CrystEngComm

ROYAL SOCIETY OF CHEMISTRY

EDITORIAL

View Article Online
View Journal | View Issue

Cite this: *CrystEngComm*, 2022, **24**, 431

Computer-aided solid form design

Susan M. Reutzel-Edens

DOI: 10.1039/d1ce90150h

rsc.li/crystengcomm

This themed collection of *CrystEngComm* on **computer-aided solid form design** features recent work on the development, application, and experimental validation of algorithms for structure-based modeling and prediction. Combined with high performance computing and rich databases, computational chemistry, machine learning and knowledge-based approaches are making it increasingly possible to optimize molecular and material properties in computers before synthesis commences. Highlighted in this collection are some of the latest examples of cutting-edge science along three of the main thrust areas of computer-aided solid form design: crystal structure modeling and prediction, the experimental validation of predicted target structures, and the use of computer modeling for material property prediction.

Crystal structure modeling and prediction

Crystal structure prediction (CSP) methods continue to develop to address what was considered in 1988 to be "one of the continuing scandals in the physical sciences", to quote John Maddox, namely the ability to predict the crystal structure of a compound from its chemical composition. Today, as the paradigm has begun to shift to computer-aided solid form design, the reliable prediction of how a molecule crystallizes, including its tendency to exhibit polymorphism, is considered the critical first step to minimizing the experimental footprint of costly, time-consuming, and even potentially hazardous (in the case of energetic materials) solid form screening. Marom *et al.* (Carnegie Mellon University) demonstrated the potential of incorporating CSP into the digital design of energetic materials. In this work, experimental structures were found by CSP for three target molecules using the random structure generator Genarris, and the genetic algorithm Gator. The ease with which known polymorph structures could be generated was related to their location on the potential energy surface, in particular the width of the energy basin in which they reside. Interestingly, an energetically competitive, high-density structure with a sheet packing motif, associated with reduced detonation power and sensitivity, was identified as a potentially viable polymorph for one of the energetic targets, DATB (2,4,6-trinitrobenzene-1,3-diamine) (DOI: 10.1039/d1ce00745a).

Inspired by the Cambridge Crystallographic Data Centre (CCDC)-sponsored blind tests of CSP, which have benchmarked the progress of the algorithms over more than two decades, and with the emergence of commercial CSP providers, structure prediction is now being widely applied across the pharmaceutical industry. However, the application of CSP to pharmaceutical molecules, which have been thoroughly experimentally screened for polymorphs, has revealed two problems: computational overprediction and experimental underestimation. Salvalaglio *et al.* (University College London) have reported the use of large-scale, physics-based molecular dynamics simulations to reduce the problem of crystal structure overprediction. Molecular dynamics (MD) simulations and enhanced sampling methods were used to simulate the crystal structures of ibuprofen at finite temperature and pressure, which effectively melted some of the lattice energy minima, while coalescing others, separated by small barriers, into a smaller number of more stable geometries. The simulation workflow not only significantly reduced the number of predicted crystal packings to a subset containing the experimentally known ibuprofen polymorphs, but it also provided quantitative insights into the emergence of conformational and orientational disorder and the persistence of intermolecular hydrogen-bonding interaction motifs at finite temperature (DOI: 10.1039/d1ce00616a).

As an alternative to the more expensive and time-consuming physics-based CSP-based approach, Abramov *et al.* (XtalPi Inc.) explored the combined use of physics-based (conductor-like screening model for real solvents, COSMO-RS) and machine learning (ML) modeling approaches for fast virtual cocrystal screening. In this work, an overall strategy to model the component miscibility (based on COSMO-RS excess enthalpy) and crystallinity (based on random forest ML) contributions to cocrystal formation proved superior for rapidly identifying the most promising cocrystal formers for ibuprofen (DOI: 10.1039/d1ce00587a).

Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, UK.
E-mail: sreutzel-edens@ccdc.cam.ac.uk

The challenges posed by computational overprediction and the experimental underestimation of polymorphism have also inspired data-driven approaches for polymorph risk assessment. Knowledge-based tools developed by the CCDC, for example, show how a given organic molecule is statistically likely to crystallize based on the 1.1+ million structures in the Cambridge Structural Database (CSD) that have clearly overcome kinetic barriers to crystal nucleation and growth. Doherty *et al.* (GlaxoSmithKline (GSK)), in reporting a survey of the GSK small molecule crystal structure database, exploited differences between the proprietary data and the CSD drug subset (CSD-DS) to build a more reliable knowledge-based polymorphism risk assessment based on hydrogen bond propensity (HBP). Adding chemically relevant structures from the proprietary GSK database to the training data set of published CSD-DS data ensured good coverage of functional groups to improve the HBP logistic regression model (DOI: 10.1039/d1ce00665g).

