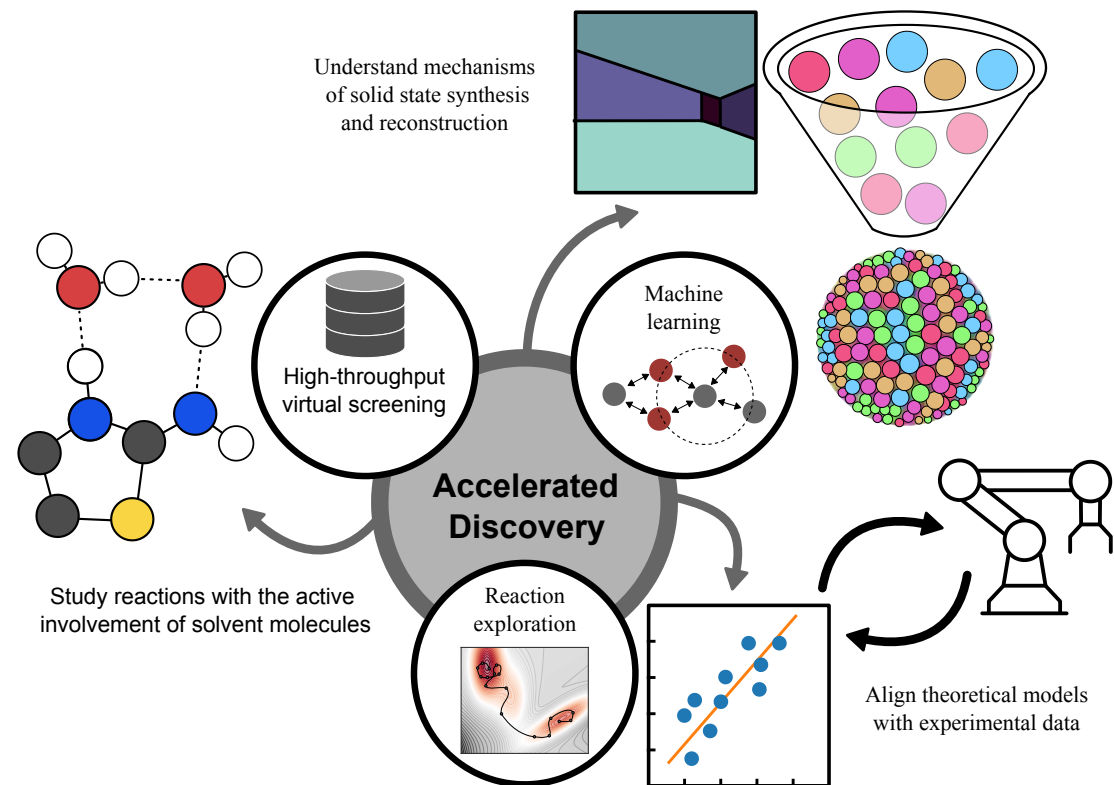


Understanding chemical reaction mechanisms with automation and machine learning

Miguel Steiner

DTU Energy / DTU Compute / CAPeX

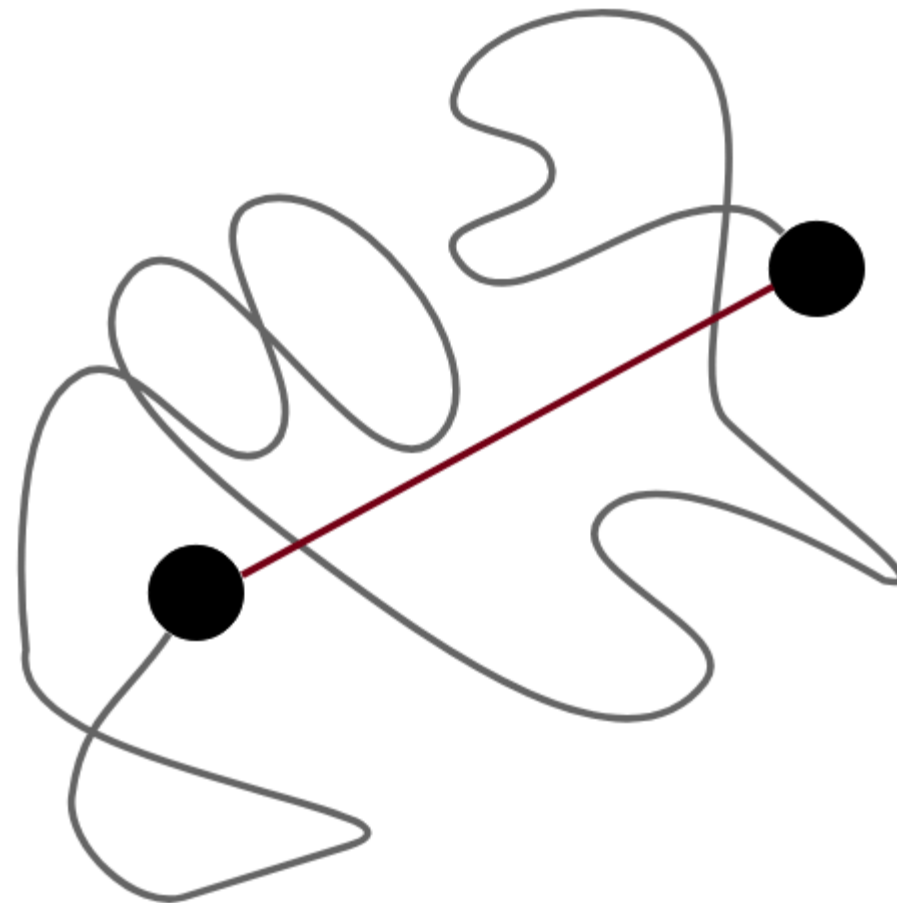


How can we accelerate discovery?

Informed design should be
faster / more efficient than
serendipity

I want to **understand chemical
reactions** and processes at the
atomistic level

Computational approaches offer
insights that are otherwise
**difficult to obtain with
experiments**

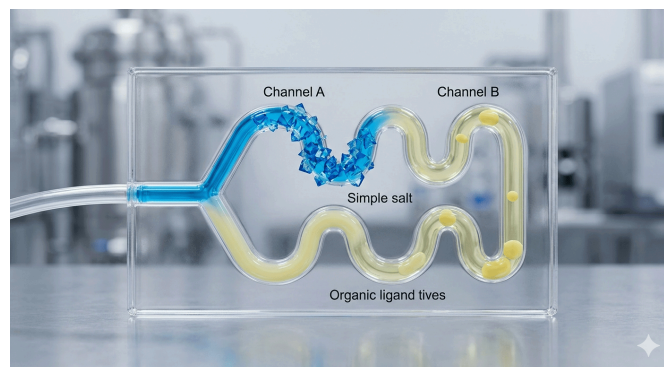


What is computational chemistry capable of?

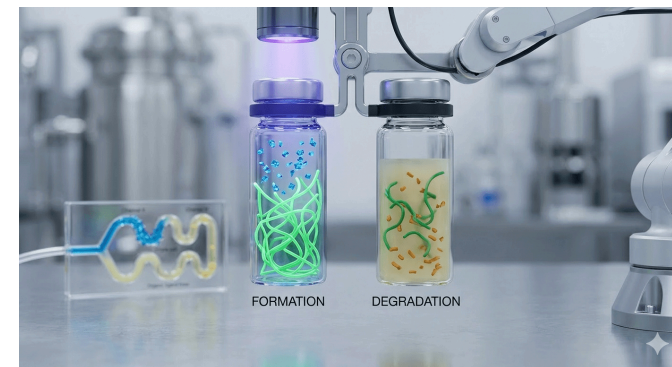
Atomistic insights for molecular reactions with **gas-phase approximations**

Strongly **limited** knowledge about **complex** systems

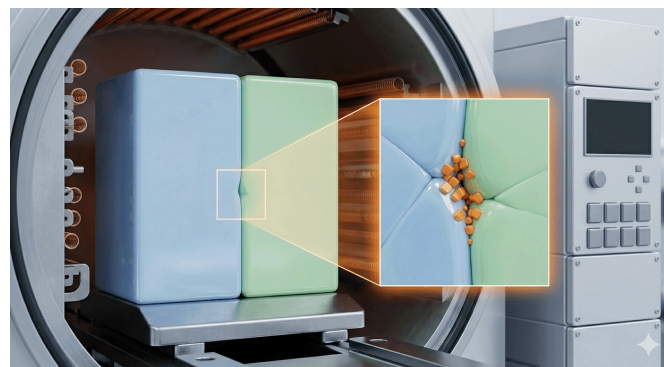
System-focused approaches, **general** accurate methods always **out of reach**



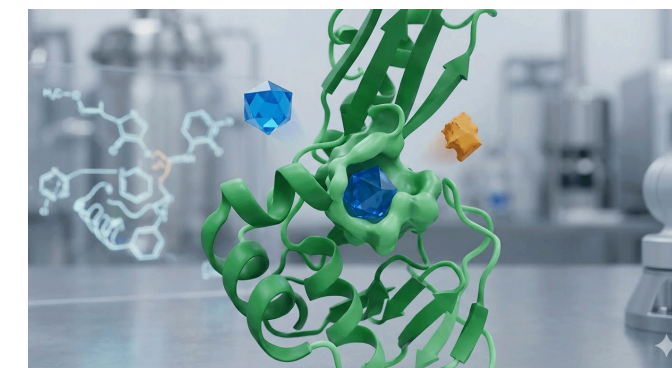
Effects of solvents & additives



Polymer formation & degradation



Solid state synthesis



Enzymes

Electrochemical reactions at triphasic interfaces

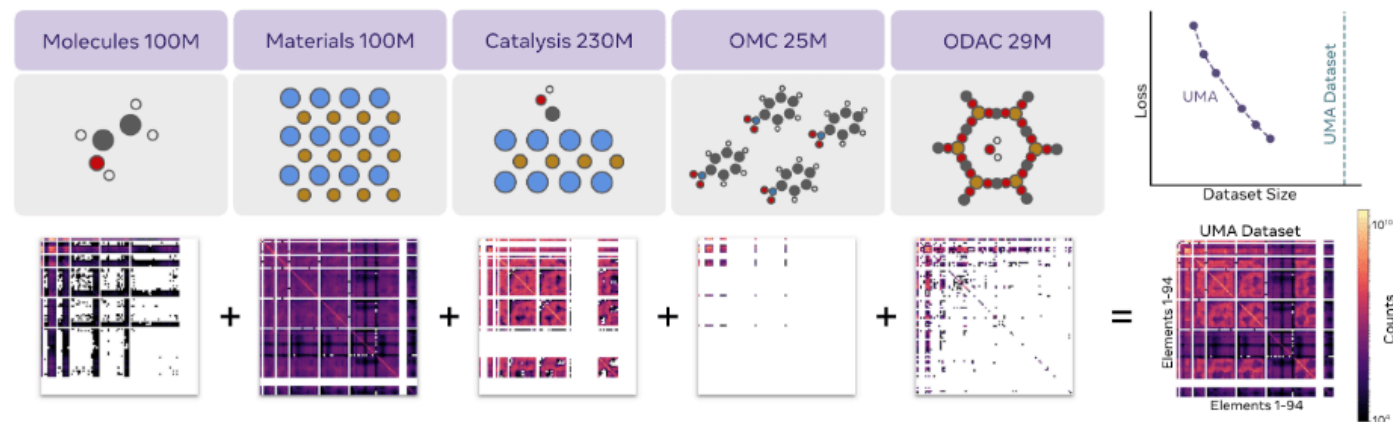
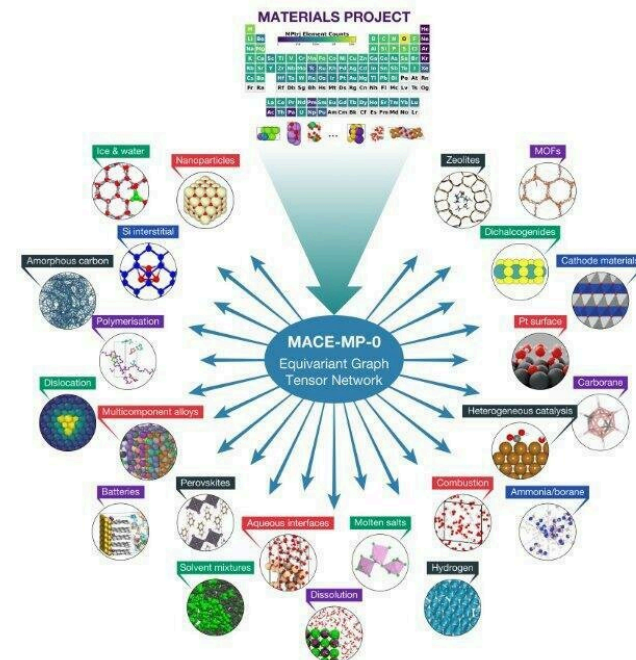


What is computational chemistry capable of?

Models are **breaking** crucial time and size **constraints**

How can we assure that we **calculate the correct things?**

- **Realistic** structural models
- **Connecting** theory and experiments
- Become source of insights: **Predictive** instead of explanatory

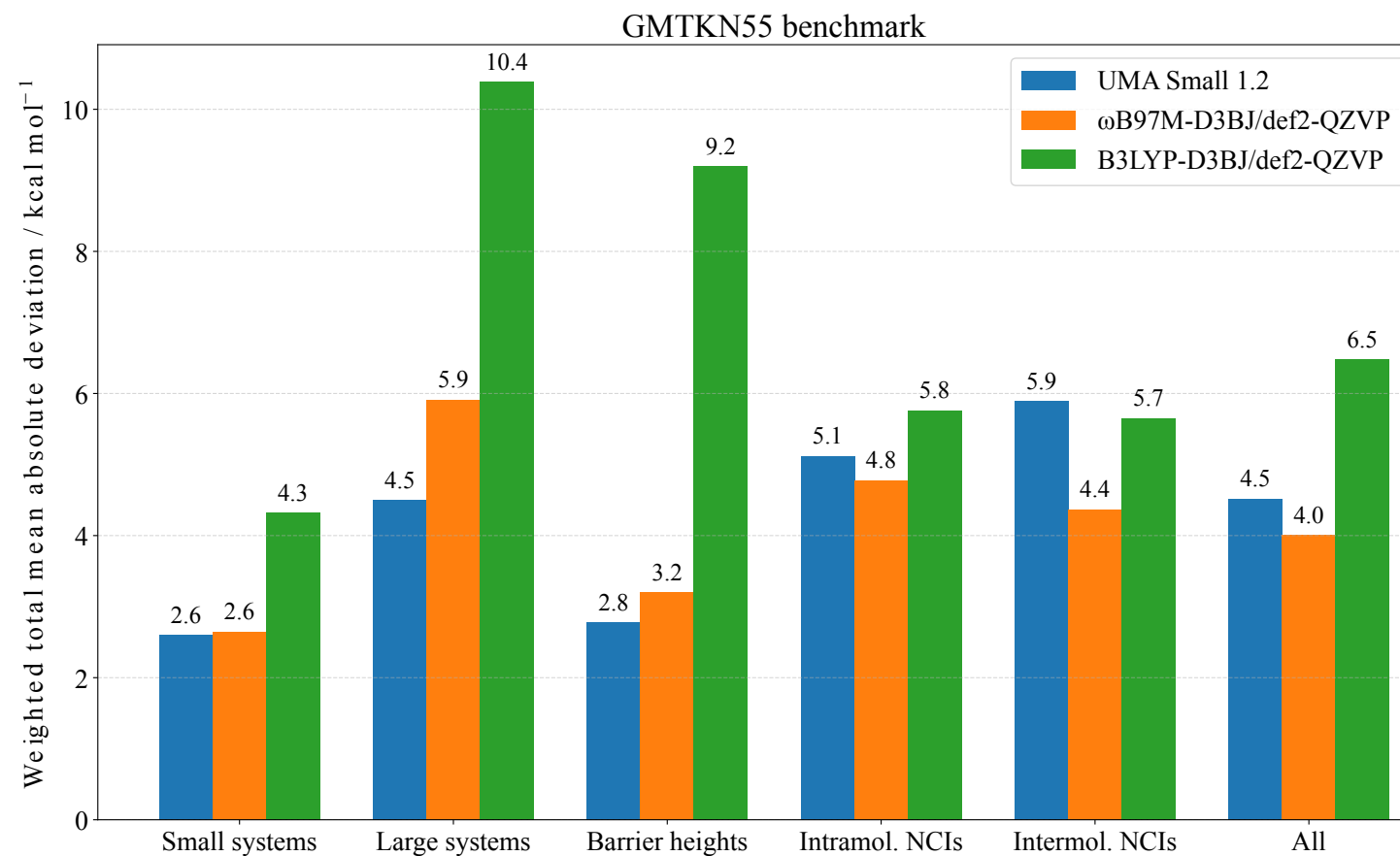


Batatia, et al. *J. Phys. Chem.*, **2025**

Wood, et al., *arXiv:2506.23971*, **2025**

MLIP are as good as their reference data

Have reached DFT accuracy



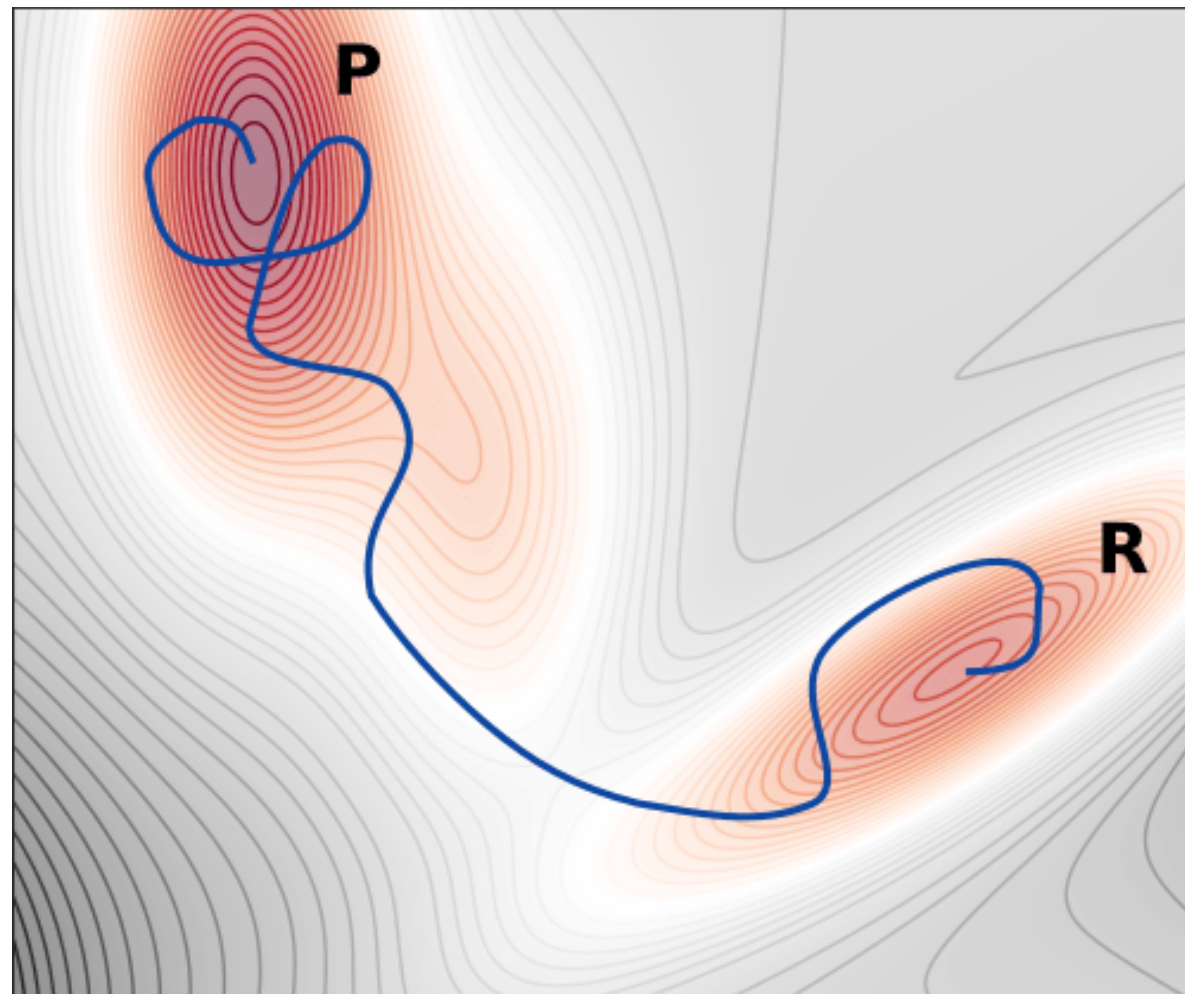
Data by Rowan Scientific Corporation

Reaction exploration is prime use case for MLIP

Reaction exploration either searching for individual transition states (TS) or ab-initio molecular dynamics

Difficult to locate TS or to enhance sampling to find unknown reaction mechanisms

Inherently large computational cost

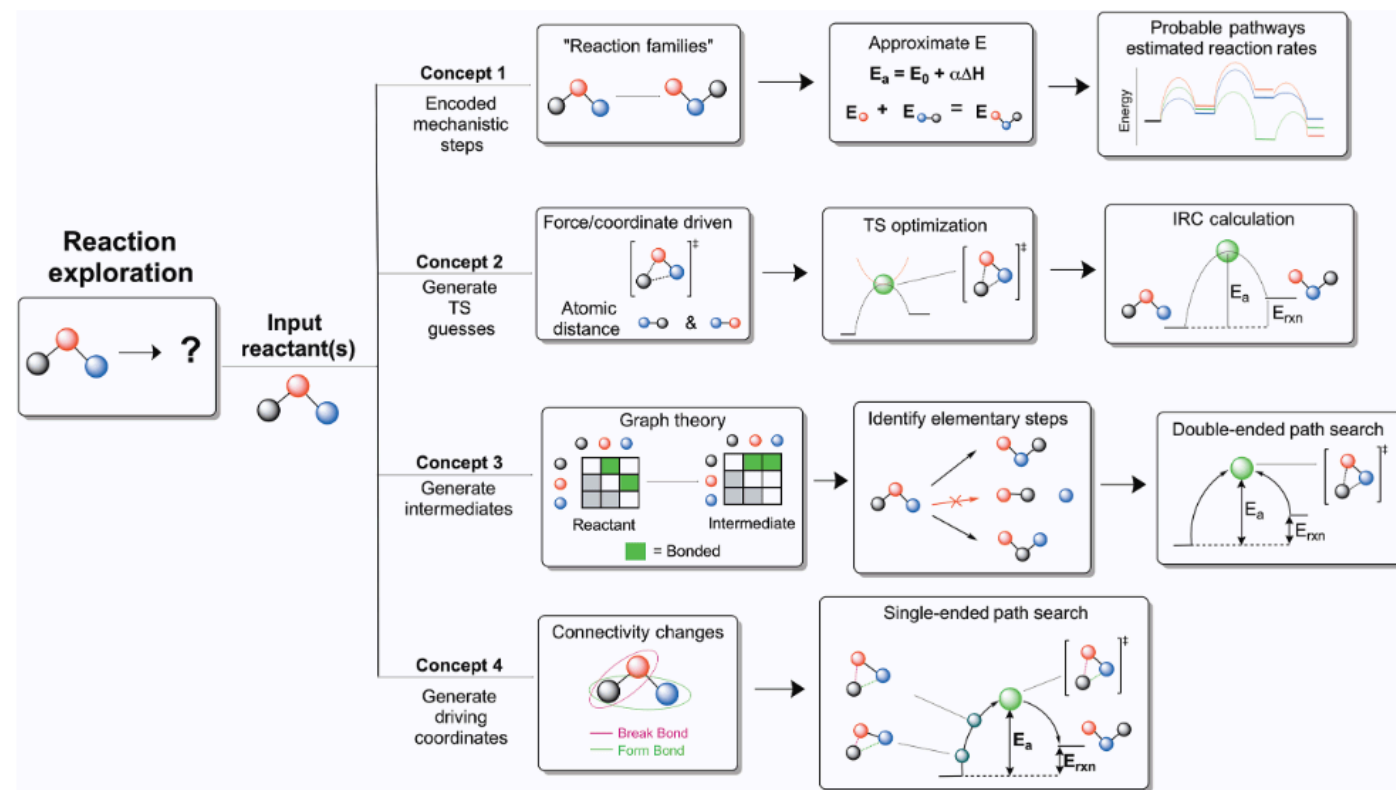


Automating reaction exploration

One elementary step is not enough: require broad coverage which requires automation

