A machine learning approach to study of thermosalient molecular crystals

Bruno Mladineo, Ivor Lončarić
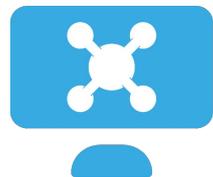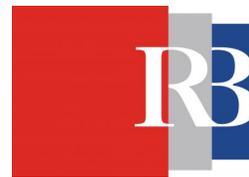
Why we use Machine Learning Interatomic Potentials?

- Computational quantum mechanical modelling methods like DFT and Coupled cluster methods are accurate but scale poorly

- Methods based on classical (empirical) potentials scale linearly but are not as accurate

Why we use Machine Learning Interatomic Potentials?

- Size scalability

- Resource efficiency

- Can use already existing data
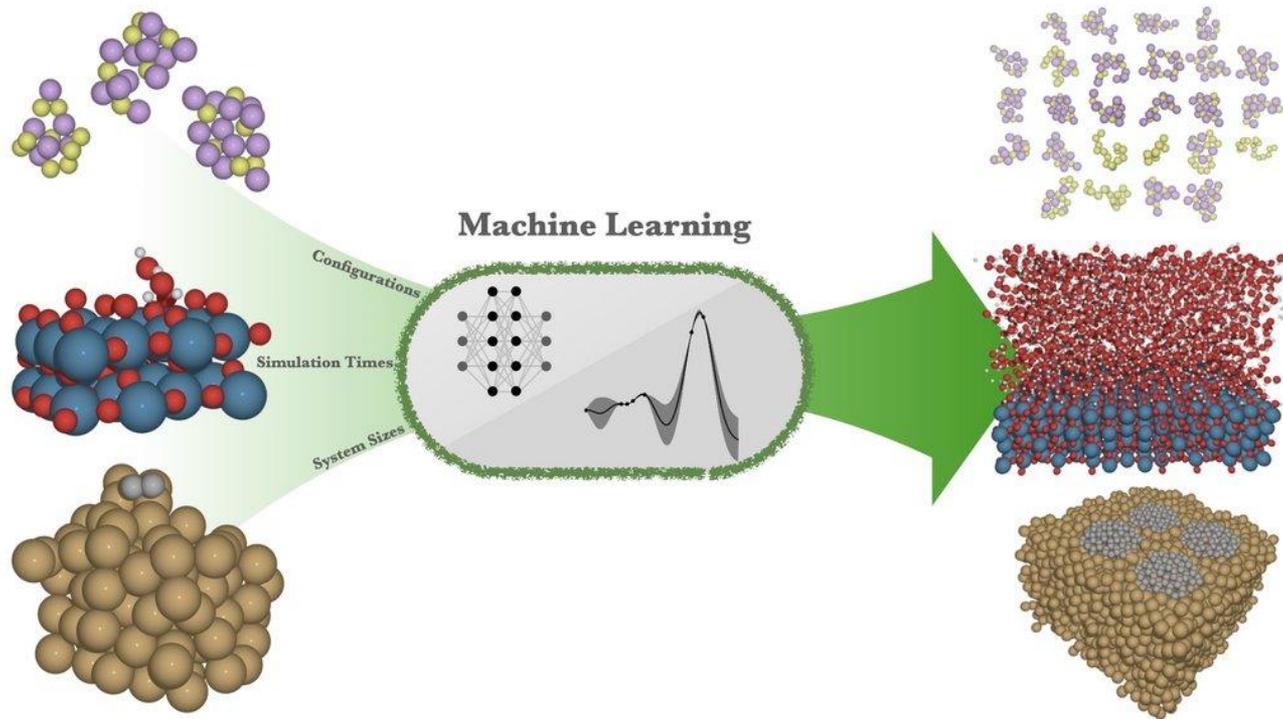
Basics of Machine Learning Interatomic Potentials - MLIPs

