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

Craig A. Bingman, Simon T.M. Allard, Eduard Bitto, Byung Woo Han, Jason G. McCoy, 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

1XRI At1g05000.1
unknown function

1XQ1 At1g07440.1
tropinone reductase-like

1XJ5 At1g23820.1

1VJH At1g24000.1
unknown function

1YCN At1g35720.1
anexin

1VJI At1g76680.1
12-oxophytodienoate reductase (OPR1)

1XMT At1g77540.1
putative acetyltransferase

1Q44 At2g03760.1
steroid sulfotransferase

1Q45 At2g06050.1
12-oxophytodienoate reductase (OPR3)

1XFI At2g17340.1
panthothenate kinase-like

1XYG At2g19940.2
unknown function

1XM8 At2g31350.1
glyoxalase II

1VMO At2g34160.1
unknown function

1Q4M At3g16990.1
TenA/THI-4 protein

Abstract

CESG solves structures of selected eukaryotic proteins. Here we report the status and performance of our integrated *WHITE ICE* system (Wisconsin HI-Throughput Extensible and Integrated Crystallization Environment) consisting of a Tecan Genesis™ crystallization platform, CrystalScore™ and CrystalFarm™ imaging systems, and Sesame, our LIMS. We also present a preliminary analysis of the Fluidigm Topaz™ microfluidic crystallization and imaging platform, and evaluate its performance relative to microliter-scale crystallization experiments. The relative performance of protein samples prepared by micro- and large-scale pipeline methods is also evaluated. 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 and Optimization on Tecan Genesis™

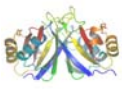
All optimizations at CESG are performed using the worklists written by the WELL module of Sesame, and executed on the Tecan Genesis™. One of the most significant challenges we faced was accurately pipetting solutions of a wide range of viscosities and surface tensions, from aqueous mixtures of organic 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 mutagensis" 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 than wild-type.

At1g07440.1



At5g48480.1

Reductive methylation (Rayment) has dramatically improved the quality of some target proteins. Two recent PDB depositions from CESG derive from reductively methylated samples (At1g07440.1, 1XQ1, above left; At5g48480.1, 1XY7, above right).

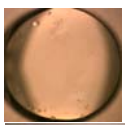
An extensive panel of "perturbation" screens have been devised and implemented at CESG. One 96-condition ion/pH perturbation panel is patterned after Nextal's screen. A 48-condition co-factor/ligand/osmolyte/detergent/organic perturbation screen has also been implemented. Although only a limited number of proteins have been subjected to perturbation salvage, in almost every case, novel crystal morphologies developed starting from a "near miss" (poor morphology, poor diffraction).

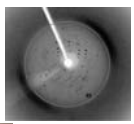
Case Studies

Orf 14914
At3g03250.1

INITIAL HIT
30%PEG 2000
200 mM K Glutamate
100 mM MES/Ac pH 5.5

Diffraction limit 6Å

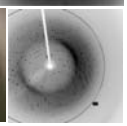




Perturbed Condition

Initial condition
With 5% DMSO
(Similar results with TMAO)

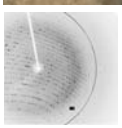
Diffraction limit 2.2Å

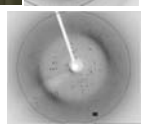


Orf 22797
At4g34215.1

Left, native crystal, highly mosaic
SeMet sample did not diffract beyond 8Å

Right, SeMet crystal, with GTP perturbant. Diffraction to ~3.5Å





Imaging and Scoring

CrystalScore™ and CrystalFarm™

CESG has used two semi-automated CrystalScore™ systems for over two years. Initially the sole pipeline imaging and scoring platform at ambient and 4ºC, CrystalScore™ is now the ambient pipeline imaging system. Approximately 0.2 TB of images have been aquired with this system to date.

CrystalFarm™ is fully automated and capable of storing up to 400 plates and imaging them on a pre-determined schedule or on demand. CrystalFarm™ has become sufficiently reliable for pipeline service, and is now our sole imaging platform for 4ºC crystallization experiments.

Diffraction screening and data collection

Home Sources

A Bruker AXS Proteum R™ CCD detector, with a BruNo™ automated sample changer on a MICROSTAR™ generator is CESG's home screening platform. The automated sample changer should help eliminate a significant bottleneck in our process, evaluating the large number of crystallization hits generated by our screen.

RT Screening

MicroRT™, MicroMounts™ and special goniometer bases designed for use with MicroRT™ capillary replacements facilitate screening of crystals directly from high-throughput microplates without first devising a cryoprotection strategy. For crystals that do not grow in an intrinsic cryosolvent, this screening methodology allows around eight crystals to be screened in the time it would take for one crystal to go through a novel cryosolvent. A conspicuous advantage of this methodology is that it rapidly reveals the native diffracting power of the crystals, prior to perturbation with cryosolvent.

