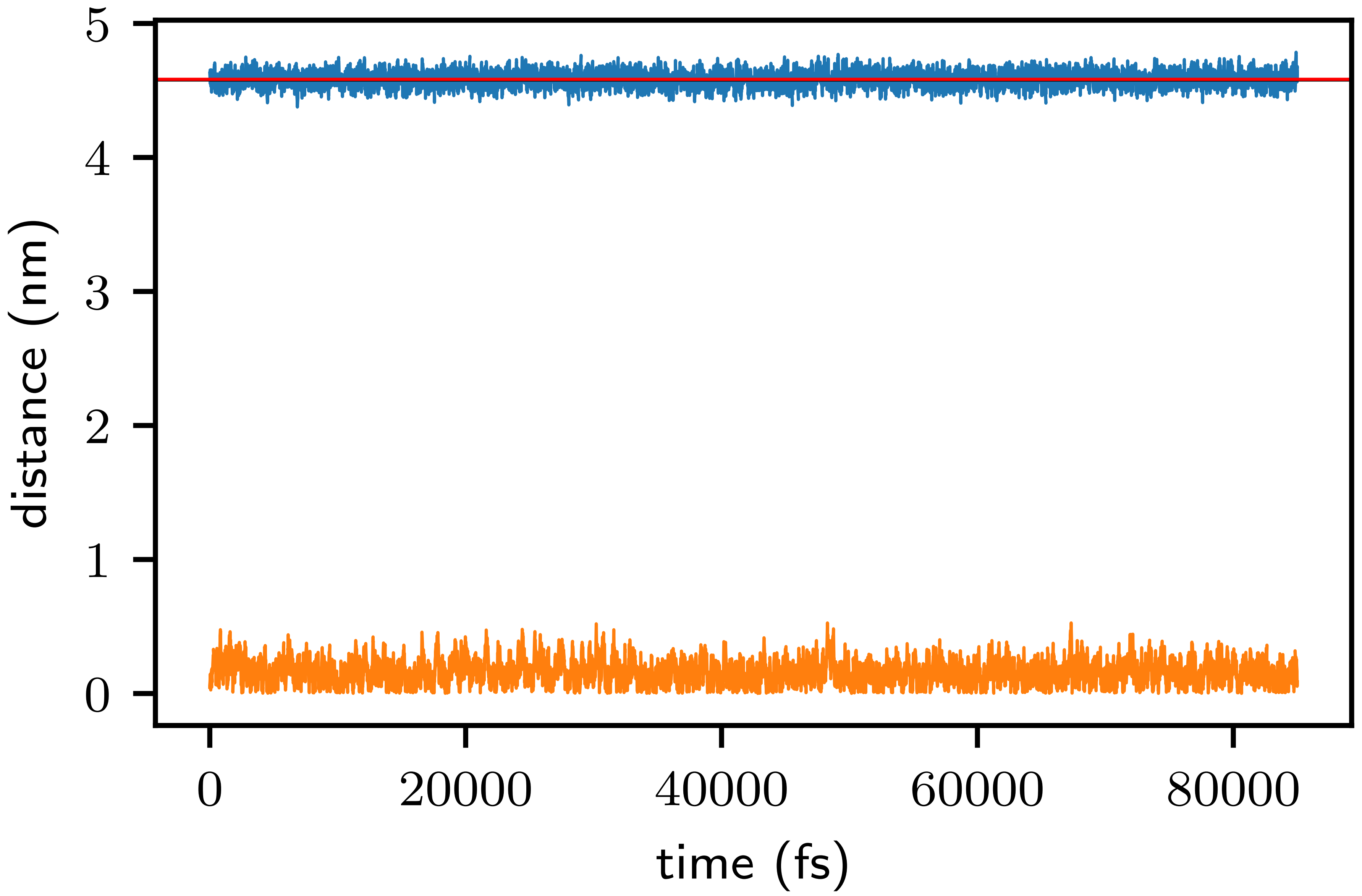


Glutamic acid

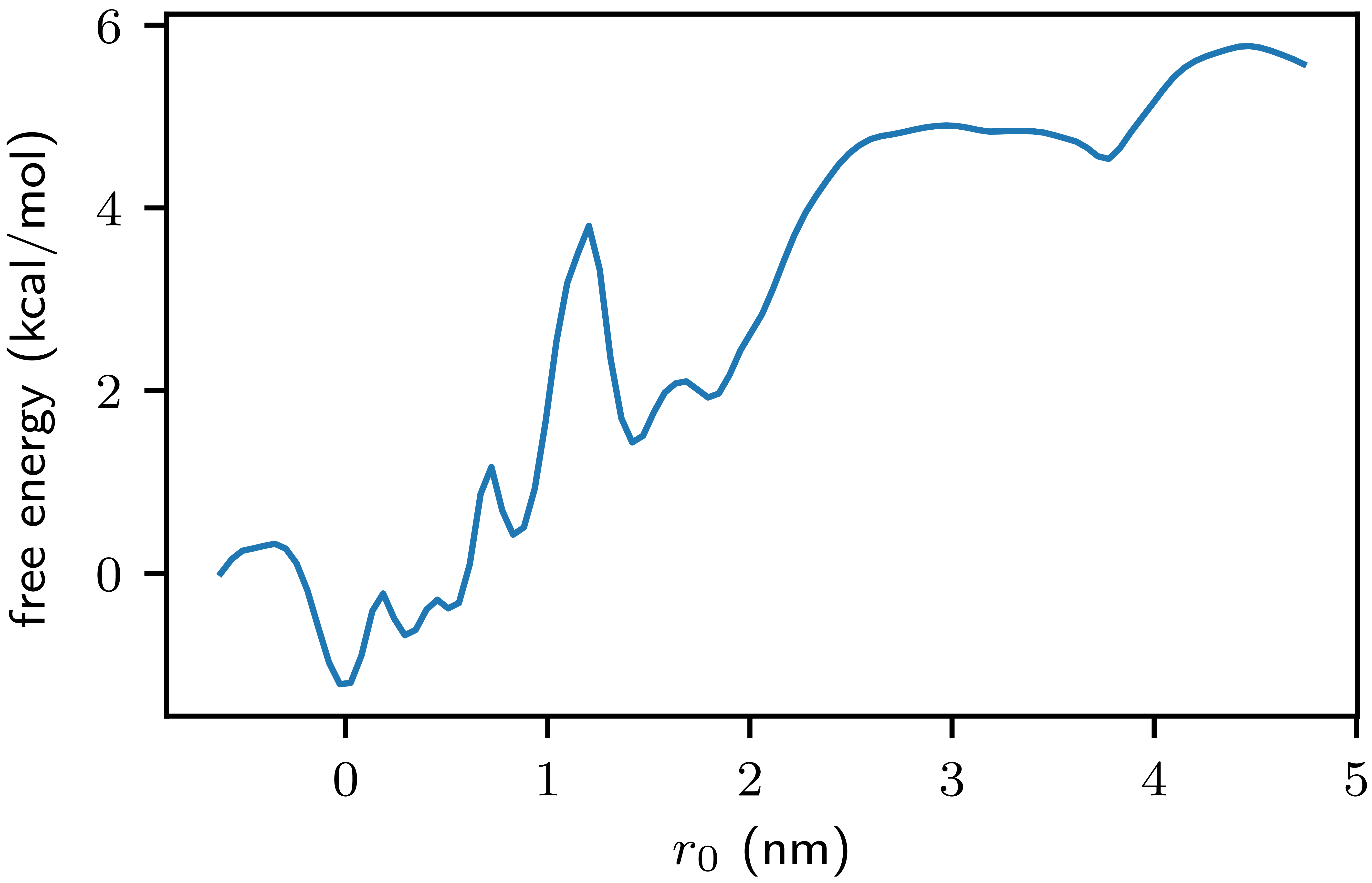




