

# Transition path sampling and the calculation of free energies of enzymatic reactions

Sree Ganesh Balasubramani

Schwartz Group  
Chemistry and Biochemistry



# Outline

- Motivations

# Outline

- Motivations
- Introduction to transition path sampling

# Outline

- Motivations
- Introduction to transition path sampling
- Equilibrium sampling within TPS

# Outline

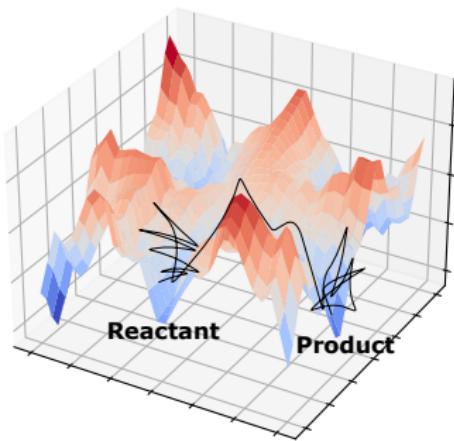
- Motivations
- Introduction to transition path sampling
- Equilibrium sampling within TPS
- Algorithm

# Outline

- Motivations
- Introduction to transition path sampling
- Equilibrium sampling within TPS
- Algorithm
- Results

# Motivations: Sampling rare but important events

# Motivations: Sampling rare but important events



## Time scales in molecular dynamics simulations

- Consider the dissociation of a weak acid in water, with a half life of say  $1\text{ ms}$
- Straightforward MD simulations with  $1\text{ fs}$  time step (takes  $1\text{ s}$  of wall time to simulate) would take  $10^{12}\text{ s}$  of wall time to observe a single dissociation event

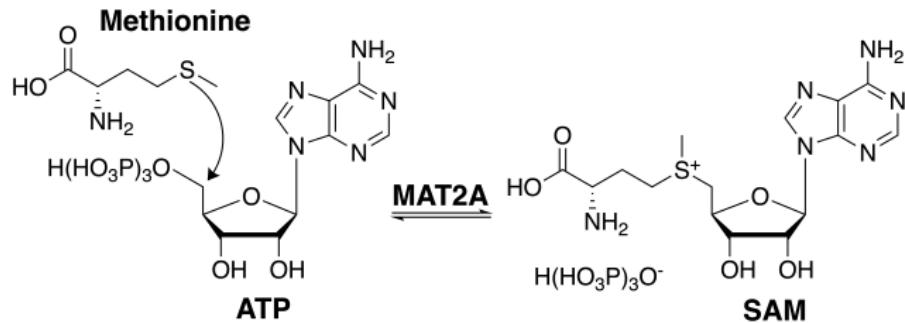
# Enzymes

# Enzymes

Enzymes are evolved to carry out complex chemical transformations routinely

# Enzymes

Enzymes are evolved to carry out complex chemical transformations routinely



# Enhanced sampling techniques

# Enhanced sampling techniques

Umbrella sampling, metadynamics, milestoning, conformational flooding, adaptive biasing force method and so on

# Enhanced sampling techniques

Umbrella sampling, metadynamics, milestoning, conformational flooding, adaptive biasing force method and so on

- Rendered MD simulations to be widely applicable in the fields of chemistry, condensed matter physics, biology and materials science

# Enhanced sampling techniques

Umbrella sampling, metadynamics, milestoning, conformational flooding, adaptive biasing force method and so on

- Rendered MD simulations to be widely applicable in the fields of chemistry, condensed matter physics, biology and materials science
- Use of predefined reaction coordinates, which can be complicated for large biomolecular systems

# Enhanced sampling techniques

Umbrella sampling, metadynamics, milestoning, conformational flooding, adaptive biasing force method and so on

- Rendered MD simulations to be widely applicable in the fields of chemistry, condensed matter physics, biology and materials science
- Use of predefined reaction coordinates, which can be complicated for large biomolecular systems
- Use of biasing potential to access regions in phase space with high free energy barriers

# Transition path sampling (TPS)<sup>1</sup>

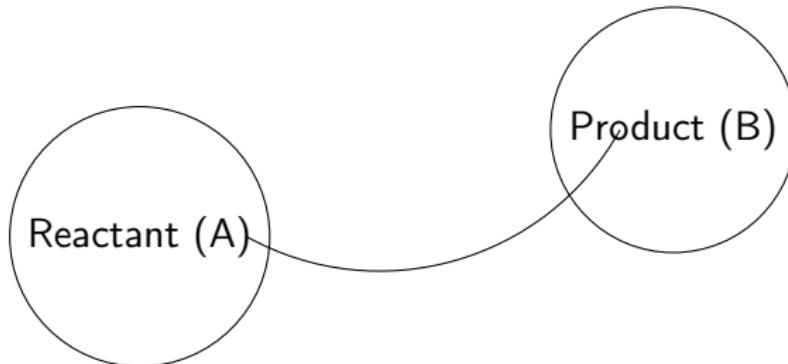
<sup>1</sup>P. G. Bolhuis et al., Annual Review of Physical Chemistry 53, 291 (2002).

# Transition path sampling (TPS)<sup>1</sup>



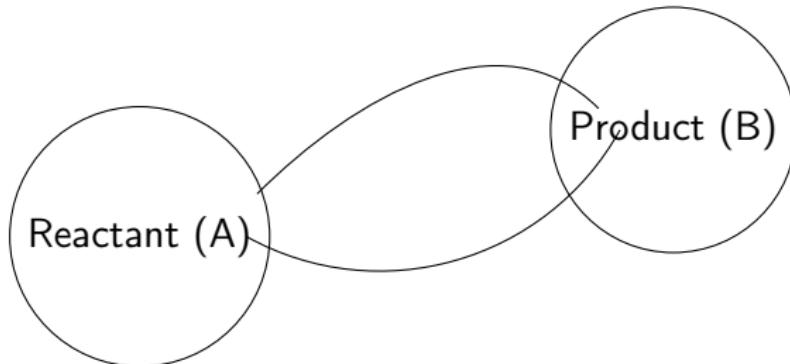
- Focus on pathways connecting long lived reactant and product states
- Reaction coordinate and bias potential free description
- Monte Carlo algorithm that is easy to implement
- Transition state, reaction coordinates, kinetic parameters

# Transition path sampling (TPS)<sup>1</sup>



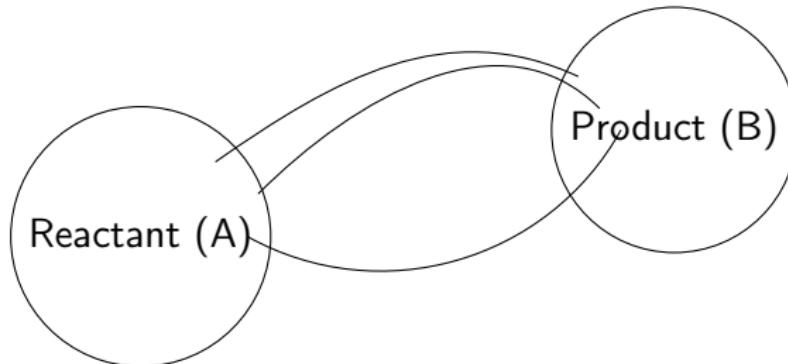
- Focus on pathways connecting long lived reactant and product states
- Reaction coordinate and bias potential free description
- Monte Carlo algorithm that is easy to implement
- Transition state, reaction coordinates, kinetic parameters

# Transition path sampling (TPS)<sup>1</sup>



- Focus on pathways connecting long lived reactant and product states
- Reaction coordinate and bias potential free description
- Monte Carlo algorithm that is easy to implement
- Transition state, reaction coordinates, kinetic parameters

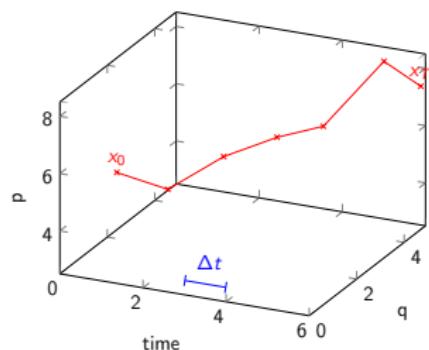
# Transition path sampling (TPS)<sup>1</sup>



- Focus on pathways connecting long lived reactant and product states
- Reaction coordinate and bias potential free description
- Monte Carlo algorithm that is easy to implement
- Transition state, reaction coordinates, kinetic parameters

<sup>1</sup>P. G. Bolhuis et al., Annual Review of Physical Chemistry 53, 291 (2002).

