Molecular Structure Determination, High-End Computing, Discovery & Innovation

Russ Miller

Cyberinfrastructure Lab, SUNY-Buffalo

Hauptman-Woodward Med Res Inst

NSF, NIH, DOE, NIMA, NYS, Dell

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

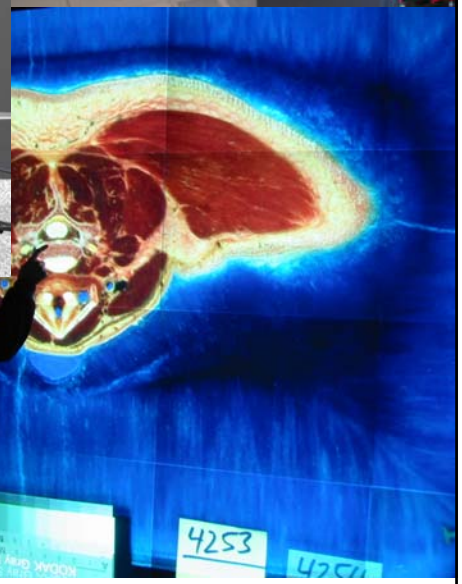
Research Activities

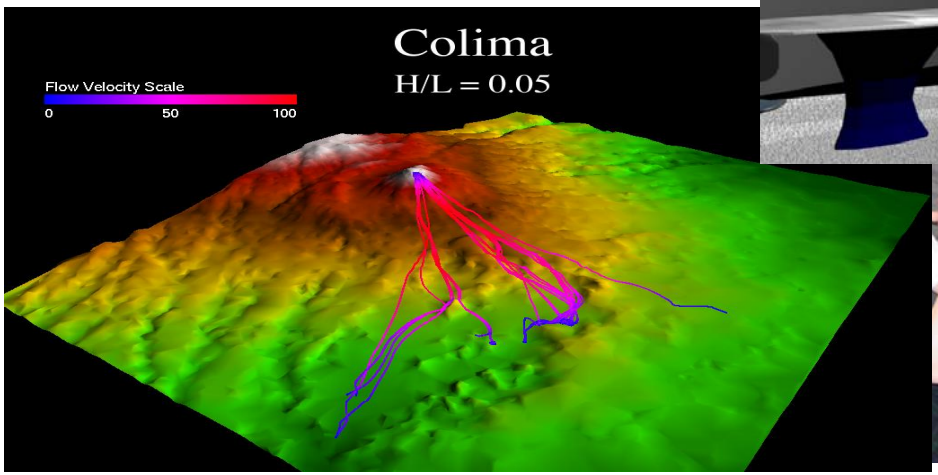
- **Theory/Algorithms**
  - **Fundamental Problems, Data Movement, Computational Geometry, Image Analysis**
  - **Mesh, Pyramid, Hypercube, PRAM, Reconfigurable Mesh, CGM**
- **Experimentation**
  - **Distributed- and Shared-Memory Machines**
  - **Computational Geometry, NP-Hard Approximation Algorithms, Image Analysis**
- **Applications**
  - **Molecular Structure Determination**
- **Systems**
  - **Grid Computing**

"Science is a Team Sport"

Center for Computational Research (CCR): 1998-2006

- **Founding Director (1998-2006)**
- **Facts & Figures**
  - **Top Academic HPC Center in World**
  - **~25 TF of HPC**
  - **~600 TB of High-End Storage**
  - **~30 FTEs Staff**
  - **140 Projects Annually**
  - **1200+ publications**
  - **Software, Media, Algorithms, Consulting, Training, …**





CCR Highlights (1998-2006)

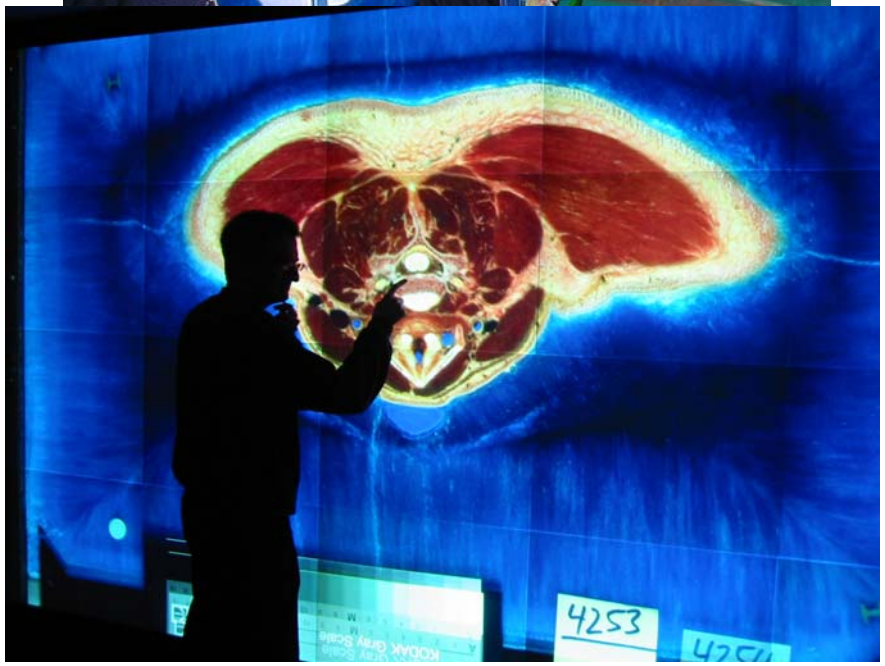
- Provide HE-Comp
- Provide HE-Vis + AGN
- Special Purpose Systems
  - Bioinformatics
  - Data Warehouse / Mining
- Support Local/National Efforts – Industry + Acad
- Create jobs in WNY
- Certificate Program
- Workshops + Tours
  - Campus, Industry
  - High-School
- Urban Planning & Design
- MTV Videos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

Major Compute/Storage Resources

- Dell Linux Cluster (10TF peak)
  - 1600 Xeon EM64T Processors (3.2 GHz)
  - 2 TB RAM; 65 TB Disk
  - Myrinet / Force10
  - 30 TB EMC SAN
- Dell Linux Cluster (2.9TF peak)
  - 600 P4 Processors (2.4 GHz)
  - 600 GB RAM; 40 TB Disk; Myrinet
- Dell Linux Cluster (6TF peak)
  - 4036 Processors (PIII 1.2 GHz)
  - 2TB RAM; 160TB Disk; 16TB SAN
- IBM BladeCenter Cluster (3TF peak)
  - 532 P4 Processors (2.8 GHz)
  - 5TB SAN
- SGI Intel Linux Cluster (0.1TF peak)
  - 150 PIII Processors (1 GHz)
  - Myrinet
- SGI Altix3700 (0.4TF peak)
  - 64 Processors (1.3GHz ITF2)
  - 256 GB RAM
  - 2.5 TB Disk
- Apex Bioinformatics System
  - Sun V880 (3), Sun 6800
  - Sun 280R (2)
  - Intel PIIIs
  - Sun 3960: 7 TB Disk Storage
- HP/Compaq SAN
  - 75 TB Disk; 190 TB Tape
  - 64 Alpha Processors (400 MHz)
  - 32 GB RAM; 400 GB Disk

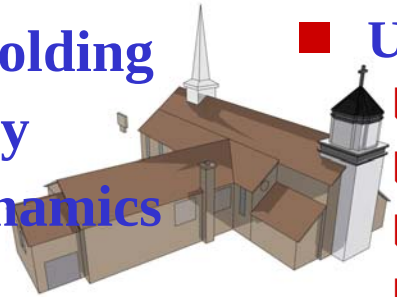
CCR Visualization Resources

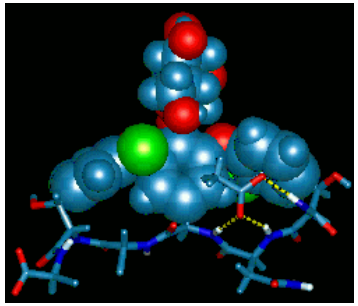
- **Fakespace ImmersaDesk R2**
  - Portable 3D Device
  - Onyx2: 6 R10000 @ 250MHz
  - 2 IR2 Pipes; 3 64MB texture memory mgrs.
- **3D Passive Stereo Display**
  - VisDuo ceiling mounted system
- **Tiled-Display Wall**
  - 20 NEC projectors: 15.7M pixels
  - Screen is 11'×7'
  - Dell PCs with Myrinet2000
- **Access Grid Nodes (2)**
  - Group-to-Group Communication
  - Commodity components
- **SGI Reality Center 3300W**
  - Dual Barco's on 8'×4' screen
  - Onyx300: 10 R14000 @ 500MHz
  - 2 IR4 Pipes; 1 GB texture mem per pipe



CCR Research & Projects (Simulation & Modeling)

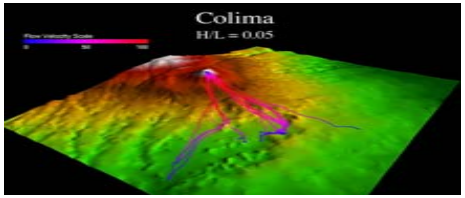
- Archaeology
- Bioinformatics/Protein Folding
- Computational Chemistry
- Computational Fluid Dynamics
- Data Mining/Database
- Earthquake Engineering
- Environ Modeling & Simulation
- Grid Computing
- Molecular Structure Determination
- Physics

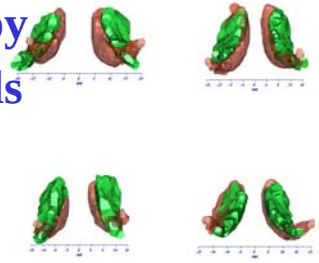




- Videos: MTV
- Urban Simulation and Viz
  - StreetScenes
  - I-90 Toll Barrier
  - Medical Campus
  - Peace Bridge

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





CCR Funding (1998-2006)

- **CCR-Enabled to SUNY-Buffalo**
  - **$170M External Funds**
  - **$140M In-Kind Contributions**
- **CCR-Enabled to WNY**
  - **$200M External Funds**
- **Federal Appropriations**
- **New York State Appropriations**
- **Local WNY Foundations**
- **In-Kind Contributions (Dell, SGI, Sun, etc.)**
- **Grants (NSF, NIH, DOE, etc.)**
- **Projects with Local Companies**
- **Government Projects**
- **SUNY-Buffalo: staff and space**


Sen. Clinton


Sen. Schumer


Rep. Reynolds


Gov. Pataki


M. Capellas


Michael Dell

HEC/CI in the 21st Century

- **Empower students to compete in knowledge-based economy**
- **Embrace digital data-driven society**
- **Accelerate discovery and comprehension**
- **Embrace relationships between a wide variety of organizations**
- **Provide increased Education, Outreach, and Training**
- **Enhance virtual organizations**

HEC/CI in the 21st Century: Implementation

- **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**
- **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)**

Research environment that supports internet-based computation with goal of deriving novel scientific theories and generating knowlege; Core of modern simulation and modeling; Provides entirely new methods of investigation

Real-Time Visualization

StreetScenes: Real-Time 3D Traffic Simulation

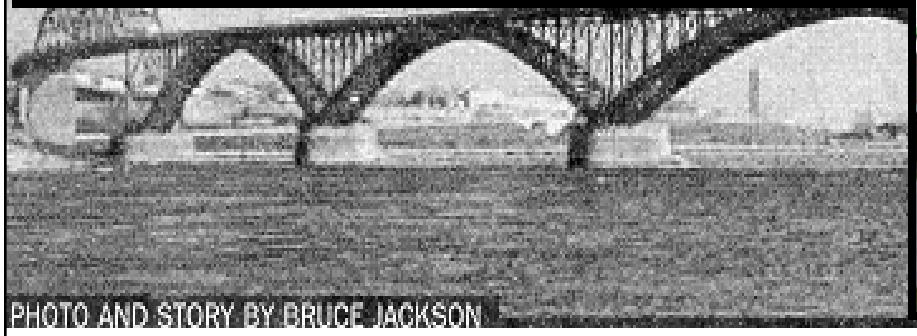
- Accurate local landmarks: Bridges, Street Signs, Business, Homes
- Can be viewed from driver's perspective
- Real-Time Navigation
- Works with
  - Corsim
  - Synchro
- Generate AVI & MOV
- Multiple Simultaneous
  - Traffic Loads
  - Simulation
  - Varying POV

Animation & Simulation

Rendered Scenes

Peace Bridge Visualization: Animation & Simulation

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



MTV

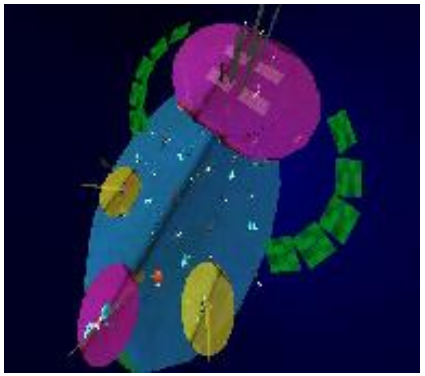
IBC Digital & CCR

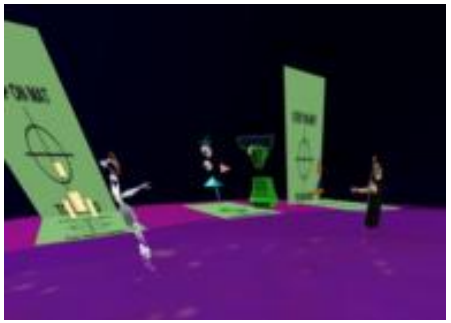
Song: I'm OK (I Promise)
Band: Chemical Romance
Gaming Environment: Death Jr.

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

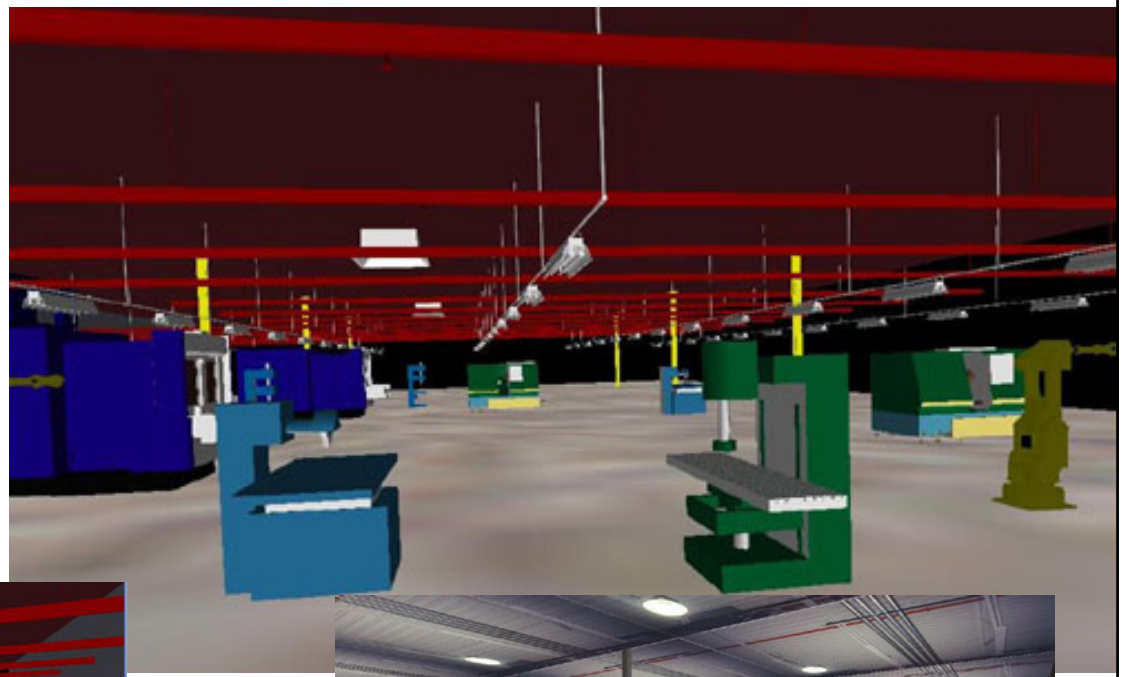


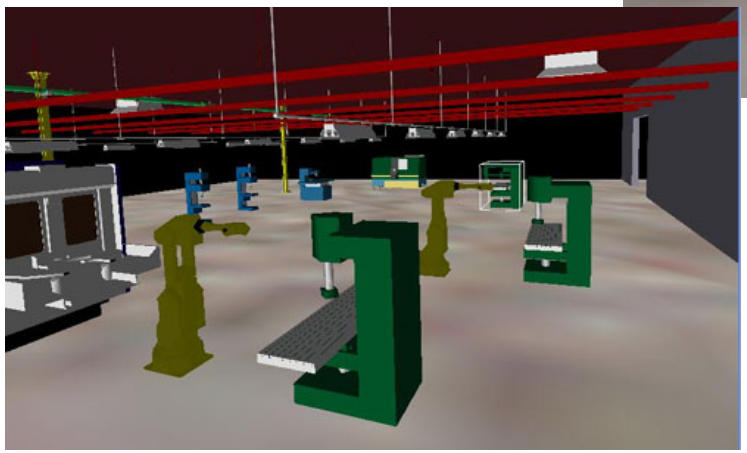


J. Anstey

VR-Fact!

