Rational Reduction of Computationally Predicted Crystal Energy Landscapes with Molecular Dynamics

Nicholas F. Francia[2], Edgar Ohlenovics[1], Louise Price[3], Sally L. Price[3] and Matteo Salvalaglio[1]*

*m.salvalaglio@ucl.ac.uk
1 Chemical Engineering, University College London, London, United Kingdom,
2 The Cambridge Crystallographic Data Centre (CCDC), Cambridge, United Kingdom,
3 Chemistry, University College London, London, United Kingdom.

Crystal Structure Prediction (CSP) methods should ideally identify the operative conditions to obtain a specific polymorph starting only from the molecular structure of their building blocks. Unfortunately, even if CSP methods have dramatically evolved over the past decade, the field is still far from achieving this vision. While successful predictions of the most stable polymorphs of organic crystals based on accurate lattice (free) energy estimates are becoming increasingly common, the number of predicted putative polymorphs remains a standing issue.
Typically, the overall number of potential structures computationally identified as putative polymorphs usually grossly exceeds the number of polymorphs observed experimentally [1]. This problem in the field is typically referred to as "overprediction".
A reason for CSP methods over-predicting possible polymorphs is that temperature increases molecular motions so that some structures that appear as distinct lattice energy minima merge into the same free energy minimum at finite temperature [2].
Here we discuss a physics-driven computational procedure [3-5] to reduce the number of relevant local minima of crystal energy landscapes based on a systematic and scalable application of molecular dynamics and enhanced sampling simulations.
To identify persistent crystal structures, we perform classical molecular dynamics simulations at finite temperature on CSP-generated crystal structures. Unstable configurations are then automatically removed by comparing the distribution of internal orientations against the random arrangement typical of a disordered melted state.
To identify crystal structures that convert to a common finite-temperature state, we carry out a clustering analysis based on probabilistic fingerprints that capture information on the relative position, orientation, and conformation of molecules within a dynamic crystal supercell. Differences in the probabilistic fingerprints concur to define a dissimilarity metric between finite-temperature structures, which is then used to perform unsupervised clustering of finite temperature structures.
In this contribution, we discuss the application of this method to systems of increasing size and complexity (urea, succinic acid, ibuprofen, sulflower, persulforated coronene, olanzapine and others), spanning from a few dozens to thousands of structures, thus approaching the size and complexity of real-world CSP applications.
We note that, in all cases, we substantially reduce the crystal energy landscapes while consistently retaining the experimentally observed crystal structures.

References

[1] DA Bardwell, CS Adjiman, YA Arnautova, E Bartashevich, SXM Boerrigter, et al., 2011, Acta Cryst. B 67:535-55
[2] Price SL. 2013, Acta Cryst. B. 69:313-28.
[3] Francia, N.F., Price, L.S., Nyman, J., Price, S.L. and Salvalaglio, M., 2020. *Crystal Growth & Design*, *20*(10), pp.6847-6862.
[4] Francia, N.F., Price, L.S. and Salvalaglio, M., 2021, *CrystEngComm*, *23*(33), pp.5575-5584.
[5] IJ Sugden, NF Francia, T Jensen, CS Adjiman, M Salvalaglio, 2022, CrystEngComm 24 (39), 6830-6838