



Enrico Trizio, Peilin Kang & Michele Parrinello

The committor wonderland

Computing the committor function for enhanced sampling simulations of transition states

Talk overview

nature computational science

Article <https://doi.org/10.1038/s43588-024-00645-0>

Computing the committor with the committor to study the transition state ensemble

Received: 12 January 2024 | Peilin Kang & Enrico Trizio & Michele Parrinello

Accepted: 14 May 2024 | Published online: 5 June 2024 | Check for updates

The study of the kinetic bottlenecks that hinder the rare transitions between long-lived metastable states is a major challenge in atomistic simulations. Here we propose a method to explore the transition state ensemble, which is the set of configurations that the system visits as it transitions from one metastable basin to another. We base our method on the committor function and the variational principle that it obeys. We find its minimum through a self-consistent procedure that starts from information limited to the initial and final states. Right from the start, our procedure allows the sampling of very many transition state configurations. With the help of the variational principle, we perform a detailed analysis of the transition state ensemble, ranking quantitatively the degrees of freedom mostly involved in the transition and enabling a systematic approach for the interpretation of simulation results and the construction of efficient physics-informed collective variables.

Many important physico-chemical transformations such as crystalization, chemical reactions and protein folding take place on a time scale that is not directly accessible to microscopic simulations. These processes are referred to as rare events and are hindered by kinetic bottlenecks that occur whenever the metastable basins are separated. Such kinetic bottlenecks are often overcome when the metastable basins are separated by a high-free-energy set of configurations that we refer to as the transition state ensemble (TSE). Finding and analyzing this region is of utmost theoretical and practical importance. For example, it comes to chemical reactions, as it provides precious information about reaction mechanisms, rates, or when dealing with proteins, it can provide insights into their dynamics.

In the vast rare-event literature^{1–3}, the determination of the TSE is usually the culmination of the simulation. In this Article, instead, we propose a method to explore the transition state ensemble directly. The theoretical tool that allows this change of perspective is the committor function $q(\mathbf{x})$, introduced by Kolmogorov⁴. Given two metastable states, A and B, $q(\mathbf{x})$ gives the probability that, starting from configuration \mathbf{x} , the system ends in B without having first passed to A, and is, as a consequence, the complementarily needed to identify the TSE as the set of configurations for which $q(\mathbf{x}) = 1$ (refs. 13–24). Once $q(\mathbf{x})$ is known, properties such as the transition rate between A and B, the density of reactive trajectories or the reactive fluxes can be calculated. Unfortunately, although the committor $q(\mathbf{x})$ is challenging to find in the transition state sampling literature, it has been mostly estimated for curated sets of points via committor analysis. However, such an approach can be computationally expensive and often dependent on the choice of the criteria used to identify the TSE. Moreover, either a trajectory is assigned to either basin A or B (ref. 20).

An alternative to such an approach is to solve the so-called “Kolmogorov equation”, where $q(\mathbf{x})$ can be determined as the solution of a partial differential equation⁵ that obeys the boundary conditions $q(\mathbf{x}_A) = 0$ and $q(\mathbf{x}_B) = 1$, where \mathbf{x}_A and \mathbf{x}_B denote two configurations belonging to the two basins, respectively. Unfortunately, solving such multidimensional equations for real systems poses insurmountable problems. However, under the hypothesis of overdamped dynamics, the solution of the Kolmogorov equation can also be obtained by following a variational approach that amounts to minimizing functional

Atomistic Simulations, Italian Institute of Technology, Genova, Italy. ²Department of Materials Science, Università di Milano-Bicocca, Milano, Italy.
✉ e-mail: michele.parrinello@iit.it

Nature Computational Science | Volume 4 | June 2024 | 451–460

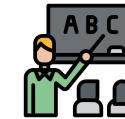
451

Kang, Trizio, and Parrinello, Nat. Comp. Sci. 2024

Why do we care about transition states?



Transition state in conventional enhanced sampling



Obtaining the committor and the transition state ensemble



Examples and analysis

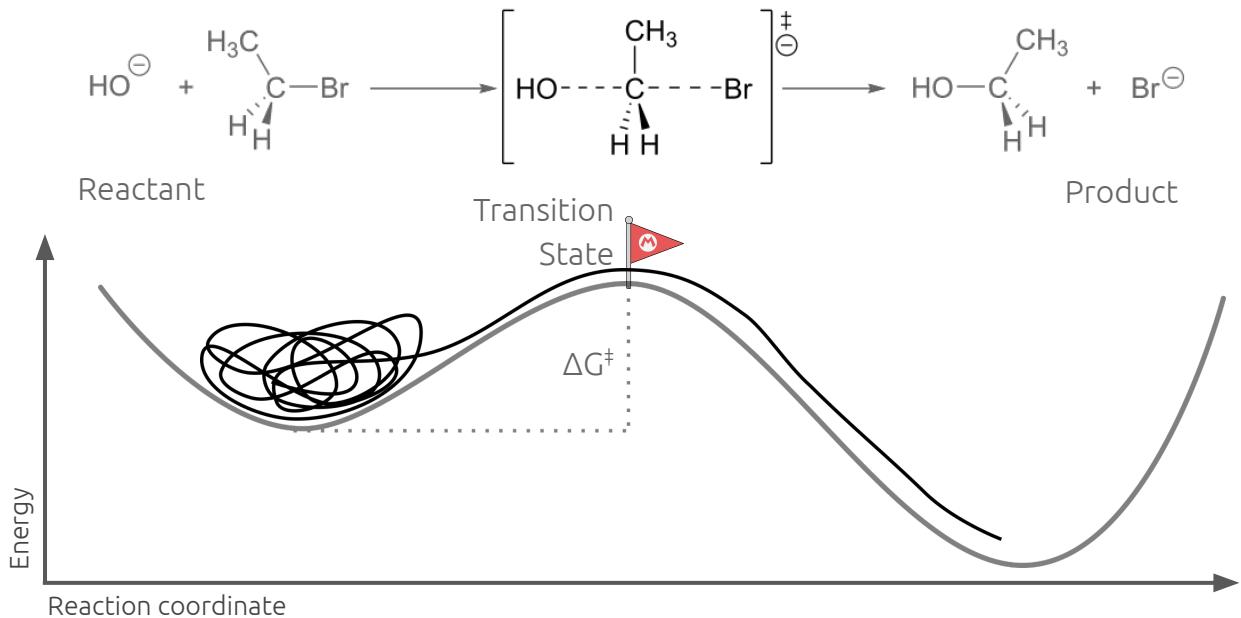


The transition state, why do we like it so much?

The holy grail in chemistry

Transition state (wikipedia)

In chemistry, the transition state (TS) of a chemical reaction is a particular configuration along the reaction coordinate. It is defined as the state corresponding to the **highest potential energy** along this reaction coordinate.



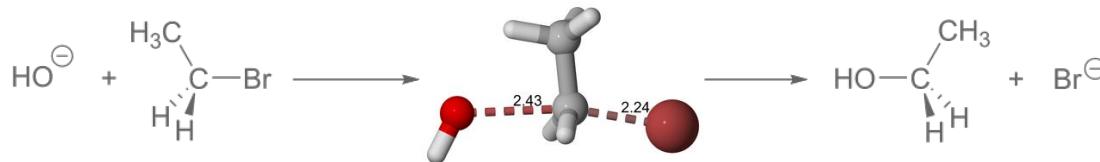
The TS is an elusive entity in simulations as it sits on the barrier that lead to the timescale problem

The transition state, why do we like it so much?

Reaction mechanisms and rates

Reaction mechanisms

E.g. SN₁ or SN₂?