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





Kesh

Collaborative Visualization Environments

- Enable distributed collaboration via software developed at CCR
- Enable visualization and interaction with data across a geographically disparate network topology
- Integrate multiple data sources:
  - Scientific
  - Multimedia
- Research Topics
  - Distributed databases
  - OpenGL 3D programming
  - 3D Modeling
  - Character animation
  - User interaction
  - Virtual Reality

A. Ghadersohi, R. Miller, M. Green

Western New York

Some Facts

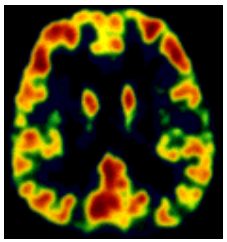
Buffalo, New York

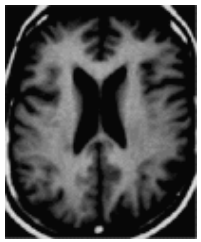
- The Queen City: 2nd Largest City in NYS
- City of Lights
  - First U.S. city to have electric street lights
  - Pan American Exposition (1901)
    - Pres. McKinley Shot

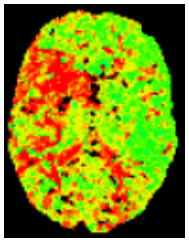
- Architecture
  - Frederick Law Olmsted
  - Frank Lloyd Wright
- Underground Railroad
  - Slaves escaped to freedom in Canada
- Four straight Super Bowl appearances
- Culinary Delights
  - Beef on Weck, Pizza, Fish Fries
  - (Buffalo) Wings: Anchor Bar, 1964
- Health Problems
  - Heart Disease/Stroke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

Recent Biomedical Advances (Buffalo, NY)

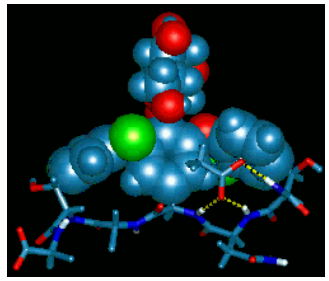
- PSA Test (screen for Prostate Cancer)
- Avonex: Interferon Treatment for Multiple Sclerosis
- Artificial Blood
- Nicorette Gum
- Fetal Viability Test
- Edible Vaccine for Hepatitis C
- Timed-Release Insulin Therapy
- Anti-Arrythmia Therapy
  - Tarantula venom







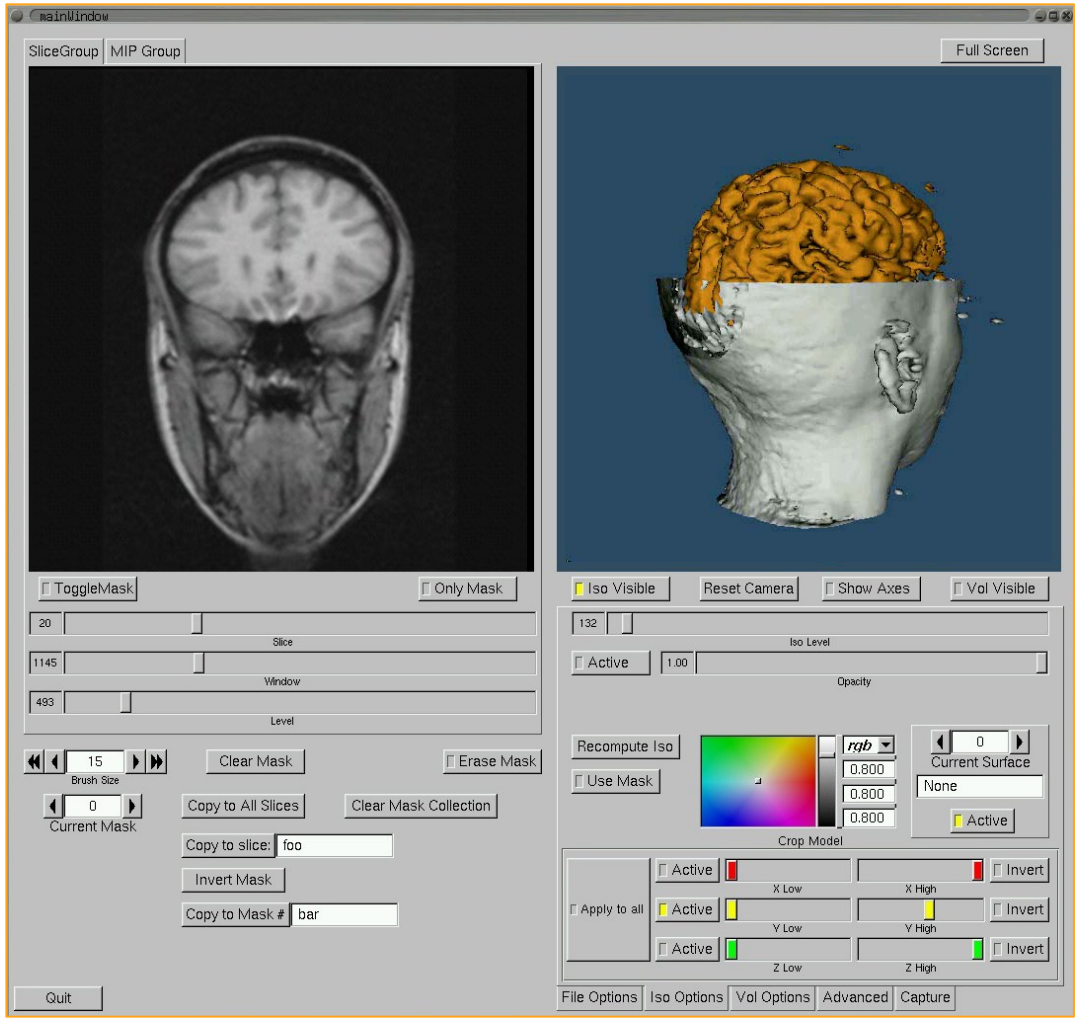
- Direct Methods Structure Determination
  - Listed on "Top Ten Algorithms of the 20th Century"
- Vancomycin
- Gramacidin A
- High Throughput Crystallization Method: Patented
- NIH National Genomics Center: Northeast Consortium
- Howard Hughes Medical Institute: Center for Genomics & Proteomics



Scientific Visualization

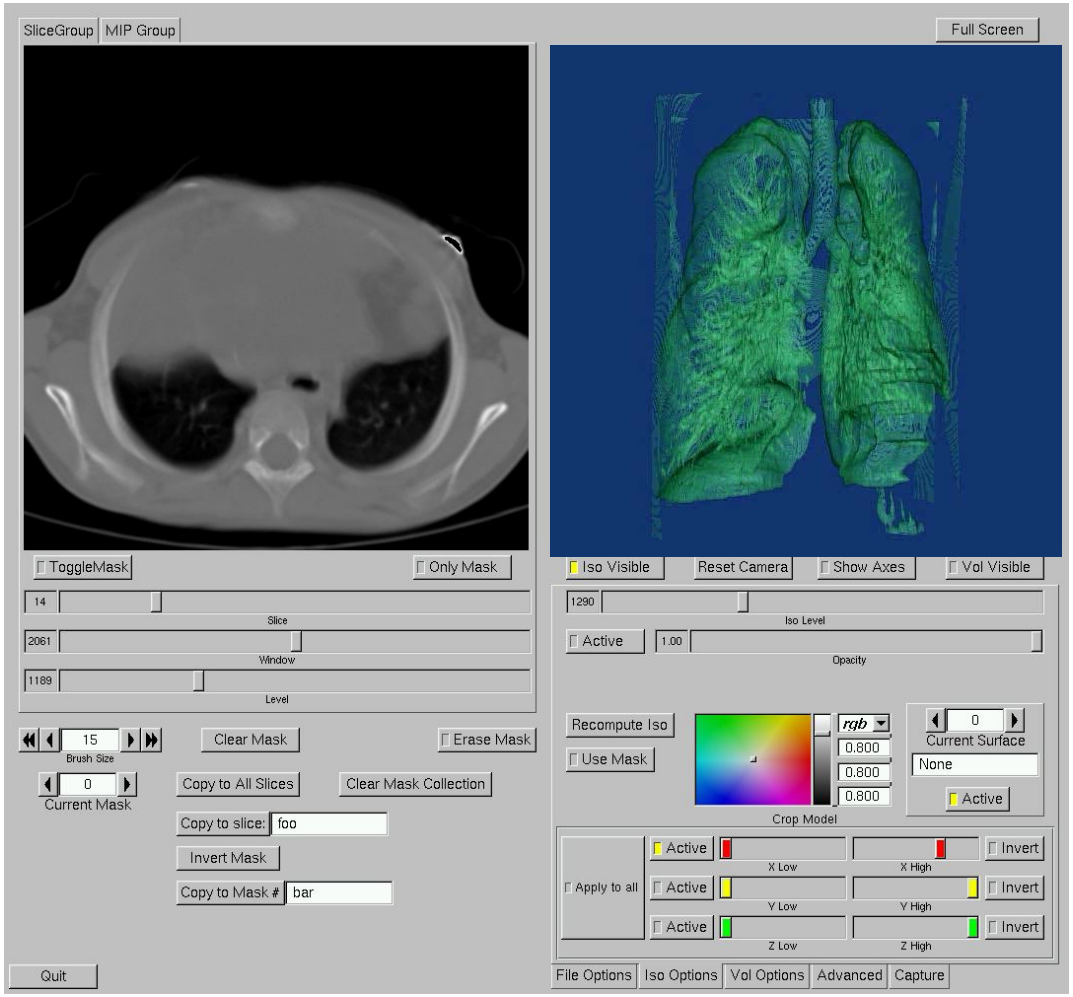
Multiple Sclerosis Project

- **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



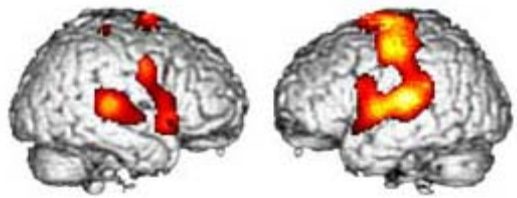
M. Innus

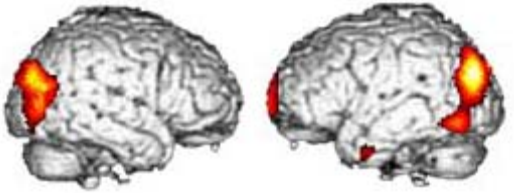
Mapping Brain Activity

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

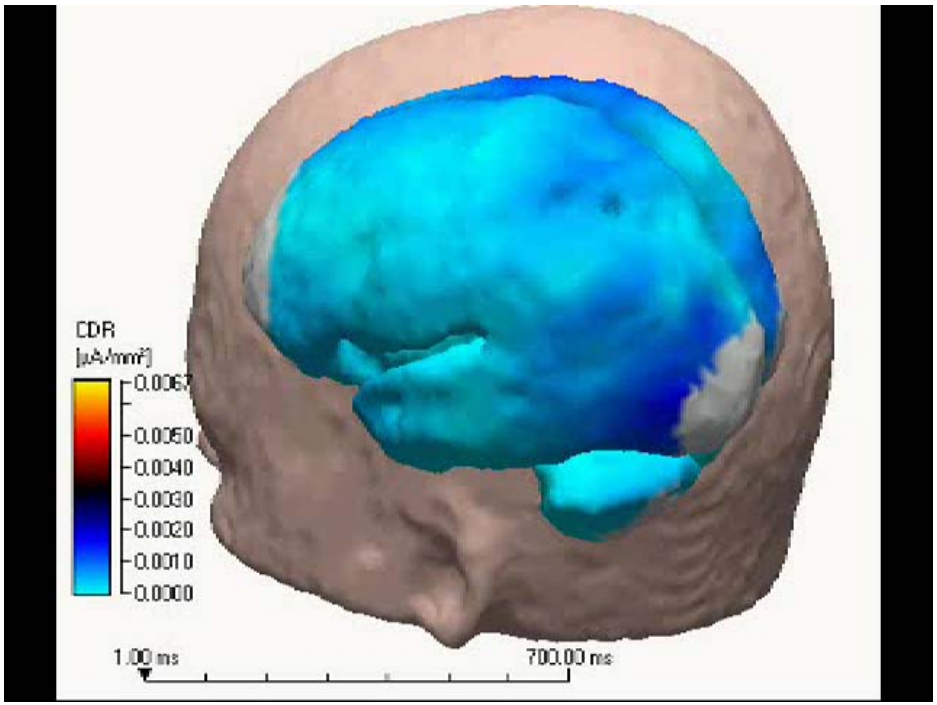
Sites Activated

Sites Deactivated





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



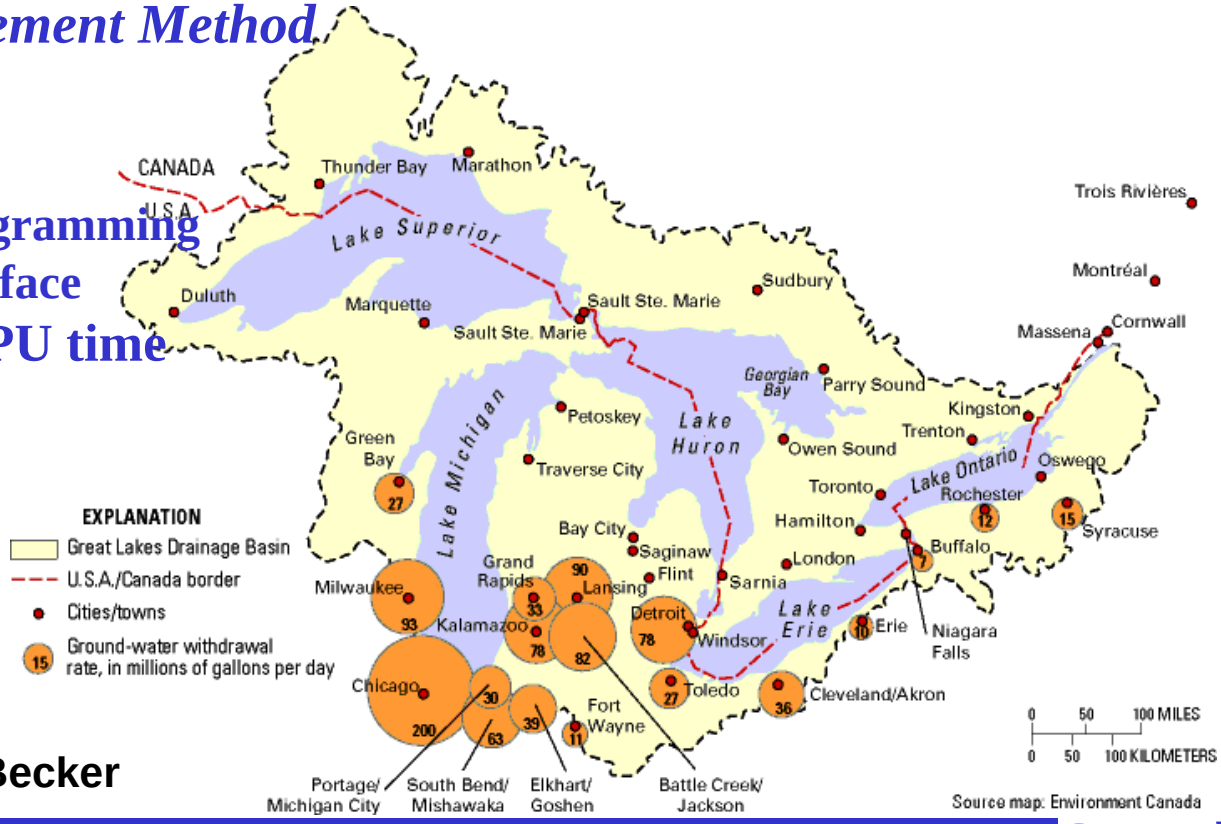
A. Lockwood

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**

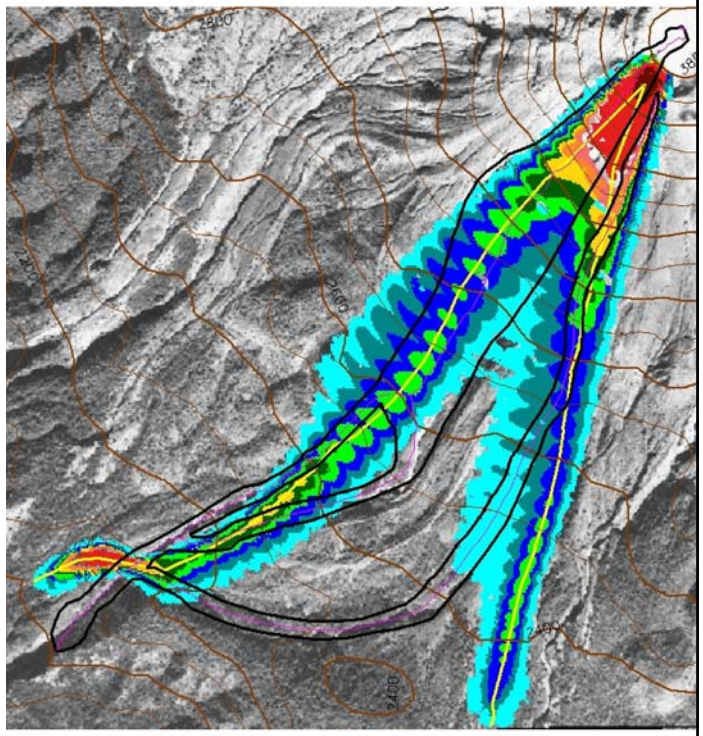


A. Rabideau, I. Jankovic, M. Becker

Avalanches, Volcanic and Mud Flows

Geology, Math, 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**



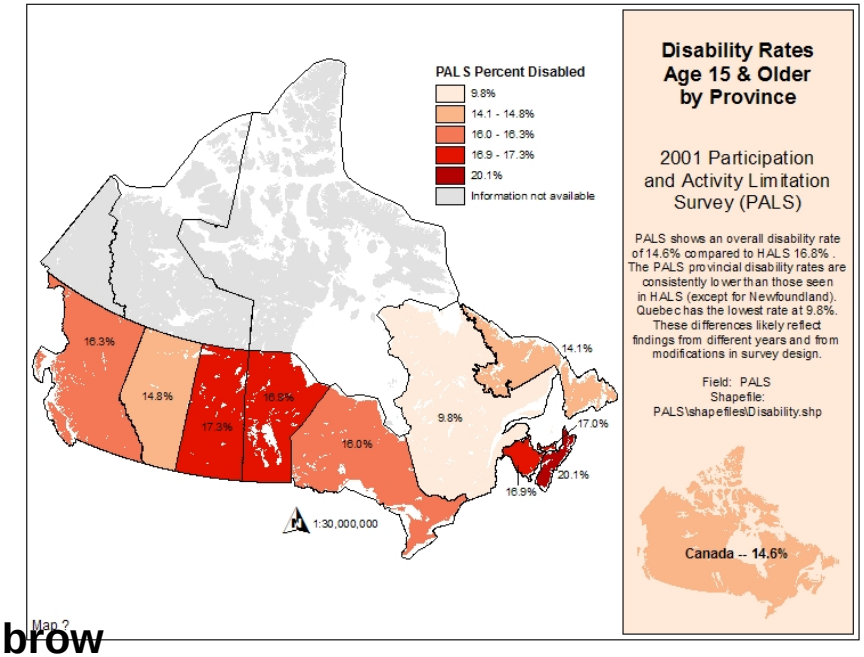
Flow models of Colima volcano
In Mexico – courtesy Rupp et. al.'06

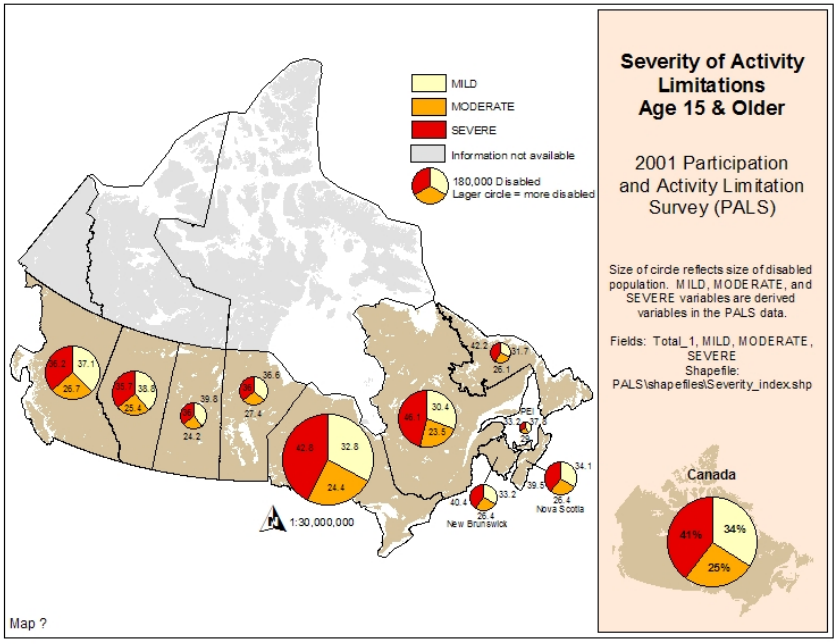
A. Patra, B. Pitman, M. Sheridan, M. Jones

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.

