Machine Learning of QM Lattice Energies for Molecular Crystal Structure Prediction

Rebecca J Clements, University of Southampton
rjc1g14@soton.ac.uk

Skills 4 Scientists

Sponsored by: optibrium, dotmatics knowledge solutions

1. Introduction

- Correcting the relative ranking of crystal polymorphs using higher level QM methods with a fragment-based model within a radial cut off:

$$E'_{latt} = E^{(ll)}_{latt} + \sum_{i}^{molecules} \left(E^{(hl)}_{ref,i} - E^{(ll)}_{ref,i}\right)$$

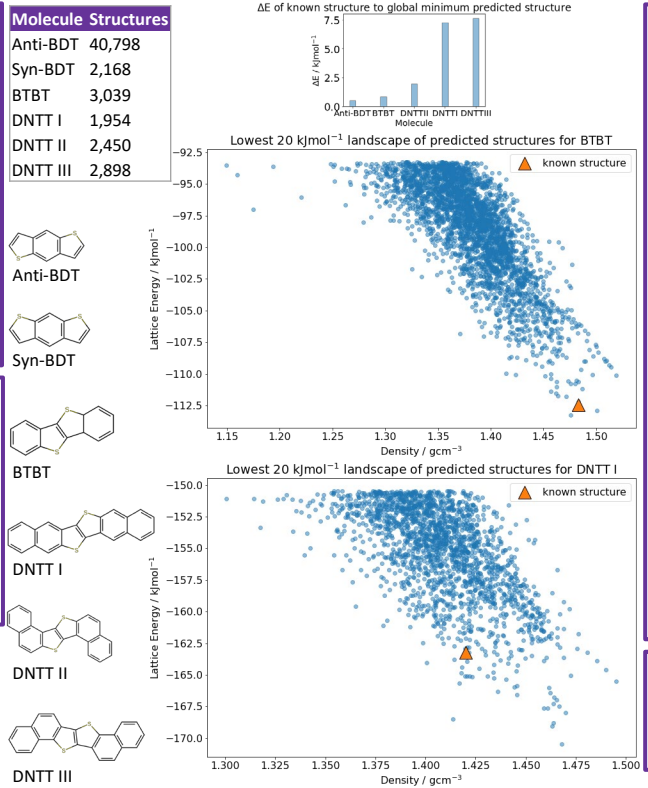
E_{latt} = lattice energy; ll = lower level method; hl = higher level; method; ref = reference molecule;

- Providing large datasets of dimers for ML model
- Extending earlier work by Dr. David McDonagh [1]

2. Motivation

- More accurate methods to distinguish between the energies of the crystal polymorphs predicted
- Allows accurate wavefunction methods to be applied to lattice energy evaluation
- Significantly reduced computational cost
- Larger training sets and transferable data store

Molecule	Structures
Anti-BDT	40,798
Syn-BDT	2,168
BTBT	3,039
DNTT I	1,954
DNTT II	2,450
DNTT III	2,898



3. Results

- Working on QM calculations in the CSPy code, including scheme for identification of identical dimers and storage of energies, descriptors and selection of training sets
- Data collection: 6 thiophene molecules (bottom left) with crystal structures predicted (middle and bottom right examples)
- Large number of structures in data set (top left); force field ranking (ΔE = change in lattice energy) of known CSD structure [2] varies over molecules (top right)
- Force field provides excellent structural agreement with known structure, so that a single-point energy correction is possible
- BTBT (middle right) shows good initial ranking but known structure is not global minimum. DFT-based energy correction without ML: slightly worse energy ranking (+1.2/+1.5 kJmol^{-1} for PBE/B3LYP(-D3,6-31+G**) for space group 14 predictions

4. Further Work

- Investigate QM energy corrections at other levels of theory with ML and extend to further thiophenes

Acknowledgements: Thanks to Prof Graeme Day and research group, the NGCM CDT for training, and the Iridis5 HPC.

References: [1] David McDonagh, Chris-Kriton Skylaris, and Graeme M Day, Journal of Chemical Theory and Computation, 15(4):2743–2758, Apr 2019. [2] Colin R Groom, Ian J Bruno, Matthew P Lightfoot, and Suzanna C Ward. Acta crystallographica Section B, Structural science, Crystal Engineering and Materials, 72(Pt 2):171–179, Apr 2016.

AI^3 Science Network$^+$

PSDS PHYSICAL SCIENCES DATA-SCIENCE SERVICE

RSC INTEREST GROUP CHEMICAL INFORMATION AND COMPUTER APPLICATIONS GROUP