# Trajectories in phase space

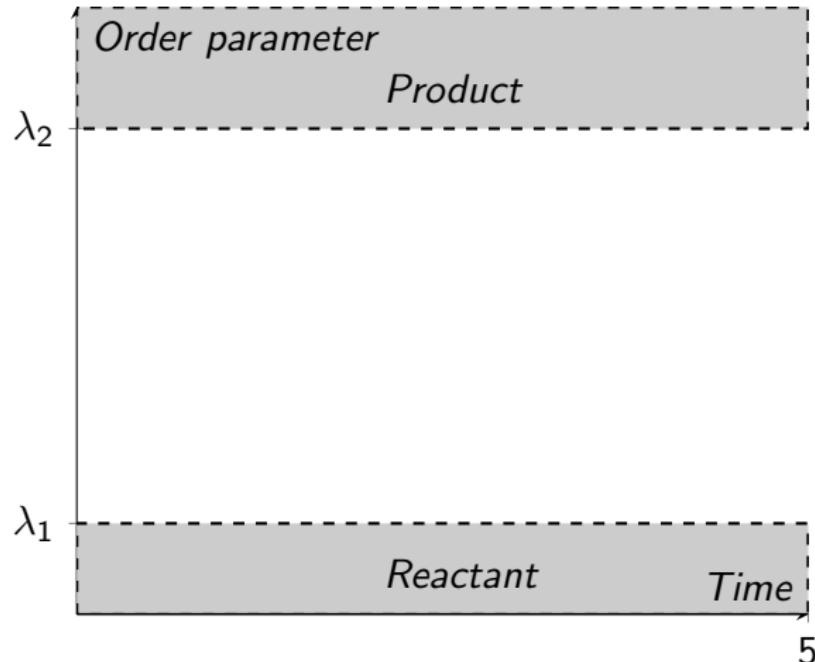


## Definitions

- $x_0$ : initial state,  $T$  : time length of trajectory,  $\Delta t$  : time step,  $x_{i\Delta t}$  : time slice ( $i \in \{0, 1, \dots\}$ )
- $x = \{\vec{q}, \vec{p}\}$  where  $\vec{q}$  : generalized coordinates and  $\vec{p}$  : generalized momenta

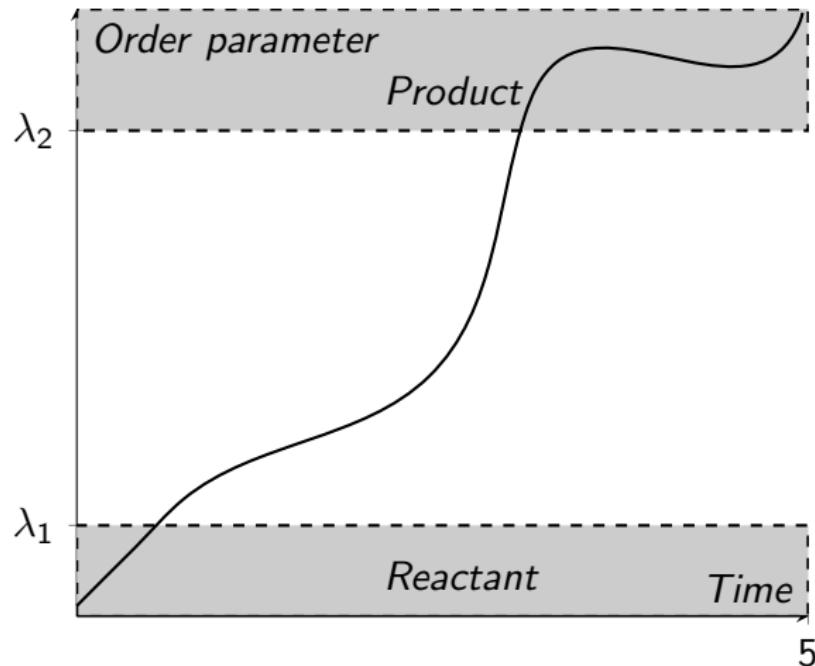
# Monte Carlo in path space: the shooting algorithm

- Start with an initial biased reactive trajectory



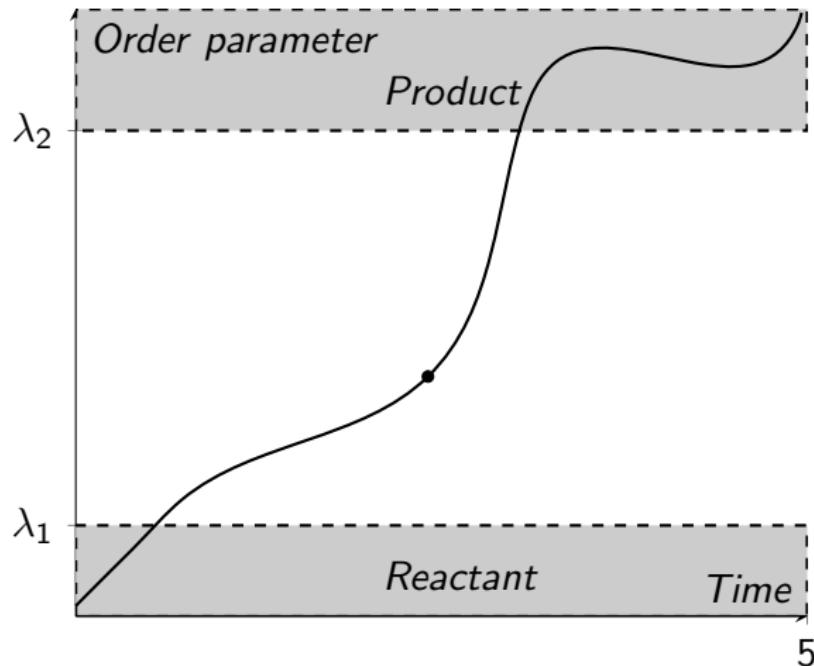
# Monte Carlo in path space: the shooting algorithm