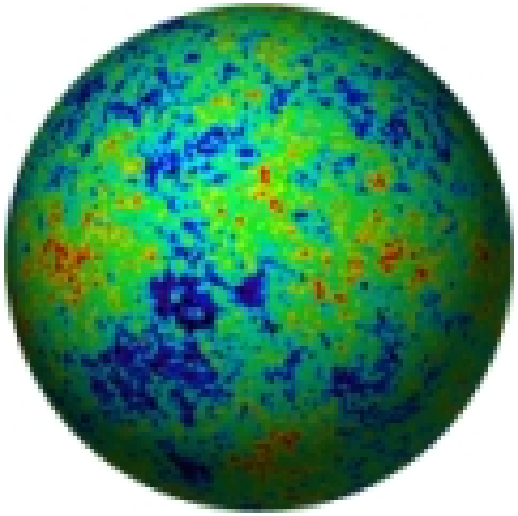


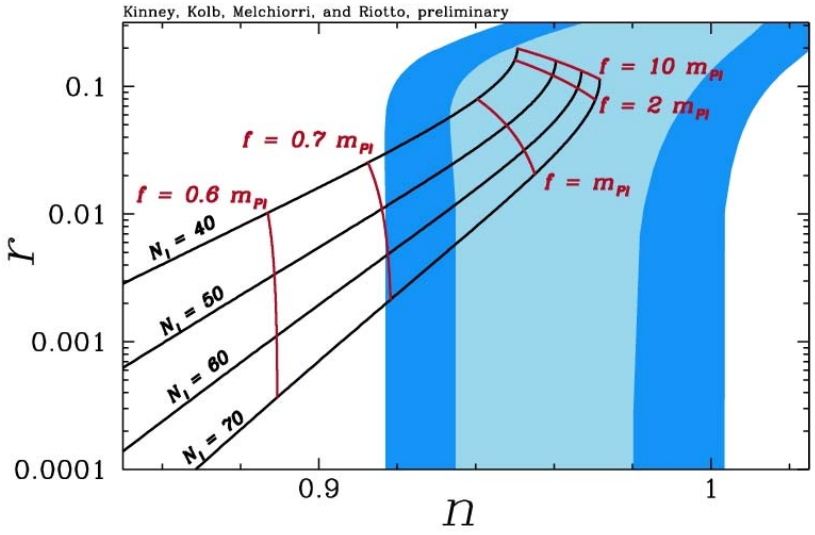


E. Zubrow

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***



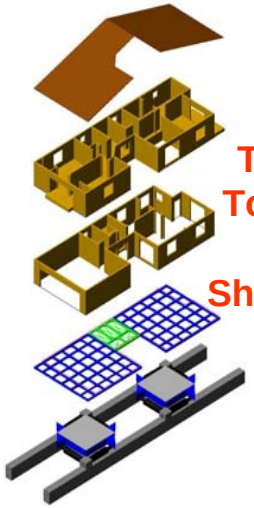


W. Kinney

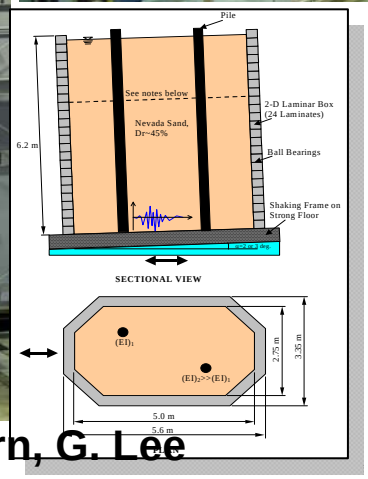
UB's Structural Engineering and Earthquake Simulation Laboratory (SEESL)

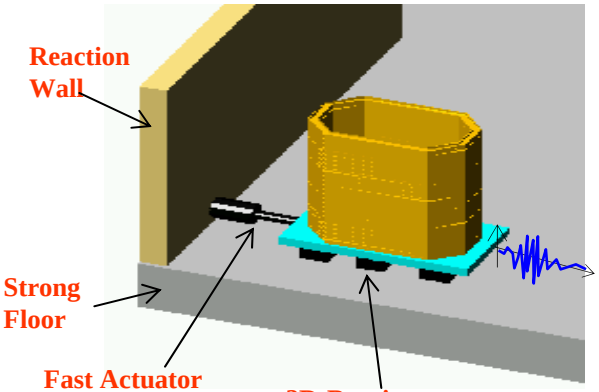
Structural Engineering

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



Two-story Townhouse on Twin Shake Tables



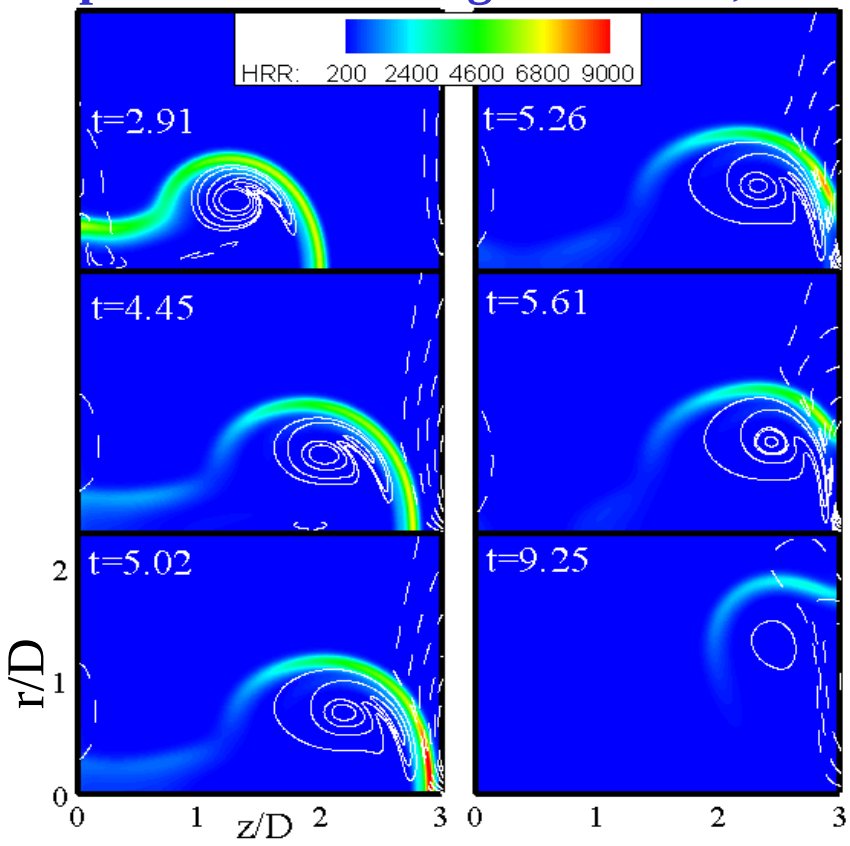


2-D Geotechnical Laminar Box Tests of Pile Foundations Subjected to Soil Liquefaction

M. Bruneau, A. Reinhorn, G. Lee

Understanding Combustion

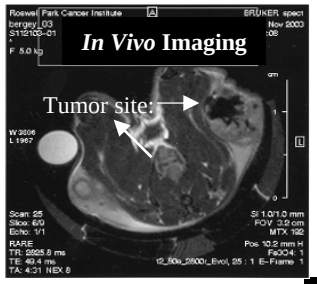
- Flame-wall interaction modeling for a non-premixed flame propelled by a vortex ring.
- In this figure different time instants are shown during the interaction. White line contours and color contours represent vortex ring and flame, respectively.
- Key Features:
  - Modeling of Detailed GRI3. Mechanism 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



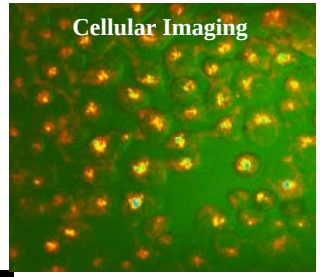
Institute for Lasers, Photonics, and Biophotonics

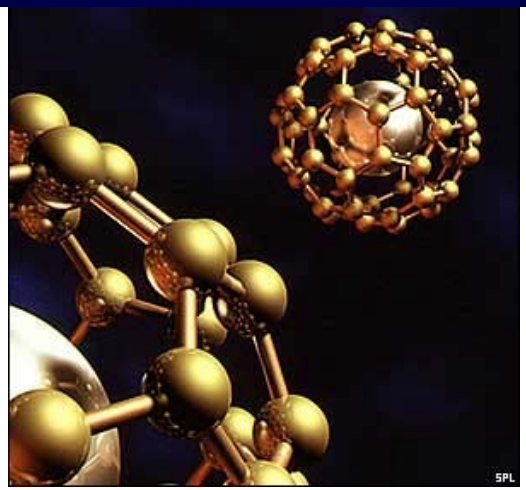
Nanomedicine Program

World class Research Program Melding Nanotechnology with Biomedical Sciences



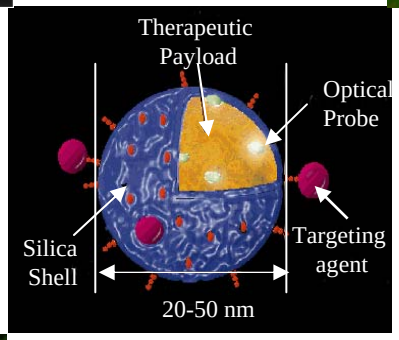
Multi-Modal Imaging



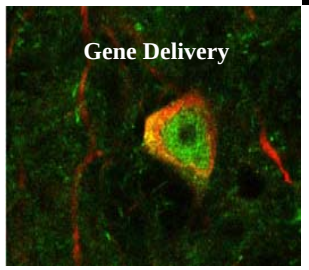


Building from the Bottom Up

BioCompatibility/ Distribution



In Vivo Sensing



Targeted Therapy

State of the Art Molecular Imaging and Nanocharacterization Facilities

- Multiphoton Laser Scanning System
- Confocal Imaging including FRET, FLIM & FRAP analysis
- Coherent Anti-Stokes Raman Imaging
- Optical Trapping/Dissection
- Advanced Laser Systems

P. Prasad

www.biophotonics.buffalo.edu

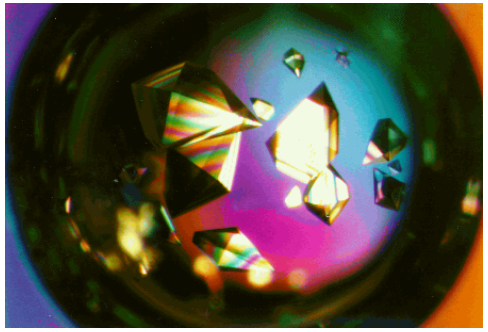
"Leading the Way to Technology through Innovation"

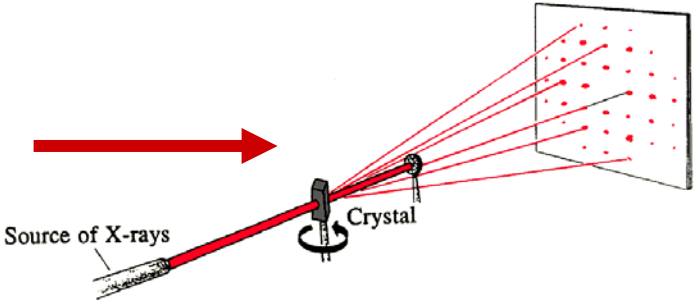
Shake-and-Bake

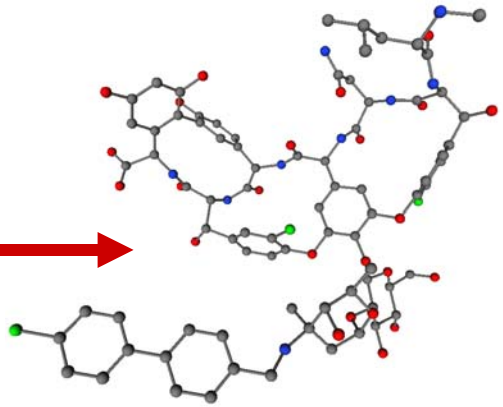
Molecular Structure Determination from X-Ray Crystallographic Data

Molecular Structure Determination via *Shake-and-Bake*

- ***SnB* Software by UB/HWI**
  - **IEEE "Top Algorithms of the Century"**
- **Worldwide Utilization**
- **Critical Step**
  - **Rational Drug Design**
  - **Structural Biology**
  - **Systems Biology**
- **Vancomycin**
  - **"Antibiotic of Last Resort"**
- **Current Efforts**
  - **Grid**
  - **Collaboratory**
  - **Intelligent Learning**



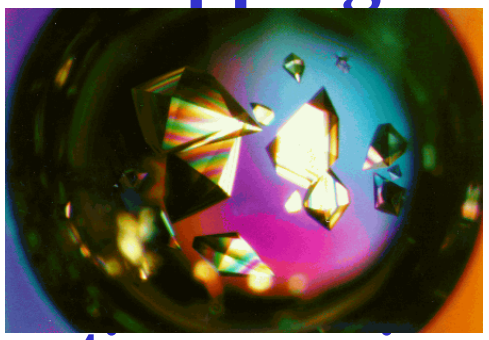




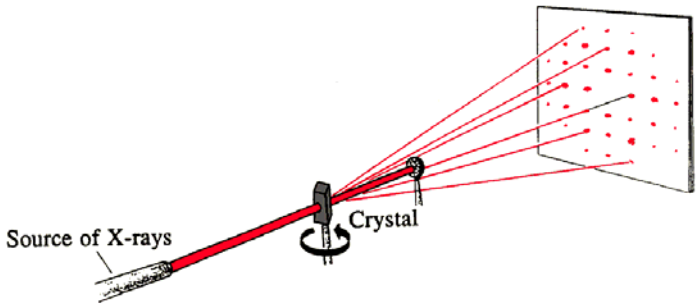
1. Isolate a single crystal 2. Perform the X-Ray diffraction experiment 3. Determine the crystal structure

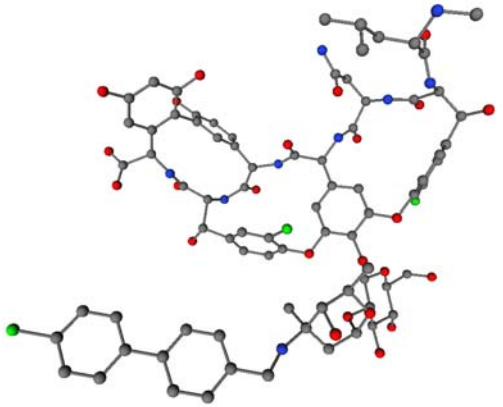
X-Ray Crystallography

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



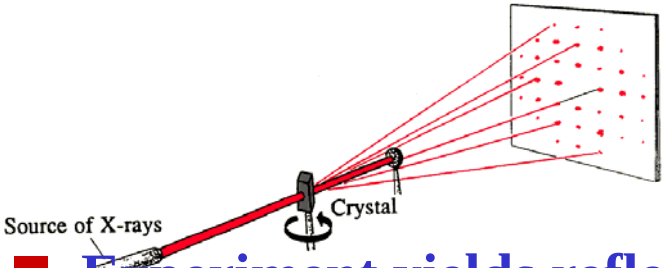
- **Procedure:**
  1. **Isolate a single crystal.**
  2. **Perform the X-Ray diffraction experiment.**



