

Data and Diversity Driven Development of a Shotgun Crystallisation Screen using the Protein Data Bank

G. Abrahams¹, J. Newman²

¹No affiliation given, ²No affiliation given

abrahams.gabi@gmail.com

Protein crystallisation has, for decades, been a critical and restrictive step in macromolecular structure determination via X-ray diffraction. Crystallisation typically involves a multi-stage exploration of the available chemical space, beginning with an initial sampling (screening) followed by iterative refinement (optimisation). Effective screening is important for reducing the number of optimisation rounds required, minimising the amount of sample required and reducing the cost and time required to determine a structure. Here, we propose an initial screen (Shotgun II) which has been derived from an analysis of recent Protein Data Bank (PDB) data and compare it with the previously derived (2014) Shotgun I screen. We clarify that the Shotgun approach entails finding the commercial crystallisation conditions which cover the most diverse space of proteins found in the PDB - which can be mapped to the well known Maximum Coverage problem in computer science. With this realisation we are able to apply a more effective algorithm for selecting conditions, such that the Shotgun II screen outperforms the Shotgun I screen both in protein coverage and quantity of data input. Our in-house data demonstrates that the Shotgun I screen, compared with alternatives, has been remarkably successful over the seven years it has been in use, indicating that Shotgun II is also likely to be a highly effective screen.

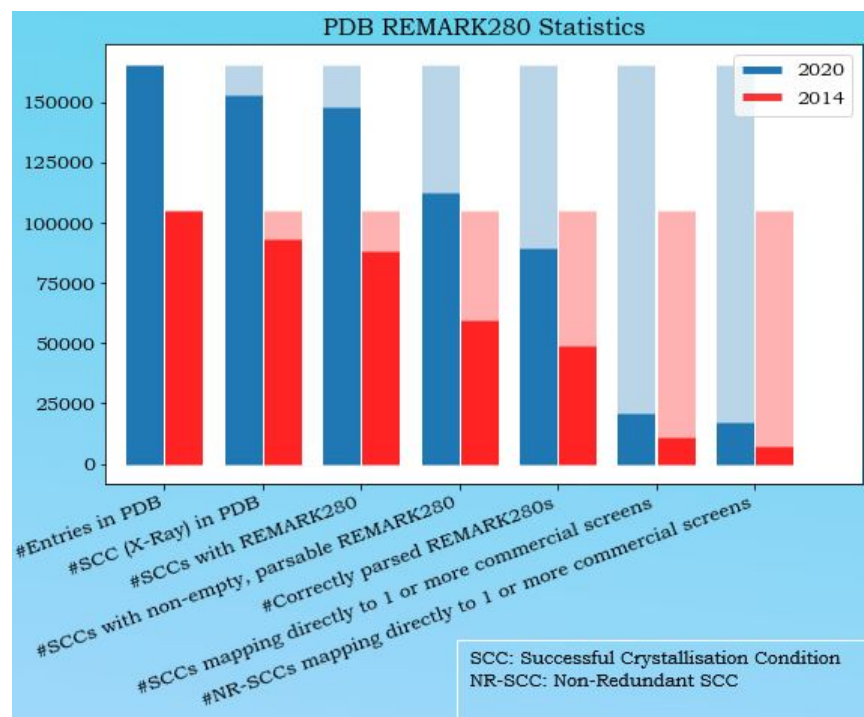


Figure 1