Reaction rates

E.g. Fast or slow?

$$k \propto e^{-\Delta G^\ddagger}$$



Smaller ΔG^\ddagger = Faster

Larger ΔG^\ddagger = Slower



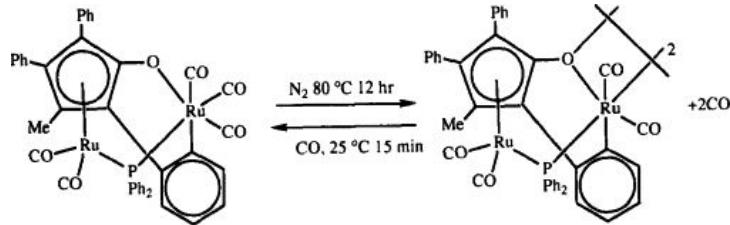
Let me enjoy this until it lasts :)

The transition state ensemble

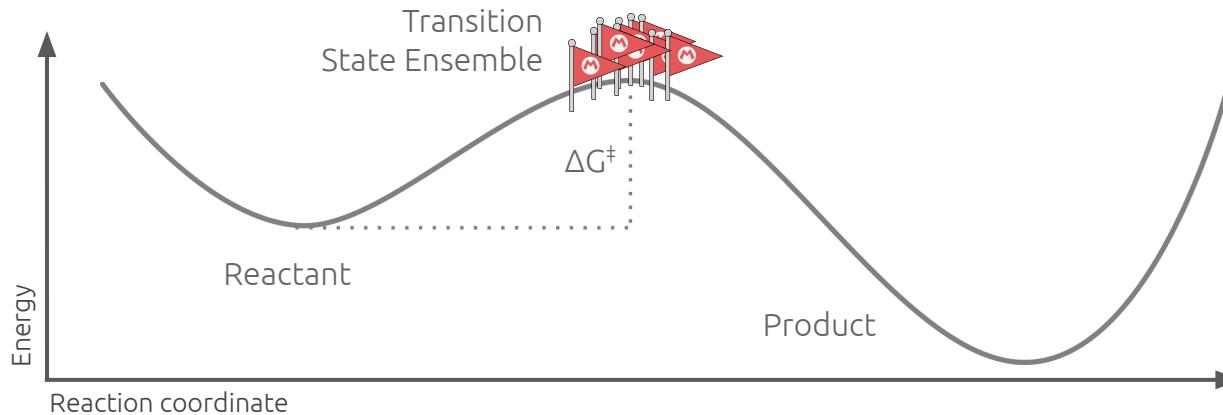
From a state to an ensemble



Complex chemical reaction



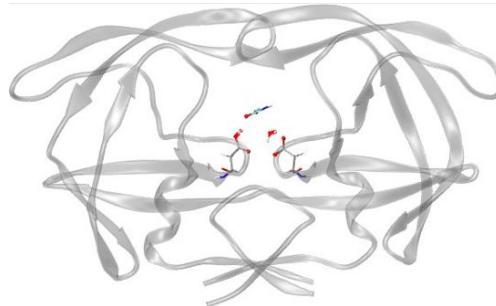
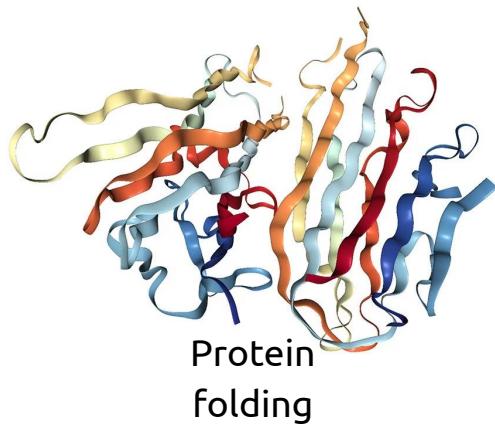
When the systems become more complex, it is wiser to refer to a transition state ensemble



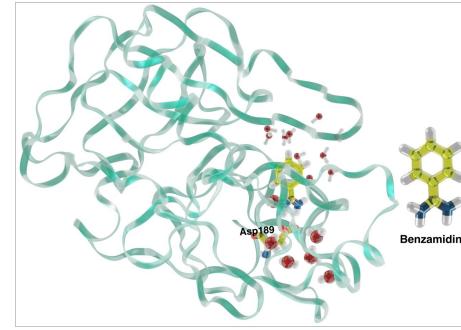
Relevant not only in chemistry

Some quick examples from biophysics

And it can easily get worse and worse..



Enzymatic
reactions

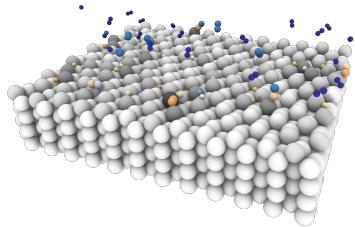


Ligand binding
Drug design

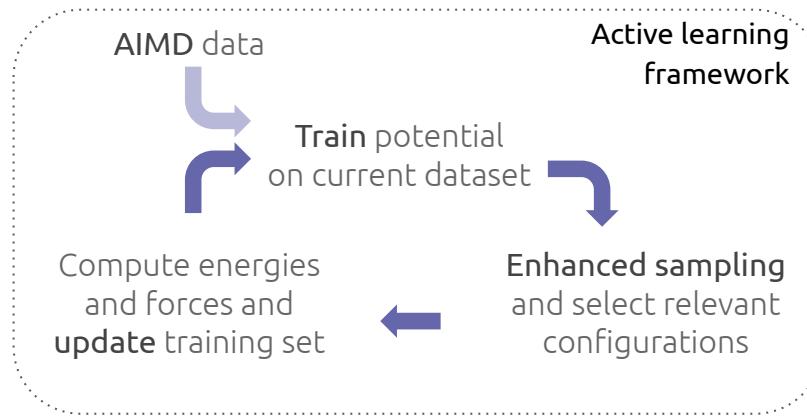
Crucial in training machine learning potentials

Some examples from our group

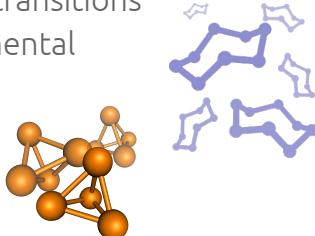
Haber-Bosch-related catalytic reactions for NH₃ production and decomposition



Bonati *et al.*, PNAS, 2023
Tripathi *et al.*, ACS Cat. 2024
Perego *et al.*, ChemRxiv. 2024
Perego and Bonati, ChemRxiv. 2024

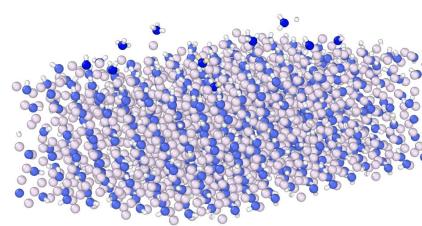


Phase transitions in elemental liquids



Yang, Karmakar, Parrinello, PRL, 2021
Yang, Trizio, Parrinello, Chem. Sci. 2024

Alternative catalysts for NH₃ production and decomposition



Yang, Raucci, Parrinello, Nat. Cat., 2023
Mambretti *et al.*, ACS Cat. 2024
Gardini, Raucci, Parrinello, ChemRxiv

The committor function

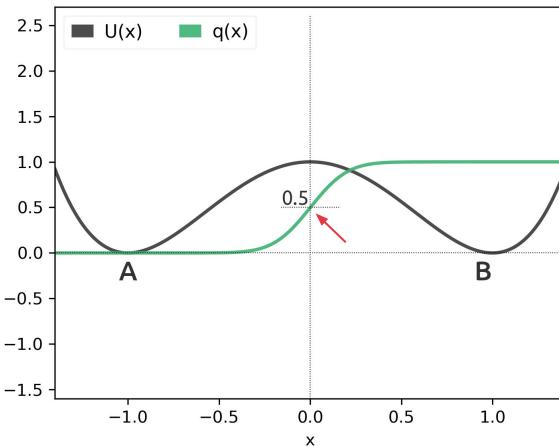
How can we formalize the concept of transition state ensemble?

