Session IX - Scoring Methods

Tuesday, July 5 - afternoon

S9-L1

WHAT'S IN A DROP? MOVING FROM IMAGES TO OUTCOMES

Janet Newman[1], Thomas Carroll[2], Vincent J Fazio[3], Rongxin Li[4], Christopher J Russell[5], Chris Watkins[6], David Ratcliffe[7]

[1]CSIRO Manufacturing (Biomedical) 343 Royal Parade, Parkville VIC 3054
[2]University of Melbourne, Parkville VIC 3010
[3]CSIRO Minerals Private Bag 10, Clayton South VIC 3169
[4]Data61, Marsfield NSW 2122
[5]CSIRO Scientific Computing, North Ryde NSW 2113
[6]CSIRO Scientific Computing Private Bag 10, Clayton South VIC 3169
[7]Data 61, Acton ACT 2601

State of the art protein crystallization is a numbers game: as it is unlikely that the conditions under which any given macromolecule will crystallize can be deduced a priori, conditions must instead be found by experimentation.

Crystallization is a time-dependent trial and error sampling of the extremely large space of possible crystallization conditions: large number of conditions are tested, and each experiment is observed (often by imaging) at several time points. The ultimate goal is to have a consistent machine generated score for each image describing the outcome and then to correlate image similarity with condition similarity, building up an accurate picture of the phase diagram for any system. This would enable conditions for crystallization to be located, even if the initial set of experiments did not sample the appropriate set of experimental conditions in the space of all possible conditions.

Currently, automation is used routinely to miniaturize the experiments and to capture their results, but not to interpret the results of the experiments. We are interested in different approaches to using machine learning to interpret the results of crystallization experiments – what tools have already been developed, and how can they be best implemented in a practical and timely way? We will discuss progress of implementation, and compare and contrast existing approaches to automation of scoring. Finally, we will discuss the steps we are taking to find relationships between the experimental conditions and the outcomes of those experiments.

S9-L2

AUTOMATED SCORING OF CRYSTALLISATION EXPERIMENTS USING MULTIPLE IMAGES

J. Wilson

Departments of Chemistry and Mathematics, University of York, York, UK

Finding the conditions that will produce diffraction quality crystals can require many crystallization experiments. The use of robots has increased the number of experiments performed in most laboratories and, in structural genomics centres, tens of thousands of experiments can be produced each day. Visual inspection is becoming increasingly impractical and automated imaging systems are now used routinely to record the results of these experiments. Image analysis software has been developed a number of research groups [1-4] to provide scores, allowing the images from crystallization trials to be examined in order of merit and reducing the number that need to be examined by eye. However, scoring individual images does not take advantage of the fact that each experiment is assessed regularly over a period of time. As each new image is produced, further information about the experiment becomes available and changes between images can be encoded as additional features for classification. The more information that can be obtained, the greater the likelihood of correct classification and, in addition to analysis of the time-course images as a sequence, the information gained from UV imaging is considered. For example, although the drop in figure 1 is easily identified in the greyscale gradient magnitudes, the drop in figure 2 cannot be found. However, the additional information from the UV image taken at the same time, allows a mask to be found so that further processing is restricted to the crystallisation drop.

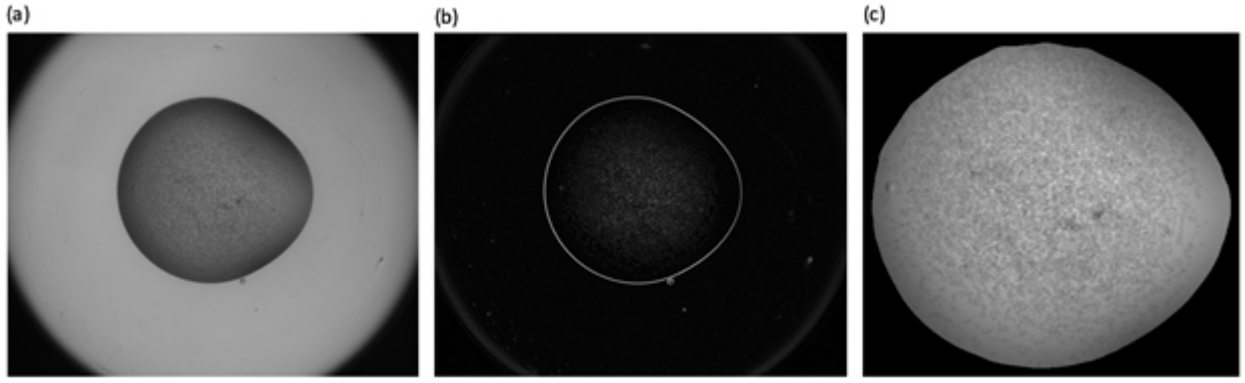

Figure 1. The greyscale gradient magnitudes (b) can be used to provide a mask for the crystallisation drop in (a).

