

SBGrid Update

April 10th, 2014

Piotr Sliz

attending AHM



Mick
Timony



Ben
Eisenbraun



Andrew
Morin



Jason
Key



Justin
O'Connor



Paul
Sanschagrin

- Wide Search Molecular Replacement
 - Update
 - Future Plans
- Low resolution refinement with DEN
- SBGrid

Wide Search Molecular Replacement (WSMR)

- WSMR provides a specialized tools for difficult X-ray crystallography structure determination problems.
- WSMR should be used in cases where other structure determination approaches fail.
 - Detecting models with low sequence identity
 - Selecting the best model from a large library
 - Detecting contaminants (often not reported, but important for convergence of the project!)
- The method was described by Stokes and Sliz in PNAS (2010)
- The method is specific to proteins. We do not offer RNA or DNA support at this point.

WSMR - job requirements

- In 2010-11 we anticipated to support ~20 WSMR jobs submission per calendar year (ie 20 unique end-users)
- For each WSMR job we execute ~100,000 molecular replacement (MR) searches.
- Each MR search requires direct access to three files:
 - Experimental data: X-ray diffraction dataset (this is identical for all MR searches within a single WSMR job)
 - Model: A protein databank file (each MR will access a unique file)
 - An input file
- WSMR returns a graph with Log Likelihood Gain scores and Translation Function scores. A small set of protein databank files with best scores are then chosen for structure refinement.

WSMR - User Interface

- Jobs submitted through SBGrid Portal Interface
- Portal open to all members of the community
- After internal evaluation all jobs dispatched to OSG
- A graph with WSMR is returned to a user

The screenshot displays the SBGrid Science Portal interface. At the top, there is a navigation bar with links for Home, Applications, Resources, Support, and About, along with Register and Login options. The main content area features a central banner with three sections: 'Wide Search Molecular Replacement' (WSMR), 'Deformable Elastic Network' (DEN), and 'SBGrid Consortium'. WSMR is described as searching up to 100,000 domains for molecular replacement models given X-ray structure factors. DEN is described as producing high resolution structures from low resolution X-ray data by only refining known restraints and the use of homology models. The SBGrid Consortium provides a distribution of over 200 scientific software packages with monthly updates to member labs. To the right of the text is a diagram illustrating the DEN refinement process, showing a protein structure being refined from a low-resolution state to a high-resolution state. Below the banner, a paragraph states that the SBGrid Science Portal provides access to web-enabled structural biology applications, data sharing facilities, biological data sets, and other resources valuable to the computational structural biology community. This is in addition to the core SBGrid program of software distribution and scientific computing technical support. The page is divided into two main sections: 'Services' and 'Consortium'. The 'Services' section lists the following: SBGrid has been funded by the NSF to encourage the adoption of advanced computational methods, grid computing facilities, and collaborative data-sharing systems within two communities: Structural biology labs which are part of the SBGrid consortium and Biomedical researchers in New England. As such, we provide the following services and facilities: Consulting expertise to enable advanced computational methods for researchers who are in need; Issuance of digital certificates and Virtual Organization membership to enable grid access; Access to shared data storage facilities with user-driven access control; Access to large scale compute clusters, either locally (within SBGrid) or nationally (via Open Science Grid); Assistance commissioning grid computing resources or deployment of applications onto a grid infrastructure. The 'Consortium' section describes NEBioGrid as a consortium of bio-medical researchers in the New England area and is hosted at Harvard Medical School by the Sliz Lab, in cooperation with the SBGrid Core team. The consortium has four main goals: Share experience, requirements and best practice for federated, collaborative e-Science; Facilitate access to shared cyberinfrastructure, such as Open Science Grid and TeraGrid; Provide a web portal interface to shared cyberinfrastructure; Develop and deploy tools and computational workflows specific to our community. A logo for the NSF is shown in the bottom right corner, with text indicating funding from the National Science Foundation.

Register | Login

SBGrid Science Portal

Home Applications Resources Support About

Wide Search Molecular Replacement
WSMR searches up to 100,000 domains for molecular replacement models given X-ray structure factors.

Deformable Elastic Network
DEN produces high resolution structures from low resolution X-ray data by only refining known restraints and the use of homology models.

SBGrid Consortium
SBGrid provides a distribution of over 200 scientific software packages with monthly updates to member labs.