- Start with an initial biased reactive trajectory



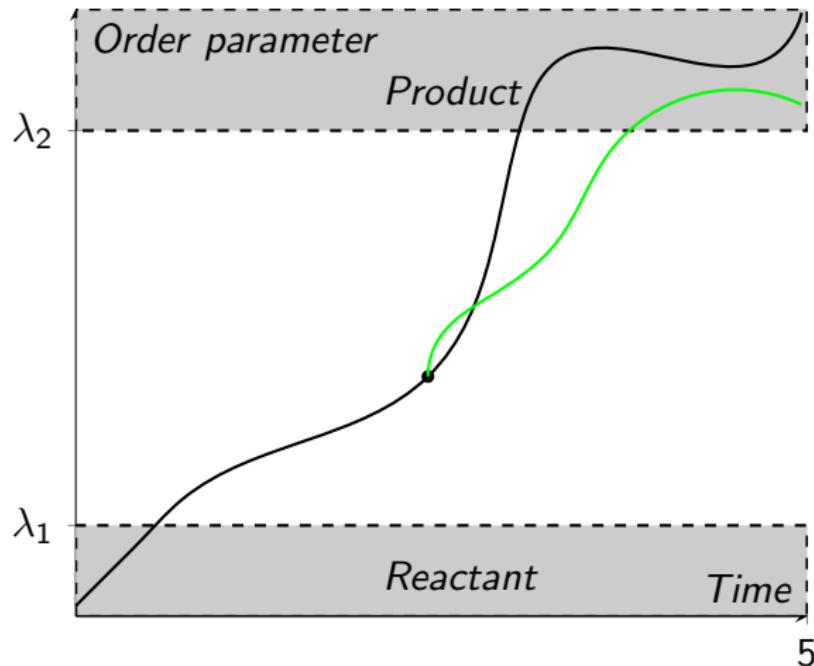
# Monte Carlo in path space: the shooting algorithm

- Start with an initial biased reactive trajectory



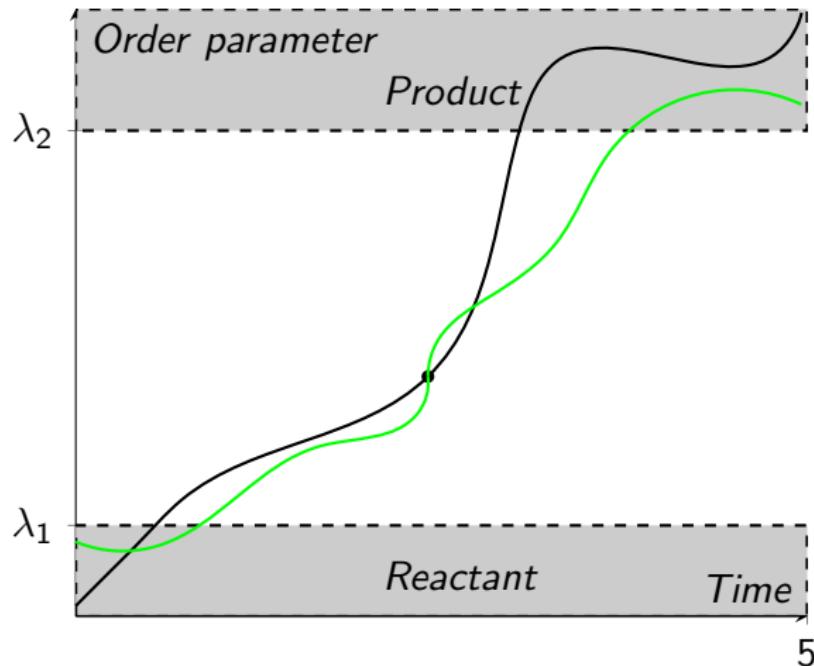
# Monte Carlo in path space: the shooting algorithm

- Start with an initial biased reactive trajectory



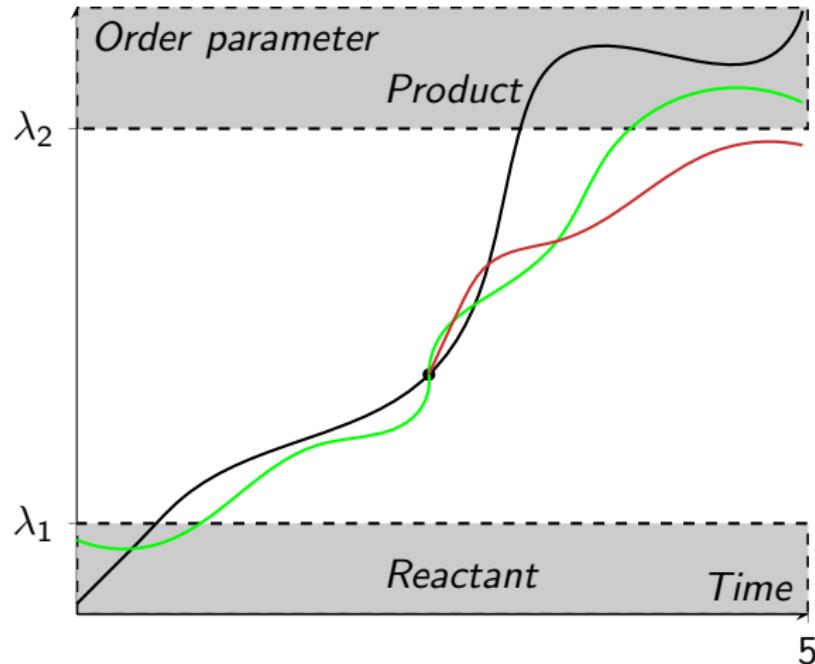
# Monte Carlo in path space: the shooting algorithm

- Start with an initial biased reactive trajectory



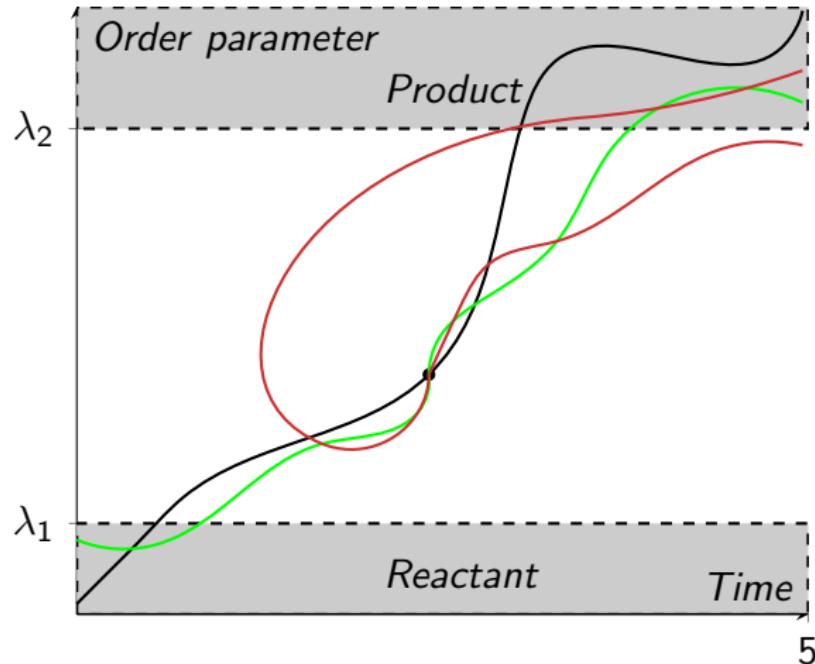
# Monte Carlo in path space: the shooting algorithm