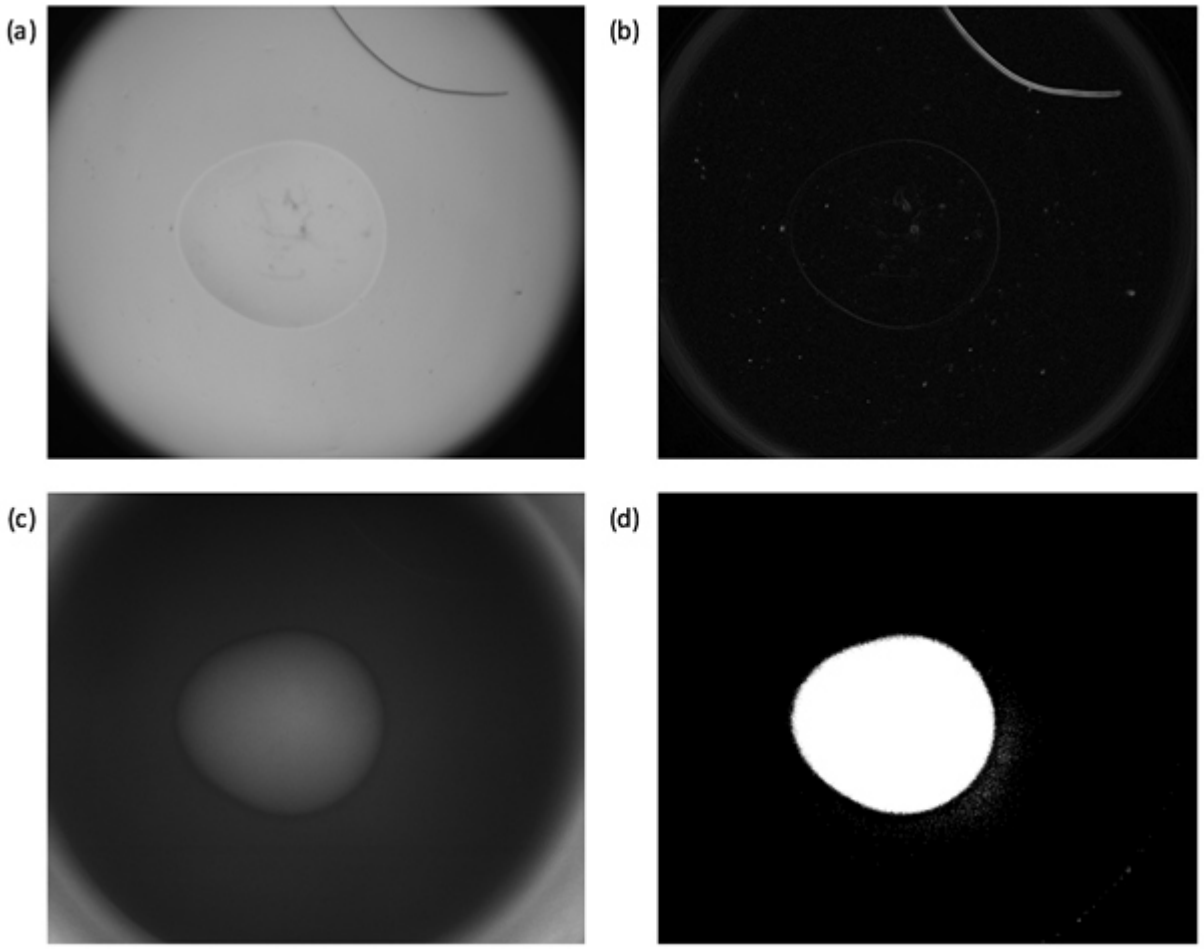

Figure 2. The lack of a defined boundary in (a) does not allow the drop to be identified by the gradient magnitudes (b). However, the UV image provides additional information that can be used to mask the drop.

1. S. Buchala, J. Wilson. *Acta Cryst. D. Biol. Crystallogr.*, **D64** (2008), pp. 823-833.
2. J.T. Ng, J.T., C. Dekker, M. Kroemer, M. Osborne, F.von Delft. *Acta Cryst. D. Biol.Crystallogr.*, **70** (2014), pp. 2702-2718.
3. C. Cumbaa, I. Jurisica, I. *J. Struct. Funct. Genomics*, **11** (2010), pp. 61–69.
4. S. Pan, G. Shavit, M. Penas-Centeno, D.H. Xu, L. Shapiro, R. Ladner, E. Riskin, W. Hol, D. Meldrum. *Acta Crystallogr. D. Biol. Crystallogr*. **62** (2006), pp. 271–279.

