Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

Future Artificial Intelligence Research

AI for Physics and Chemistry

Istituto Italiano di Tecnologia
Genova

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

We need good catalysts to:

Generate H_2

Capture CO_2

Reduce the environmental cost of many energy intensive chemical processes

All of this should be done at an industrial scale

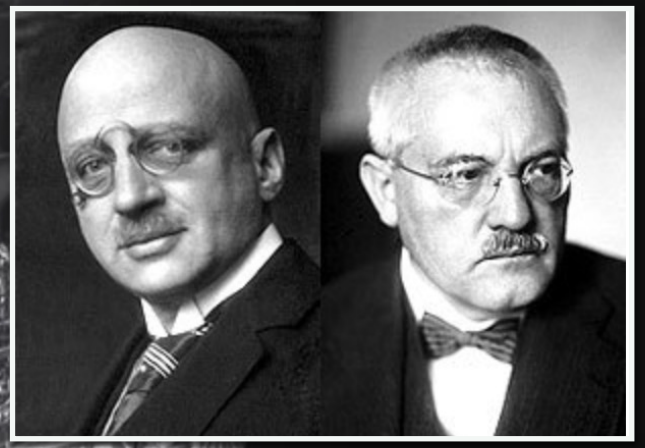
Finanziato dall'Unione europea
NextGenerationEU