One popular way to mathematically formalize the concept of TSE on a potential surface $U(x)$ with two metastable states A and B is the **committor function** $q(x)$



Über die analytischen Methoden in der
Wahrscheinlichkeitsrechnung.

Von
A. Kolmogoroff in Moskau.



For a given x $q(x)$ is the **probability**
that starting from x we end in B
before passing by A

Conventionally, the TSE is identified as
the region where $q \sim 0.5$

TSE \longleftrightarrow $q \sim 0.5$

The committor function

Accessing the committor is a real challenge

Experimental approach

Committor analysis: many trajectories are started from a curated set of points to estimate their committor value

- Computationally expensive
- Often dependent on the criteria used to define whether a trajectory is committed to A or B

Theoretical approach (Kolmogorov strikes again)

Under the hypothesis of overdamped Langevin dynamics (but it can be easily extended)



Nasty differential problem

$$\begin{cases} \nabla U \cdot \nabla q - \beta^{-1} \Delta q = 0 & \mathbf{x} \in \Omega \setminus (A \cup B) \\ q(\mathbf{x}) = 0 & \mathbf{x} \in A \\ q(\mathbf{x}) = 1 & \mathbf{x} \in B \end{cases}$$

Impossible

Unpleasant variational problem

$$\min_q K : K = \frac{1}{Z} \int_{\Omega \setminus (A \cup B)} |\nabla q(\mathbf{x})|^2 e^{-\beta U(\mathbf{x})} d\mathbf{x} + \text{b.c.}$$

Difficult

A closer look at the variational approach

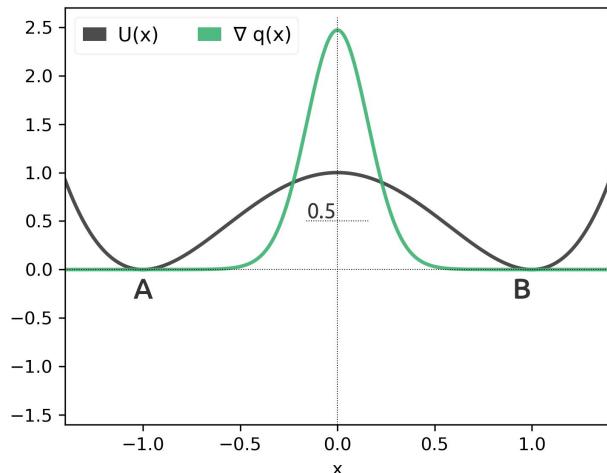
Why is it still challenging?

Variational problem for the committor

$$\min_q K : K = \frac{1}{Z} \int_{\Omega \setminus (A \cup B)} |\nabla q(\mathbf{x})|^2 e^{-\beta U(\mathbf{x})} d\mathbf{x} = \left\langle |\nabla q(\mathbf{x})|^2 \right\rangle_{U(\mathbf{x})}$$

$$q(\mathbf{x}_A) = 0 \quad q(\mathbf{x}_B) = 1$$

Variational functional



Boundary conditions

The gradients ∇q are **strongly localized** on the TS...

...but do we remember the good old **timescale problem** of MD simulations?

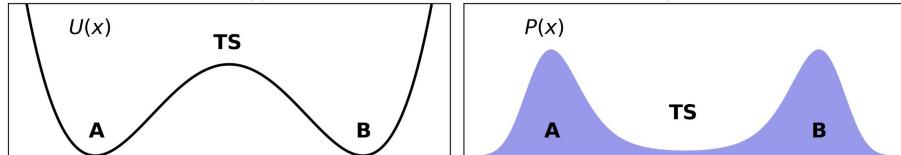
Most of the contribution to the variational problem come from the region that is **hard to sample!**

Transition states and molecular dynamics

Like all nice things it's difficult to obtain

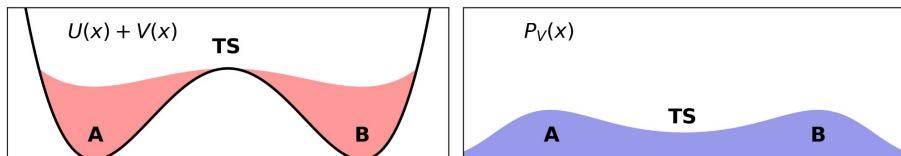
Standard MD

Good luck with that



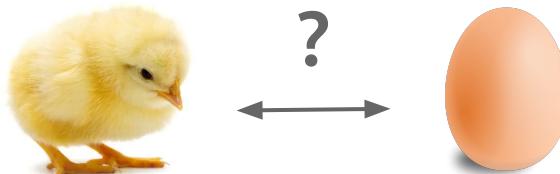
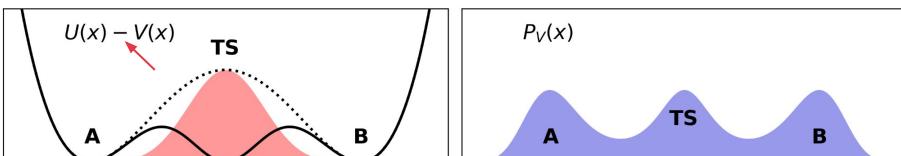
How about enhanced sampling? MetaD, OPES..

Even when basins are filled the TS is **still** unfavoured limiting its sampling



What would be nice to have?

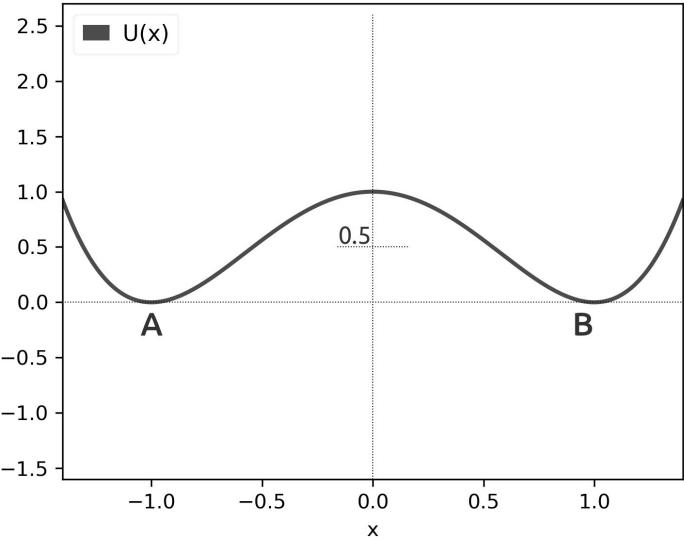
A bias that focuses sampling **on** the TS region by design!



To build the bias we need to know the TS
To know the TS we need to build the bias

