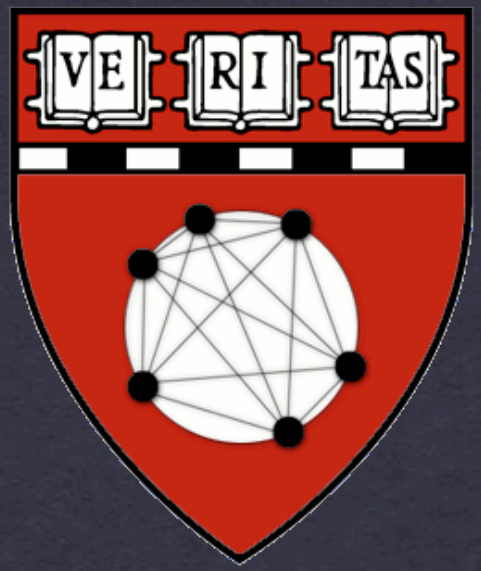
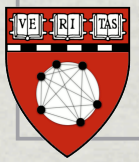


Prospects for Science on OSG: Structural Biology Portal and Applications

SBGrid Research Coordination Network
Harvard Medical School

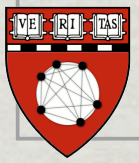
SBGrid

- 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



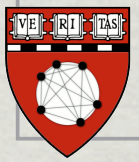
SBGrid PIs and Affiliations

- **87 PIs:** Natalia Beglova, Pamela Bjorkman, Stephen Blacklow, Titus Boggon, Demetrios Braddock, Timothy Fenn, Rick Cerione, Chazin , Bing Chen, Yifan Cheng, James Chou, Jon Clardy, Bil Clemons, John Collier, Brian Crane, Steve Ealick, Mike Eck, Martin Egli, Brandt Eichman, Tom Ellenberger, Qing Fan, Daved Fremont, Barbara Furie, K. Christopher Garcia, Rachelle Gaudet, Tamir Gonen, Niko Grigorieff, Ya Ha, David Harrison, Stephen Harrison, Katya Heldwein, James Hogle, Theodore Jardetzky, Grant Jensen, David Jeruzalmi, Moonsoo Jin, Daniel Kahne, Ailong Ke, Amir Khan, Tom Kirchhausen, Borden Lacy, Andres Leschziner, Gan-Guang Liou, Elias Lolis, Roderick MacKinnon, Mark Mayer, Keith Miller, JJ Miranda, Teru Nakagawa, Niconowicz , Kanagalaghatta Rajashankar, Melanie Ohi, Robert Oswald, Colin Parrish, Anjana Rao, Tom Rapoport, Douglas Rees, Karin Reinisch, William Royer, Montserrat Samso, Sanders , Joseph Schlessinger, Brenda Schulman, Shamoo Yousif, Fred Sigworth, Piotr Sliz, Holger Sondermann, Ben Spiller, Timothy Springer, Henning Stahlberg, Michael Stone, Manal Swairjo, Yizhi Tao, Erik Toth, Jarrod Smith, Greg Verdine, Hector Viadiu, Gerhard Wagner, Loren Walensky, Suzanne Walker, Christopher Walsh, Thomas Walz, Jia-huai Wang, Waterman, John Williams, Stephen Wong, Fenny Zhou
- **34 Institutes:** BIDMC BU BWH CALTECH CMCD COLUMBIA CU DFCI HHMI HMS HU IDI MGH NHRI NIH RCK RFUMS RU SJ STANFORD TCH TC TJU TU UCDAVIS UCSD UCSF UMASS UMSM UW VU WESTERNU WUSTL YSM



Motivation for Grid

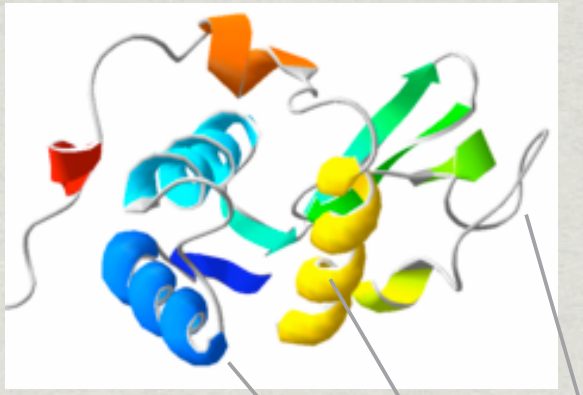
- Because **computational requirements** continue to be a bottle neck
- Because **complexity of tools** impedes quality and efficiency of scientific investigation
- Because some affiliated labs don't have large compute clusters available to them
- Because **new computationally intensive methods** are being developed



