

Discovery 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**

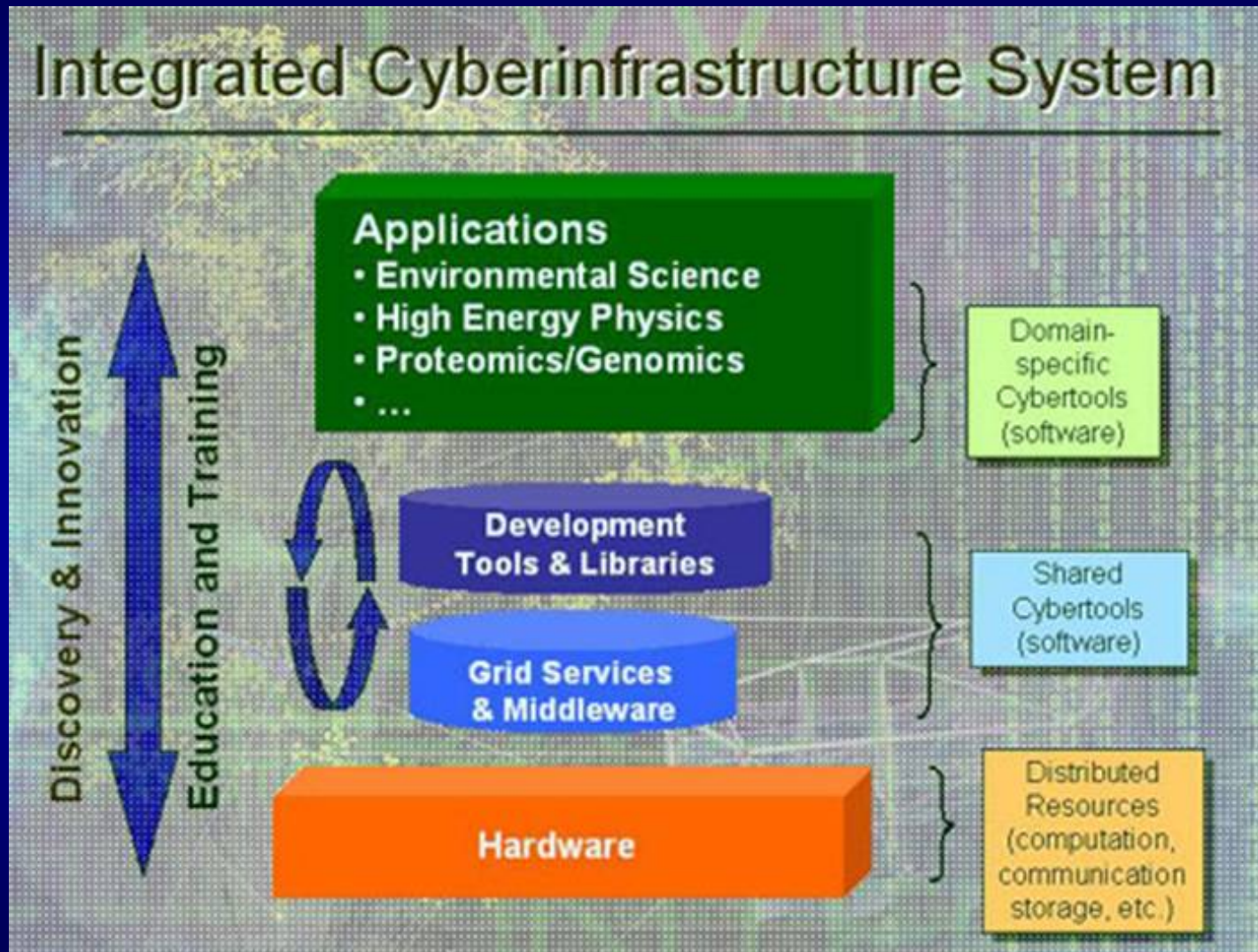


Academia in the 21st Century: Medium-Level View

