

# Structural Biology on the Grid

## SBGrid Research Coordination Network

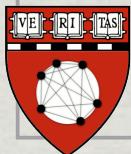
### Harvard Medical School

# **Summary**

**SBGrid Background  
Portal Infrastructure  
Molecular Replacement  
Objectives and Priorities**

# SBGrid

- 📍 Consortium of structural biology labs
  - Involved in structure determination of (primarily) proteins
  - X-Ray crystallography, NRM, EM
- 📍 87 member labs across the US
  - 28 at Harvard
- 📍 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>



## Welcome to SBGrid

[News](#) | [Services](#) | [History & Staff](#)



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](#)

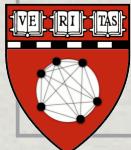
### Latest News

- [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](#)



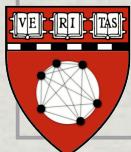
# SBGrid Services

- Application optimization
- Application packaging and automated distribution/  
update
- Help desk
- Centralized license management
- Soon:
  - Grid portal for SB applications
  - Gateway to OSG



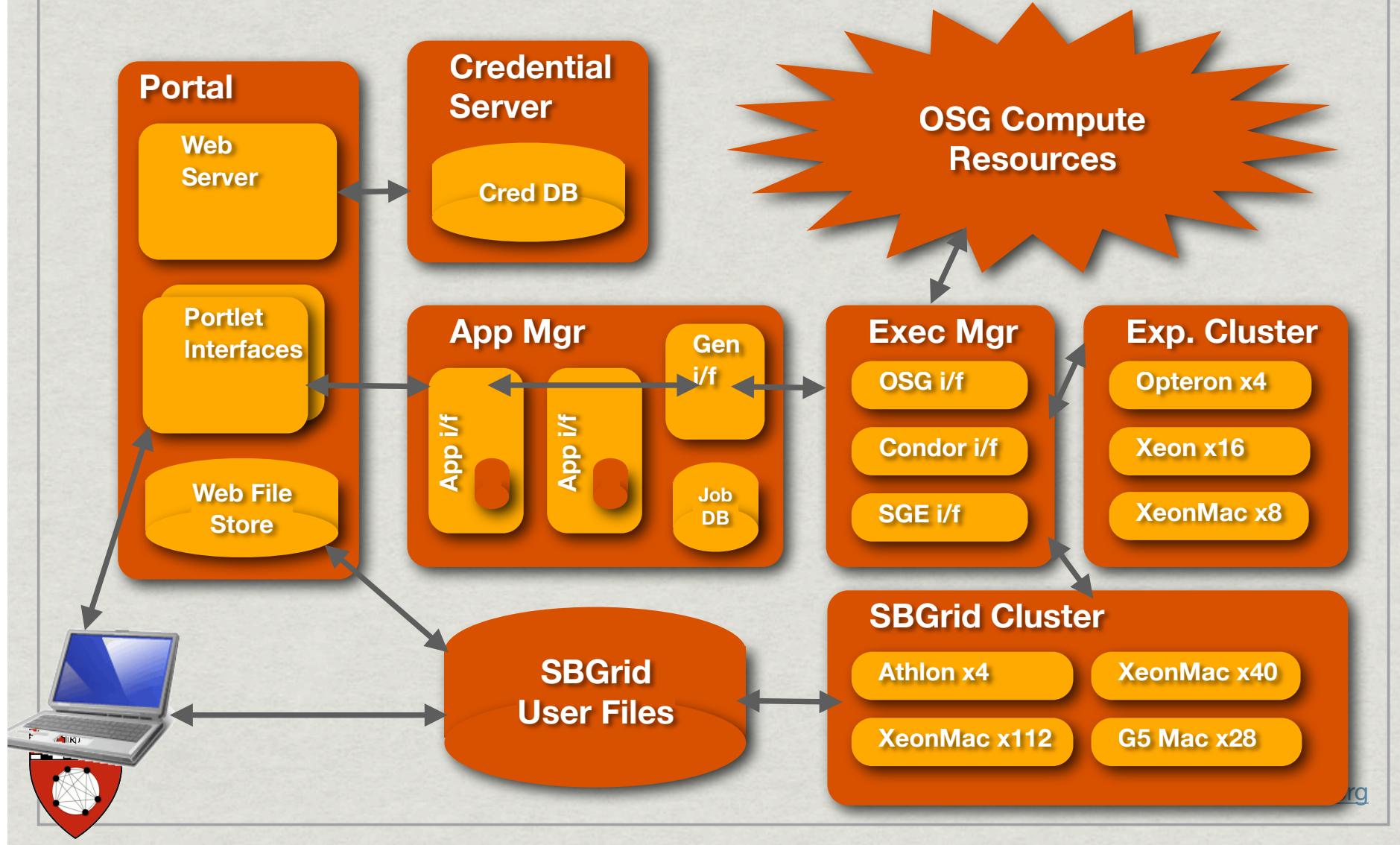
# Motivation for Grid

- ↳ 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

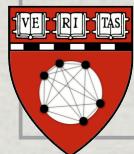
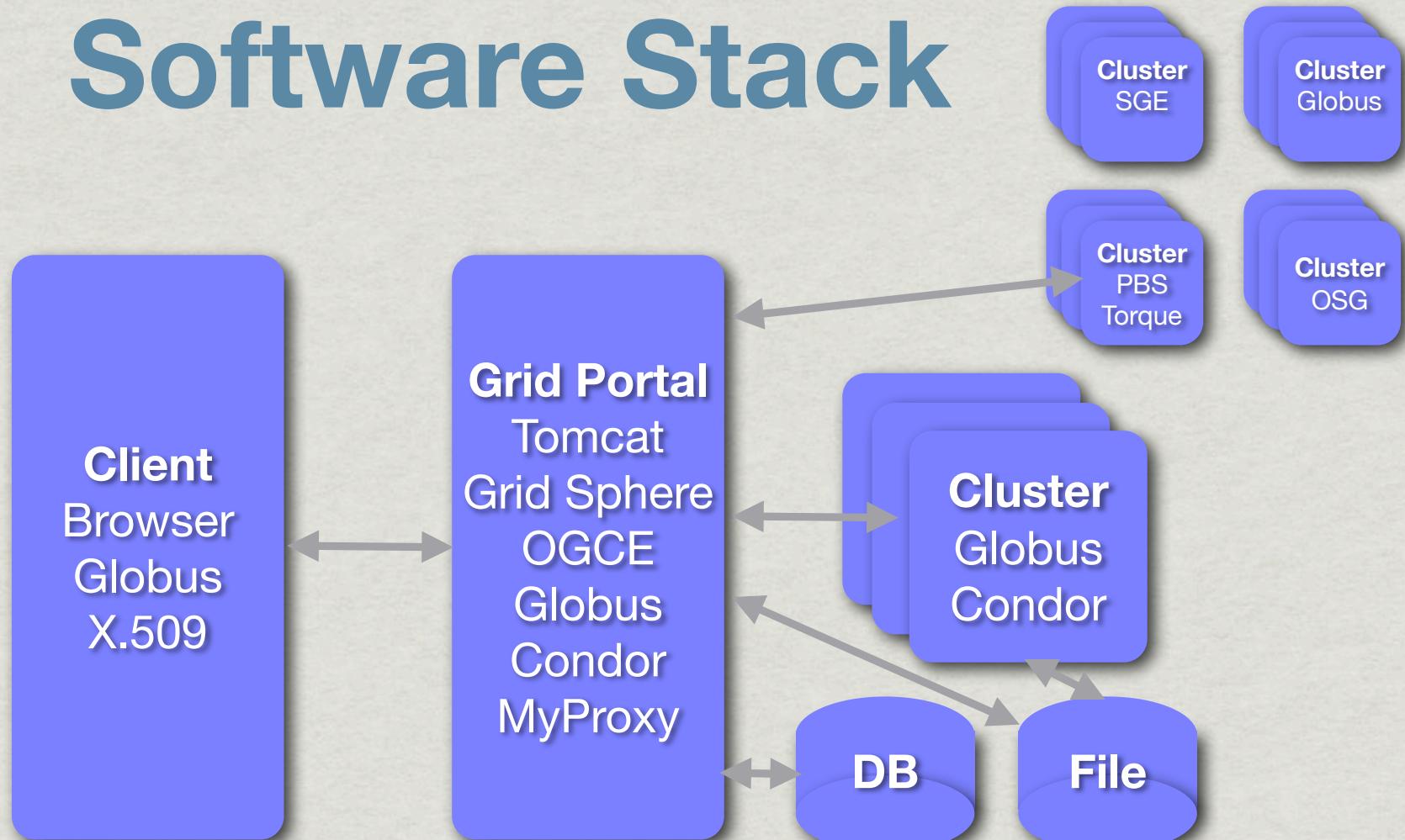


