

Poster Presentation

MS102.P06

The \$100 million question: Do we need to apply informatics to crystallization?

J. Newman¹, V. Fazio¹, B. Schmidt¹, T. Peat¹

¹CSIRO, Biophysics program, Parkville, Australia

The PDB is currently growing at a rate of about 9000 structures annually, and 90% of these have been determined by X-ray diffraction methods. Each structure is the result of one or more crystals. Not every protein crystallises nor do all crystals diffract well enough; it has been estimated that of every 10 proteins that are purified, four will show some sign of crystallisation and one will crystallise robustly enough to obtain a structure¹. Using these numbers, and making some educated guesses (for example, that most proteins are tested in 1000 crystallisation trials¹) these 8000 structures represent 80,000 purified proteins, and 80,000,000 crystallisation trials which are set up each year. The cost of consumables, chemicals and direct labour to set up those trials varies, but can be estimated to be \$0.1- \$1, excluding the cost of the protein sample and any automation, suggesting that the structural biology community spends between 10 and 100 million each year on crystallisation. Any tools or insight that we can get from data mining or taking a computational approach to rationalize this process may not only profoundly change structural biology, but will make it much less expensive as well. We will discuss approaches to data mining, data standards, and software tools to enable a more rational approach to the process of crystallogensis.

[1] Newman, J., Bolton, E.E., Müller-Dieckmann, J., et al. , 2012. *Acta Cryst.* F68, 253–258.

Keywords: Crystallization, Standardization, Distance metrics