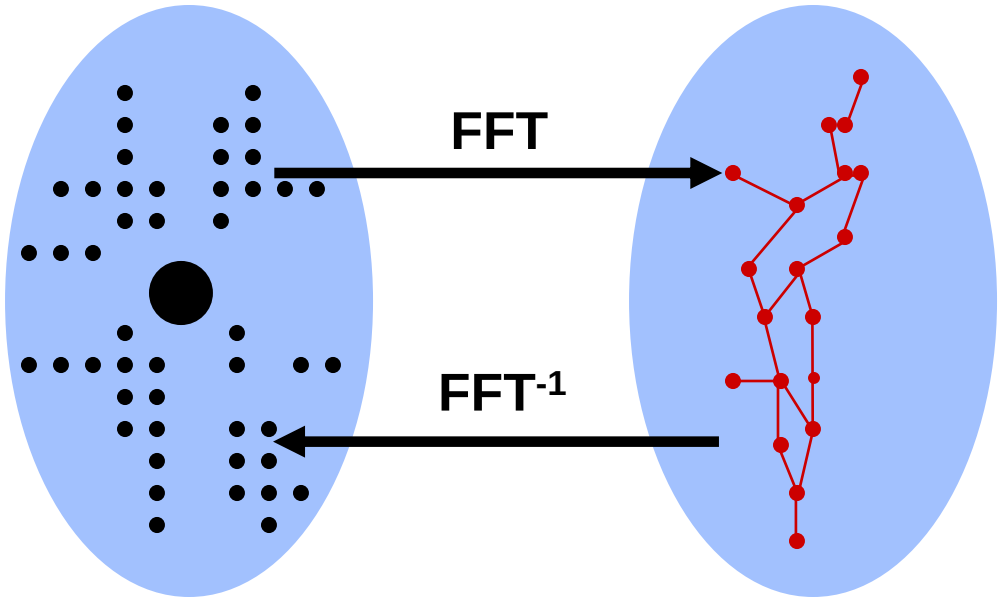


  3. **Determine molecular structure that agrees with diffration data.**

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.**



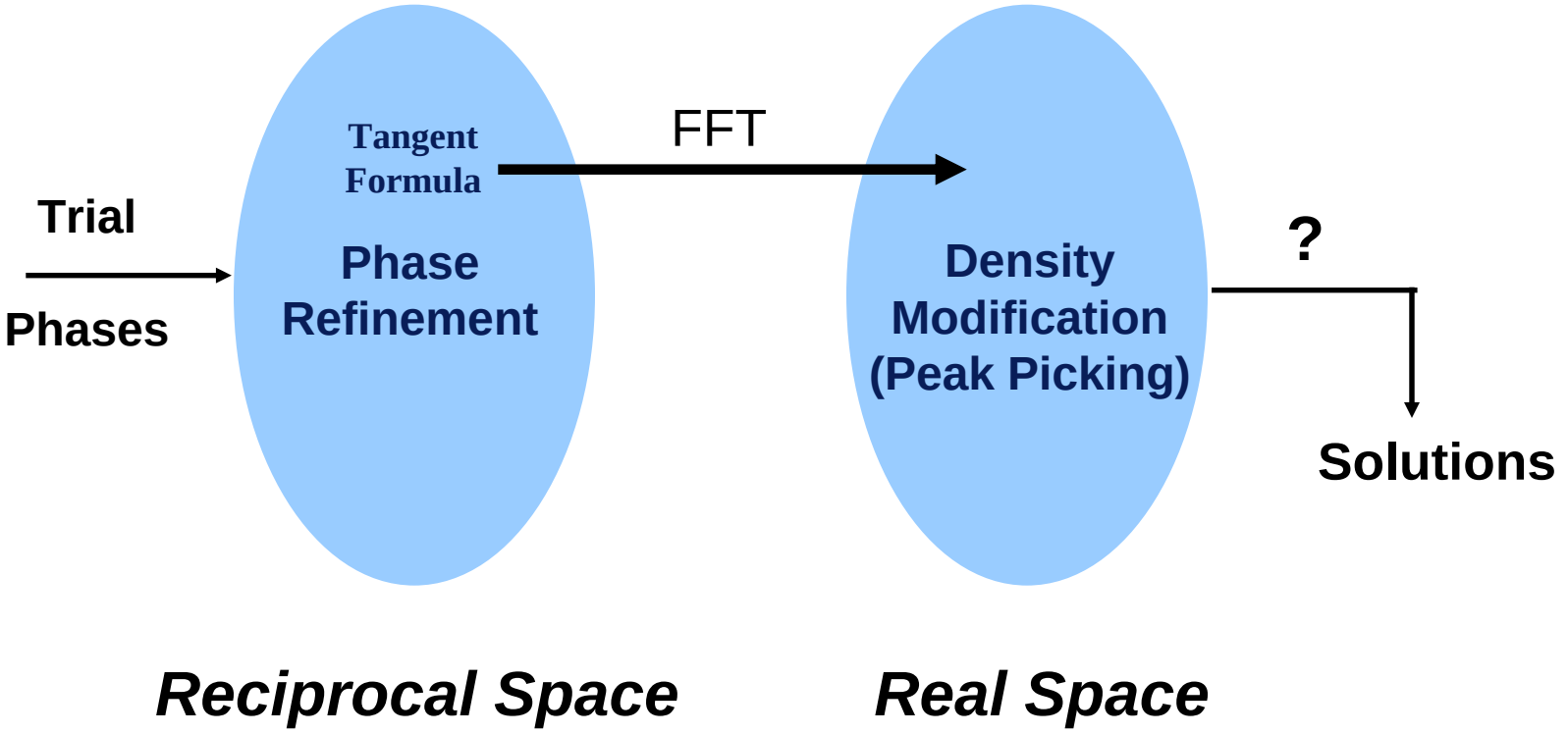
The Phase Problem

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

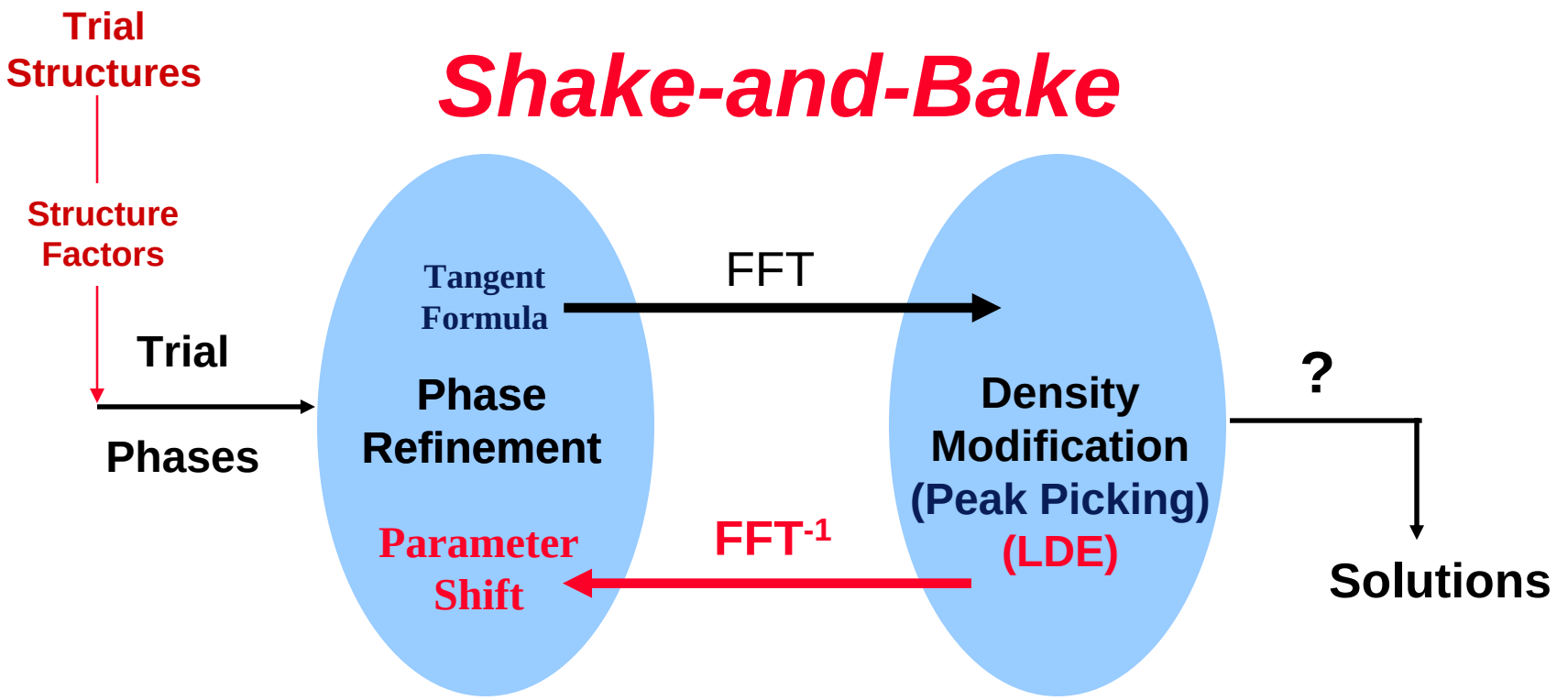
Overview of Direct Methods

- **Probability theory gives information about certain linear combinations of phases.**
  - **In particular, the triples $\phi_H + \phi_K + \phi_{-H-K} = 0$ with high probability.**
- **Probabilistic estimates are expressed in terms of normalized structure factor magnitudes ($|E|$).**
- **Optimization methods are used to extract the values of individual phases.**
- **A multiple trial approach is used during the optimization process.**
- **A suitable figure-of-merit is used to determine the trials that represent solutions.**

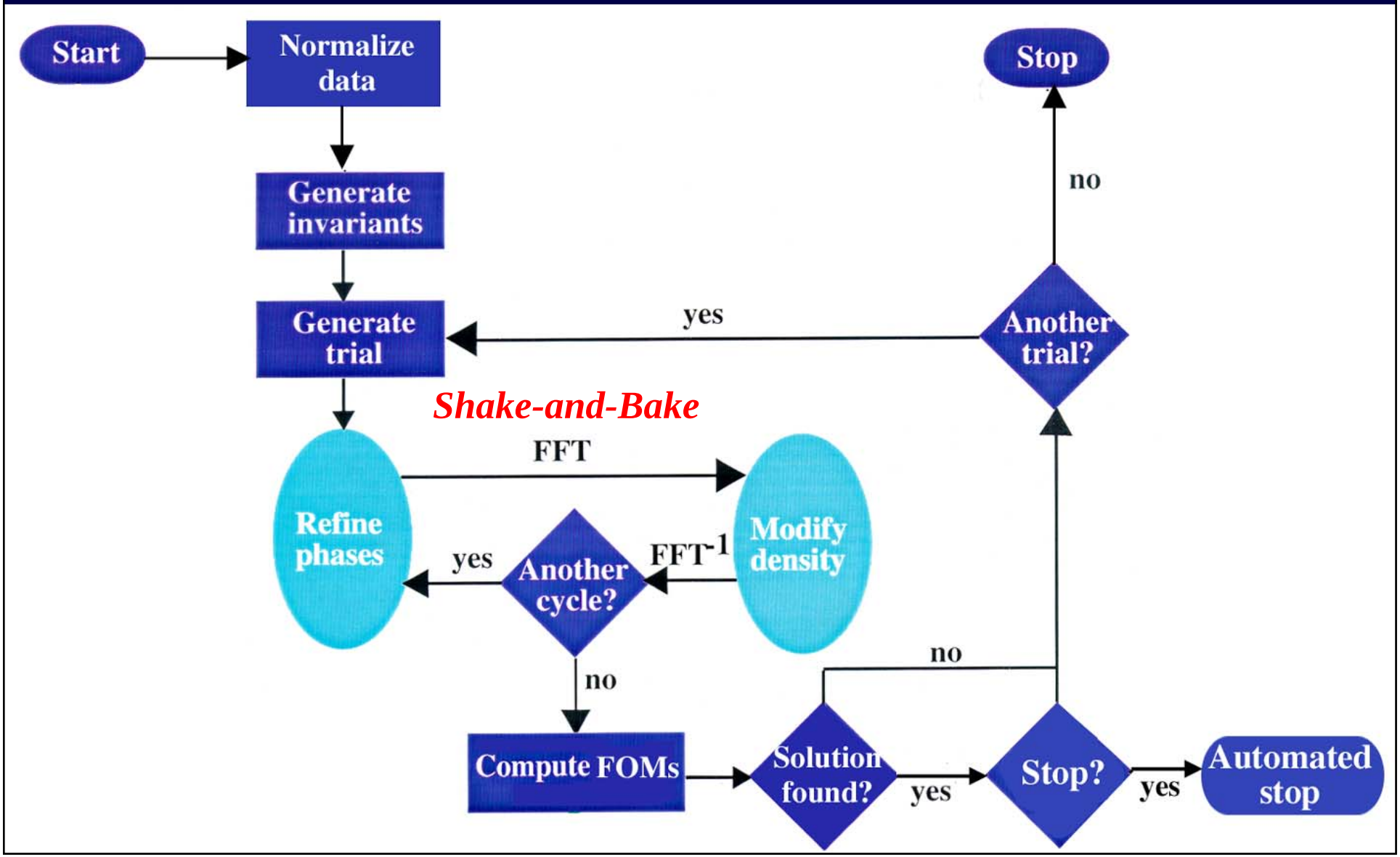
Conventional Direct Methods



Shake-and-Bake Method: Dual-Space Refinement



A Direct Methods Flowchart



Generate Triplet Invariants

Reflections

Rank	h	k	l	|E|
1	0	3	4	4.65
2	0	7	30	3.67
3	5	1	1	3.67
4	8	8	5	3.26
5	6	0	1	3.15
⋮	⋮	⋮	⋮	⋮
10n=840	7	0	3	1.33
841	2	4	30	1.33

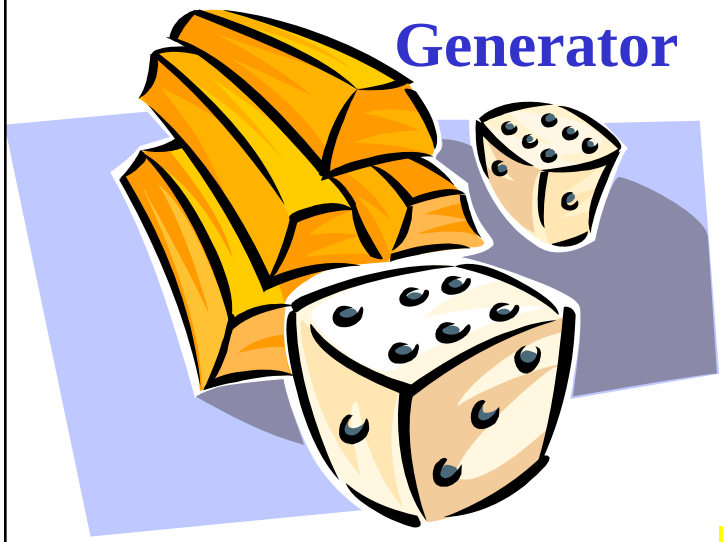
Triplets

Rank	H	K	-H-K	A
1	1	4	45	3.90
2	1	3	165	3.52
3	3	5	17	3.37
4	1	3	289	3.16
5	1	28	40	3.09
⋮	⋮	⋮	⋮	⋮
100n=840	19	259	734	0.71
8401	142	179	283	0.71

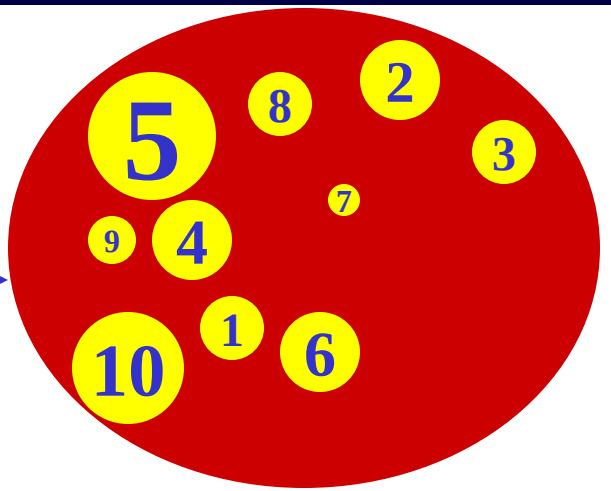
n = 84 unique atoms

Getting Started: Random Atoms

Random Number Generator



***n* = 10 atoms**
(30 coordinates)



Structure Factor Calculation

ϕ_1 ϕ_2
ϕ_3 ϕ_4
ϕ_5 ϕ_6
ϕ_7 ϕ_8
ϕ_9 ϕ_{10}

Useful Relationships for Multiple Trial Phasing

Tangent Formula

$$\tan\phi_H = \frac{-\sum_K |E_K E_{-H-K}| \sin(\phi_K + \phi_{-H-K})}{\sum_K |E_K E_{-H-K}| \cos(\phi_K + \phi_{-H-K})}$$

Parameter Shift Optimization

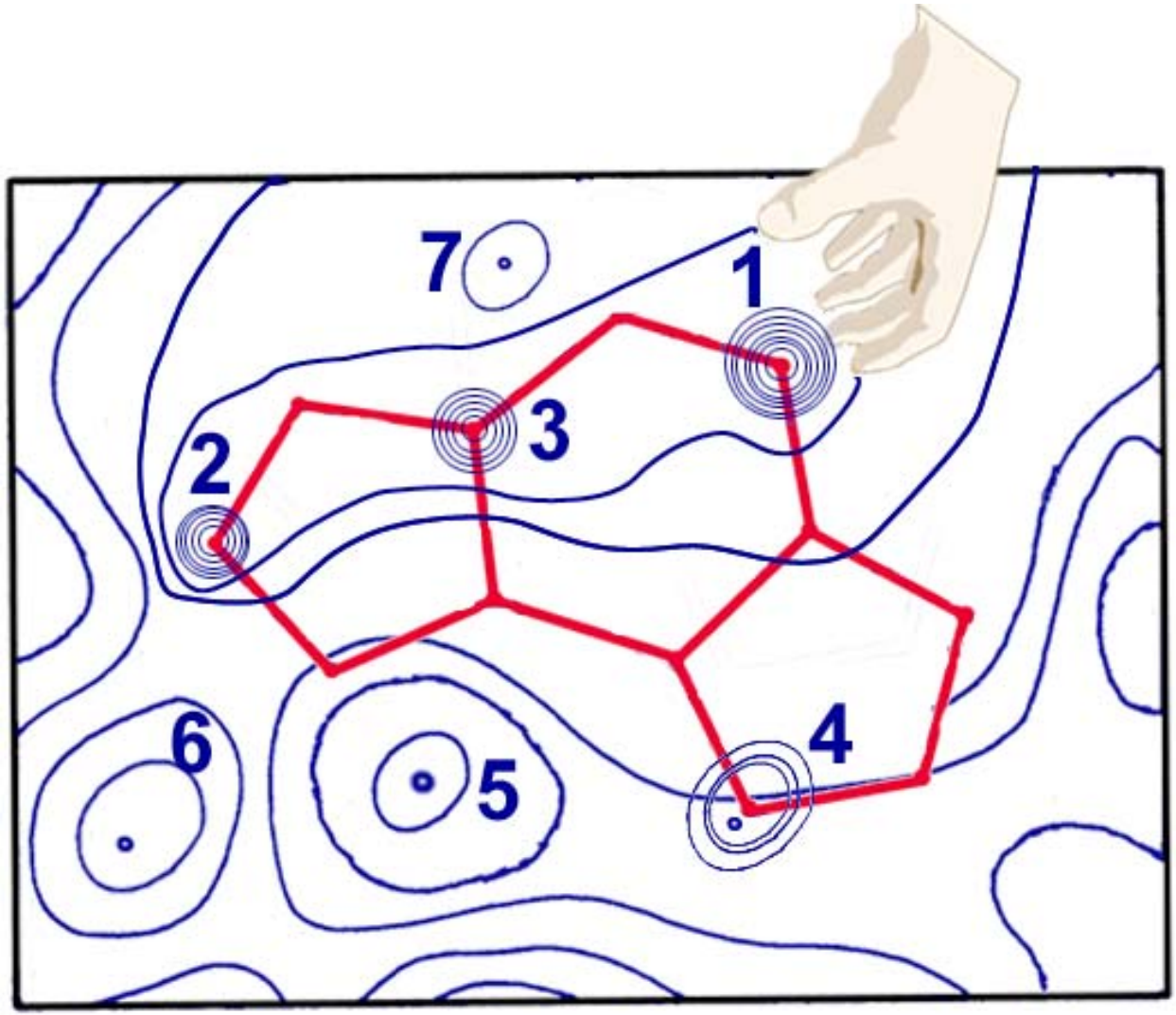
$$R(\phi) = \frac{1}{\sum_{H,K} W_{HK}} \sum_{H,K} W_{HK} \left(\cos\Phi_{HK} - \frac{I_1(W_{HK})}{I_0(W_{HK})} \right)^2$$