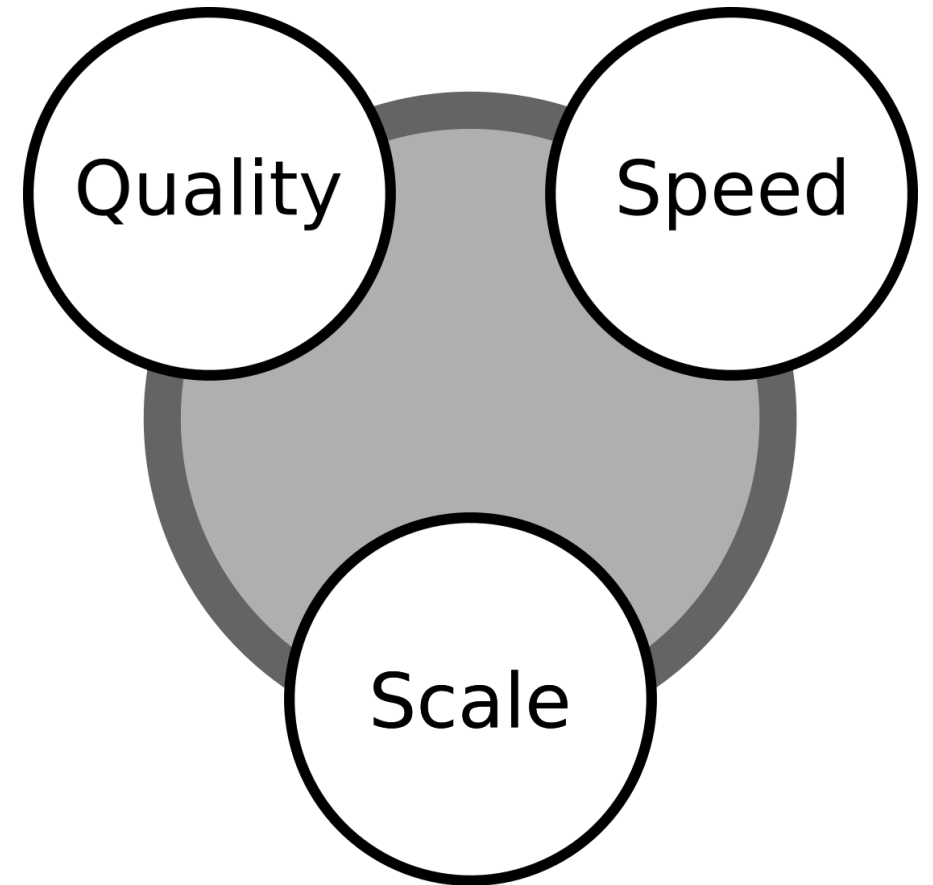
Initially driven by top-level approaches

Systematic workflows developed over time



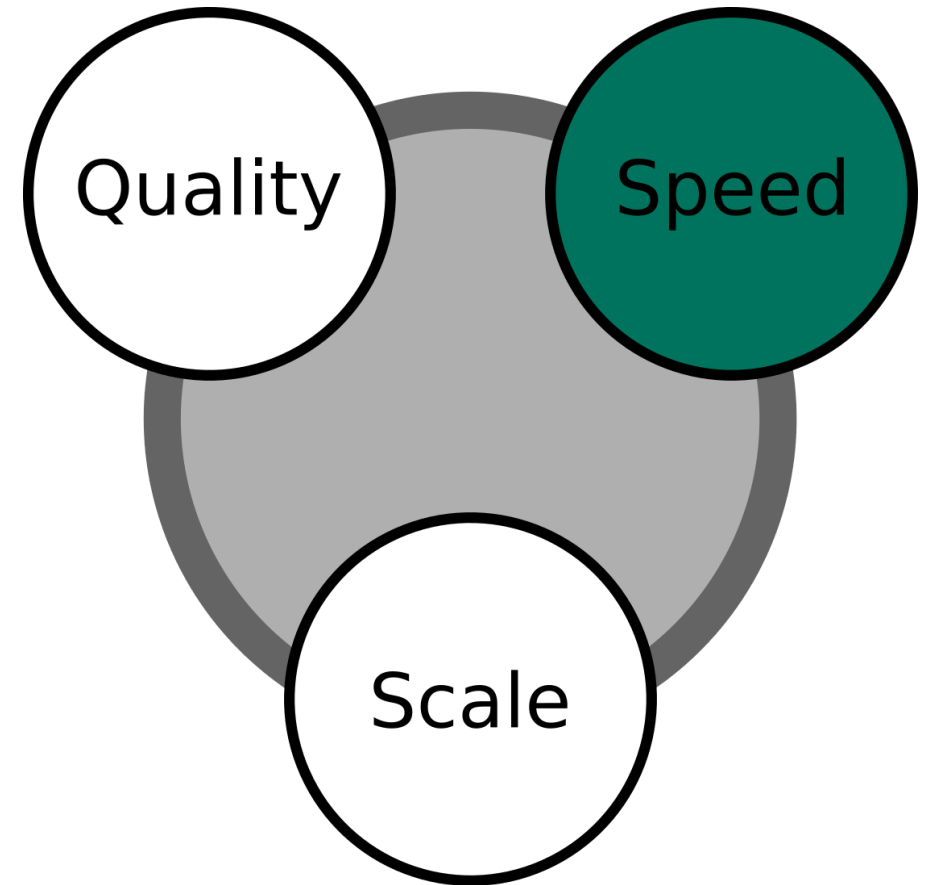
Steering explorations

Explorations require careful **balance**, already for simple systems, even more for complex systems



Steering explorations

Explorations require careful **balance**, already for simple systems, even more for complex systems

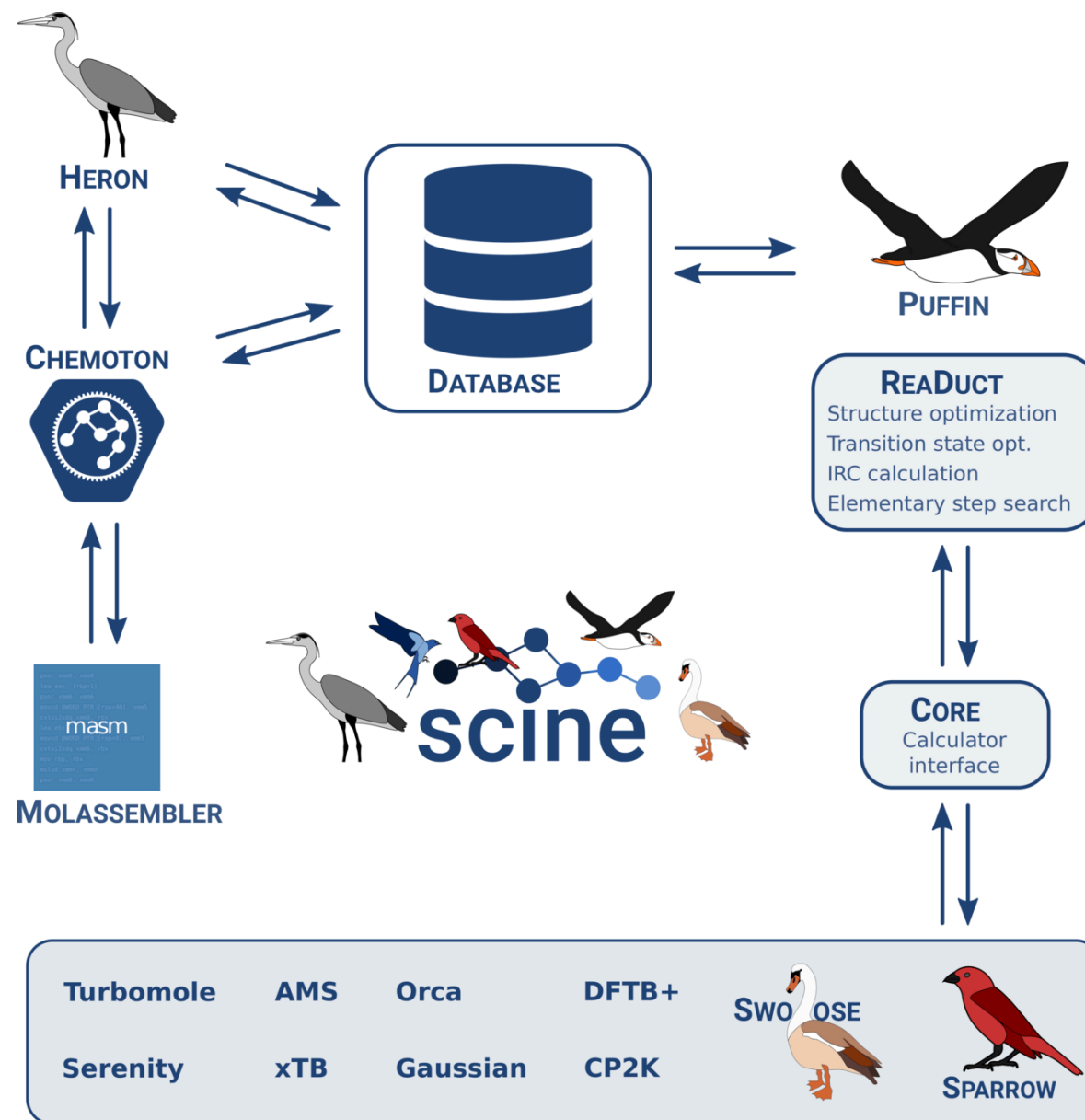


SCINE framework

Modular and **open-source**
reaction exploration

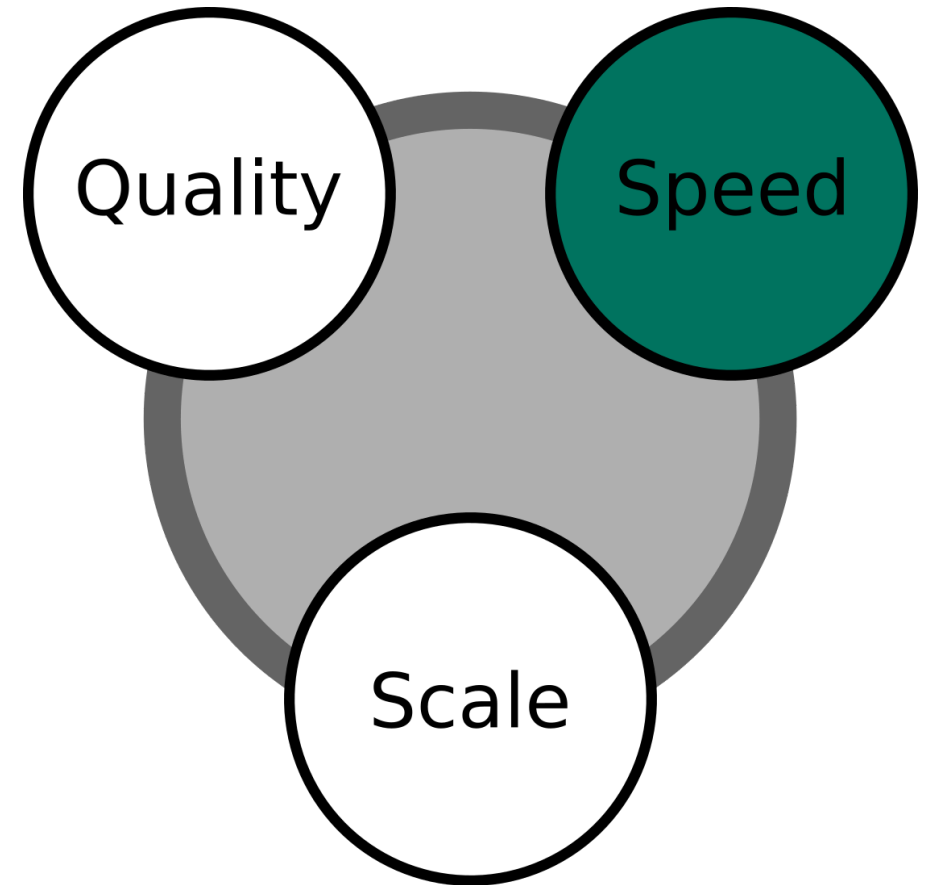
Flexible regarding the applied
optimization algorithm, reaction
sampling, and **electronic
structure**

Minimal chemical **constraints**



Steering explorations

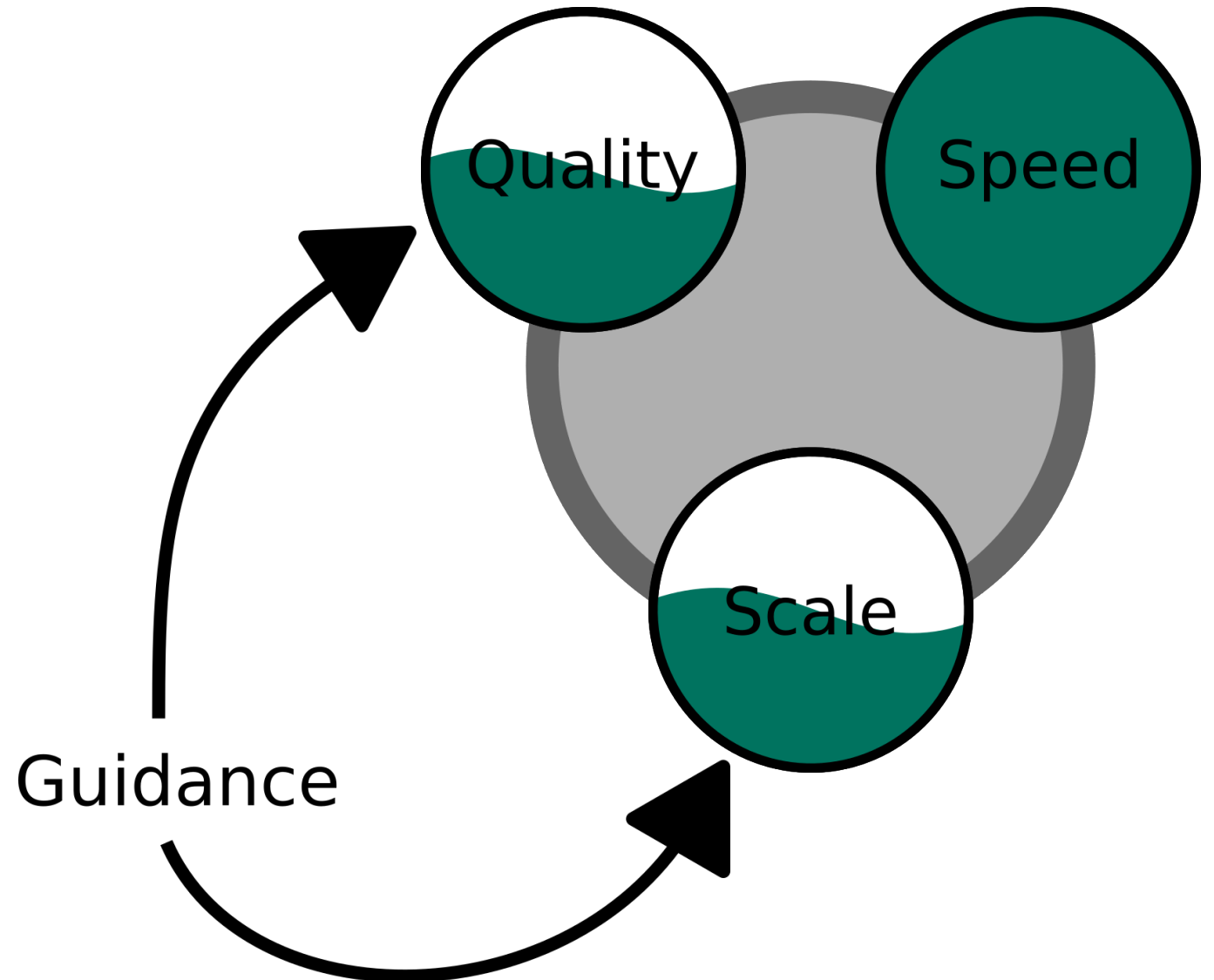
Explorations require careful **balance**, already for simple systems, even more for complex systems



Steering explorations

Explorations require careful **balance**, already for simple systems, even more for complex systems

This balance **can come from guiding** the automated algorithms



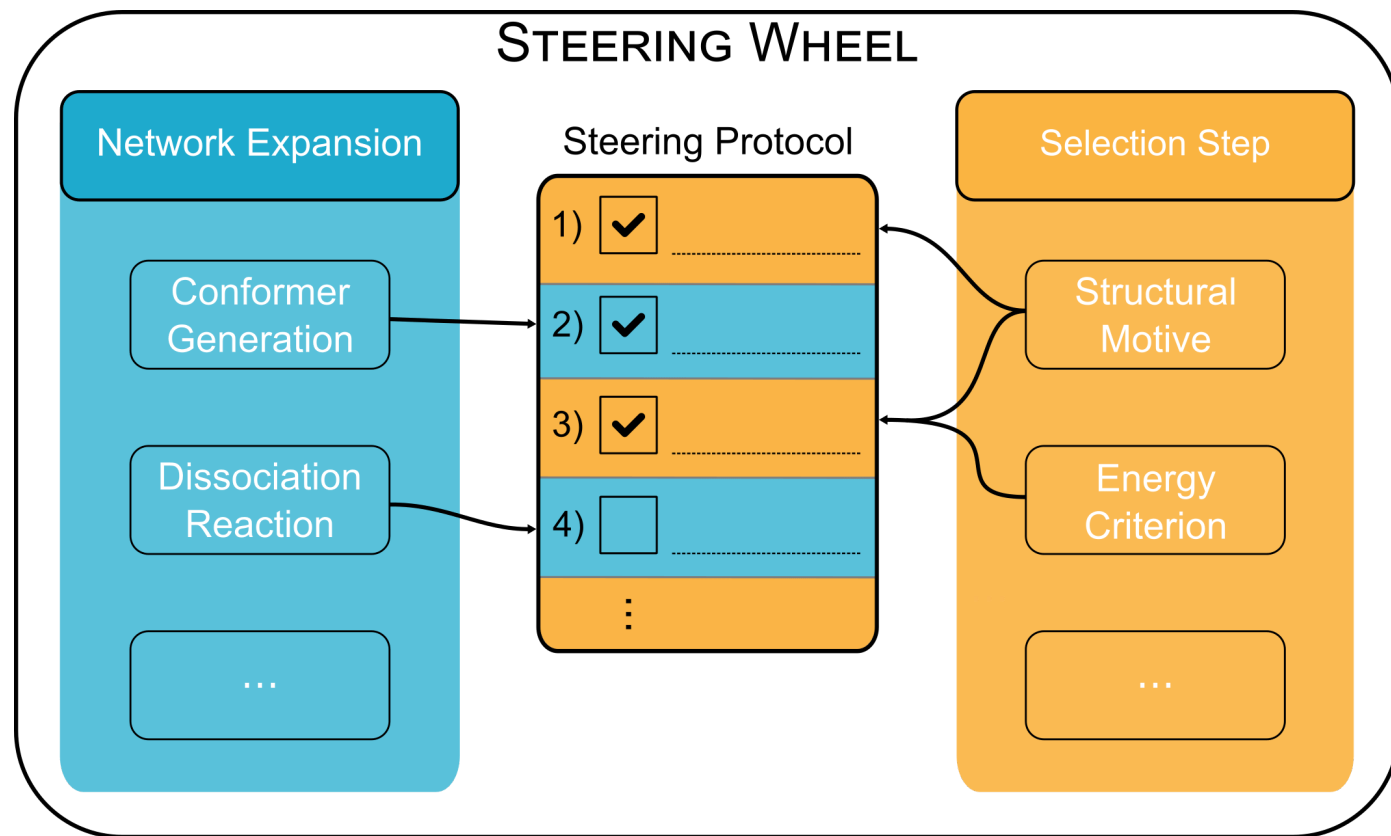
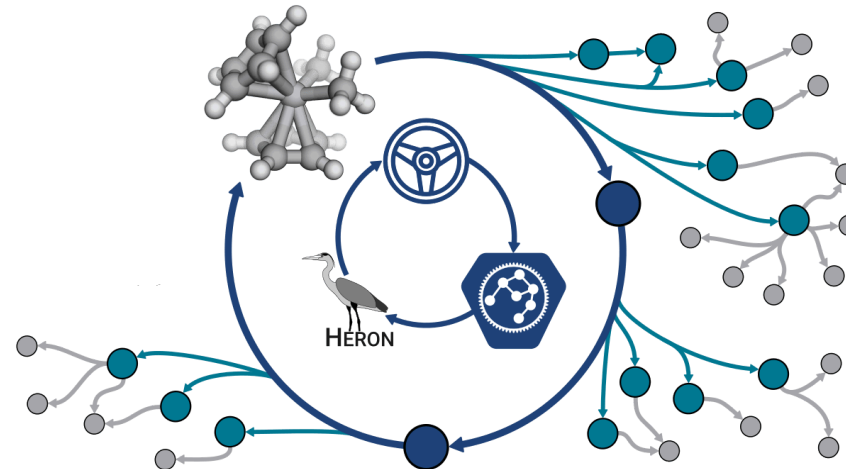
Guiding automated algorithms

