Structural Biology on the Grid
SBGrid Research Coordination Network
Harvard Medical School
Summary

SBGrid Background
Grid Outreach
Portal Infrastructure
Molecular Replacement
Objectives and Priorities
SBGrid

- Consortium of structural biology labs
  - Involved in structure determination of (primarily) proteins
  - X-Ray crystallography, NMR, Electron Microscopy
- 87 member labs across the US
  - 28 at Harvard & Boston Academic/Medical hub
- 90 software packages
- Modest local cluster resource
  - 300 cores across several clusters (Intel, Mac, AMD)
- Now developing web-based portal interfaces to key apps

Ian Stokes-Rees, http://sbgrid.org
The SBGrid Team

Ian Stokes-Rees, http://sbgrid.org
Welcome to SBGrid

Structural Biology Grid (SBGrid) is a computing collaboration of several X-ray crystallography, NMR and electron microscopy laboratories. Participating laboratories include groups primarily at Harvard Medical School, Harvard University and Yale Medical School, but our alumni often remain members after becoming principal investigators at other institutions.

Explore SBGrid.org

- See what SBGrid can do for you
- See the labs currently affiliated with SBGrid
- Browse the software that SBGrid configures for its affiliates
- Learn more about the benefits of Grid Computing
- See how SBGrid is leveraging the power of OS X for scientific applications
- Download publically-available software released by SBGrid

Latest News

- SBGrid User Summit: Quo Vadis Structural Biology?
  SAVE THE DATE! The first ever SBGrid user summit ‘Quo Vadis Structural Biology?’ will take place on May 5th and 6th 2008 in Boston, Massachusetts. We will have a number of talks and workshops focused (read more...)
- Grid computing specialist joins SBGrid
- Experimental cluster acquired under NSF grant
- Apple Worldwide Developers Conference
- New Consortium Members
- SBGrid Receives Taplin Funds for Discovery Award for Grid Computing

Ian Stokes-Rees, http://sbgrid.org
SBGrid Services

- Application optimization
- Application packaging and automated distribution/update
- Help desk
- Centralized license management

Soon:
- Grid portal for SB applications
- Gateway to OSG
Motivation for Grid

Because **computational requirements** continue to be a bottle neck

Because **complexity of tools** impedes quality and efficiency of scientific investigation

Because some affiliated labs don’t have large compute clusters available to them

Because **new computationally intensive methods** are being developed
Portal Infrastructure
SBGrid Architecture

Portal
- Web Server
- Portlet Interfaces
- Web File Store

Credential Server
- Cred DB

App Mgr
- App i/f
- Gen i/f
- Job DB

OSG Compute Resources
- Exec Mgr
  - OSG i/f
  - Condor i/f
  - SGE i/f
- Exp. Cluster
  - Opteron x4
  - Xeon x16
  - XeonMac x8

SBGrid Cluster
- Athlon x4
- XeonMac x40
- XeonMac x112
- G5 Mac x28

SBGrid User Files

Web File Store

User Files

SBGrid

Web Server

Base i/f

Portlet

Interfaces

Web

Server

Store
Observations

- Fairly deep software stack
- Development and debugging difficult
- Documentation of existing components often sketchy
- New domains: hard to predict what users want or need
Grid Outreach
Bringing New Users to the Grid

- SBGrid consortium
  - VO administrators for existing community
  - Setting up new users with OSG accounts
  - Portal infrastructure to access grid resources and expose “Bio” applications
  - Eventually aim to bring member lab computing resources into OSG

- Northeastern University
  - Partnering with bio-informatics group to introduce OSG
  - Providing 6 month internship for undergraduates to deploy/expand SBGrid

- Boston Latin School
  - Public exam school, one of the top in US
  - Running computational biology seminars and job shadowing
  - Summer grid computing internship

Ian Stokes-Rees, http://sbgrid.org
Boston Latin School half
day computational
structural biology seminar
Molecular Replacement
Fourier series for electron density is a sum of contributions from individual reflections.

Ian Stokes-Rees, http://sbgrid.org
Phase Problem

$F_{\text{HKL}}$

- **Amplitudes**
  - can be measured
  - $\sim$ sq rt of intensity

- **Frequency**
  - Fixed and known from X-ray source

- **Phase**
  - Unknown!
Molecular Replacement

Homologous or incomplete model:

Combining model phases with experimental intensities will reveal details of missing elements

Typically 30% identity and 1/3 of a structure required.
Experimental Phasing

Transglycosylase

Lysozyme

PHAGE

3 months Too Slow!

Ian Stokes-Rees, http://sbgrid.org
Welcome to the RCSB PDB

The RCSB PDB provides a variety of tools and resources for studying the structures of biological macromolecules and their relationships to sequence, function, and disease.

The RCSB is a member of the wwPDB whose mission is to ensure that the PDB archive remains an international resource with uniform data.

This site offers tools for browsing, searching, and reporting that utilize the data resulting from ongoing efforts to create a more consistent and comprehensive archive.

Information about compatible browsers can be found here.

A narrated tutorial illustrates how to search, navigate, browse, generate reports and visualize structures using this new site. (This requires the Macromedia Flash player download.)

Comments? info@rcsb.org

Molecule of the Month: Small Interfering RNA (siRNA)

Double-stranded RNA is often a sign of trouble. Our transfer RNA and ribosomes do contain little hairpins that are double-stranded, but most of the free forms of RNA, messenger RNA molecules in particular, are single strands. Many viruses, however, form long stretches of double-stranded RNA as they replicate their genomes. When our cells find double-stranded RNA, it is often a sign of an infection, and they mount a vigorous response that often leads to death of the entire cell. However, plant and animal cells also have a more targeted defense that attacks the viral RNA directly, termed RNA interference.

More ...

Previous Features
Molecular Replacement Structures

21583

Ian Stokes-Rees, http://sbgrid.org
Example from Harrison Lab, Harvard Medical School

Our roadmap:
- Expand the Antibody Library to incorporate new structures
- Setup computations through a portal
- Configure molecular replacement applications with more advanced options (e.g. rigid body refinement).


Ian Stokes-Rees, http://sbgrid.org
CASE 2:

Blind Molecular Replacement

Welcome to SCOP: Structural Classification of Proteins.
1.73 release (November 2007)

34494 PDB Entries. 1 Literature Reference. 97178 Domains. (excluding nucleic acids and theoretical models).
Folds, superfamilies, and families statistics here.
New folds superfamilies families.
List of obsolete entries and their replacements.

Ian Stokes-Rees, http://sbgrid.org
Objectives and Priorities
Grid Computing for Biologists

• Ease of use is number one concern
  • Portal infrastructure
  • Single Sign On
  • Single point of access

• Integration of diverse resources
  • Local VDT/OSG-based clusters
  • Local SGE clusters
  • OSG
  • Member labs’ compute resources

• Secure processes and data

• Data availability
  • WebDAV
  • SCP/SFTP
  • HTTP(S)

• Storage management
  • Meta-data facilities
  • File catalogue

• Advanced Users
  • APIs
  • Scripting
  • Workflows
Thank you! Questions?

Ian Stokes-Rees, Research Associate
SBGrid, Harvard Medical School
ijstokes@crystal.harvard.edu

http://sbgrid.org

Check out our website and email us with any questions.