Discovery & Innovation via Cyberinfrastructure

Russ Miller

Cyberinfrastructure Lab, SUNY-Buffalo

Hauptman-Woodward Med Res Inst

NSF, NIH, DOE, NIMA, NYS, HP

www.cse.buffalo.edu/faculty/miller/CI/

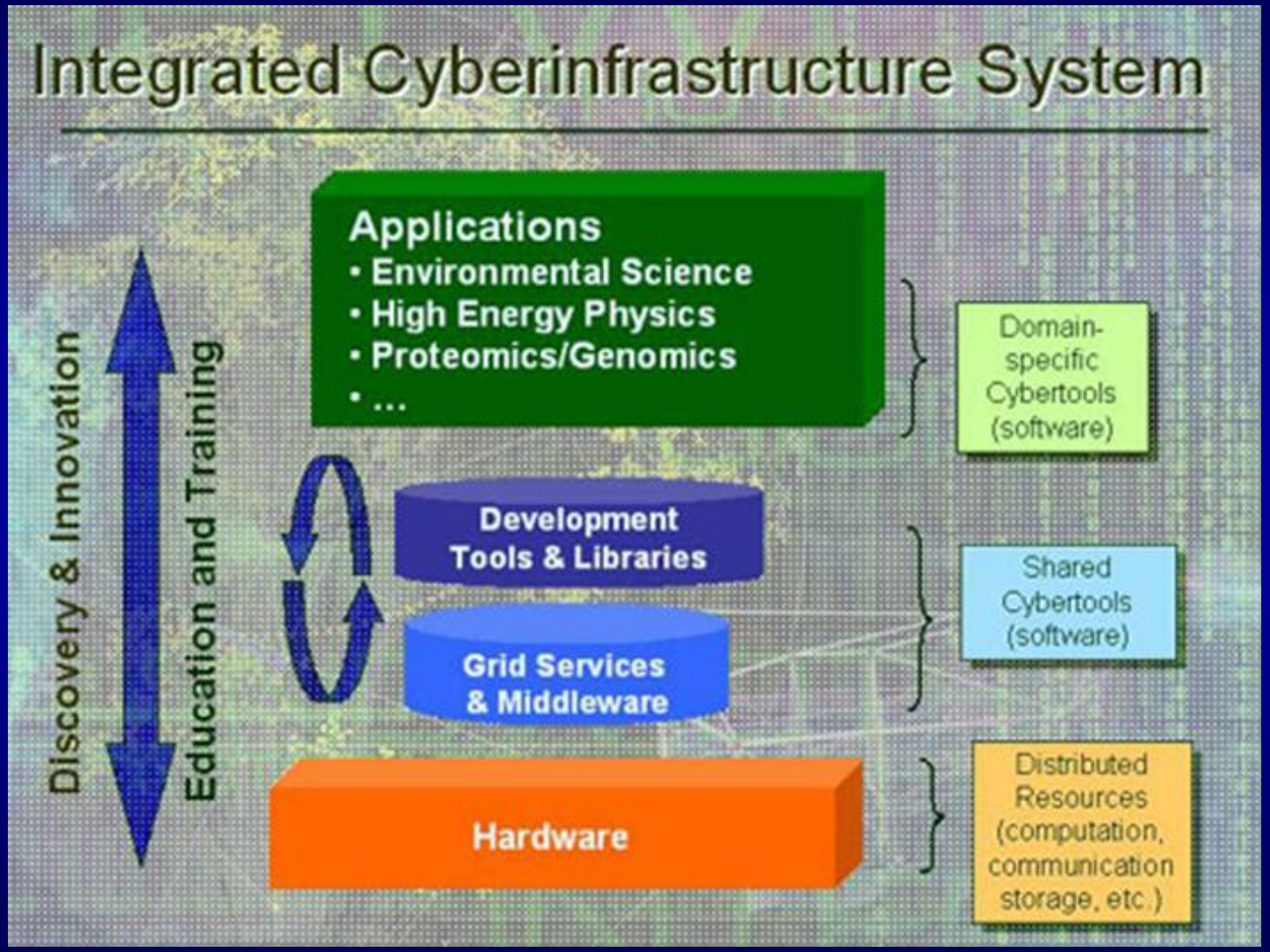
Academia in the 21st Century: High-Level View

- **Empower students to compete in knowledge-based economy**
- **Embrace digital data-driven society**
- **Accelerate discovery and comprehension**
- **Enhance virtual organizations**
- **Provide increased education, outreach, and training**
- **Enhance and expand relationships between academia and the corporate world**

Academia in the 21st Century: Medium-Level View

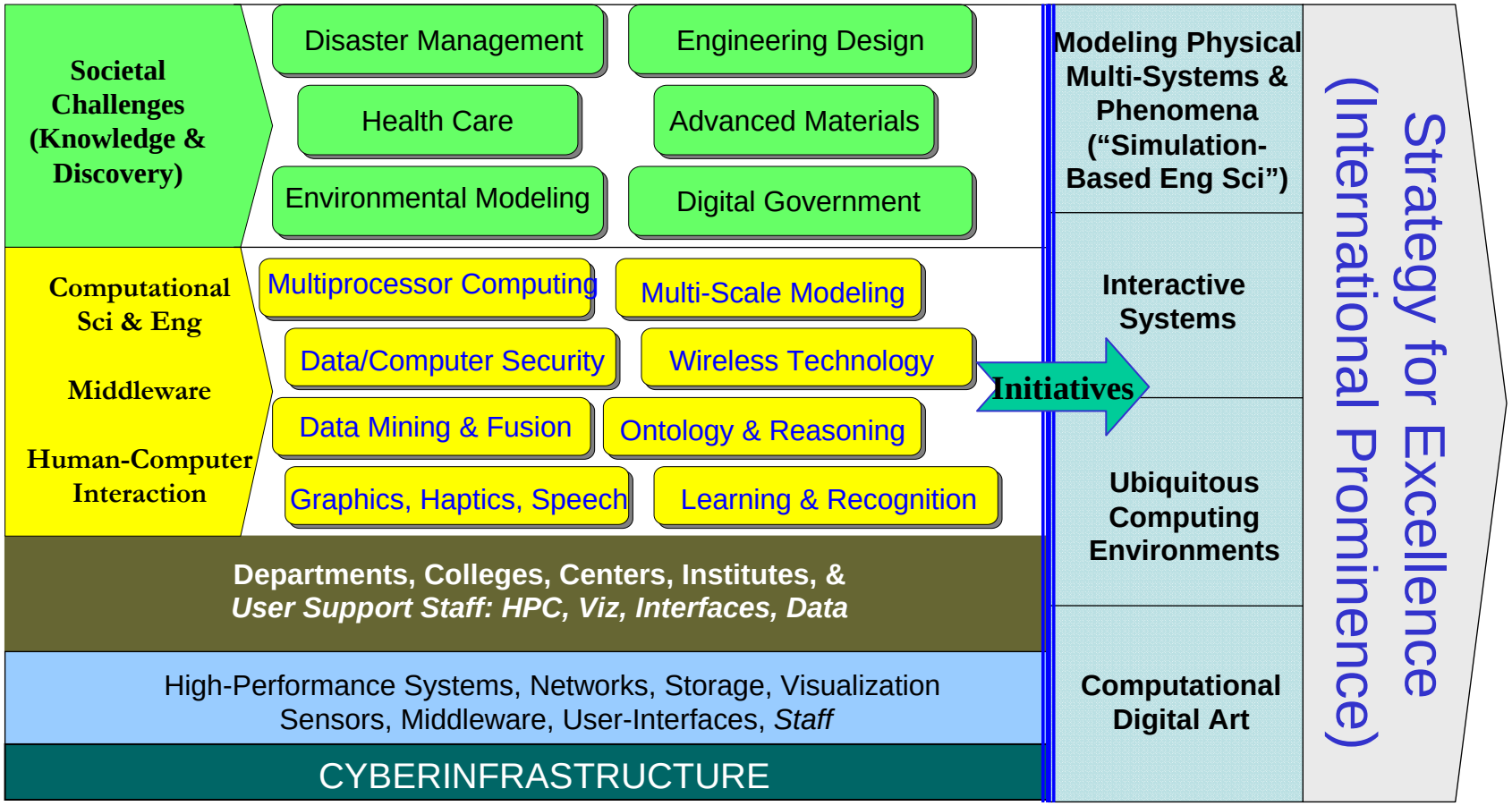
- **Create links between enabling technologists and disciplinary users**
- **Improve efficiency of knowledge-driven applications in myriad disciplines**
  - **New Techniques**
  - **New Algorithms**
  - **New Interactions (people & systems)**
- **Support HPC infrastructure, research, and applications**
- **Deliver high-end cyberinfrastructure to enable efficient**
  - **Collection of data**
  - **Management/Organization of data**
  - **Distribution of data**
  - **Analysis of data**
  - **Visualization of data**

NSF Integrated Cyberinfrastructure



NSF Director Arden L. Bement: "leadership in cyberinfrastructure may determine America's continued ability to innovate – and thus our ability to compete successfully in the global arena."

Academic Computing Initiative: Inverted Umbrella (Sample)

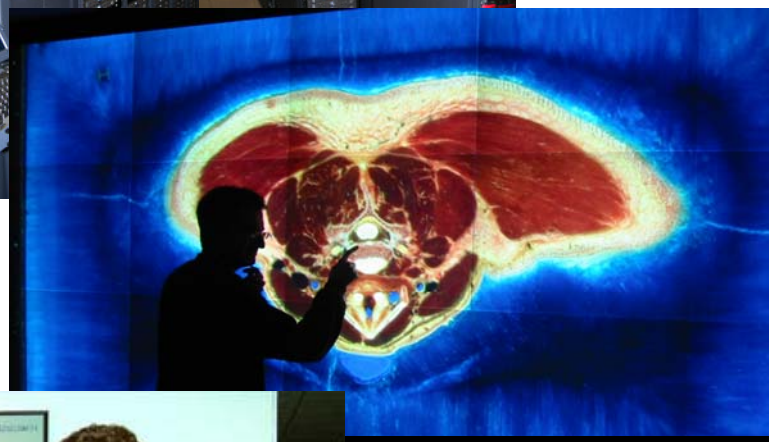


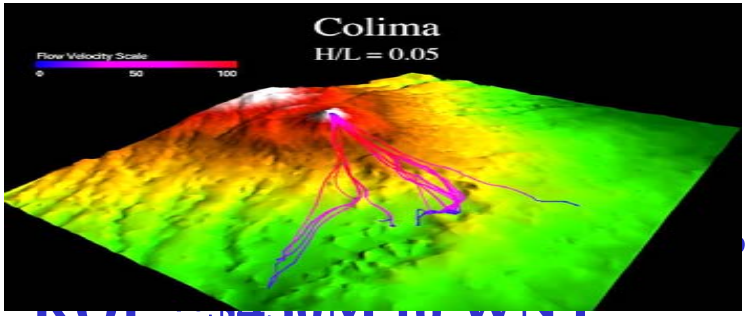
HPC/CI/CSE Initiative

- Must be Pervasive Across the Entire University
- Must Remove Barriers
- Groups Must Interact
  - Research Groups
  - Support Staff
  - Students
  - Departments
  - Colleges
- Issues
  - Tenure & Promotion
  - University vs Colleges vs Departments vs Faculty vs Centers/Institutes vs Degrees vs Courses
- Details are University Dependent

Center for Computational Research (CCR)

- **Founding Director (1998-2006)**
- **Facts & Figures**
  - **Top Academic HPC Center in World**
  - **~25 TF of HPC**
  - **~600 TB of High-End Storage**
  - **Significant Visualization**
  - **Special-Purpose Systems**
  - **~30 FTEs Staff**
  - **140 Projects Annually**
- **EOT**
- **ROI: ~$450M to WNY**



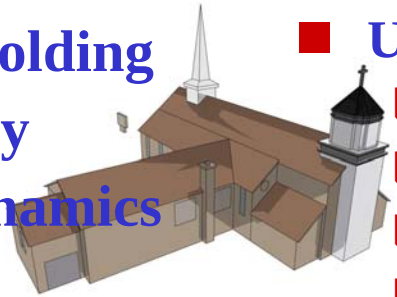


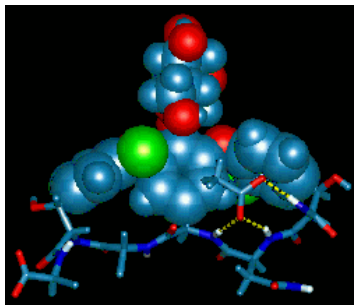
CCR Highlights (1998-2006)

- Provide HE-Comp
- Provide HE-Vis + AGN
- Special Purpose Systems
  - Bioinformatics
  - Data Warehouse / Mining
- Support Local/National Efforts – Industry + Acad
- Create jobs in WNY
- Certificate Program
- Workshops + Tours
  - Campus, Industry
  - High-School
- Urban Planning & Design
- MTV Videos
- Peace Bridge, Med Campus
- Olmsted Parks, Thruway
- NYS Agencies
- Elected Officials
- Magnet on Campus
- Significant Funds
- Numerous Awards
- Significant Publicity

CCR Research & Projects

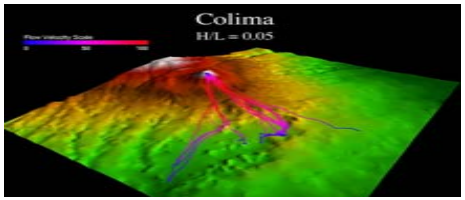
- Archaeology
- Bioinformatics/Protein Folding
- Computational Chemistry
- Computational Fluid Dynamics
- Data Mining/Database
- Earthquake Engineering
- Environ Modeling & Simulation
- Grid Computing
- Molecular Structure Determination
- Physics

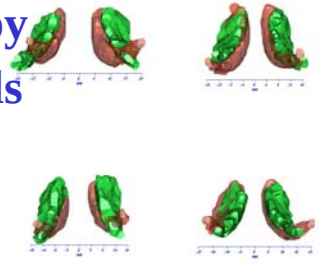




- Videos: MTV
- Urban Simulation and Viz
  - StreetScenes
  - I-90 Toll Barrier
  - Medical Campus
  - Peace Bridge

- Accident Reconstruction
- Scientific Viz
  - Dental
  - Surgery
  - MRI/CT Scan
  - Confocal Microscopy
  - Crystallization Wells
  - Collaboratories





Real-Time Visualization

StreetScenes: Real-Time 3D Traffic Simulation

- Accurate local landmarks: Bridges, Street Signs, Business, Homes
- Can be viewed from driver's perspective
- Real-Time Navigation
- Works with
  - Corsim
  - Synchro
- Generate AVI & MOV
- Multiple Simultaneous
  - Traffic Loads
  - Simulation
  - Varying POV

Animation & Simulation

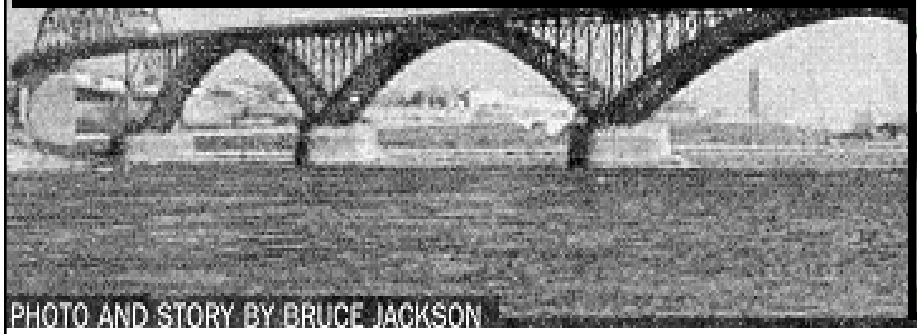
Rendered Scenes

Visualization in Planning Studies

Bergmann Associates: Indian Street Bridge - Florida

Peace Bridge Visualization: Animation & Simulation

- **Proposed Options**
  - **Relocate US plaza**
  - **Build a 3-lane companion span**



Public Forum

MTV

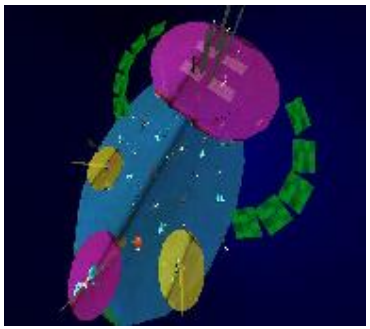
IBC Digital & CCR

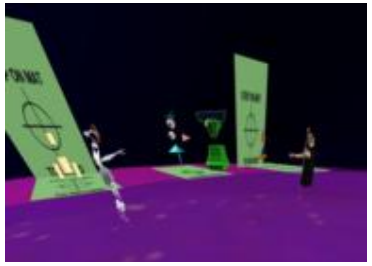
Song: I'm OK (I Promise)
Band: Chemical Romance
Gaming Environment: Death Jr.