A new approach to enhanced sampling

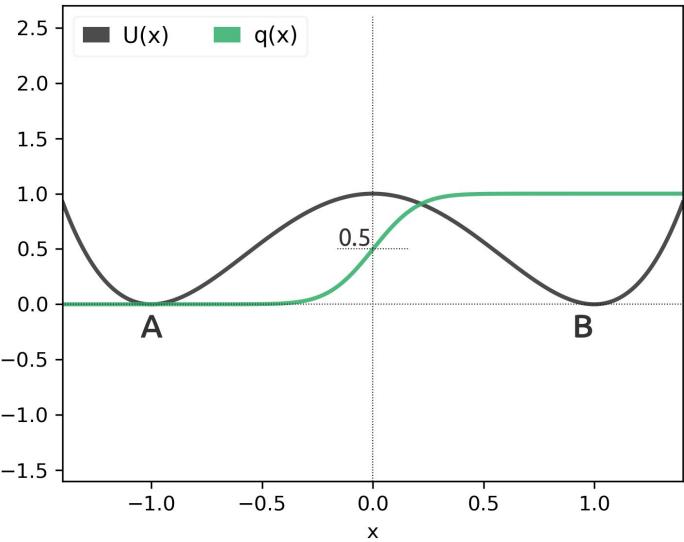
A committer-inspired solution



Let's go back to our beloved
double-well potential

A new approach to enhanced sampling

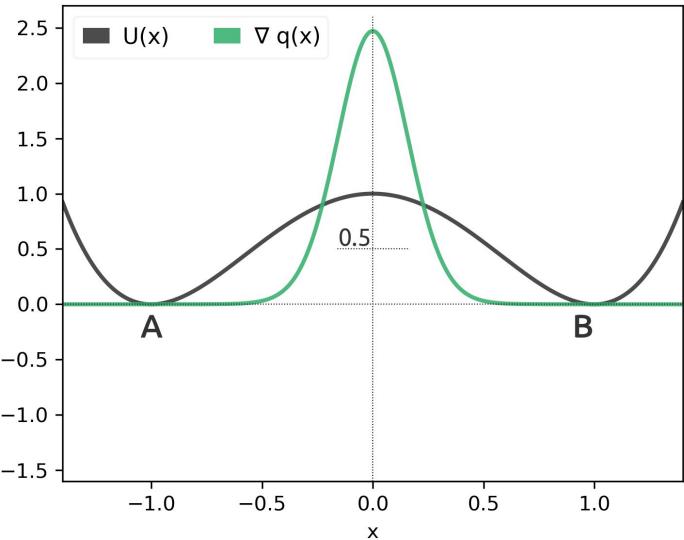
A committer-inspired solution



Let's assume for a second that we know the committer for our system

A new approach to enhanced sampling

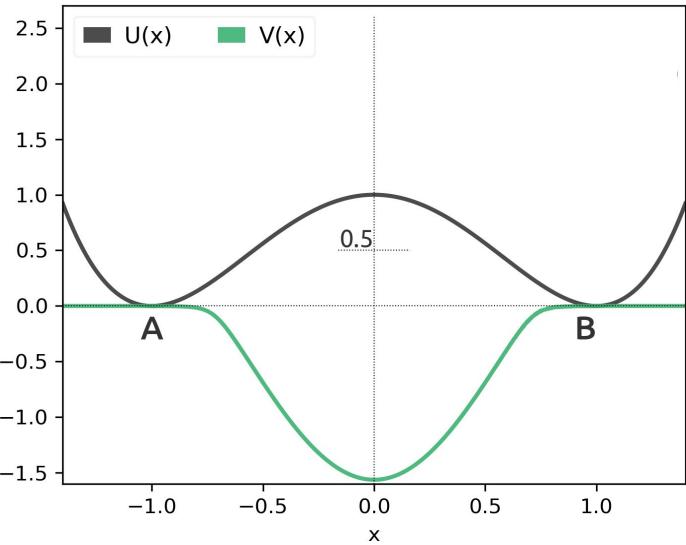
A committor-inspired solution



The gradients of the committor ∇q are **strongly localized** on the TS by design...

A new approach to enhanced sampling

A committer-inspired solution



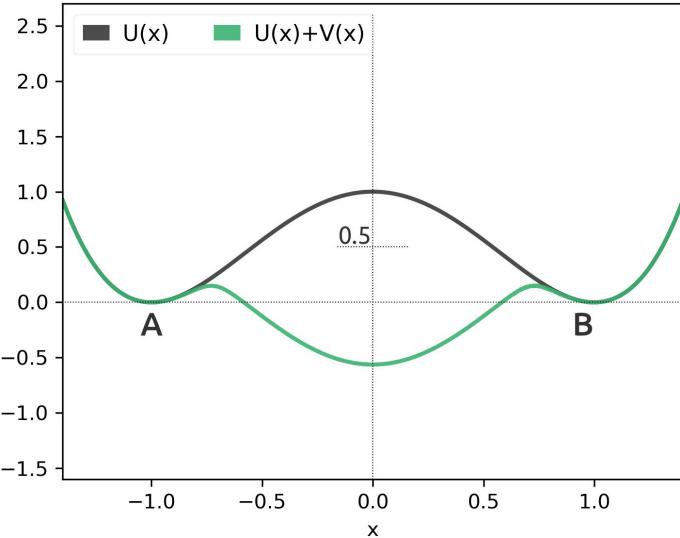
Kolmogorov bias

We could write our TS-oriented bias
as a function of ∇q !

$$V_K(x) = -\frac{\lambda}{\beta} \log(|\nabla q(x)|^2 + \epsilon)$$

A new approach to enhanced sampling

A committor-inspired solution

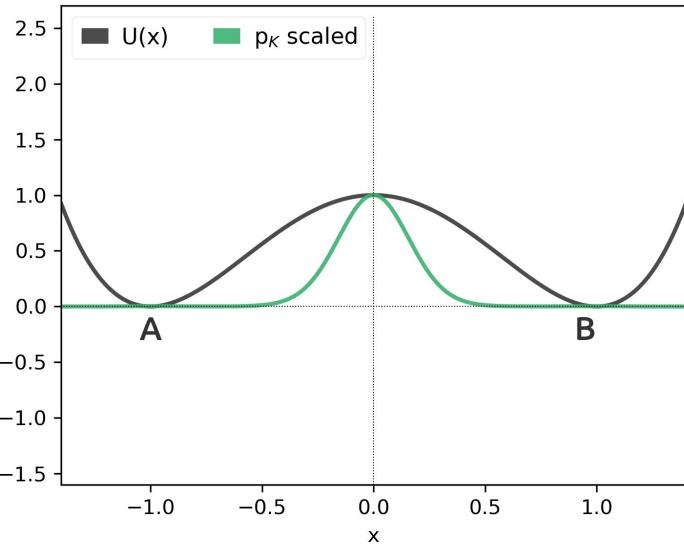


By applying our bias we can turn the TS-maximum into a minimum that can be **effortlessly sampled!**

$$V_K(\mathbf{x}) = -\frac{\lambda}{\beta} \log(|\nabla q(\mathbf{x})|^2 + \epsilon)$$

A new approach to enhanced sampling

A committor-inspired solution



Kolmogorov distribution

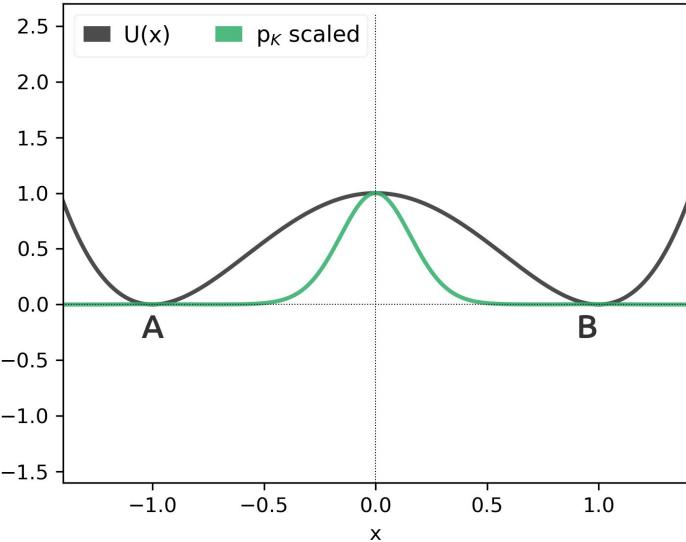
This way we can also **extend the conventional TSE definition** to also include the probability that the included configurations are **actually sampled**

$$p_{\mathcal{K}}(\mathbf{x}) = \frac{e^{-\beta U_{\mathcal{K}}(\mathbf{x})}}{Z_{\mathcal{K}}} \quad \text{with} \quad U_{\mathcal{K}}(\mathbf{x}) = U(\mathbf{x}) + V_{\mathcal{K}}(\mathbf{x})$$
$$Z_{\mathcal{K}} = \int d\mathbf{x} e^{-\beta U_{\mathcal{K}}(\mathbf{x})}$$

(This will be clearer in the following non-1D examples)