Steering explorations

Framework to guide
exploration campaigns

Objectives are fully
customizable and workflow
is **reproducible** by design

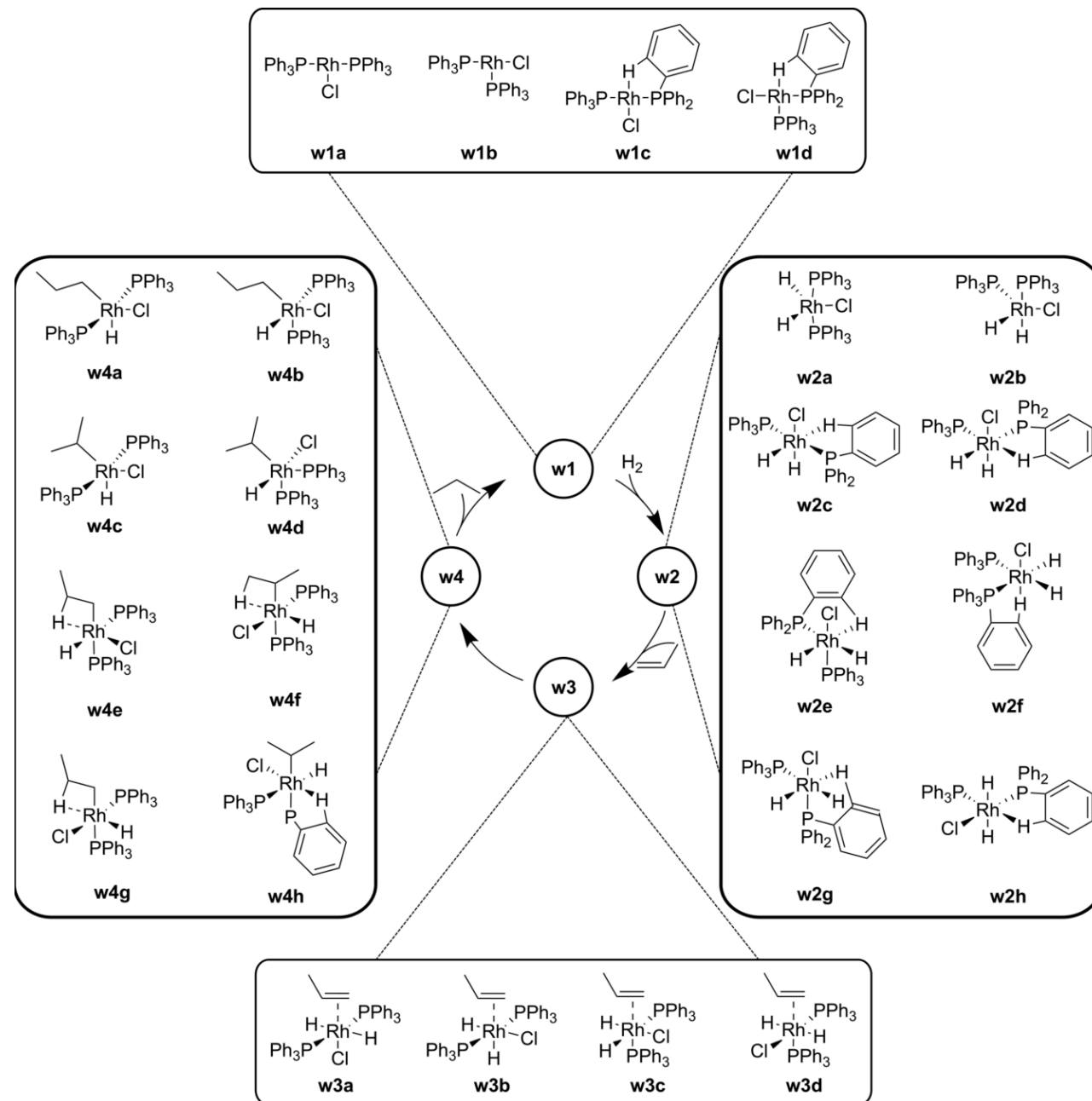
Explored reaction space can
be **enhanced afterwards**



Wilkinson hydrogenation

Expanding reaction path from small alley to a highway

Besides literature intermediates, found most variations in ligand sphere in depth-first exploration



Nanoscopic structures

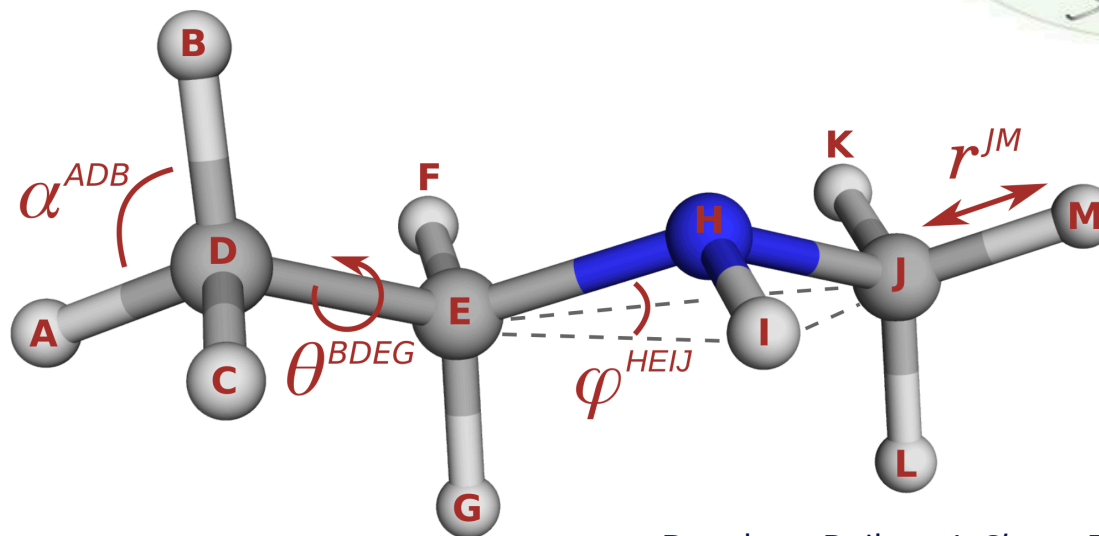
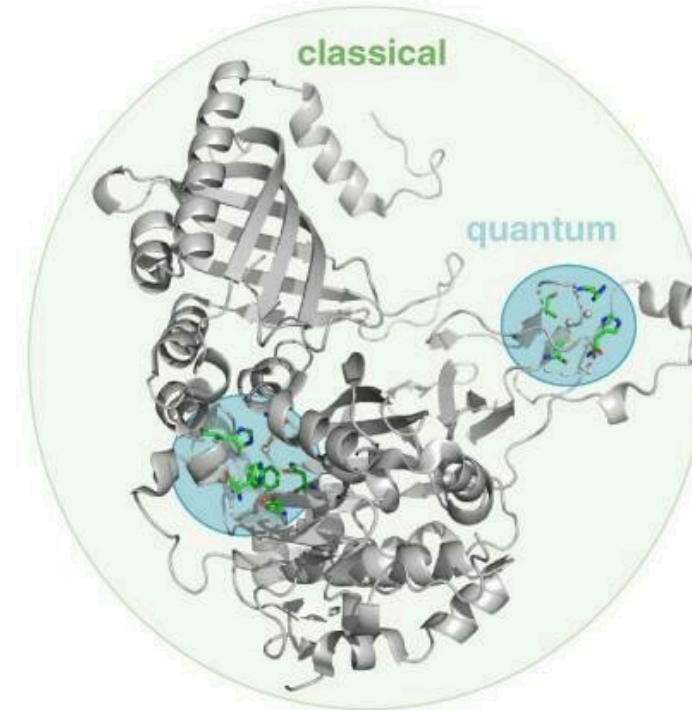
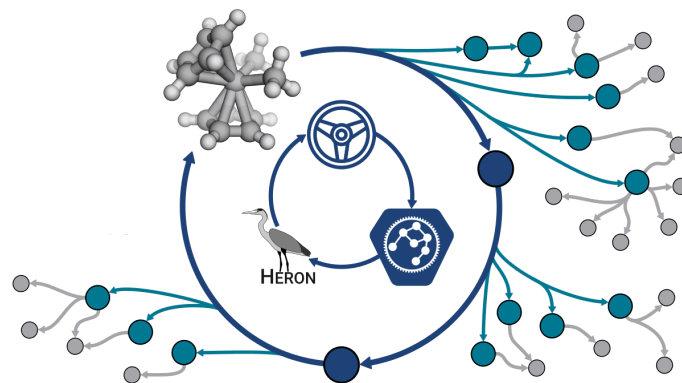
Can we now explore **all possible chemical structures** in a steered manner?

But if they are too large?

→ Hybrid methods (QM/MM)

If there is no MM

→ System-focused model (SFAM)



Brunken, Reiher, *J. Chem. Theory Comput.*, **2020**

Brunken, Reiher, *J. Chem. Theory Comput.*, **2021**

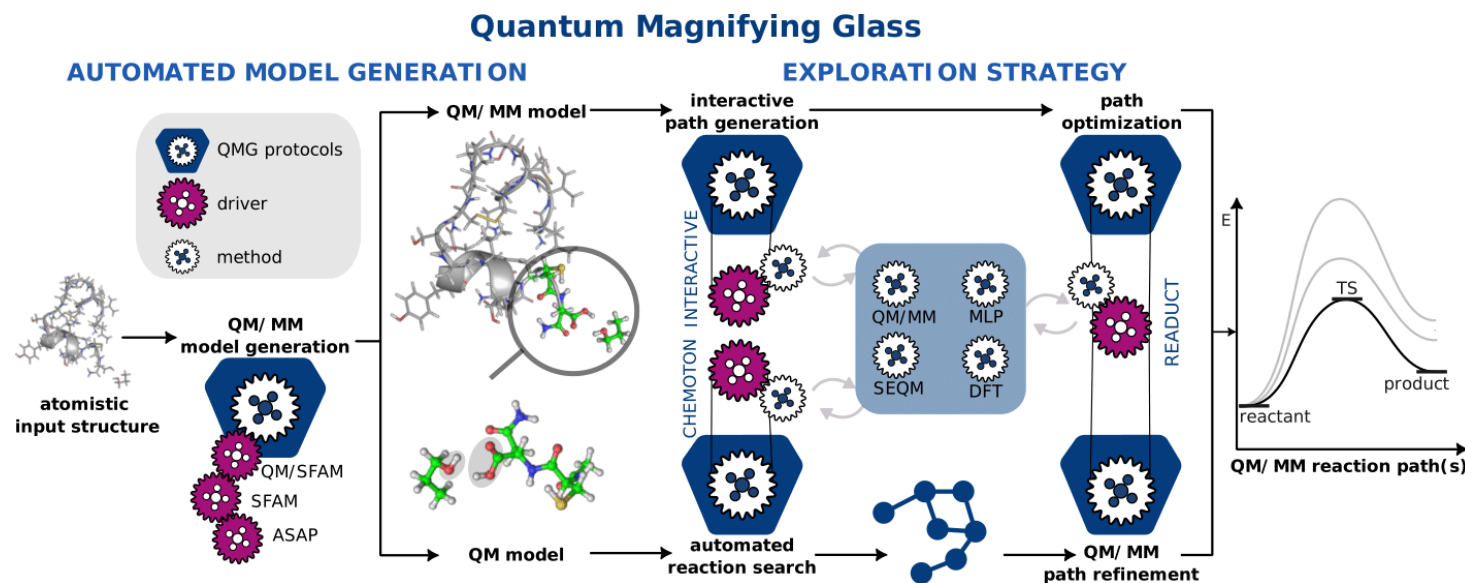
Csizi, Reiher, *WIREs Comput. Mol. Sci.*, **2023**

Quantum Magnifying Glass

Extend reaction explorations to nanoscopic systems with QM/MM hybrid models

Steered exploration of extracted sub-system or

Real-time calculations to generate reaction paths

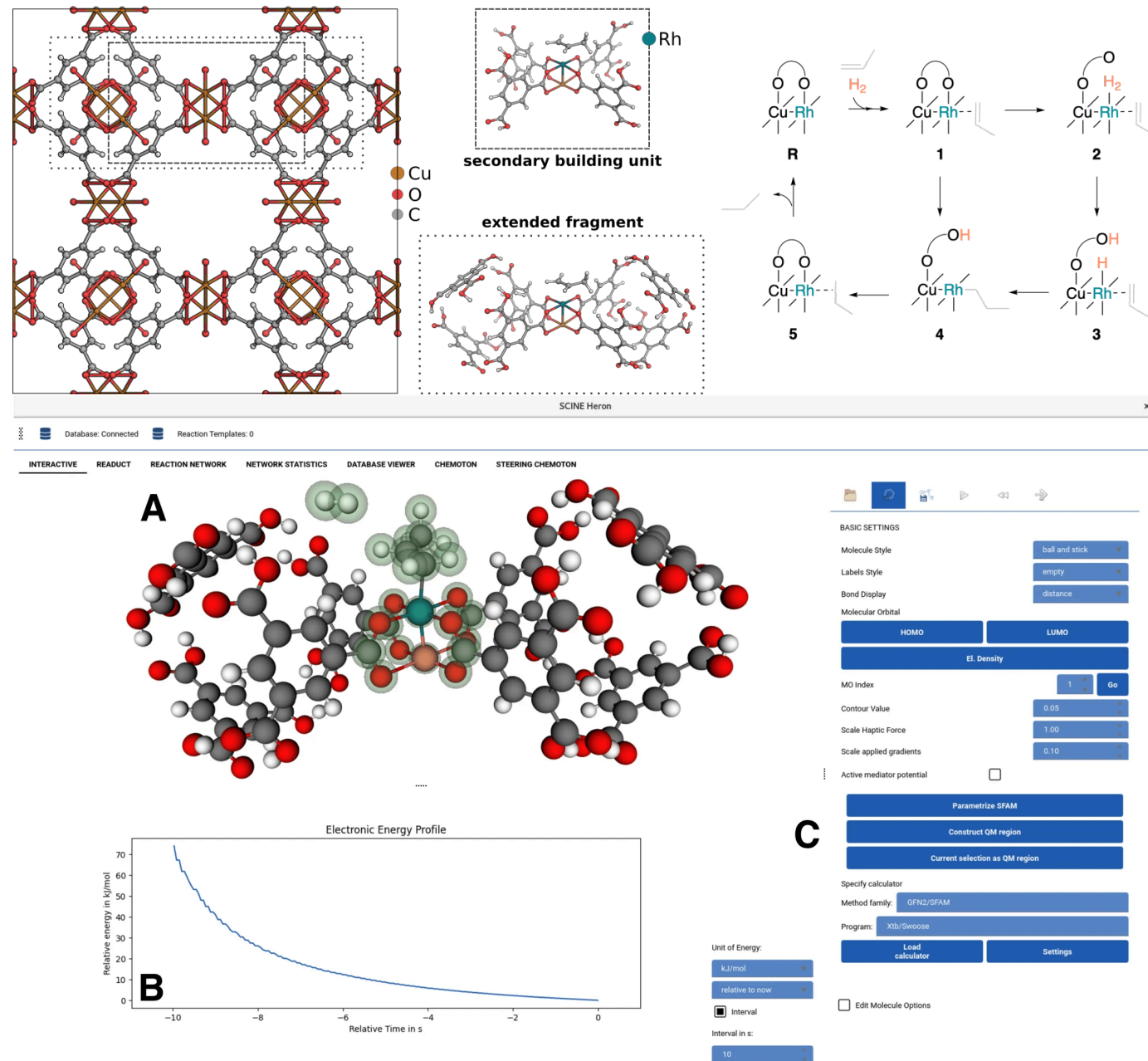


Quantum Magnifying Glass

Example for **real-time** approach:

Metal organic framework catalyst HKUST-1 analog

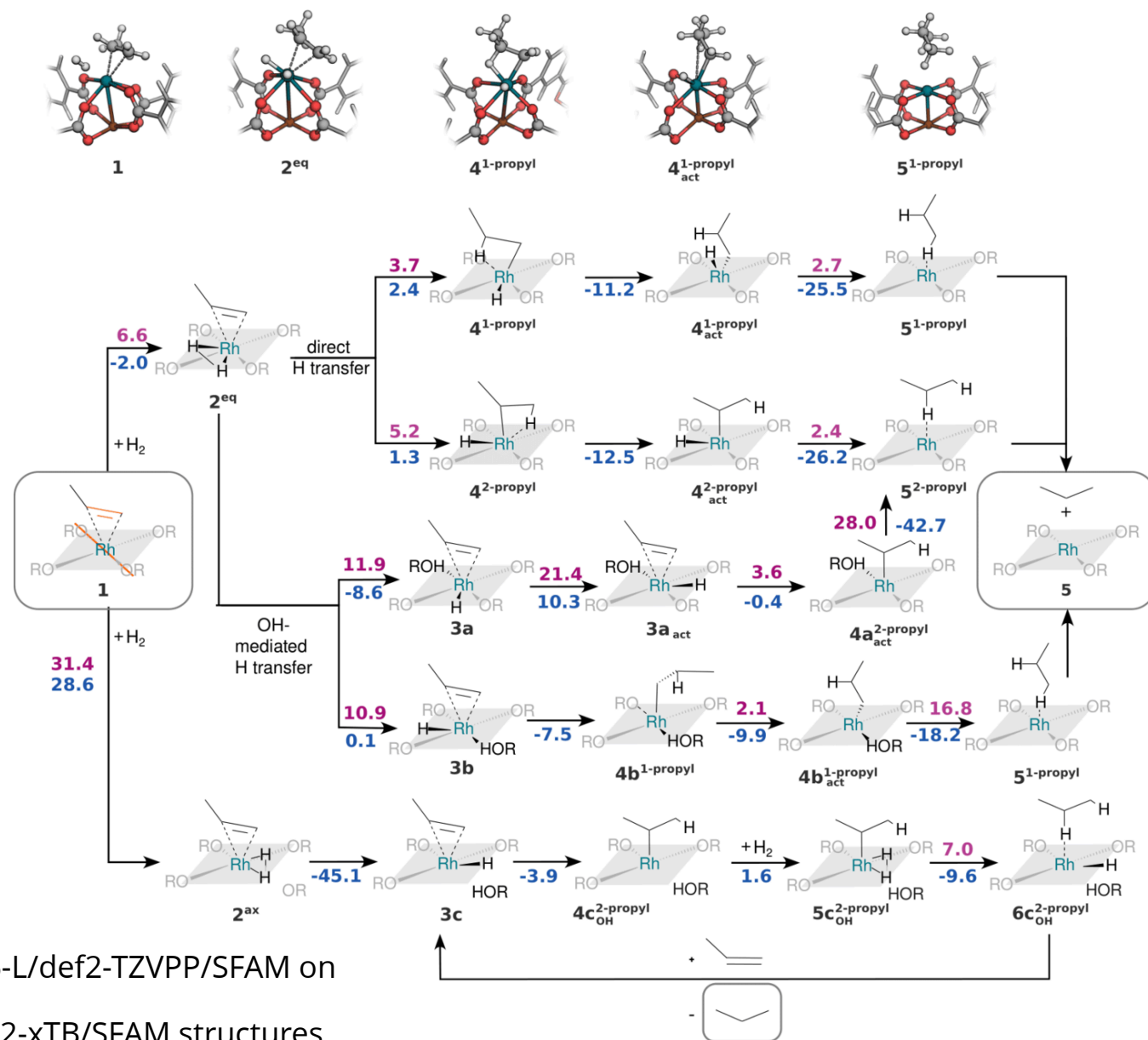
Study propylene hydrogenation with **interactive exploration**



Quantum Magnifying Glass

With **intuitive** and fast exploration, a **broader reaction space** than in previous studies can be mapped out

Able to confirm previous results and **find new path**, in agreement with experiment



M06-L/def2-TZVPP/SFAM on

GFN2-xTB/SFAM structures

Activation energy

Reaction energy

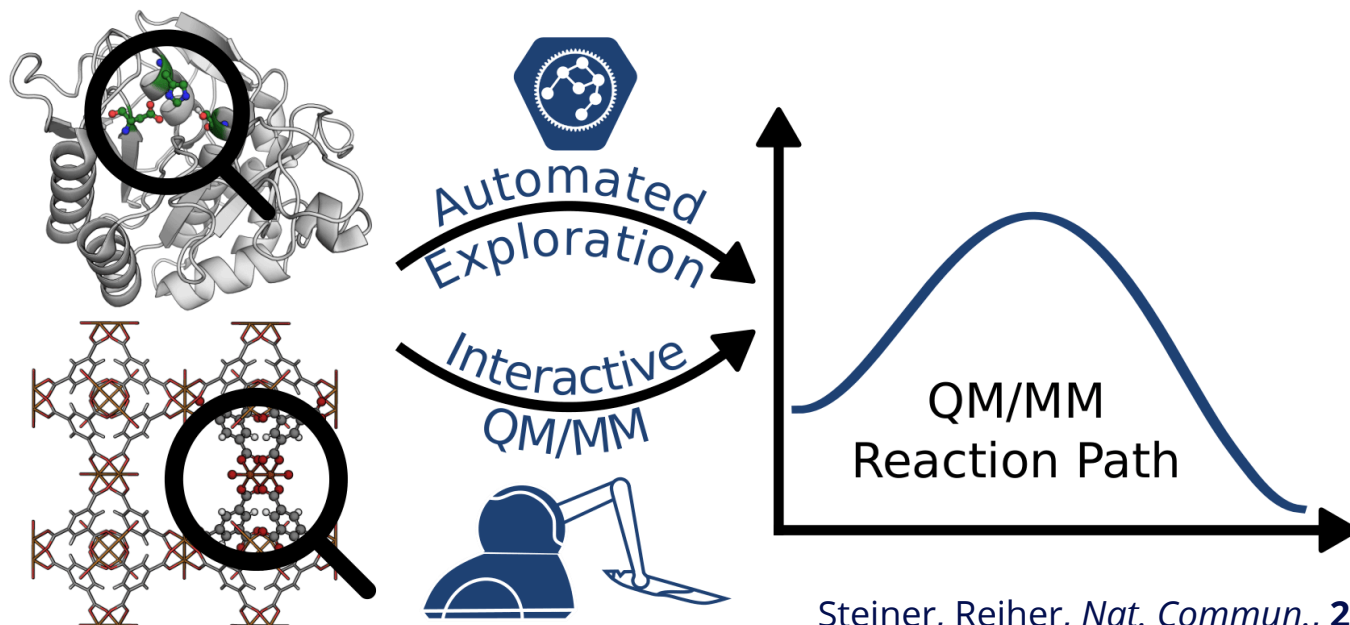
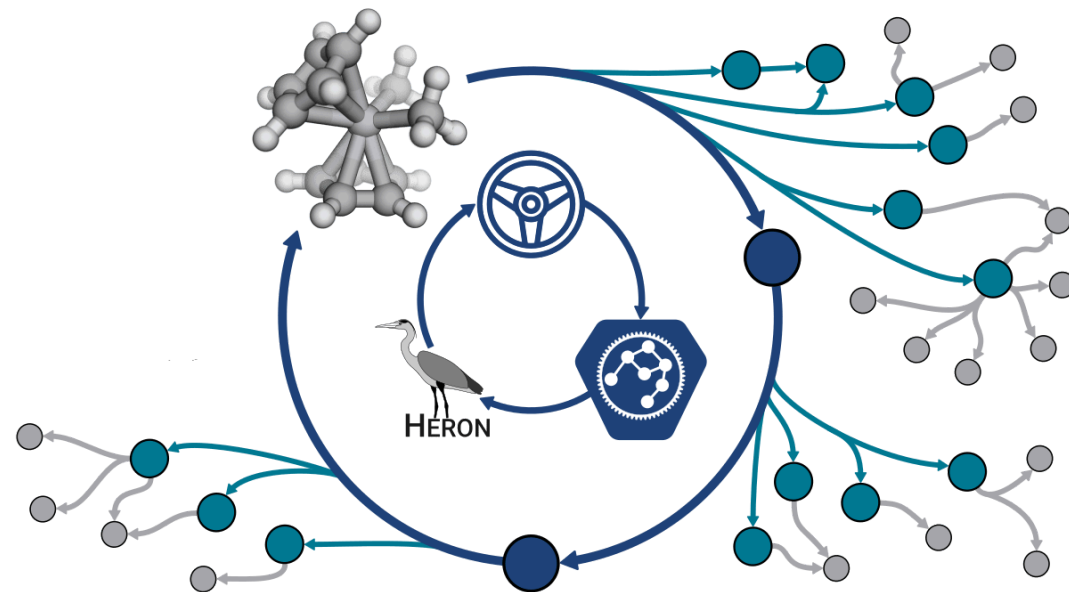
Csizi, Steiner, Reiher, *Nat. Commun.*, **2024**