positions & chemical species $\Longrightarrow$ energies, forces & stresses

Behler, Jörg, and Michele Parrinello. "Generalized neural-network representation of high-dimensional potential-energy surfaces." *Physical review letters* 98.14 (2007): 146401.
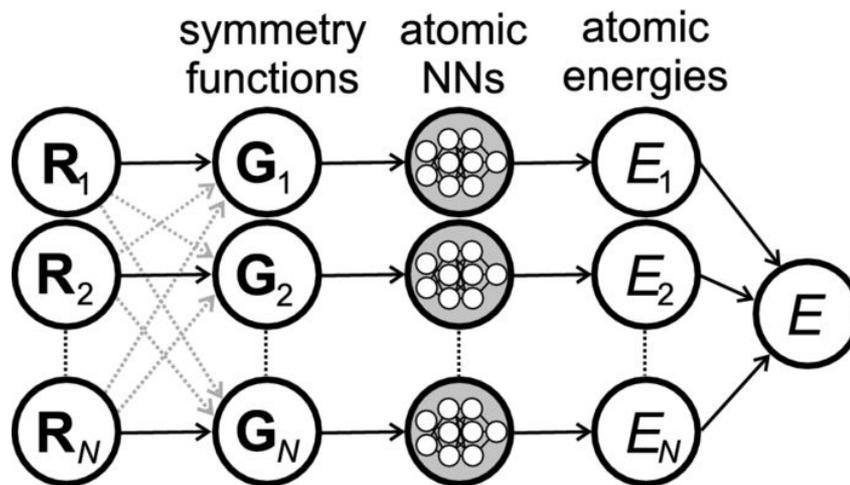
Modern MLIP models

- Continuous-filter convolutional message passing graph neural networks : Duvenaud et al.[1](2015), Schnet[2] (2017)

- E(3)-equivariant MLIPs : NequIP[3](2022), Allegro[4](2023)

(1)   Duvenaud, David, et al. 'Convolutional Networks on Graphs for Learning Molecular Fingerprints'. arXiv, 3 November 2015. http://arxiv.org/abs/1509.09292.
(2)   Schütt, Kristof T., et al. 'SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions'. arXiv, 19 December 2017. http://arxiv.org/abs/1706.08566.
(3)   Batzner, Simon, et al. 'E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials'. *Nature Communications* 13, no. 1 (4 May 2022): 2453. https://doi.org/10.1038/s41467-022-29939-5.
(4)   Musaelian, Albert, et al. 'Learning Local Equivariant Representations for Large-Scale Atomistic Dynamics'. *Nature Communications* 14, no. 1 (3 February 2023): 579. https://doi.org/10.1038/s41467-023-36329-y.

E(3)-equivariant MLIPs
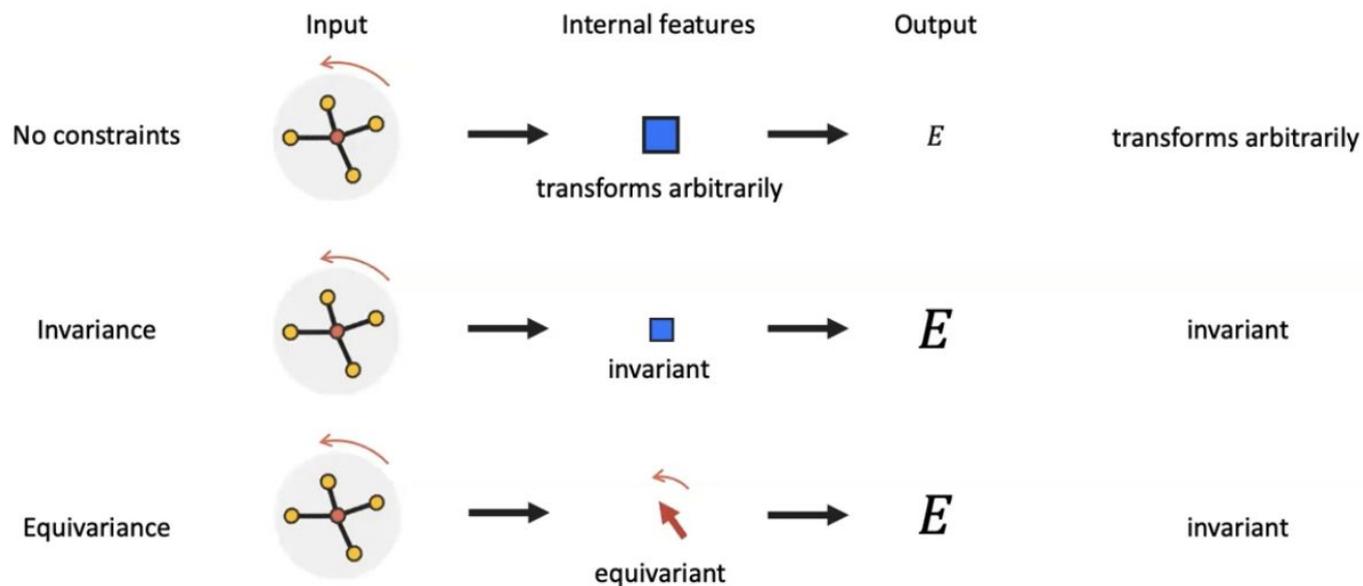
better data efficiency + accuracy