# Portal Infrastructure

# SBGrid Architecture

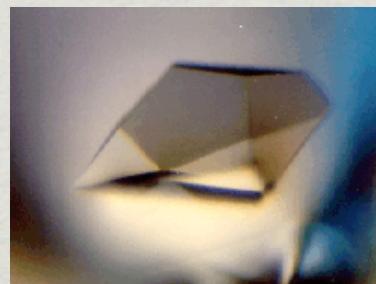


# Software Stack

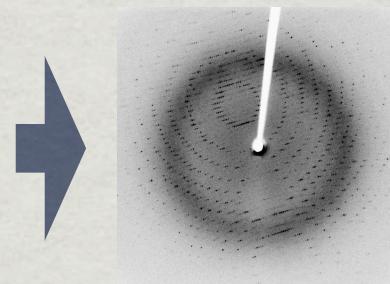


# Molecular Replacement

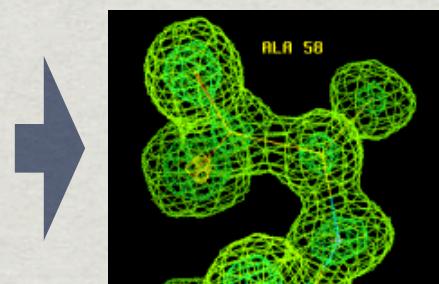
# Macromolecular Crystallography



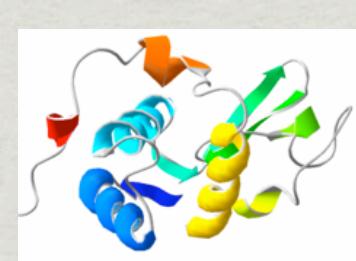
CRYSTAL



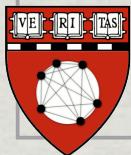
RECIPROCAL LATTICE



ELECTRON DENSITY MAP



MODEL

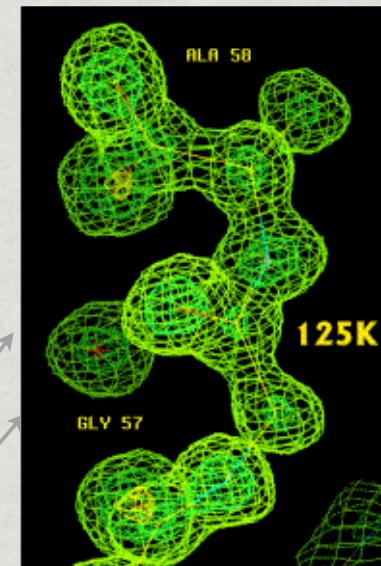


Ian Stokes-Rees, <http://sbgrid.org>

## Fourier Transform:



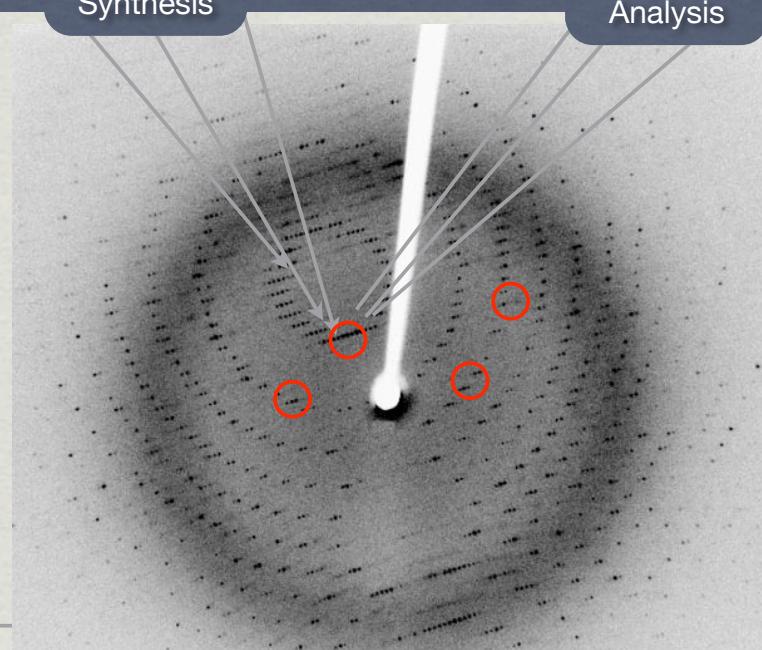
Real  
Space



Fourier  
Synthesis

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

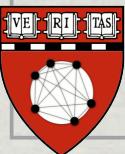
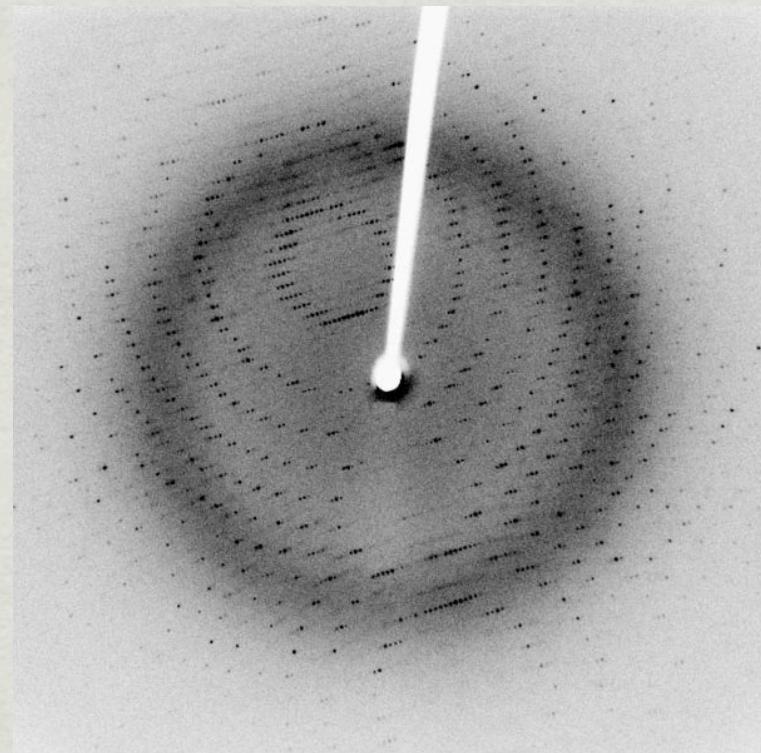
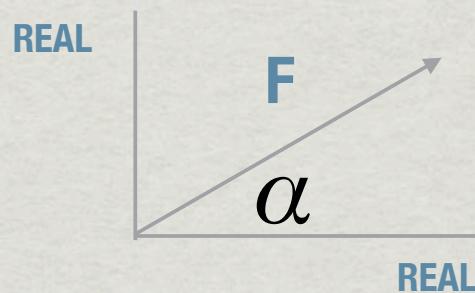
- can be measured
- $\sim \text{sq rt of intensity}$