Ministero dell'Università e della Ricerca

Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA

FAIR Future Artificial Intelligence Research

150 million tons of Ammonia per year produced via the Haber–Bosch process

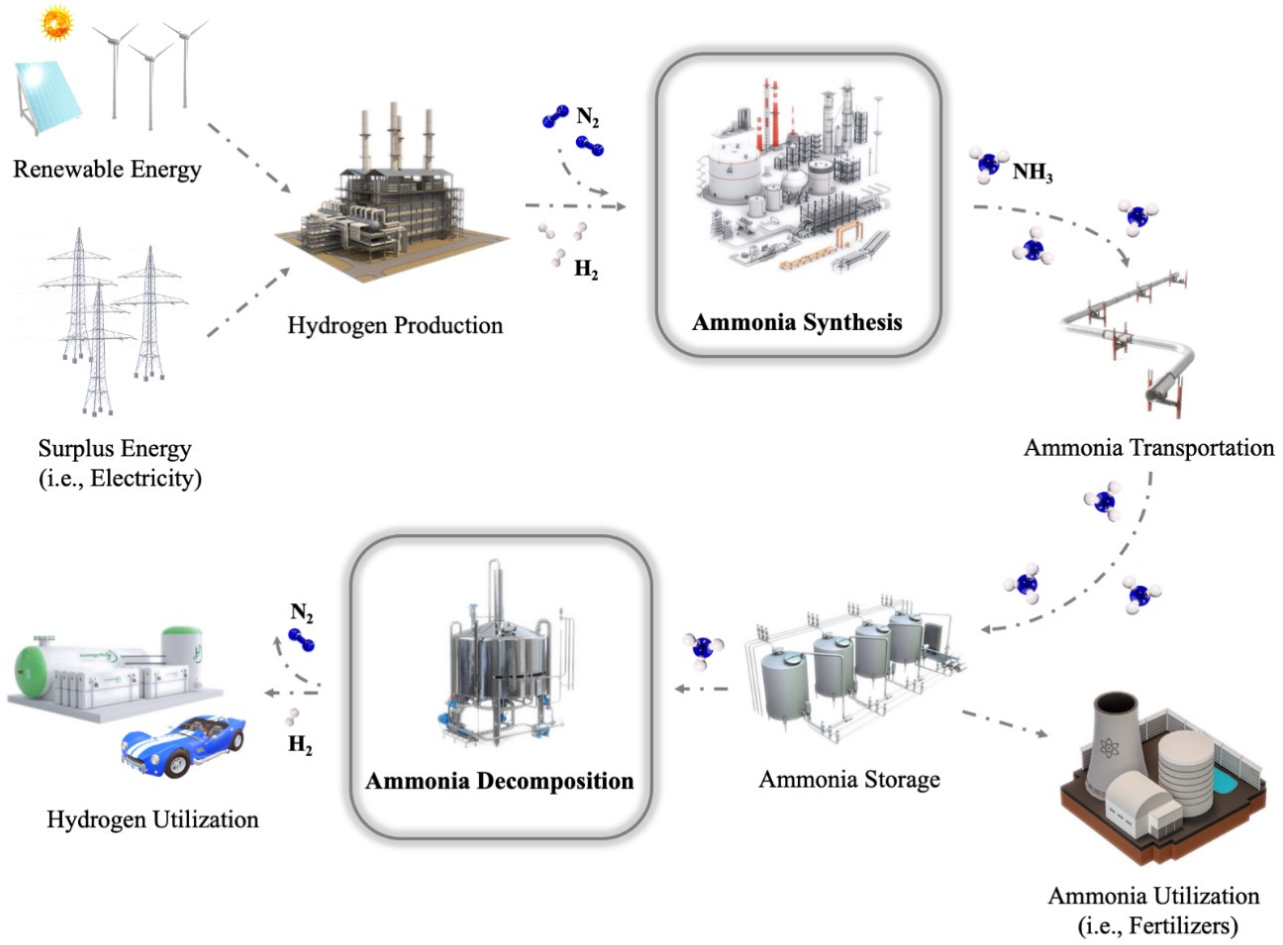


Consumes 2% of the world energy supply

Responsible for 1.6% of world CO_2 emission

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

The ammonia cycle



Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

The challenge

Industrial catalysis takes place at high temperature and pressure.

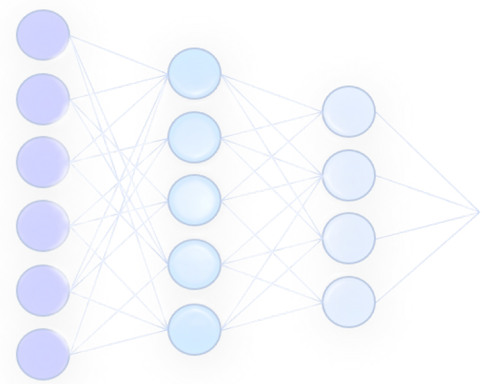
Experiments are difficult

Modeling the high temperature reactive environment of a catalyst is challenging

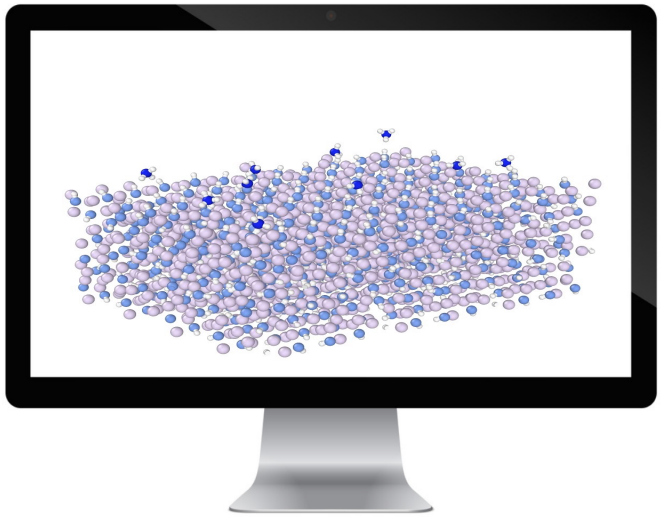
The operando behaviour is inferred from low temperature experiments and theory