(Jumping) Organic molecular crystals

The system

Reversible Thermosalient
Effect with an Immense
Negative Compressibility
(Jumping phenomenon)

Phase 1

Phase 2

Phase 3
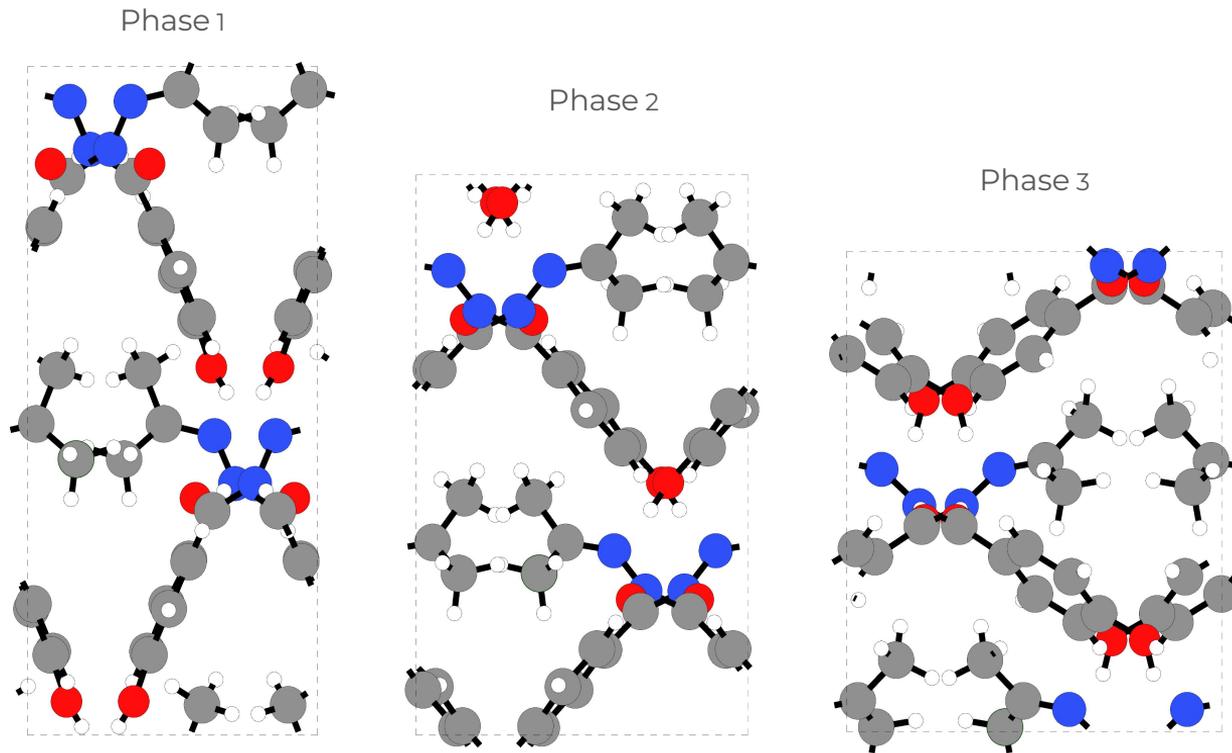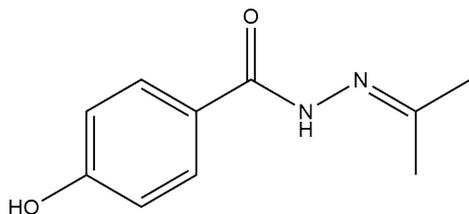
(1)   Lončarić, Ivor, Jasminka Popović, Vito Despoja, Sanja Burazer, Ivan Grgičević, Dean Popović, and Željko Skoko. 'Reversible Thermosalient Effect of
        N'-2-Propylidene-4-Hydroxybenzohydrazide Accompanied by an Immense Negative Compressibility: Structural and Theoretical Arguments Aiming toward the Elucidation of
        Jumping Phenomenon'. *Crystal Growth & Design* 17, no. 8 (2 August 2017): 4445–53. https://doi.org/10.1021/acs.cgd.7b00785.

