

Computational Organic Crystal Structure Prediction, an evolving challenge?

S. L. Price¹

¹*Department of Chemistry, University College London, 20 Gordon St., London WC1H 0AJ*

s.l.price@ucl.ac.uk

The original aim of computational organic crystal structure prediction (CSP) was to see if it was possible to predict the crystal structure of an organic molecule from the chemical diagram. The “from the chemical diagram” was to ensure that the method could be used prior to the synthesis of the molecule, so the methodology should not use any experimental data, at least not on the specific molecule. The aim of seeking “the” crystal structure was because most people ignored polymorphism at that time, with the crystal structure being that of the first crystal that could be grown that was of a suitable quality for single crystal X-ray diffraction. Thus the theory behind CSP is that the molecule will crystallize in the most thermodynamically stable crystal structure. This assumption is ideal for computers, requiring a search for the crystal structure that gave the most stable structure relative to the isolated molecules. The thermodynamic stability was approximated by the lattice energy, the energy difference between a hypothetical, infinite, perfect, static crystal and infinitely-separated, static molecules in their lowest energy conformation. This zeroth order CSP₀ was envisioned as an aid to the design of new functional organic materials, to estimate whether an organic molecule would crystallise with the desired physical properties prior to synthesis. Progress towards this aim has been charted by the CCDC’s blind tests of CSP [1].

It was the interest in polymorphism, particularly the late discovery of more thermodynamically stable polymorphs leading to the “disappearance” of the previously apparently stable forms, which made CSP an industrially desirable technology. It became apparent that for most molecules, there were many local minimum in the crystal energy that were close in stability to the most stable, some of which corresponded to known polymorphs. Thus CSP developed into a technology for aiding polymorph screening, showing the range of possible crystal structures of a molecule, and aiding the determination of structures from powder diffraction, solid state NMR and generally assisting in understanding the solid form landscape [2]. However, CSP usually generates more thermodynamically plausible structures than known polymorphs, raising the question of why don’t we find more polymorphs [3]?

There has been progress on this over-prediction problem, with the recent emergence of methods of determining whether minima in the static lattice energy remain distinct when the dynamic motions of the molecules at practical temperatures are considered. The accuracy of calculations of the thermodynamics of organic crystals is rapidly improving, and needs validation against experimental data, an aim of the COST action BEST-CSP. In some cases, having a CSP-generated structure may suggest a method of crystallisation, such as sublimation onto a template. Our ability to calculate the physical properties of a crystal from its structure, such as spectra, morphology, solubility and mechanical properties, are also progressing. However, the digital design of an industrial crystallisation process from the molecular diagram remains hampered by the accuracy of the calculations and inability to predict nucleation, growth and transformation rates [4].

We have recently produced the CPOSS209 data set of crystal structures of 20 medium sized organic molecules, containing the known polymorphs, structures derived from those of closely related molecules and structures that a CSP study suggests are putative polymorphs [5]. It is hoped that this data set will be useful for preliminary work on testing new computational models for the thermodynamic stability and suggesting which thermodynamically plausible structures could actually be observed. The ideal crystal structure prediction code would predict all the structures that could be crystallised (and how this could be done), and no more [6].

[1] Hunnisett, L. M. et al. The seventh blind test of crystal structure prediction: structure ranking methods. (2024) *Acta Cryst. B*, **80**, 548.

[2] Price, S. L.; Reutzler-Edens, S. M. The potential of computed crystal energy landscapes to aid solid-form development. (2016) *Drug Discovery Today*, **21**, 912.

[3] Price, S. L. Why don't we find more polymorphs? (2013) *Acta Cryst. B*, **69**, 313.

[4] Burcham, C. L.; Doherty, M. F.; Peters, B. G.; Price, S. L.; Salvalaglio, M.; Reutzler-Edens, S. M.; Price, L. S.; Addula, R. K. R.; Francia, N.; Khanna, V.; Zhao, Y. Pharmaceutical Digital Design: From Chemical Structure through Crystal Polymorph to Conceptual Crystallization Process. (2024) *Crystal Growth & Design*, **24**, 5417.

[5] Price, L.; Paloni, M.; Salvalaglio, M.; Price, S. L. One size fits all? Development of the CPOSS209 Data Set of Experimental and Hypothetical Polymorphs for Testing Computational Modeling Methods. (2025) *Crystal Growth & Design*, <https://doi.org/10.1021/acs.cgd.5c00255>.

[6] Price, S. L. Is zeroth order crystal structure prediction (CSP₀) coming to maturity? What should we aim for in an ideal crystal structure prediction code? (2018) *Faraday Discussions*, **211**, 9.

I wish to thank the many collaborators from the start of the CPOSS project whose work has led to the many combined experimental and computational papers on www.cposs.org.uk that are behind this overview lecture. In particular, Dr Louise Price (no relation) has been my technical and administrative assistant and master of our database of CSP structures for over 20 years.