MS02-P06

ContaMiner and *ContaBase*: Automated identification of unwantedly crystallized protein contaminants

Stefan Arold[1], Arnaud Hungler[1], Afaque Momin[1], Kay Diederichs[2]

1. Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
2. Fachbereich Biologie, Universität Konstanz, Konstanz, Germany

email: stefan.arold@kaust.edu.sa

Solving the phase problem in protein X-ray crystallography relies heavily on the identity of the crystallized protein, especially when molecular replacement (MR) methods are used. Yet, it is not uncommon that a contaminant crystallizes instead of the protein of interest. Such contaminants may be proteins from the expression host organism, protein fusion tags or proteins added during the purification steps. Many contaminants co-purify easily, crystallize and give good diffraction data. Identification of contaminant crystals may take time, since the presence of the contaminant is unexpected, and its identity unknown. We have established a webserver (*ContaMiner*) and a contaminant database (*ContaBase*), both available at strube.cbrc.kaust.edu.sa/contaminer/, to allow fast MR-based screening of crystallographic data against currently 76 known contaminants from more than seven different expression systems [1]. Here we present the latest developments of *ContaMiner*. Novel features include use of *UglyMol* [2] for an interactive online visualization of electron density maps of MR solutions, the possibility for screening of 'custom contaminants', and features for improved speed and performances. We are currently using *ContaMiner* to scan all PDB entries with suspiciously poor refinement statistics, and will present the results obtained.

References:

[1] Hungler, A., Momin, A., Diederichs, K. & Arold, S.T., ContaMiner and ContaBase: A web server and database for early identification of unwantedly crystallized protein contaminants. (2016) J. Appl. Cryst. 49, 2252-2258

[2] Wojdyr, M., UglyMol, a WebGL macromolecular viewer focused on the electron density. (2017) J. of Open Source Software 2. DOI: 10.21105/joss.00350

Keywords: contaminant, molecular replacement, web server

MS02-P07

Ensembling for molecular replacement: making the most of your distant homologues

Ronan Keegan[1], Daniel Rigden[2], Stuart McNicholas[3], Eugene Krissinel[1], Jens Thomas[2], Adam Simpkin[3], Felix Simkovic[2], Martyn Winn[1], Keith Wilson[4]

1. CCP4, STFC Rutherford Appleton Laboratory, Didcot, United Kingdom
2. Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom
3. Synchrotron Soleil, Paris, France
4. YSBL, University of York, York, United Kingdom

email: ronan.keegan@stfc.ac.uk

In the process of structure determination from a macromolecular crystallography diffraction experiment, the most successful approach to solving the well-documented phase problem has been and remains Molecular Replacement (MR). Typically, MR exploits the structural similarity between proteins that are evolutionarily related to derive the necessary phases from a known structure for an unknown target. The majority of proteins in the Protein Data Bank (PDB) have been phased in this way and, as this database expands, the chance of finding a suitable homologue for use in MR increases. Despite its success, the method requires a high degree of structural similarity between the homologue and the target in order for it to act as a suitable proxy for the target's phases. Given the availability of several structurally-similar homologues, an ensemble of such homologues can help to reduce this sensitivity. In addition, the structural alignment of these models to generate an ensemble can reveal common structural motifs or cores that are likely to be also present in the target. We present here two recent developments in the CCP4 suite designed to generate and exploit ensembles for use in MR. MrBUMP [1], an automated MR pipeline, can retrieve, align and truncate sets of known homologues to produce a set of ensembles. Alternatively, where there may be only one suitable homologue available, AMPLE [2], originally designed to exploit the use of ab initio generated search models in MR, can also now make use of the CONCOORD [3] application to produce a set of ensembles from a single known homologue. In addition, we present a related development in the molecular graphics application, CCP4mg. Through integration with the MrBUMP application, CCP4mg can assist in the generation of ensemble models by enabling the user to visually inspect and interactively edit the structural models.

References:

[1] Keegan, R.M. et al. (2018). Acta Cryst. D74, 167-182
[2] Rigden, D.J. et al. (2018). Acta Cryst. D74, 183-193
[3] de Groot, B.L. et al. (1997). Proteins 29: 240-251

Keywords: Molecular Replacement, Structure Solution, Macromolecular Crystallography