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

AI comes to the rescue



ML-based interatomic potentials

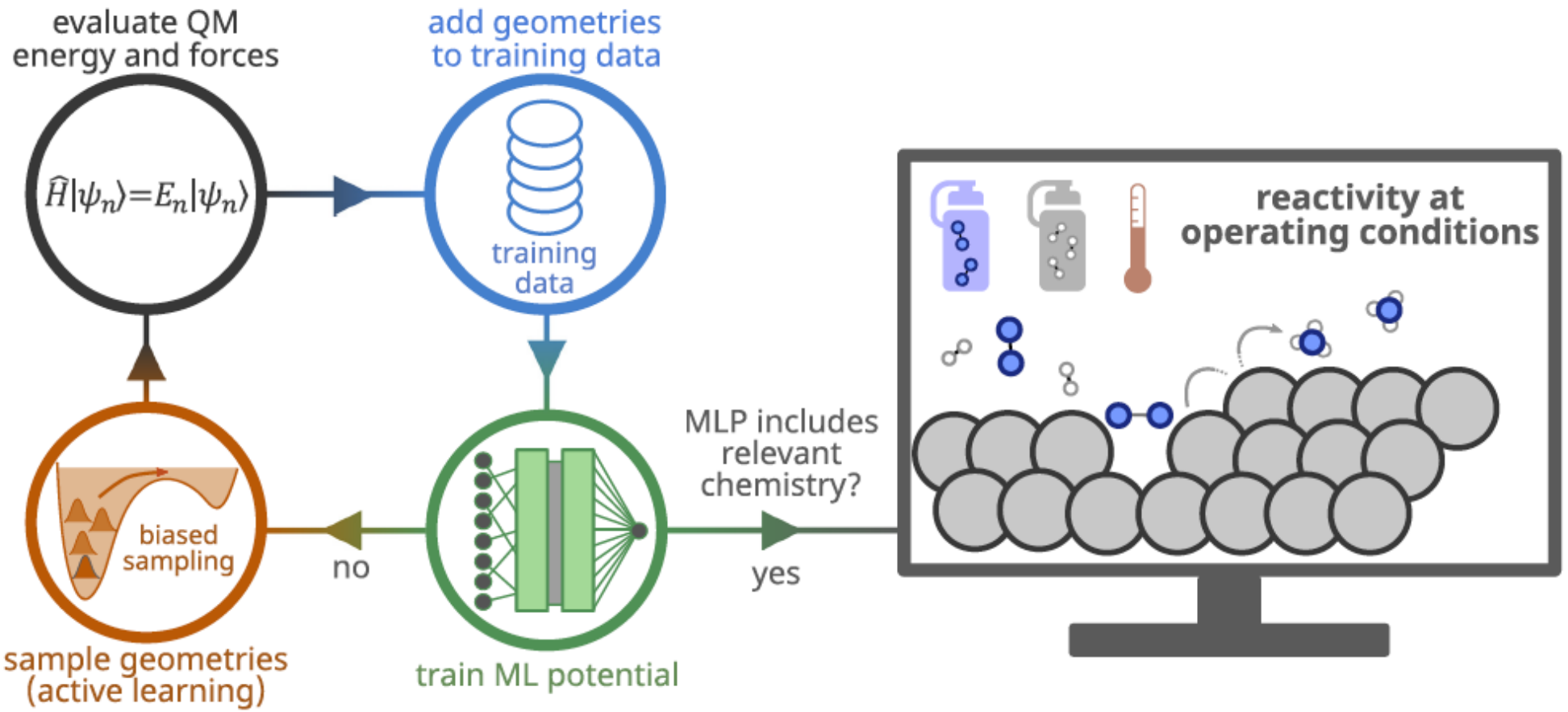


Study large systems (>1k atoms) on long time scales (> 100 ns) with DFT accuracy

Unravelling the workings of a catalyst *in operando* conditions through **machine-learning accelerated molecular dynamics**