- Start with an initial biased reactive trajectory



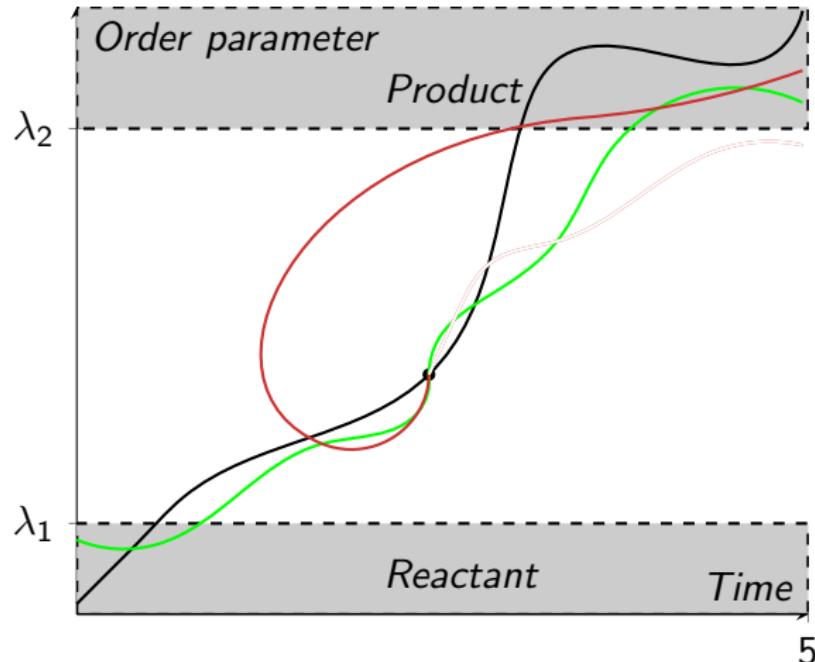
# Monte Carlo in path space: the shooting algorithm

- Start with an initial biased reactive trajectory



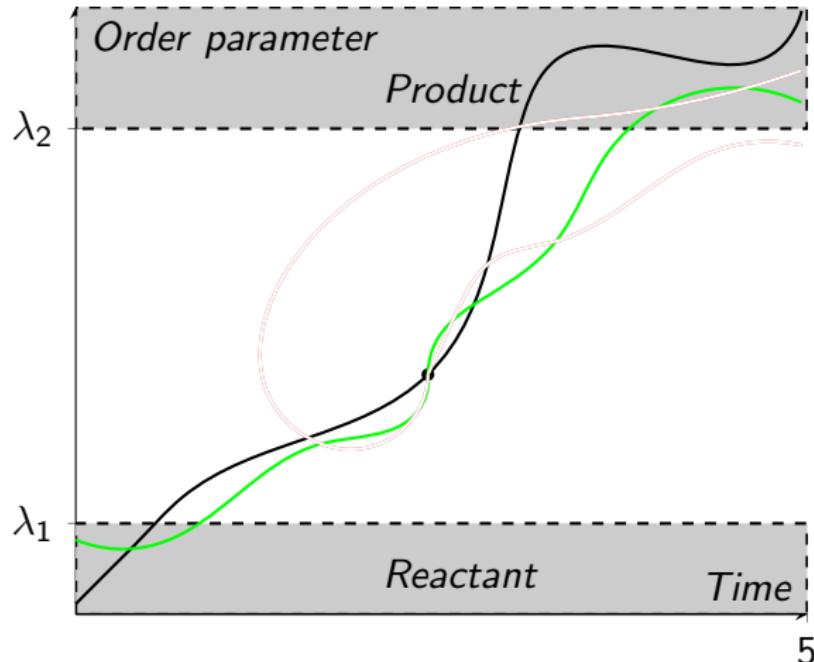
# Monte Carlo in path space: the shooting algorithm

- Start with an initial biased reactive trajectory

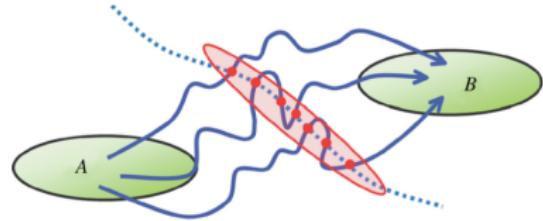


# Monte Carlo in path space: the shooting algorithm

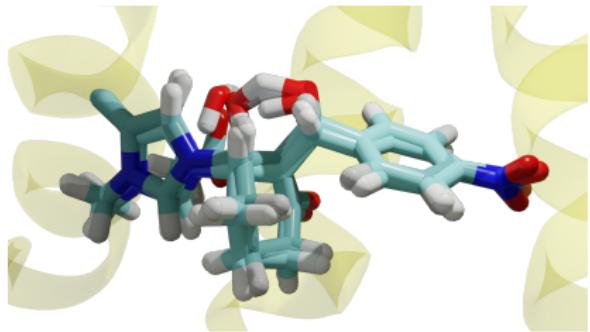
- Start with an initial biased reactive trajectory



# Separatrix and committor analysis<sup>2</sup>



**Figure:** Committor analysis for 1<sup>st</sup> TS from trajectory no. 139 of the old variant



**Figure:** Committor analysis for 2<sup>nd</sup> TS from trajectory no. 139 of the old variant

<sup>2</sup>C. Dellago et al., "Transition path sampling methods", in *Computer simulations in condensed matter systems: from materials to chemical biology volume 1*, edited by M. Ferrario et al. (Springer Berlin Heidelberg, Berlin, Heidelberg, 2006), pp. 349–391. ↗ ↘ ↙ ↘

# Speeding up TPS calculations and analysis

Using neural networks to classify molecular states

## TPS ensemble

Pathways connecting  $\mathcal{A}$  and  $\mathcal{B}$  in phase space have an associated probability distribution of

$$\mathcal{P}_{\mathcal{AB}}[x_T] = h_{\mathcal{A}}(x_0)\mathcal{P}[x(T)]h_{\mathcal{B}}(x_T)/Z_{\mathcal{AB}}(T)$$

where

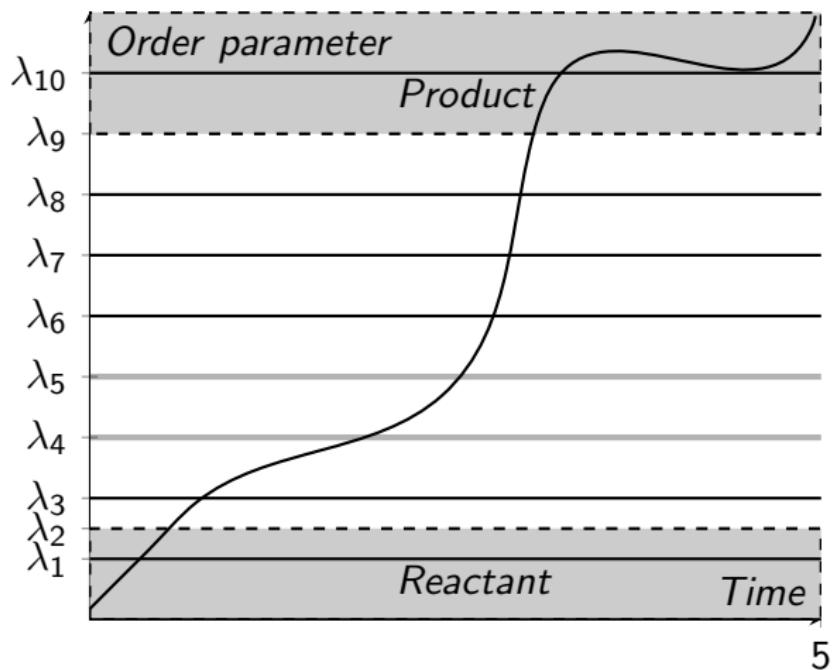
$$h_{\mathcal{A}/\mathcal{B}}(x) = \begin{cases} 1, & \text{if } x \in \mathcal{A}/\mathcal{B} \\ 0, & \text{otherwise} \end{cases}$$

and  $Z_{\mathcal{AB}}(T)$  is the normalization factor for this probability distribution function.