A new approach to enhanced sampling

A committor-inspired solution



The Kolmogorov functional can be evaluated on the biased potential!
This way we focus on the TSE by design

$$u_{\mathcal{K}}(\mathbf{x}) = U(\mathbf{x}) + V_{\mathcal{K}}(\mathbf{x}) \quad \xrightarrow{\mathcal{K}[q(\mathbf{x})] = \frac{Z_{\mathcal{K}}}{Z_B}} \quad \mathcal{K}[q(\mathbf{x})] = \left\langle \frac{1}{|\nabla q(\mathbf{x})|^2} \right\rangle_{u_{\mathcal{K}}}^{-1}$$
$$Z_{\mathcal{K}} = \int d\mathbf{x} e^{-\beta u_{\mathcal{K}}(\mathbf{x})}$$

A new approach to enhanced sampling

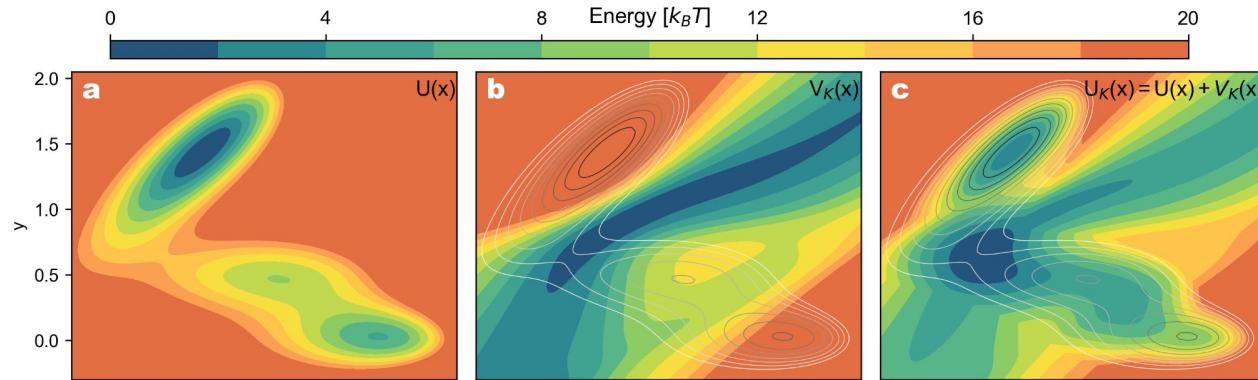
Short summary adding one dimension

Given the committor $q(x)$ for a potential $U(x)$
We can build a **static** bias

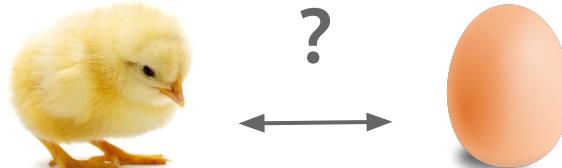
$$V_K(x) = -\frac{\lambda}{\beta} \log(|\nabla q(x)|^2 + \epsilon)$$

That focuses the sampling on the TS by design

$$p_K(x) = \frac{e^{-\beta U_K(x)}}{Z_K}$$



But, how can we actually get the committor function?

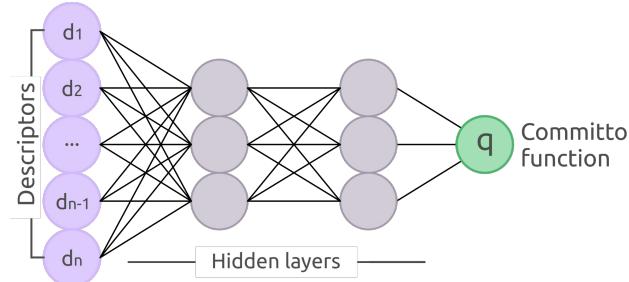


To build the bias we need obtain the committor
To obtain the committor we need to build the bias

Machine learning the committor function

The machinery

We parametrize the committor function as a **neural network** taking as inputs a set of **physical descriptors**. This also gives us smoothness and automatic derivatives.



In the objective function for the optimization we enforce the boundary conditions and the variational principle

$$\longrightarrow L = L_v + \alpha L_b$$

$$L_v = \frac{1}{N^n} \sum_i^{N^n} w_i |\nabla_u q(x_i)|^2$$

Variational loss term
Whole dataset

$$L_b = \frac{1}{N_A} \sum_{i \in A}^{N_A} (q(x_i))^2 + \frac{1}{N_B} \sum_{i \in B}^{N_B} (q(x_i) - 1)^2$$

Boundary loss term
Labeled dataset

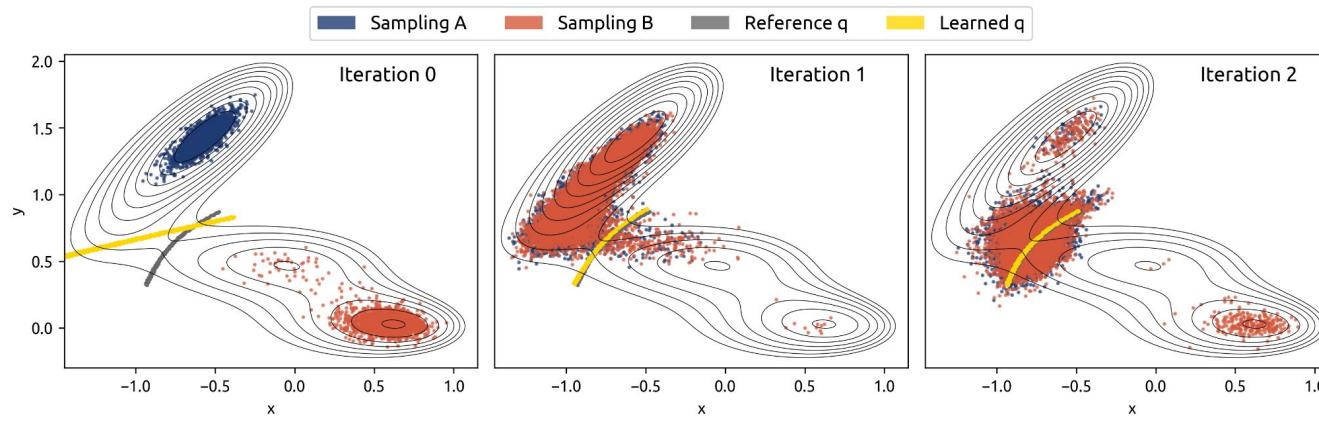
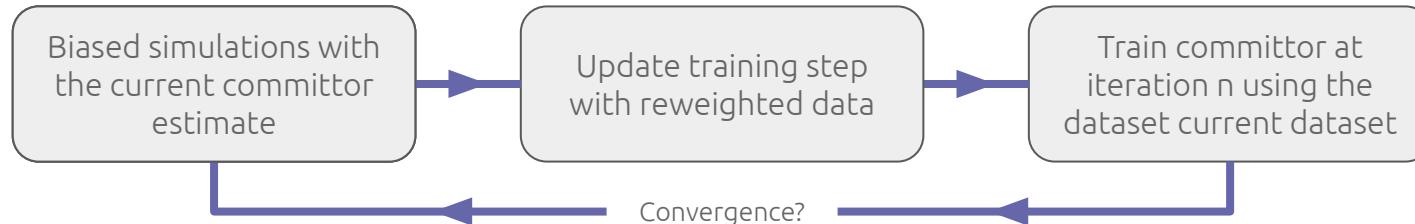
All the code is **open-source** and available through the **mlcolvar library**



Machine learning the committor function

Making the approach iterative and self-consistent

We still need a practical way to combine together the different elements of the story