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

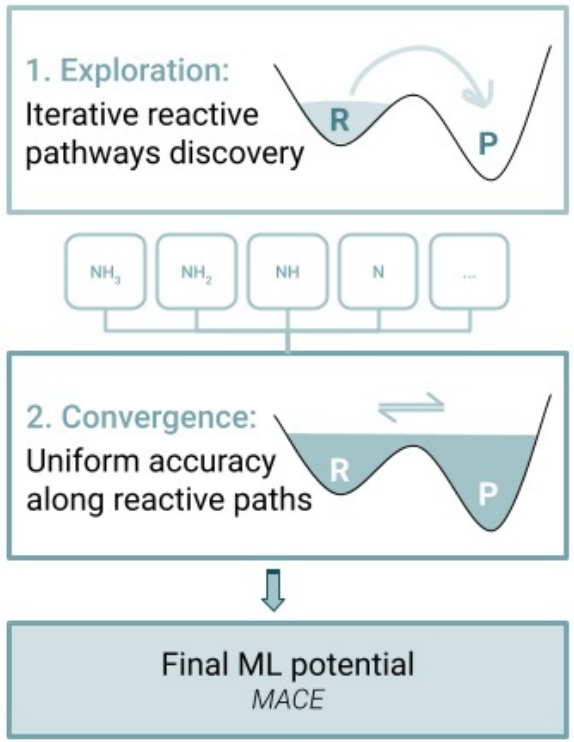
Veronique Van Speybroeck[a,1]

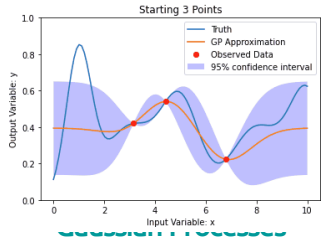


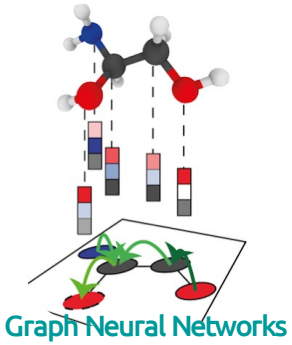
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

Data-efficient machine learning potentials

Combine machine learning, enhanced sampling and active learning (AL) to construct training datasets for *reactive potentials* in a data-efficient way







Build uniformly accurate MLP along reactive pathways with just ~ 1k reference calculations per reaction step

	Standard	Data Efficient
NH3 decomposition	110000 (Fe)	5000 (FeCo)
Lateral interactions	230000 (Fe)	8000 (FeCo)

20-30x less calculations!

Perego & Bonati, ChemRxiv (2024)

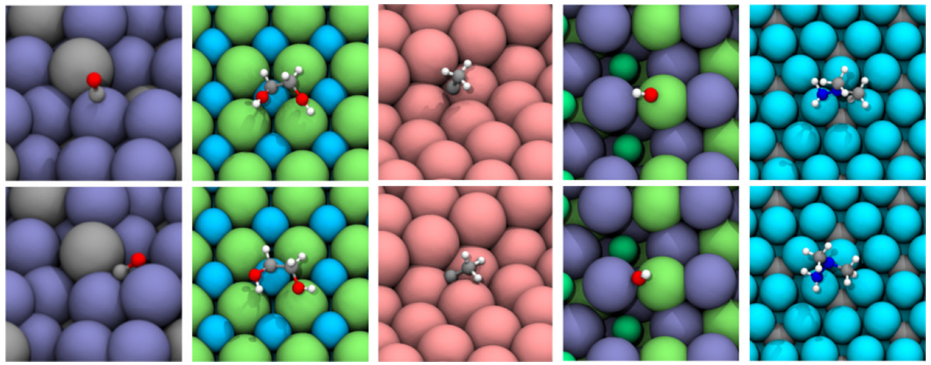
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

Transfer learning from atomistic foundation models

How do we learn efficiently ML potentials for atomistic systems with few and/or costly reference data?

Leverage the availability of large datasets

- "Foundation models" based on graph neural networks (GNNs) trained, e.g. on the OC20 dataset
- Not easy to fine-tune them to specific systems



e.g. OPEN CATALYST:
270 millions of DFT calculations

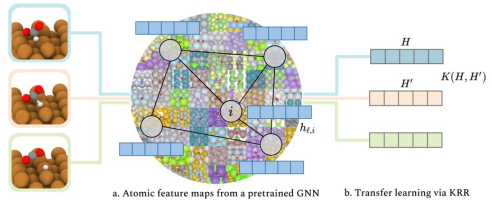
Falk, Bonati, Novelli, Parrinello, Pontil, *NeurIPS* 2023