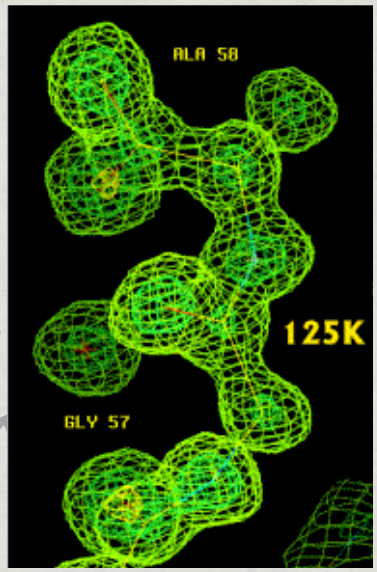
Molecular Replacement

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

Fourier Transform:



Real Space



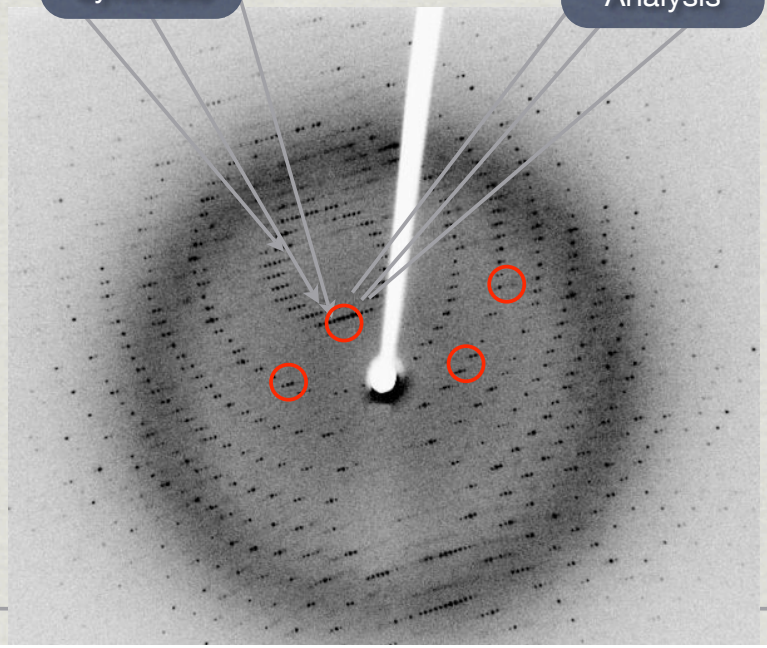
Simple diffraction waves

Fourier Synthesis

TEXT

Fourier Analysis

Reciprocal Space



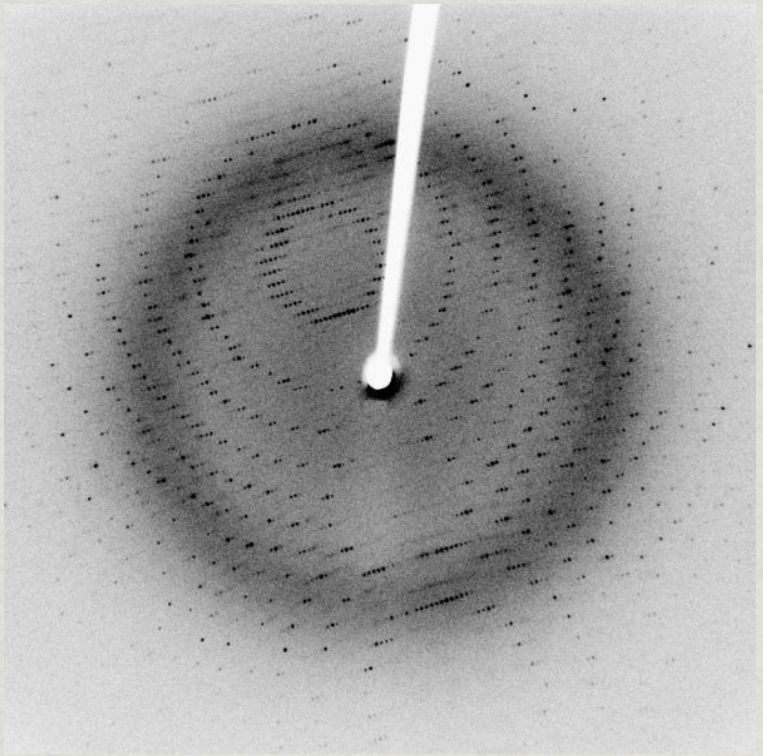
Fourier series for electron density is a sum of contributions from individual reflections.