where $|E_H| \propto |F_H|$ normalized in resolution shells

Invariants: $\Phi_{HK} = \phi_H + \phi_K + \phi_{-H-K} \approx 0$

Weights: $W_{HK} = A_{HK} = 2N^{-1/2} |E_H E_K E_{-H-K}|$

Peak Picking



Scoring Trial Structures: *SnB* FOMs

1. The minimal function (R(Φ) or Rmin)

2. $R_{cryst} = \sum \big| \, | E_o | - k | E_c | \, \big| / \sum | E_o |$

   where the scale factor $k = \sum | E_o | / \sum | E_c |$

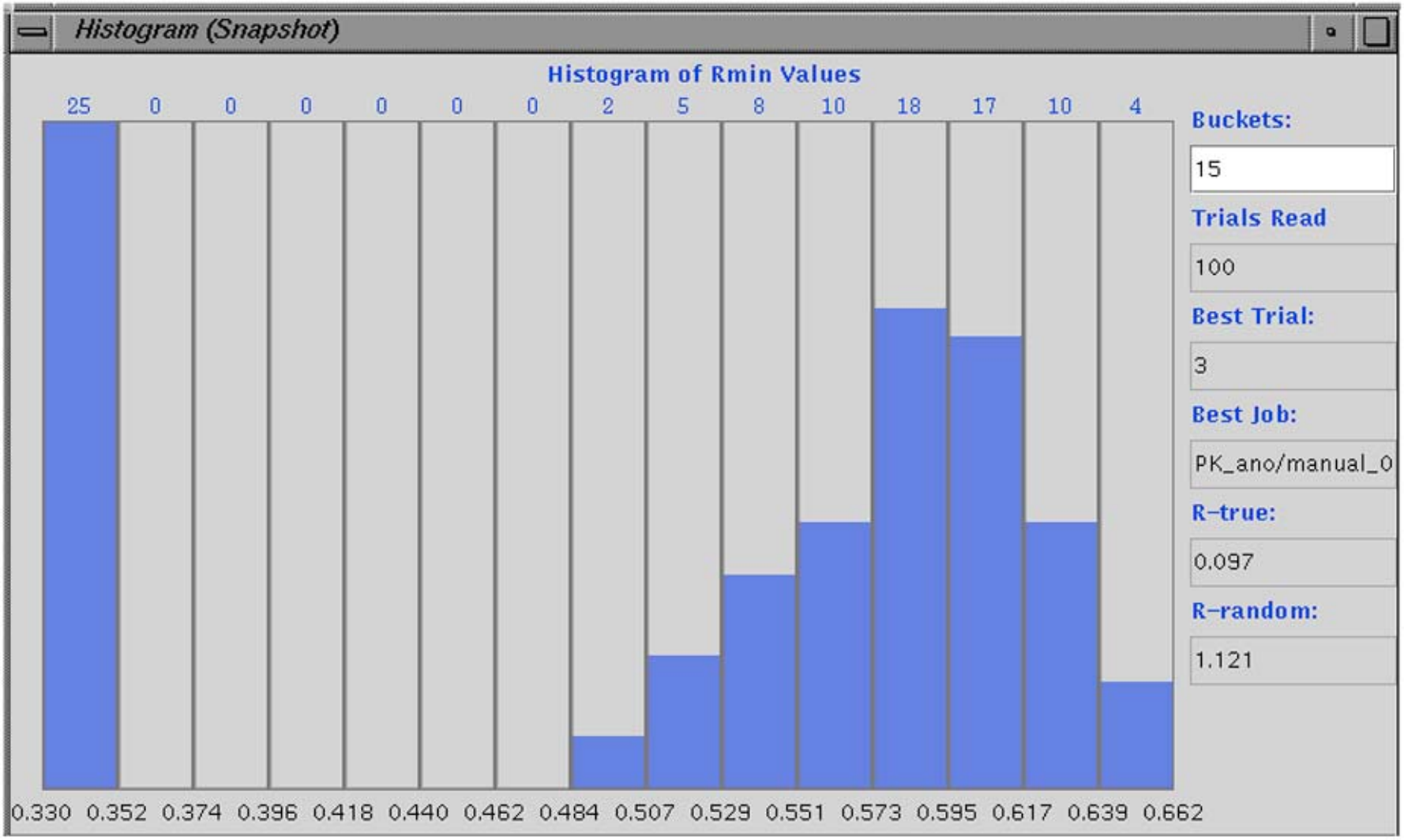
3. Correlation Coefficient (CC)

$$CC = [\sum w E_o^2 E_c^2 \sum w - \sum w E_o^2 \sum w E_c^2] /$$

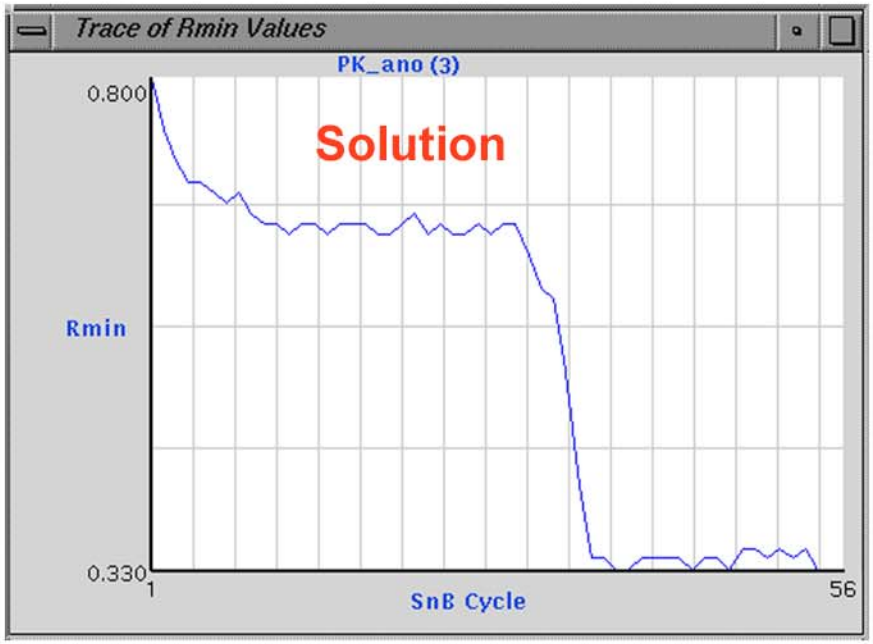
$$\{[\sum w E_o^4 \sum w - (\sum w E_o^2)^2][\sum w E_c^4 \sum w - (\sum w E_c^2)^2]\}^{1/2}$$

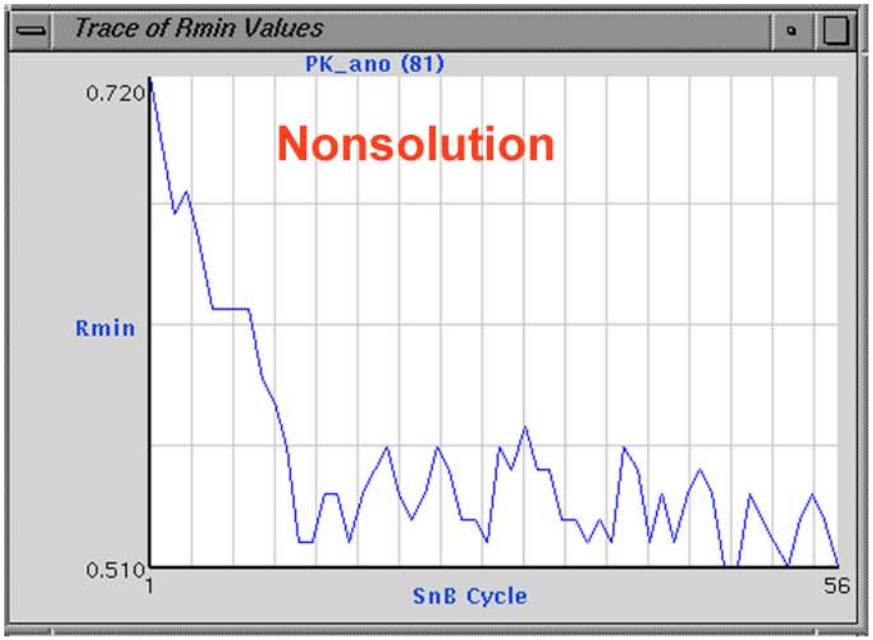
where weights $w = 1/[0.04 + \sigma^2(E_o)]$

Ph8755: *SnB* Histogram

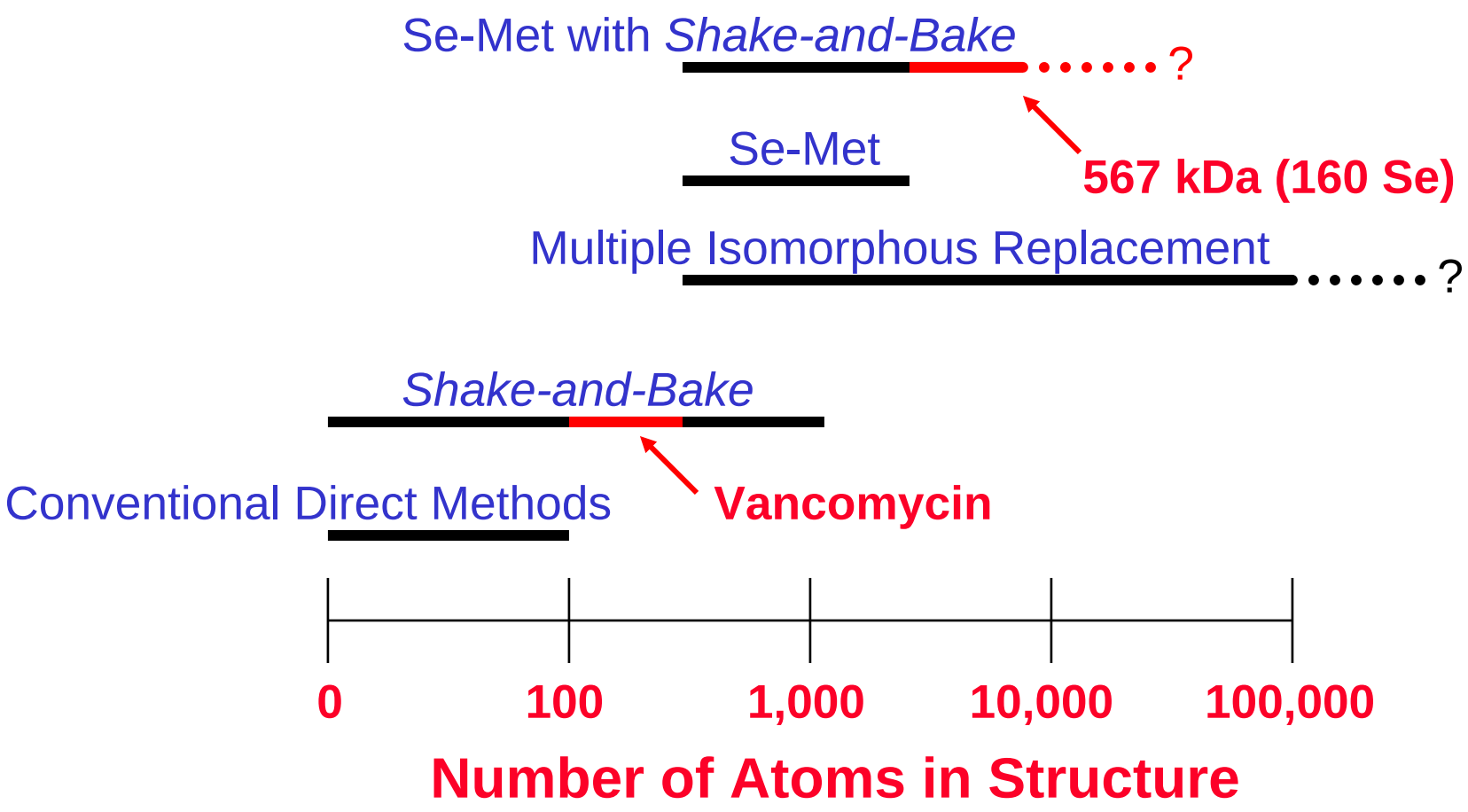


Minimal Function Traces





Phasing and Structure Size

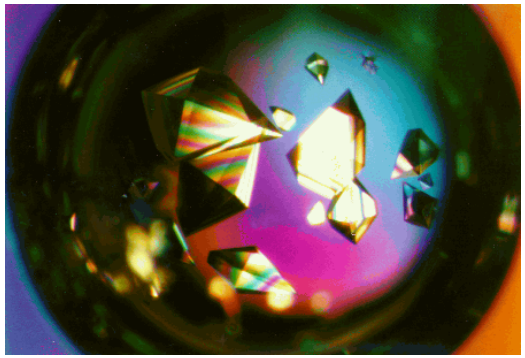


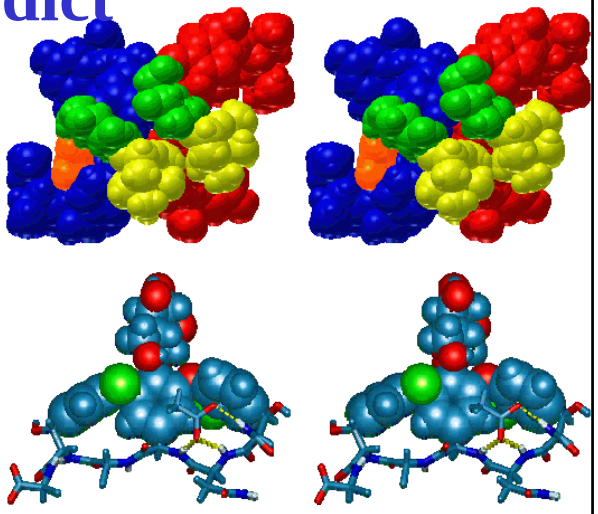
Shake-and-Bake Applications: Structure Size and Data Resolution

- **Basic Data (Full Structure)**
  - ~750 unique non-H atoms (equal)
  - ~2000 such atoms including 8 Fe's
  - 1.1-1.2Å data (equal atom)
  - 1.3-1.4Å data (unequal atoms, sometimes)
- **SAS or SIR Difference Data (substructures)**
  - 160 Se (567 kDa / ASU)
  - 3-4Å data
  - 5Å truncated data have also worked

Vancomycin

- **Interferes with formation of bacterial walls**
- ***Last line of defense* against deadly**
  - **streptococcal and staphylococcal bacteria strains**
- **Vancomycin resistance exists (Michigan)**
- **Can't just synthesize variants and test**
- **Need structure-based approach to predict**
- **Solution with *SnB* (*Shake-and-Bake*)**
  - **Pat Loll**
  - **George Sheldrick**





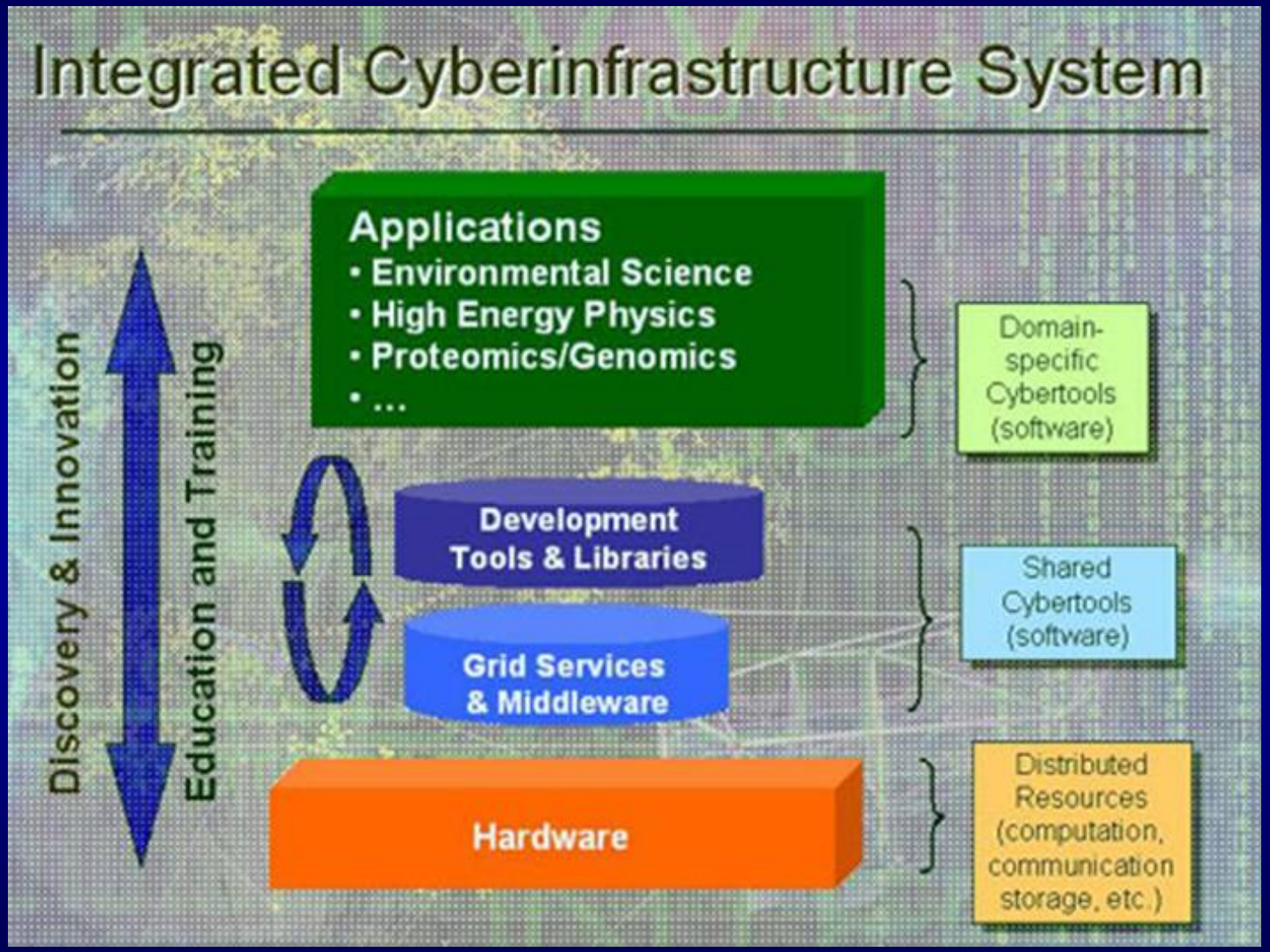
Miller's Cyberinfrastructure Laboratory (MCIL)