## Frequency

- Fixed and known from X-ray source

## Phase

- Unknown!

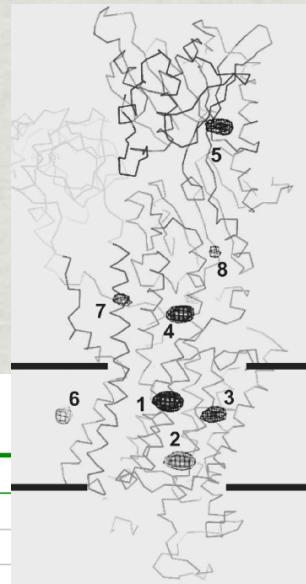


# Heavy Atom Derivatives

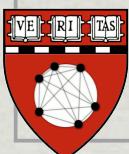
Successful crystal soaks

Table S1 | Data collection and refinement statistics

	Native	Xenon	Ta <sub>3</sub> Br <sub>14</sub>	SelenoMet	2'-I-ADP	EMP
<b>Data Collection</b>						
Space group	C2	C2	C2	C2	C2	C2
Cell dimensions						
a, b, c (Å)	161.279 103.955 181.013	161.391 103.839 181.948	160.895 104.163 181.937	161.399 103.886 181.644	160.621 105.045 181.245	161.510 104.074 181.546
α, β, γ (°)	90.000 97.987 90.000	90.000 97.471 90.000	90.000 97.849 90.000	90.000 97.501 90.000	90.000 98.021 90.000	90.000 97.672 90.000
Resolution (Å)	30-3.0	30-3.3	30-3.8	30-3.3	30-3.1	30-3.3
R <sub>sym</sub> or R <sub>merge</sub> *	8.1 (56.9)	10.9 (45.7)	10.8 (27.8)	8.1 (35.1)	7.9 (45.9)	9.9 (53.3)
I/σI	20.9 (2.18)	14.5 (2.00)	13.3 (3.96)	16.7 (2.83)	19.5 (1.99)	15.0 (2.40)
Completeness (%)	99.1 (87.9)	98.0 (85.9)	95.2 (73.9)	97.6 (79.2)	98.0 (79.9)	100 (100)
Redundancy	7.4 (5.3)	7.1 (5.1)	6.6 (5.5)	6.0 (4.4)	6.2 (5.1)	6.4 (5.9)



Many heavy atom derivatives needed: time consuming

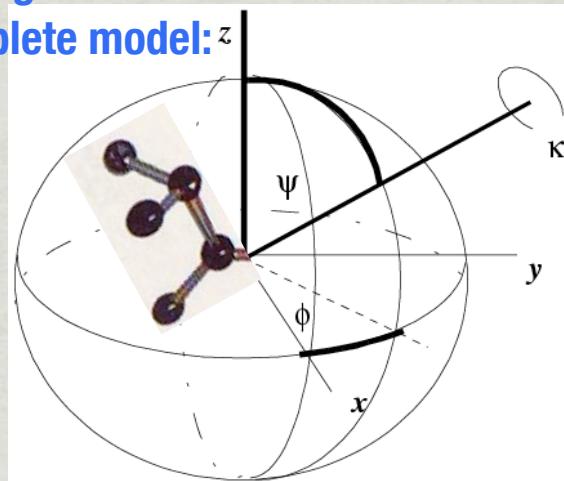


Dawson et al. Structure of a bacterial multidrug ABC transporter. Nature (2006) vol. 443 (7108) pp. 180-5