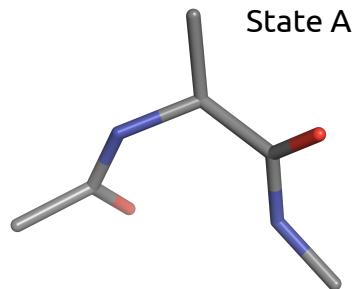


Examples, finally

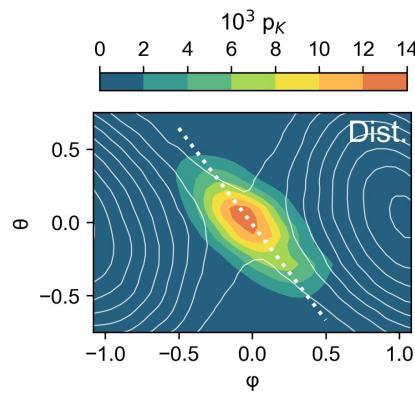
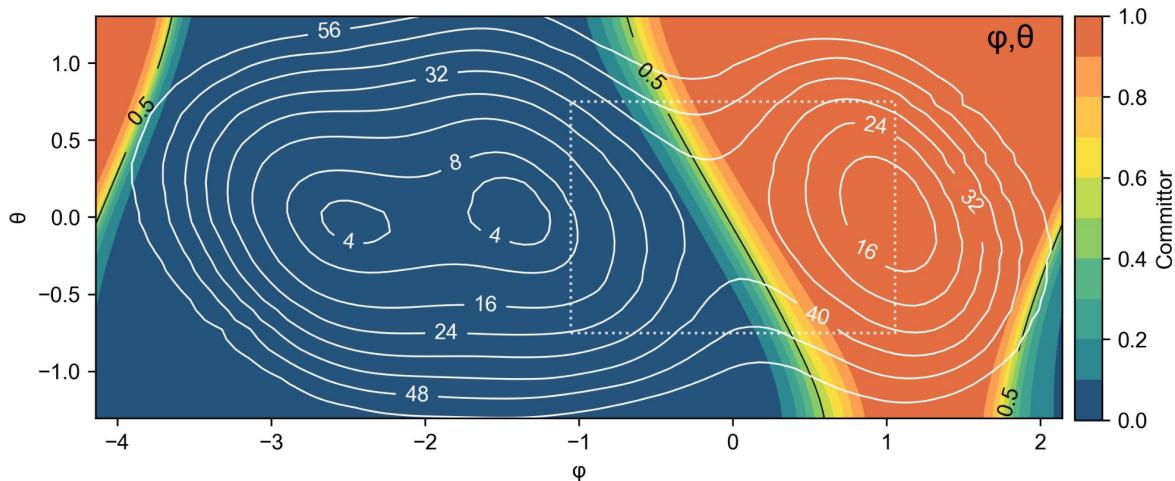
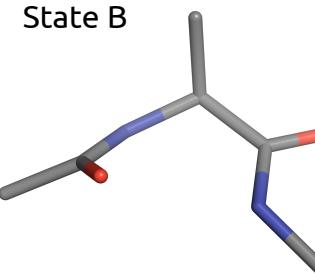
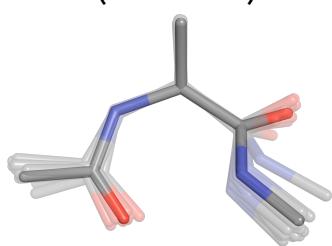
And hints on how to get the best out of our new TSE data

Alanine dipeptide in vacuum

Probably our favourite example ever

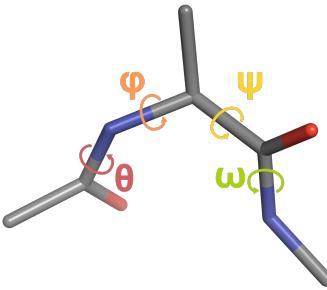


Transition state
(ensemble)



Assessing the relevance of descriptors

The lower the K_m the better

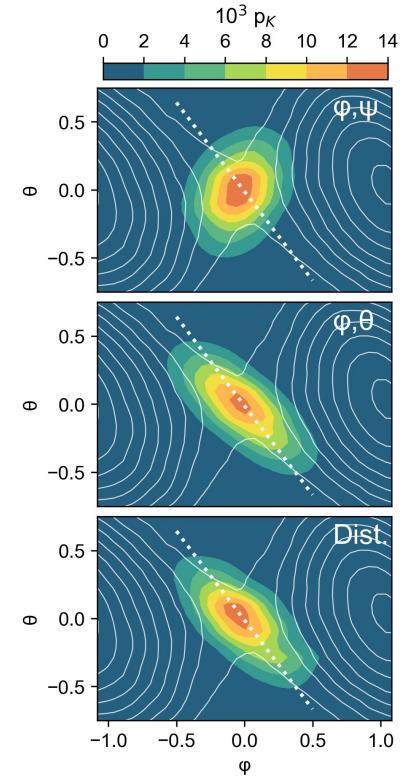


We can pretend we do not know anything about the system and go blind with the angles

- We add **psi** and get slightly better
this is enough for enhanced sampling
- We add others but nothing changes
- We add **theta** and we get **much better!**
this was seen with TPS as well

We can go even blinder and use all the heavy atoms distances in the molecule!

Descriptors \mathbf{x}	K_m
φ, ψ^*	7.6 ± 0.1
φ, ω	8.2 ± 0.4
φ, ψ, ω	7.1 ± 0.2
φ, θ^*	3.4 ± 0.1
φ, ψ, θ	2.7 ± 0.1
φ, θ, ω	3.0 ± 0.1
$\varphi, \psi, \theta, \omega$	2.6 ± 0.1
45 distances*	1.1 ± 0.1



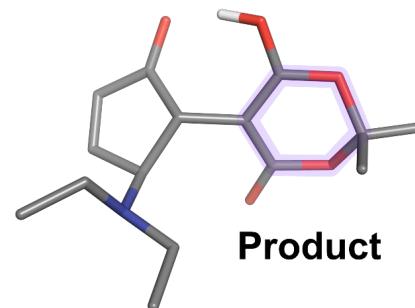
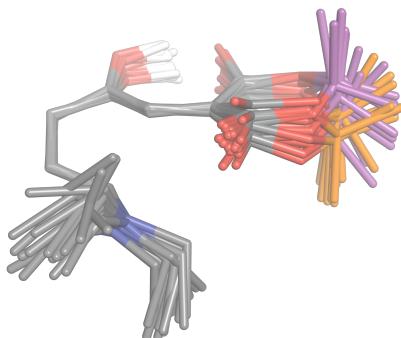
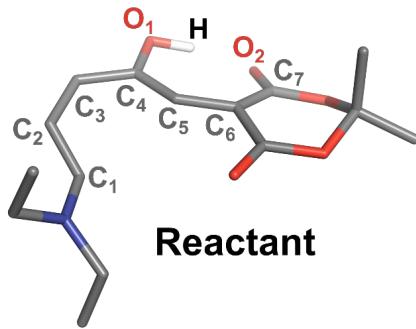
Concerted proton transfer in DASA

A first proper TS ensemble

We have a proton transfer and the simultaneous formation of a ring



When we get the committor we obtain an ensemble of configurations that are not so easy to interpret