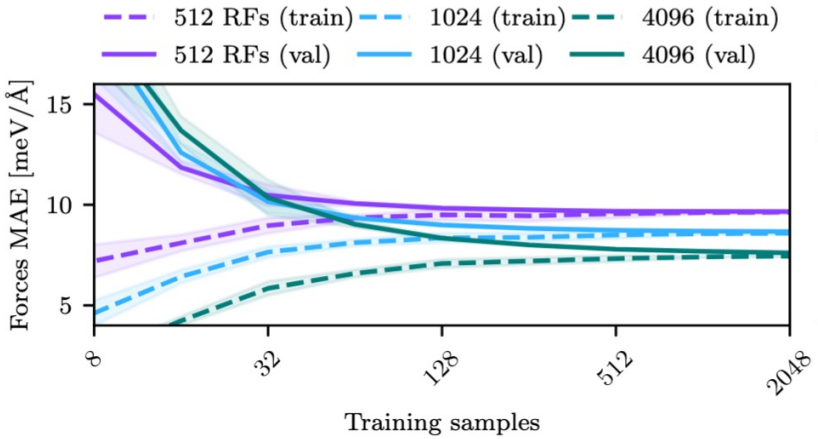
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

Fine tuning of atomistic foundation models for molecular dynamics

a. Extract a representation from a pre-trained GNN atomistic foundation model

b. Learn energy and forces for MD via large-scale kernel techinques



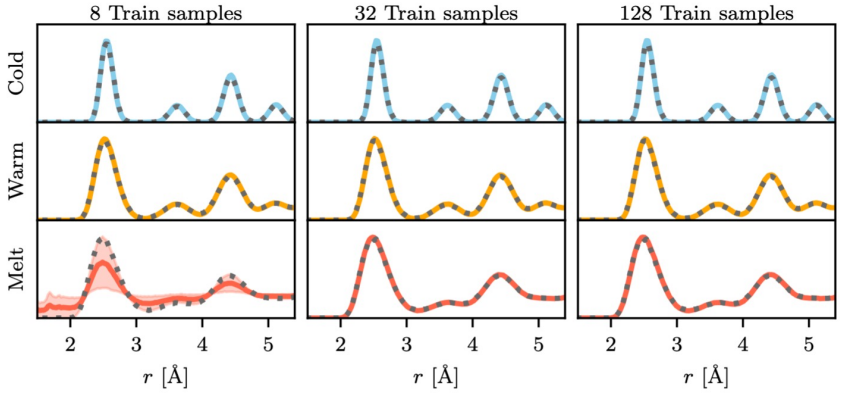


Accurate and data-efficient force predictions

Train time on 4xA100 GPUs

# Samples	Train time [s]
8	5.8 ± 0.4
32	8.4 ± 0.5
128	16.6 ± 0.9
512	46.4 ± 0.9
2048	163.8 ± 0.6

Requires ~seconds to fine-tune models



Reliable MD simulations with just a handful of training samples

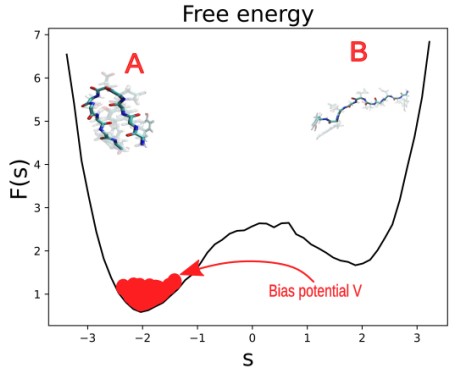
Jorro, Novelli, Bonati, Meanti, Rosasco, Parrinello, Pontil, *in preparation*

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

Learning dynamics from static calculations

Langevin dynamics

$$dX_t = -\frac{1}{\gamma m}\nabla U(X_t)dt + \sqrt{\frac{2k_BT}{\gamma m}}dW_t$$