Virtual Reality

Alive on the Grid: PAAPAB

- Networked art application for CAVE
  - Users from around the world
  - First performance 2001
- Dance-floor environment
  - Inhabited by life-size puppets
  - Dance with each other
  - Synchro
- Recording Booth
  - User enters booth
  - User dances
  - System records dance from tracking on head and hands
  - Dance mapped to Avatar

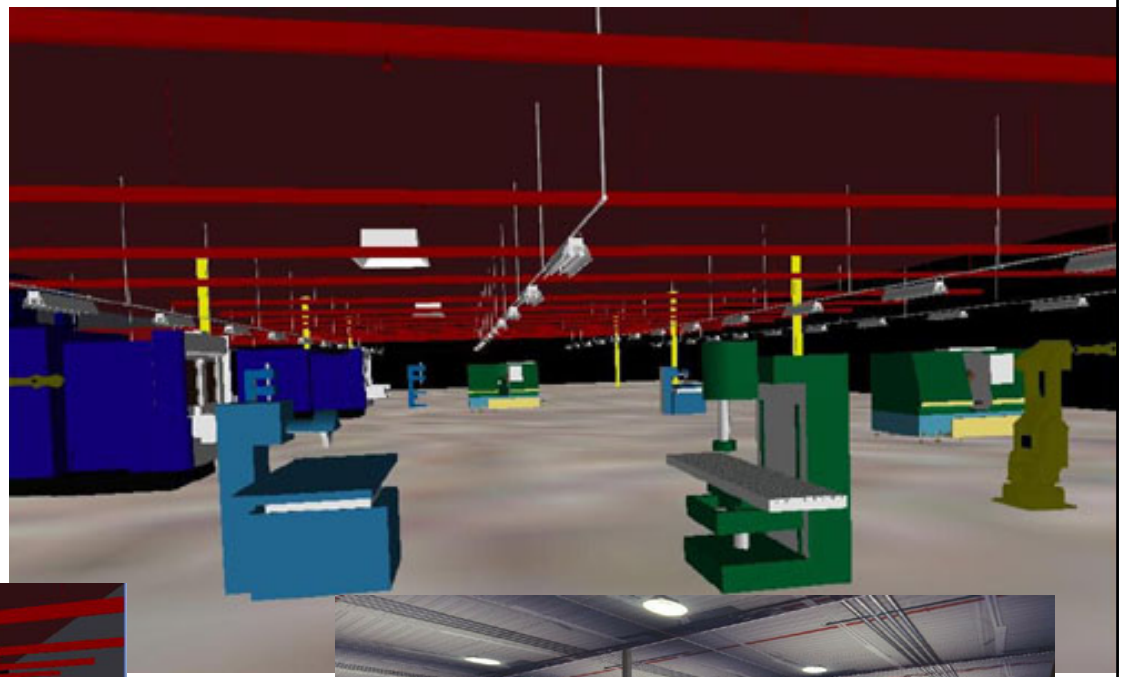


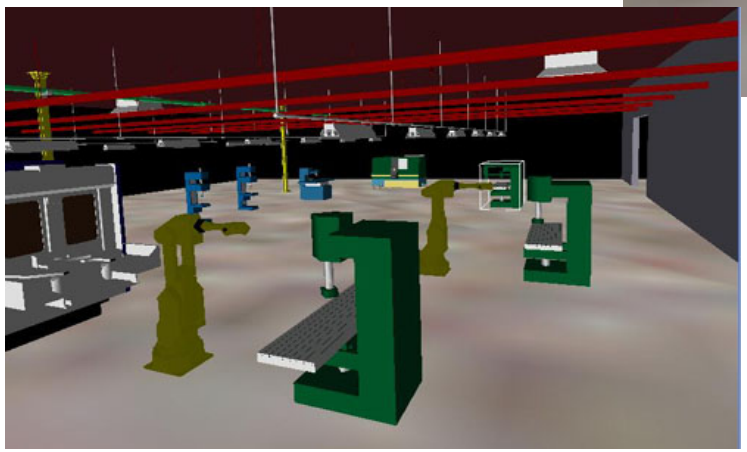


J. Anstey

VR-Fact!

- Interactive virtual factory
- Creates digital mock-up of factory
- Drag & place modular machines
- Mathematical algorithms for consistency checks





Kesh

Western New York

Some Facts

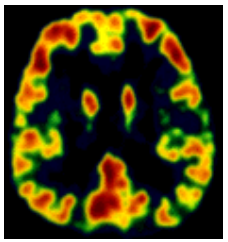
Buffalo, New York

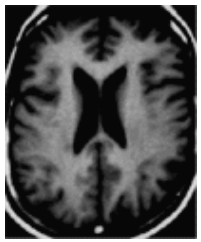
- The Queen City: 2nd Largest City in NYS
- City of Lights
  - First U.S. city to have electric street lights
  - Pan American Exposition (1901)
    - Pres. McKinley Shot

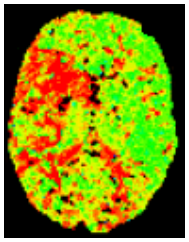
- Architecture
  - Frederick Law Olmsted
  - Frank Lloyd Wright
- Underground Railroad
  - Slaves escaped to freedom in Canada
- Four straight Super Bowl appearances
- Culinary Delights
  - Beef on Weck, Pizza, Fish Fries
  - (Buffalo) Wings: Anchor Bar, 1964
- Health Problems
  - Heart Disease/Stroke
  - Multiple Sclerosis

Recent Biomedical Advances (Buffalo, NY)

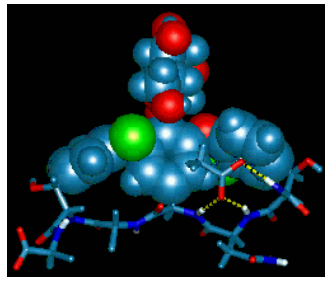
- PSA Test (screen for Prostate Cancer)
- Avonex: Interferon Treatment for Multiple Sclerosis
- Artificial Blood
- Nicorette Gum
- Fetal Viability Test
- Edible Vaccine for Hepatitis C
- Timed-Release Insulin Therapy
- Anti-Arrythmia Therapy
  - Tarantula venom







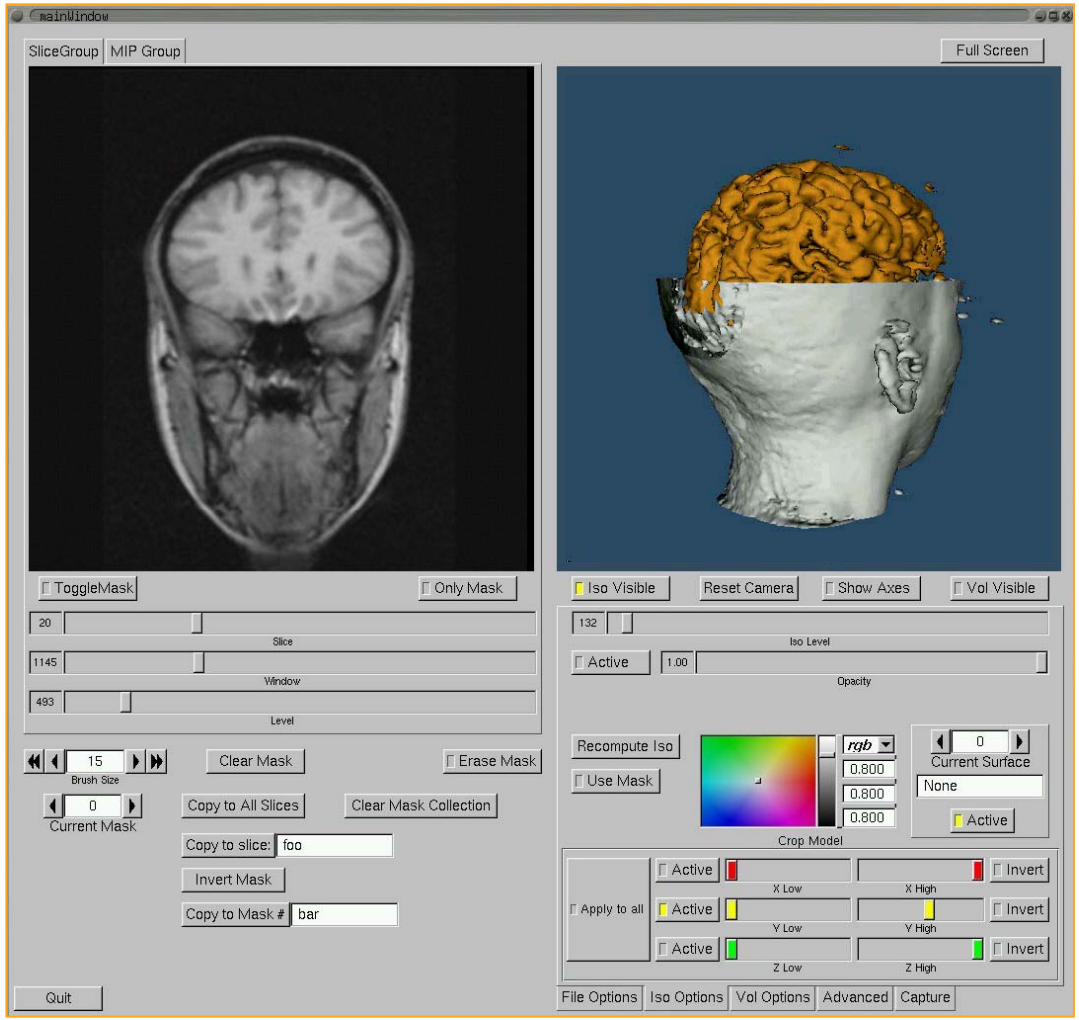
- Direct Methods Structure Determination
  - Listed on "Top Ten Algorithms of the 20th Century"
- Vancomycin
- Gramacidin A
- High Throughput Crystallization Method: Patented
- NIH National Genomics Center: Northeast Consortium
- Howard Hughes Medical Institute: Center for Genomics & Proteomics



Scientific Visualization

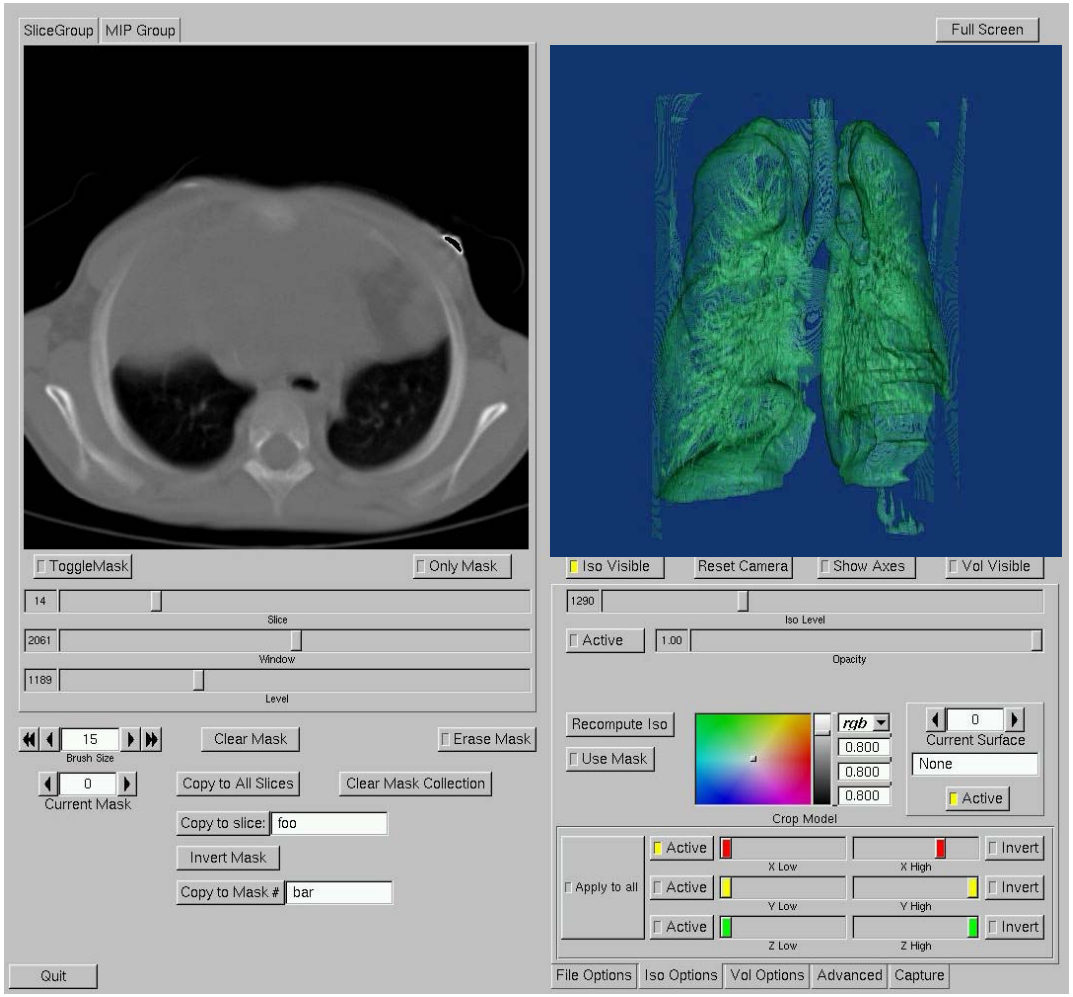
Multiple Sclerosis Project

- **Collaboration with Buffalo Neuroimaging Analysis Center (BNAC)**
  - **Developers of Avonex, drug of choice for treatment of MS**
- **MS Project examines patients and compares scans to healthy volunteers**



3D Medical Visualization

- Reads data output from a CT or MRI Scan
- Collaboration with Children's Hospital
- Visualize multiple surfaces and volumes
- Export images, movies or CAD file
- Pre-surgical planning
- Runs on a PC



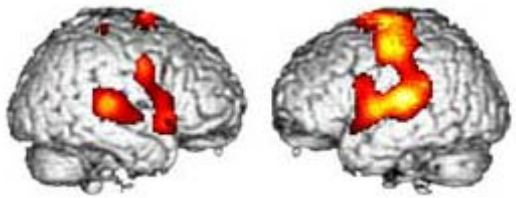
M. Innus

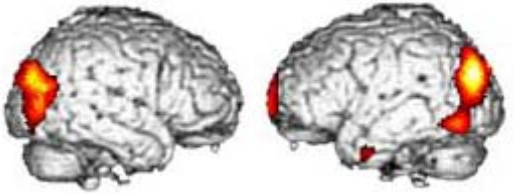
Mapping Brain Activity

Positron emission tomography (PET), shows sites activated and deactivated as subjects decide whether a sound is a target or not.

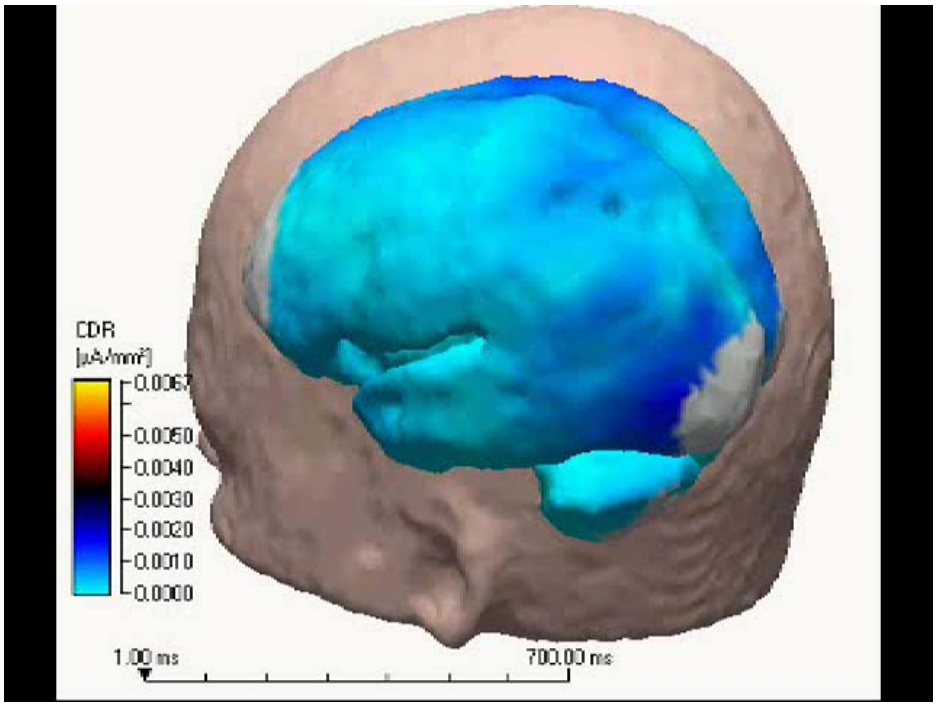
Sites Activated

Sites Deactivated





Current density maps of brain surface (1–700 ms after target) show dynamic pattern of brain activity during decision-making process.