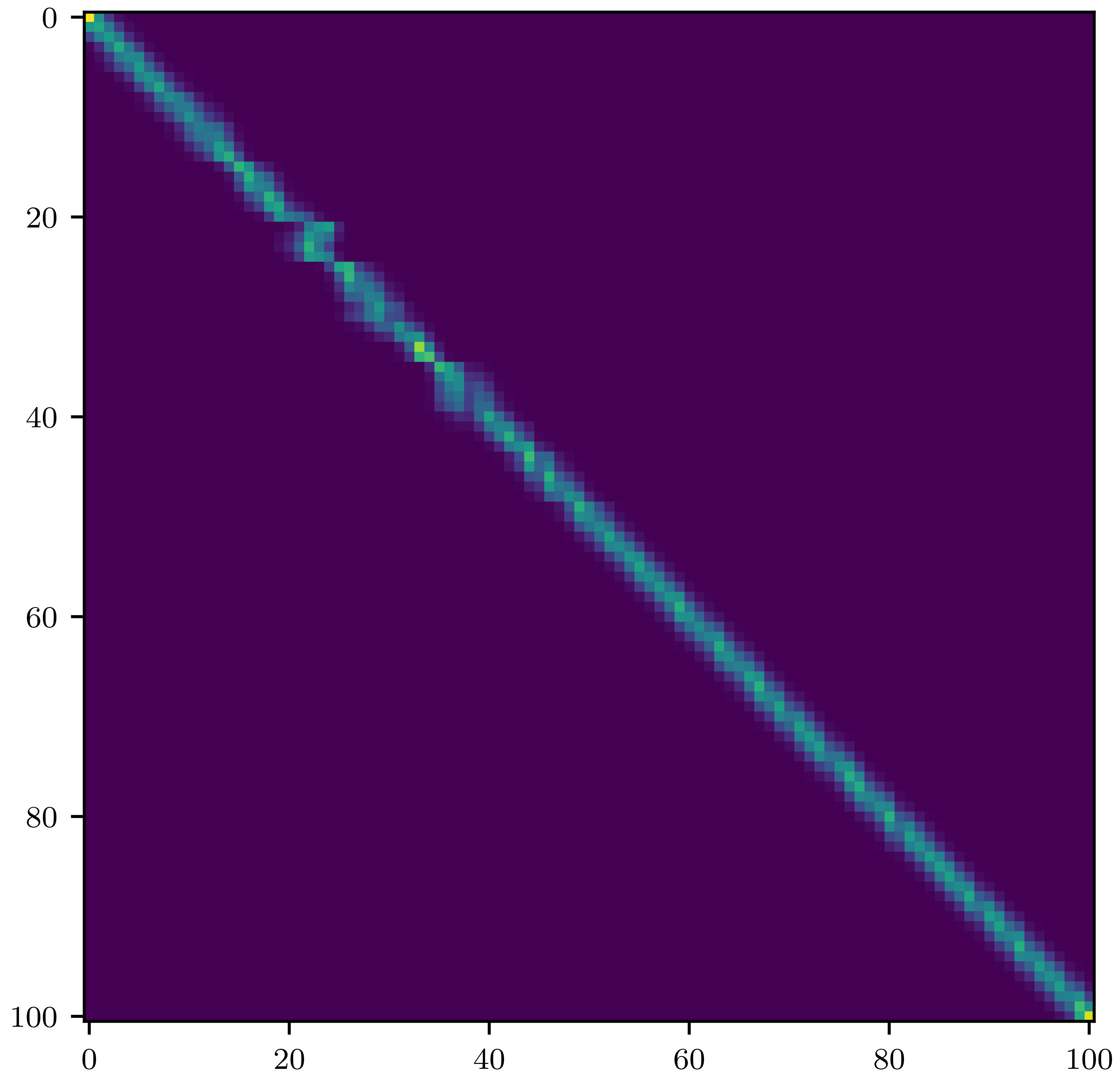




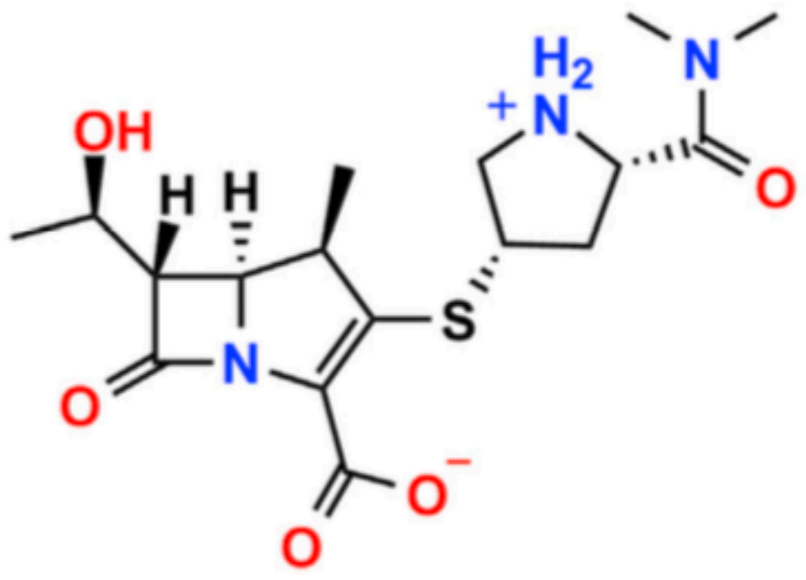