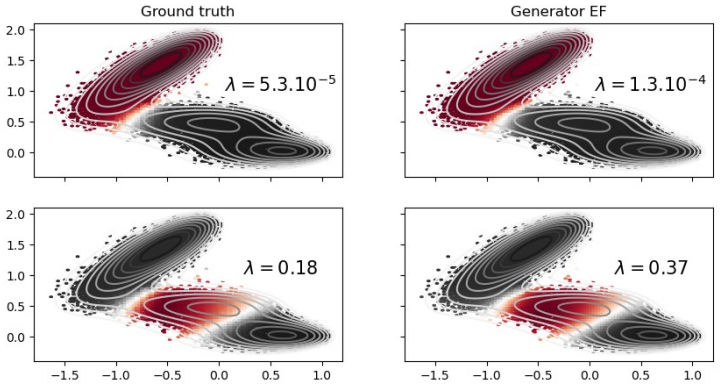


Biased Langevin dynamics

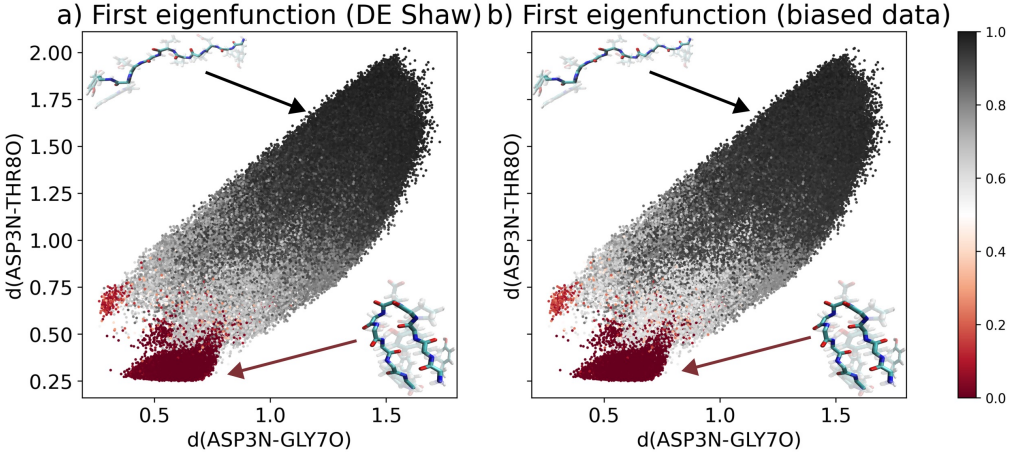
$$dX_t = -\frac{1}{\gamma m}\nabla (U(X_t) + V(X_t))dt + \sqrt{\frac{2k_BT}{\gamma m}}dW_t$$

Neural Network to get actual dynamics

Validation on a toy model :