Tiled-Display Wall in CCR

MCIL Overview

- **"Cyberinfrastructure (CI) is a comprehensive phenomenon that involves creation, dissemination, preservation, and application of knowledge" (NSF)**
- **Working Philosophy**
  - **CI sits at core of modern simulation & modeling**
  - **CI allows for new methods of investigation to address previously unsolvable problems**
- **Focus of MCIL is on development of *algorithms, portals, interfaces, middleware***
- **Goal of MCIL is to free end-users to do disciplinary work**
- **Funding (2001-pres)**
  - **NSF: ITR, CRI, MRI**
  - **NYS appropriations**
  - **Federal appropriations**

NSF Integrated Cyberinfrastructure

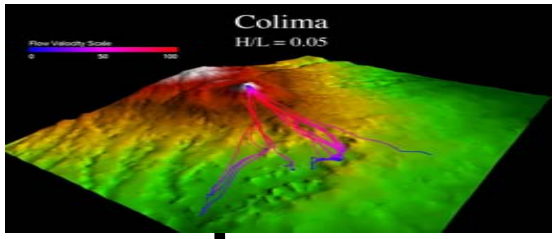


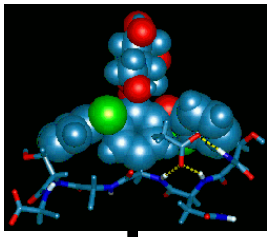
NSF Director Arden L. Bement: "leadership in cyberinfrastructure may determine America's continued ability to innovate – and thus our ability to compete successfully in the global arena."

MCIL Equipment (57.5 TF; 37 TB)

- **Production Equipment**
  - **Magic (50+ Tops; 12,000+ cores: Largest Compute System in WNY, NYS Grid, OSG – based on GPUs)**
    - **Dell Intel Head Node; Dell Intel Worker Nodes; 13 NVIDIA Tesla S1070s, Dell 15 TB Storage**
  - **Dell Workstations**
- **Experimental Equipment**
  - **Clusters**
    - **Head Nodes: Dell 1950s (Intel Dual Core Xeon)**
    - **Workers: Dell Intel 2×4s, Intel 1×2s, & AMD 2×2s**
    - **NVIDIA S870s**
  - **Virtual Memory Machines (2 × Dell Intel 4×4)**
  - **Dell GigE Managed Switches; InfiniBand Switches**
  - **22 TB Dell Storage (2)**
  - **Condor Flock (35 Intel/AMD)**

Grid Computing Tutorial



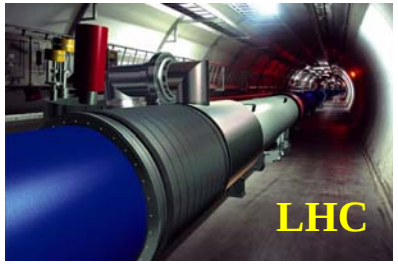


Data Acquisition

Advanced Visualization

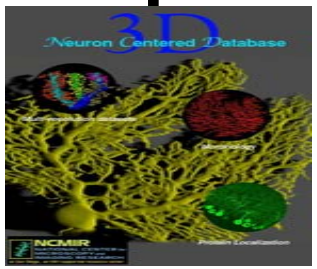
Analysis

Imaging Instruments



LHC

Computational Resources



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

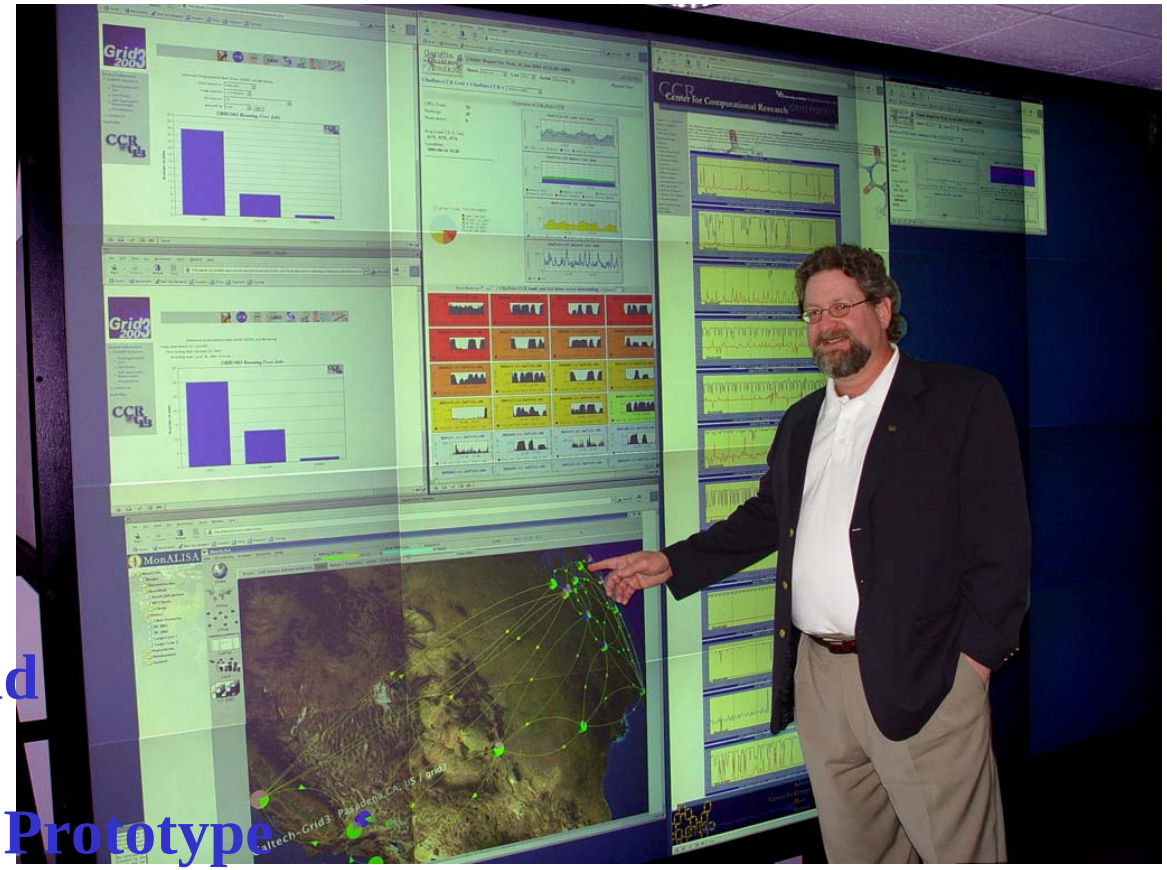
- TeraGrid (NSF)
  - Integrates High-End Resources
  - High-Performance (Dedicated) Networks
  - 11 Sites; 1.2PF, 4PB Disk, 30PB Tape
  - 100+ Databases Available
- OSG (DOE, NSF)
  - High-Throughput Distributed Facility
  - Open & Heterogeneous
  - Biology, Computer Science, Astrophysics, LHC
  - 57 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!

Evolution of MCIL Lab Projects

- **Buffalo-Based 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**
- **Western New York Grid**
  - **Heterogeneous System: Hardware, Networking, Utilization**
  - **Buffalo, Geneseo, Hauptman-Woodward, Niagara**
- **New York State 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}**

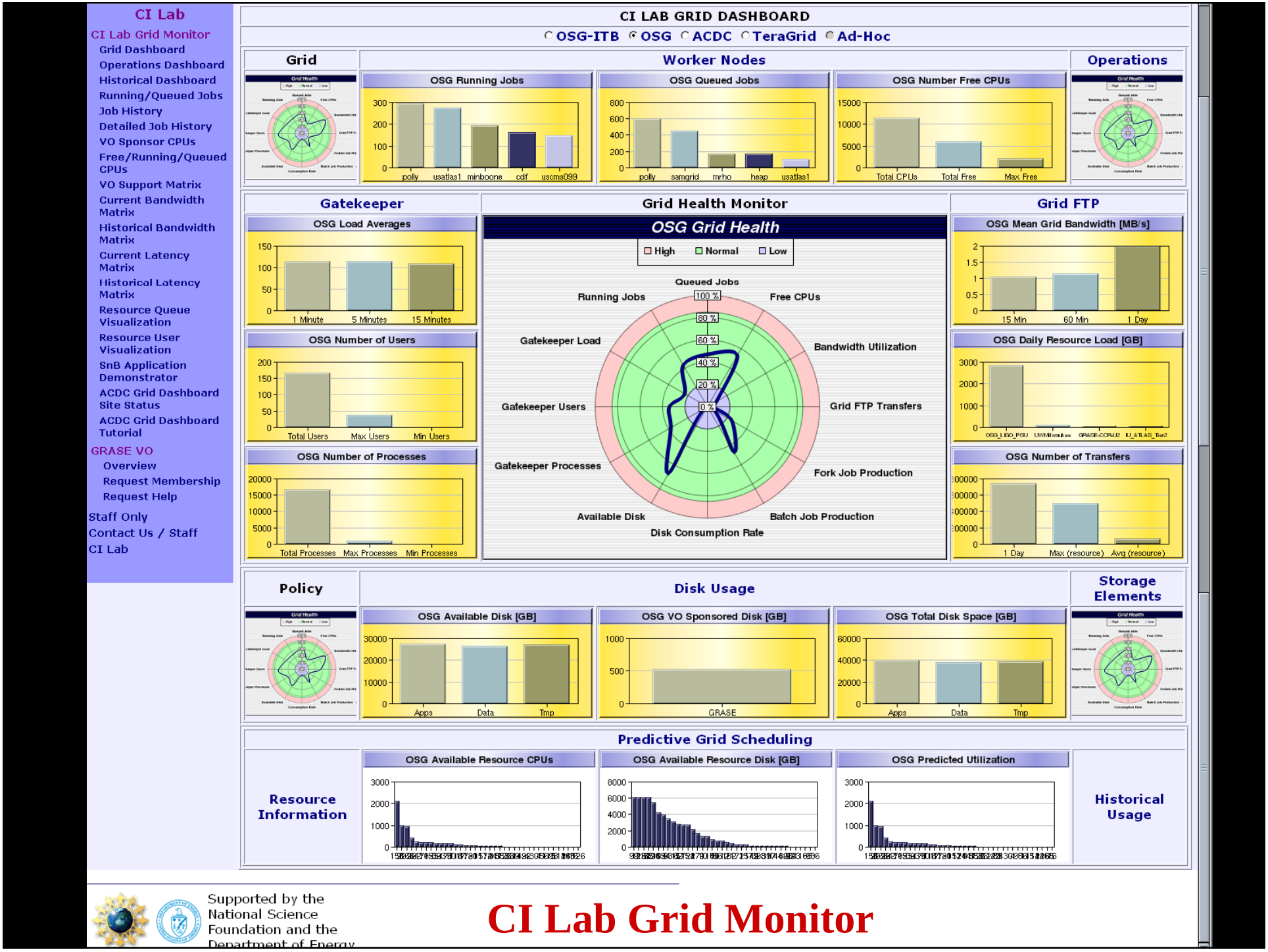
MCIL Lab Collaborations

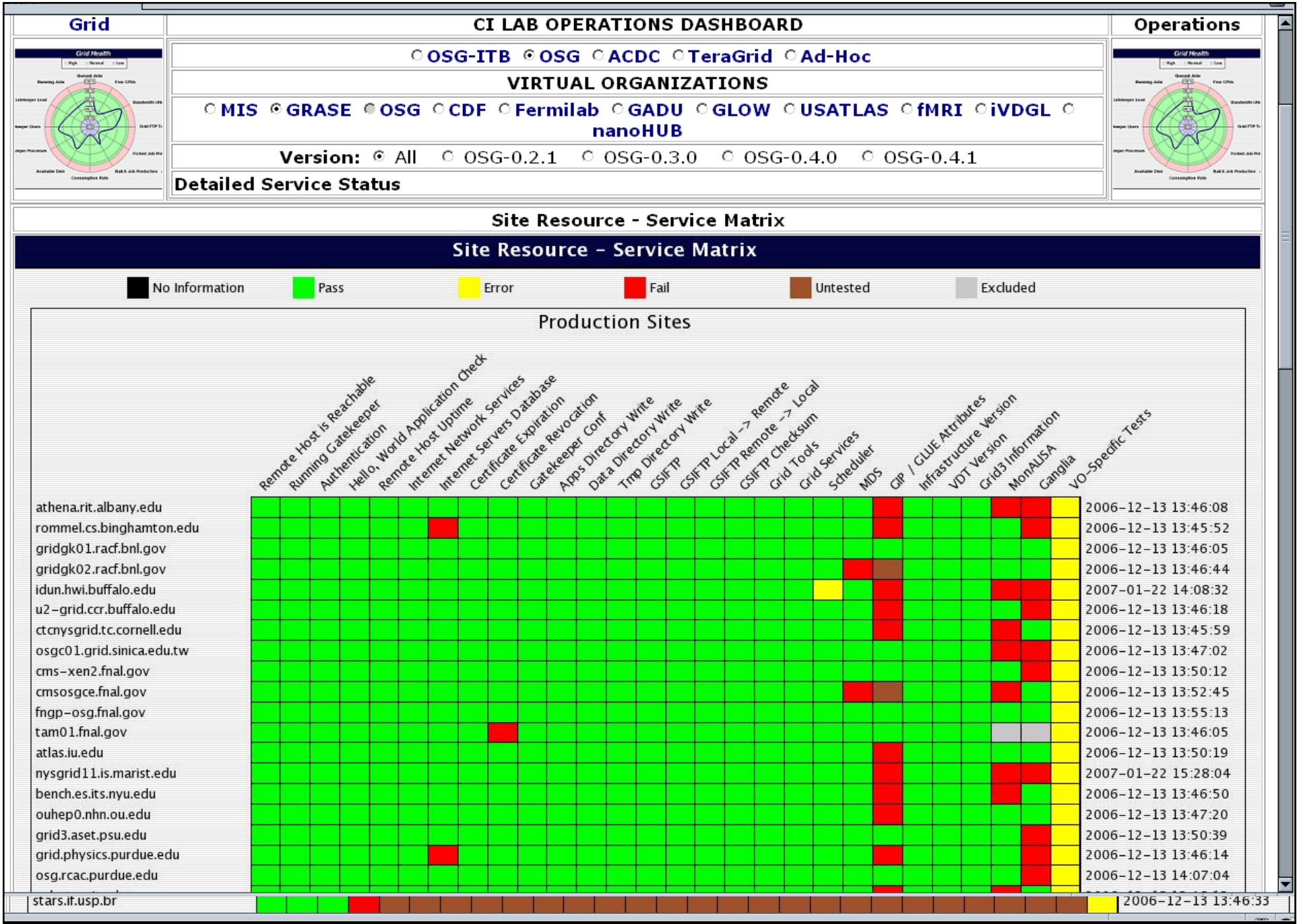
- High-Performance Networking Infrastructure
- Grid3+ Collaboration
- iVDGL Member
  - Only External Member
- Open Science Grid
  - GRASE VO
- NYS CI Initiative
  - Executive Director
  - Various WGs
- Grid-Lite: Campus Grid
  - HP Labs Collaboration
- Innovative Laboratory Prototype
  - Dell Collaboration



MCIL Lab Projects

- **Lightweight Grid Monitor (Dashboard)**
- **Predictive Scheduler**
  - **Define quality of service estimates of job completion, by better estimating job runtimes by profiling users.**
- **Dynamic Resource Allocation**
  - **Develop automated procedures for dynamic computational resource allocation.**
- **High-Performance Grid-Enabled Data Repositories**
  - **Develop automated procedures for dynamic data repository creation and deletion.**
- **Integrated Data Grid**
  - **Automated Data File Migration based on profiling users.**
- **Grid Portal**


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



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*

- **Problem Statement**
  - **Use all available resources for determining a single structure**
- **Grid Enabling Criteria**
  - **Run on heterogeneous set of resources**
  - **Store results in *SnB* database**
  - **Mine database (and automagically deploy new jobs) to improve parameter settings**
- **Runtime Parameters Transparent to User**
  - **Assembling Necessary Files**
  - **Number of Processors**
  - **Trials per Processor**
  - **Appropriate Queue and Running Times**

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

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.