Schröder 2010 DEN refinement

Launch DEN

The SBGrid Science Portal provides access to web-enabled structural biology applications, data sharing facilities, biological data sets, and other resources valuable to the computational structural biology community. This is in addition to the core SBGrid program of software distribution and scientific computing technical support.

Services

SBGrid has been funded by the NSF to encourage the adoption of advanced computational methods, grid computing facilities, and collaborative data-sharing systems within two communities:

- Structural biology labs which are part of the SBGrid consortium
- Biomedical researchers in New England

As such, we provide the following services and facilities:

- Consulting expertise to enable advanced computational methods for researchers who are in need.
- Issuance of digital certificates and Virtual Organization membership to enable grid access.
- Access to shared data storage facilities with user-driven access control.
- Access to large scale compute clusters, either locally (within SBGrid) or nationally (via [Open Science Grid](#)).
- Assistance commissioning grid computing resources or deployment of applications onto a grid infrastructure.

Consortium

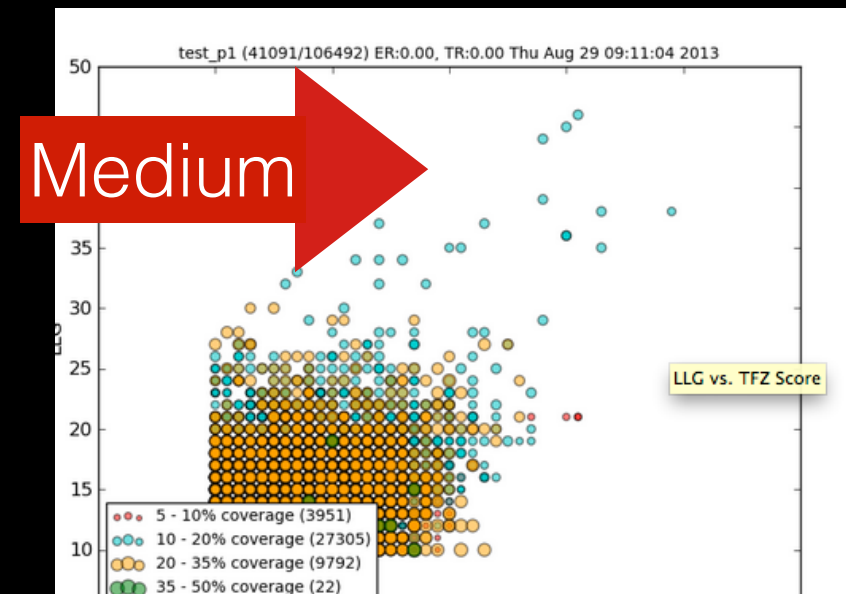
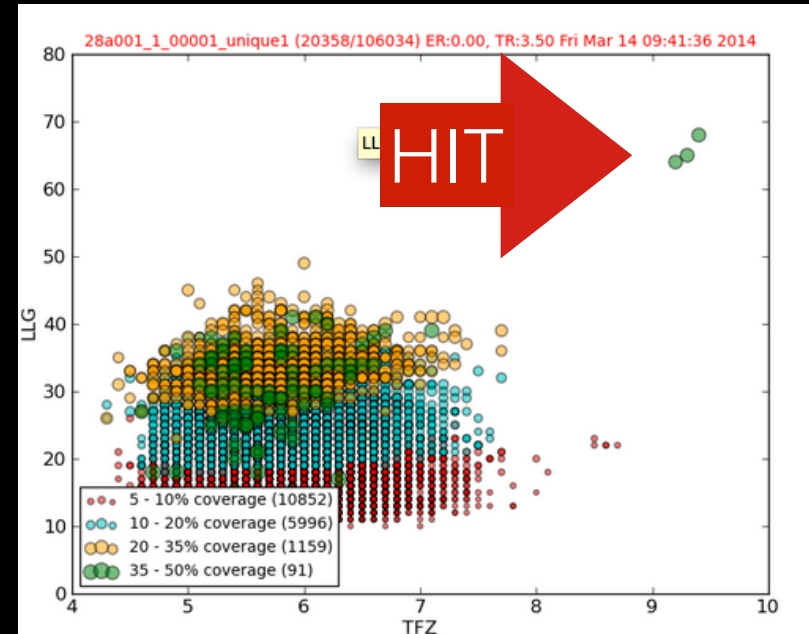
NEBioGrid is a consortium of bio-medical researchers in the New England area and is hosted at [Harvard Medical School](#) by the [Sliz Lab](#), in cooperation with the SBGrid Core team. The consortium has four main goals:

- Share experience, requirements and best practice for federated, collaborative e-Science
- Facilitate access to shared cyberinfrastructure, such as [Open Science Grid](#) and [TeraGrid](#).
- Provide a web portal interface to shared cyberinfrastructure
- Develop and deploy tools and computational workflows specific to our community

With funding from the National Science Foundation

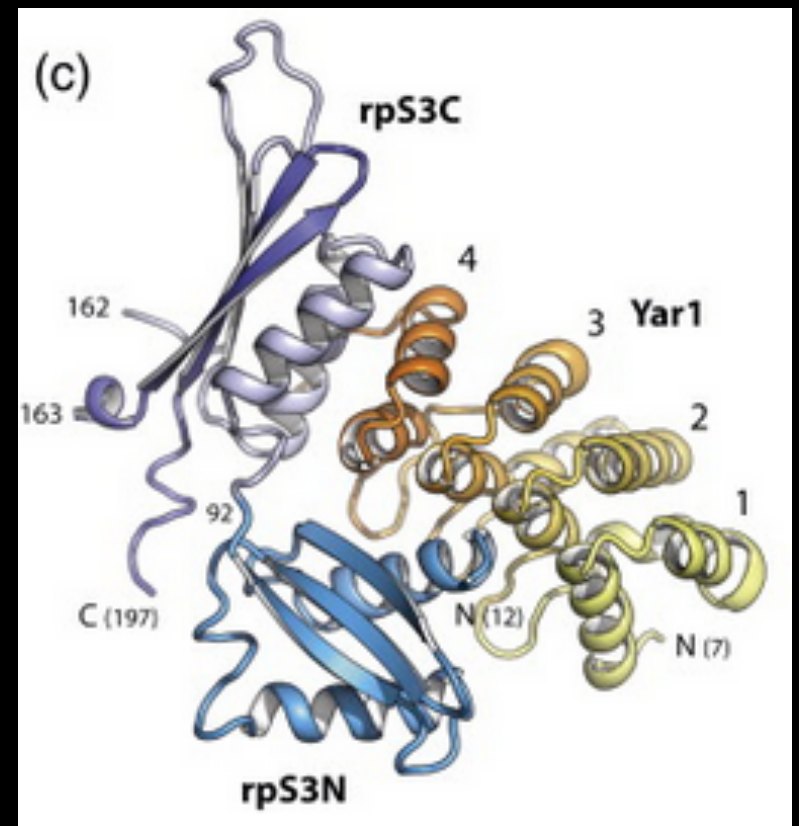
WSMR Statistics

- Between April 1st 2013 and March 30th 2014 SBGrid supported 97 WSMR Jobs.
- 10 jobs were flagged as “Hits”
- 15 jobs were flagged as “Medium”



