MS34-P12 Using matched-molecular structures in the Cambridge Structural Database for crystal engineering

Anthony M. Reilly[1], Ilenia Giangreco[1], Jason C. Cole[1], Gregory P. Shields[1], Neil Feeder[1], Colin R. Groom[1]

1. The Cambridge Crystallographic Data Centre

email: reilly@ccdc.cam.ac.uk

With over 800,000 small-molecule crystal structures, the Cambridge Structural Database (CSD) features a wealth of information for understanding molecular packing and crystal engineering. Within this vast dataset are numerous pairs of molecules related by a single well-defined structural transformation. In drug discovery, analysis of such pairs is used to understand how structural modifications will affect the properties of a molecule. This type of investigation is not common for crystal structures, but has the potential to provide critical insights into how simple chemical transformations affect the structure and properties of solid forms.

In this presentation, we outline the process of finding over 100 million matched molecular pairs in the CSD. We will then highlight the potential of this dataset to inform on crystal engineering in a variety of ways. One example is to explore how transformations around acyclic single bonds (*e.g.* a terminal methyl group becoming a Cl atom) affect the packing of a molecule. There are over 12,000 such transformations represented in the CSD, which allows us to determine which functional-group changes are more or less likely to disrupt or change the packing of a molecule in the solid state.

Keywords: CSD, matched molecular pairs, isostructurality

MS34-P13 Rational engineering of protein surfaces to improve crystallization while preserving solubility

Shikha Singh[1]

1. Biological Sciences, Columbia University in the city of New York

email: shikhasingh.bhu@gmail.com

Obtaining diffraction quality crystals remains the bottleneck for structure elucidation of most proteins using X-ray crystallography, which is the most precise method for atomic-resolution structure determination. A better understanding of crystallization should therefore help us identify both problematic areas of the process and potential solutions to this critical barrier. The simple observation that some proteins crystallize with minimal effort, while others are entirely resistant to crystallization, suggests that crystallization propensity is an intrinsic property of an individual protein. Based on our previously published work, we hypothesized that the principal determinant of this propensity is the prevalence on the protein surface of low-entropy epitopes with high intermolecular interaction potential. To find such “crystal packing epitopes”, we developed a simple topological scheme to identify local sequence motifs that are statistically overrepresented in crystal-packing interfaces in the Protein Data Bank. All residues within a minimum contact distance between chains are identified and then grouped into an ascending hierarchy ranging from the simplest elementary binary interacting epitopes to complete binary interprotein interaction interfaces. For counting and averaging purposes, protein chains are redundancy-downweighted to account for homologous chains forming similar crystals, as evaluated by a dot-product-like Packing Similarity Score. We show that introducing such “overrepresented crystal packing epitopes” into ten crystallization-resistant proteins generally improves their crystallization propensity without reducing their stability or solubility in aqueous salt or polyethylene glycol solutions. In some cases, their solubility is even improved, demonstrating that protein crystallizability and solubility are fundamentally separable properties. The most effective crystal-packing epitopes identified in our studies to date contain high-entropy polar sidechains constrained in low-entropy conformations by local interactions, so the mutations that we have used to improve crystallization generally involve introducing rather than removing high-entropy sidechains from the protein surface. Our results open a new approach to probabilistic engineering of protein surfaces to simultaneously enhance their stability, solubility, and crystallization propensity.

Keywords: protein crystallization, protein engineering, x-ray structure determination, crystallization-epitope engineering, data mining