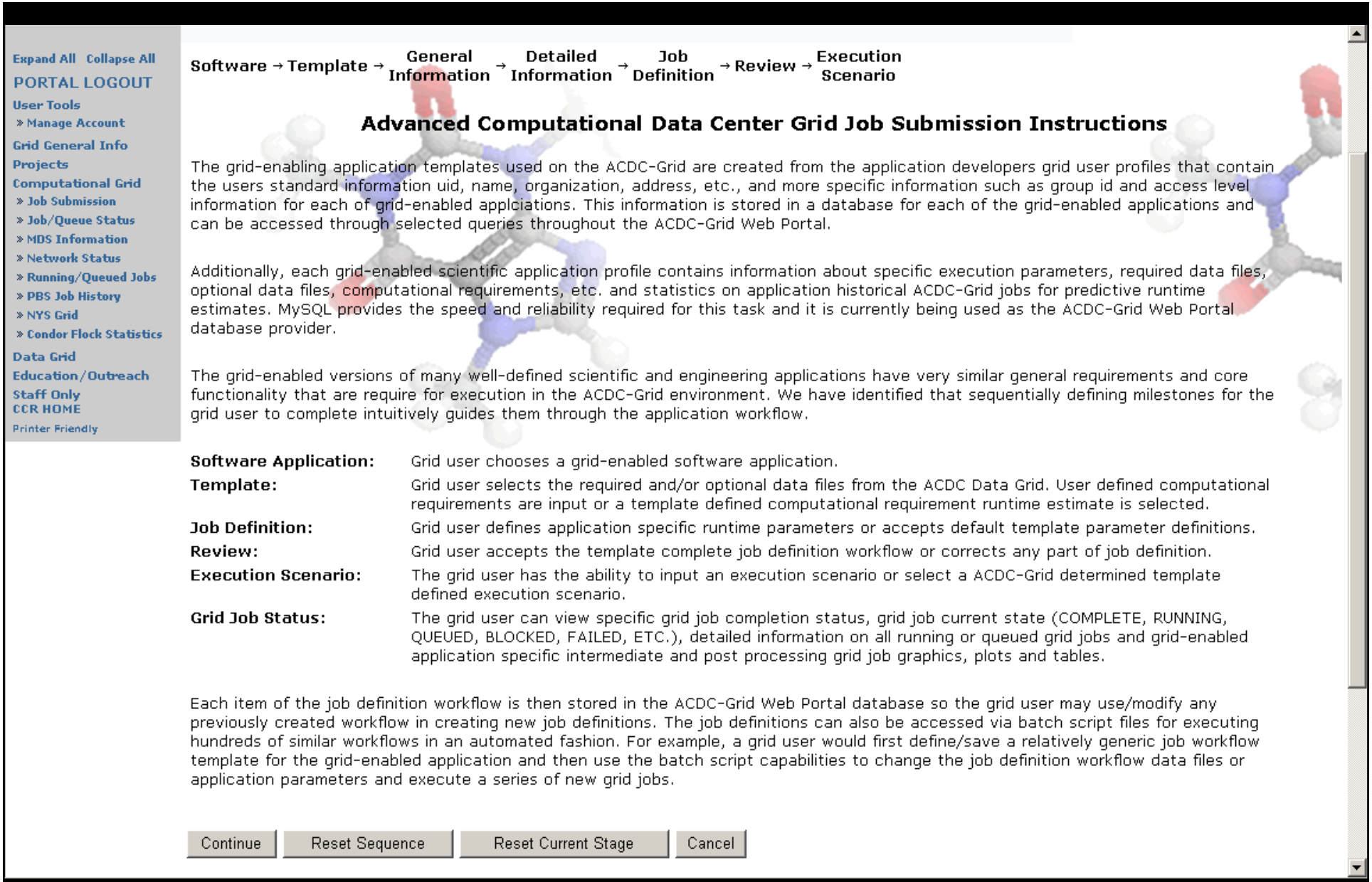
The grids developed by the CI Lab support teaching and research activities, as well as providing infrastructure that includes high-end data, computing, imaging, grid-enabled software, all of which relies on the New York State Research Network (**NYSERNet**).

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.

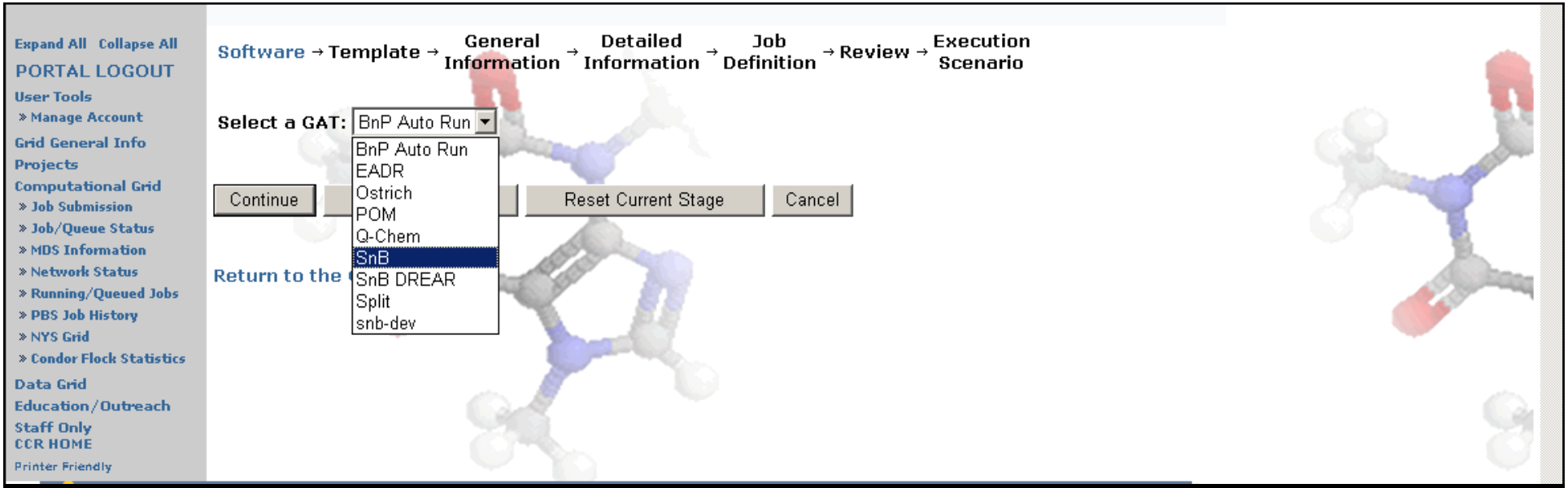
BnP

Software : BnP
Field : Protein crystal structure determination

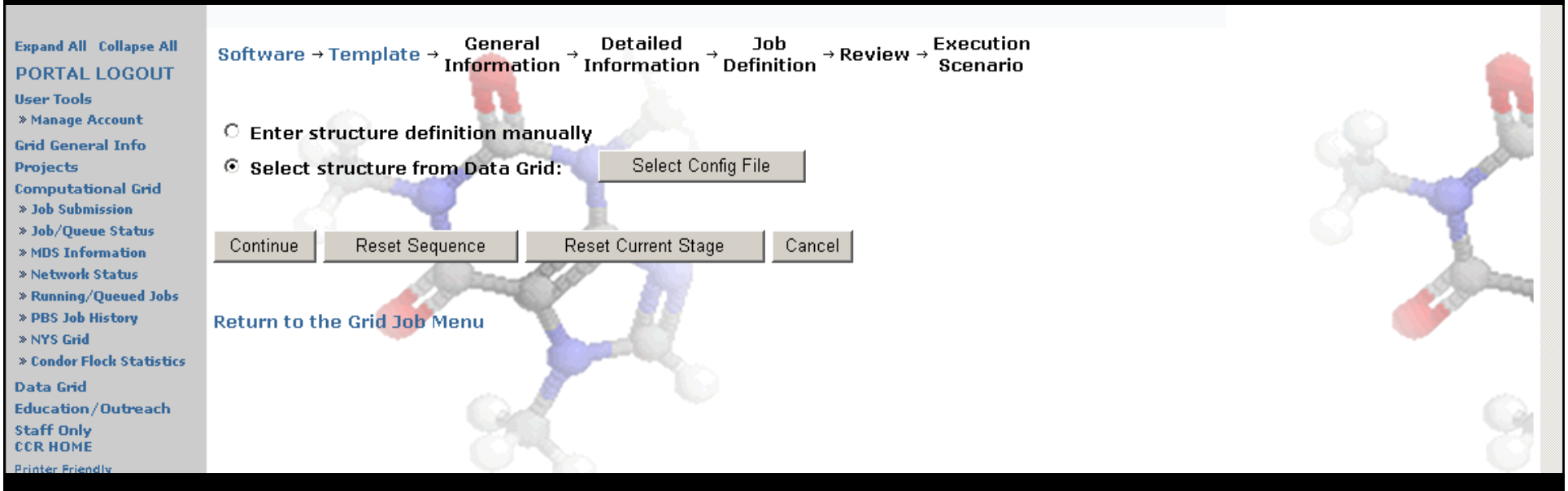
Startup Screen for CI Lab Grid Job Submission



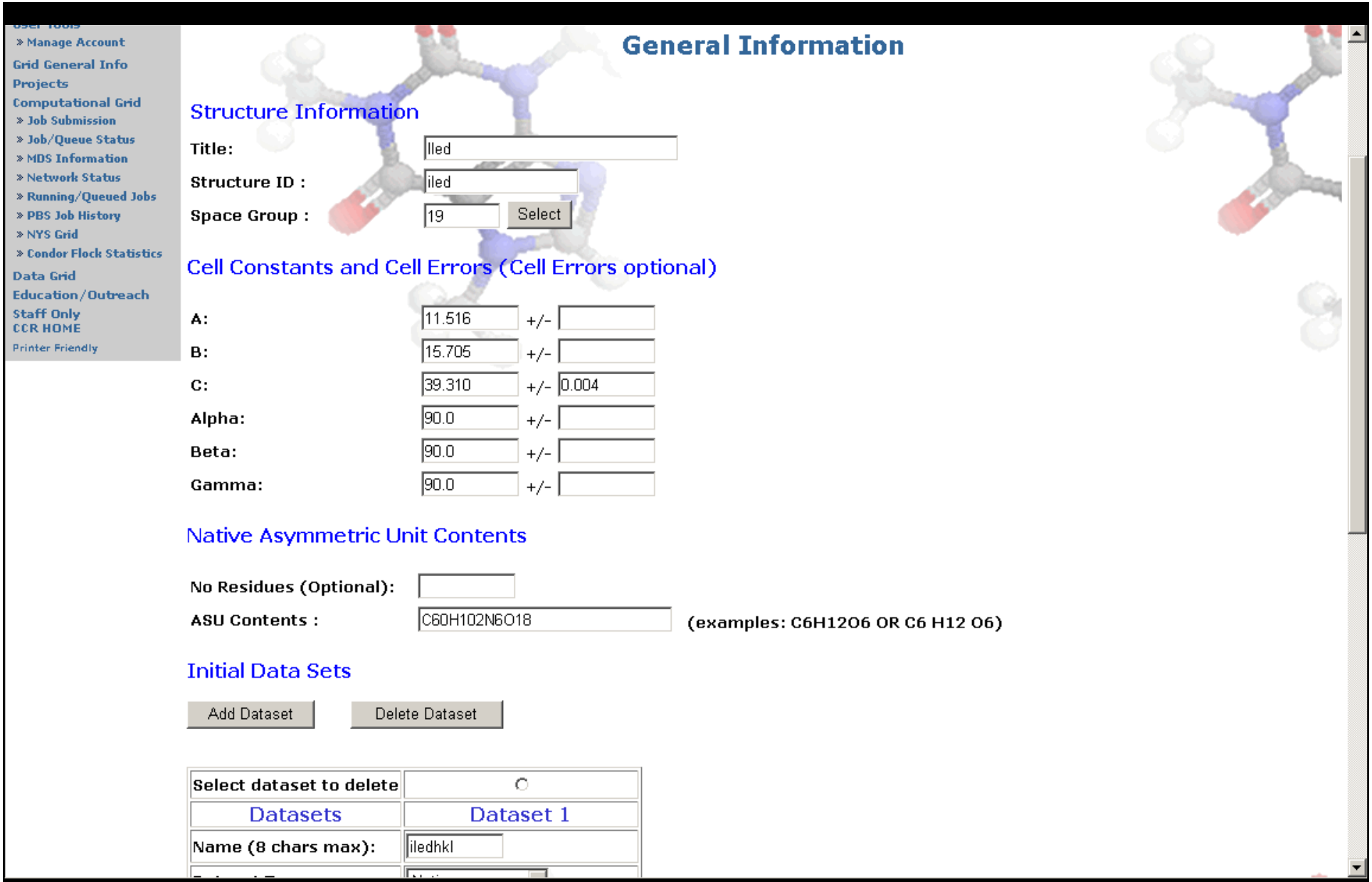
Instructions and Description for Running a Job on ACDC-Grid



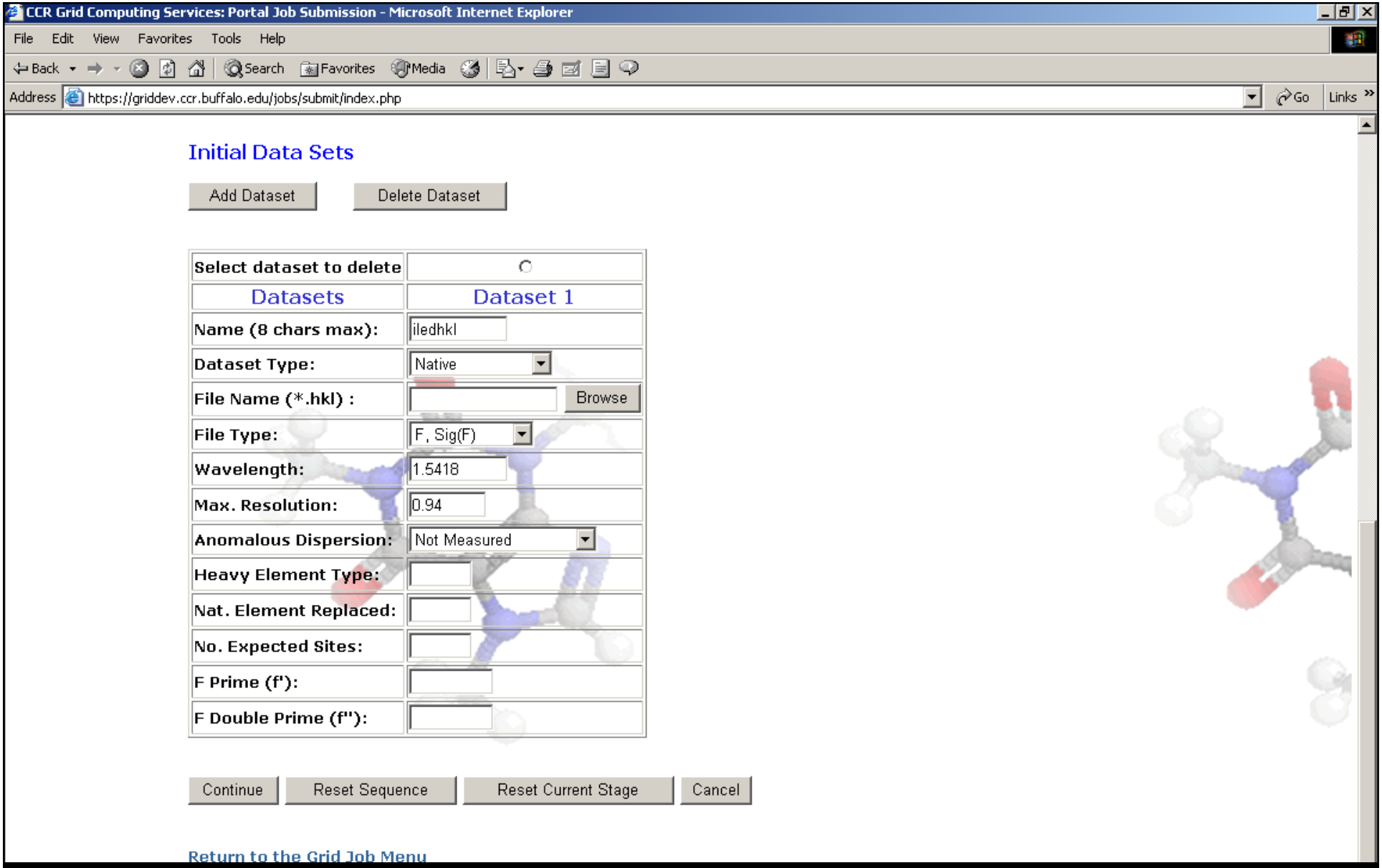
Software Package Selection



Full Structure / Substructure Template Selection



Default Parameters Based on Template

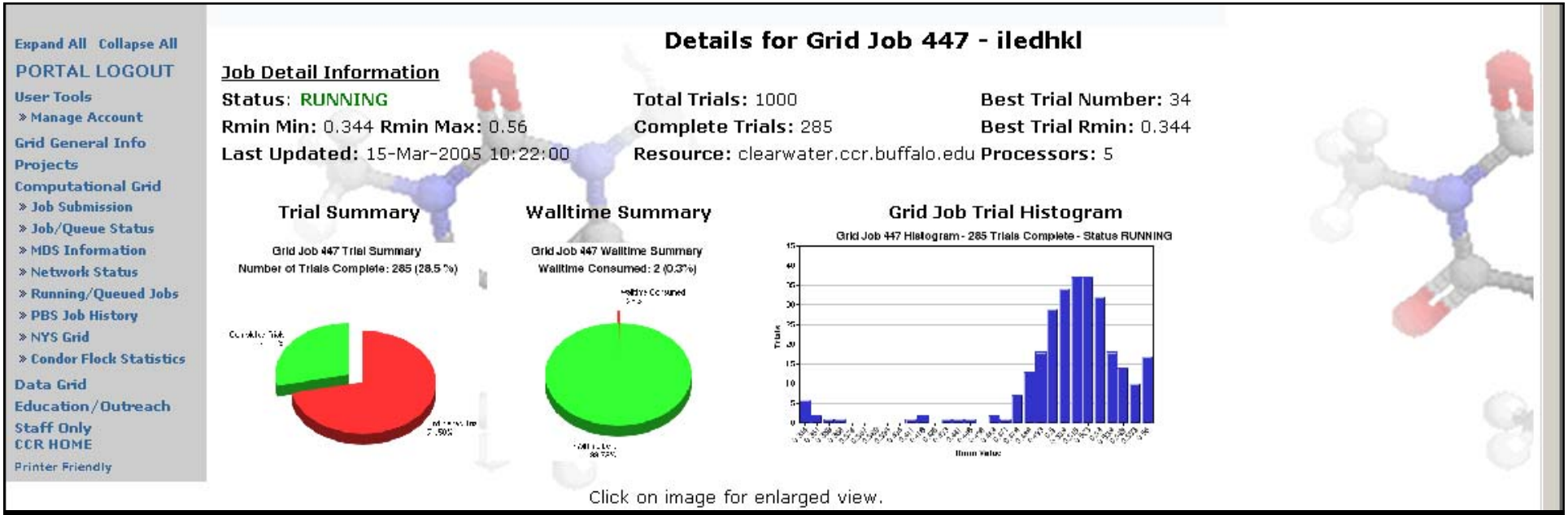


Default Parameters (cont'd)