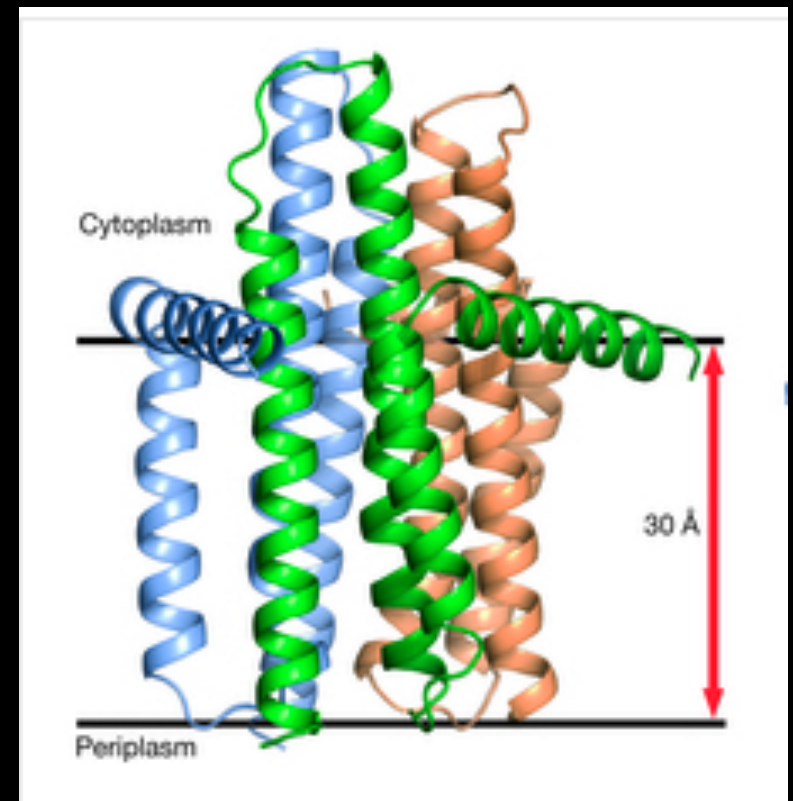
WSMR Citations

- Yeast Ribosomal Protein rpS3 in complex with Yar1
 - ▶ The crystal structure of rpS3 in complex with Yar1 was solved using wide search molecular replacement [REF] followed by manual rebuilding in Coot
 - ▶ JMB 2013.



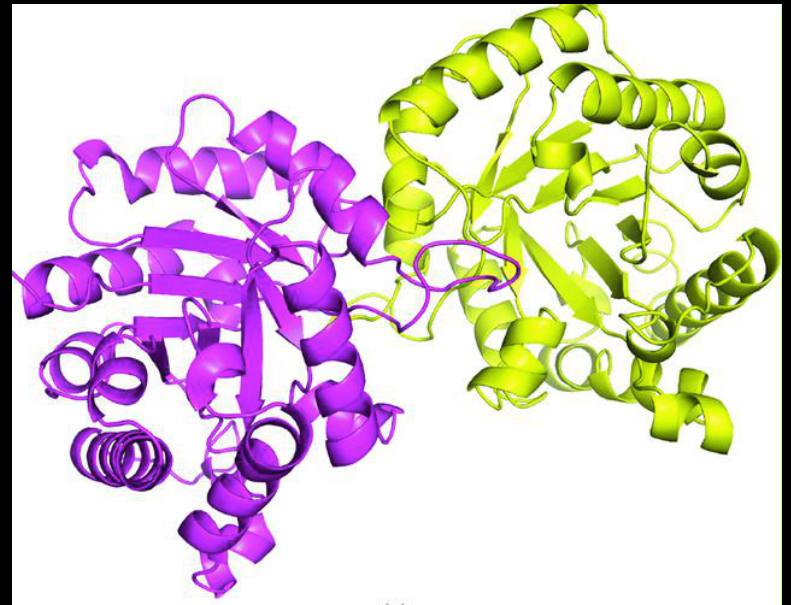
WSMR Citations

- Integral Membrane diacylglycerol kinase
 - Nature 2013.



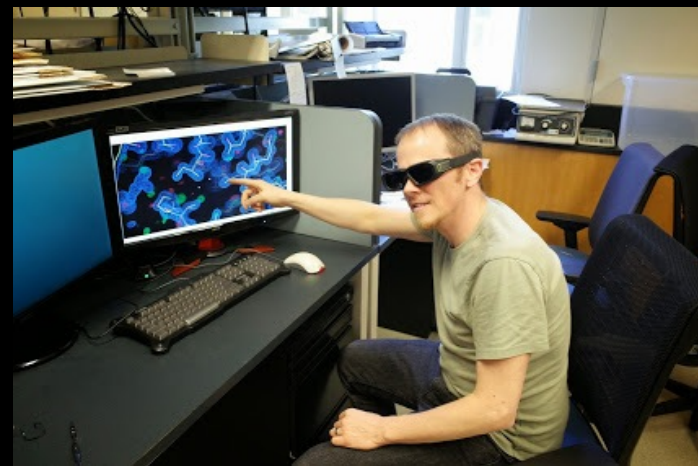
WSMR Citations

- Triosephosphate isomerase
 - ▶ WSMR yielded more than 100 hits, with the best being the triosephosphate isomerase from *E. coli*.
 - ▶ Acta Cryst. F



Summer 2014 - FAB Library

- Immunoglobulin Fab fragments
 - ▶ Used as crystallization scaffolds
 - ▶ Have a unique elbow angle and therefore are tricky to use as MR models
 - ▶ Large library of Fab available in PDB
- Summer Project (in collaboration with Harrison)
 - ▶ Analyze existing FAB dataset
 - ▶ Expand the dataset with MD
 - ▶ Create a specialized library for WSMR
- Participants:
 - ▶ Jason Key, PhD (research computing specialist)
 - ▶ An undergraduate student from Swarthmore College