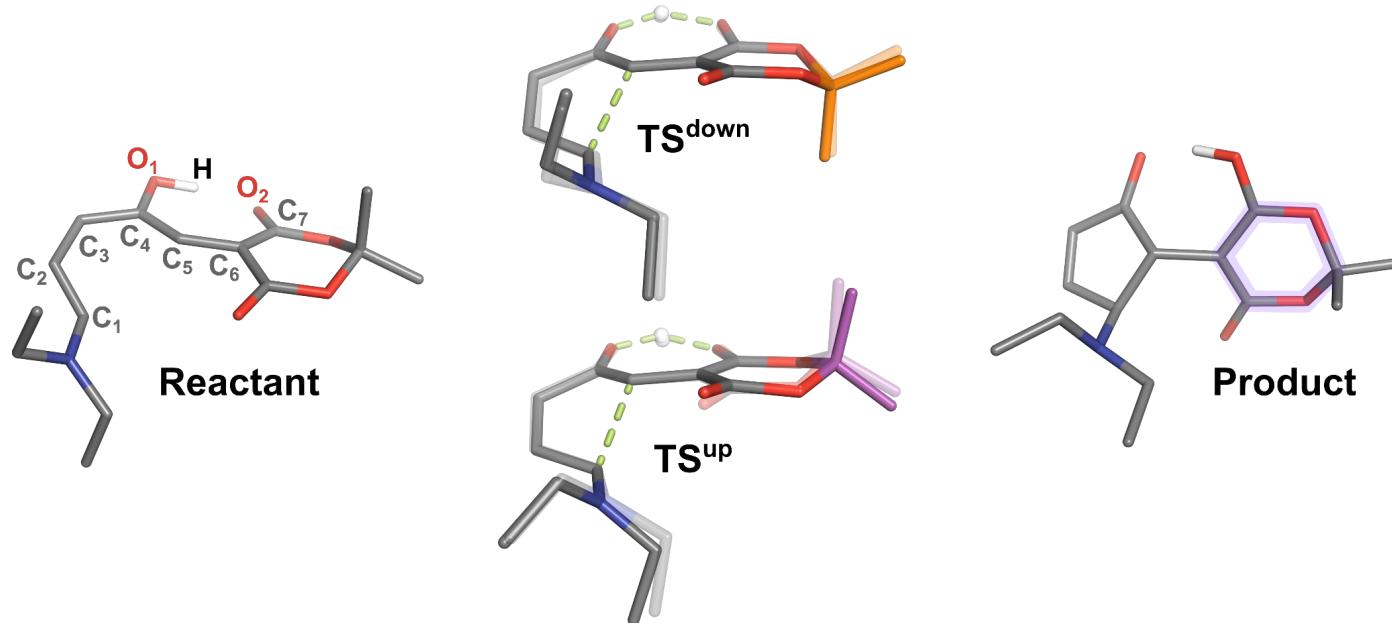


We can still use distances as descriptors

Using clustering to understand the TSE

Extensive sampling opens many doors

But having extensive sampling of the TSE allows us to apply further data-driven analysis techniques to sort it out!

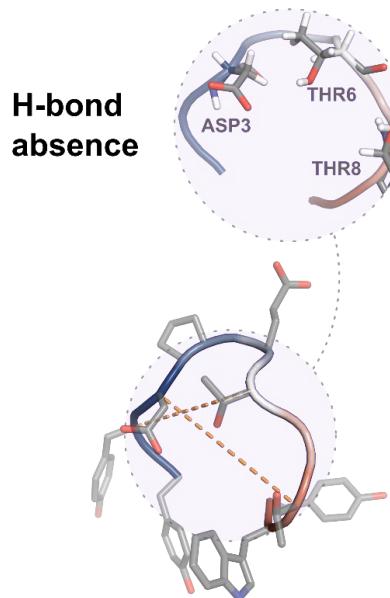


Analyzing two clusters medoid configurations is much better than staring at trajectories for hours

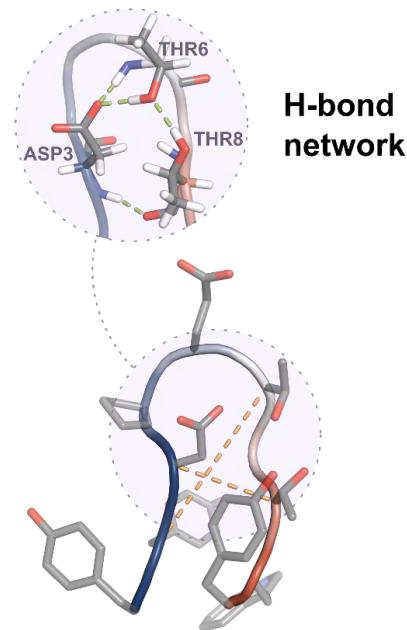
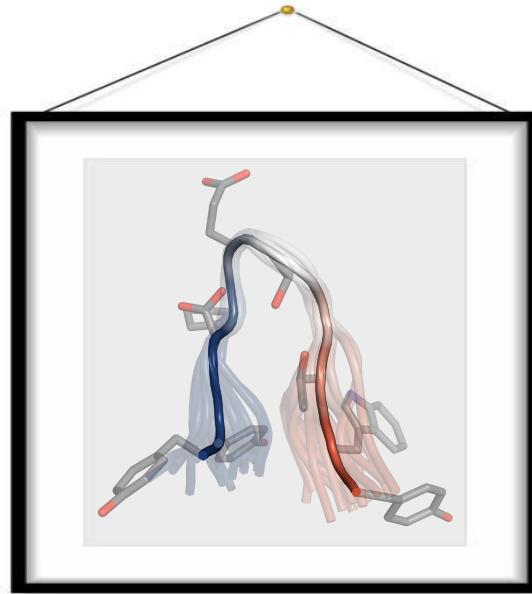
Folding of chignolin in water

My favourite protein

Can we still get the TSE for such a system?