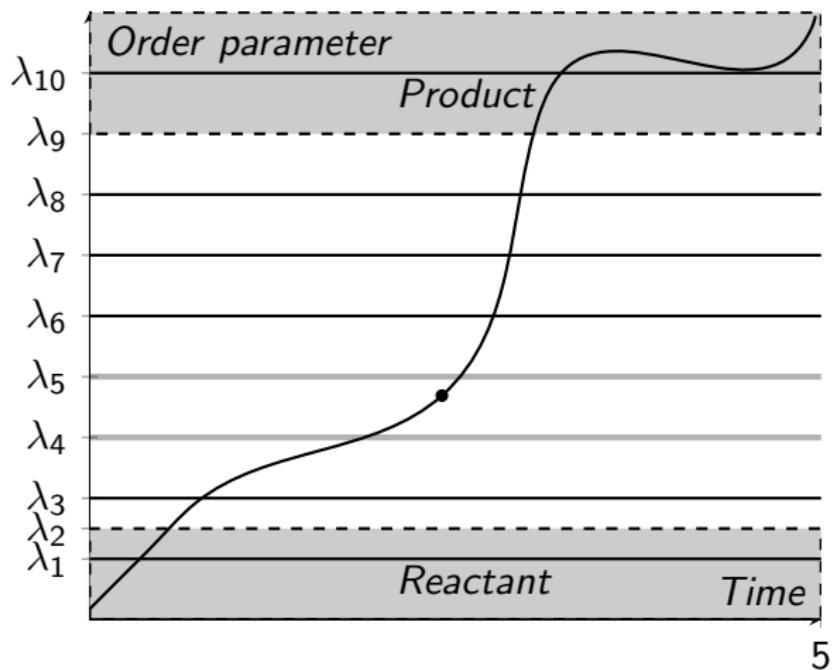
At equilibrium the distribution of pathways is different as transition pathways undersample the regions  $\mathcal{A}$  and  $\mathcal{B}$

## Equilibrium sampling in trajectory space: The algorithm

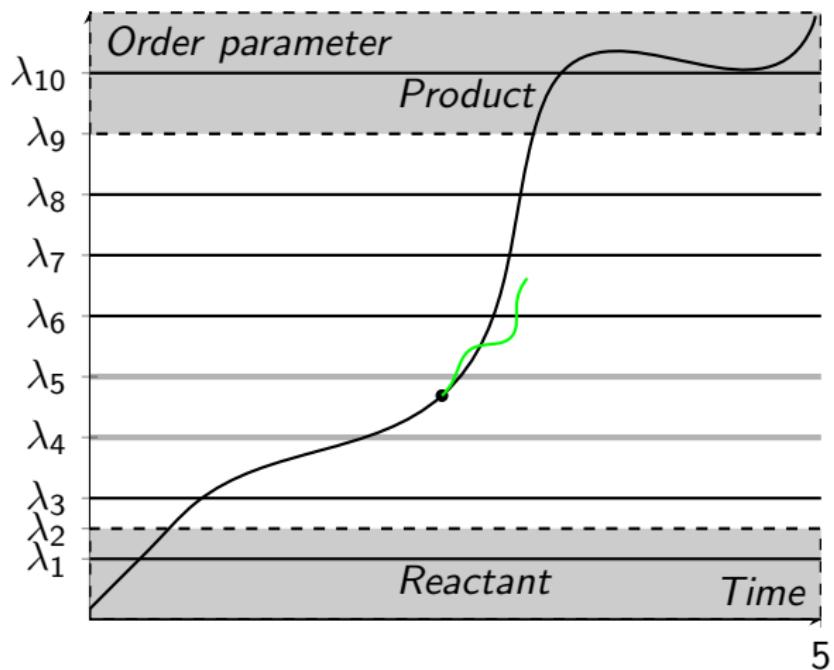
## Equilibrium sampling in trajectory space: The algorithm



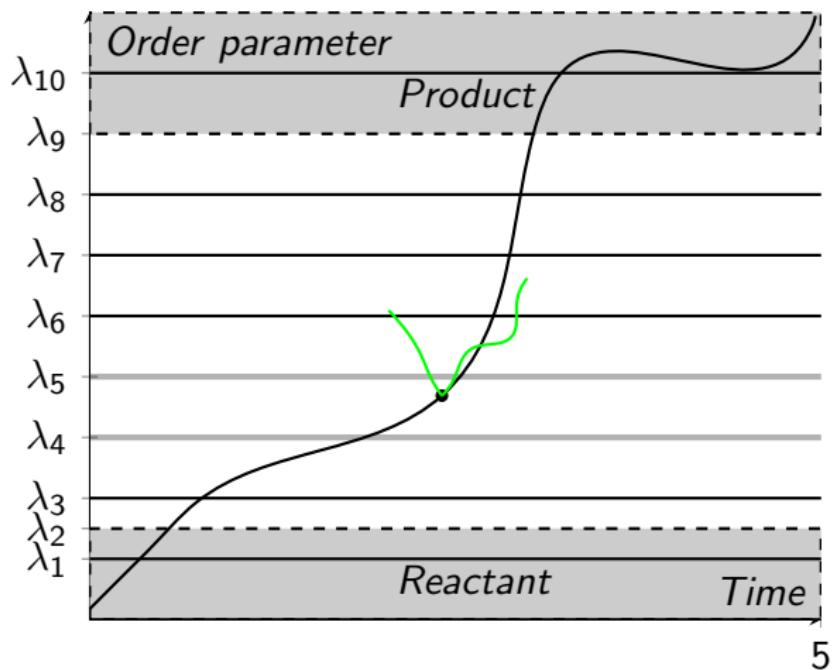
## Equilibrium sampling in trajectory space: The algorithm



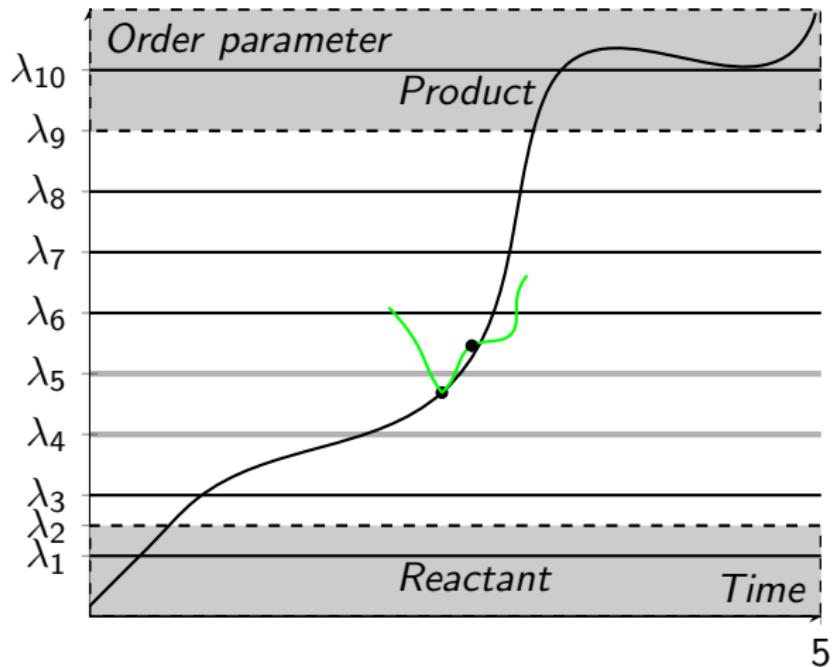
## Equilibrium sampling in trajectory space: The algorithm