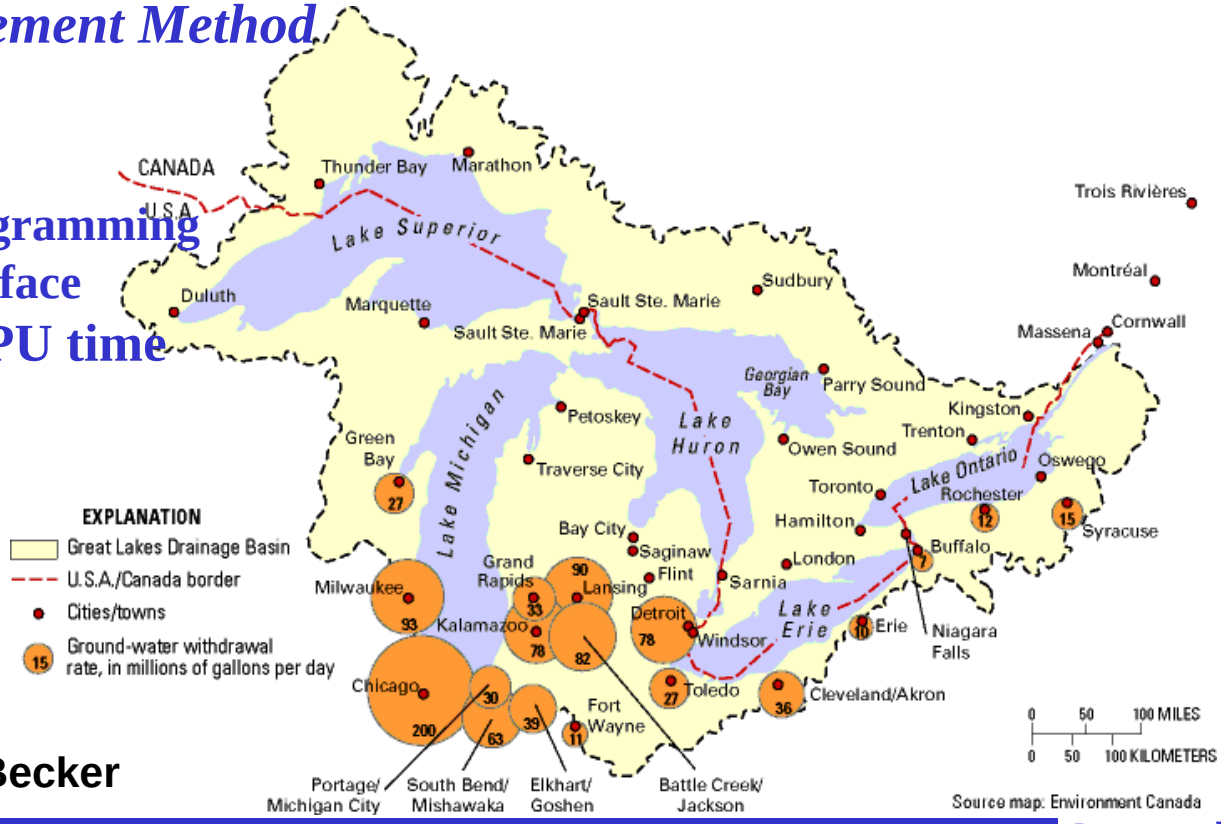
A. Lockwood

Science & Engineering

Small Subset of Projects

Groundwater Flow Modeling

- Regional scale modeling of groundwater flow and contaminant transport (Great Lakes)
- Ability to include all hydrogeologic features as independent objects
- Based on *Analytic Element Method*
- Key features:
  - Highly parallel
  - Object-oriented programming
  - Intelligent user interface
- Utilized 42 years of CPU time on CCR computers in 1 calendar year

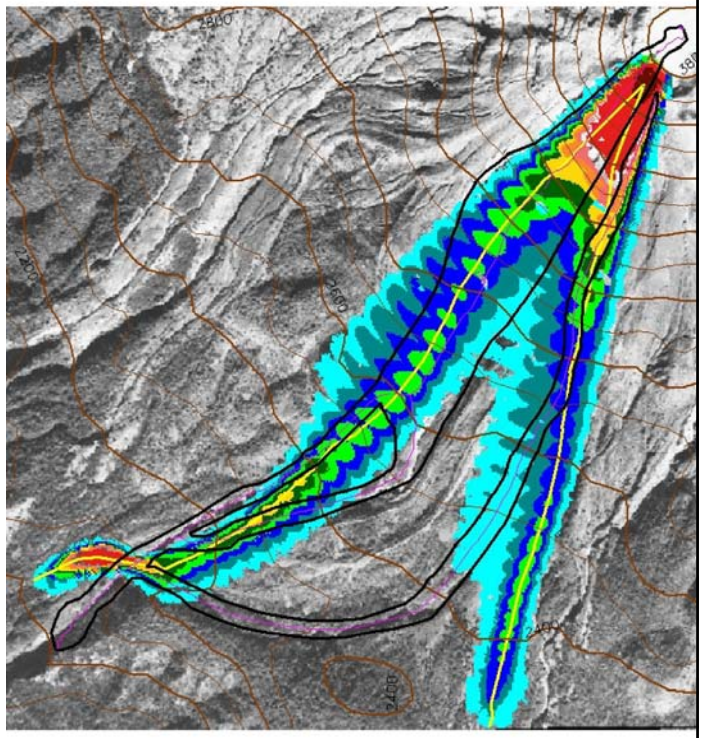


A. Rabideau, I. Jankovic, M. Becker

Avalanches, Volcanic and Mud Flows

Geology, Engineering

- **Modeling of Volcanic Flows, Mud flows (flash flooding), and avalanches**
- **Integrate information from several sources**
  - **Simulation results**
  - **Remote sensing**
  - **GIS data**
- **Present information to decision makers using custom visualization tools local & remote**
- **GRID enabled for remote access**
- **Key Features**
  - **Parallel Adaptive Computation**
  - **Integrated with GIS System for flows on natural terrain**



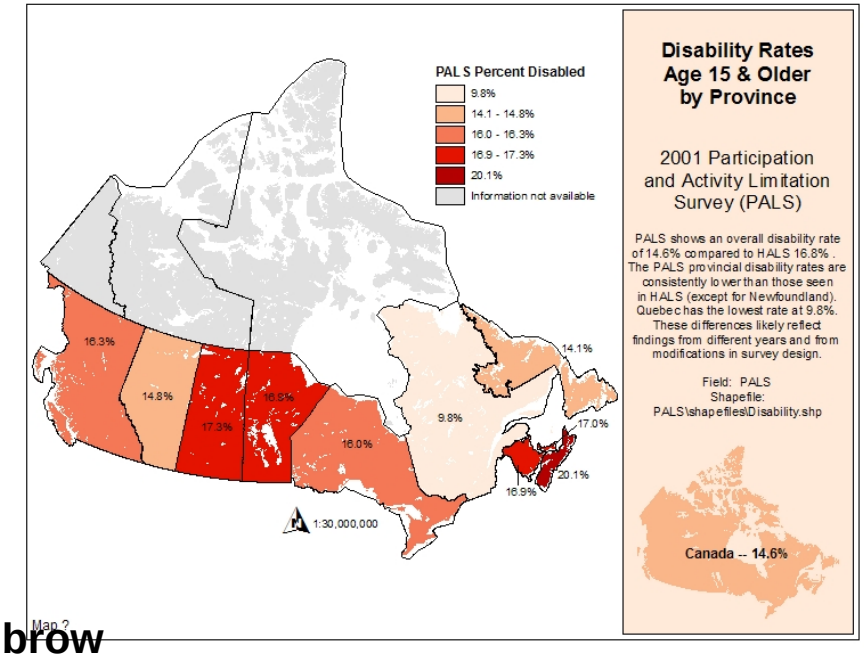
Flow models of Colima volcano
In Mexico – courtesy Rupp et. al.'06

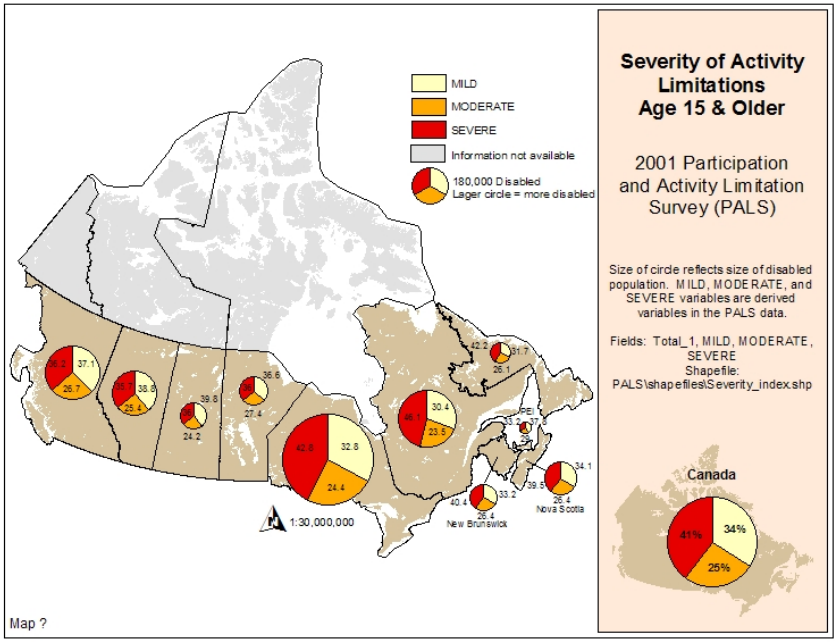
A. Patra, B. Pitman, M. Sheridan, M. Jones

Anthropology

Literacy & Disability in Canada

- **Exploring the relationship between illiteracy & disability across the Canadian landscape**
- **Social Systems GIS Lab in the Dept. of Anthropology is working with researchers from York University & the Canadian Abilities Foundation.**
- **Sponsored by The Adult Learning & Literacy Directorate of the Ministry of Human Resources & Social Development Canada.**

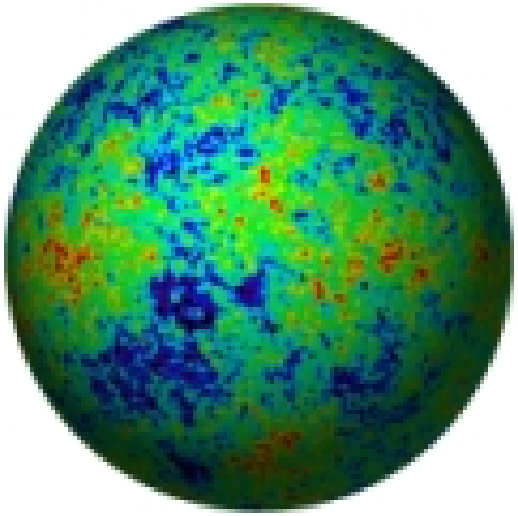


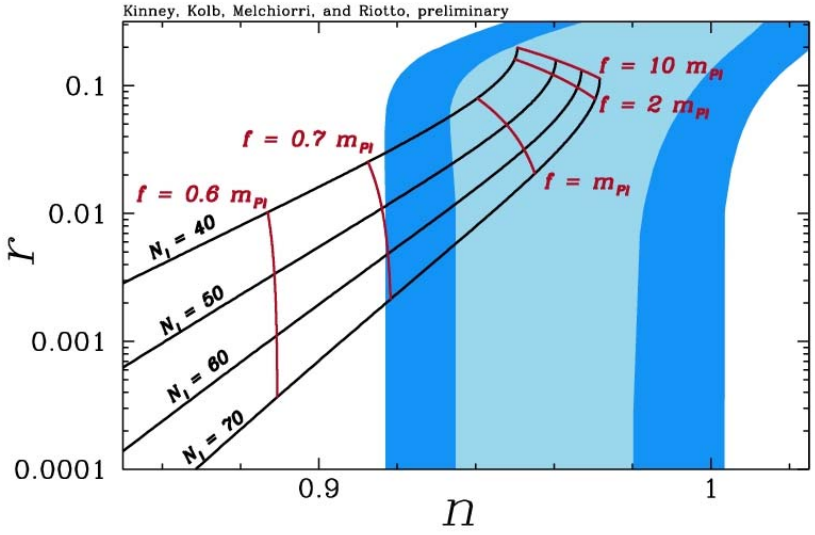


E. Zubrow

Cosmological Parameter Estimation

- **Wealth of new precision cosmological data**
- **WMAP Cosmic Microwave Background Measurement**
- **Sloan Digital Sky Survey: 3-D map of a million galaxies**
- **Interpret implications of data for models of the first trillionth of a second of the universe: *inflation***
- ***Monte Carlo Markov Chain data analysis: stochastic exploration of many-dimensional parameter spaces***





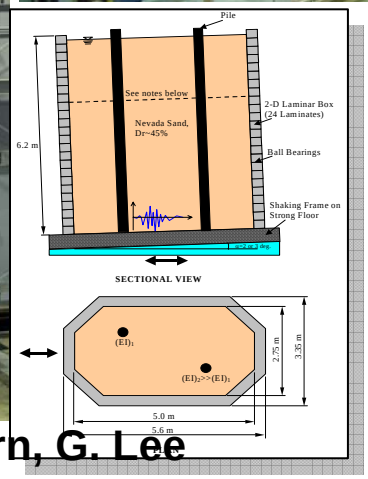
W. Kinney

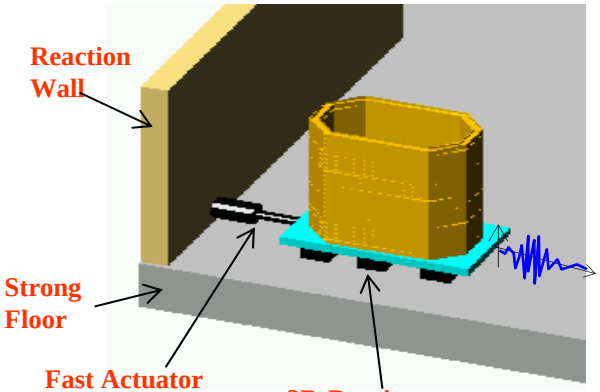
UB's Structural Engineering and Earthquake Simulation Laboratory (SEESL)

Structural Engineering

NEESWood: Development of a Performance-Based Seismic Design for Woodframe Construction:

Two-story Townhouse on Twin Shake Tables



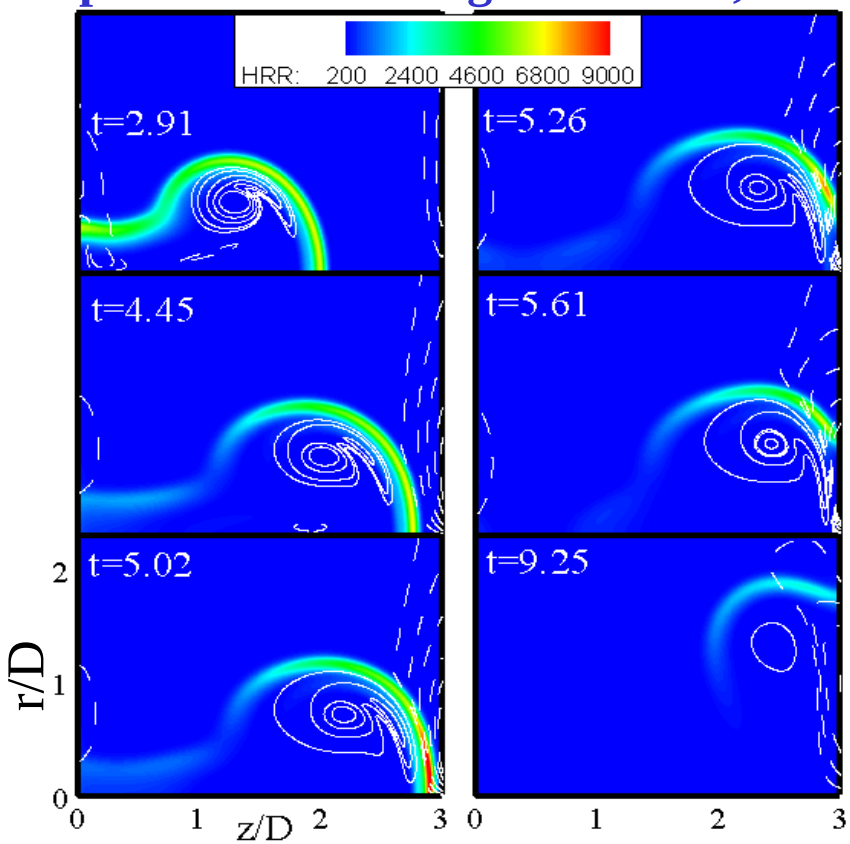


2-D Geotechnical Laminar Box Tests of Pile Foundations Subjected to Soil Liquefaction

M. Bruneau, A. Reinhorn, G. Lee

Understanding Combustion

- Flame-wall interaction modeling for a non-premixed flame propelled by a vortex ring.
- In this figure different time instants are shown during the interaction. White line contours and color contours represent vortex ring and flame, respectively.
- Key Features:
  - Modeling of Detailed GRI3. Mechanism for Methane Combustion
  - Parallel algorithm using mpi
  - 85-90% Parallel efficiency for up to 64 processors
- FWI study is important to determine
  - Engine Design
  - Quenching Distances
  - Flame Structure
  - Unburned hydrocarbon
  - Maximum Wall heat fluxes

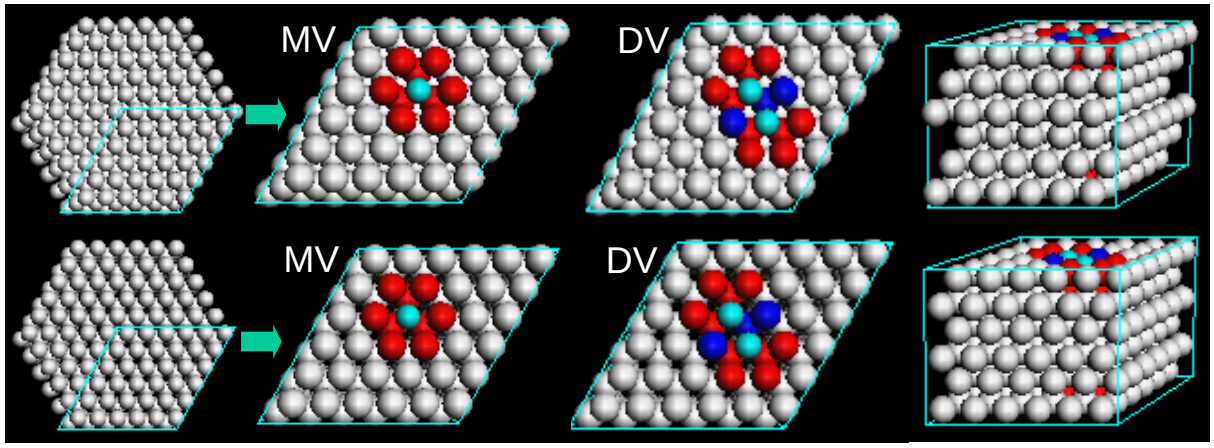


C. Madnia

Computational Materials Science

- Molecular and mesoscale modeling used to understand the behavior of materials
- Example application: Electromigration
- Strong electrical currents cause movement of atoms in metal
- Result is large defects that lead to failure of electrical connection
- Consequences can be catastrophic
- Interdisciplinary experimental/ modeling studies leading to understanding of behavior

Photos of metal lines that have developed voids (above) and hillocks (below) due to electromigration. (source: www.nd.edu)



Simulation cells of solids with mono- and di-vacancies (light blue spheres), highlighting atoms neighboring the defects.

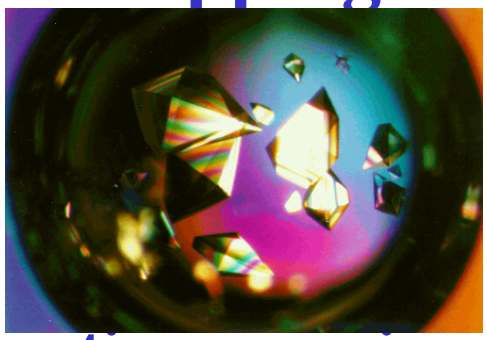
D. Kofke

Shake-and-Bake

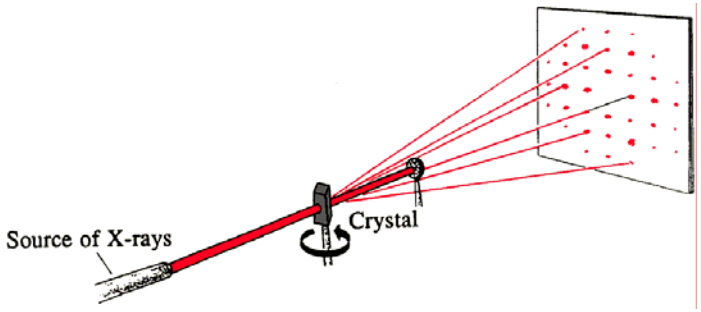
Molecular Structure Determination from X-Ray Crystallographic Data

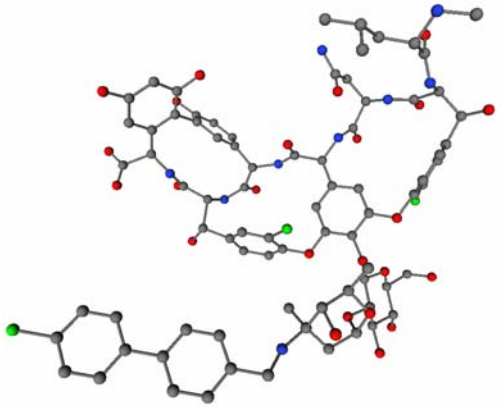
X-Ray Crystallography

- **Objective: Provide a 3-D mapping of the atoms in a crystal.**



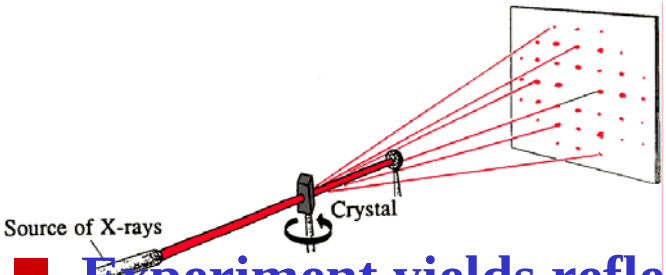
- **Procedure:**
  1. **Isolate a single crystal.**
  2. **Perform the X-Ray diffraction experiment.**



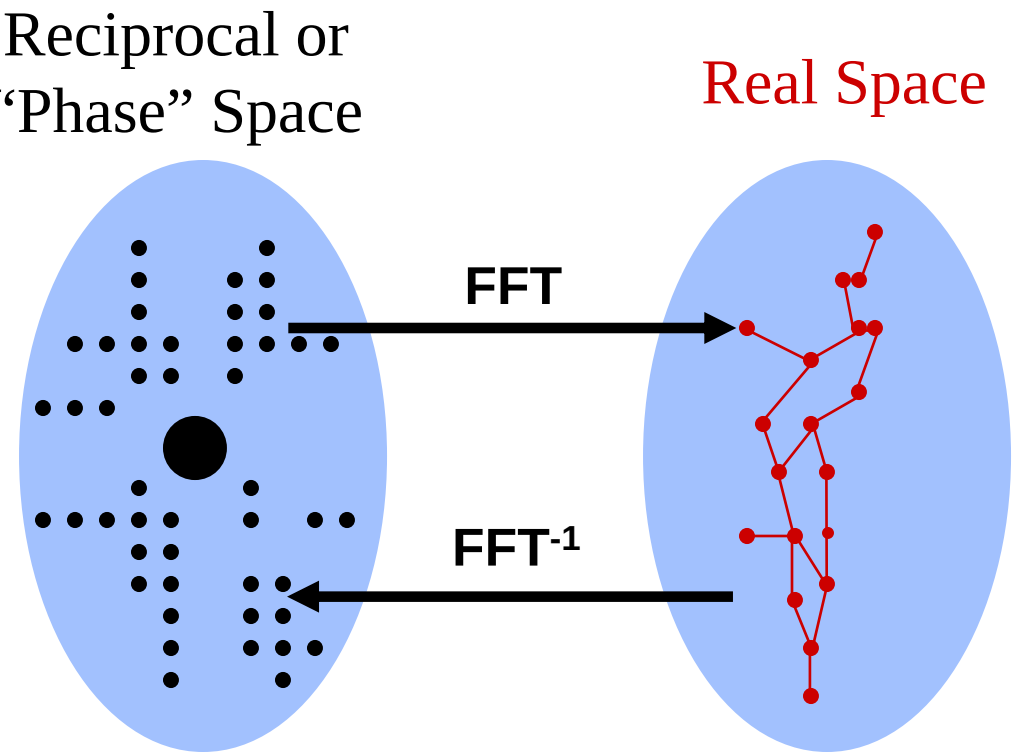


  3. **Determine molecular structure that agrees with diffration data.**

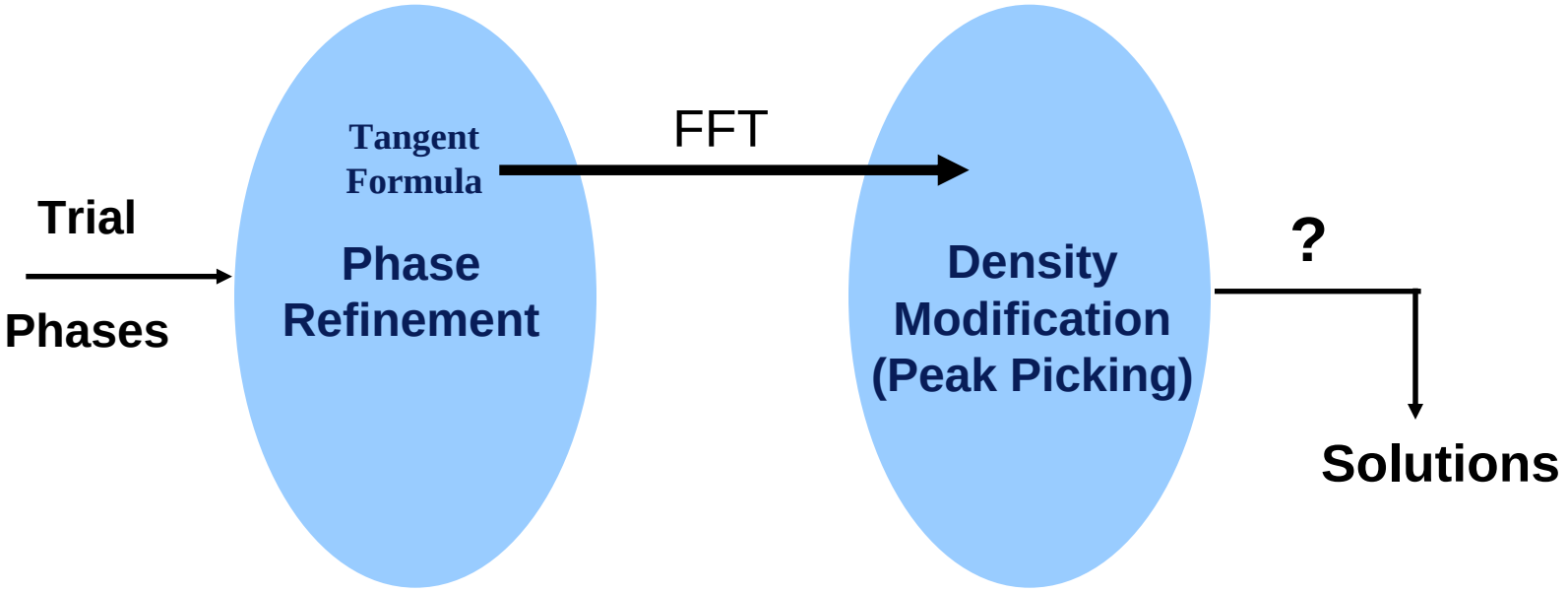
X-Ray Data & Corresponding Molecular Structure



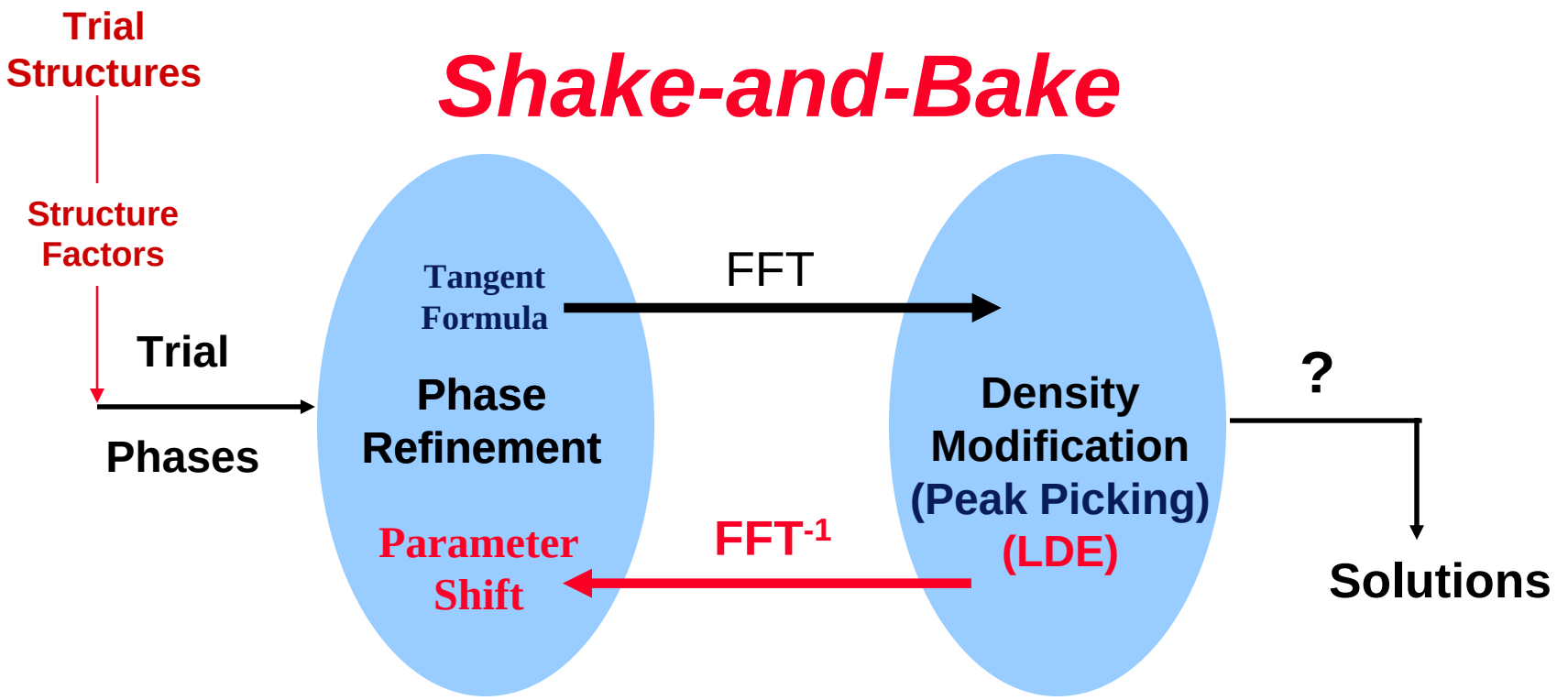
- **Experiment yields reflections and associated intensities.**
- **Underlying atomic arrangement is related to the reflections by a 3-D Fourier transform.**
- ***Phase angles are lost in experiment.***
- ***Phase Problem*: Determine the set of phases corresponding to the reflections.**