Steering automated approaches

Achieves an **initial reaction network** based on previous hypothesis and on-the-fly adaptation

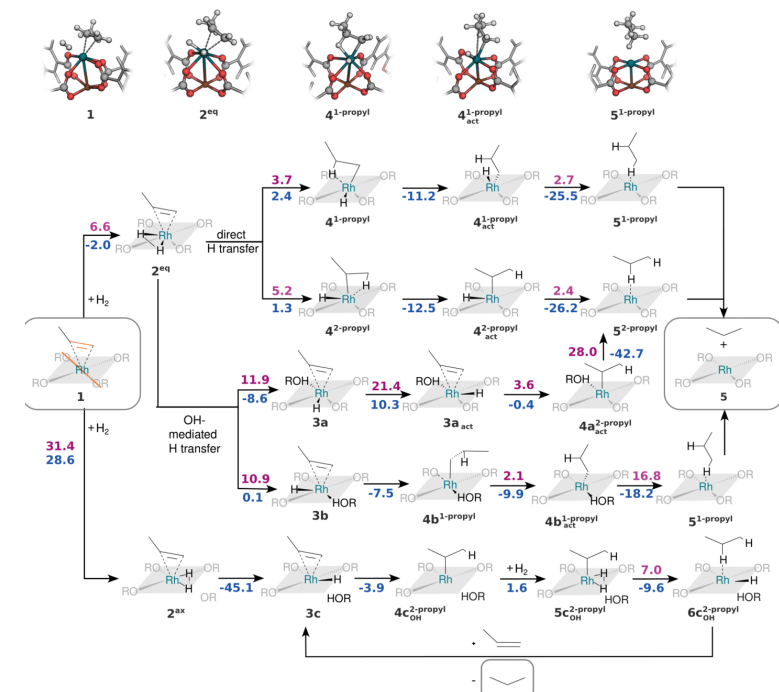
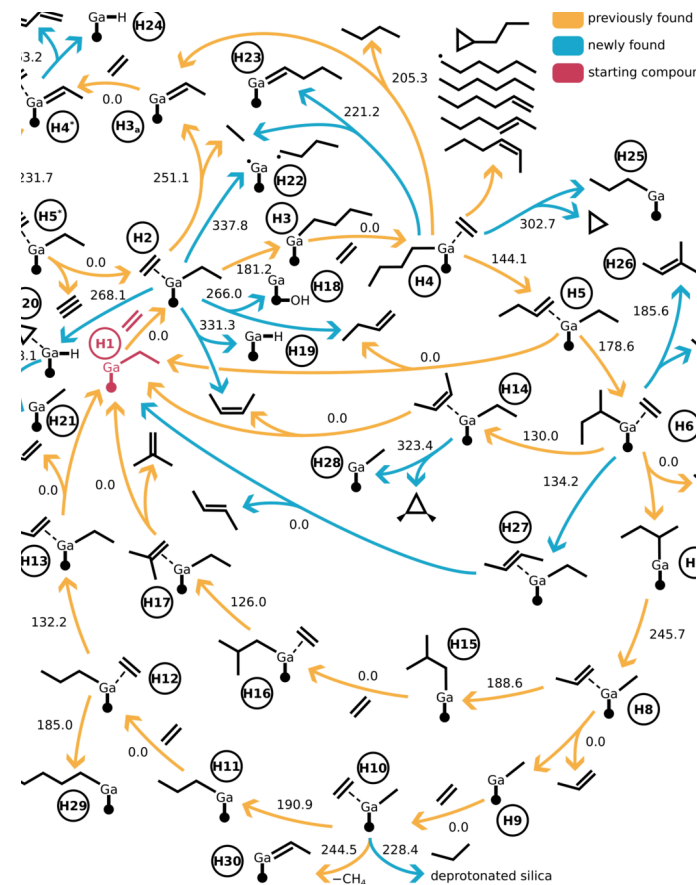
No restriction in the underlying chemistry or underlying computational potential

Broadly applicable



Learn (from) the data

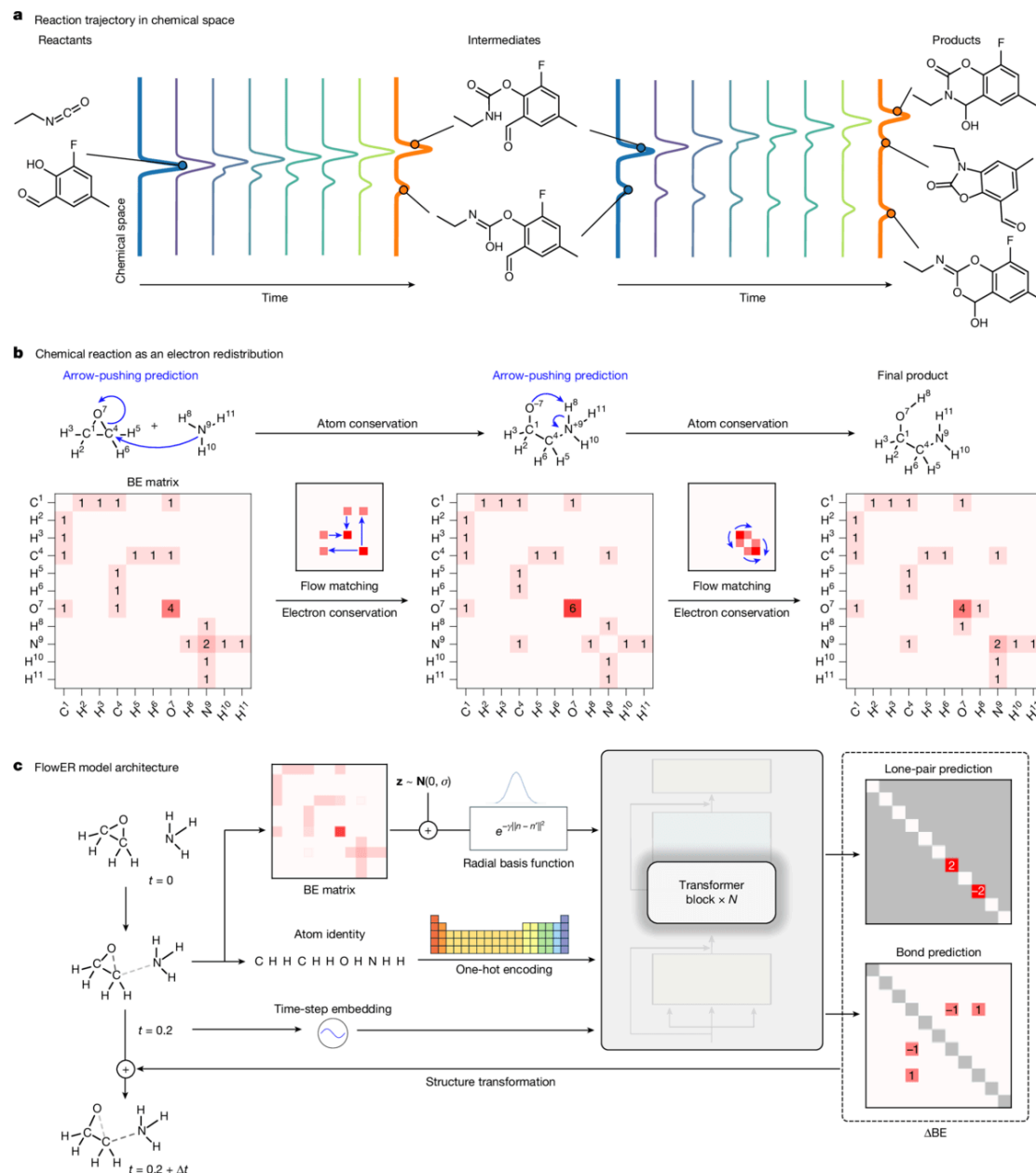
Can we **automate the guidance** as well?

Csizi, Steiner, Reiher, *Nat. Commun.*, **2024**

Mechanisms from the literature

Top-level literature based approaches **start to incorporate mechanistic approaches**

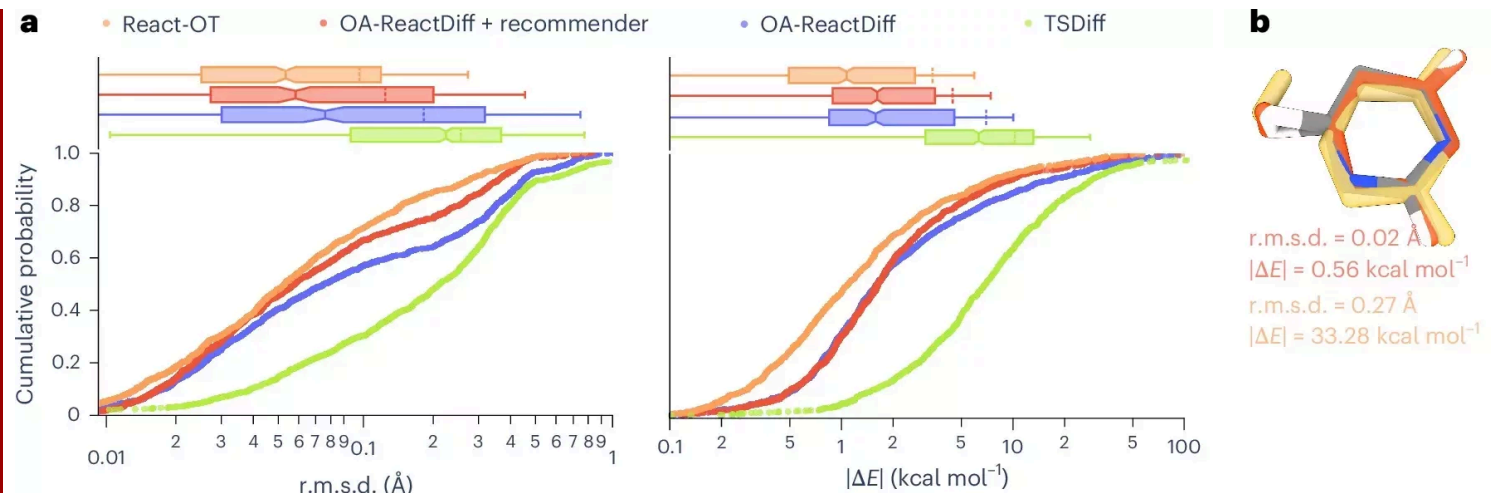
Still limited in chemical scope and **biased by available literature**, but potential for more



Can't we learn transition states directly?

With enough data **transition states can be learned**, but currently too much data required to be applicable

Need a good model for sampling and good sampling for a model



Transition-1x

Schreiner et al., *Sci. Data*, **2022**

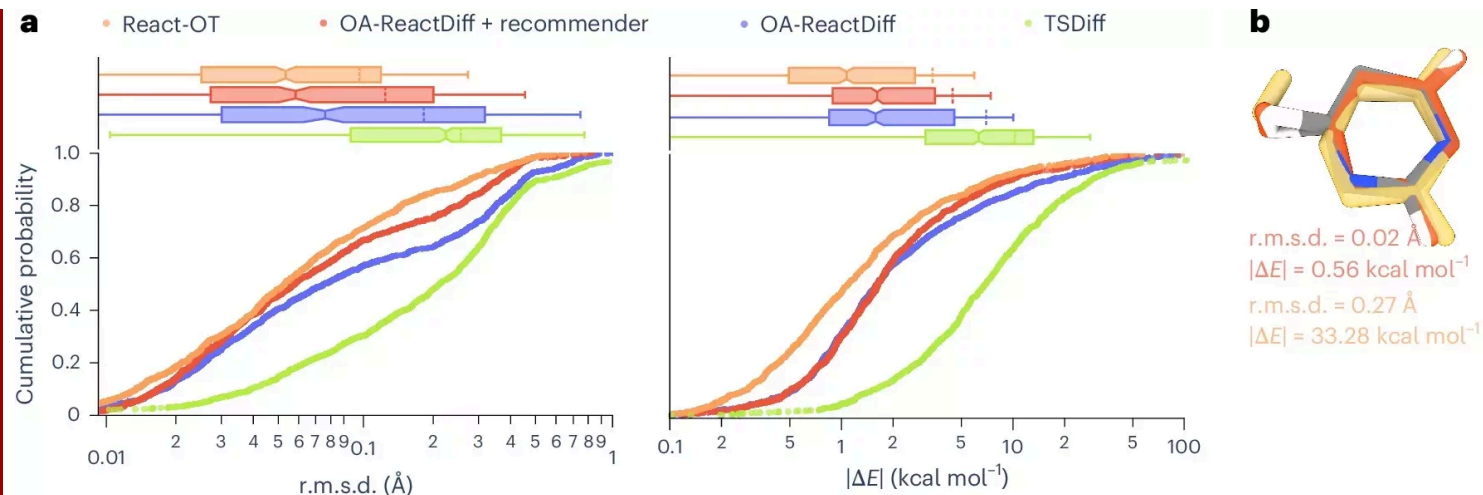
Duan et al., *Nat. Mach. Intell.*, **2025**

Can't we learn transition states directly?

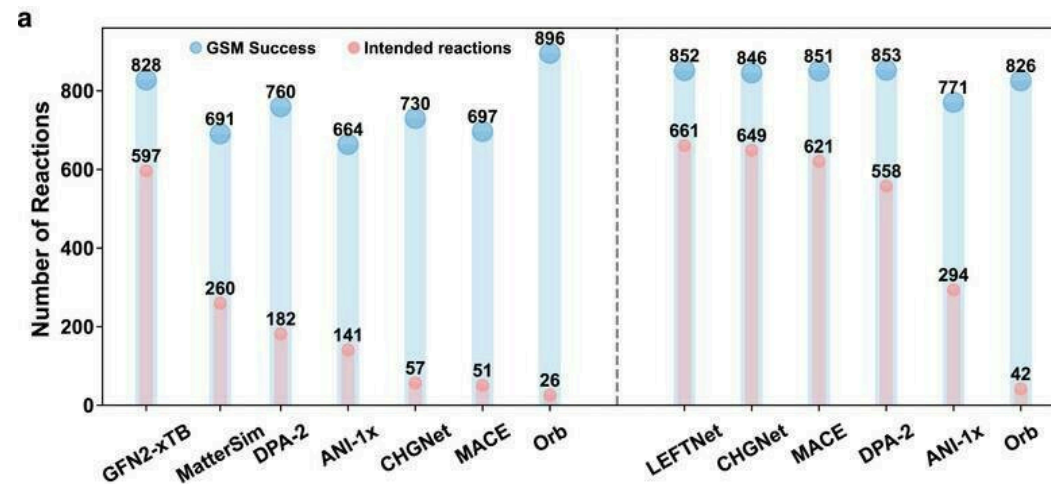
With enough data **transition states can be learned**, but currently too much data required to be applicable

Keep conventional algorithms and use MLIPs?
TS were out-of-distribution, but actually things are changing fast

Need algorithms for mechanism discovery!



Transition-1x



Schreiner et al., *Sci. Data*, **2022**

Zhao et al., *Adv. Sci.*, **2025**

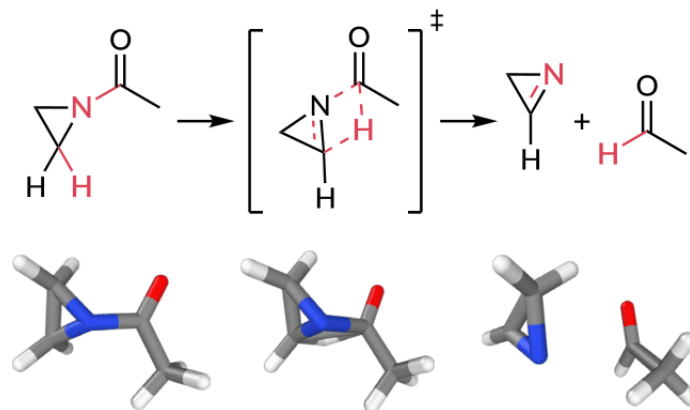
Duan et al., *Nat. Mach. Intell.*, **2025**

Circumvent chicken and egg problem

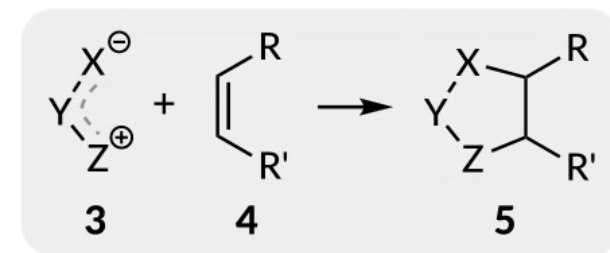
Alternative learning strategy based on **on-the-fly energy calculations** during training

Self-supervised improving model guided by **physical feedback**

Continuous path expressed as deviation from interpolation of two stable states



Transition-1x



[3+2] cycloaddition

Schreiner, Bhowmik, Vegge, Busk, Winther, *Sci. Data*, **2022**

Stuyver, Jorner, Coley, *Sci. Data*, **2023**

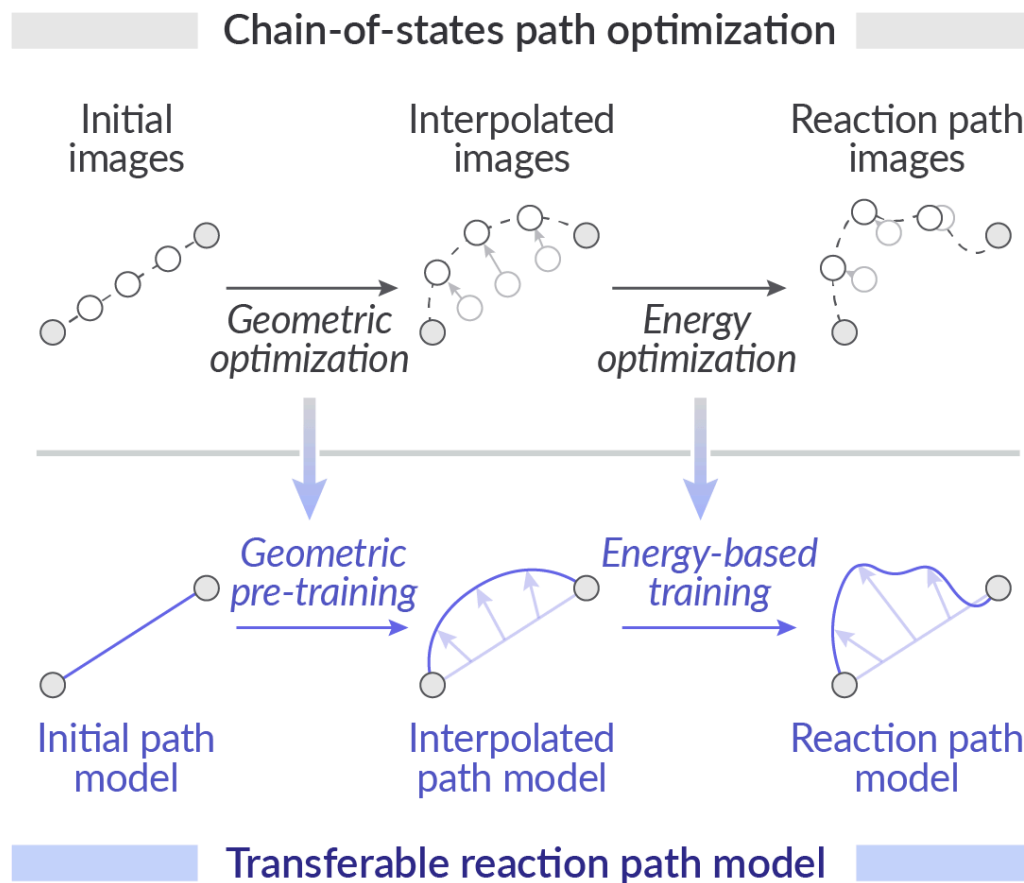
Nam, Steiner, Misterka, Yang, Singhal, Gómez-Bombarelli *J. Phys. Chem. Lett.*, **2025**

Circumvent chicken and egg problem

Alternative learning strategy based on **on-the-fly energy calculations** during training