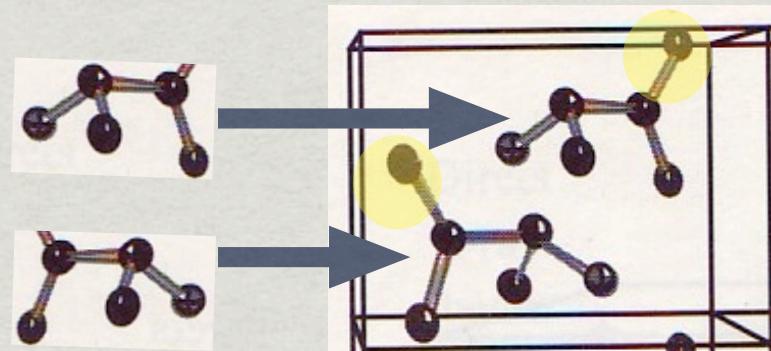
Ian Stokes-Rees, <http://sbgrid.org>

# Molecular Replacement

Homologous or incomplete model:



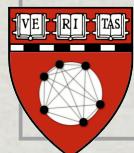
Rotational Alignment



Translational Alignment

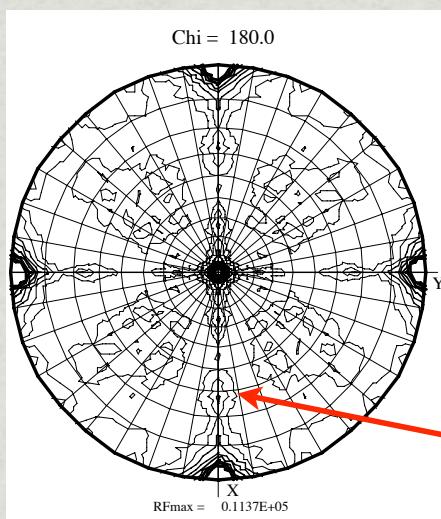
- Combining model phases with experimental intensities will reveal details of missing elements

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



# Cross-Rotation Function

		NUMBER OF RF PEAKS : 10							
		THETA	PHI	CHI	ALPHA	BETA	GAMMA	RF	RF/SIGMA
SOL_RF	1	146.34	-139.17	155.19	55.63	65.55	153.97	0.1453E+07	3.69
SOL_RF	2	132.29	165.77	54.08	56.82	39.30	265.27	0.1409E+07	3.57
SOL_RF	3	126.82	159.77	57.42	51.60	45.23	272.06	0.1395E+07	3.54
SOL_RF	4	147.99	-141.48	160.47	49.99	62.99	152.96	0.1387E+07	3.52
SOL_RF	5	154.58	-138.22	162.21	51.61	50.18	148.05	0.1285E+07	3.26
SOL_RF	6	127.83	148.20	69.52	35.15	53.52	278.74	0.1265E+07	3.21



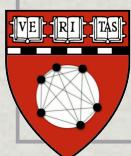
model with 30% identity, maybe?

1-24 hours per model

	NUMBER OF PEAKS : 50							
	ALPHA	BETA	GAMMA	THETA	PHI	CHI	RF	RF/SIGMA
1	49.88	82.51	104.56	41.97	62.66	160.85	7748.	7.07
2	102.03	27.25	29.60	14.88	126.21	133.08	4409.	4.02
3	141.78	36.81	105.22	158.24	-71.72	116.84	4346.	3.97
4	50.12	44.24	104.89	22.60	62.61	156.88	3881.	3.54
5	148.79	85.37	27.42	42.70	150.68	177.22	3812.	3.48

YES!

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



# Translation Function

	NUMBER OF PEAKS : 50										
	IX	IY	IZ	XFRAC	YFRAC	ZFRAC	XORT	YORT	ZORT	DENS	DENS/SIGMA
1	32	55	79	0.333	0.457	0.410	16.45	34.12	50.41	0.2875	11.48
2	32	45	79	0.335	0.379	0.410	16.54	28.31	50.41	0.1590	6.35
3	32	50	79	0.334	0.417	0.411	16.51	31.18	50.55	0.1518	6.06
4	32	55	62	0.331	0.458	0.322	16.35	34.23	39.67	0.1356	5.41
5	21	55	79	0.218	0.455	0.410	10.78	34.03	50.45	0.1347	5.38

Obvious solution with a 100% identical model

	NUMBER OF PEAKS : 50										
	IX	IY	IZ	XFRAC	YFRAC	ZFRAC	XORT	YORT	ZORT	DENS	DENS/SIGMA
1	25	27	25	0.262	0.224	0.131	12.98	16.72	16.15	0.1388	5.02
2	11	43	42	0.114	0.361	0.219	5.64	26.98	26.90	0.1180	4.27
3	11	37	83	0.118	0.307	0.434	5.86	22.95	53.38	0.1151	4.16
4	21	49	56	0.218	0.409	0.294	10.80	30.53	36.14	0.1136	4.10
5	23	27	32	0.238	0.227	0.168	11.78	16.93	20.61	0.1125	4.07

Possible solution - needs further testing



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](http://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.