Potential of mean force



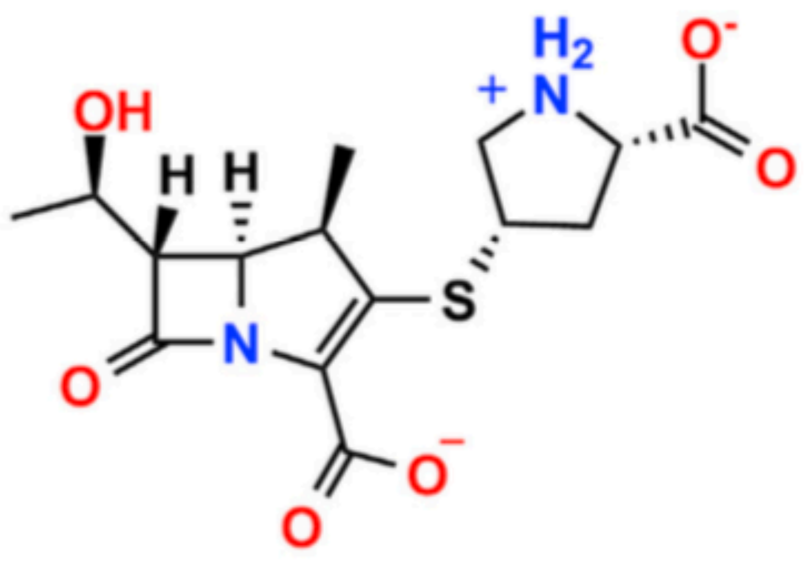
Overlap Matrix