Self-supervised improving model guided by **physical feedback**

Continuous path expressed as deviation from interpolation of two stable states



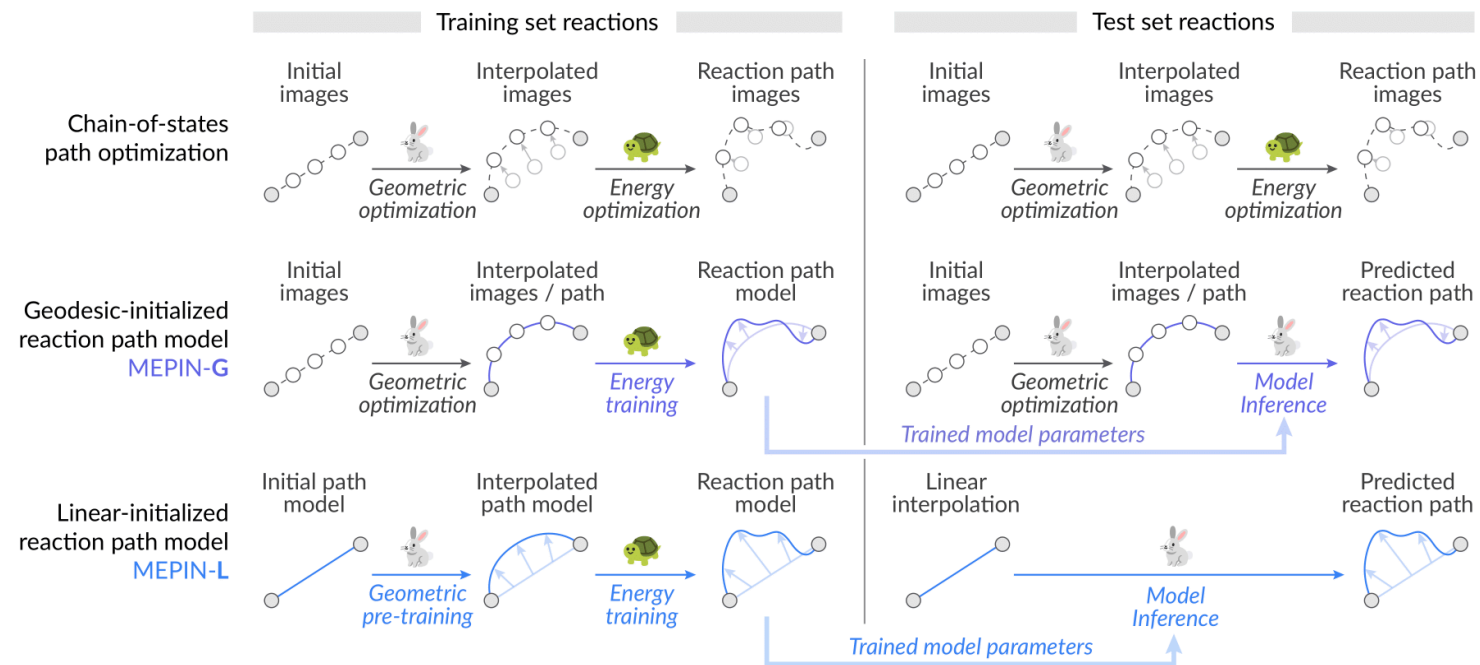
$$\mathcal{L}_{\text{flux}}[x(t)] = \frac{1}{\beta} \log \int_0^1 \exp(\beta U(x(t))) \|\dot{x}(t)\| dt$$

Circumvent chicken and egg problem

Alternative learning strategy based on **on-the-fly energy calculations** during training

Self-supervised improving model guided by **physical feedback**

Continuous path expressed as deviation from interpolation of two stable states

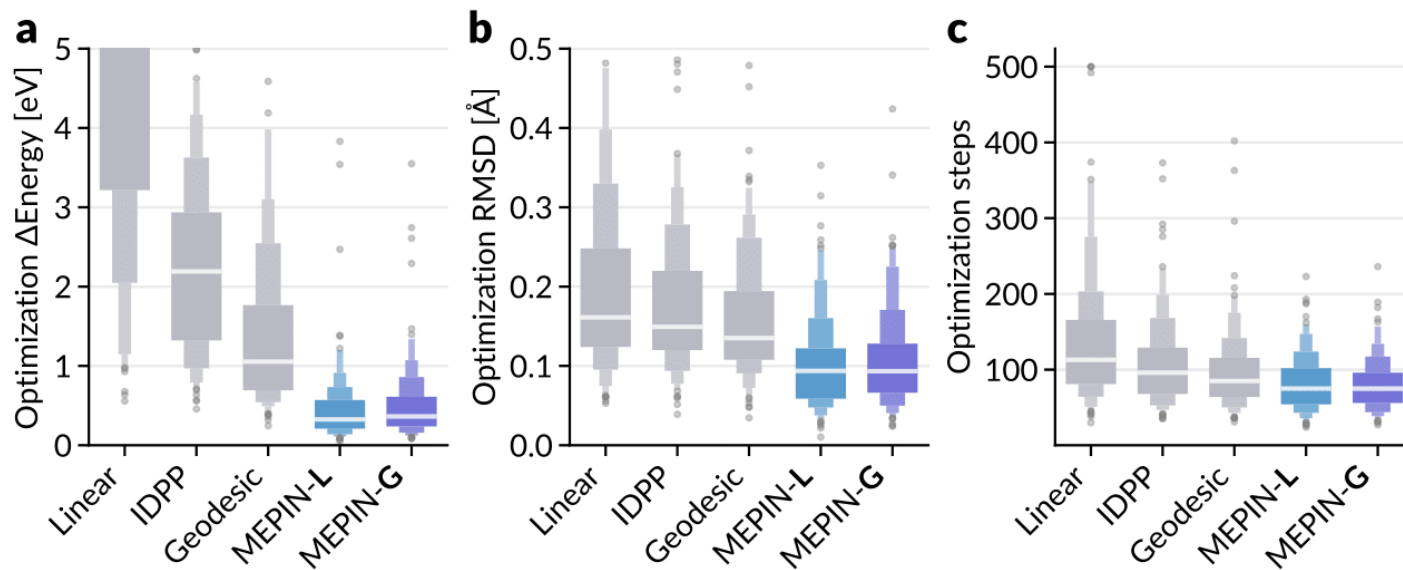


$$\mathcal{L}_{\text{flux}}[x(t)] = \frac{1}{\beta} \log \int_0^1 \exp(\beta U(x(t))) \|\dot{x}(t)\| dt$$

MEPIN benchmarks

Model can transfer learned differences to test set

Models that learn energy difference or TS directly are better, but they have more data available



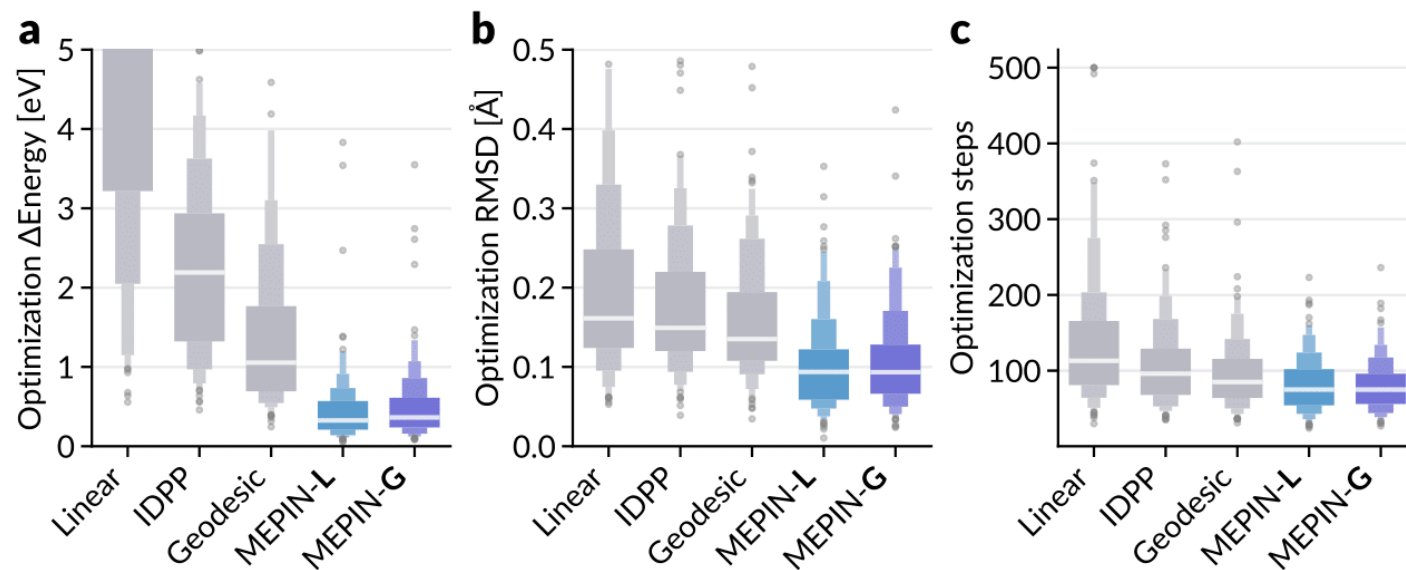
MEPIN limitations

Only time-independent calculations

Full MEP learning is expensive, and MEP not always required

Want:

Generative model for TS ensemble based on MD to be able to study more complex systems



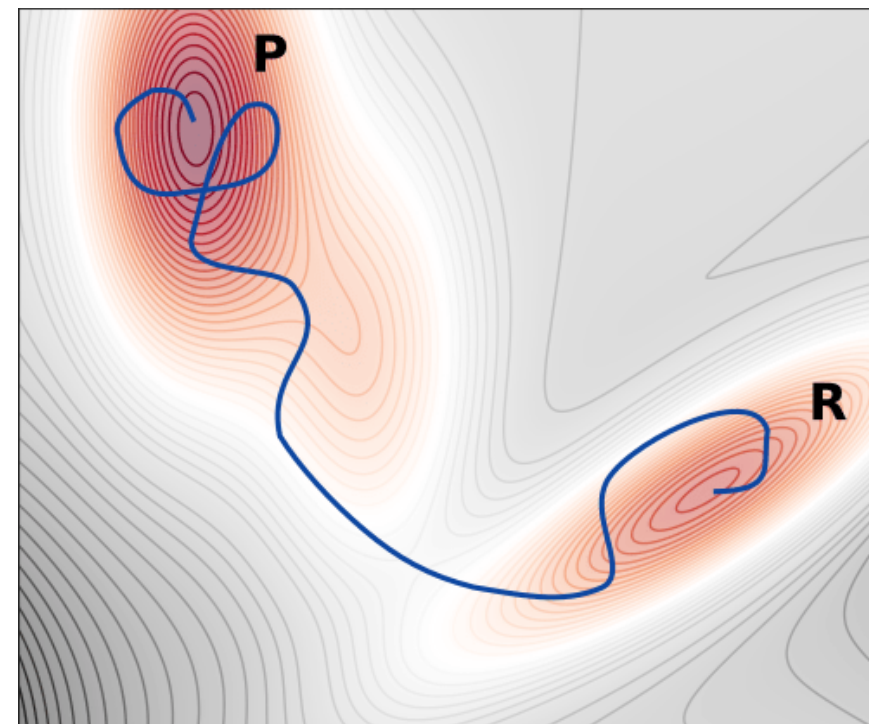
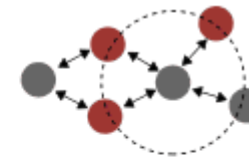
Learn reaction sampling from scratch

Learn bias potential with differentiable simulations

Batches of MD simulations biased by ML potential

Update weights to achieve transitory paths

$$U(\mathbf{x}, \theta) = U_0(\mathbf{x}) + B(\mathbf{x}, \theta)$$

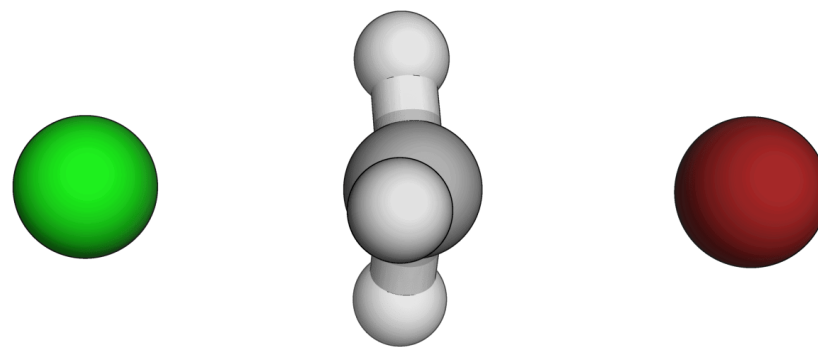
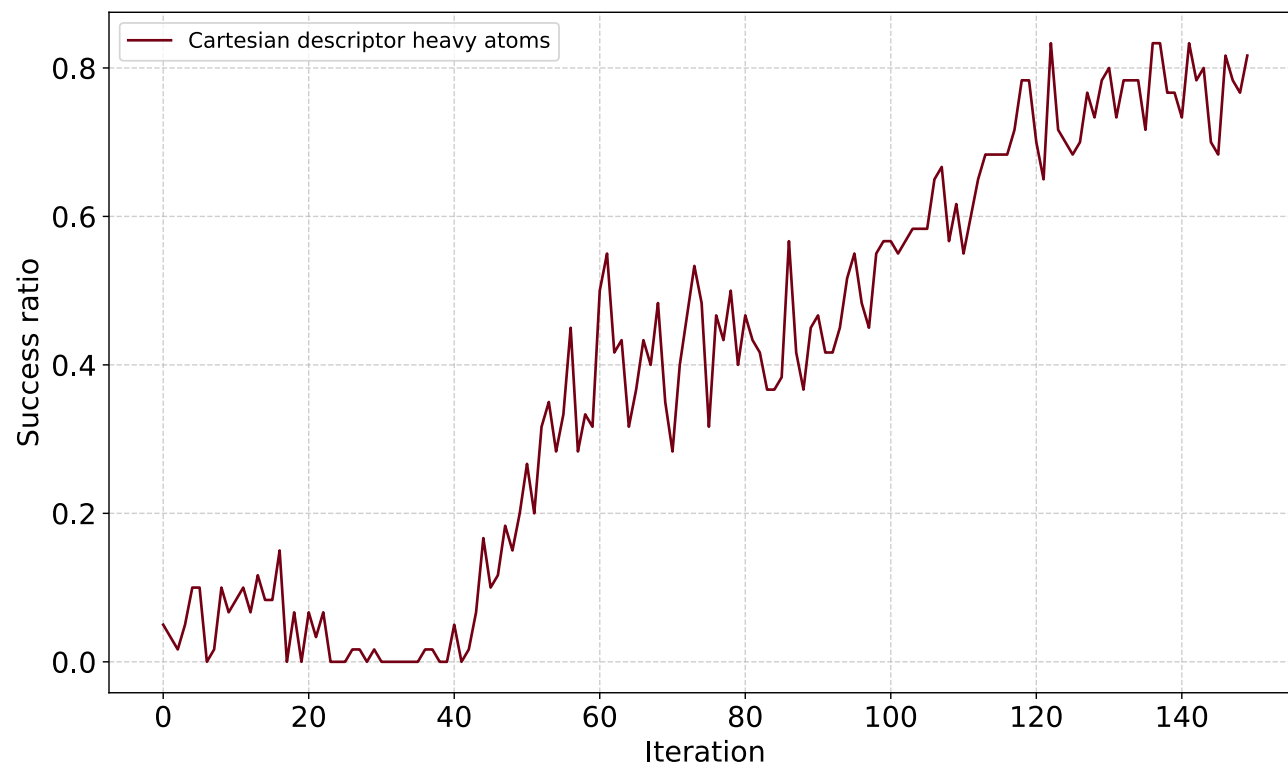


Learn reaction sampling from scratch

Learn bias potential with differentiable simulations

Batches of MD simulations biased by ML potential

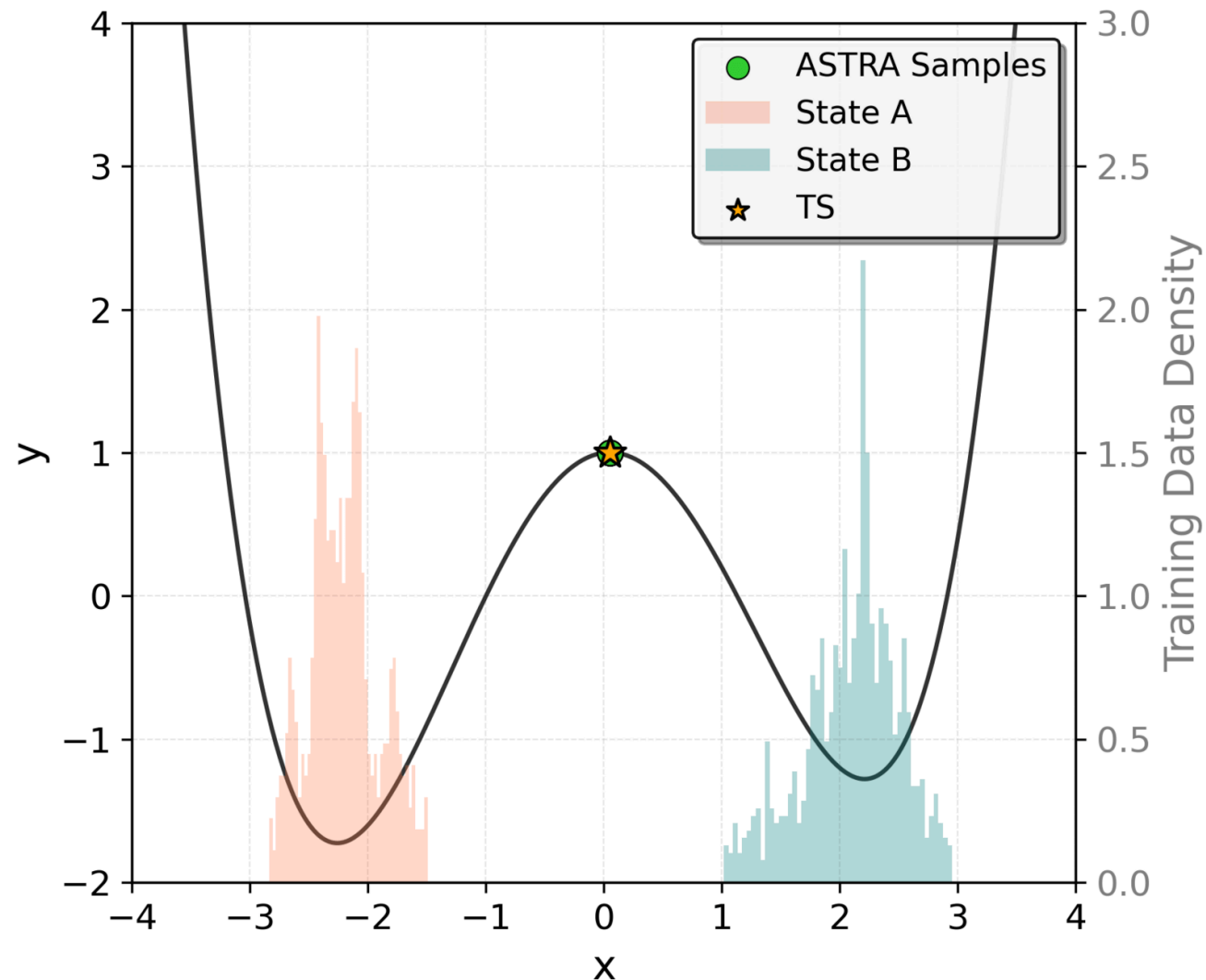
Update weights to achieve transitory paths



ASTRA

A Priori Sampling of Transition States with Guided Diffusion

Generative approach with diffusion model **driven by** state classification and **physical forces**

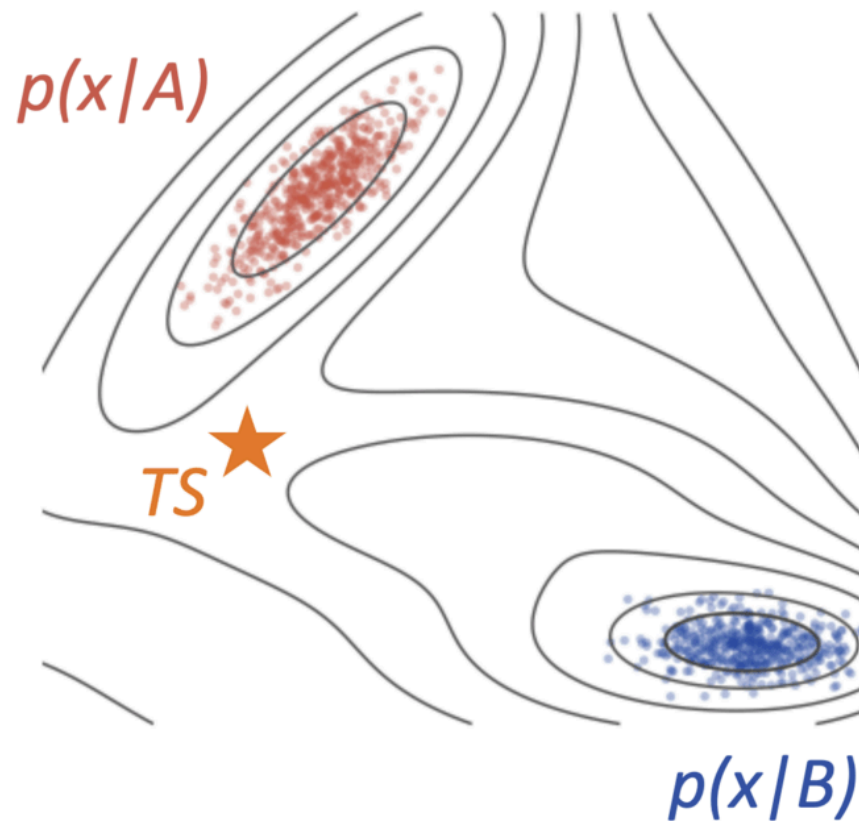


Lim, Yang, Pinede, Steiner, Du, Gómez-Bombarelli,
under review in Nat. Mach. Intell., **2026**

ASTRA

A Priori Sampling of Transition States with Guided Diffusion

Generative approach with diffusion model **driven by** state classification and **physical forces**

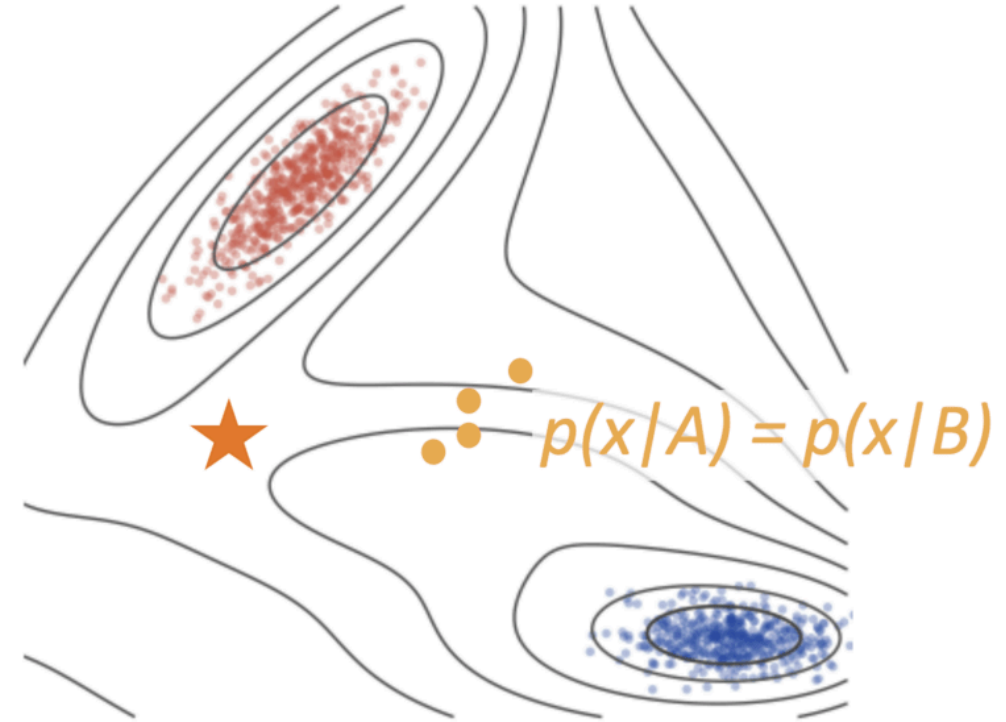


Lim, Yang, Pinede, Steiner, Du, Gómez-Bombarelli,
under review in Nat. Mach. Intell., **2026**