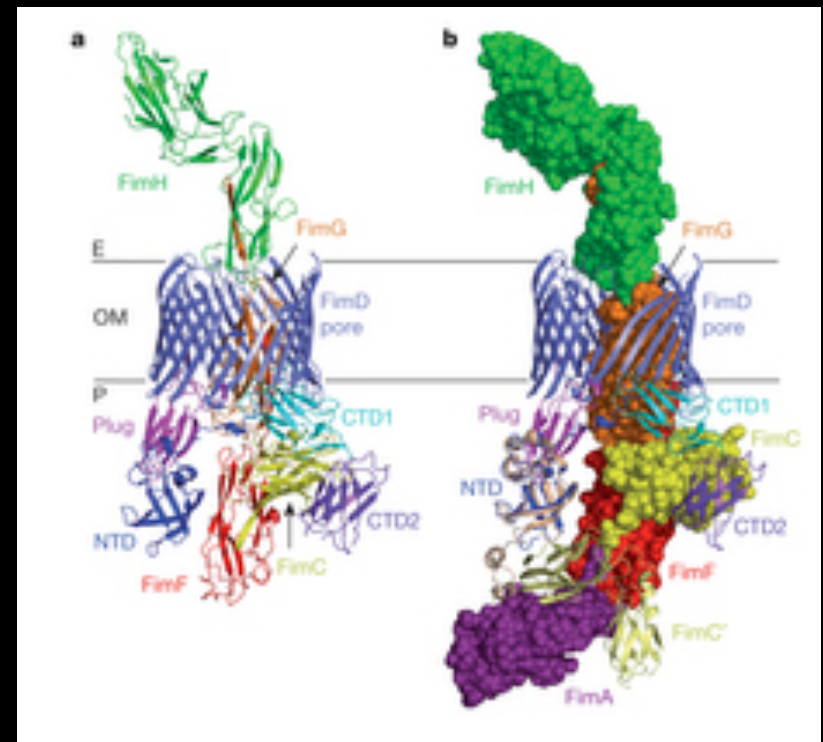


DEN - low resolution refinement

- O'Donovan et al., 2012
- Low-resolution refinement for difficult/large macromolecular assemblies.

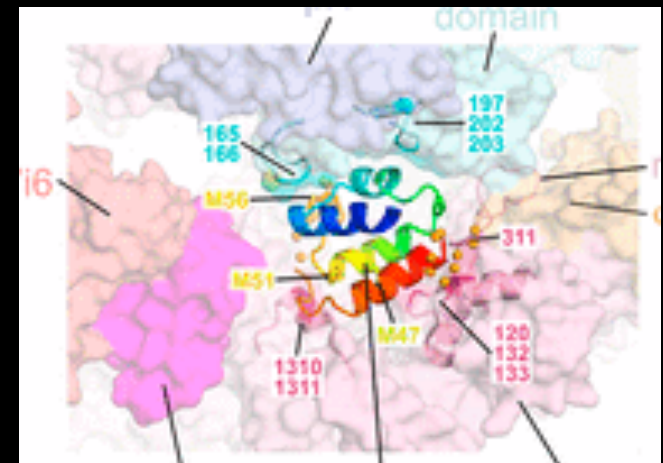
DEN - citations

- Outer membrane assembly
 - FimD
- “After a two-dimensional grid search for the DEN parameters using the SBGrid Science Portal web service, DEN parameters γ and ω were set to 0.2 and 100, respectively”
- Nature, 2013



DEN - citations

- T7 inhibition of RNA polymerase
- “deformable elastic network refinement with noncrystallographic symmetry restraints using CNS 1.3 performed on the Structural Biology Grid cluster”
- PNAS, 2013



SBGrid publication



- “SBGrid also provides the wider structural biology community with access to supercomputing facilities across the US in collaboration with **the Open Science Grid (OSG: Pordes et al., 2007)** via the SBGrid Science Portal. Currently, SBGrid provides access to two grid-enabled services: the Wide-Search Molecular Replacement (WSMR; Stokes-Rees and Sliz, 2010) service for determining crystallographic phase using the Phaser program (McCoy et al., 2007), and the Deformable Elastic Network (DEN) service for refining low-resolution electron density data (O’Donovan et al., 2012; Schröder et al., 2010).”

2014/15

- Continue to utilize OSG for WSMR and DEN.
- Evaluate OSG Connect. e.g, can OSGC can provide a workstation-level tool for OSG job submission?