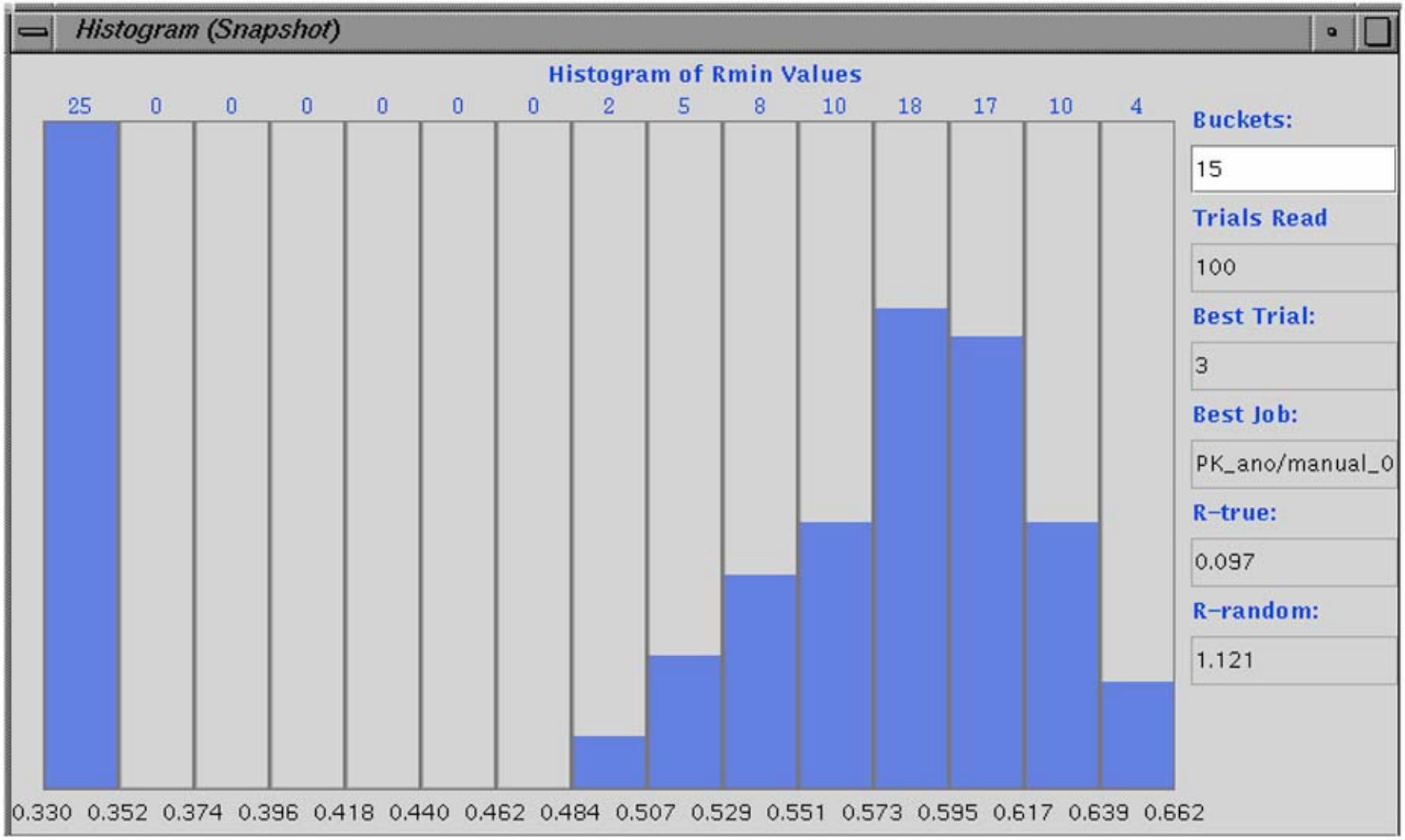
Conventional Direct Methods



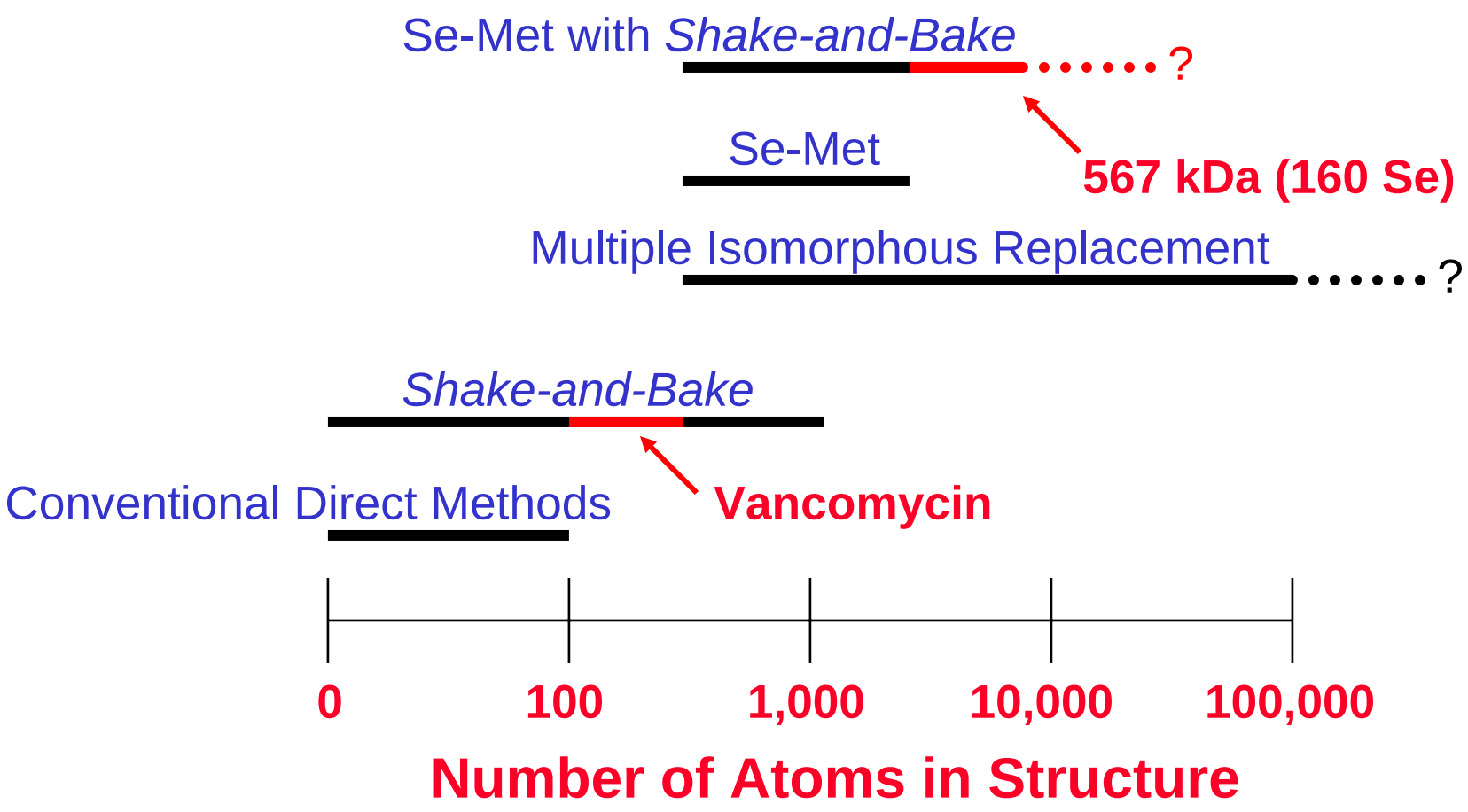
Shake-and-Bake Method: Dual-Space Refinement



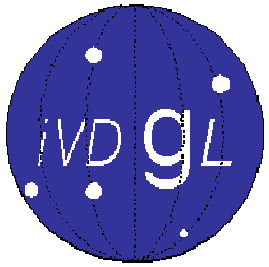
Ph8755: *SnB* Histogram

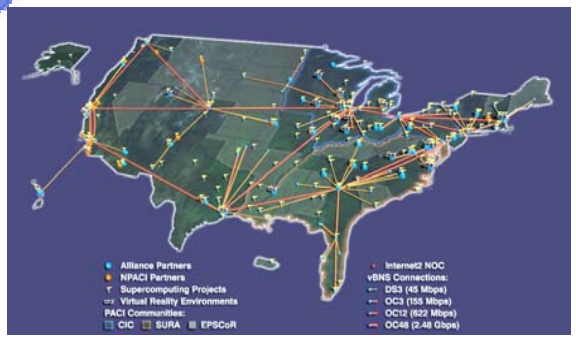


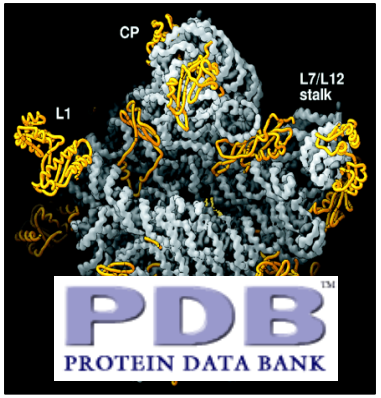
Phasing and Structure Size

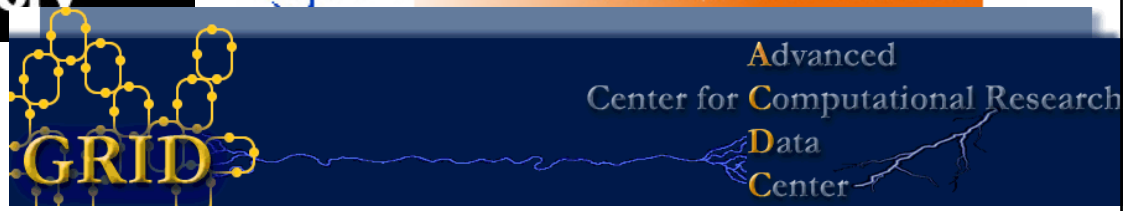


Grid Computing

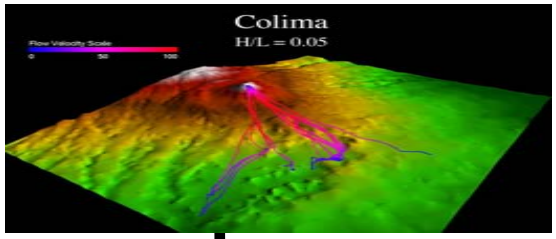




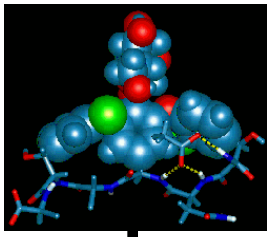




Grid Computing Overview



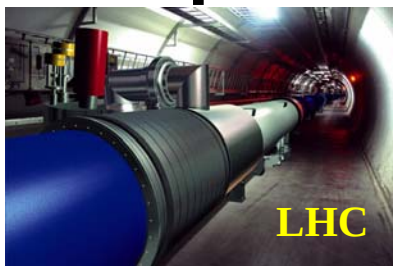
Data Acquisition

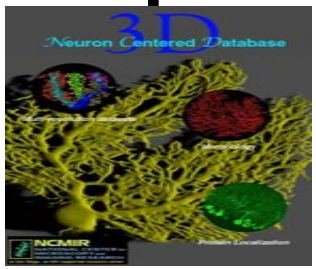


Advanced Visualization

Analysis

Imaging Instruments





Large-Scale Databases

- Coordinate Computing Resources, People, Instruments in Dynamic Geographically-Distributed Multi-Institutional Environment
- Treat Computing Resources like Commodities
  - Compute cycles, data storage, instruments
  - Human communication environments
- No Central Control; No Trust

Major Grid Initiatives

- TeraGrid (NSF)
  - Integrates High-End Resources
  - High-Performance (Dedicated) Networks
  - 9 Sites (?); 250TF & 30PB (?)
  - 100+ Databases Available
- OSG (DOE, NSF)
  - High-Throughput Distributed Facility
  - Open & Heterogeneous
  - Biology, Computer Science, Astrophysics, LHC
  - 57 Compute Sites; 11 Storage Sites;
  - 10K CPUS; 6PB
- EGEE: Enabling Grids for E-SciencE (European Commision)
  - Initial Focus on CERN (5PB of Data/Year)
    - High-Energy Physics and Life Sciences
  - Expanded Focus Includes Virtually All Scientific Domains
  - 200 Institutions; 40 Countries
  - 20K+ CPUs; 5PB; 25,000 jobs per day!

Miller's Cyberinfrastructure Lab

- **CI sits at core of modern simulation & modeling**
- **CI allows for new methods of investigation to address previously unsolvable problems**
- **Focus on development of *algorithms*, *portals*, *interfaces*, *middleware***
- **Free end-users to do disciplinary work**
- **Funding (2001-pres): NSF ITR, NSF CRI, NSF MRI, NYS, Fed**
- **Experimental Equipment (Dell/Lenovo): 1.25 TF Clusters, 140 Cores (Intel/AMD), 4 TB Internal Storage, GigE, IB, Condor Flock (35 Intel/AMD), 22 TB Storage (2)**
- **Production Equipment (Dell): Workstations, 15 TB Storage, CCR equipment**

Evolution of CI Lab Projects

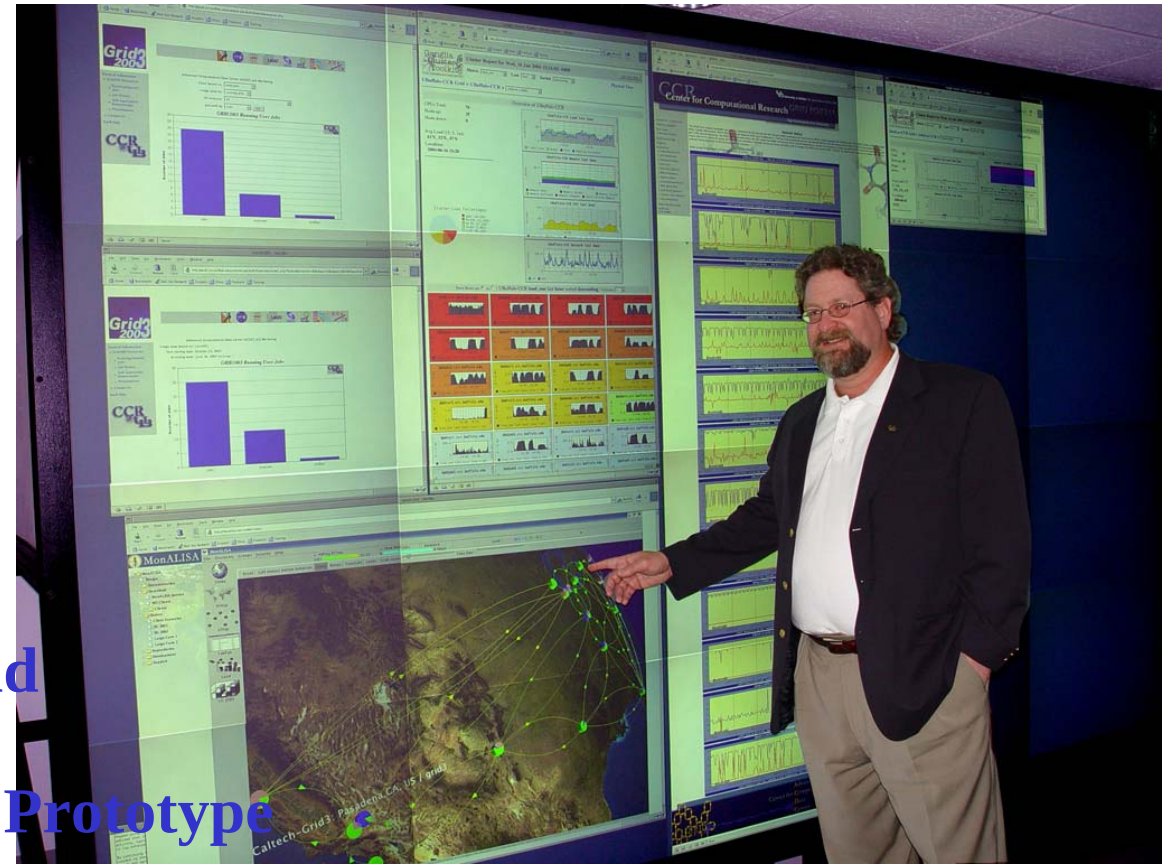
- **ACDC-Grid**
  - **Experimental Grid: Globus & Condor**
  - **Integrate Data & Compute, Monitor, Portal, Node Swapping, Predictive Scheduling/Resource Management**
  - **GRASE VO: Structural Biology, Groundwater Modeling, Earthquake Eng, Comp Chemistry, GIS/BioHazards**
  - **Buffalo, Buffalo State, Canisius, Hauptman-Woodward**
- **WNY Grid**
  - **Heterogeneous System: Hardware, Networking, Utilization**
  - **Buffalo, Geneseo, Hauptman-Woodward, Niagara**
- **NYS Grid**
  - **Extension to Hardened Production-Level System State-Wide**
  - **Albany, Binghamton, Buffalo, Geneseo, Canisius, Columbia, HWI, Niagara, [Cornell, NYU, RIT, Rochester, Syracuse, Marist], {Stony Brook, RPI, Iona}**