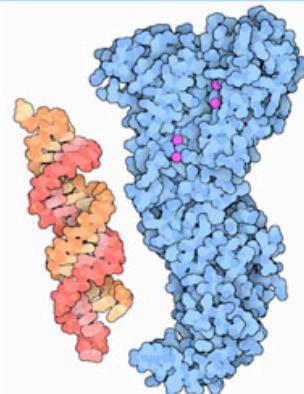
This site offers tools for browsing, searching, and reporting that utilize the data resulting from ongoing efforts to create a more consistent and comprehensive archive.

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](mailto: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>



# Molecular Replacement Structures

PDB ID or keyword   Author MOLECULAR REPLACEMENT Site Search | Advanced Search

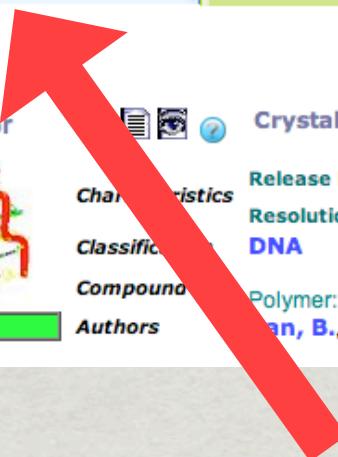
Are you missing data updates? The PDB archive has moved to <ftp://ftp.wwpdb.org>.  
For more information click [here](#).

Help 21583 Structure Hits 9953 Citations 4602 Ligand Hits 20 Web Page Hits

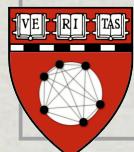
Advanced Keyword Query for: molecular replacement  
1 2 3 4 5 .. 2159 ↗

1m6r      Crystal structure of rGd(CGCGC) forming hexamer Z-DNA duplex with 5'-(rG) overhang

**Characteristics**: Release Date: 25-Feb-2003 Exp. Method: X Ray Diffraction  
Classification: Resolution: 1.54 Å  
Compound: DNA  
Authors: Polymer: 1 Molecule: 5'-R(\*G)D(\*CP\*GP\*CP\*GP\*CP\*G)-3' Chains: A,B  
    Ian, B., Sundaralingam, M.



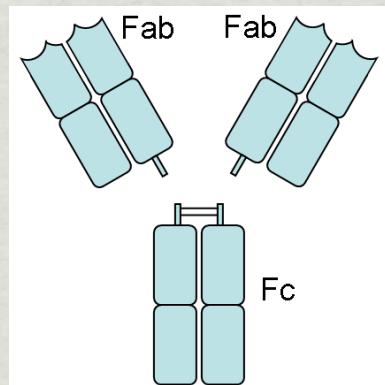
21583



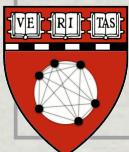
Ian Stokes-Rees, <http://sbgrid.org>

# MolRep: Antibodies (I)

- Antibodies can help with crystallization of soluble proteins, RNA and membrane proteins (by making rigid crystal packing contacts)
- Once crystallized antibodies can be placed in the unit cell by molecular replacement and will provide phasing information about the studied molecule (RNA, membrane protein, etc)
- 855 models of antibody are deposited in the Protein Data Bank. Due to hinge variability it is difficult to predict which antibody molecule will provide a clear molecular replacement solution.
- Solution: try all of them

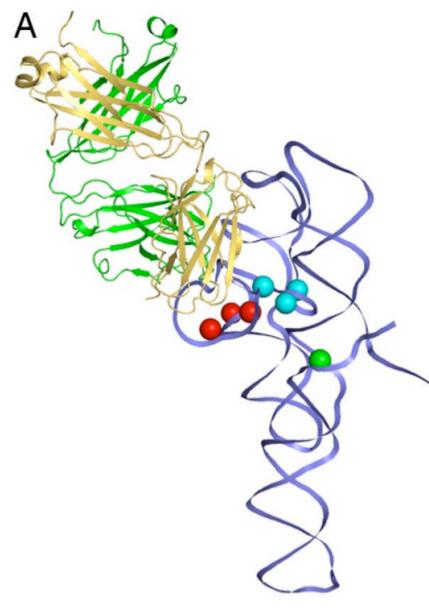


Ian Stokes-Rees, <http://sbgrid.org>



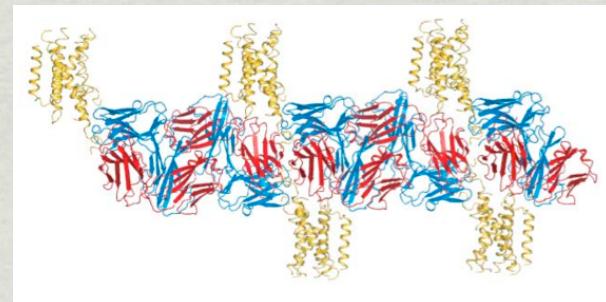
# MolRep: Antibodies (II)

## RNA molecules



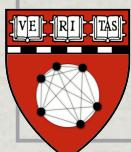
Ye et al. Synthetic antibodies for specific recognition and crystallization of structured RNA. Proc Natl Acad Sci USA (2008) vol. 105 (1) pp. 82-7

## Membrane Proteins



crystals, diffraction images were obtained by microcrystallography. The structure of the  $\beta_2$ AR365–Fab5 complex was solved by molecular replacement, using separate constant and variable Fab domain structures as search models. Coordinates and structure factors are deposited in the Protein Data Bank (accession codes 2R4R for  $\beta_2$ AR365–Fab5 and 2R4S for  $\beta_2$ AR24/365–Fab5).