Initial dataset

Creation of a dataset using
Normal Mode Sampling

- Which adds ~700
  normal mode
  structures to dataset
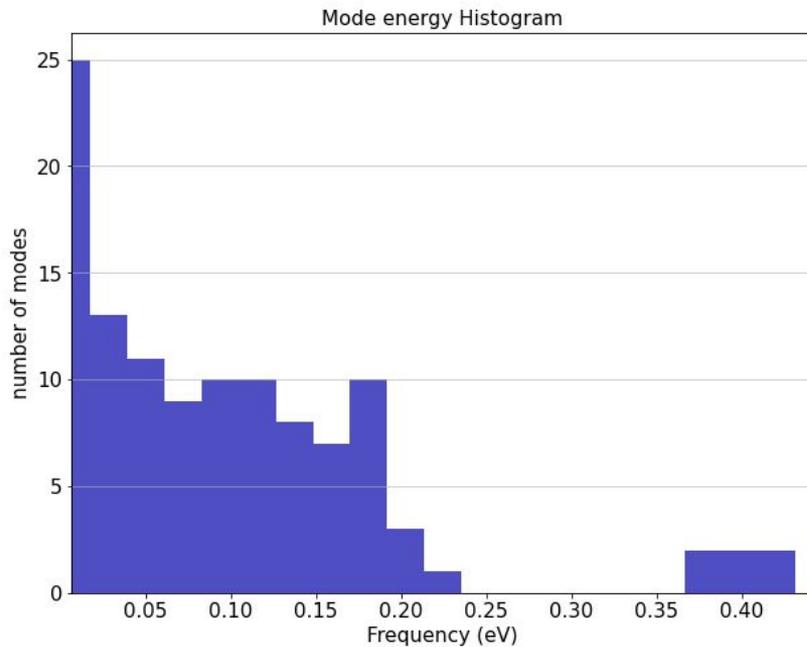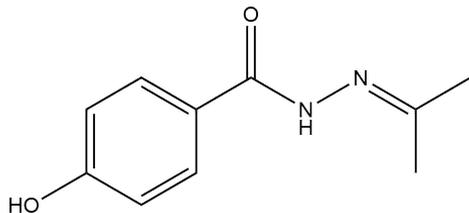
NequIP Active Learning

- Also called optimal experimental design.

- Using a Query by committee[1] (QBC) from 6 models to get an uncertainty estimation

- Running a NPT MD at different temperatures (100-400K) to generate new structures

- Energy and forces of structures with the highest uncertainty are calculated using DFT and added to the dataset

