Poster

Demystifying CCP4 Cloud concepts: safeguarding structure determination through data structure and automation

Maria Fando, Eugene Krissinel, Andrey Lebedev, Oleg Kovalevskyi, Jools Wills, Ronan Keegan, Charles Ballard, Robert Nicholls, Ville Uski, David Waterman

Scientific Computing Department, Science and Technology Facilities Council UK, Didcot, OX11 0FA, United Kingdom

maria.fando@stfc.ac.uk

Advances in crystallography software have made structural determination more accessible and less reliant on understanding complex principles of macromolecular crystallography (MX). This opens up opportunities for researchers and eases access for those who are new to the field. While newcomers benefit from streamlined processes, they may be unfamiliar with efficient utilisation of computational tools, which may lead to confusion and potential mistakes in structure determination.

To address these challenges, Collaborative Computational Project Number 4 in Protein Crystallography (CCP4) [1] invests significantly in the development of graphical user interfaces, educational initiatives, and comprehensive software documentation. CCP4 graphical interfaces: CCP4i2 and CCP4 Cloud, facilitate easy access to over a hundred CCP4 programs.

CCP4 Cloud exploits the data-driven paradigm for MX project development, where decision-making relies on available data rather than predefined sequences of steps and data-in-promise [2]. CCP4 Cloud operates in terms of data objects, rather than files; an object may include information from several files. This design simplifies the overall data logistics by eliminating the burden of managing multiple files in job interfaces. As a result, the specification of job input becomes rather minimalistic and automatic, reducing the chance of making mistakes. Using data objects instead of files eliminates attempts of invoking tasks for which there is no data available. As a consequence, this allows implementing intelligent suggestions of follow-up tasks based on available data and combining tasks in workflows for common structure solution scenarios. Using automatic project workflows in CCP4 Cloud mitigates risks of suboptimal approach to structure determination and associated failures.

The data-driven approach naturally leads to organising structure solution projects into trees of jobs, based on data flows between tasks. This representation makes projects easy to navigate, analyse and revisit in future. Special data object: Structure Revision, was designed with properties that ensure unambiguous project branching. This presentation will discuss the data object model in CCP4 Cloud and how it streamlines crystallographic project development through effective data management; structured project organisation and varying levels of automation, enabling a highly efficient and ergonomic approach to structure solution process and safeguarding users from making common mistakes.

[1] Agirre, J., Atanasova, M., Bagdonas, H., Ballard, C. B., Baslé, A., Beilsten-Edmands, J., Borges, R. J., Brown, D. G., Burgos-Mármol, J. J., Berrisford, J. M., Bond, P. S., Caballero, I., Catapano, L., Chojnowski, G., Cook, A. G., Cowtan, K. D., Croll, T. I., Debreczeni, J. É., Devenish, N. E., Dodson, E. J., … Yamashita, K. (2023). The CCP4 suite: integrative software for macromolecular crystallography. Acta crystallographica. Section D, Structural biology, 79(Pt 6), 449–461. https://doi.org/10.1107/S2059798323003595

[2] Krissinel, E., Lebedev, A.A., Uski, V., Ballard, C.B., Keegan, R.M., Kovalevskiy, O., Nicholls, R.A., Pannu, N.S., Skubak, P., Berrisford, J., Fando, M., Lohkamp, B., Wojdyr, M., Simpkin, A.J., Thomas, J.M.H., Oliver, C., Vonrhein, C., Chojnowski, G., Basle, A., Purkiss, A., Isupov, M.N., McNicholas, S., Lowe, E., Trivino, J., Cowtan, K., Agirre, J., Rigden, D.J., Uson, I., Lamzin, V., Tews, I., Bricogne, G., Leslie, A.G.W. & Brown, D.G. (2022) CCP4 Cloud for structure determination and project management in macromolecular crystallography. Acta Cryst. D78: 1079-1089; doi:10.1107/S2059798322007987