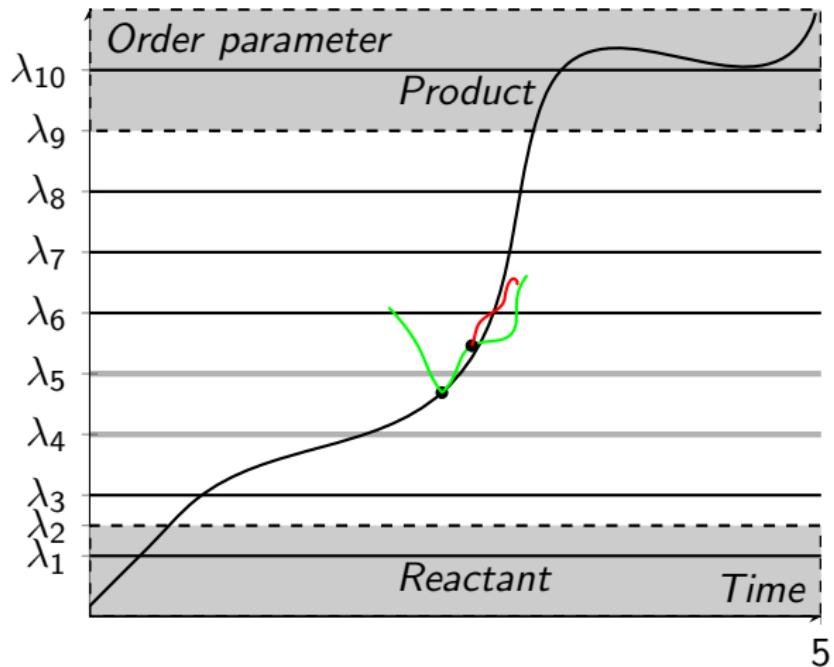
## Equilibrium sampling in trajectory space: The algorithm



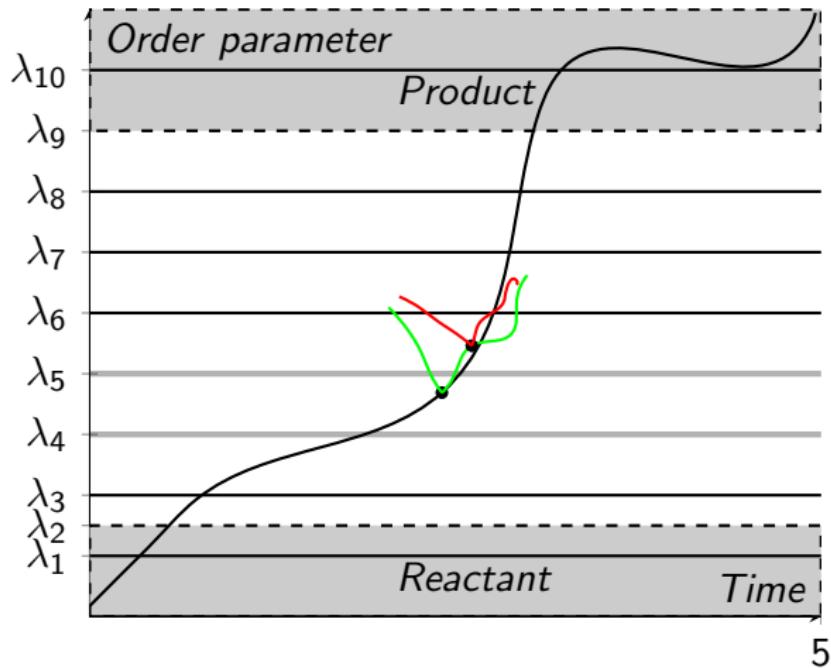
# Equilibrium sampling in trajectory space: The algorithm



# Equilibrium sampling in trajectory space: The algorithm



# Equilibrium sampling in trajectory space: The algorithm



## Free energies from TPS

$$A[\lambda_i] = -k_B T \log(P(\lambda_i)) + \text{constant}$$

The probability of  $P(\lambda_i)$  is defined by

$$P(\lambda_i) = \int d\mathbf{v} \rho(\vec{q}) \delta(\lambda_i - \tilde{\lambda}(\vec{q}))$$

In practice, the calculation of  $P(\lambda_i)$  proceeds through histogram based techniques.

# QM/MM simulations

CHARMM force field consists of intramolecular terms such as

$$\begin{aligned} E_{int} = & \sum_{bonds} k_r(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 \\ & + \sum_{dihedrals} k_\phi(1 + \cos(n\phi - \delta)) \\ & + \sum_{imp. dihed.} k_\psi(\psi - \psi_0)^2 + \sum_{Urey-Bradley} k_{UB}(r_{1,3} - r_{1,3;0}) \end{aligned}$$

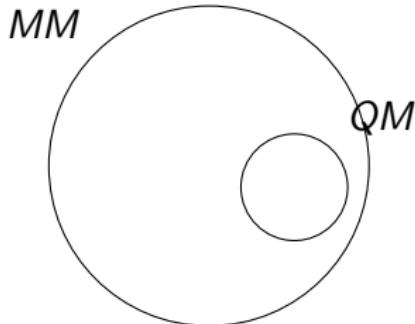
# QM/MM simulations

CHARMM force field consists of intramolecular terms such as

$$\begin{aligned} E_{int} = & \sum_{bonds} k_r(r - r_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 \\ & + \sum_{dihedrals} k_\phi(1 + \cos(n\phi - \delta)) \\ & + \sum_{imp. dihed.} k_\psi(\psi - \psi_0)^2 + \sum_{Urey-Bradley} k_{UB}(r_{1,3} - r_{1,3;0}) \end{aligned}$$

and intermolecular terms such as

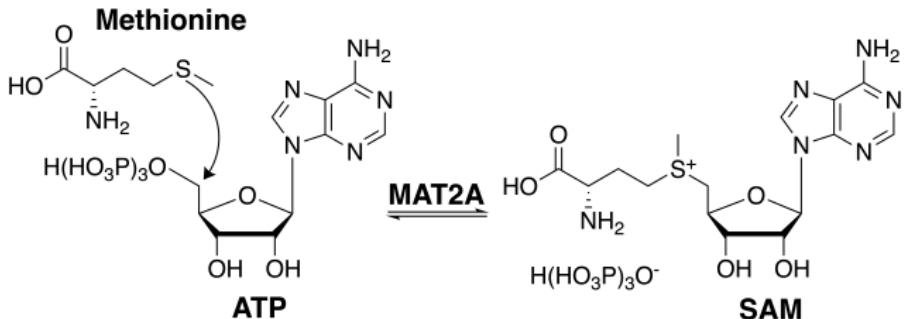
$$\sum_{nonbonded} \left( \frac{q_i q_j}{4\pi\epsilon r_{ij}} - E_{min} \left[ \left( \frac{R_{min}}{r_{ij}} \right)^{12} - \left( \frac{R_{min}}{r_{ij}} \right)^6 \right] \right)$$



QM methods: semiempirical (PM3, AM1, etc.), density functional approximations ( $\mathcal{O}(N^3)$ )

Stitch the QM and MM regions together with generalized hybrid orbital (GHO) scheme