(1)   Christoph Schran, Krystof Brezina, Ondrej Marsalek; Committee neural network potentials control generalization errors and enable active learning. J. Chem. Phys. 14 September 2020; 153 (10): 104105. https://doi.org/10.1063/5.0016004
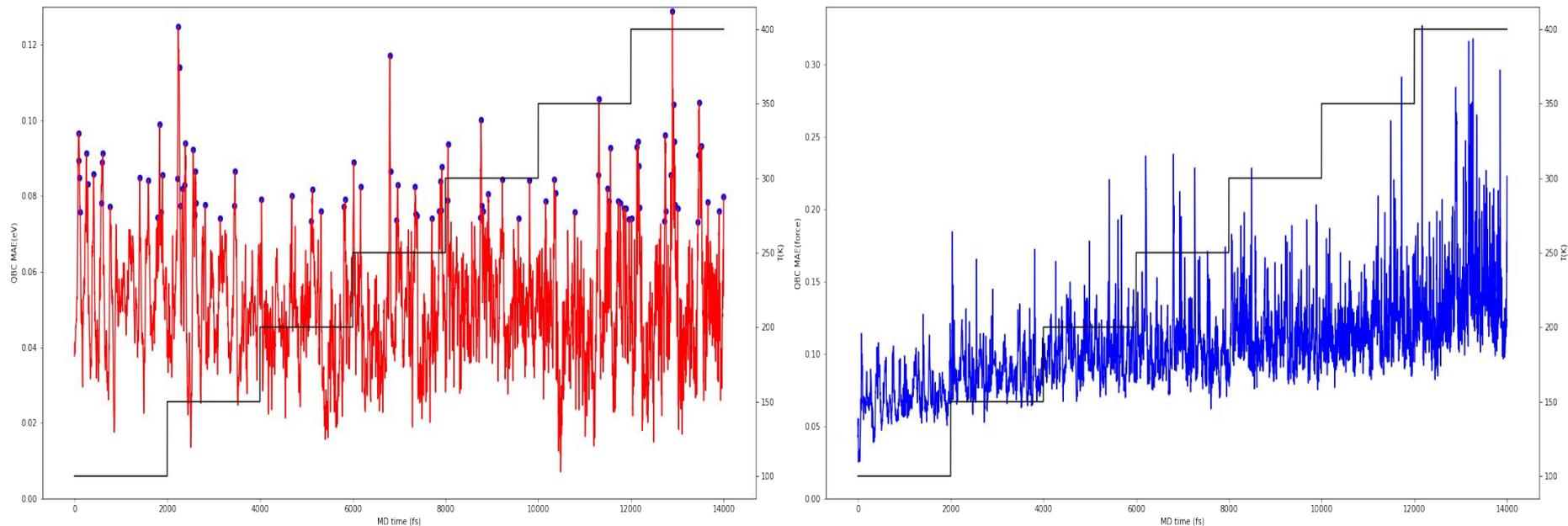
GEN 0

Graph of energy and force uncertainty (QBC MAE) of an NPT MD and the selected structures (blue dots) and the set thermostat temperature during the MD (black line)