Experimental validation

The prediction of energetically competitive structures has motivated experimentalists to find them. Braun (University of Innsbruck) reported an experimental investigation targeting computationally generated structures of the cocrystal former, 3-hydroxybenzoic acid. The discovery of a third polymorph has shown that expanding the scope of experimental solid form screening can pay dividends in producing novel forms. However, it as yet remains impossible to predict if a given structure will crystallize, let alone how to target it in an experimental design (DOI: 10.1039/d1ce00159k).

Cruz-Cabeza *et al.* (University of Manchester), in reporting the discovery of the ninth polymorph of the well-studied tolfenamic acid (TFA), have shown that when it comes to polymorph discovery, special techniques are no surrogate for careful observation. The new TFA polymorph, which had been predicted computationally from a previous CSP study, crystallised concomitantly with the more common forms I and II by conventional cooling crystallization from isopropanol. The late appearance of form IX was explained in terms of its overall higher intrinsic rugosity relative to those of forms I and II, building on the idea that polymorphic forms with rougher surfaces may have higher energy barriers to cross for nucleation, as well as the absence of aromatic stacking interactions leading to slower crystal growth relative to the competing polymorphs (DOI: 10.1039/d1ce00343g).

Material properties

To achieve theory-driven crystal engineering, where the properties of molecular materials are optimized in a computer, the ability to reliably predict crystal structures must be met with accurate methods to predict the solid state properties of interest. Feng *et al.* (Dalhousie University) applied a novel computational methodology to model the photoluminescent behaviour of ROY in eight of its polymorphs and a series of 9-acetylanthracene cocrystals, showing how isolated-molecule and dispersion-bound periodic boundary DFT calculations could be combined in a cost-effective way. Whereas the polymorph-dependent photoluminescence (color zoning) of ROY was found to be controlled primarily by intramolecular geometry, and hence could be modelled for isolated crystal conformers, the periodic crystal environment needed to be accounted for in cases where the emission properties are driven by intermolecular charge transfer, as seen in the cocrystals (DOI: 10.1039/d1ce00383f).

Optimizing the physical–chemical properties of organic crystals requires a molecular level understanding not only of the bulk crystal structure, but also of the particle characteristics and surface chemistry, where in many cases most of the action takes place. With respect to surface chemistry, Rantanen *et al.* (University of Copenhagen) combined atomic force microscopy with MD simulations to examine the solid–solution interface of paracetamol in water–ethanol mixtures, showing the dramatic effects that the solvent can have on surface crystallinity and hydrophobicity. Mediated by dynamic heterogeneous disordered surface layers at the solid–solution interface, different critical surface properties could be manipulated by choice of the solvent composition (DOI: 10.1039/d1ce00209k).

The collection of papers in this themed collection highlights the enormous progress all-around in developing and applying crystal modeling and prediction to mimimize the risk of late-appearing polymorphs and to optimize the bulk and surface properties of organic materials. These recent papers build on great contributions toward computer-aided solid form design, some of which were highlighted in the 2020 Editor's Collection under the same name. However, despite the considerable attention paid to the subject in recent years, we are far from being able to engineer material properties in a computer to, say, the level of fidelity required by the aviation industry for designing airplanes. Until physics-based computational approaches achieve the required level of accuracy and affordability to not just predict a thermodynamically feasible crystal structure and its properties, but also the experimental conditions to make it (for which better understanding of crystal nucleation and growth will be required), trial-and-error experimentation will continue to feature prominently in organic crystal design and material property optimization.

The progress toward computer-aided solid form design has nonetheless generated much excitement and interest across industry and academia. Whether the goal is to move to more material-sparing approaches, maximize R&D efficiency, inform decision making, or minimize downstream risk, fit-for-purpose modeling and simulation tools have already proven invaluable. Experimental data of the highest quality will, of course, continue to underpin the entire digital design enterprise, providing the required benchmarks for

computational algorithm development, as well as inputs to data-driven (ML and informatics) approaches and the all-important validation of modeling and prediction tools. For ML and informatics-based material property predictions, the models will only be as good as the underlying data, which speaks to the need for making good data FAIR (findable, accessible, interoperable and retrievable) and ensuring that it is representative. The 1.1+ million crystal structures in the CSD, for example, provide excellent substrates for ML model building, however, some chemistries are sparse, evidenced by the improvement in model predictions with inclusion of proprietary structural data. Beyond this, critical gaps still exist because negative data, which are helpful for testing hypotheses and required for binary classification models, oftentimes go unpublished and provisions (funding, collaboration with expert experimentalists) are not always made for collecting high-quality experimental data.

Digital approaches to solid form property design and optimization will inevitably improve as tools and algorithms to reliably and efficiently model energies (thermodynamics) and account for nucleation and growth kinetics under crystallization process relevant conditions evolve. This themed collection has highlighted a number of developments, which continue to lay the foundation for the future state of computer-aided solid form design, whereby the right crystal form is designed for the right application, allowing the right material to be produced the first time.