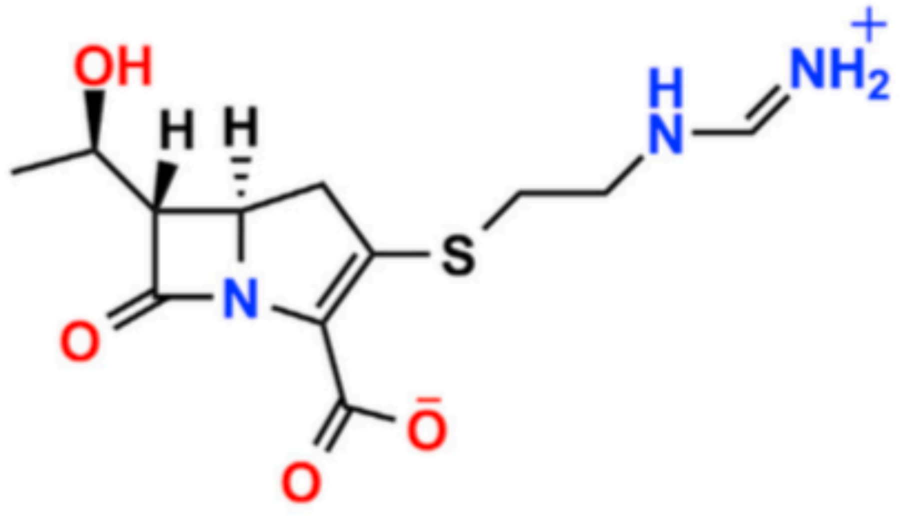
SET OF INTEREST



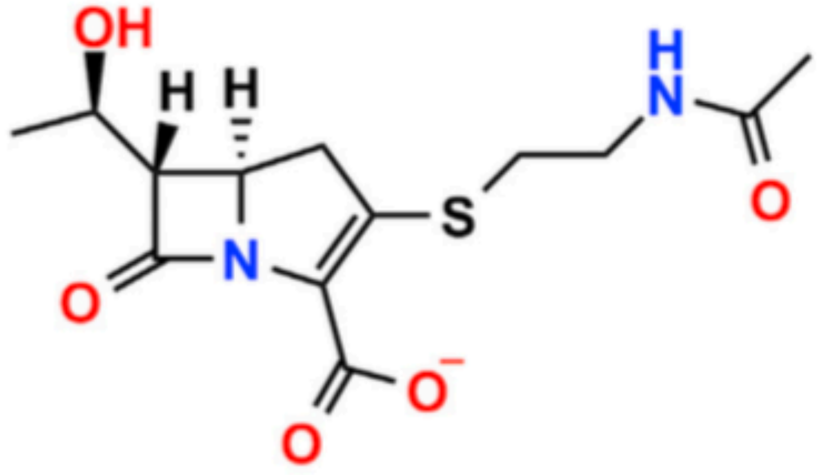