- **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**

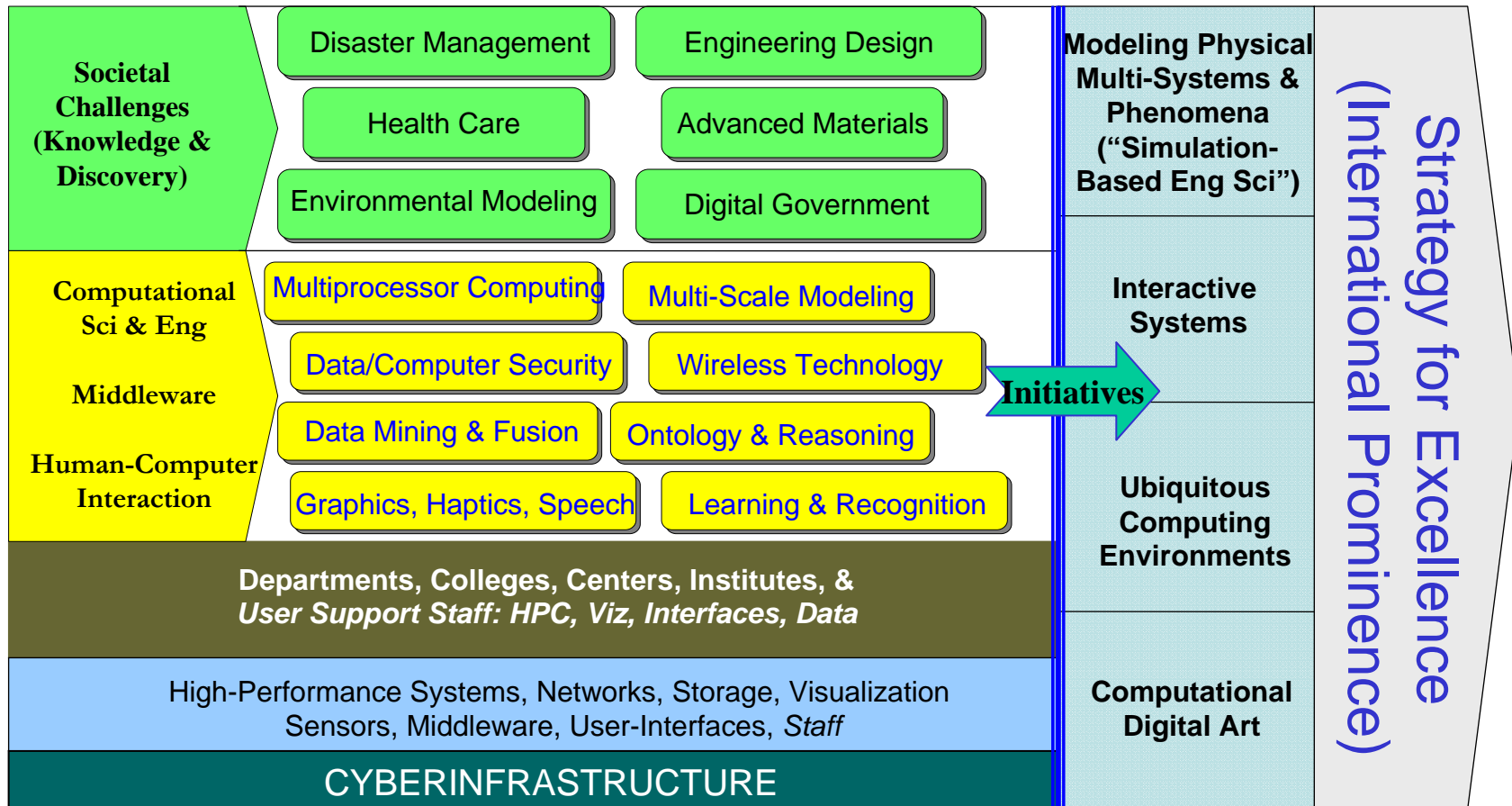


NSF Integrated Cyberinfrastructure



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)