# Human adenosyl methionine enzyme

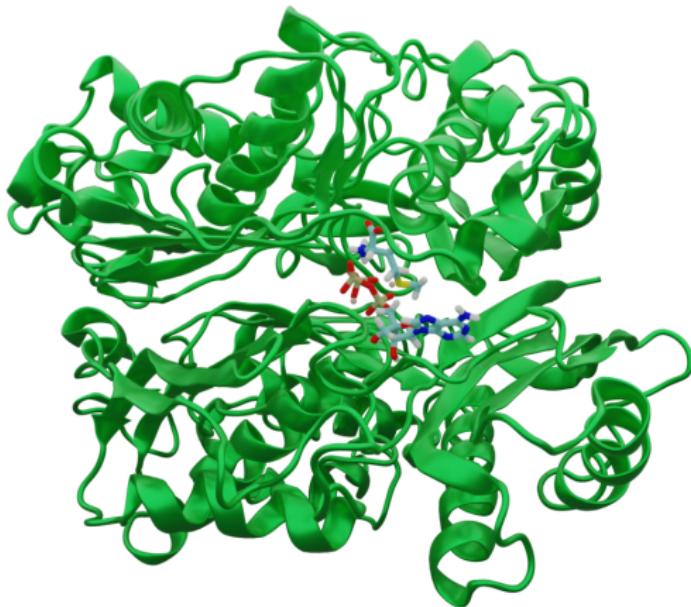


SAM is an essential metabolite which is distributed to almost all body tissues and fluids, is a universal methyl donor, and is of fundamental importance to the metabolism of compounds such as hormones, neurotransmitters, proteins, and nucleic acids.

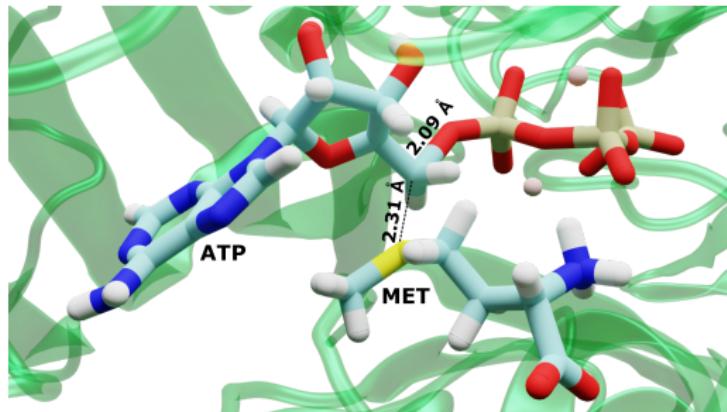
## System preparation

- The energies of both the systems were minimized using 50 steps of the steepest descent method, followed by 2000 steps of the adopted basis Newton-Raphson method where only classical molecular mechanics was used for the dynamics.
- The minimized systems were heated slowly to 300 K for 35 ps beginning with harmonic constraints on all atoms except on the H atoms and the TIP3P water molecules with gradual reduction of the restraint forces.
- 15 ps of equilibration was carried out starting with harmonic restraint forces followed by 20 ps of constraint free equilibration to prepare the systems for TPS simulations.

## System preparation



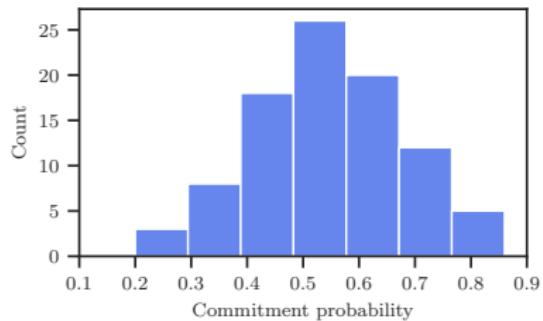
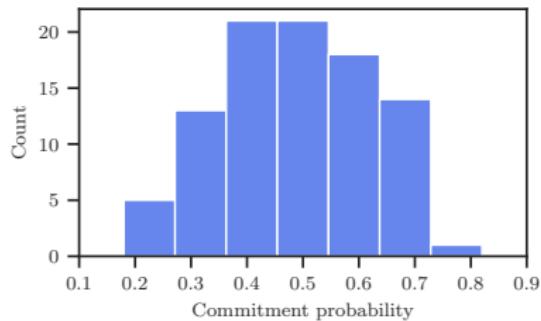
# Committor analysis and transition state



	$d_{SC}$ (Å)	$d_{OC}$ (Å)
Mean	2.35	2.12
Stddev	0.04	0.03

Experimental results:  $d_{SC}$  : 2.03 Å and  $d_{OC}$  : 2.32 Å

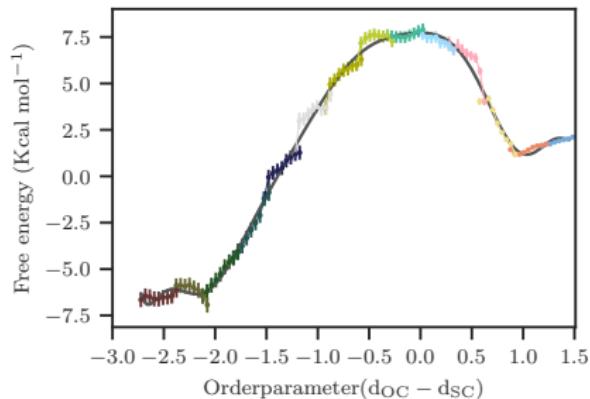
# Committor distribution analysis



**Figure:** Committor distribution analysis for obtaining the reaction coordinate of the MAT2A catalyzed reaction. The figure on the left has the QM region constrained and the figure on the right has the QM region along with the Gln113, Ser114, Arg249 and Arg264 residues constrained.

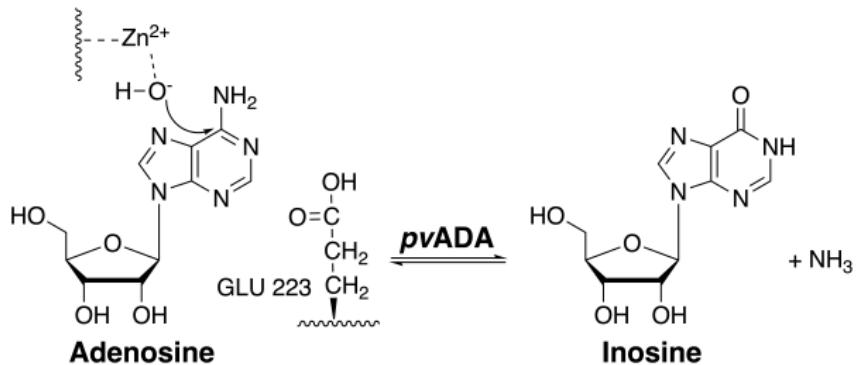
# Free energies

- 20 overlapping windows, 2500 trajectories of 20 fs length in each window



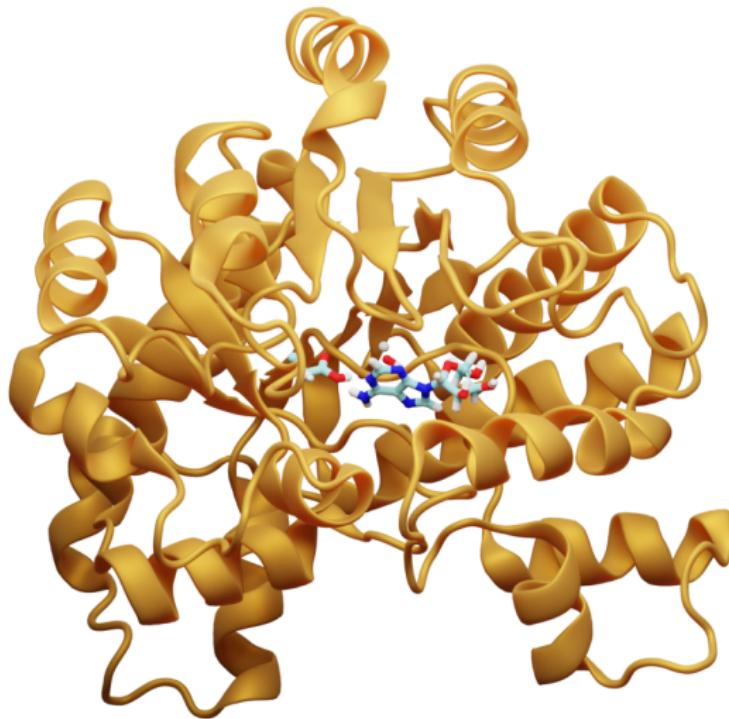
Free energy barrier ( <i>kcal mol<sup>-1</sup></i> )	
Experimental	17.27
Calculated	16 <sup>3</sup>

