MS11-P03 | SIMBAD: STRUCTURE BASED SEARCH MODEL IDENTIFICATION FOR MOLECULAR REPLACEMENT USING THE PDB DATABASE.

Simpkin, Adam (University of Liverpool, Liverpool, GBR); Simkovic, Felix (University of Liverpool, Liverpool, GBR); Thomas, Jens (University of Liverpool, Liverpool, GBR); Savko, Martin (synchrotron SOLEIL, Gif-sur-Yvette, FRA); Lebedev, Andrey (STFC-UKRI, Didcot, GBR); Uski, Ville (CCP4, STFC, Oxford, GBR); Wojdyr, Marcin (CCP4, STFC, Global phasing, Oxford, GBR); Ballard, Charles (CCP4, STFC, Oxford, GBR); Shepard, William (synchrotron SOLEIL, Gif-sur-Yvette, FRA); Rigden, Daniel (University of Liverpool, Liverpool, GBR); Keegan, Ronan (CCP4, STFC, Oxford, GBR)

The conventional approach to identifying search models in molecular replacement (MR) is to use the sequence of the target to search against a database of known structures. This approach is based on the assumption that sequence similarity is a useful guide to structural similarity. While largely true, this strategy is not always effective: for example, conformational changes may mean that the most structurally similar Protein Data Bank (PDB) entries to the unknown structure are not those most closely related in sequence. A conceptually straightforward alternative approach to this problem would be to perform full MR on everything in the PDB. This strategy, however, would be extremely slow. SIMBAD is a CCP4 development that provides an alternate route to identify suitable search models. A rotation search is performed on every entry in the MoRDa database, a non-redundant derivative of the PDB containing ~100,000 structural domains. Through assessing the scores of these rotation searches, suitable search models can be obtained relatively quickly.

The SIMBAD pipeline also allows users to test if a contaminant has been crystalised in place of the desired protein. The use of SIMBAD post data collection on synchrotron beamlines has helped to avoid misspent research effort by quickly identifying these contaminants.

Recent developments include the use of Phaser and ensembles to enhance the sensitivity of SIMBAD as well as the deployment SIMBAD on CCP4 cloud resources.