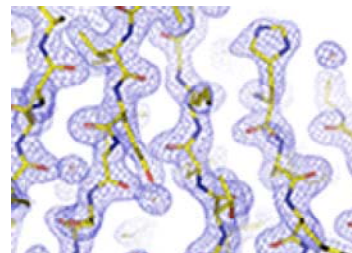


X-ray Crystallography

PRINT

The X-RAY CRYSTALLOGRAPHY Team, headed by George Phillips, Jr., PhD, is one of two groups responsible for determining three-dimensional protein structures. This team focuses on protein crystal production and analysis and has solved over 70 X-ray structures.



Goals for CESH X-ray Crystallography Team

- Target eukaryotic proteins for structure determination according to three criteria: unique sequences, medical relevance, and outside user requests.
 - Reduce amount of protein required for crystallization trials.
 - Improve yield of diffraction-quality crystals.
 - Rapid determination of three-dimensional protein structures.
- [Bioinformatics](#)
 - [Cell-Free](#)
 - [Cloning](#)
 - [Crystallography](#)
 - [Expression Testing](#)
 - [NMR](#)
 - [Protein Production](#)
 - [Protein Purification](#)
 - [Quality Assurance](#)
 - [Sesame](#)

Overview of X-ray Crystallography at CESH

The crystallomics pipeline capabilities include micro-scale screening on fluidic chips ([Fluidigm Topaz](#)), [Tecan Genesis](#) hanging / sitting drop robotics, and [Bruker Crystal Farm](#) plate handling and imaging. Over a million crystallization experiments have been conducted, with all images and their scores stored in a database for archival and research purposes. Structures are typically solved by selenomethionyl phasing using data collected at the [Advanced Photon Source at Argonne National Laboratory](#).

During the first few years, CESH focused on *Arabidopsis thaliana* as the source organism for selection of fold-space targets. This choice was made due to the relatively high quality of the sequencing of that organism and the completeness of a first draft of its resulting gene model. Genes were cloned from genomic DNA in intronless or cDNA libraries. Since that time, a large number of sequence-verified cDNA's have become available from the Mammalian Gene Collection (MGC), including primarily human and mouse full length cDNA's, but also from the genus *Ratus* (rat), *Bos* (cow), *Xenopus* (frog) and *Danio* (zebrafish). Over 50,000 full length clones are now available. Because these genes are verified as being expressed, they represent a tremendous source of starting materials, especially for structural genomics efforts where the attrition rates in the early stages of the pipeline are high.

The [Arabidopsis Biological Resource Center \(ABRC\)](#) has also begun distributing full length cDNA clones, and CESH has been able to continue projects that failed at the cloning step using libraries. More recently we have extended the general strategy of using thermophilic organisms as a source of protein sequences for structure determination. We have chosen *Cyanidioschyzon merolae* and *Galderia sulfuraria*, two "red" algal species with genomic data and gene models that are under development. These organisms grow well at temperatures up to 55°C. One-hundred and ninety-two clones have been chosen based on unique structures; many have been expressed and are under evaluation. The first protein from *Galderia* has recently been solved. The success rate for crystallization of proteins is somewhat better than workgroups from other organisms.

CESG develops crystals of proteins using an integrated platform of micro- and macro-scaled crystallization experiments. Microfluidic chips are employed for initial screening when precious reagents are used or when the initial amount of protein is small. Sitting drop experiments are used when at least 5 mg of protein is available. The largest labor saving implementation at CESH is not automating the setup of crystallization trials, but rather is imaging the experiments: several million images have been taken since the inception of CESH. Automatic plate handling and image capture improves the quality and reproducibility of the images, which are scored by trained students on a numerical scale. Optimization of

Center for Eukaryotic Structural Genomics - X-ray Crystallography

crystallization experiments follow, also using automatic image capture. CESG employs two Crystal Farm systems, one operated at 4°C and the other at 20°C.

Methods Being Investigated and Publications

+ Standard Protocol and Instruments

- Crystallization Screening and Optimization with Tecan Genesis (Corning EX/384, 192 condition screen)
- Crystallization Optimization with Gilson/Cyberlab C-250
- Crystallization Imaging with CrystalScore System (20°C and 4°C)
- CrystalFarm Automated Tray Hotel and Imaging System
- CCD detector, high-brilliance rotating anode, automated sample changer
- Home source diffraction with R-axis 4, conventional rotating anode source
- SeMet MAD at BioCARS, SBC-CAT and COM-CAT (APS)

+ Integrated Robotic Crystallization and Optimization

CESG has developed a tightly coupled system for initial screening and optimization of crystallization conditions that utilizes a uniform set of stock solutions and methods for robotic handling. Outcomes of initial crystallization screening experiments are currently recorded in [CrystalScore](#) databases and in the Well module of the [Sesame \(LIMS\)](#) system.