ASTRA

A Priori Sampling of Transition States with Guided Diffusion

Generative approach with diffusion model **driven by** state classification and **physical forces**

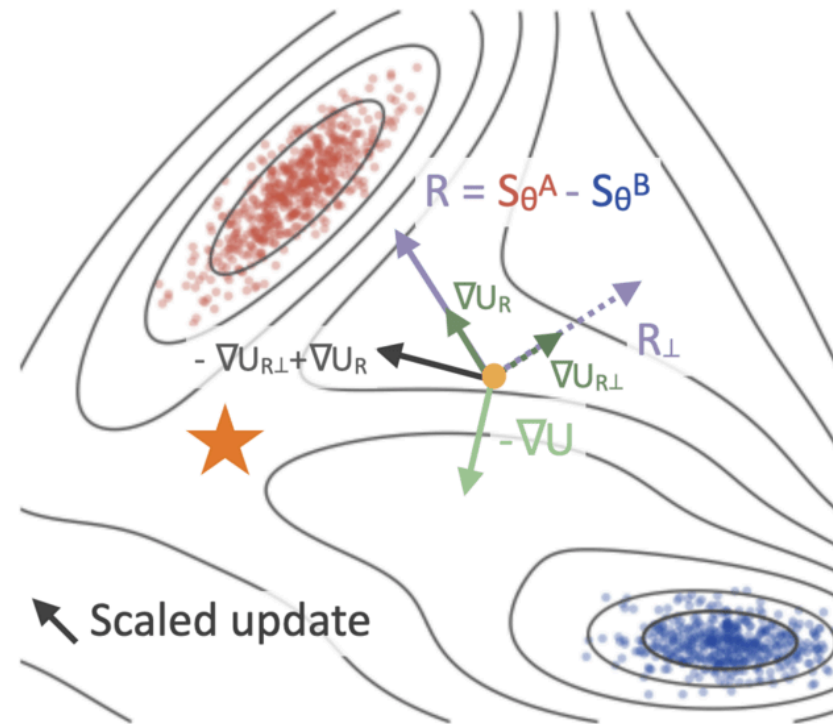


Lim, Yang, Pinede, Steiner, Du, Gómez-Bombarelli,
under review in Nat. Mach. Intell., **2026**

ASTRA

A Priori Sampling of Transition States with Guided Diffusion

Generative approach with diffusion model **driven by** state classification and **physical forces**

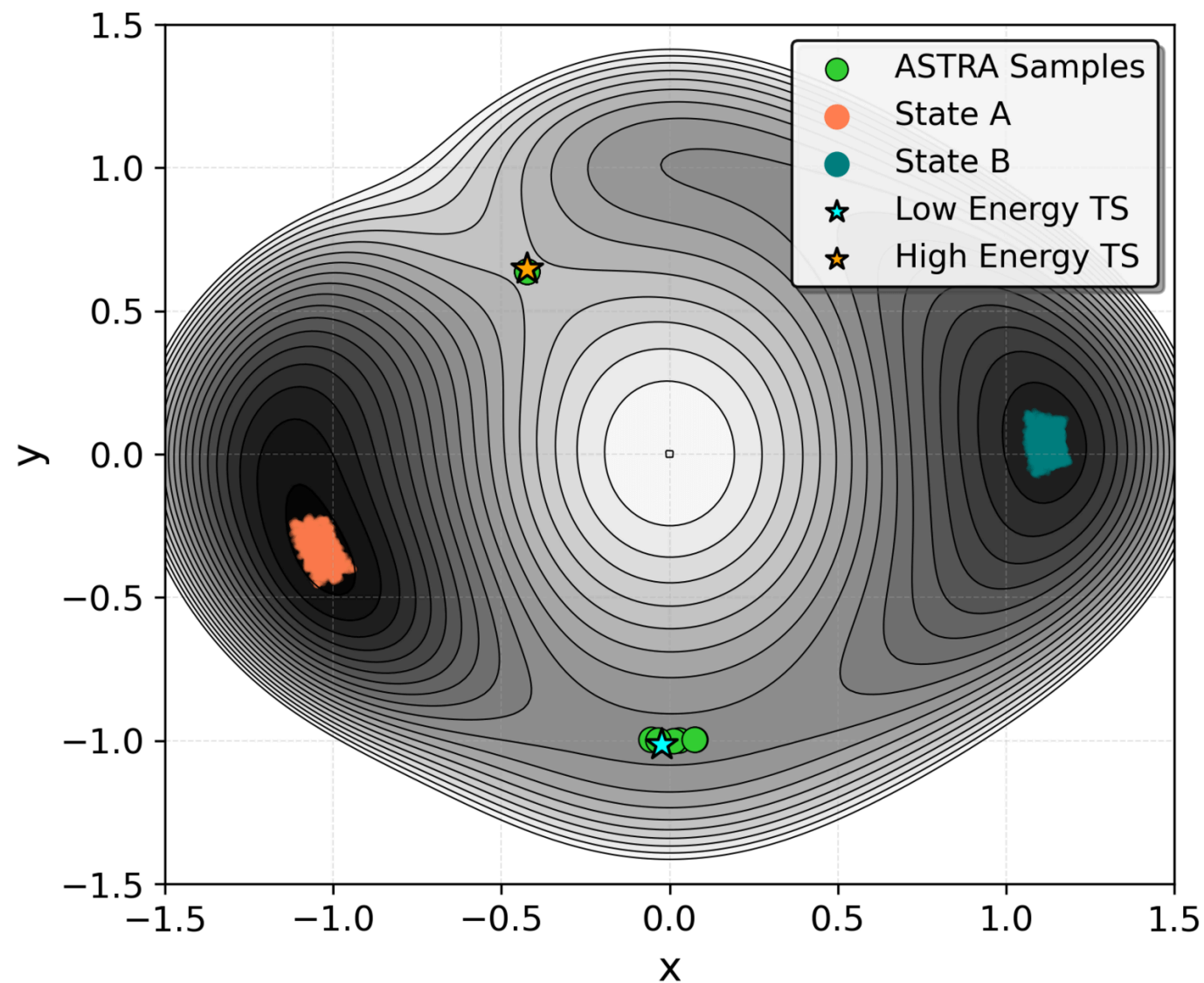


Lim, Yang, Pinede, Steiner, Du, Gómez-Bombarelli,
under review in Nat. Mach. Intell., **2026**

ASTRA

A Priori Sampling of Transition States with Guided Diffusion

Generative approach with diffusion model driven by state classification and physical forces

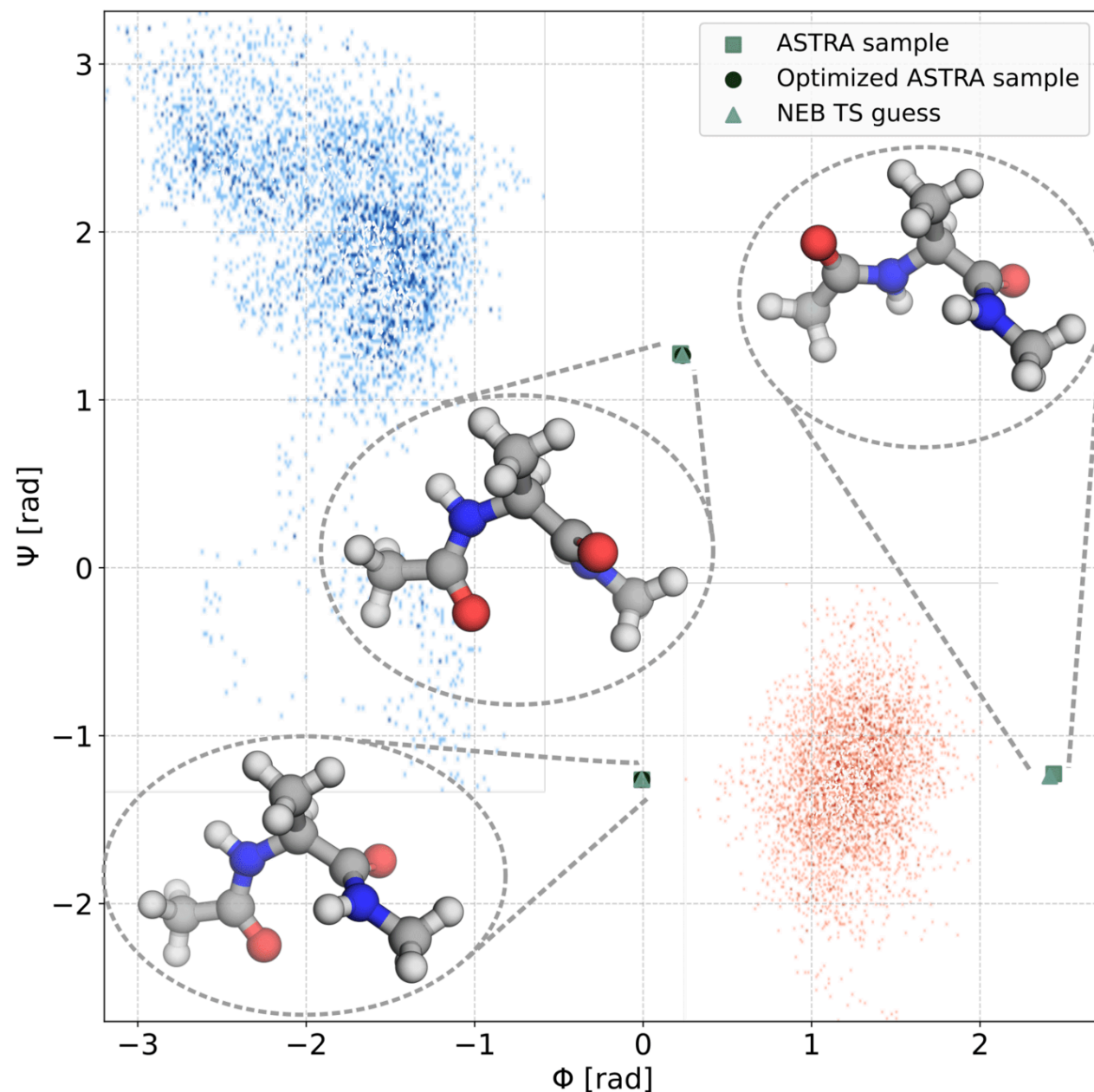


Lim, Yang, Pinede, Steiner, Du, Gómez-Bombarelli,
under review in Nat. Mach. Intell., **2026**

ASTRA

Generative approach with
diffusion model driven by state
classification and physical forces

Can reproduce baseline
optimization results for
conformational reorientation



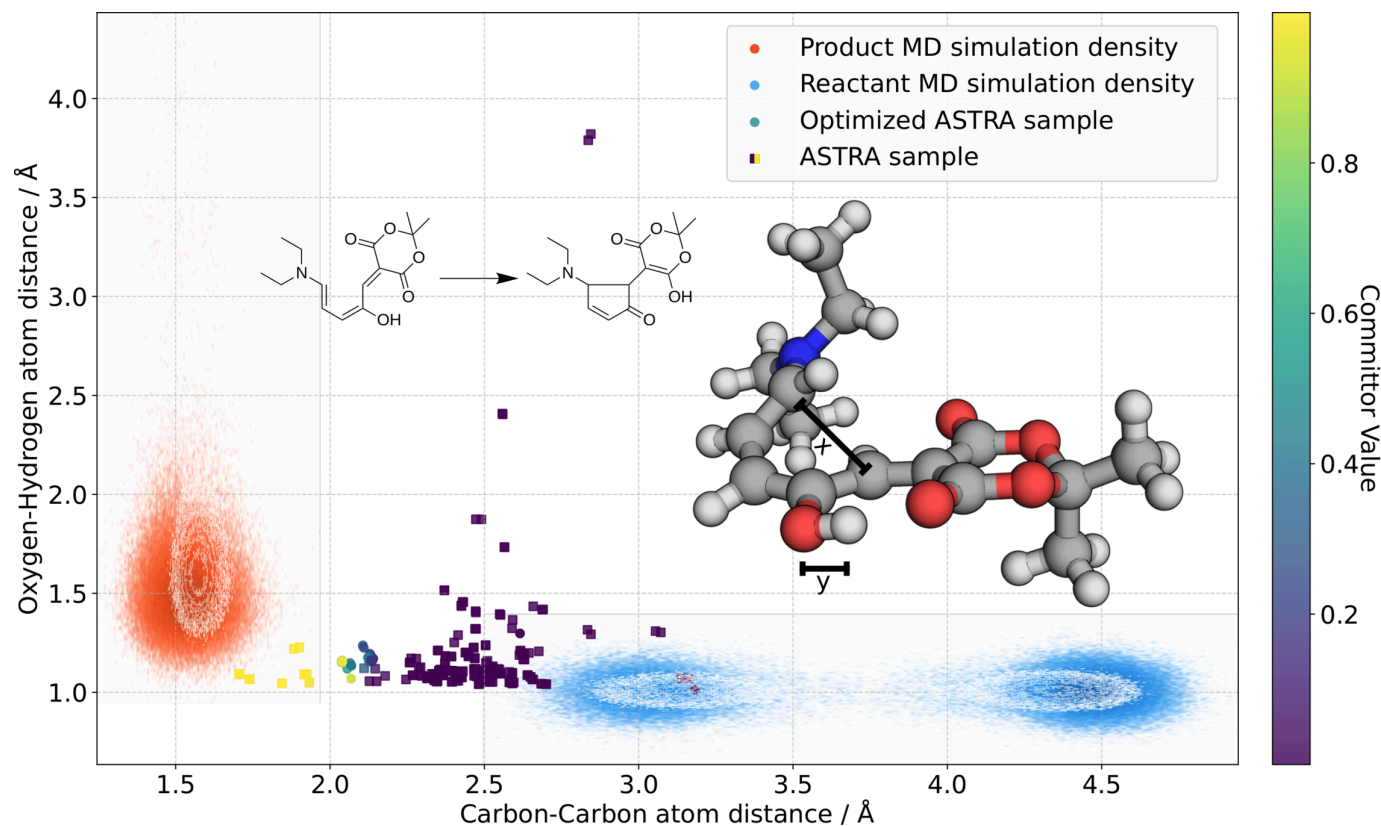
Lim, Yang, Pinede, Steiner, Du, Gómez-Bombarelli,

under review in Nat. Mach. Intell., **2026**

ASTRA

Generative approach with diffusion model driven by state classification and physical forces

Generates structures in TS region for chemical reaction



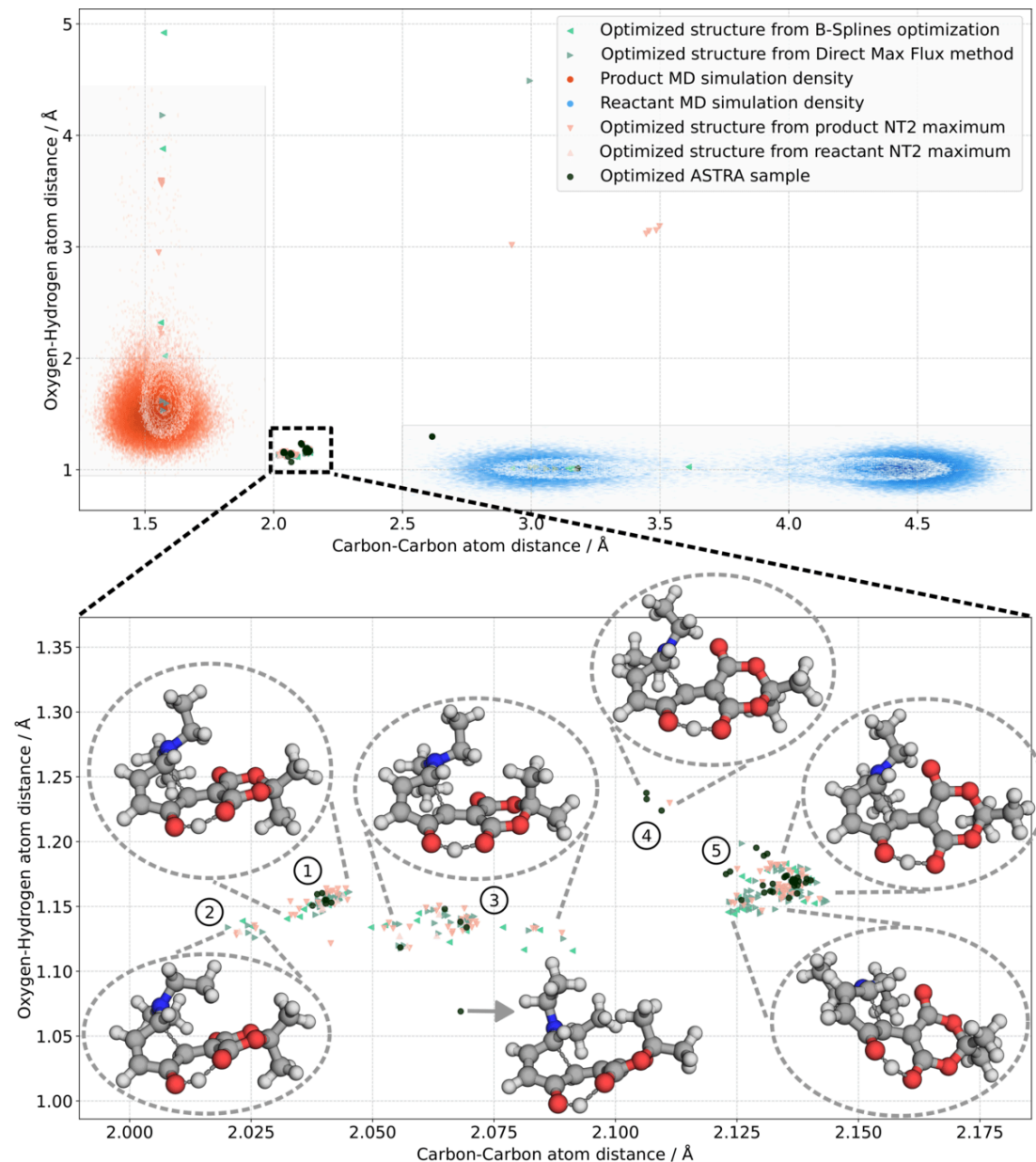
Lim, Yang, Pinede, Steiner, Du, Gómez-Bombarelli,
under review in Nat. Mach. Intell., **2026**

ASTRA

Generative approach with
diffusion model driven by state
classification and physical forces

Generates structures in TS region
for chemical reaction

Covers same conformers as
multiple baseline methods



Lim, Yang, Pinede, Steiner, Du, Gómez-Bombarelli,

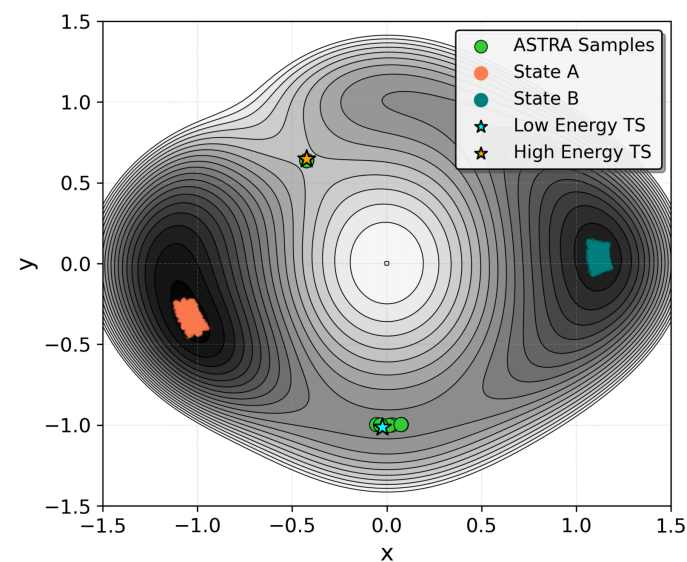
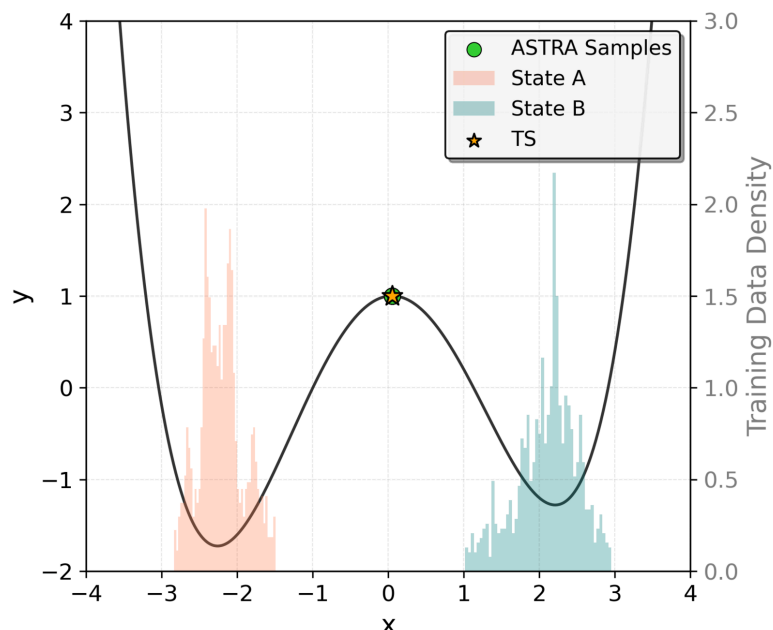
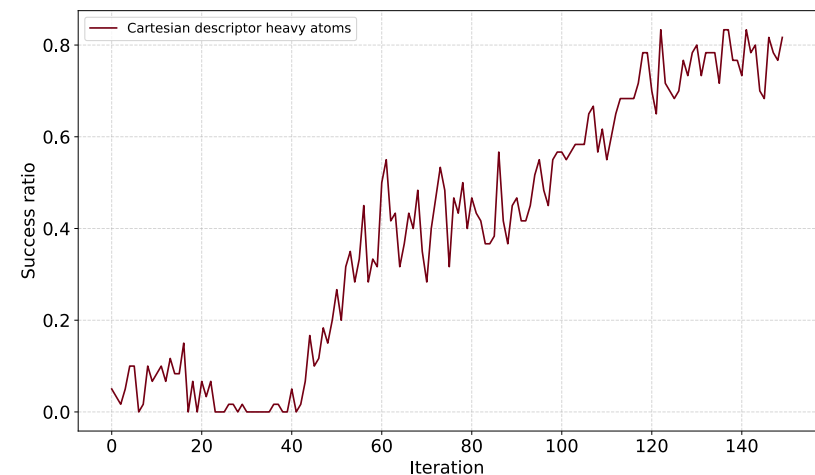
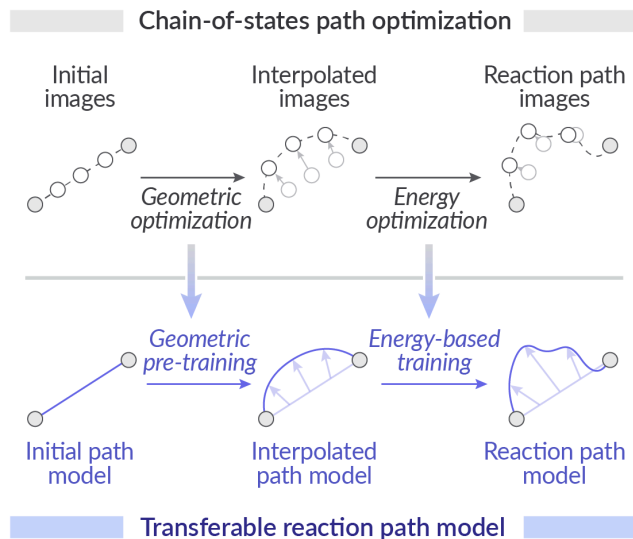
under review in Nat. Mach. Intell., **2026**