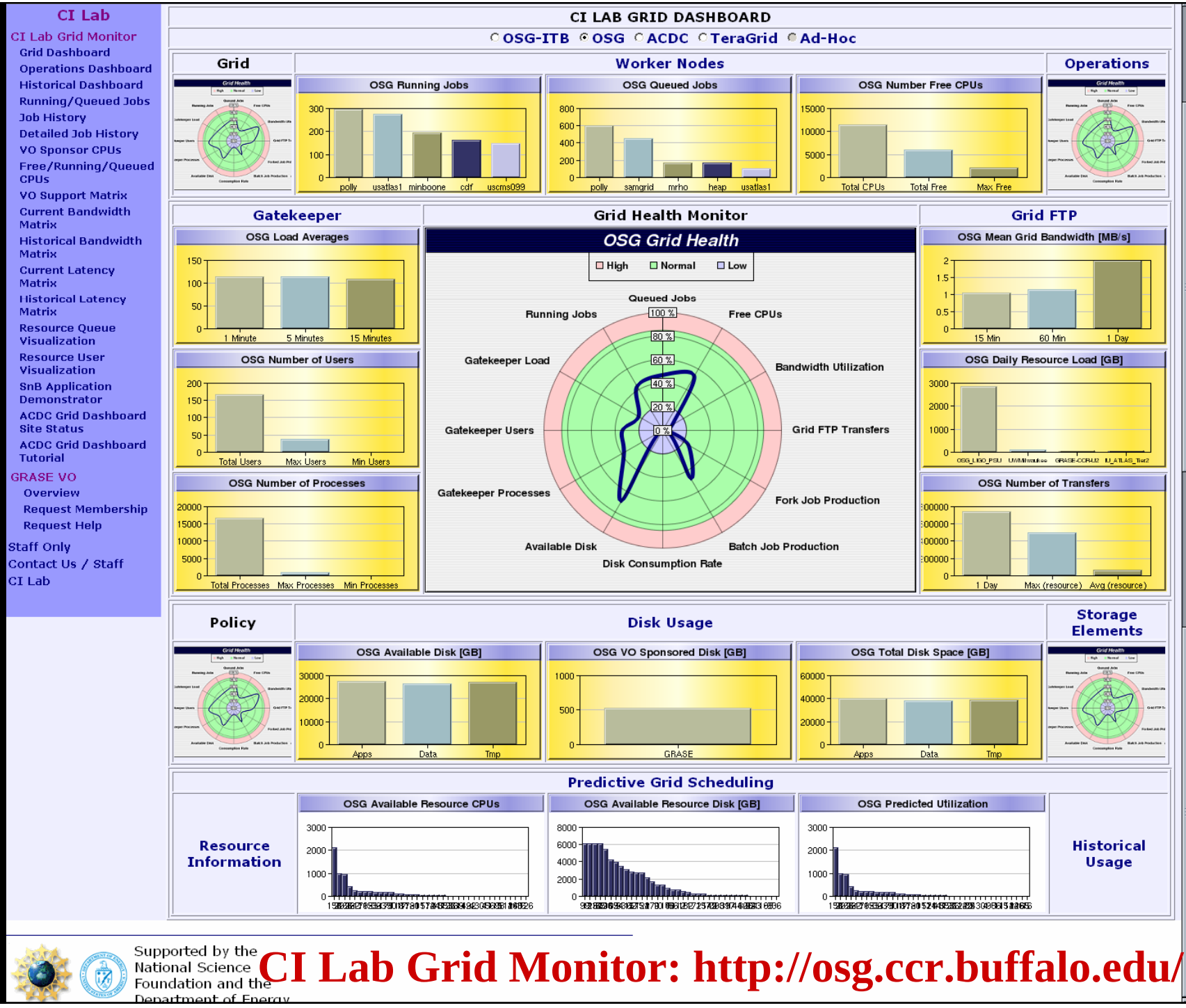
NYS Grid Resources

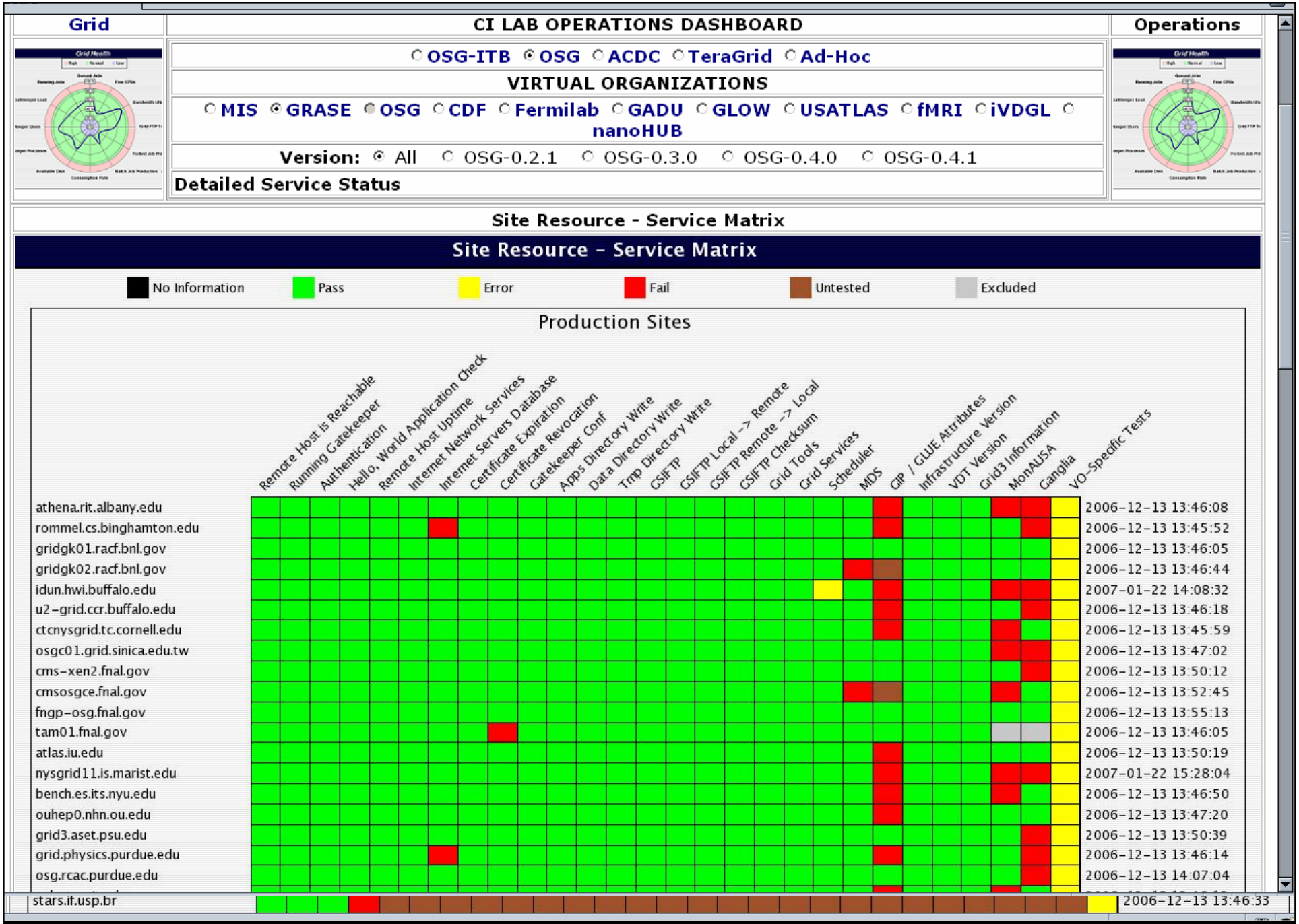
- Albany: 8 Dual-Processor Xeon Nodes
- Binghamton: 15 Dual-Processor Xeon Nodes
- Buffalo: 1050 Dual-Processor Xeon Nodes
- Cornell: 30 Dual-Processor Xeon Nodes
- Geneseo State: Sun/AMD with 128 Compute Cores
- Hauptman-Woodward Institute: 50 Dual-Core G5 Nodes
- Marist: 9 P4 Nodes
- Niagara University: 64 Dual-Processor Xeon Nodes
- NYU: 58 Dual-Processor PowerPC Nodes
- RIT: 4 Dual-Processor Xeon Nodes
- Syracuse: 8 Dual-Processor Xeon Nodes

CI Lab Collaborations

- High-Performance Networking Infrastructure
- Grid3+ Collaboration
- iVDGL Member
  - Only External Member
- Open Science Grid
  - GRASE VO
- NYSGrid.org
  - NYS CI Initiative
  - Executive Director
  - Various WGs
- Grid-Lite: Campus Grid
  - HP Labs Collaboration
- Innovative Laboratory Prototype
  - Dell Collaboration




CI Lab
CI Lab Grid Monitor
Grid Dashboard
Operations Dashboard
Historical Dashboard
Running/Queued Jobs
Job History
Detailed Job History
VO Sponsor CPUs
Free/Running/Queued CPUs
VO Support Matrix
Current Bandwidth Matrix
Historical Bandwidth Matrix
Current Latency Matrix
Historical Latency Matrix
Resource Queue Visualization
Resource User Visualization
SnB Application Demonstrator
ACDC Grid Dashboard Site Status
ACDC Grid Dashboard Tutorial
GRASE VO
Overview
Request Membership
Request Help
Staff Only
Contact Us / Staff
CI Lab
CI LAB GRID DASHBOARD
OSG-ITB OSG ACDC TeraGrid Ad-Hoc
Grid
Worker Nodes
Operations
OSG Running Jobs
polly usatlas1 minboone cdf uscms099
OSG Queued Jobs
polly samgrid mrho heap usatlas1
OSG Number Free CPUs
Total CPUs Total Free Max Free
Gatekeeper
OSG Load Averages
1 Minute 5 Minutes 15 Minutes
OSG Number of Users
Total Users Max Users Min Users
OSG Number of Processes
Total Processes Max Processes Min Processes
Grid Health Monitor
OSG Grid Health
High Normal Low
Queued Jobs
Running Jobs
Free CPUs
Gatekeeper Load
Bandwidth Utilization
Gatekeeper Users
Grid FTP Transfers
Gatekeeper Processes
Fork Job Production
Available Disk
Batch Job Production
Disk Consumption Rate
Grid FTP
OSG Mean Grid Bandwidth [MB/s]
15 Min 60 Min 1 Day
OSG Daily Resource Load [GB]
OSG Number of Transfers
1 Day Max (resource) Avg (resource)
Policy
Disk Usage
Storage Elements
OSG Available Disk [GB]
Apps Data Tmp
OSG VO Sponsored Disk [GB]
GRASE
OSG Total Disk Space [GB]
Apps Data Tmp
Predictive Grid Scheduling
Resource Information
OSG Available Resource CPUs
OSG Available Resource Disk [GB]
OSG Predicted Utilization
Historical Usage
Supported by the National Science Foundation and the Department of Energy
CI Lab Grid Monitor: http://osg.ccr.buffalo.edu/



CI Lab Operations Dashboard

Grid-Enabling Application Templates (GATs)

- **Structural Biology**
  - ***SnB* and *BnP* for Molecular Structure Determination/Phasing**
- **Groundwater Modeling**
  - ***Ostrich*: Optimization and Parameter Estimation Tool**
  - ***POMGL*: Princeton Ocean Model Great Lakes for Hydrodynamic Circulation**
  - ***Split*: Modeling Groundwater Flow with Analytic Element Method**
- **Earthquake Engineering**
  - ***EADR*: Evolutionary Aseismic Design and Retrofit; Passive Energy Dissipation System for Designing Earthquake Resilient Structures**
- **Computational Chemistry**
  - ***Q-Chem*: Quantum Chemistry Package**
- **Geographic Information Systems & BioHazards**
  - ***Titan*: Computational Modeling of Hazardous Geophysical Mass Flows**

Grid Enabled *SnB*

- **Required Layered Grid Services**
  - **Grid-enabled Application Layer**
    - ***Shake – and – Bake* application**
    - **Apache web server**
    - **MySQL database**
  - **High-level Service Layer**
    - **Globus, NWS, PHP, Fortran, and C**
  - **Core Service Layer**
    - **Metacomputing Directory Service, Globus Security Interface, GRAM, GASS**
  - **Local Service Layer**
    - **Condor, MPI, PBS, Maui, WINNT, IRIX, Solaris, RedHat Linux**

https://grid.ccr.buffalo.edu/

Mail Home Netscape

New Tab | CCR Grid Computing Services:

Cyberinfrastructure Laboratory

Grid Portal

Dr. Russ Miller
UB Distinguished Professor of Computer Science & Engineering

CI Lab
Grid Portal Info
- Overview
- Portal Login
- Grid Account Info

Computational Grid
- Job Submission
- Job/Queue Status
- MDS Information
- Network Status
- Running/Queued Jobs
- PBS Job History
- Condor Flock Statistics
- GAT/Resource Matrix

Data Grid
- Data Grid Tree
- Data Grid Upload
- Data Grid Download
- Data Grid File Manager
- Data Grid Replica Manager
- Data Grid Simulator
- Data Grid Admin Tools
- Data Grid Admin File Tools

Contact Us / Staff
CI Lab
Staff Only

Welcome to the Cyberinfrastructure Laboratory Grid Portal

The **Cyberinfrastructure Laboratory**, in conjunction with the **Center for Computational Research**, has created an integrated Data and Computational Grid. This site is devoted to a Grid Portal that provides access to applications that can be run on a variety of grids. A related site contains a **Grid Monitoring System** designed by the Cyberinfrastructure Laboratory.

Applications may be run on the Cyberinfrastructure Laboratory's **ACDC Grid**, **Western New York Grid**, and **New York State Grid**, which includes computational and data storage systems from dozens of institutions throughout the State of New York.

The applications available to the users cover a variety of disciplines, including Bioinformatics, Computational Chemistry, Crystallography and Medical Imaging, to name a few.