GEN 4
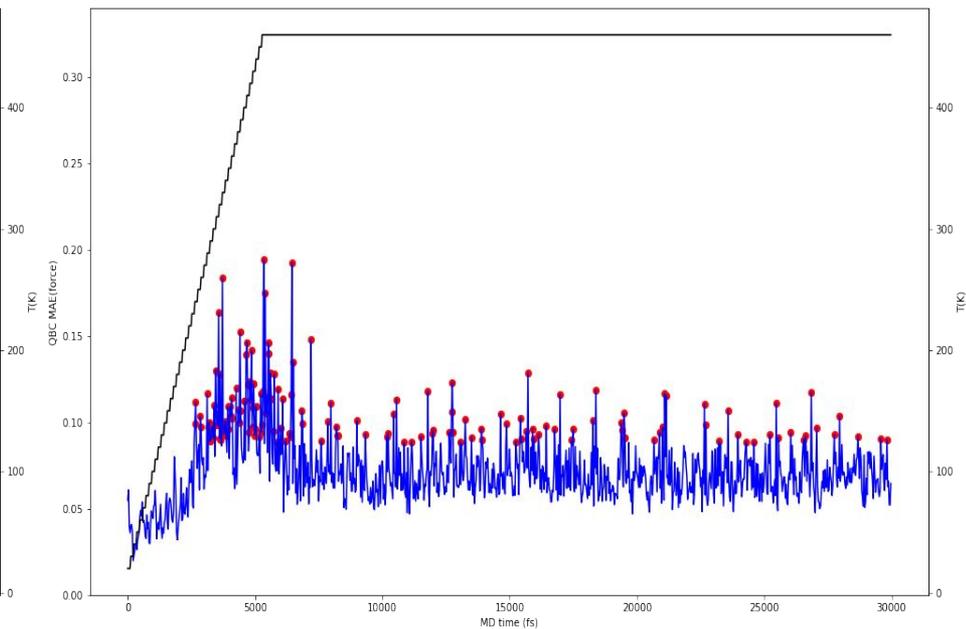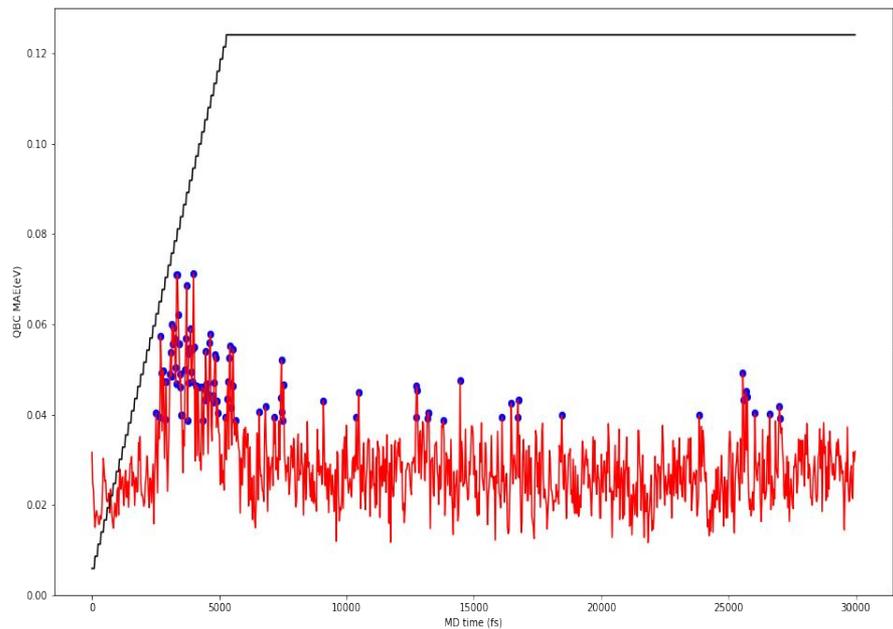
Graph of energy and force uncertainty (QBC MAE) of an NPT MD and the selected structures (blue dots) and the set thermostat temperature during the MD (black line)