Learning reaction pathways

Can **learn collective variables** from scratch

MEPIN: **on-the-fly** data

ASTRA: Augment diffusion model with **physical forces**



Nam, Steiner, Misterka, Yang, Singhal, Gómez-Bombarelli, *J. Phys. Chem. Lett.*, **2025**
Lim, Yang, Pinede, Steiner, Du, Gómez-Bombarelli, *arXiv*, **2026**

Pioneer Center for Accelerating P2X Materials Discovery

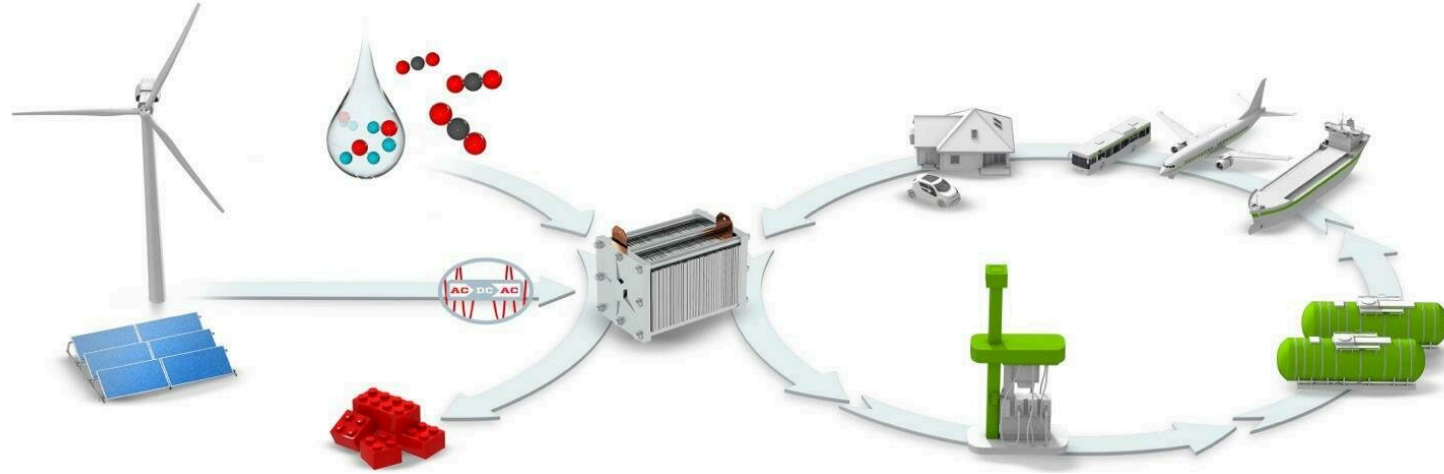
Accelerating Materials Discovery

Green transition needs more efficient, durable and scalable advanced materials for Power-to-X

Dependency on scarce materials, and too low efficiency and reliability

13 year Pioneer Center to accelerate P2X materials discovery

Tejs Vegge, DTU
Frede Blaabjerg, AAU



Four Complementary “X-trails”

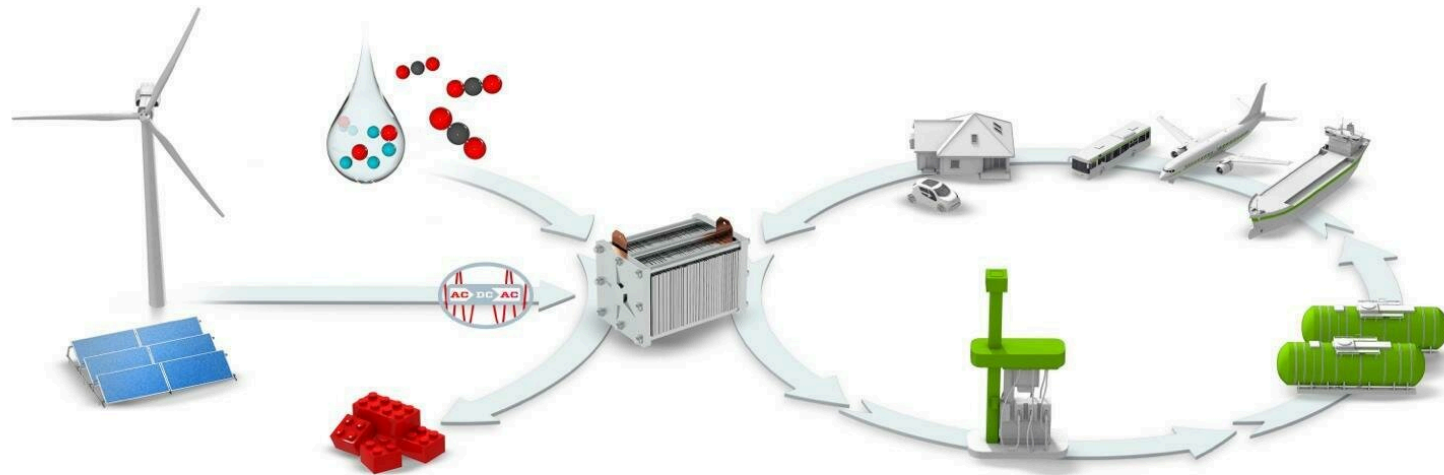
Focus on electrocatalysis at intermediate temperatures and pressures

X1: Water electrolysis and hydrogen production

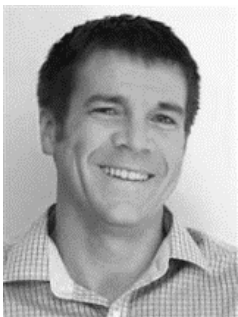
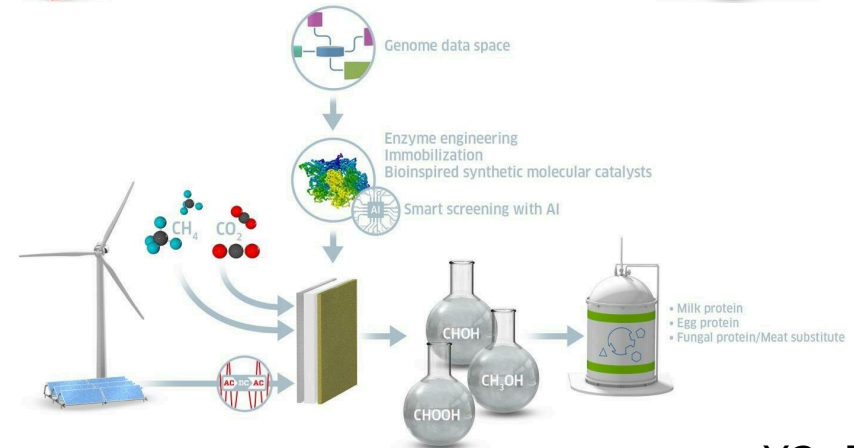
X2: Liquid fuels for aviation and maritime applications