S9-L3

A GENETIC ALGORITHM FOR THE OPTIMIZATION OF PROTEIN CRYSTALLIZATION SCREENING

Samyam Acharya, Marc L Pusey, Ramazan S Aygun, Imren Dinc

[1]*DataMedia Research Lab, Computer Science Department, Univ. of Alabama in Huntsville, Huntsville, Alabama 35899, United States*
[2]*iXpressGenes, Inc., 601 Genome Way, Huntsville, Alabama 35806, United States*
sa0066@uah.edu

Protein crystallization screening focuses on determining the factors crucial for successful protein crystallization. The protein crystallization may require large number of parameters to be considered for setting up of cocktails that would yield suitable large crystals for X-ray data collection [1, 2]. These parameters include types of reagents, ionic strengths, types of salts, pH value of buffers, temperature, etc. [3]. Our goal is to implement a genetic algorithm which isolates combinations of reagents and concentrations that have a higher degree of synergy and potentially offer better crystalline outcome.

Combinations of reagents along with their concentrations are mapped into binary coded strings called *chromosomes* (not to be confused with biological chromosomes). Each chromosome represents a cocktail which is a certain set of buffer, pH, salts, etc. The length of a chromosome depends on the number of reagents we take into consideration for a particular experiment. Using expert score from previously conducted experiments, we identify new conditions generated by the algorithm in successive iterations. Undesired conditions, such as those that are known to cause phase separation and precipitates, are removed and favourable conditions are paired to produce the next generation of conditions. The top ranked conditions produced by the algorithm will be evaluated with respect to experiments conducted based on our associative experimental design [4].

The advantage of using genetic algorithm for protein crystallization screening lies in the ability of the algorithm to handle large number of parameters in an uneven search space environment. With this approach, we can employ selective pairing of conditions (chromosomes), which could be useful in identifying precipitant synergy for obtaining crystals and antergy (pairs that produce no crystals) and thus narrow down the screening process. The output conditions generated by the algorithm will be evaluated using the *Bin – Recall Metric* [4].

1. J. Jancarik and S.-H. Kim, *Sparse matrix sampling: a screening method for crystallization of proteins,* Journal of applied crystallography, vol. 24, no. 4, pp. 409–411, 1991.
2. A. McPherson and B. Cudney, *Optimization of crystallization conditions for biological macromolecules,* Structural Biology and Crystallization Communications, vol. 70, no. 11, pp. 1445–1467, 2014.
3. A. McPherson, *Crystallization of Biological Macromolecules*. Cold Spring Harbor Laboratory Press, 1999.
4. I. Dinç, M. L. Pusey, R. S. Aygün, *Protein Crystallization Screening Using Associative Experimental Design*, Bioinformatics Research and Applications, 2015.

S9-L4

AN ADD-ON DEVICE FOR AUTOMATED *IN SITU* DIFFRACTION SCREENING

A. Criswell[1], C. Acheson[1], P. Le Magueres[1], T. Hendrixson[1], C.Yang[2], Z. Zijian[3], J.D. Ferrara[1]

[1]*Rigaku Oxford Diffraction, 9009 New Trails Drive. The Woodlands, TX, 77389, USA*
[2]*Beijing Liangjia Deshi Technol.o Ltd. Co., Unit 209, Nonglinwei Build., 9, Haidian District, Beijing, P.R. China*
[3]*Rigaku Beijing Corp., 2601A, Tengda Plaza, No. 168, Xizhimenwai Ave. Haidian District, Beijing, 100044, P.R. China*
angela.criswell@rigaku.com

In situ screening of crystallization trials continues to be a topic of importance. *In situ* screening allows one to determine if the crystal-like object in a trial is salt, protein or something else of interest. Furthermore, crystal quality can be assessed before cryoprotection is attempted, providing a means to better optimize this step. Finally, some crystals are well-behaved enough that complete or nearly-complete data sets can be collected *in situ,* eliminating the need for harvesting and cryoportection altogether.

In this presentation we will show results from a newly developed automated *in situ* screening system designed specifically to work with legacy Rigaku diffractometers, extending their utility. The system is easy to install and deinstall so one can go from collecting data on cryo-protected samples to *in situ* screening in minutes. The system uses the same software as the Rigaku Oxford Diffraction PXScanner, CrystalEyes, making the transition between dedicated and part time systems easy.