Academic Computing Initiative: Organization

- **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**

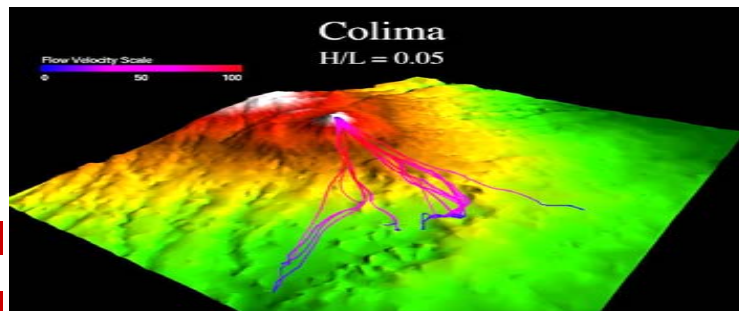


Center for Computational Research (CCR)

- **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**



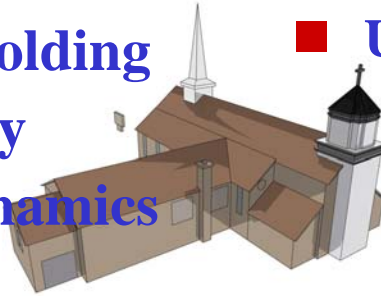
CCR Highlights (1998-2006)

- 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

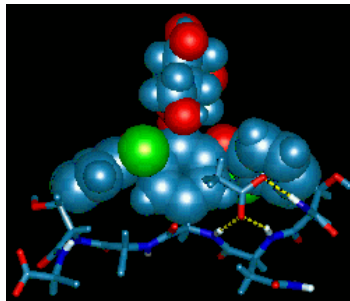
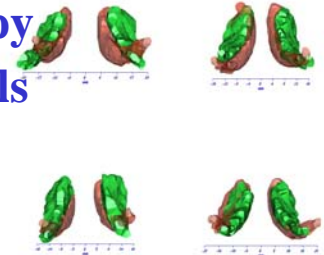
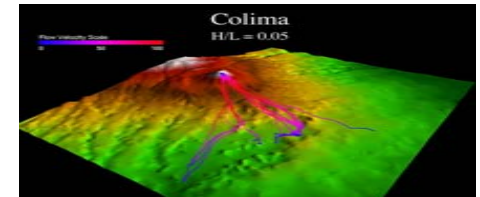


- Videos: MTV
- Urban Simulation and Viz
 - StreetScenes
 - I-90 Toll Barrier
 - Medical Campus
 - Peace Bridge



- 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

Williamsville Toll Barrier Improvement Project



Initial Photo Match incorporating real and computer-generated components



University at Buffalo *The State University of New York*

Cyberinfrastructure Laboratory

CI Lab

Peace Bridge Visualization: Animation & Simulation

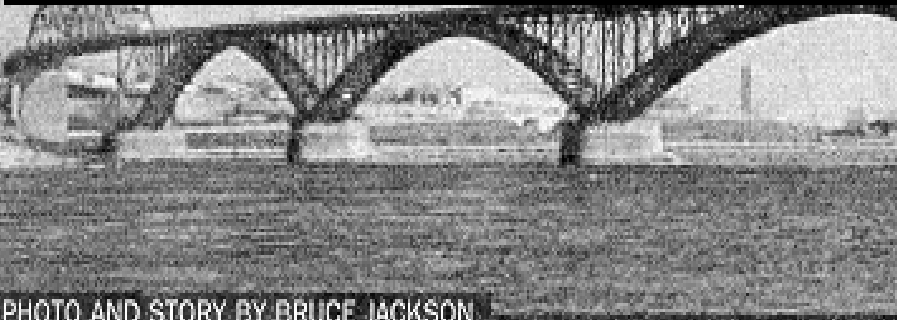


PHOTO AND STORY BY BRUCE JACKSON

■ Proposed Options

- ❑ Relocate US plaza
- ❑ Build a 3-lane companion span & rehab existing bridge



MTV

IBC Digital & CCR

Song: I'm OK (I Promise)

Band: Chemical Romance

Gaming Environment: Death Jr.