Phonon and free energy calculations - Harmonic approximation
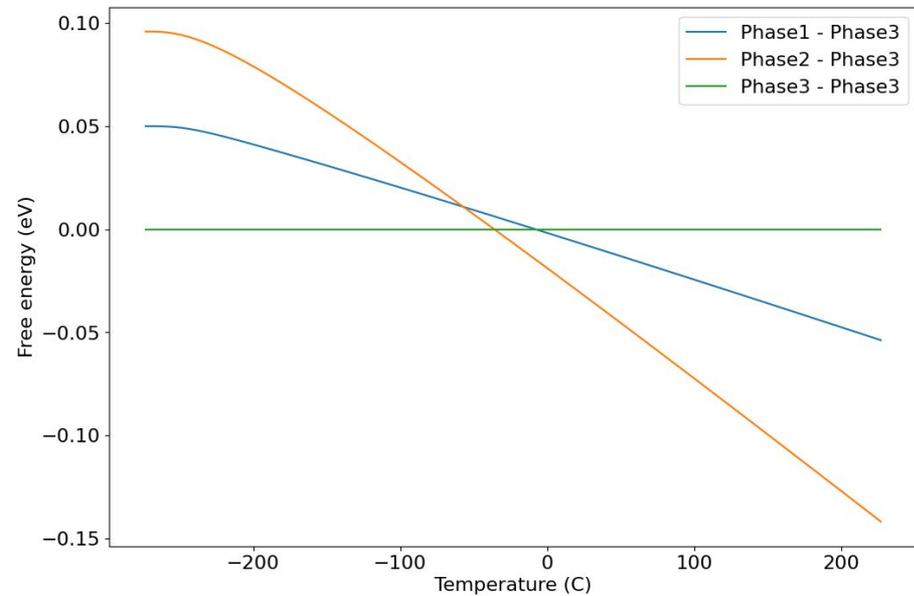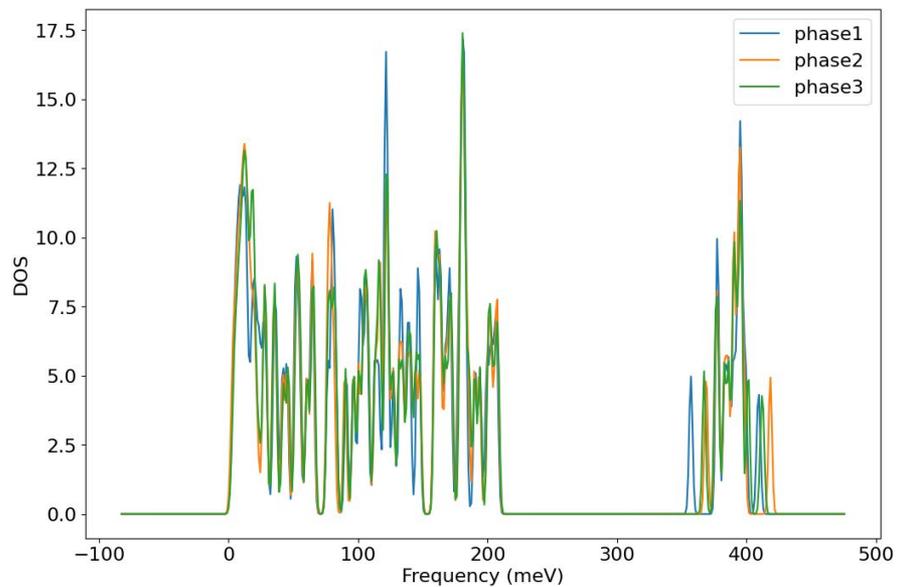
Free energy calculations - Quasi harmonic approximation screening by **pressure**
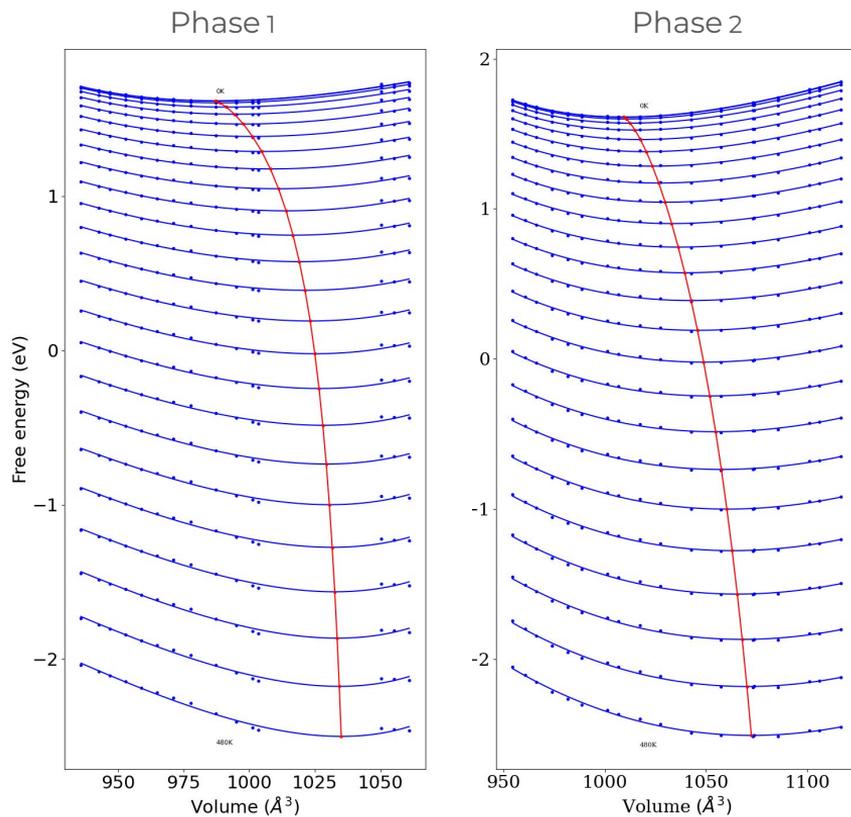
Comparison with experiment