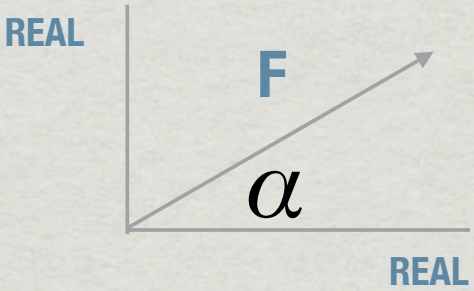
Ian Stokes-Rees, http://sbgrid.org

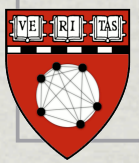
Phase Problem

F_{HKL}

- Amplitudes
  - can be measured
  - ~ sq rt of intensity
- Frequency
  - Fixed and known from X-ray source
- Phase
  - Unknown!



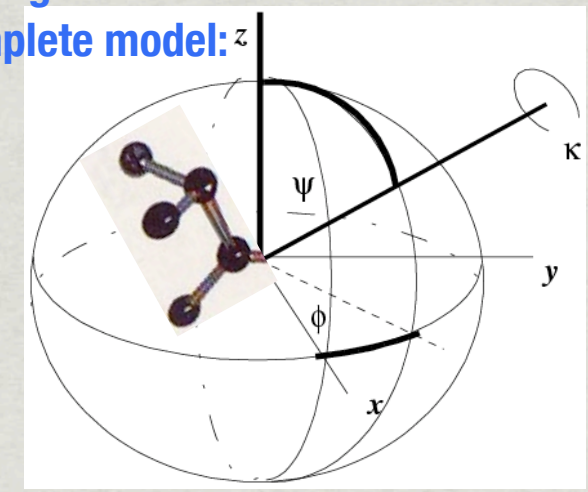




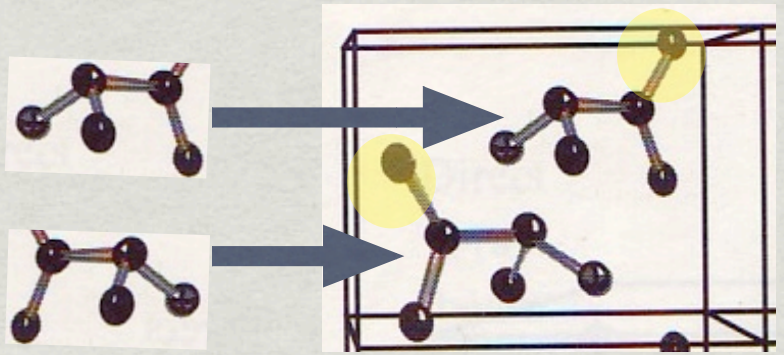
Ian Stokes-Rees, http://sbgrid.org

Molecular Replacement

Homologous or incomplete model:



Rotational Alignment



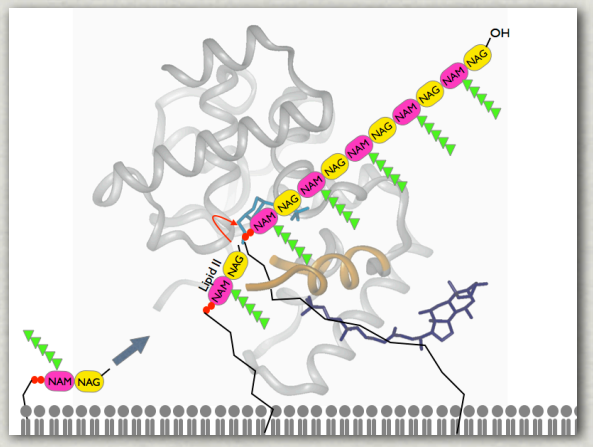
Translational Alignment

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

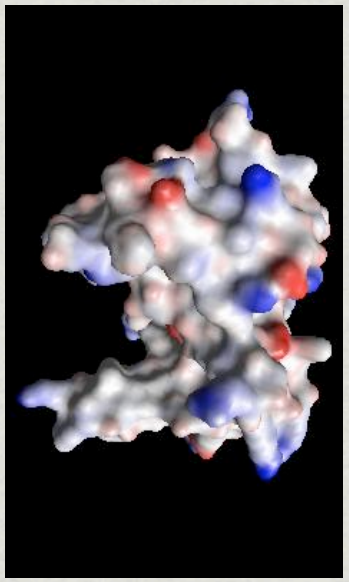
Typically 30% identity and 1/3 of a structure required.