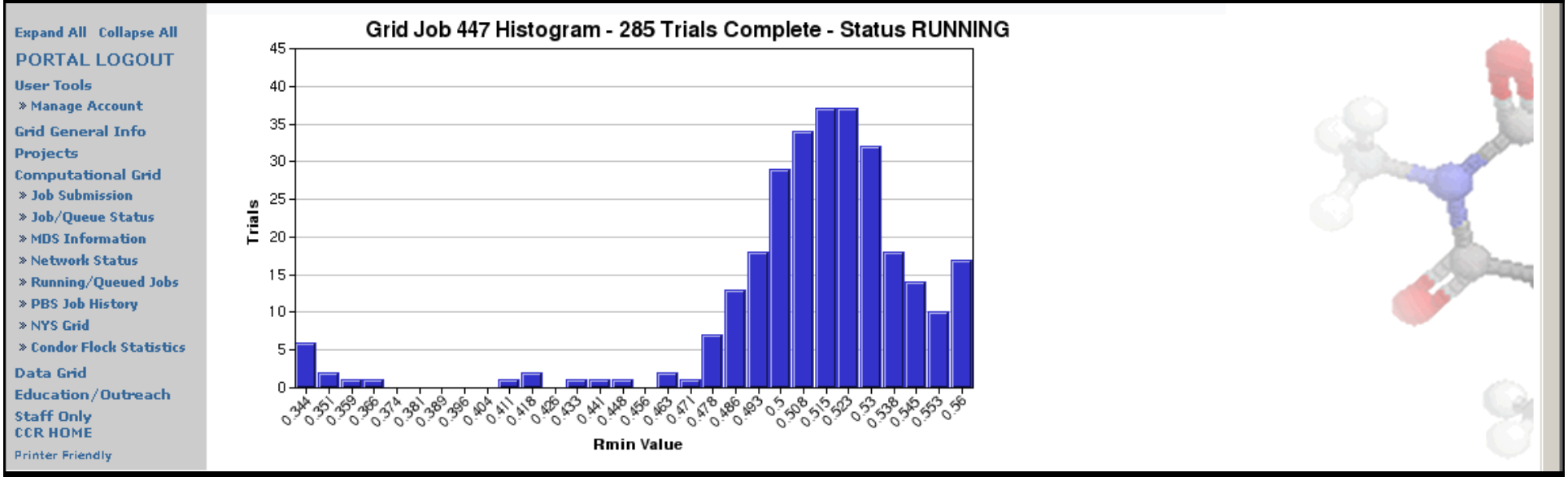
SnB Job Review

Grid Job ID:	447
Selected resource:	clearwater.ccr.buffalo.edu
Number of processors:	5
Wallclock time requested:	720
Number of triplet invariant to use:	8400
Start Phases From:	Random Atoms
Random seed (prime):	11909
Number of trials:	1000
Starting Trial:	1
Input Phase File:	Unused
Input Atom File:	Unused
Keep complete (every trial) peak file? :	Yes
Number of Shake-and-bake cycles:	20
Keep complete (every cycle) trace file? :	No
Terminate trials failing the R-Ratio test? :	No
R-Ratio cutoff:	Unused
Phase Refinement Method:	Parameter Shift(Fast)
Number of passes through phase set:	3
Phase shift:	90.0
Number of shifts:	2
Number of peaks to select:	84
Minimum interpeak distance:	3
Minimum distance between symmetry-related peaks:	3.0
Number of special position peaks to keep:	0
Fourier grid size:	0.31
Perform extra cycles with more peaks? :	No
Number of extra cycles:	Unused
Number of peaks:	Unused
Trials for E-Fourier filtering (fourier refinement)? :	None
Number of cycles:	Unused
Number of peaks:	Unused
Minimum |E|:	Unused

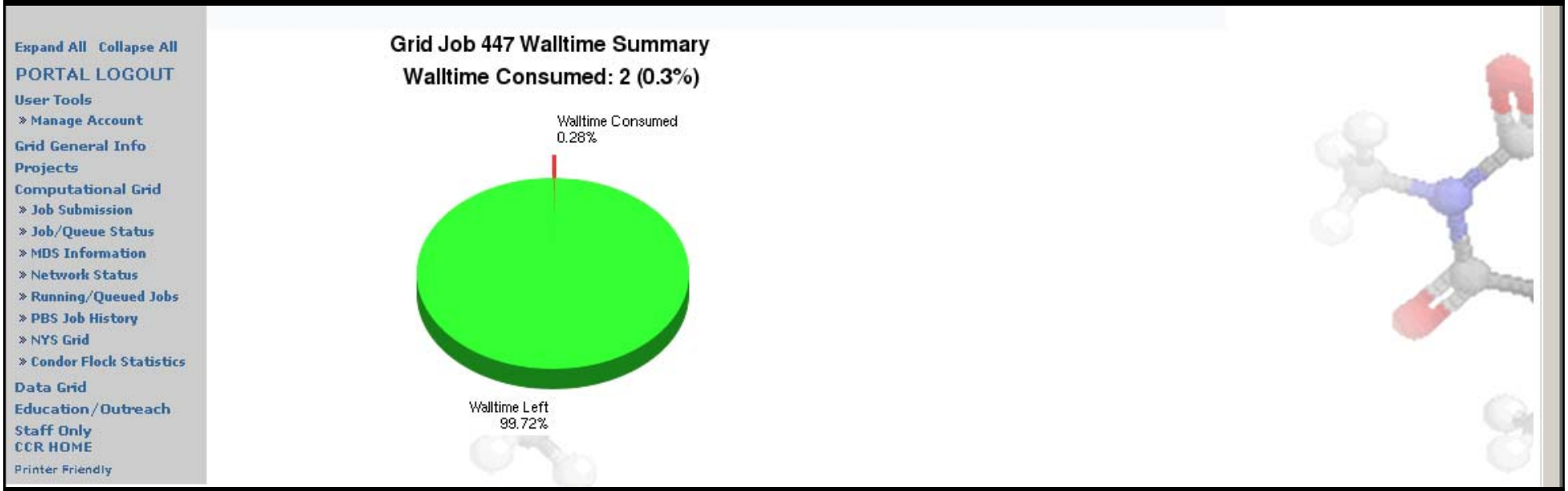
SnB Review (Grid job ID: 447)



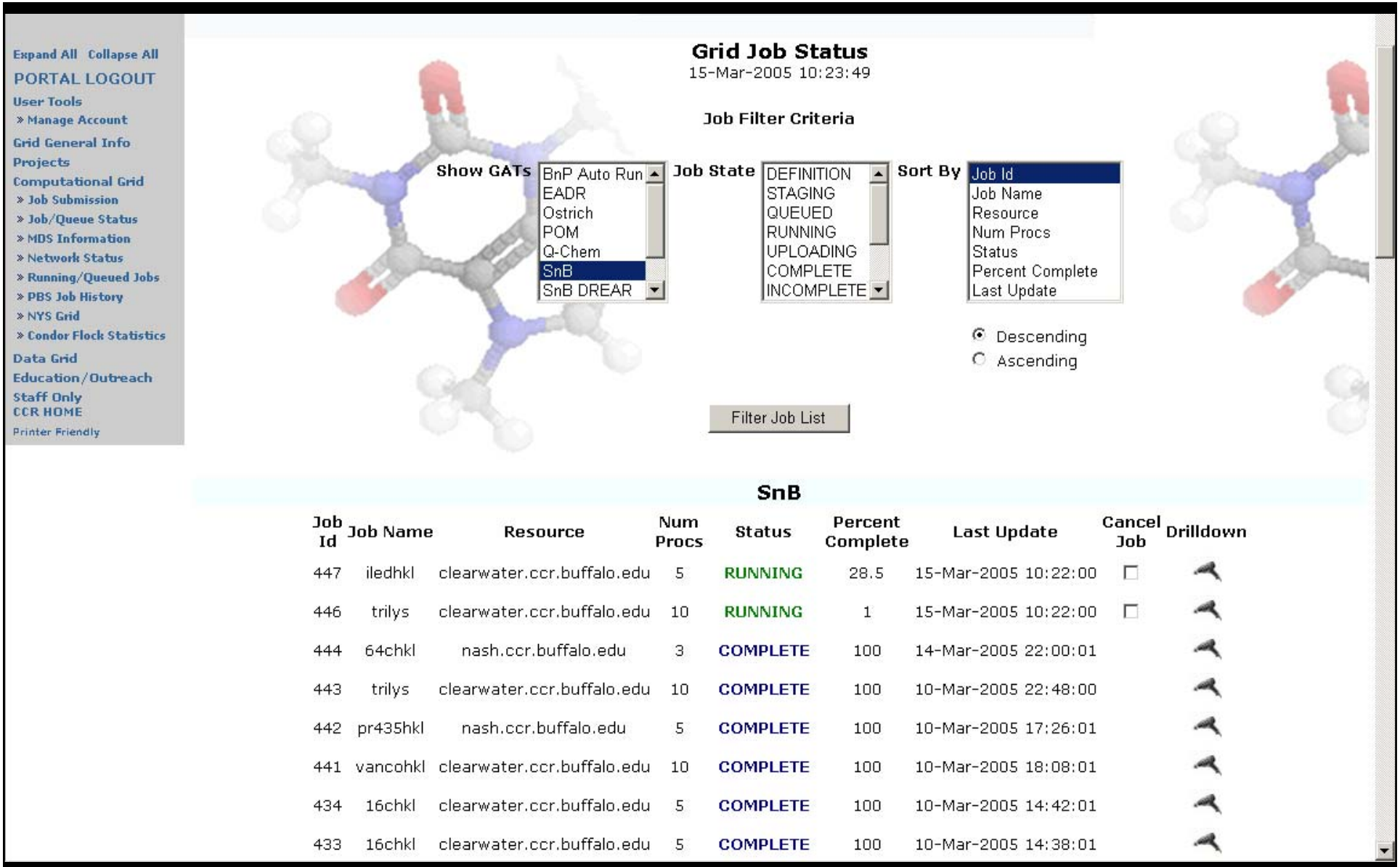
Graphical Representation of Intermediate Job Status



Histogram of Completed Trial Structures

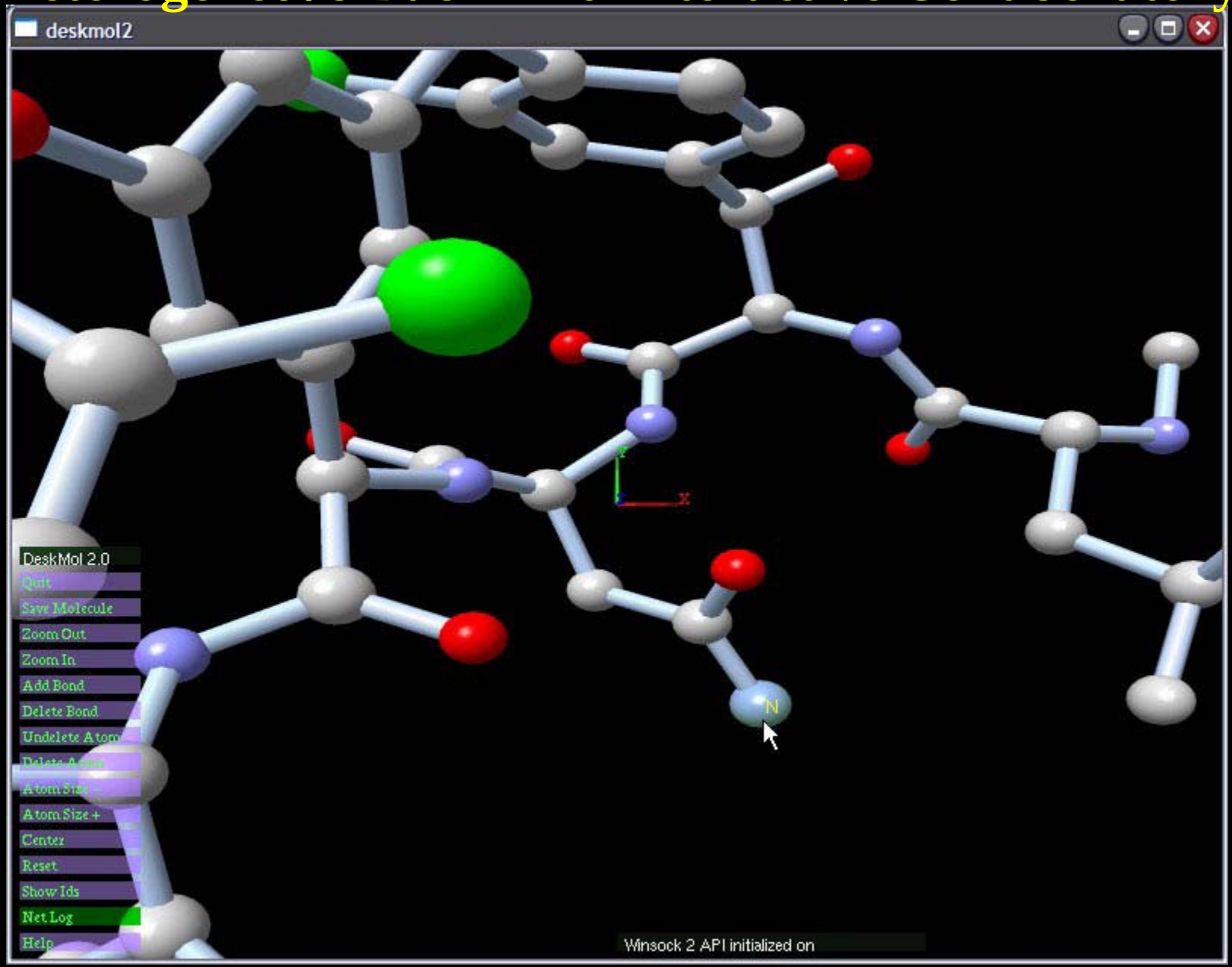


Walltime Summary Chart

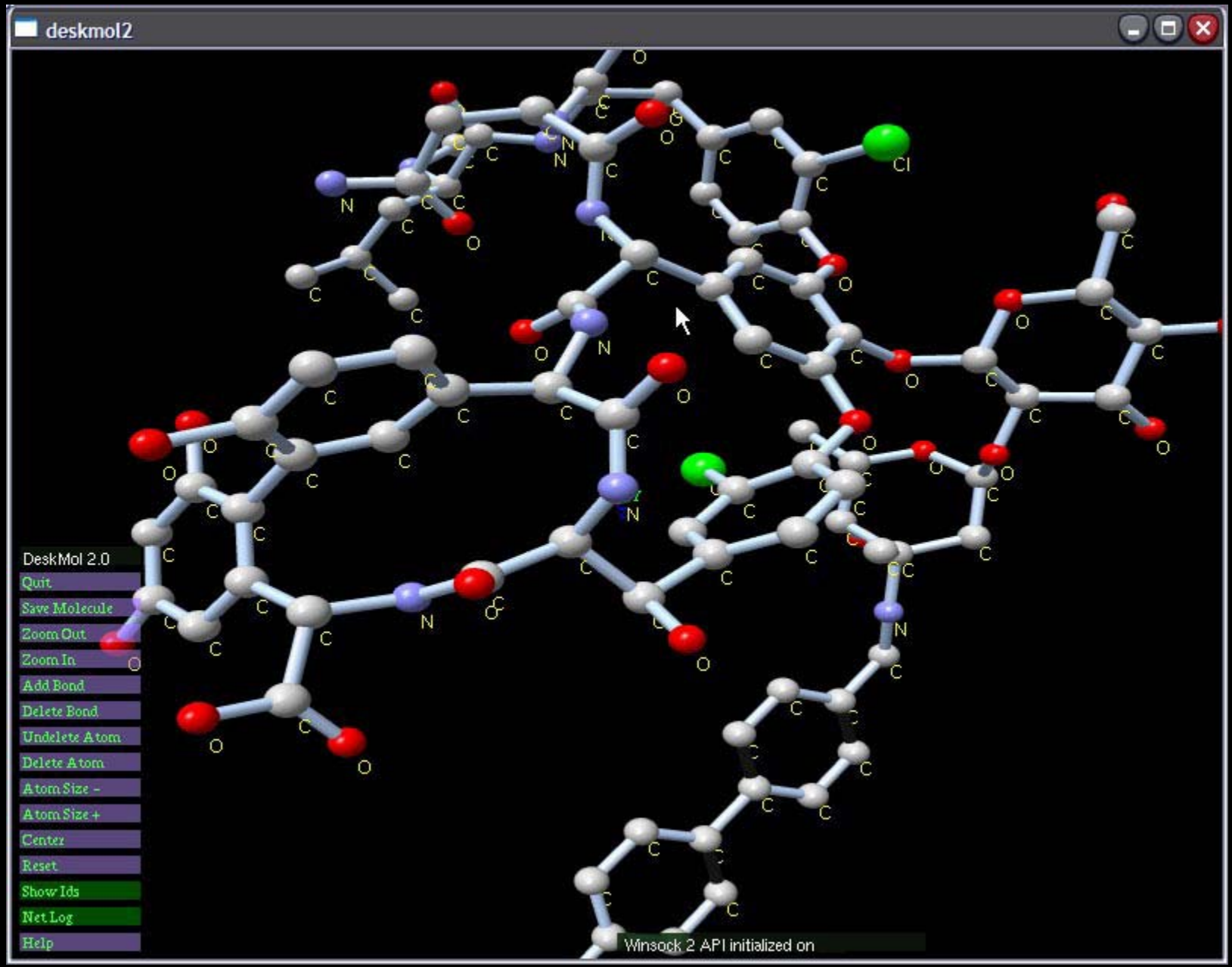


Status of Jobs

Heterogeneous Back-End Interactive Collaboratory



User starts up – default image of structure.



Molecule scaled, rotated, and labeled.

New York State Grass Roots Cyberinfrastructure Initiative

- **Miller's NYS Grid used as fundamental infrastructure.**
- **Currently an initiative of NYSERNet.**
- **Open to academic and research institutions.**
- **Mission Statement: To create and advance collaborative technological infrastructure that supports and enhances the research and educational missions of institutions in NYS.**
- **Enable Research, Scholarship, and Economic Development in NYS.**
- **Currently, no significant utilization.**

Outreach

- **Pilot HS Program in Computational Science**
  - **Year long extracurricular activity at Mount St. Mary's, City Honors, and Orchard Park HS**
  - **Produce next generation scientists and engineers**
  - **Students learn Perl, SQL, Bioinformatics**
  - **$50,000 startup funding from Verizon, PC's from HP**

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
CCR
CCR