Experimental values
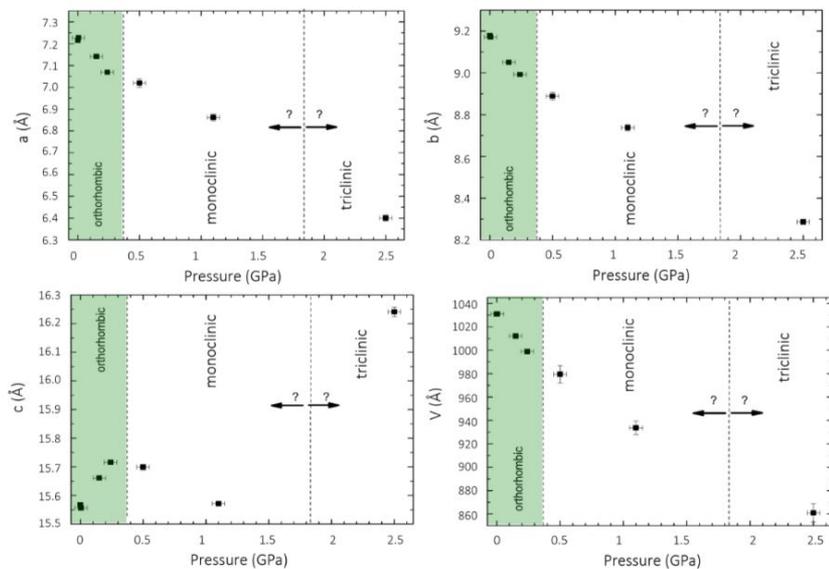
Figure 17 Lattice parameters of N'-2-propylidene-4-hydroxybenzohydrazide Form I as a function of hydrostatic pressure determined by in-situ high-pressure single crystal diffraction.

Calculated with MLIP
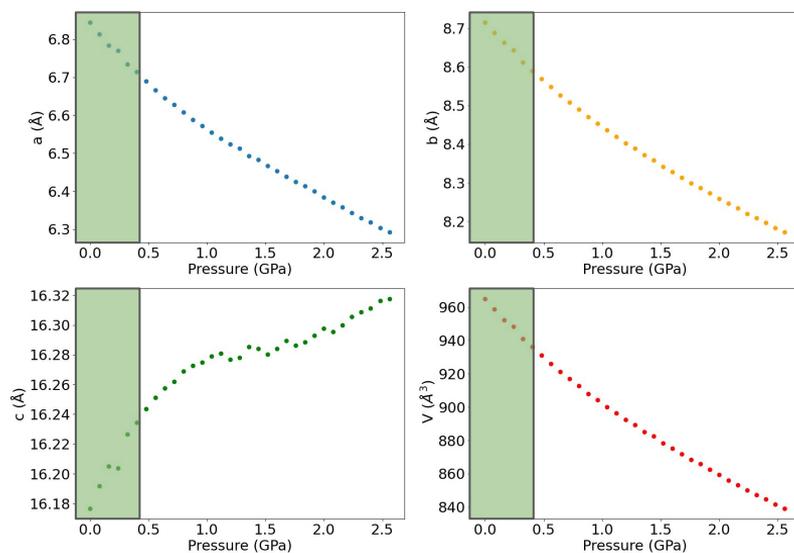
Lattice parameters as a function of pressure for phase 1

Summary

- MLIPs are accurate, and allow efficient and fast computations

- Possibility for HTC
    - E-V curves require ~40 structure minimisations of 50-1000 steps each
    - Harmonic approximation requires ~150 calculations with supercells (2700 atoms)
    - Workflow can be done in a day on a PC

- Faster feedback loop for methodology development

Thank you for your attention!