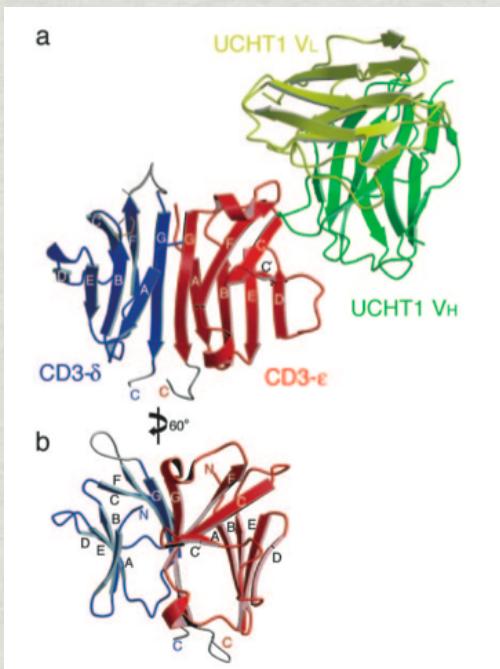
Rasmussen et al. Crystal structure of the human beta2 adrenergic G-protein-coupled receptor. Nature (2007) vol. 450 (7168) pp. 383-7



855 models - which one to use?

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



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

## CASE 2:

# Blind Molecular Replacement

*Structural Classification of Proteins*



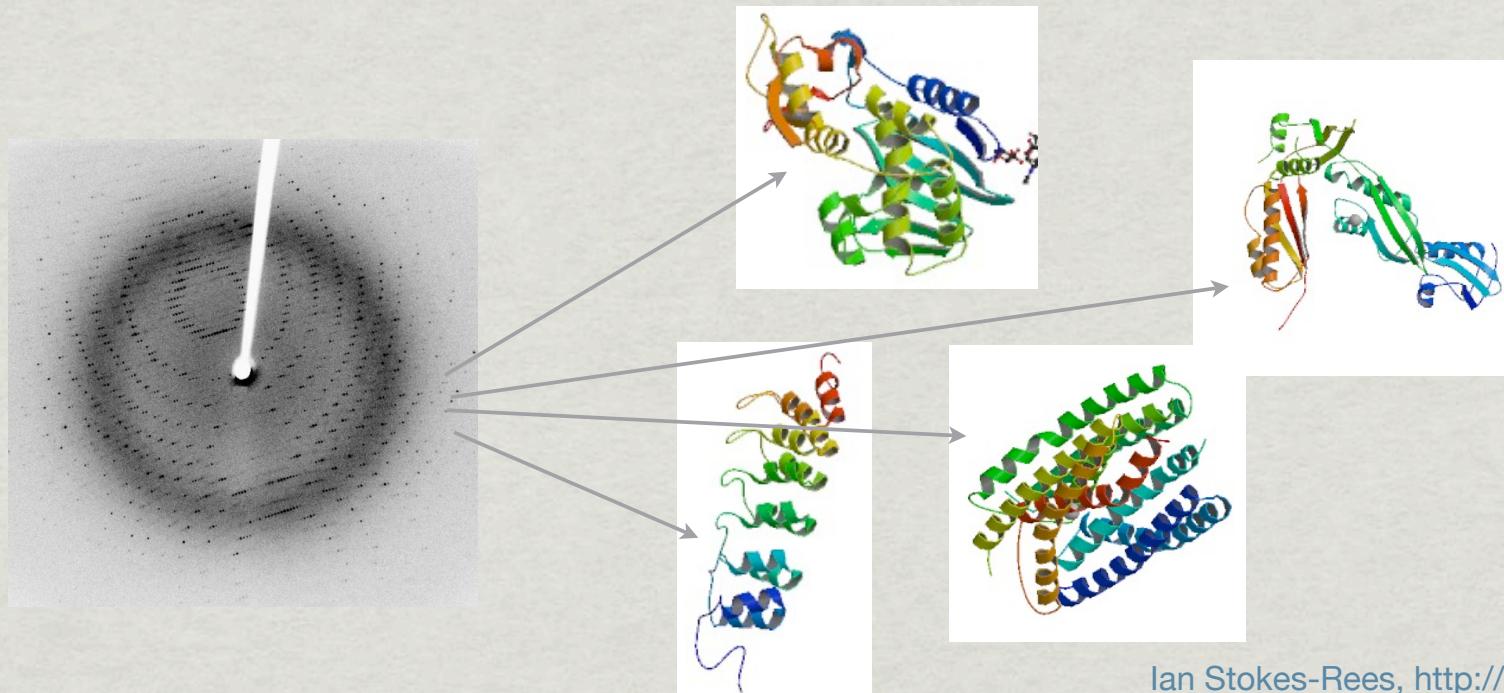
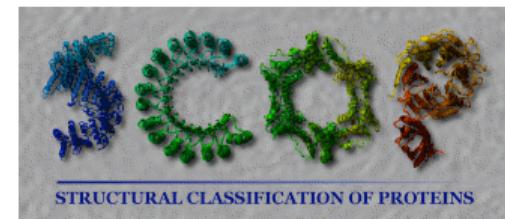
Welcome to SCOP: Structural Classification of Proteins.