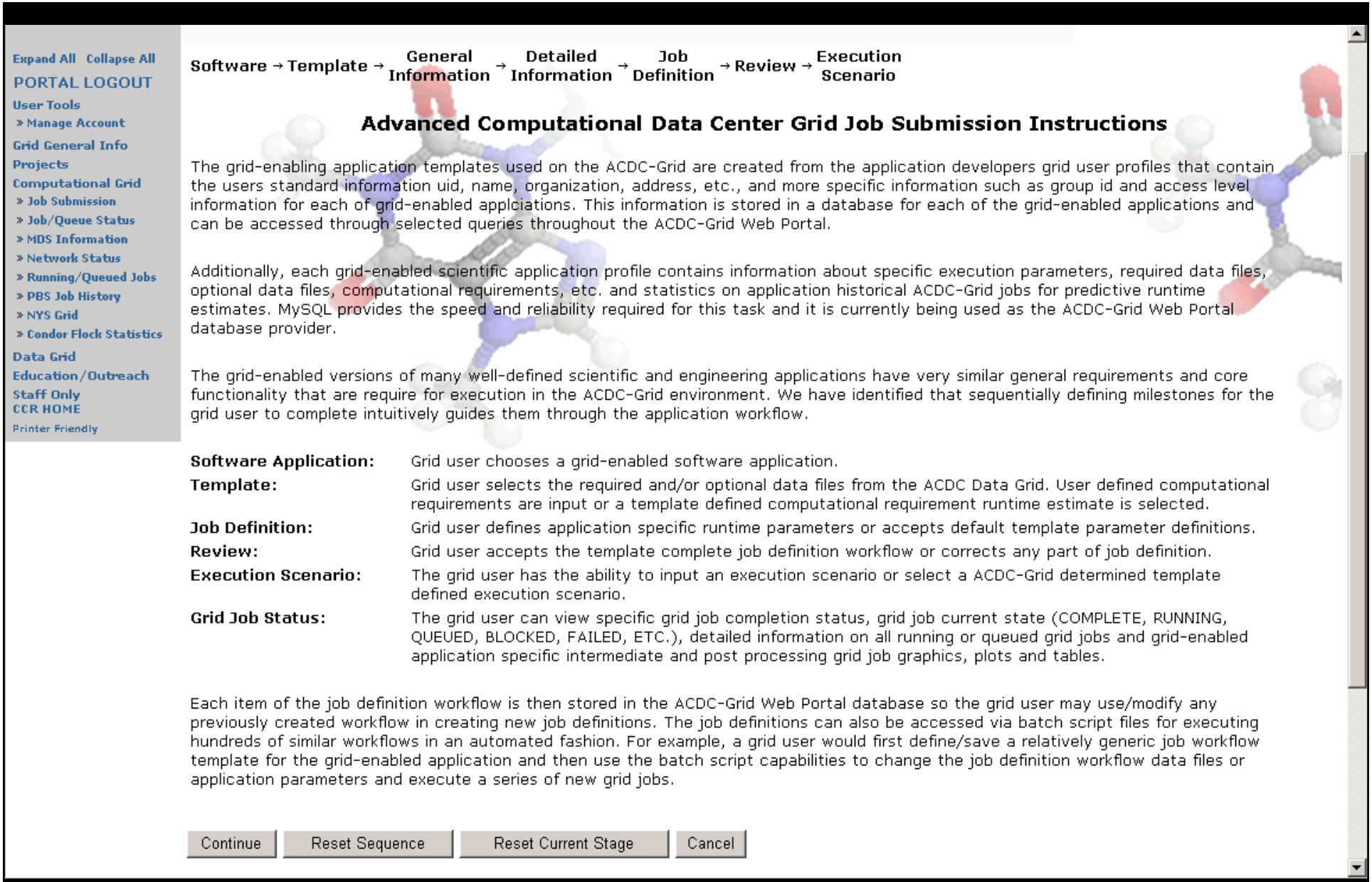
The grids developed by the CI Lab support teaching and research activities, as well as providing infrastructure that includes high-end data, computing, imaging, grid-enabled software, all of which relies on the New York State Research Network (**NYSERNet**).

This work is funded by the National Science Foundation (ITR, MRI, CRI), three program projects from The National Institutes of Health, and the Department of Energy.

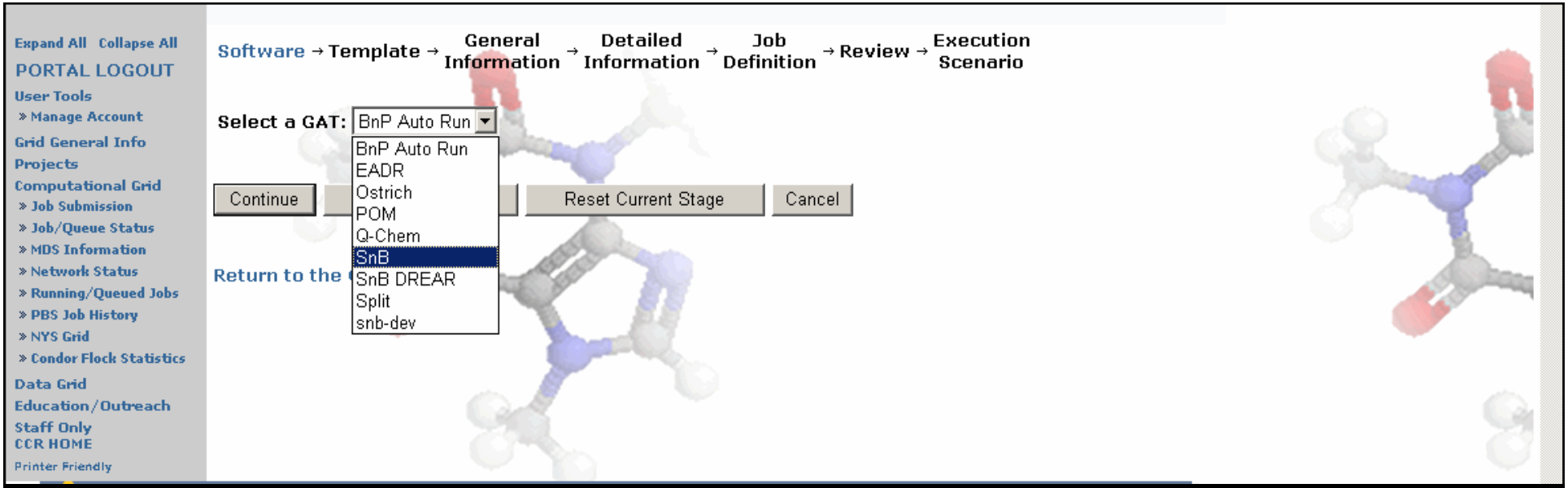
BnP

Software : BnP
Field : Protein crystal structure determination

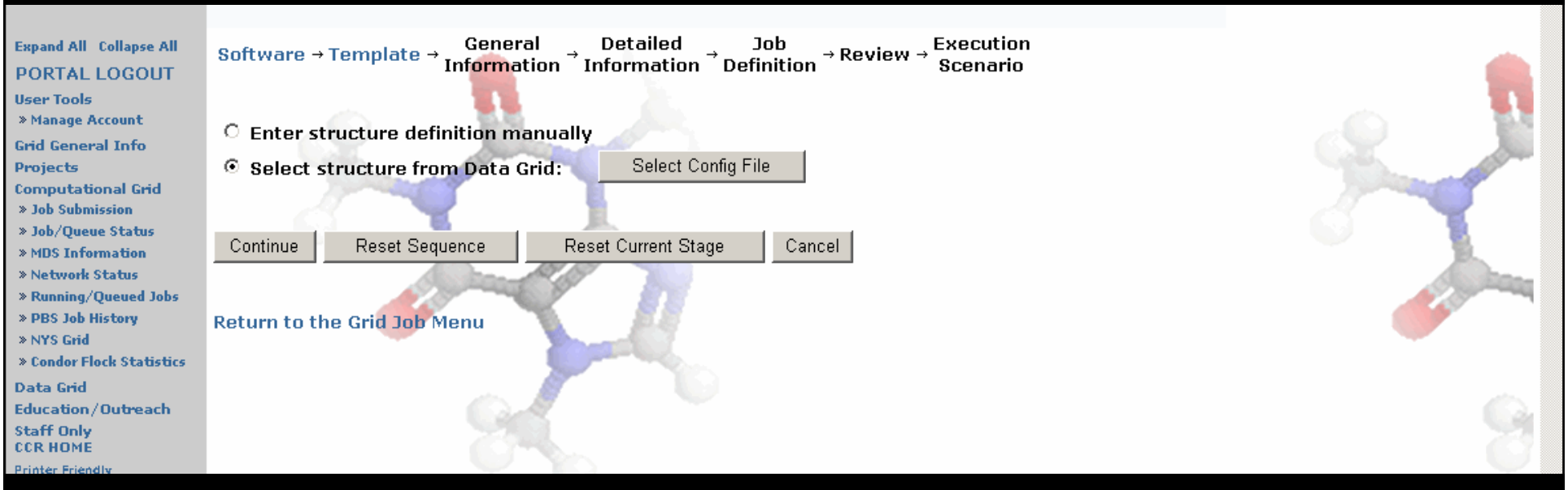
Startup Screen for CI Lab Grid Job Submission



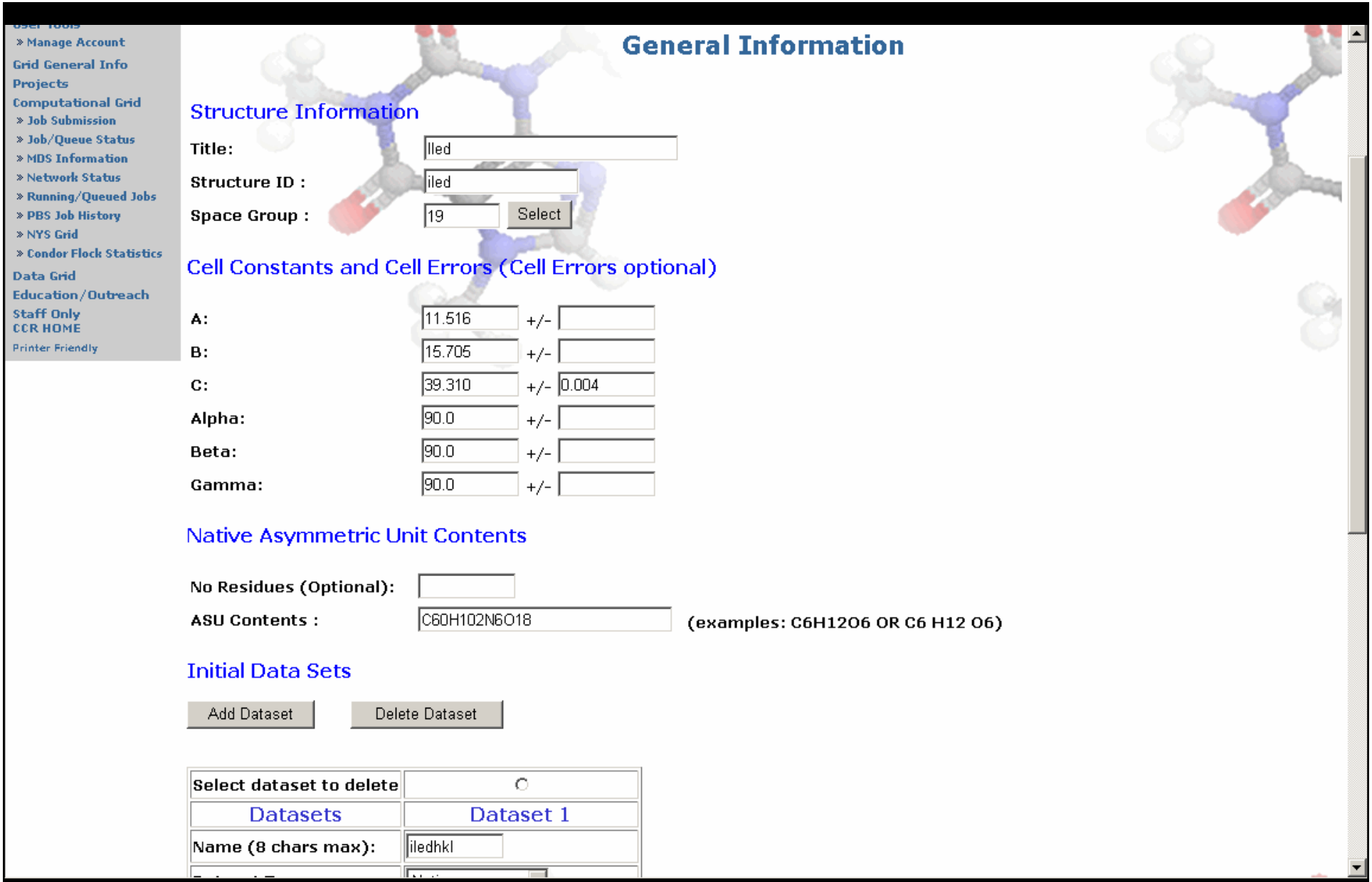
Instructions and Description for Running a Job on ACDC-Grid



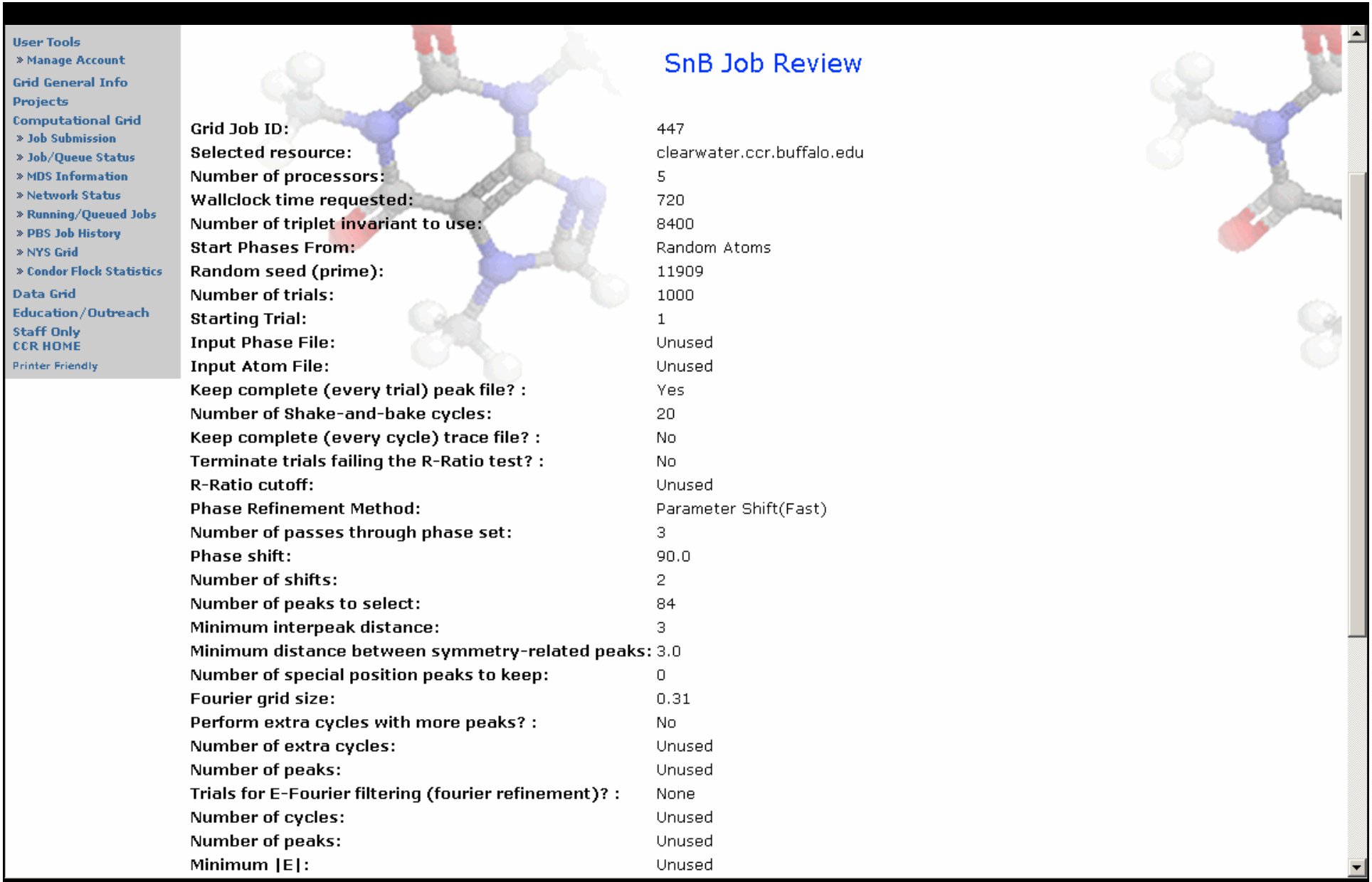
Software Package Selection



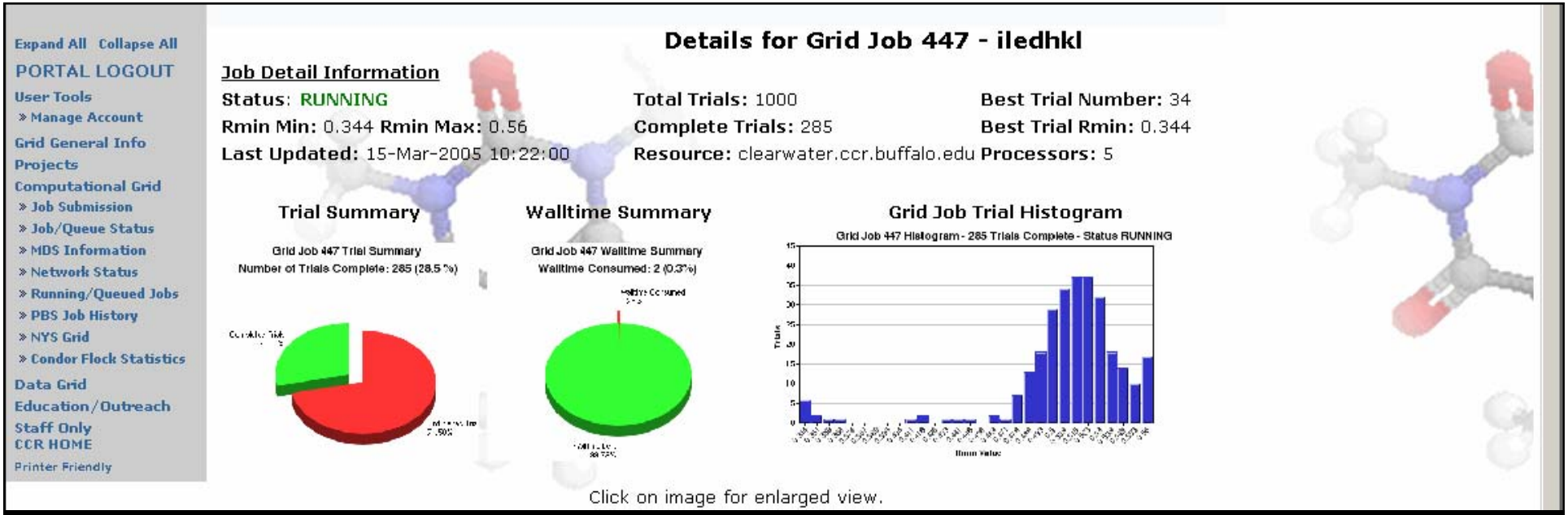
Full Structure / Substructure Template Selection