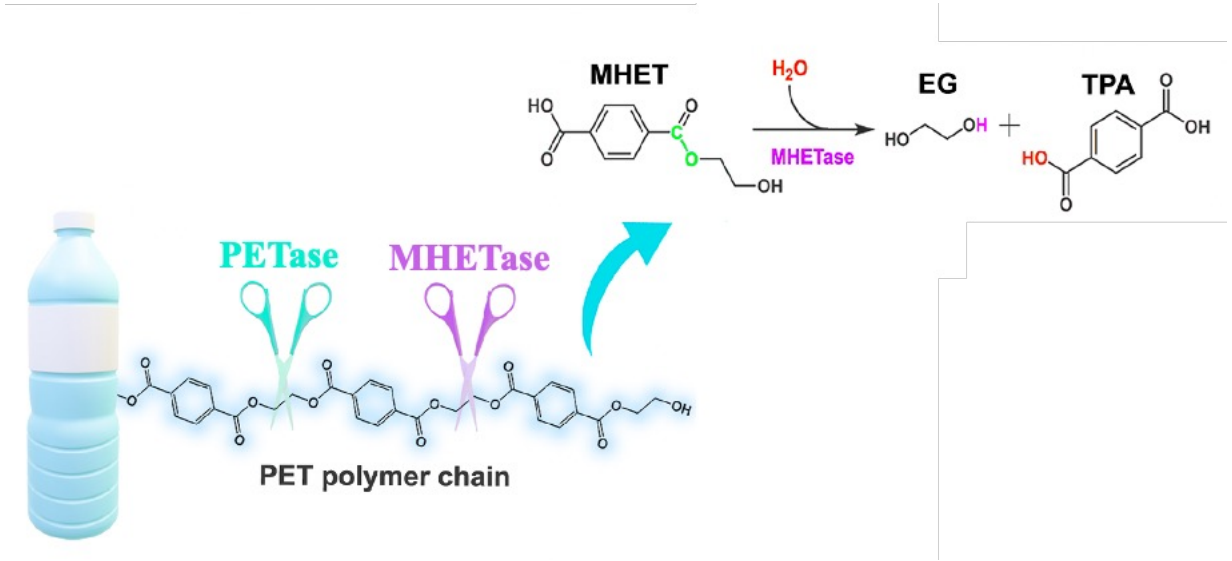


Large scale dataset : chignolin miniprotein folding



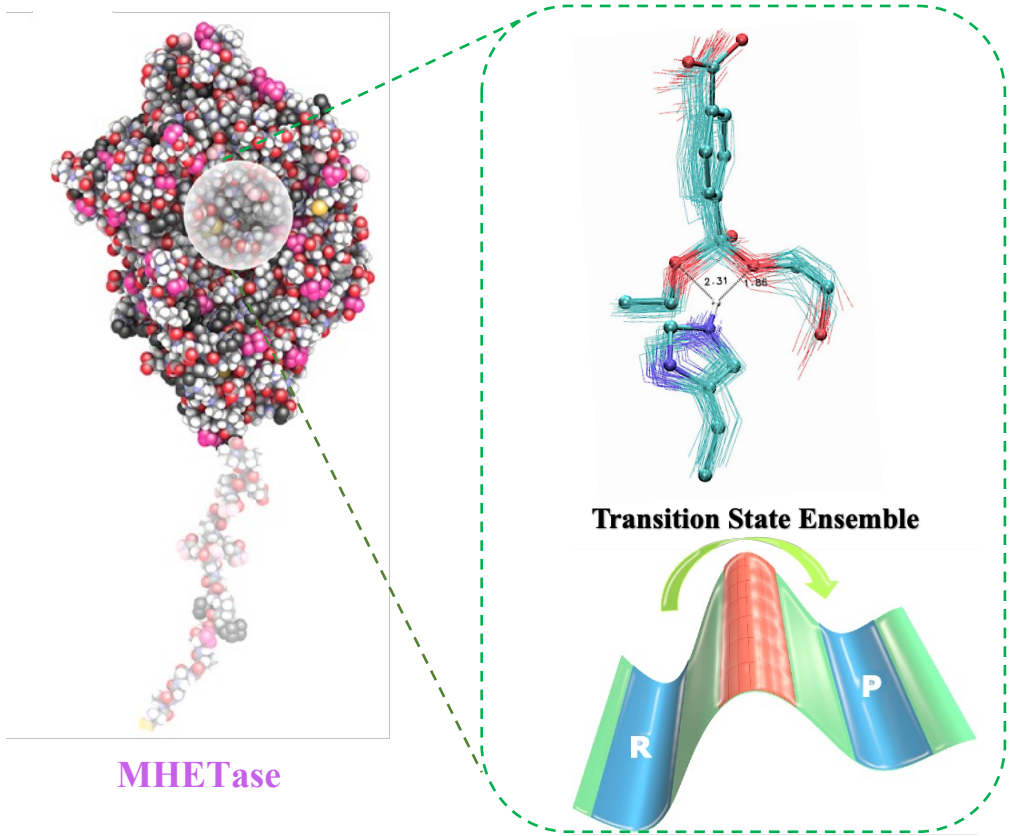
Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
FAIR Future Artificial Intelligence Research

Computational Design of PET Degrading Enzymes



MHETase shows a strong ability to degrade the PET monomer at room temperature, converting it back into the precursors used in PET production[1]

Designing mutations to stabilize the **Transition State Ensemble** via a ML-based committor function[2]



[1] ACS Catal. 2021, 11, 10416 [2] Nat. Comput. Science 2024, 4, 451