Using 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 compatibility with microscale Fluidigm, mesoscale Mosquito, and macroscale Tecan instruments for crystallization experiments and the Crystal Farm imaging robots illustrates the adaptability of the system. Software has been developed for managing optimization of initial hits that is well integrated.

+ Automatic Interpretation of Electron Density Maps (ACMI)

One particularly time-consuming step in protein crystallography is interpreting the electron density map; that is, fitting a complete molecular model of the protein into a 3D image of the protein produced by the crystallographic process. In poor-quality electron density maps, the interpretation may require a significant amount of a crystallographer's time. We have investigated automating the time-consuming initial backbone trace in poor-quality density maps. We describe ACMI (Automatic Crystallographic Map Interpreter), which uses a probabilistic model known as a Markov field to represent the protein. Residues of the protein are modeled as nodes in a graph, while edges model pairwise structural interactions. Modeling the protein in this manner allows the model to be flexible, considering an almost infinite number of possible conformations, while rejecting any that are physically impossible. Using an efficient algorithm for approximate inference 'belief propagation' allows the most probable trace of the protein's backbone through the density map to be determined. We have tested ACMI on a set of density maps (at 2.5 to 4.0 Å resolution) and have shown that ACMI offers a more accurate backbone trace than current approaches.

DiMaio, F., Shavlik, J., Phillips, G.N., Jr. (2003) **Using pictorial structures to identify proteins in X-ray crystallography density maps.** *Working Notes of the ICML Workshop on Machine Learning in Bioinformatics*, August.

DiMaio, F., Shavlik, J., Phillips, G.N., Jr. (2006) **A probabilistic approach to backbone tracing in electron density maps.** *Bioinformatics* 22(14):e81-9. [[16873525](#)]

DiMaio, F., Soni, A., Phillips, G.N., Jr. and Shavlik, J. (2008) **Improved methods for template-matching in electron density maps using spherical harmonics.** *IEEE-BIBM 2007 Conference Proceedings* (in press).

+ Ensemble Refinements of Protein Structures

X-ray crystallography typically uses a single set of coordinates and B-factors to describe macromolecular conformations. Refinement of multiple copies of the entire structure has been previously used in specific cases as an alternative means of representing structural flexibility. Here, we systematically validate this method using simulated diffraction data, and find ensemble refinement produces better representations of the distributions of atomic positions in the simulated structures than single conformer refinements. Comparison of principal components calculated from the refined ensembles and simulations shows that concerted motions are captured locally, but correlations dissipate over long distances. Ensemble refinement is also used on 50 experimental structures of varying resolution, and leads to decreases in R-free, implying that improvements in the representation of flexibility observed for the simulated structures may apply to real structures. These

Center for Eukaryotic Structural Genomics - X-ray Crystallography

improvements in the representation of normally observed conformational ensembles may apply to real structures. These gains are essentially independent of resolution or data-to-parameter ratio, suggesting even structures at moderate resolution can benefit from ensemble refinement.

Levin, E. Kondrashov, D., Phillips, G.N., Jr. (2007) **Ensemble refinement of protein crystal structures: validation and application.** *Structure* 15(9):1040-52. |[17850744](#)|

[+ Functional Studies By X-ray Crystallography](#)

Structures solved by CESG range from proteins of completely unknown function to those with a putatively assigned function based on structure or sequence homology to those with a highly defined function. Functional studies may be carried out by scientists interested in a specific solved protein. X-ray crystallography is occasionally used by CESG to acquire additional functional information on its targets. Crystal structures of complexes between the protein and a substrate analogue or cofactor can provide important clues as to the functional identity of the protein, the specific determinants for ligand specificity, and snapshots of the catalytic cycle that can reveal enzyme mechanism.

[+ Future Instrumentation](#)

- [Advanced Computer Graphics Development](#)
- [Robotic Crystal Manipulator](#)

X-ray Crystallography Team

- George Phillips, Jr., PhD (PI)
- Craig A. Bingman, PhD (Team Leader)
- Angel Villahoz-Baleta, PhD

Center for Eukaryotic Structural Genomics (CESG), Department of Biochemistry
University of Wisconsin-Madison, 445 Henry Mall, Madison, WI 53706 | [Map](#) |
Telephone: 608.263.2183 Fax: 608.890.1942 Email: cesginfo@biochem.wisc.edu

| [Biochemistry](#) | [Private](#) | [PDB](#) | [PSI](#) | [PSI-KB](#) | [PSI-MR](#) | [Sesame](#) | [UW-Madison](#) |
Copyright © 2006 The Board of Regents of the University of Wisconsin System