Experimental Phasing

Transglycosylase



Lysozyme



PHAGE

3 months **Too Slow!**

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

RCSB PDB
PROTEIN DATA BANK

A MEMBER OF THE PDB

An Information Portal to Biological Macromolecular Structures

As of Tuesday Feb 19, 2008 there are 49048 Structures | PDB Statistics

CONTACT US | HELP | PRINT PAGE PDB ID or keyword Author Site Search | Advanced Search

Home | Search

- Home
- Getting Started
- **Download Files**
- **Deposit and Validate**
- **Structural Genomics**
- **Dictionaries & File Formats**
- **Software Tools**
- **General Education**
- **Site Tutorials**
- BioSync
- **General Information**
- Acknowledgements
- Frequently Asked Questions
- Report Bugs/Comments

Quick Tips:

Visit mmcif.rcsb.org for detailed information about the macromolecular Crystallographic Information File (mmCIF) data dictionary.

Are you missing data updates? The PDB archive has moved to ftp://ftp.wwpdb.org.
For more information click here.

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.

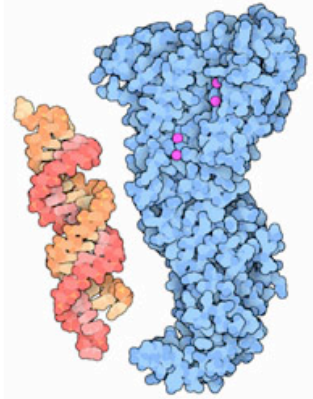
This site offers tools for browsing, searching, and reporting that utilize the data resulting from ongoing efforts to create a more consistent and comprehensive archive.

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**

News

- **Complete News**
- **Newsletter**
- **Discussion Forum**
- **Job Listings**

19-February-2008

Protein Sculptures on Display at Rutgers

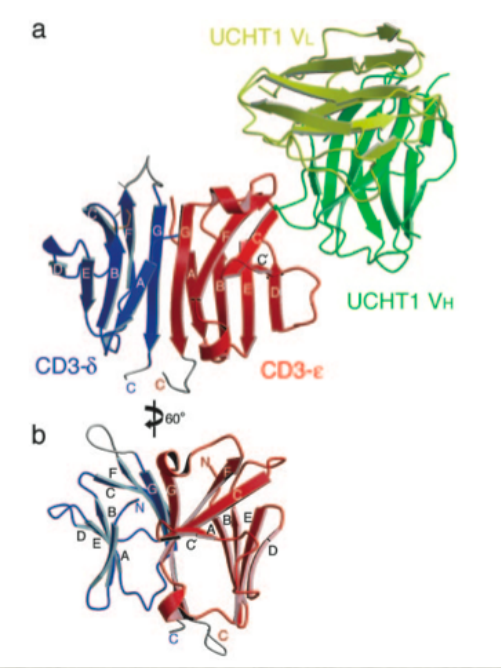
Sculptures and photographs by Julian Voss-Andreae are currently on display at Rutgers Student Center (New Brunswick, NJ) until February 22, 2008.

- Full article ...

Winter 2008 RCSB PDB Newsletter Redesigned and Published

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

Example from Harrison Lab, Harvard Medical School



complexes per asymmetric unit. We determined the structure by molecular replacement, implemented with MOLREP (32), using 244 antibody fragments as search models. We used omit maps to eliminate incorrect solutions and to verify the correct one (Protein Data Bank entry 6FAB). We built an initial model with

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).

Arnett et al. Crystal structure of a human CD3-epsilon/delta dimer in complex with a UCHT1 single-chain antibody fragment. Proc Natl Acad Sci USA (2004) vol. 101 (46) pp. 16268-73

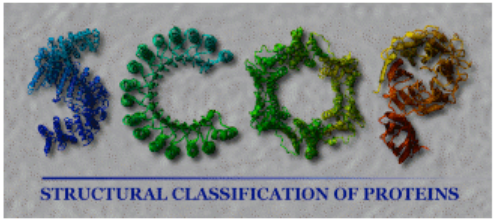
CASE 2:

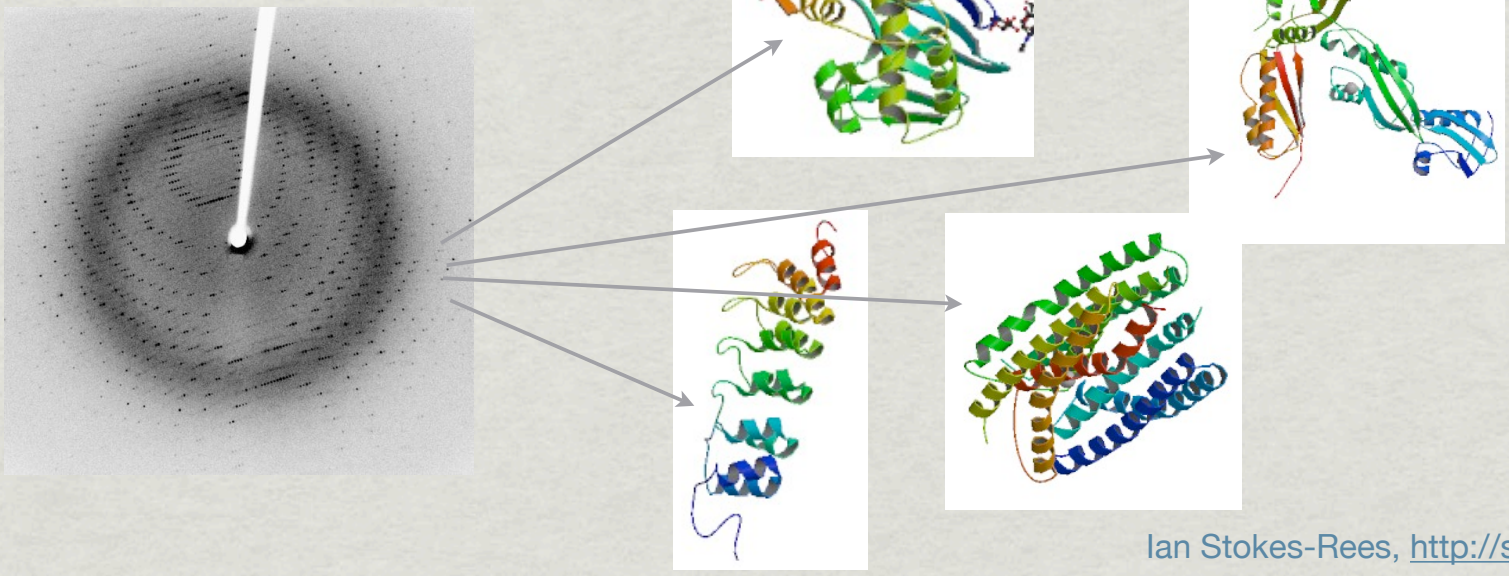
Blind Molecular Replacement

Structural Classification of Proteins

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.





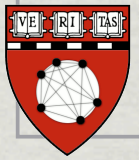
Bringing Scientists to the Grid

Or Bringing the Grid to Scientists

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

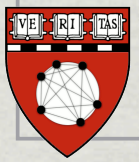
Portal Interface

- Lower the bar for adoption
  - One-stop-shop portal for community
  - NOTE: Community = VO
  - Single Sign On, and security/grid token management
- Provide multi-modal access to data
  - Web/HTTP(S), SCP/SFTP, WebDAV, FUSE/sshfs, GridFTP, slashgrid
- Streamline access and workflow to key applications
  - Make applications accessible
  - Incorporate best practice configuration and usage
- Benefit from TeraGrid (and others) tool set and portlets
  - Largely converged on Tomcat/GridSphere/OGCE



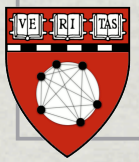
Perspectives

- There are hundreds of bio applications which can be run on the grid
- We look forward to incorporating workflow
- We need to address data management/curation issues
- We need better developer documentation
- Ease of deployment and “site” management is critical to federate smaller resources
  - **A “site” may be a desktop computer, or a lab’s “desktop cluster”**
- Dynamic VO management tools (e.g. ACL, portals, data) will be important as collaborations mature in their experience with the grid



Summary

- Structural Biology is:
  - A large and active field at the cutting edge of science
  - Full of computationally intensive problems
  - Ripe for taking advantage of a grid infrastructure
- SBGrid provides:
  - An existing community of almost 100 labs
  - Dedicated people and infrastructure to support members' computing demands
- Portal interfaces:
  - Single point of access
  - Ease of use
  - Pre-built application workflows
- Initial Applications
  - Rosetta protein prediction/design
  - Molecular Replacement
  - ... and then by user demand
- Future
  - Data curation
  - Custom workflows
  - Federate member's CPUs



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.