

Robotic crystallomics at the Center for Eukaryotic Structural Genomics (CESG)

Craig A. Bingman, Simon T.M. Allard, Eduard Bitto, Byung Woo Han, Nathan S. Rosenberg, Janelle T. Warrick, Gary E. Wesenberg, and George N. Phillips, Jr.

University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin, USA 53706-1549, <http://www.uwstructuralgenomics.org>

Abstract

CESG solves structures of selected eukaryotic proteins. Here we report status and performance of our Tecan Genesis™ and Gilson C-250™ crystallization systems and CrystalScore™ and CrystalFarm™ imaging systems, and their integration with Sesame, our LIMS.

Our coupled system for crystallization screening and optimization uses uniform stock solutions and robotic methods. Sesame is the hub of a data- and robot-rich environment, describing all crystallization trials, writing imaging templates, and generating robotic control files. Future releases will close the loop from screening to associating conditions with images to final return of scores. This system will more efficiently and consistently develop crystals.

The screening success rate for fold-space targets is over 30%, and ~80% for test targets. We report analysis of our initial screening strategy and results from a salvage pathway encompassing alternative screens, perturbation screening, reductive methylation, and mutagenesis.

Screening, optimization, and salvage

Initial Screening on Tecan Genesis

CESG has adopted the Corning 3775 plate for initial screening. This 192-condition vapor-diffusion microtiter-format plate gives good performance in robotic setup and imaging. Crystallization well solutions are collated from two deep-well master blocks in eight-channel order on the Tecan. Each tray takes 25 minutes of dedicated setup time on the Tecan, giving us a screening capacity of about 32 trays per week, assuming a 80% duty cycle in a 40 hour work-week. The Tecan has independently chillable sample and crystallization tray positions, so trays can be set up for target temperatures of 4°C to ambient.

Optimization on Tecan Genesis

All optimizations at CESG are performed using the worklists written by the WELL module of Sesame, and executed on the Tecan Genesis. One of the most significant challenges we faced was accurately pipetting solutions of a wide range of viscosities and surface tensions, from aqueous mixtures of organic solvents, to 50%-w/v high molecular weight PEGs. We resolved this issue by establishing a solution class used for optimizations with the following characteristics.

1. Zero air gap between the system fluid and solution. This was critical for aspirating extremely viscous solutions. With air-gaps, cavitation and inaccurate dispensing were inevitable.
2. Over-aspirating to protect dispensed solution from mixing with system fluid. The degree of over-aspiration varies with the target volume.
3. Extensive washing between aspirations to prevent cross contamination.
4. Slow, contact dispensing.
5. Slow, liquid detection aspiration.

Salvage

Several salvage strategies for recovering diffraction-quality crystals from samples that fail in our initial screen are in progress. These include "rational mutagenesis" to remove clusters of highly charged residues and reductive methylation, and alternative screens. A pilot project involving three variants each of three different target proteins has resulted in crystals of one protein that failed to crystallize in initial screening that is now in optimization. Overall, the mutant proteins have been less soluble in than wild-type.

Reductive methylation (Rayment) has dramatically improved the quality of some target proteins. Two recent PDB depositions from CESG derive from reductively methylated samples (A11g07440, 1XQ1, above left. A15g48480, 1XY7, above right.)

Imaging and scoring

CrystalScore

The CrystalScore is a semi-automated X,Y,Z imaging stand capable of imaging all crystallization trials in a single plate in one operation. The CrystalScore system has been in use at CESG for nearly two years, and we have accumulated approximately 0.1 TB of scored image data to date. Two systems have been used, one at 4 °C, and the other at 20 °C.



CrystalFarm

The CrystalFarm system is a temperature-controlled combination plate hotel/imaging system capable of imaging 400 trays (300 microplate+100 microplate or VDX). The CrystalFarm was delivered in January, and is currently the pipeline 4°C imaging system for the project.



Image scoring is a split responsibility between technicians and Ph.D. level crystallographers. Technicians score the first week images, and crystallographers provide final scoring.

Scoring Scheme

Although there were early plans to capture a rich set of descriptors for each experiment, in practice, data entry was too tedious. A common integral 0-10 scheme compatible with single-keystroke data entry is now used in both CrystalScore and CrystalFarm.

0	No data	5	Microcrystals
1	Clear	7	Needles
2	Phase separation	8	Plates
3	Hopeless precipitate	9	3-D crystals <0.2 mm
4	Granular precipitate	10	3-D crystals >0.2 mm
5	Spherulites or dendrites		

Diffraction screening and data collection

Home Sources

CESG currently has access to a Rigaku R-Axis IV, X-Stream cryogenic stream on a rotating anode generator. A Bruker AXS Proteum R CCD detector, with a Bruno automated sample changer on a MICROSTAR generator has been accepted. The automated sample changer should help eliminate a significant bottleneck in our process, evaluating the large number of crystallization hits generated by our screening activities.

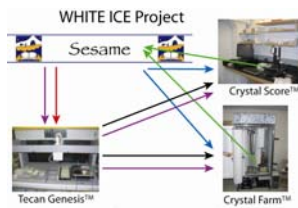
Synchrotron Sources

CESG currently collects synchrotron data sets at APS BIO-CARS, SBC-CAT and COM-CAT sectors. The University of Wisconsin has entered into a collaborative agreement with the LS-CAT consortium, and CESG is assured of access to that facility at APS upon its completion. Samples are mounted on pins geometrically conformant with the ESRF standard.