4 Meropenem



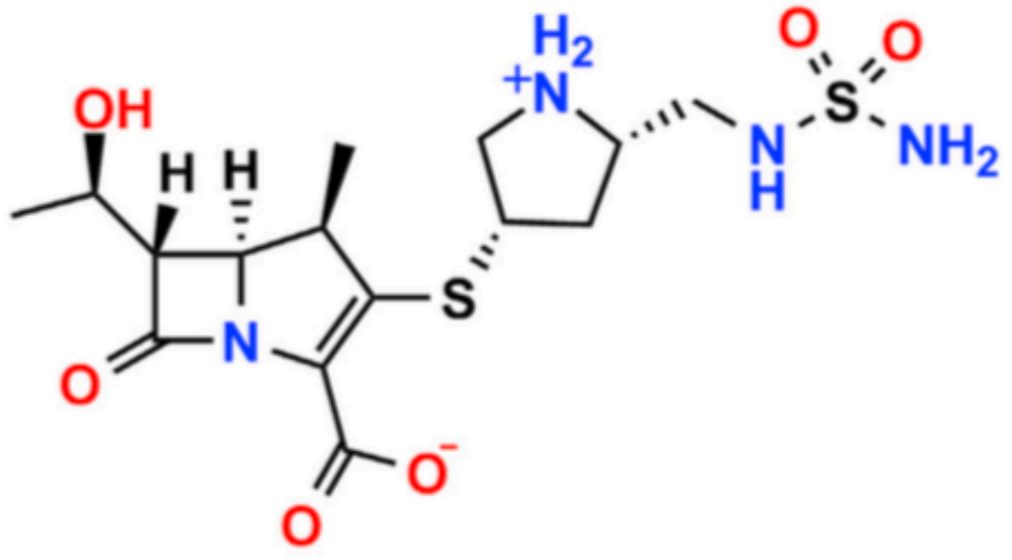
7



5 Imipenem



8



6 Doripenem

THANK YOU!