# *Plasmodium vivax* adenosine deaminase



*Plasmodium vivax* is a parasite that is responsible for the largest number of cases of malaria globally

# Equilibrium structure



# Free energies and comparison to experiments

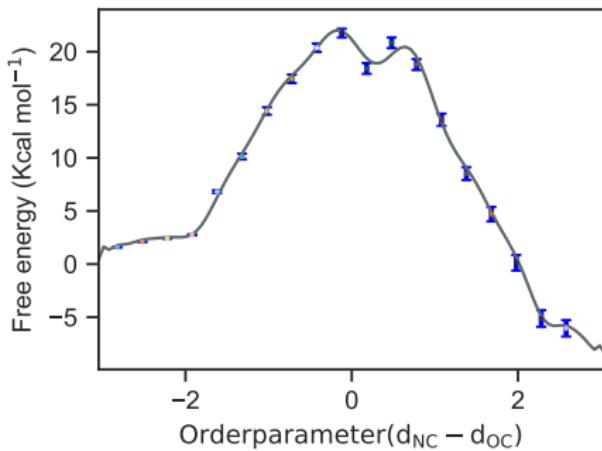
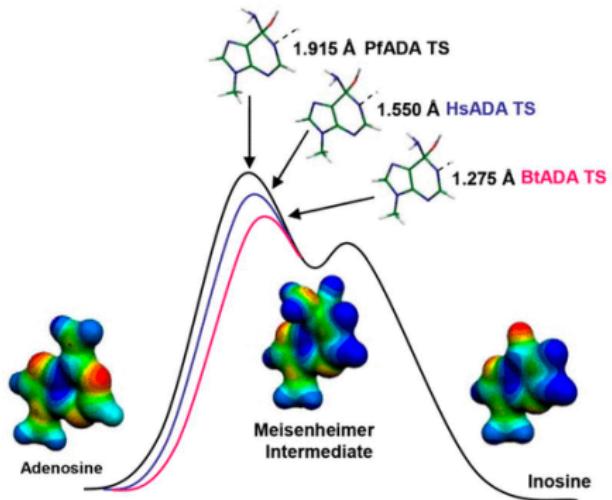


Figure: Calculated free energy barrier: 21 kcal mol<sup>-1</sup>

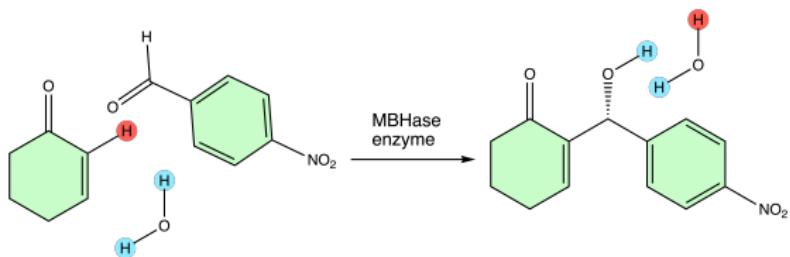
Qualitatively predicts the free energy profile as calculated from experiments for *plasmodium falciparum* adenosine deaminase.

# Project 3

## Computational design of artificial enzymes

# Directed Evolution

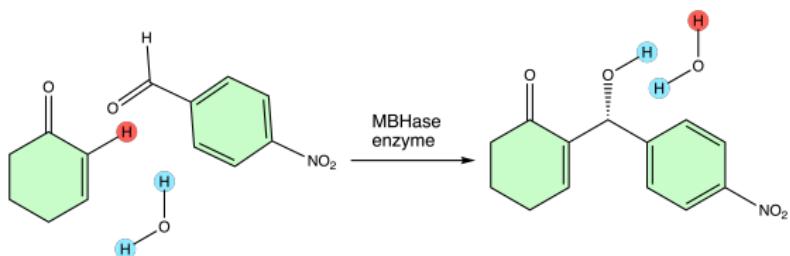
Design enzymes that catalyze the Morita-Baylis-Hillman reaction



- Transfer of two protons

# Directed Evolution

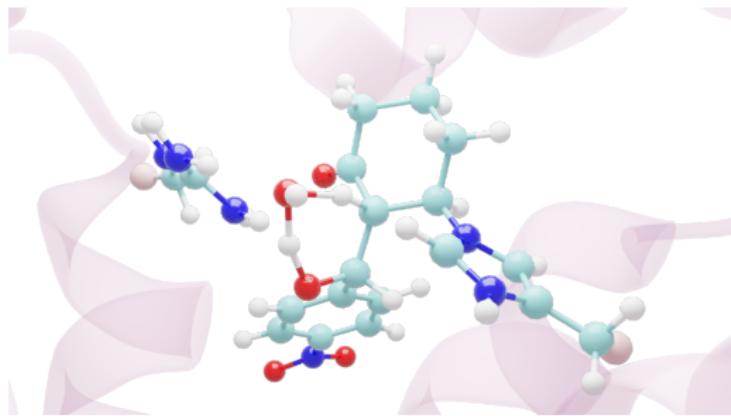
Design enzymes that catalyze the Morita-Baylis-Hillman reaction



- Transfer of two protons
- Does evolution select for quantum mechanical tunneling?

## BH32 Variant

This is the most evolved version of the previous generation



# Conclusions

- TPS based free energies are implemented

# Conclusions

- TPS based free energies are implemented
- Free energy calculations can be performed without the inclusion of biasing potentials and the knowledge of reaction coordinates

# Conclusions

- TPS based free energies are implemented
- Free energy calculations can be performed without the inclusion of biasing potentials and the knowledge of reaction coordinates
- Free energy implementation was tested for the first time on enzyme catalyzed reactions, human MAT2A enzyme and plasmodium falciparum adenosine deaminase

# Conclusions

- TPS based free energies are implemented
- Free energy calculations can be performed without the inclusion of biasing potentials and the knowledge of reaction coordinates
- Free energy implementation was tested for the first time on enzyme catalyzed reactions, human MAT2A enzyme and plasmodium falciparum adenosine deaminase
- The free energies calculated were qualitatively in good agreement with experimental measurements

# Acknowledgements

## Schwartz group

- Prof. Steven D. Schwartz
- Dr. Dimitri Antoniou
- Dr. Anthony Boldau
- Dr. Allison Smith
- Ananya Chakraborty
- Bai Hei
- Clara Frost
- Dr. Elango Munusamy

# Acknowledgements

## Schwartz group

- Prof. Steven D. Schwartz
- Dr. Dimitri Antoniou
- Dr. Anthony Boldau
- Dr. Allison Smith
- Ananya Chakraborty
- Bai Hei
- Clara Frost
- Dr. Elango Munusamy