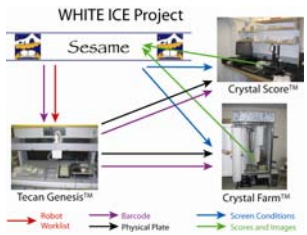
Synchrotron Sources

CESG currently collects synchrotron data sets at APS BIO-CARS, SBC-CAT and COM-CAT sectors, and will soon collect data at SER-CAT and SGX-CAT. 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 SPINE 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. Robotics and associated database tools allow for the management of crystal stock 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/Cyberlab™ 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, 236 unique proteins have been robotically screened for crystallization, ~300,000 images recorded, 88 unique proteins crystallized, 33 proteins optimized to produce diffraction quality crystals, and 28 progressed to PDB deposition as of 01/20/2005. 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 microgram quantities of protein will become available for the incremental cost of concentration. This quantity of protein would be sufficient to run several hundred crystallization experiments in the Topaz™ free-interface diffusion environment, which requires around 10 nL of sample per experiment.

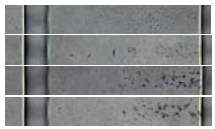
We have screened 16 unique targets and performed 88 distinct experiments. The primary objective has been to examine the suitability of Fluidigm as a "screen for crystallizability" by comparing results from free interface diffusion and vapor diffusion. We are also adapting our general screen to Fluidigm technology, by balancing water activity across individual chips. We are also evaluating the comparative efficacy of our general screen and Fluidigm reagents, and validating our micro-scale protein production process.

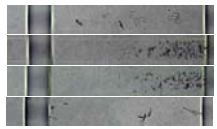
Preliminary Results

Based on the limited number of samples examined to date, a threshold result of 3 weak hits or one strong hit to trigger a large-scale growth and more extensive large-scale crystallization trials is appropriate. Under those conditions, Fluidigm screening would have produced a "go" signal for all targets crystallizable by vapor diffusion, and would have generated no false starts. Additional experiments are scheduled to more exhaustively define this threshold.

Rigorous elimination of false-positives is important for optimizing a two-tier screening strategy. We have worked with Fluidigm to improve the longevity and stability of the Topaz™ 4x96 plates. It is also important to optimize the duration of the experiment. Our preliminary results show that the vast majority of positive results are apparent within 24 hours, and that continuing imaging beyond 3-4 days generates obvious false positives.

Our preliminary results show no compelling advantage to using our 192 condition general screen or Fluidigm's proprietary reagents, after our screen was reformatted to control water activity.





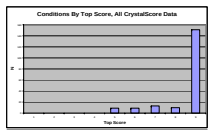
General screen performance (vapor diffusion)

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 for all temperatures and for pipeline and non-pipeline targets is presented below.

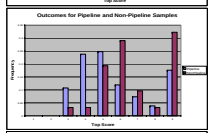
By Screen Solution

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. One-hundred fifty-one solutions have produced three-dimensional crystals (9+).This is substantially different than an analysis earlier this year, which indicated that only 96 solutions had given screening scores of 9 or higher.



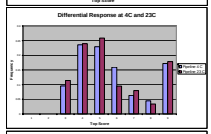
By Target

Screening pipeline targets against UW-192 consistently gives a different overall response than screening non-pipeline targets. The overall appearance of 9+ crystals for non-pipeline samples is 32%, and 18% for pipeline targets. The cause for this differential response is under investigation.



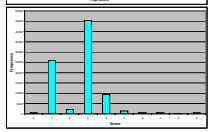
By Temperature

Overall crystallization response at both 4ºC and 23ºC 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 gave scores of 7 or higher, and only 0.7 percent gave 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.



1Q4R At3g17210.1
unknown function

1YQZ At3g21360.1
unknown function

1VK5 At3g22680.1
unknown function

1YDW At4g09670.1
unknown function

1XQ6 At5g02240.1
oxidoreductase

1VK0 At5g06450.1
unknown function

1VKP At5g08170.1
agmatine iminohydrolase

1YDH At5g11950.1
agmatine iminohydrolase

1VKV At5g18200.1
unknown function

1XY7 At5g48480.1
unknown function

1XMB At5g56660.1
panothenate kinase-like

1XQ5 Perch Hb
hemoglobin

1VM9 T4moC
toluene-4-monooxygenase system

CESG is supported by the National Institute of General Medical Sciences through the Protein Structure Initiative (P50 GM 64598).