University at Buffalo *The State University of New York*

Cyberinfrastructure Laboratory

CI Lab

Virtual Reality

Alive on the Grid: PAAPAB

■ Networked art application for CAVE

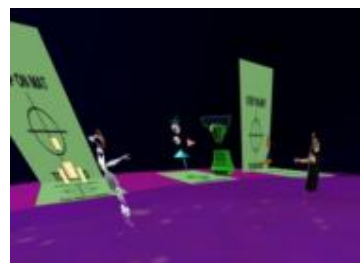
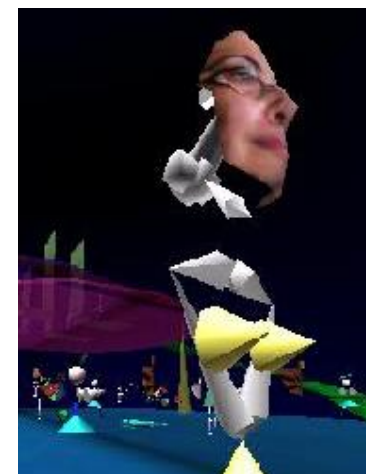
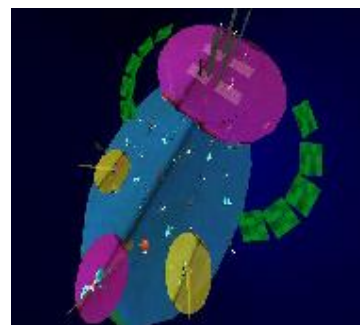
- Users from around the world
- First performance 2001

■ 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

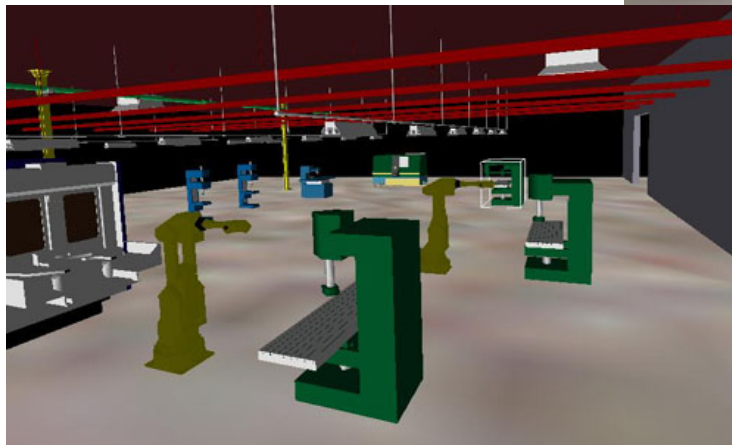
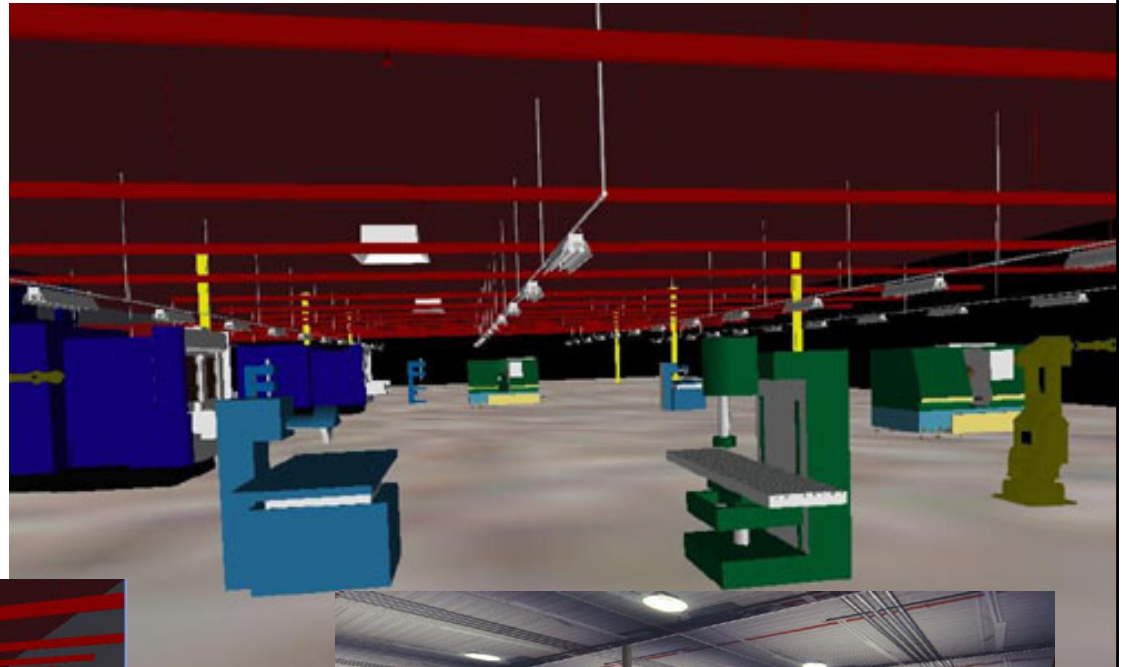