1.73 release (November 2007)

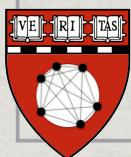
34494 PDB Entries. 1 Literature Reference. 97178 Domains. (excluding nucleic acids and theoretical models).  
Folds, superfamilies, and families [statistics here](#).

New folds [superfamilies](#) [families](#).

List of obsolete entries and their replacements.

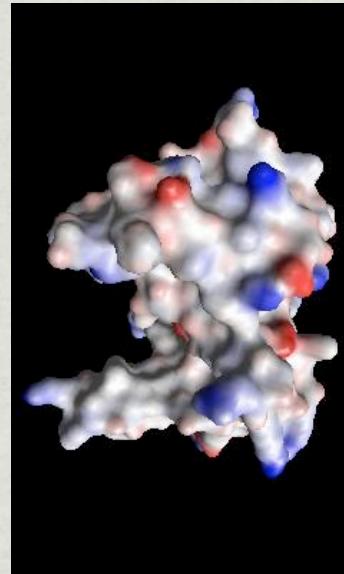
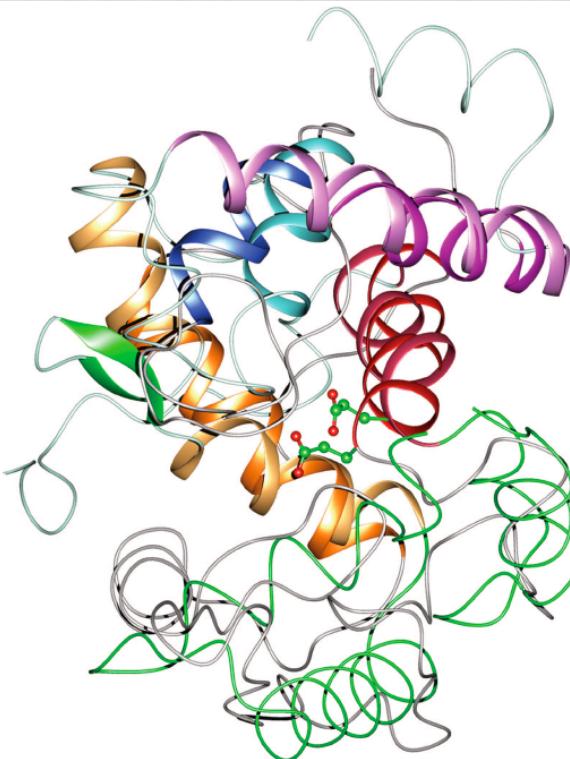
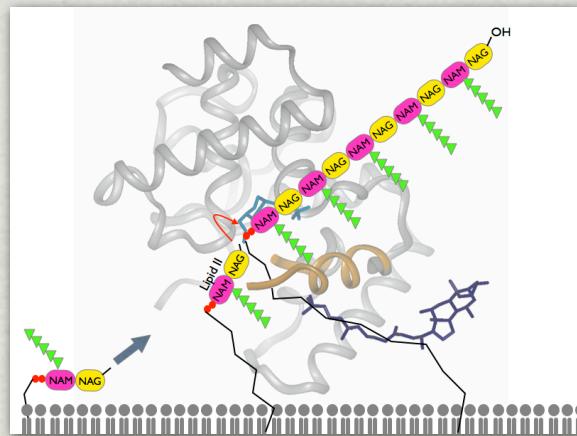


Ian Stokes-Rees, <http://sbgrid.org>



# Alternative: Experimental Phasing

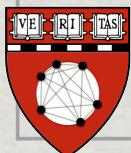
Transglycosylase



PHAGE

3 months      **Too Slow!**

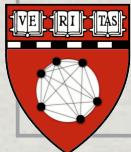
Ian Stokes-Rees, <http://sbgrid.org>



# 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](mailto:ijstokes@crystal.harvard.edu)**

**<http://sbgrid.org>**

**Check out our website and email us  
with any questions.**