X3: Sustainable commodity and specialty chemicals

X4: Electro-biocatalytic productions of chemicals, fuels & protein



X1: Søren H. Jensen,
AAU Energy



X2: Brian Seger,
DTU Physics



X3: Maria Escudero-Escribano,
ICN2



X4: Anne S. Meyer,
DTU Bioengineering

Three Methodological Research Themes

Develop new methods for materials discovery

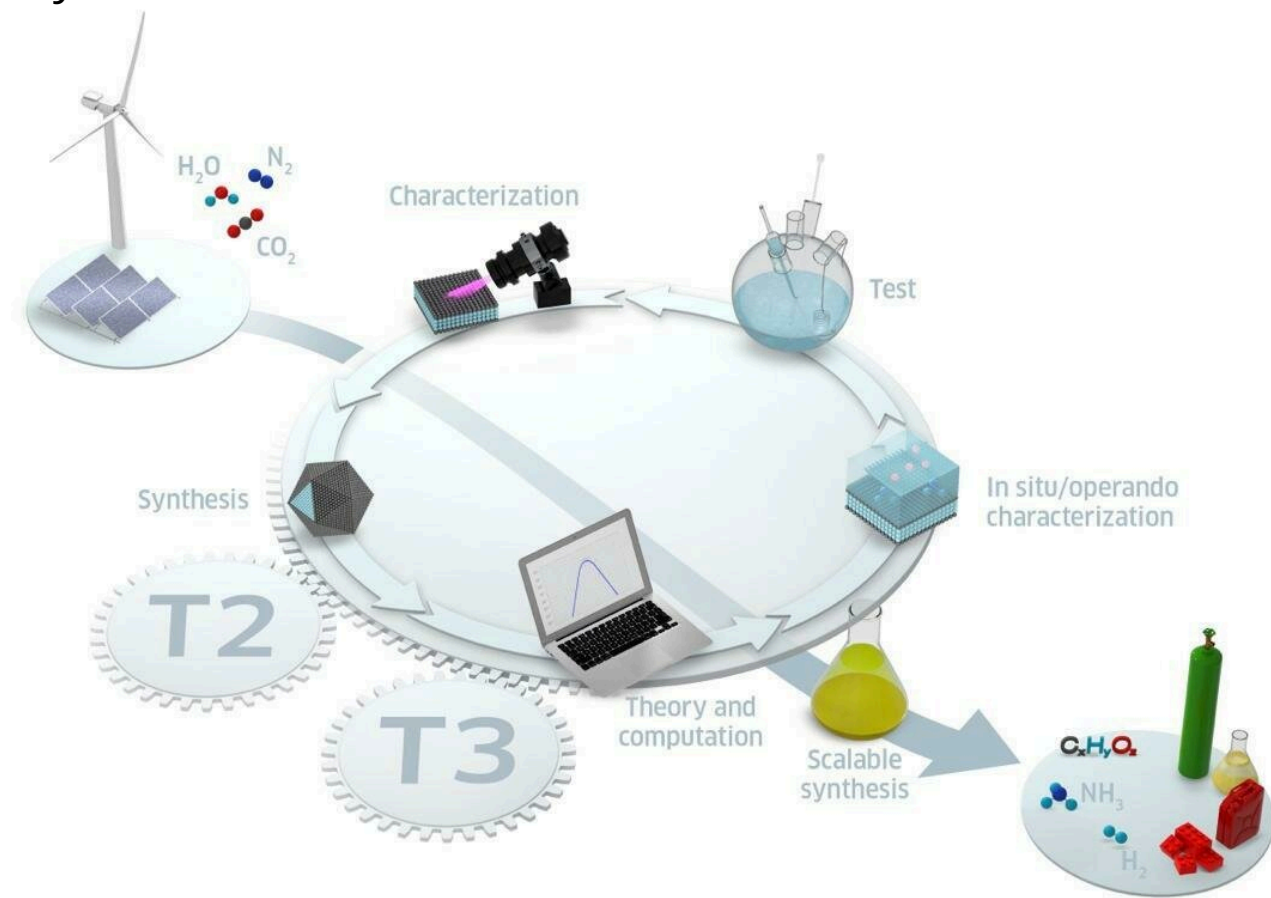
T1: An Integrated Discovery Loop

T2: An asynchronous Materials Acceleration Platform

T3: A Digital Twin across scales



T1: Jakob Kibsgaard,
DTU Physics



Three Methodological Research Themes

Develop new methods for materials discovery

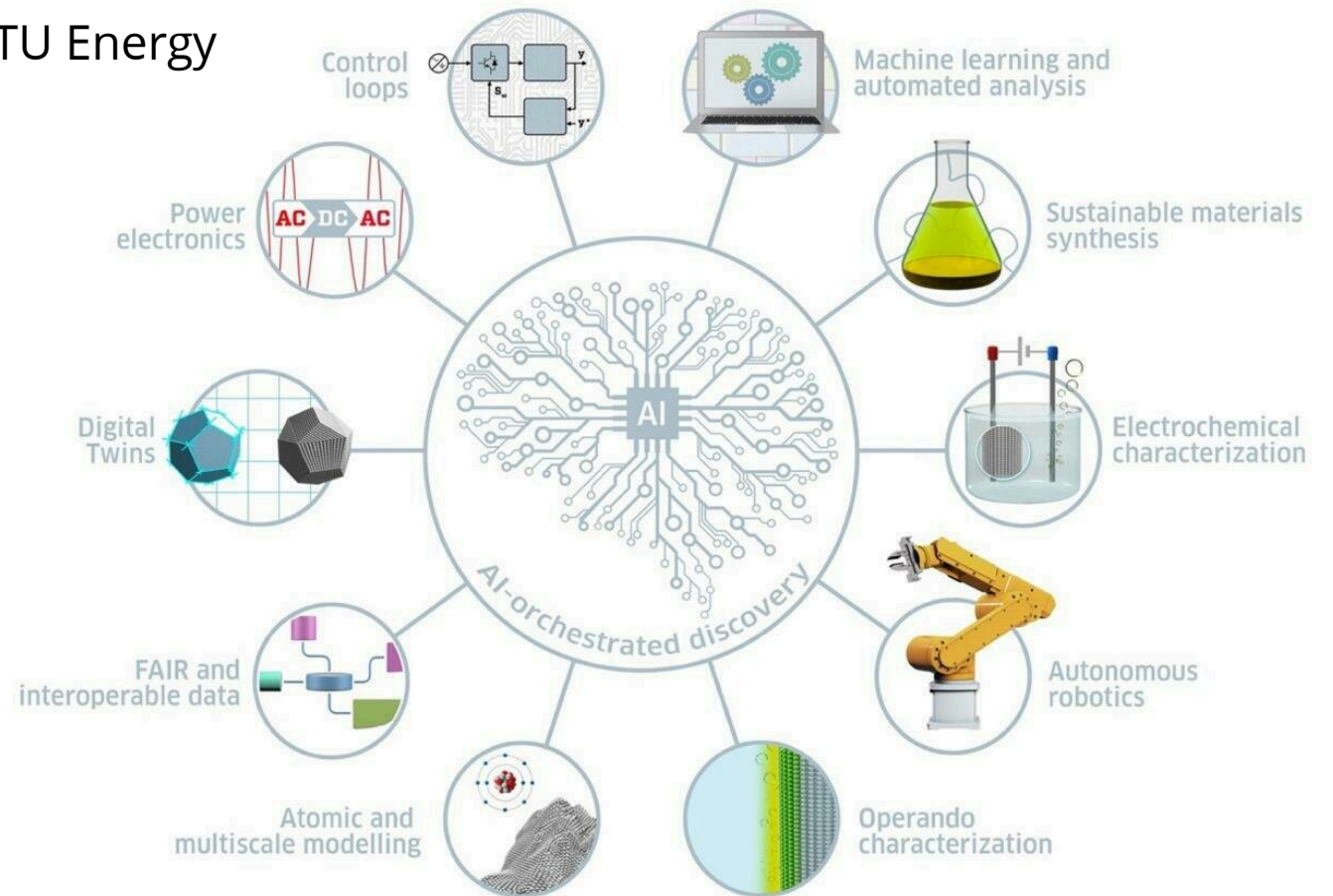
T1: An Integrated Discovery Loop

T2: An asynchronous Materials Acceleration Platform

T3: A Digital Twin across scales



T2: Tejs Vegge,
DTU Energy



Three Methodological Research Themes

Develop new methods for materials discovery

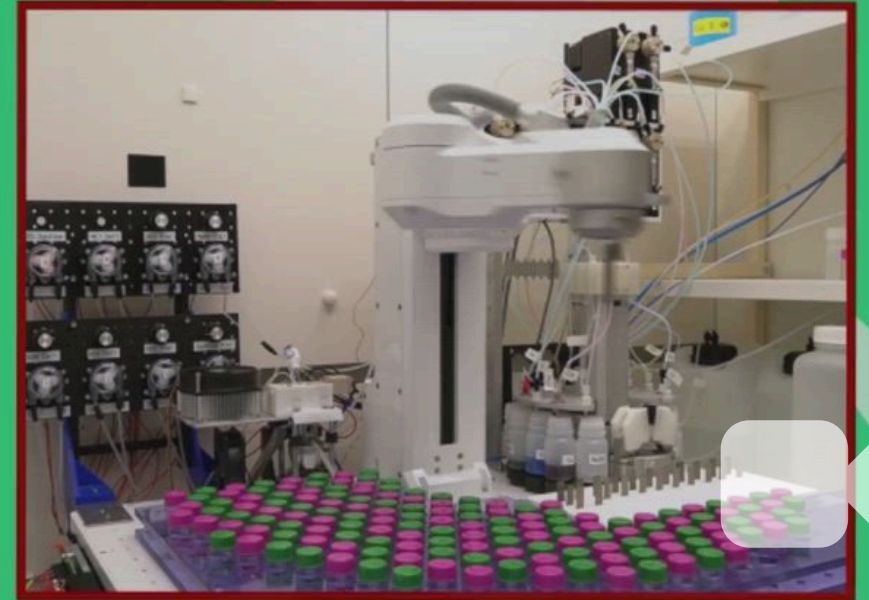
T1: An Integrated Discovery Loop

T2: An asynchronous Materials Acceleration Platform

T3: A Digital Twin across scales



T2: Tejs Vegge,
DTU Energy



Three Methodological Research Themes

Develop new methods for materials discovery

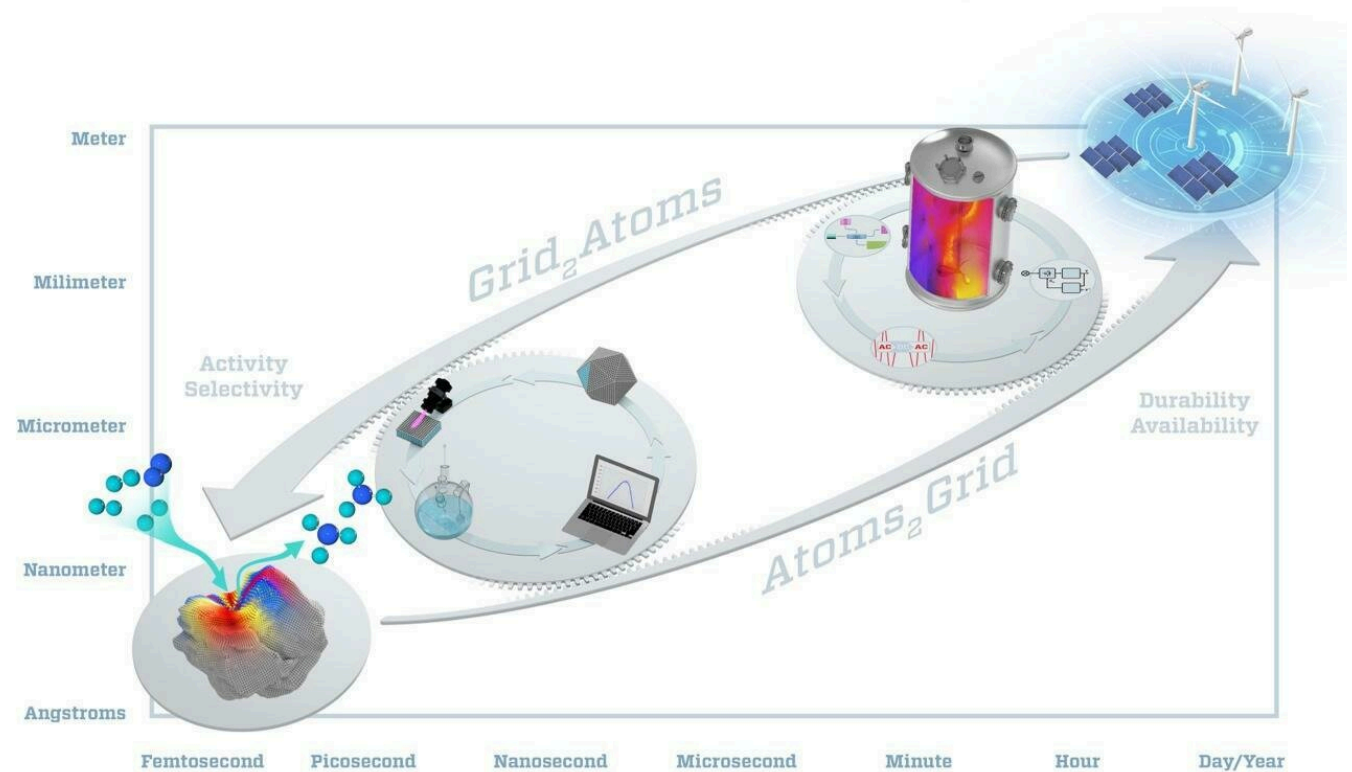
T1: An Integrated Discovery Loop

T2: An asynchronous Materials Acceleration Platform

T3: A Digital Twin across scales



T3: Frede Blaabjerg,
AAU Energy



Three Methodological Research Themes

Develop new methods for materials discovery

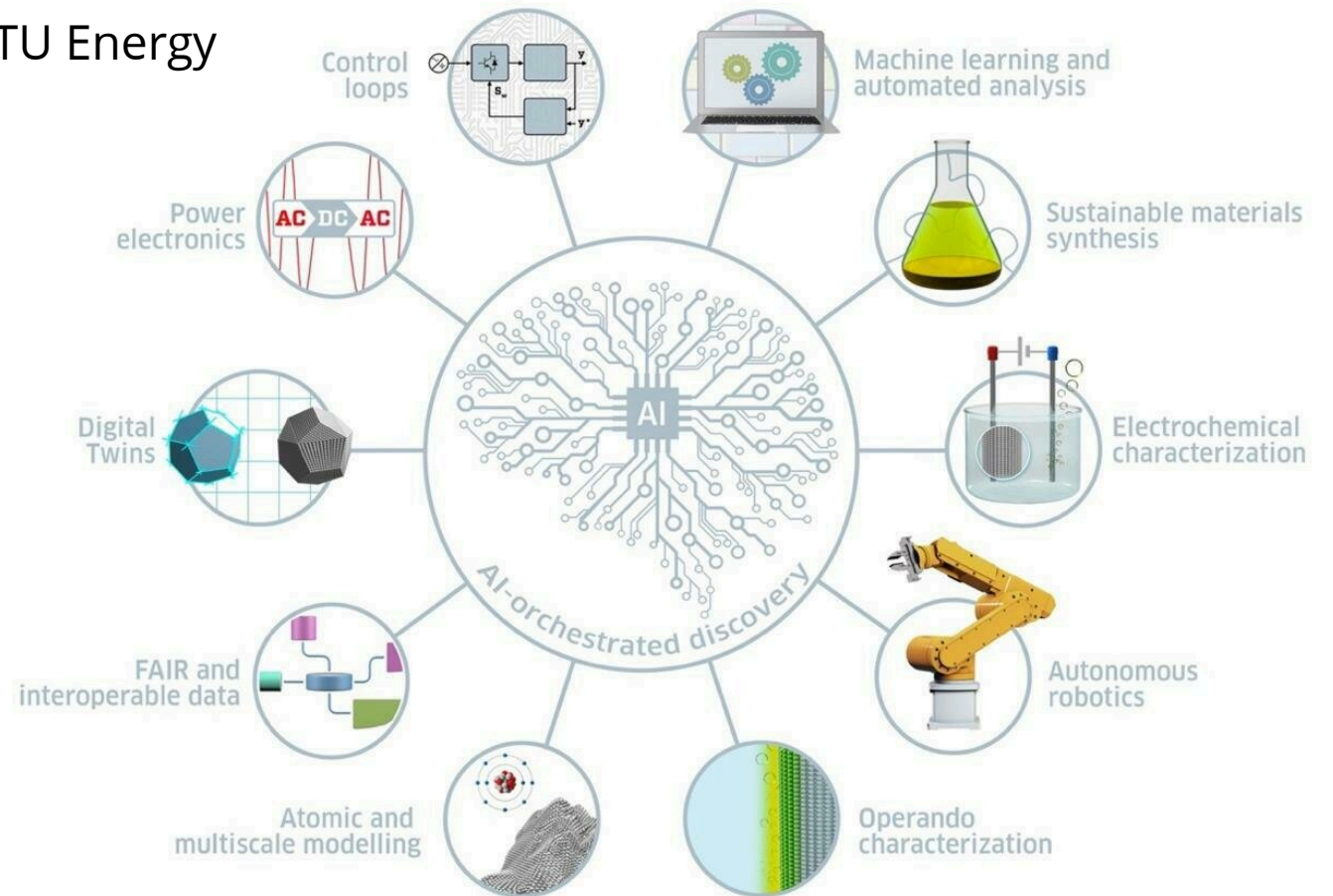
T1: An Integrated Discovery Loop

T2: An asynchronous Materials Acceleration Platform

T3: A Digital Twin across scales



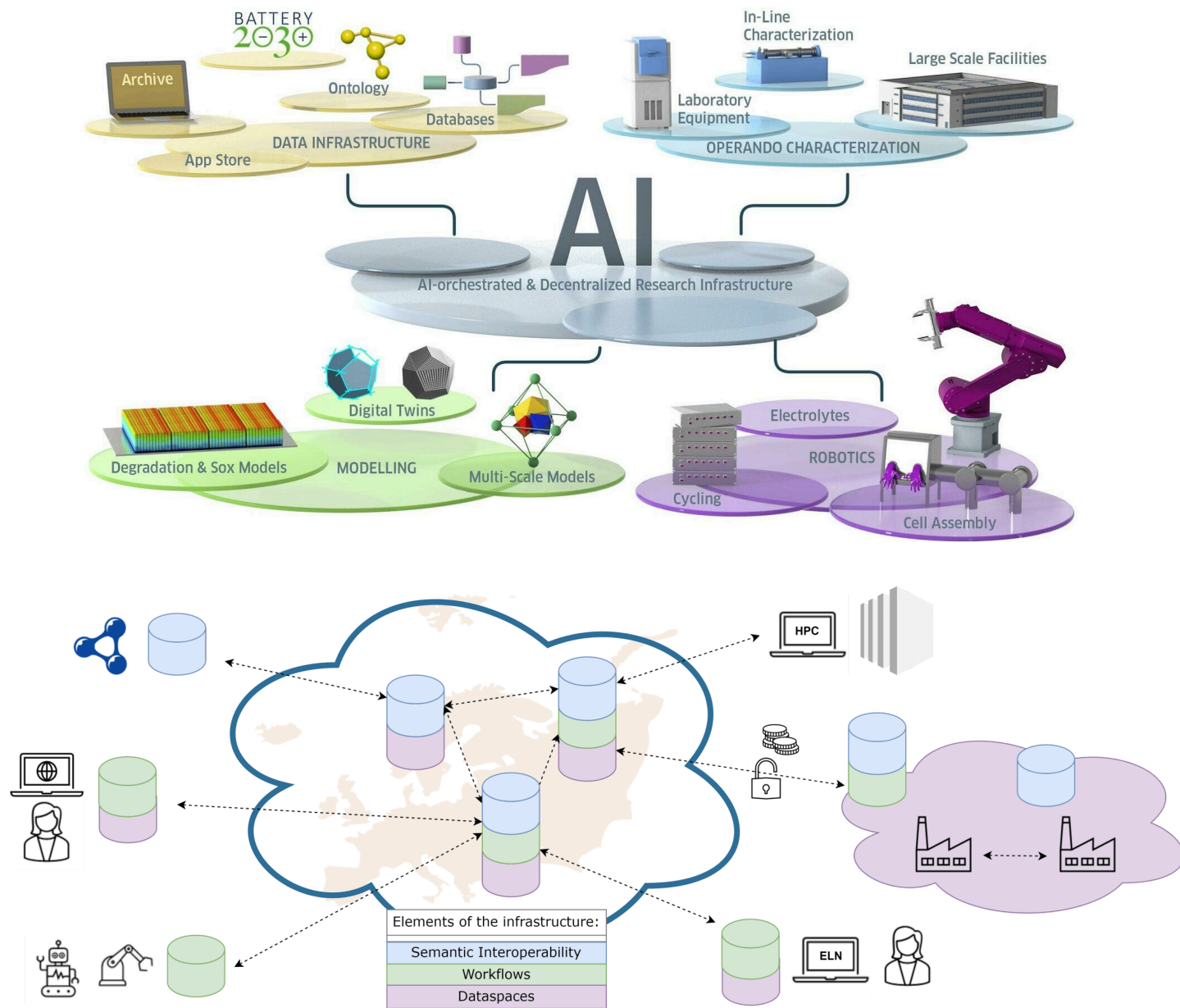
T2: Tejs Vegge,
DTU Energy



MaterialsCommons: Digital Infrastructure

Federated digital infrastructure for advanced materials and develop the components to unite the efforts in materials research

Allow scientists and engineers to visit a single entry point, and get seamless, secure and interoperable access to data repositories, using a federated infrastructure.



ScatterLab & MODEX

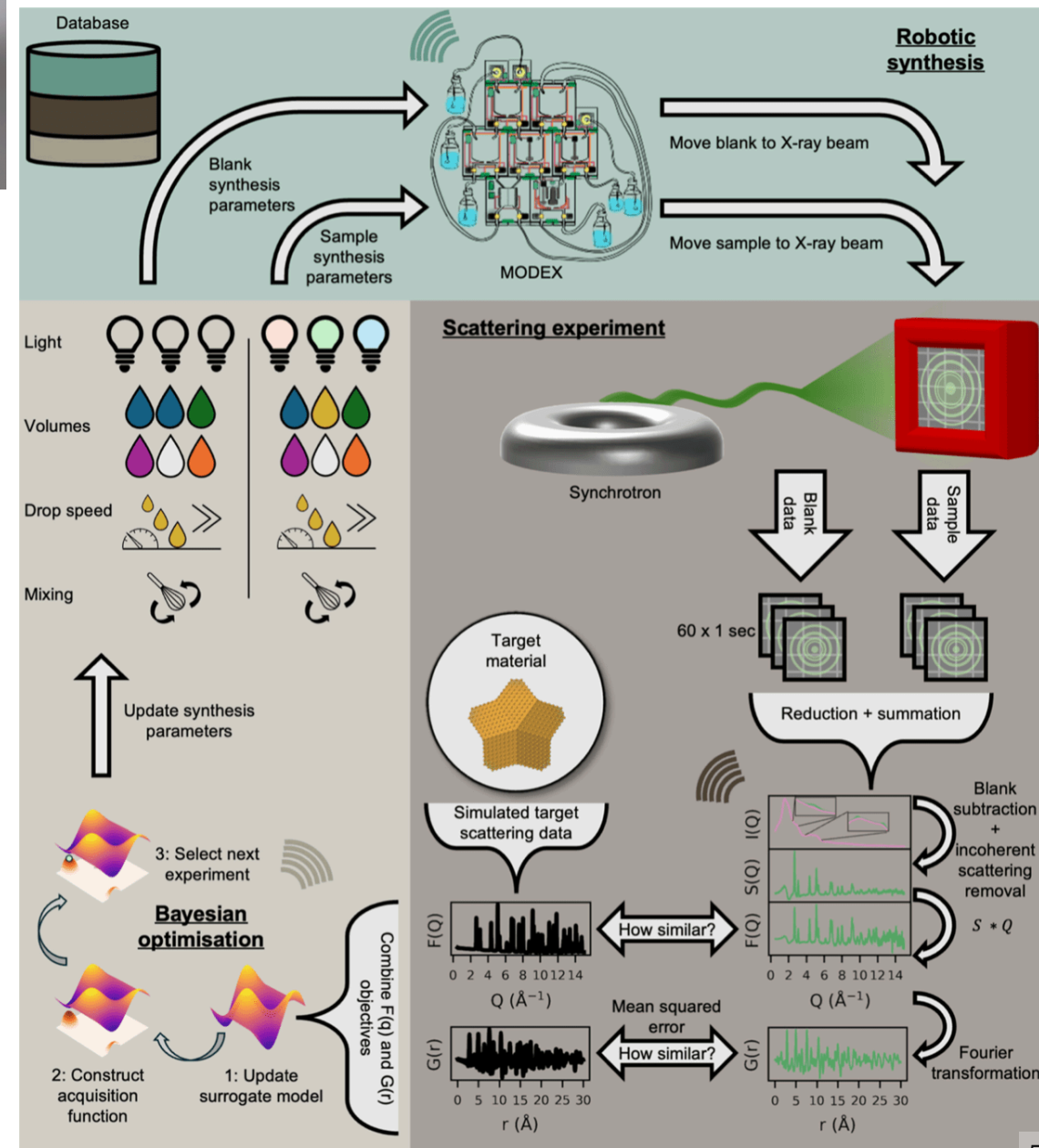
Modular setup of experiment for high adaptability and easy transport

Autonomous nanoparticle synthesis with structural

Characterization and ML prediction loop within synchrotron facility



Anker et al., ACS Nano, 2026



ScatterLab & MODEX

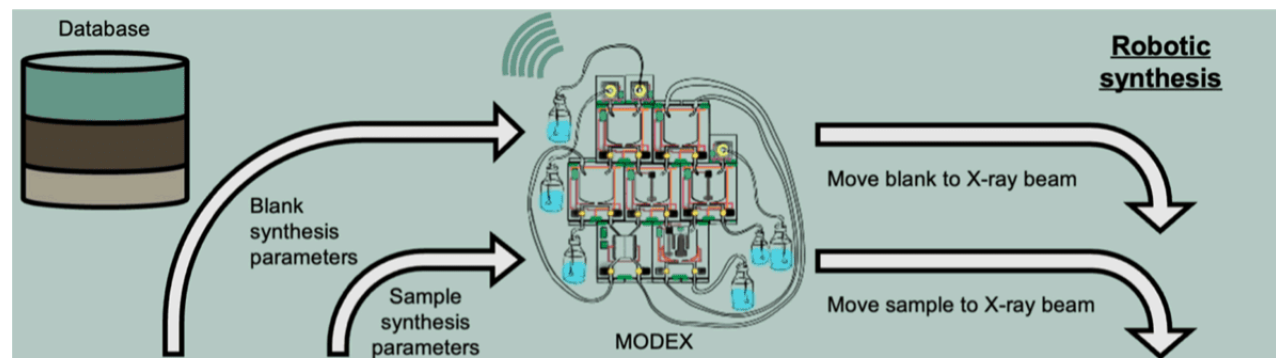
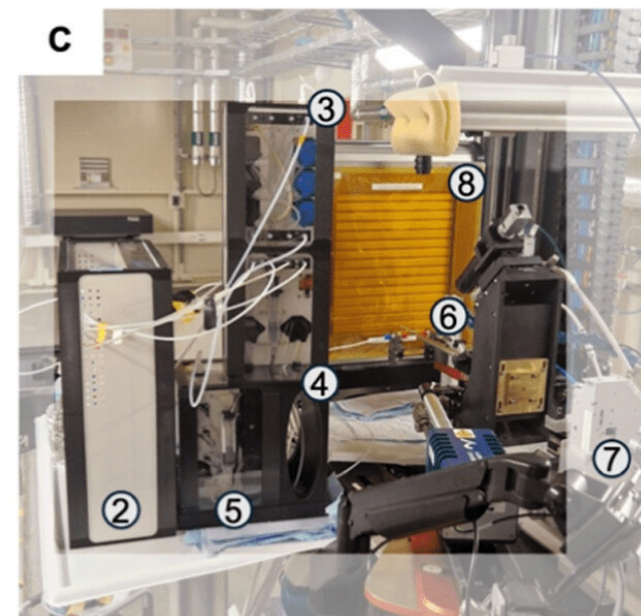
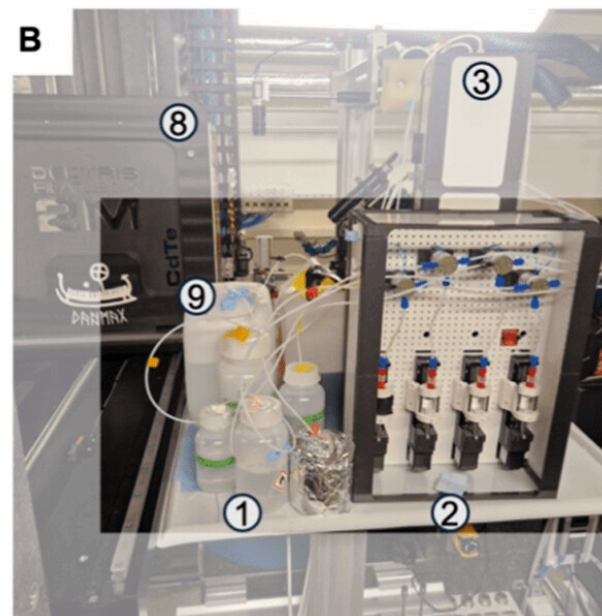
Modular setup of experiment for high adaptability and easy transport

Autonomous nanoparticle synthesis with structural

Characterization and ML prediction loop within synchrotron facility



Anker et al., ACS Nano, 2026

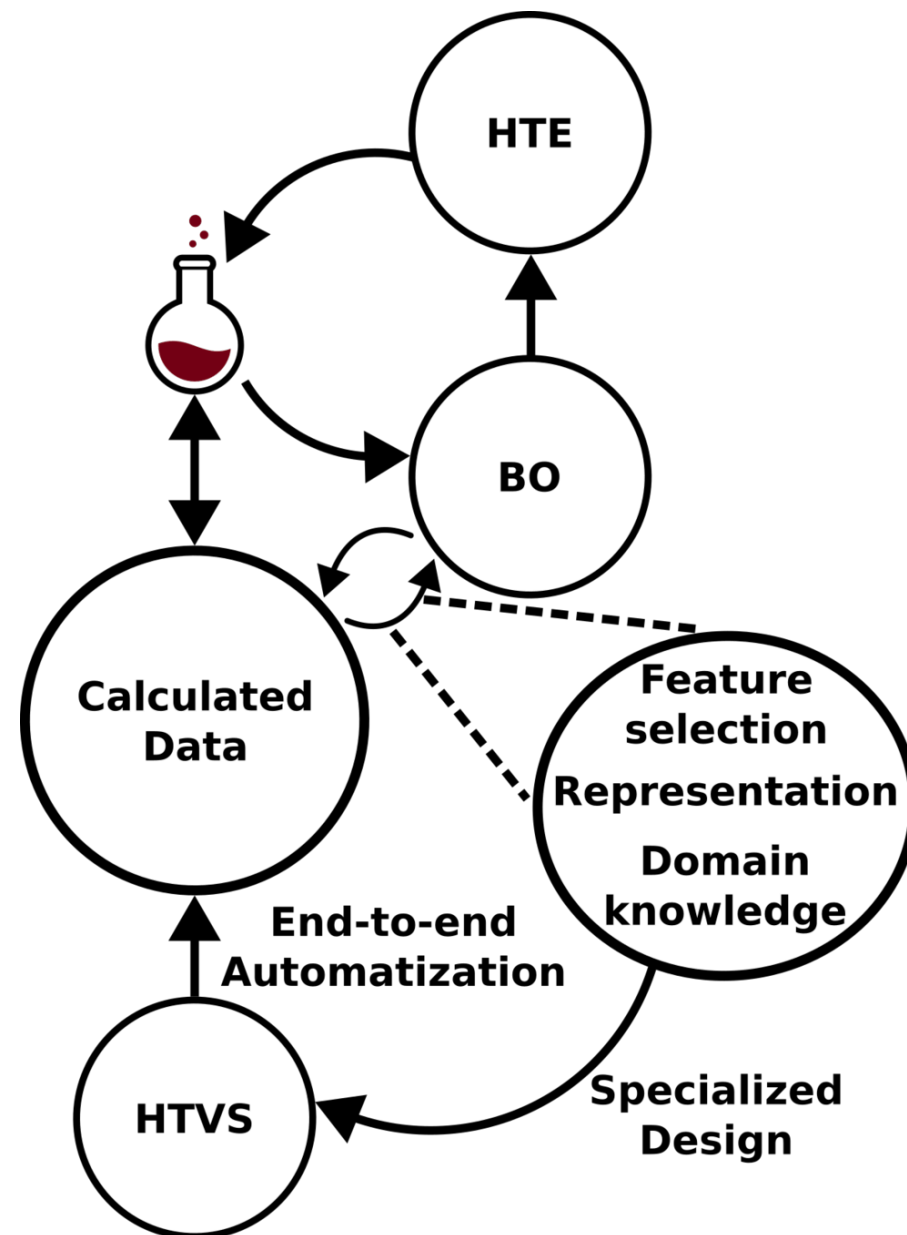


Align HTE with HTVS

Cutting-edge closed loop experiments based on **Bayesian Optimization**, which requires **surrogate model** alignment with experiment

Alignment is not guaranteed, and initial conditions are crucial

HTVS can provide much more **data**, huge benefit if we **ensure alignment with experiment**

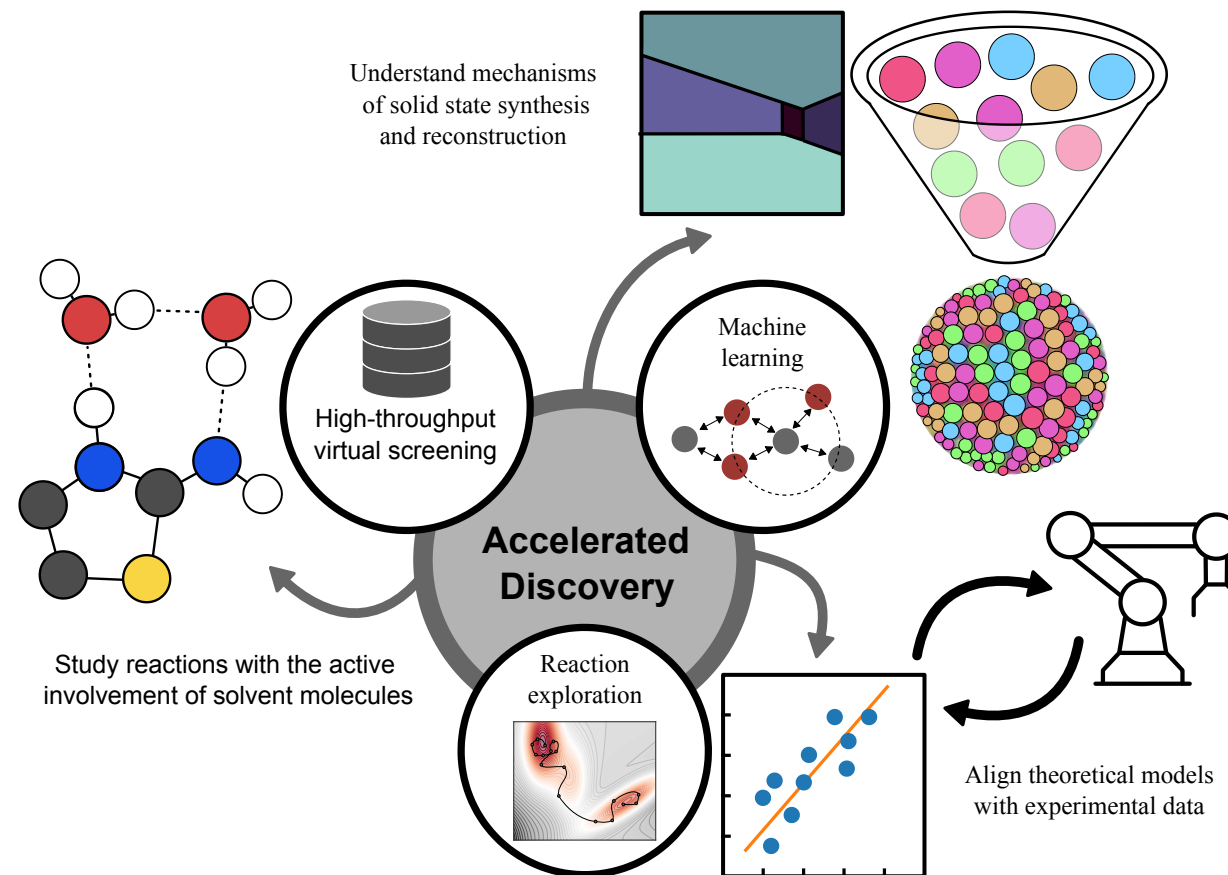


Research vision

Models are **breaking** crucial time and size **constraints**

How can we assure that we **calculate the correct things?**

- **Realistic** structural models
- **Connecting** theory and experiments
- Become source of insights:
Predictive instead of explanatory





NCCR
Catalysis



Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Prof. Markus Reiher Dr. Paul Türtscher
Dr. Jan Unsleber Dr. Katja-Sophia Csizi



Prof. Rafael Gómez-Bombarelli
Juno Nam
Dr. Soojung Yang
Hyukjun Lim
Lucas Pinède



Schweizerischer Nationalfonds
Fonds national suisse
Fondo nazionale svizzero
Swiss National Science Foundation



Prof. Tejs Vegge
 Prof. Andy S. Anker
 Dr. Jonas Busk
 Dr. Bardi Benediktsson
 Dr. Nis Fisker-Bødker
 Dr. Enzo Moretti
 Dr. Sam Walton
 Jonas H. Jensen
 Emil Frost



Understanding chemical reaction mechanisms with automation and machine learning

Miguel Steiner