J. Anstey



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**

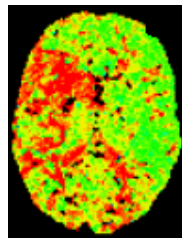
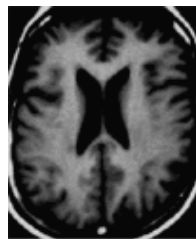
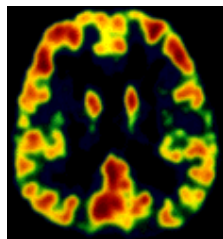
■ **Fetal Viability Test**

■ **Edible Vaccine for Hepatitis C**

■ **Timed-Release Insulin Therapy**

■ **Anti-Arrhythmia Therapy**

□ **Tarantula venom**



■ **Direct Methods Structure Determination**

□ **Listed on “Top Ten Algorithms of the 20th Century”**

■ **Vancomycin**

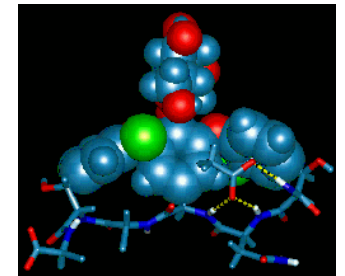
■ **Gramacidin A**

■ **High Throughput**

Crystallization Method: Patented

■ **NIH National Genomics Center: Northeast Consortium**

■ **Howard Hughes Medical Institute: Center for Genomics & Proteomics**



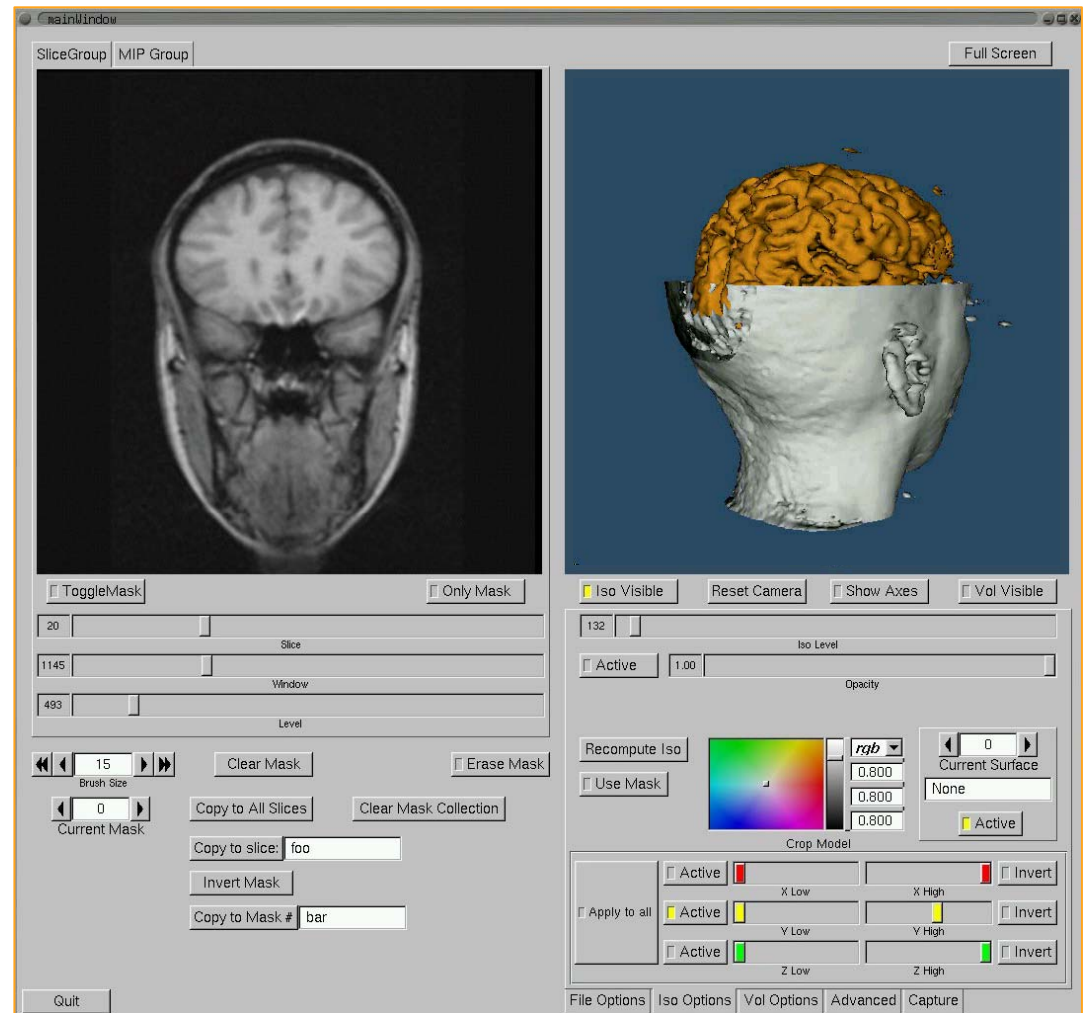
Scientific Visualization

Multiple Sclerosis Project

- Collaboration with Buffalo Neuroimaging Analysis Center (BNAC)

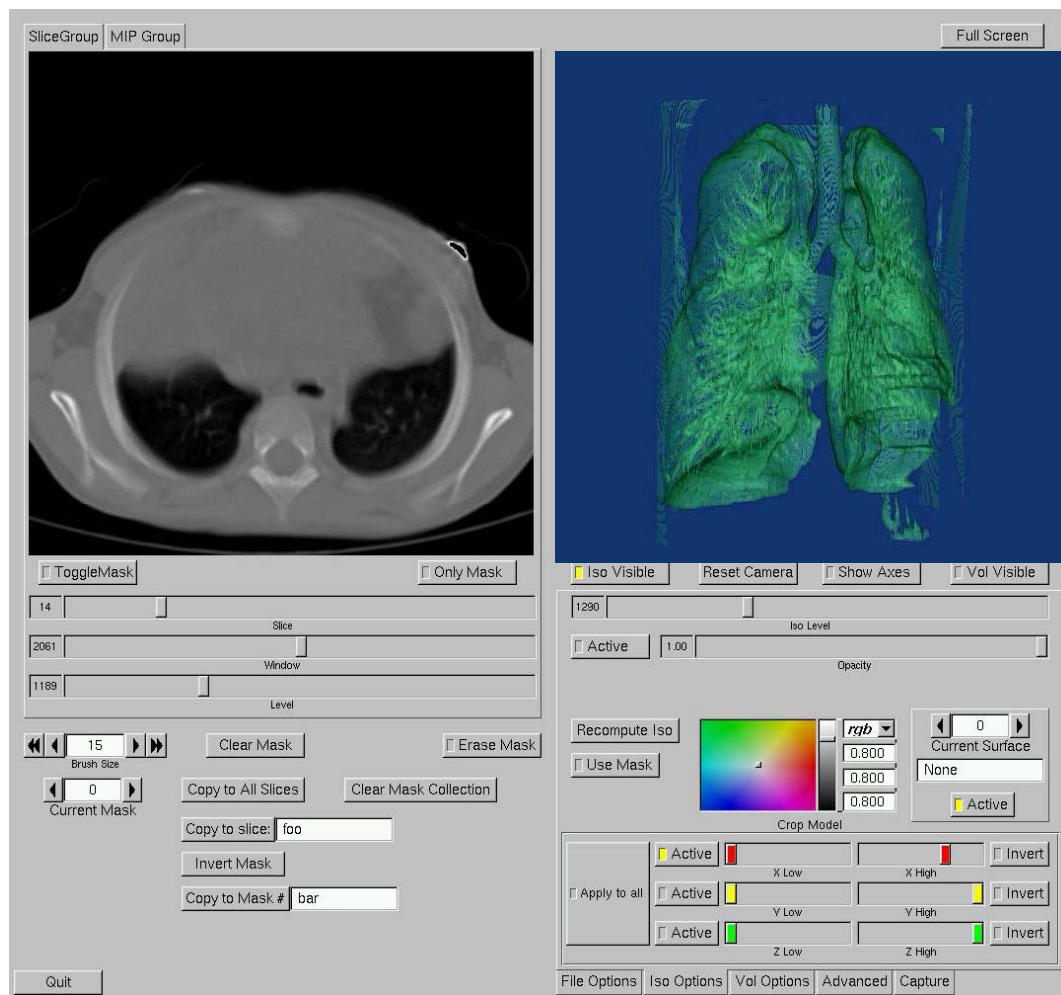
- Developers of Avonex, drug of choice for treatment of MS