WHITE ICE

Wisconsin High-Throughput Extensible and Integrated Crystallization Environment

A highly integrated environment has been developed and implemented for the generation of crystals for CESG studies. Robotic and associated database tools allow for the management of crystallizing solutions, initial screening, imaging, scoring and optimization — all coupled to the Sesame laboratory information management system. The flexibility of SESAME to accommodate writing of barcodes and files and to accept files containing conditions make for an extensible system. Our conversion from a Gilson/Cybertab robot to the Tecan system illustrates the adaptability of the system. The Well module of Sesame increasingly functions as the control center for coordinating crystallization activities at CESG. To date, 228 unique proteins have been robotically screened for crystallization, ~300,000 images recorded, 87 unique proteins crystallized, 33 proteins optimized to produce diffraction quality crystals, and 23 progressed to PDB deposition as of 11/11/2004. Many are progressing through optimizations.



Fluidigm Topaz

Motivations and Experimental Design

As our project moves to microscale protein expression and purification trials to prove targets prior to investing in large-scale growths and protein purification, 100 micogram quantities of protein will become available for the incremental cost of concentration. This quantity of protein will be sufficient to run several hundred crystallization experiments in the Topaz free-interface diffusion environment, which requires around 10 nanoliters of sample per experiment.

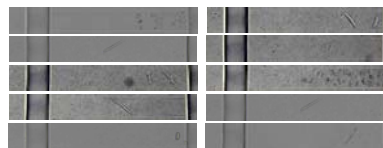
Our initial trials used eight pipeline protein samples. Four of these samples represent structures solved by CESG. The other four samples were "near misses." Two of the solved samples crystallized in a large number of conditions, whereas two samples crystallized in only one condition in our UW-192 screen. Two groups of crystallization reagents were used, UW-192, and two Fluidigm 96 condition screens.

Preliminary Results

Although we have just started this trial, it is already apparent that there is a good correlation between positive crystallization results in vapor diffusion experiments with these targets, and the results with UW-192 in Topaz microfluidic environment.

We are developing a strategy for evaluating novel leads from this system. One of the four "near miss" proteins seems to have promising leads, especially against the Fluidigm reagents.

Imaging



Our limited experience with the imaging system and auto-ranking software is generally positive. We are critically examining the joint results from Topaz and vapor diffusion trials to evaluate the false-negative rate with the auto-ranking software.

Screen performance

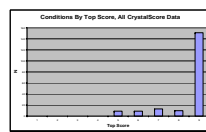
Methodology

Top unique scores for each condition for each target were extracted from all CrystalScore databases for all targets screened using UW-192 on the Tecan. Crystallization data at all temperatures and for pipeline and non-pipeline targets is presented below.

Results

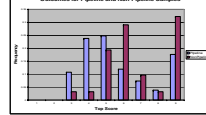
By Screen Solution

Analysis of screening outcomes for each unique solution in UW-192 at all temperatures shows that only 18 of 192 solutions have failed to produce results of needle grade or higher. 151 solutions have produced three-dimensional crystals (3+). This is substantially different than an analysis earlier this year, which indicated that only 96 solutions had given screening scores of 9 or higher.



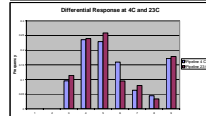
By Target

Screening pipeline targets against UW-192 consistently gives a different overall response than screening non-pipeline targets. The overall appearance of 3+ crystals for non-pipeline samples is 32%, and 18% for pipeline targets. The cause for this differential response is under investigation.



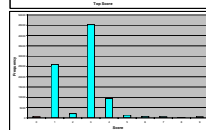
By Temperature

Overall crystallization response at both 4°C and 23°C is approximately equal. However, individual targets often crystallize at only one of the two temperatures.



By Individual Experiment

This figure illustrates the massive number of crystallization trials necessary to keep a high-throughput structure determination pipeline "fed." Only 1.5 percent of individual crystallization trials for pipeline targets give scores of 7 or higher, and only 0.7 percent give scores of 9 or 10. Individuals scoring are encouraged to report anomalous images (no droplet, droplet too small to be a creditable successful attempt) and the frequency of such failures in our process is stable at 0.6 percent, and is dominated by a few plates where not enough sample was present to complete a full screen.



1VMO A12g34160.1 unknown function



1Q4M A13g16900.1 TenA/TH4 protein



1Q4R A13g17210.1 unknown function



1VJ5 A11g23820.1 spermidine synthase



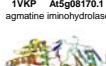
1VJH A11g24000.1 unknown function



1VJL A11g76680.1 12-oxophyllostequinone reductase (OPR1)



1Q44 A12g3760.1 steroid sulfotransferase



1VKT A15g08170.1 agmatine trimethylololase



1VQ4 A12g6050.1 12-oxophyllostequinone reductase (OPR3)



1XFI A12g17340.1 pantothenate kinase-like



1XMB A12g31350.1 pantothenate kinase-like



1VM9 T4mOC toluene-4-monooxygenase system