Unfolded State

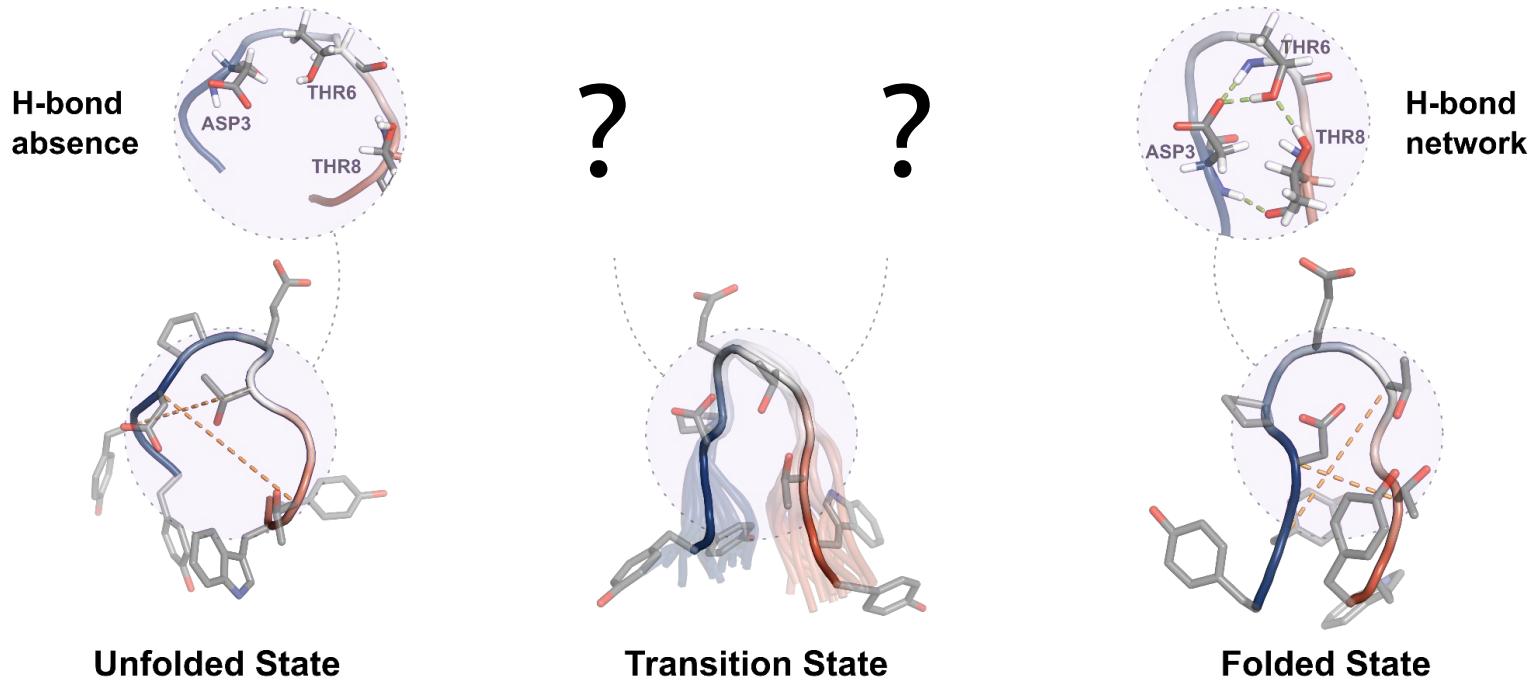


Folded State

Folding of chignolin

My favourite protein

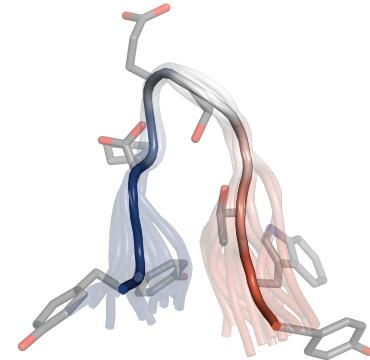
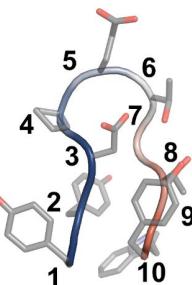
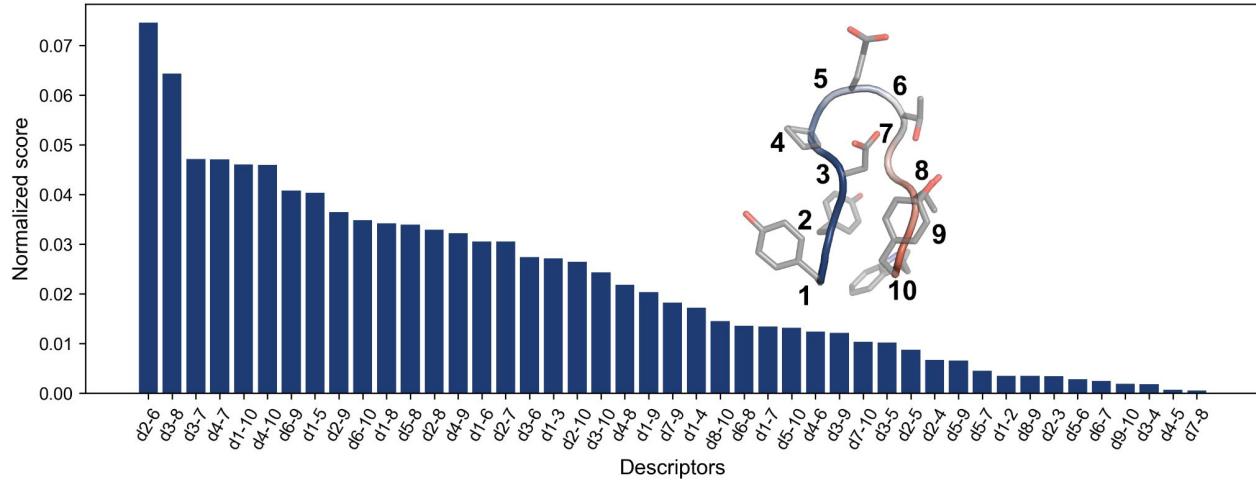
If we apply the clustering we find two clusters, but what are they?



Feature relevance to the rescue

Obtaining information back from our physics-informed model

$$r^k = \sum_{x_i \in \text{TSE}} \left| \frac{\partial q}{\partial d_j^k} \right| \sigma(d_j^k)$$

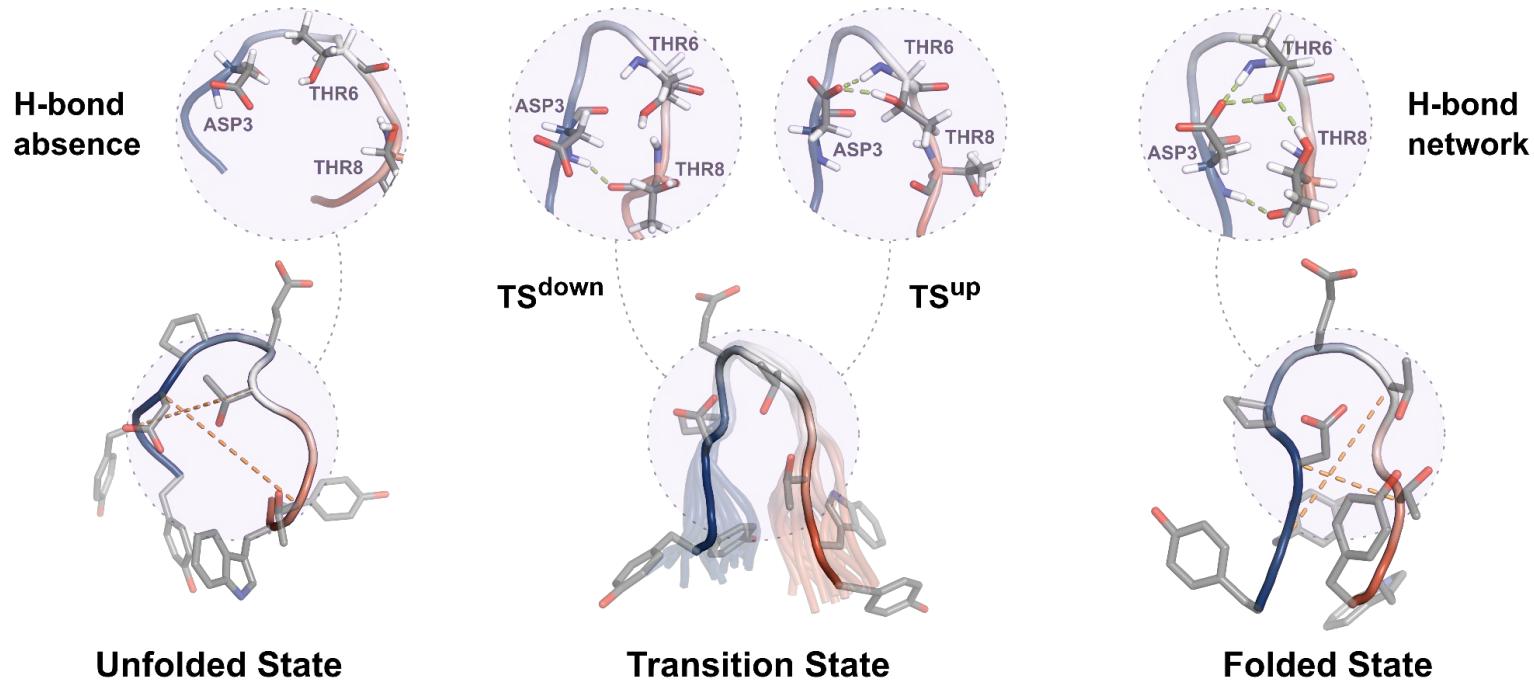


?

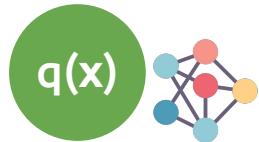
Feature relevance to the rescue

Driving our eyes where it matters

Even if the H-bonds are not explicitly encoded, the variational principle drives the model in the right direction



Wrap up



Machine learning the
committor function



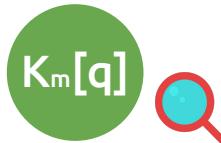
Extending the TS notion
to a proper ensemble



Extensive TSE sampling
allows data-driven analysis



TS-oriented Kolmogorov
bias potential



Variational principle to
identify relevant variables

Future perspectives

And ongoing projects



Not only expressivity
but also interpretability
of ML model



New family of enhanced
sampling methods based
on the committer



Applications to new
interesting and challenging
systems!



Get rid of descriptors to
have generalized and
transferable models



Incorporation in deep
potential active learning
frameworks



Dr. Peilin Kang



Prof. Michele Parrinello



Acknowledgments

Teamwork makes the dreamwork



 @groupparrinello