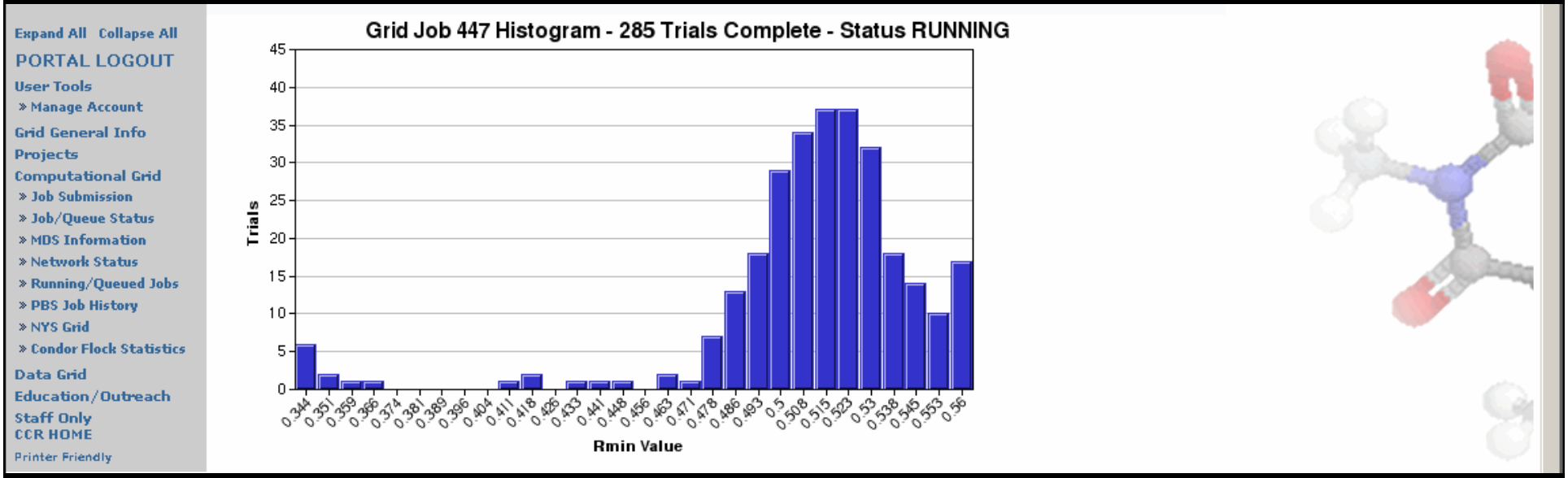
Default Parameters Based on Template



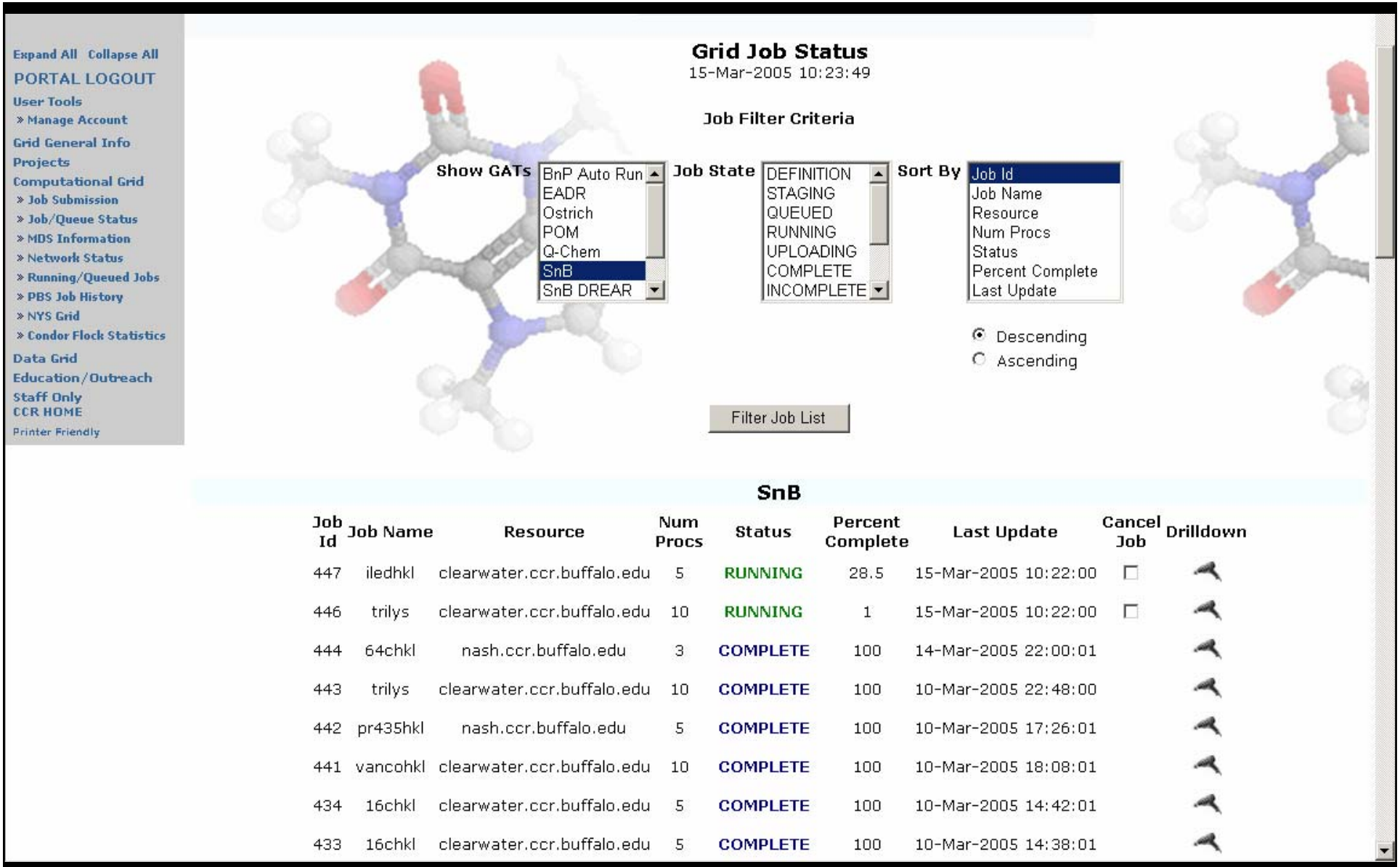
SnB Review (Grid job ID: 447)



Graphical Representation of Intermediate Job Status

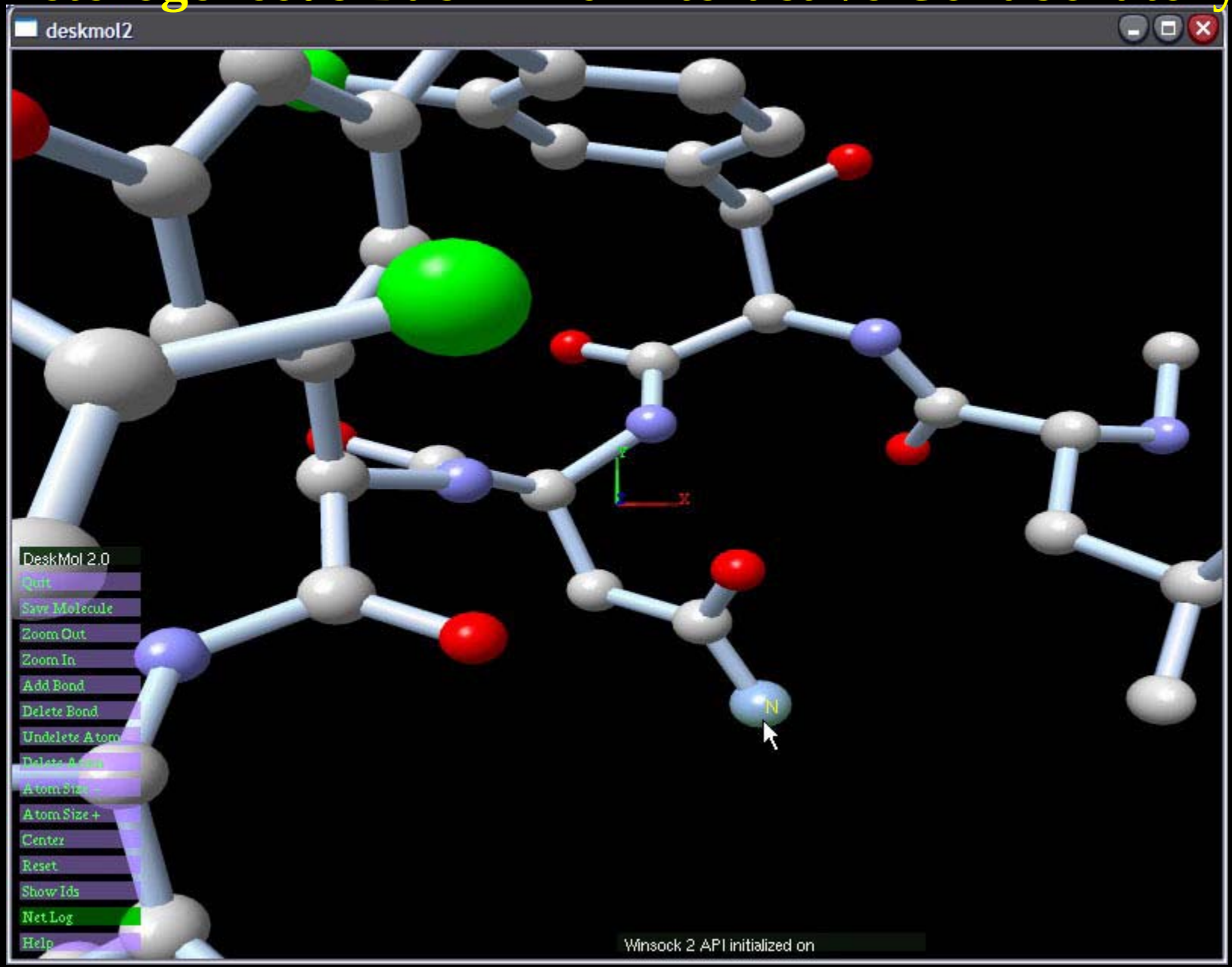


Histogram of Completed Trial Structures

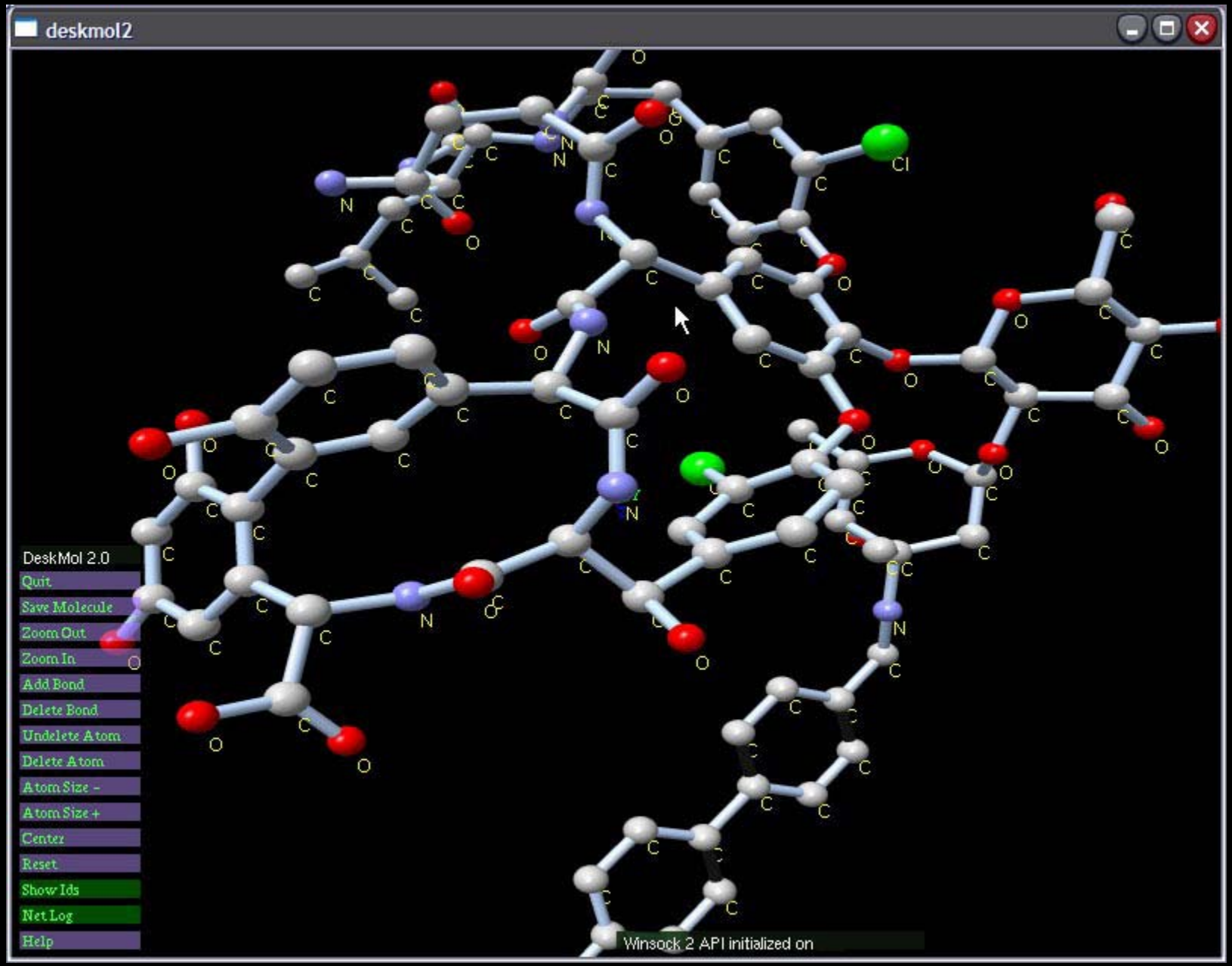


Status of Jobs

Heterogeneous Back-End Interactive Collaboratory



User starts up – default image of structure.



Molecule scaled, rotated, and labeled.

Acknowledgments

- Mark Green
- Cathy Ruby
- Amin Ghadersohi
- Naimesh Shah
- Steve Gallo
- Jason Rappleye
- Jon Bednasz
- Sam Guercio
- Martins Innus
- Cynthia Cornelius

- George DeTitta
- Herb Hauptman
- Charles Weeks
- Steve Potter

- Alan Rabideau
- Igor Janckovic
- Michael Sheridan
- Abani Patra
- Matt Jones

- NSF ITR
- NSF CRI
- NSF MRI
- NYS
- CCR

www.cse.buffalo.edu/faculty/miller