

Optimizing co-crystal screens using a data-driven machine learning method

Richard Ian Cooper¹, Simon Lawrence², Lorraine Crowley², Jerome Wicker¹

¹Dept. Of Chemistry, University Of Oxford, Oxford, United Kingdom, ²Chemistry & Analytical and Biological Chemistry Research Facility, University College Cork, Cork, Ireland
E-mail: richard.cooper@chem.ox.ac.uk

Producing multi-component crystals is a key route to altering important physicochemical properties, including bioavailability, dissolution rate and stability of active pharmaceutical ingredients (APIs).[1]

Using the results of a co-crystal screening of 34 API-like molecules with 20 common co-formers allows a training of a model using machine learning which produce a ranked list of co-formers in order to increase success rates in experiments with previously unseen APIs.[2] Descriptors of the molecular components of the putative co-crystals are built purely from 2D atomic and connectivity information, avoiding explicit consideration of the resultant structure geometry, lattice energy and any specific intermolecular interactions.

External data is used to validate the model and demonstrates that it performs well for APIs in the same chemical neighbourhood as the training data, but shows no predictive power for widely different species. A method for detecting the suitability of the training data for a proposed API is suggested.

The machine learning algorithm can train a model on a set of successful and unsuccessful co-crystallisation experiments which can then be used to guide selection of co-formers for a particular API. This is likely to save time and resources on experimental screens, enabling access to new materials with altered properties.

[1] Aakeröy, C. B., Forbes S. & Desper, J. (2009) J. Am. Chem. Soc., 131, 17048–17049.

[2] Wicker, J. G. P., Crowley, L. M., Robshaw, O., Little, E. J., Stokes, S. P., Cooper, R. I. and Lawrence, S. E. (2017) submitted

Keywords: [co-crystals](#), [machine-learning](#)