## MS42-P03

### The expected log-likelihood gain for decision making in molecular replacement

Robert D. Oeffner[1], Pavel V. Afonine[2], Claudia Millán[3], Massimo Sammito[3], Isabel Usón[3], Randy J. Read[1], Airlie J. McCoy[1]

1. Department of Haematology, University of Cambridge, Cambridge, United Kingdom
2. Lawrence Berkeley National Laboratory, Berkeley, United States of America
3. Structural Biology, Molecular Biology Institute of Barcelona, Barcelona, Spain

**email: rdo20@cam.ac.uk**

Protein crystallographers often make assumptions about the solvability of a structure by molecular replacement based on two variables: the sequence identity between the model and target and the resolution of the data. We have recently shown that the solvability of a structure by molecular replacement is, rather, predominantly dependent on four variables: the number of reflections in the data set, the fraction of the scattering for which the model accounts, the RMSD between the model and target, and the measurement errors in the data. Furthermore, the solvability can be quantified with the eLLG (McCoy et al., 2017, Oeffner et al., 2018). The eLLG is the LLGI (Read & McCoy, 2016) expected from a correctly placed model, calculated as a sum of log-likelihoods of each reflection predicted by the model but offset by the sum of log-likelihoods of a model of random atoms. Using the eLLG, the crystallographer can judge whether to pursue molecular replacement or attempt experimental phasing as the quickest path to structure solution. Other applications of the eLLG include determining search order; finding the minimal data requirements for obtaining a molecular replacement solution using a given model; and for decision making in fragment based molecular replacement, in single atom molecular replacement, and for likelihood-guided model pruning.

References:

McCoy, A. J., Oeffner, R. D., Wrobel, A. G., Ojala, J. R. M., Tryggvason, K., Lohkamp, B. & Read, R. J. (2017). Proc. Natl. Acad. Sci. U. S. A. 1–5.
Oeffner, R. D., Afonine, P. V., Millán, C., Sammito, M., Usón, I., Read, R. J. & McCoy, A. J. (2018). Acta Cryst. D74, 245-255
Read, R. J. & McCoy, A. J. (2016). Acta Crystallogr. D. 72, 375–387.

**Keywords: Likelihood, eLLG, LLGI**

## MS42-P04

### jsCoFE, a cloud system for crystallographic computations from CCP4

Eugene Krissinel[1], Andrey Lebedev[1], Charles Ballard [1], Ville Uski [1], Ronan Keegan[1]

1. CCP4, Research Complex at Harwell, STFC-RAL, Didcot, United Kingdom

**email: eugene.krissinel@stfc.ac.uk**

The Collaborative Computational Project Number 4 in Protein Crystallography (CCP4) exists to maintain, develop and provide world-class software that allows researchers to determine macromolecular structures by X-ray crystallography and other biophysical techniques. Over 37 years, the CCP4 Software was assembled and distributed as an integrated Suite of programs, installable on either user's personal PCs or centralized facilities.

Modern trends in computing suggest a fast-growing interest to mobile platforms and cloud solutions for data storage and operations in practically all areas. In context of crystallographic computing, cloud solutions become increasingly appealing in view of recent advances in automated structure solution methods, which demand for both computing power and various databases makes them less suitable for offline (local) setups. Another appealing feature of the cloud model is the simplification of software and data management, both for software provider/maintainer and end users.

Over last decade, CCP4 have developed web services for automated structure solution, which are available to crystallographic community online. In this communication, we report on the development of jsCoFE (Javascript-powered Cloud Front-End), which expands CCP4 web-services to potentially all functional CCP4 components and allows a user to keep and operate their data and whole structure solution projects on-line. jsCoFE works on all computing platforms capable of running ordinary web-browsers (including smartphones and tablets), and does not require any local data storage. Currently available functionality, apart from automated solvers, includes data merging and scaling, phasing (MR and EP), density modification, model building, refinement, ligand fitting and structure analysis. Experimental data may be either uploaded from user's device or obtained directly from data producing facility, such as a synchrotron, online (currently limited to Diamond Ltd). Visualisation tools include UglyMol and Coot, which can be also used for model building and coordinate manipulations.

Technically, jsCoFE represents a network of web-servers, which includes head node(s) for keeping user projects and overall data logistics, and computational node(s) for performing actual computations. Owing the exclusive use of http(s) protocol for all communications within jsCoFE, there is no principal restriction on geographic location of any of the nodes, which makes the system almost infinitely scalable. In the opposite extreme, all nodes may be allocated on a single machine, then jsCoFE represents itself as a mere GUI.

jsCoFE is accessible at http://ccp4serv6.rc-harwell.ac.uk/jscofe/ and is available for custom installations ranging from individual desktops to central locations such as a laboratory, University/Institute or a synchrotron.

References:

Krissinel, E., Uski, V., Lebedev, A., Winn, M. & Ballard, C. (2018) Distributed Computing for Macromolecular Crystallography, Acta Cryst D 74(2), 143-151

**Keywords: WWW, Computing, Crystallography**

## MS42-P05

# X-ray crystallography to cryo-electron microscopy: computing infrastructure in structural biology

Jason Key[1], Peter Meyer[1], Carol Herre[1], Michael Timony[1], Dimitry Filonov[1], Justin O'Connor[1], Piotr Sliz[1]

1. Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, United States of America

**email: key@hkl.hms.harvard.edu**

SBGrid is an academic research group based at Harvard Medical School dedicated to structural biology computing. Started in 2000, the group's primary initiatives are the SB-Grid consortium, a software collaborative of more than 320 structural biology labs in 20+ countries (Morin, 2013), and the SBGrid DataBank which was established in 2015 as an open research data publication system for the Structural Biology community (Meyer, 2016). The SBGrid consortium provides a supported collection of scientist-created software titles for Crystallography, NMR, Electron Microscopy, Computational Chemistry, and Structure Visualization & Analysis with workshops and webinars to connect scientists with the scientist-developers who create the software. SBGrid DataBank supports validation and reproduction of macromolecular models and development of structural biology processing methods (Meyer, 2016). It currently contains primary data sets supporting over 400 macromolecular structures. Recent advances in cryo-electron microscopy (cryo-EM) detector technology have resulted in an increasing number of X-ray crystallographers turning to cryo-EM to determine structures of biological macromolecules too intractable or too difficult to determine by crystallographic techniques. The computational infrastructure used to support x-ray crystallography is only marginally capable supporting cryo-EM computations. In particular, efficient cryo-EM computation requires more, and higher performance storage; and places significant emphasis on GPU computation. Additionally, the general principles that making primary data available supports structural validation and methods development applies even more significantly to cryo-EM than it has with x-ray crystallography. Here we will present how various parts of SBGrid infrastructure including software, data management tools, as well as local and cloud high performance computing resources contribute to establishing a robust support for crystallography and cryo-EM experiments.

References:

Morin, A. et al. (2013) eLife, 2:e01456

Meyer, P.A. et al. (2016) Nature Communications, 7:10882, 1-12.

**Keywords: Cryo-EM, Computing, GPU**