- 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



University at Buffalo The State University of New York

Cyberinfrastructure Laboratory

CI Lab

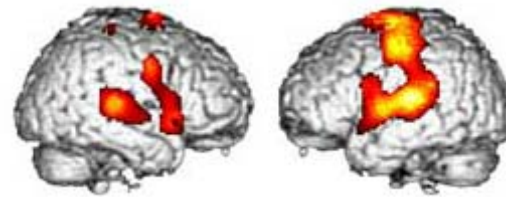
Mapping Brain Activity

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

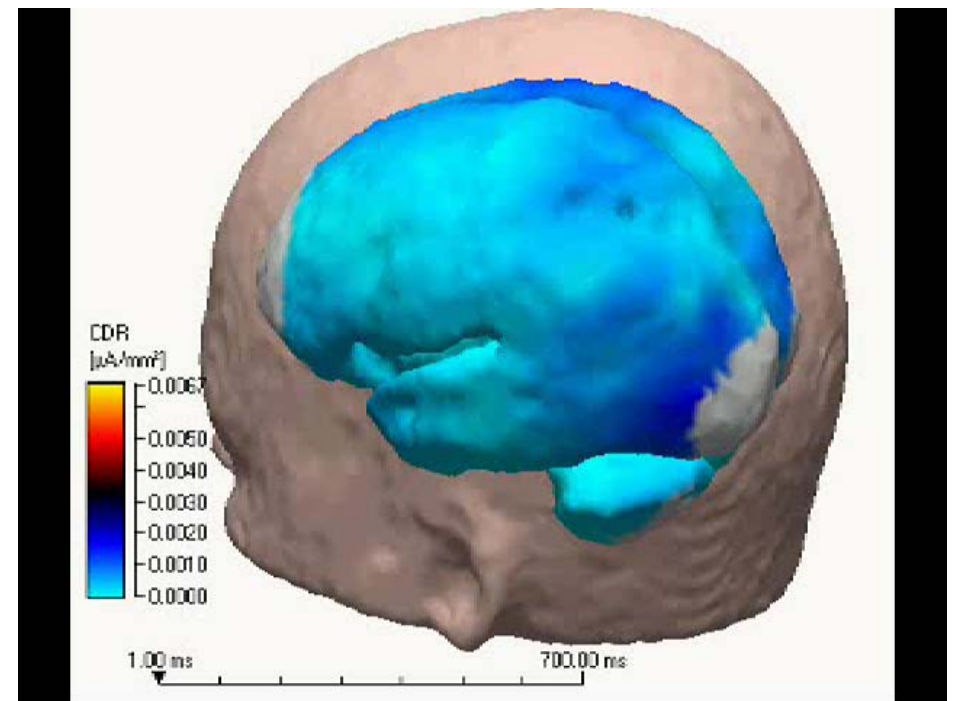
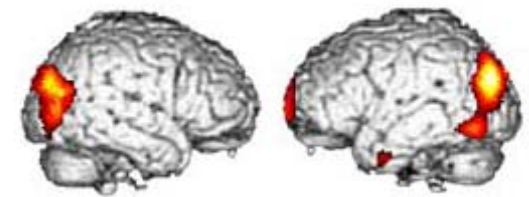
Current density maps of brain surface (1–700 ms after target) show dynamic pattern of brain activity during decision-making process.

A. Lockwood

Sites Activated



Sites Deactivated

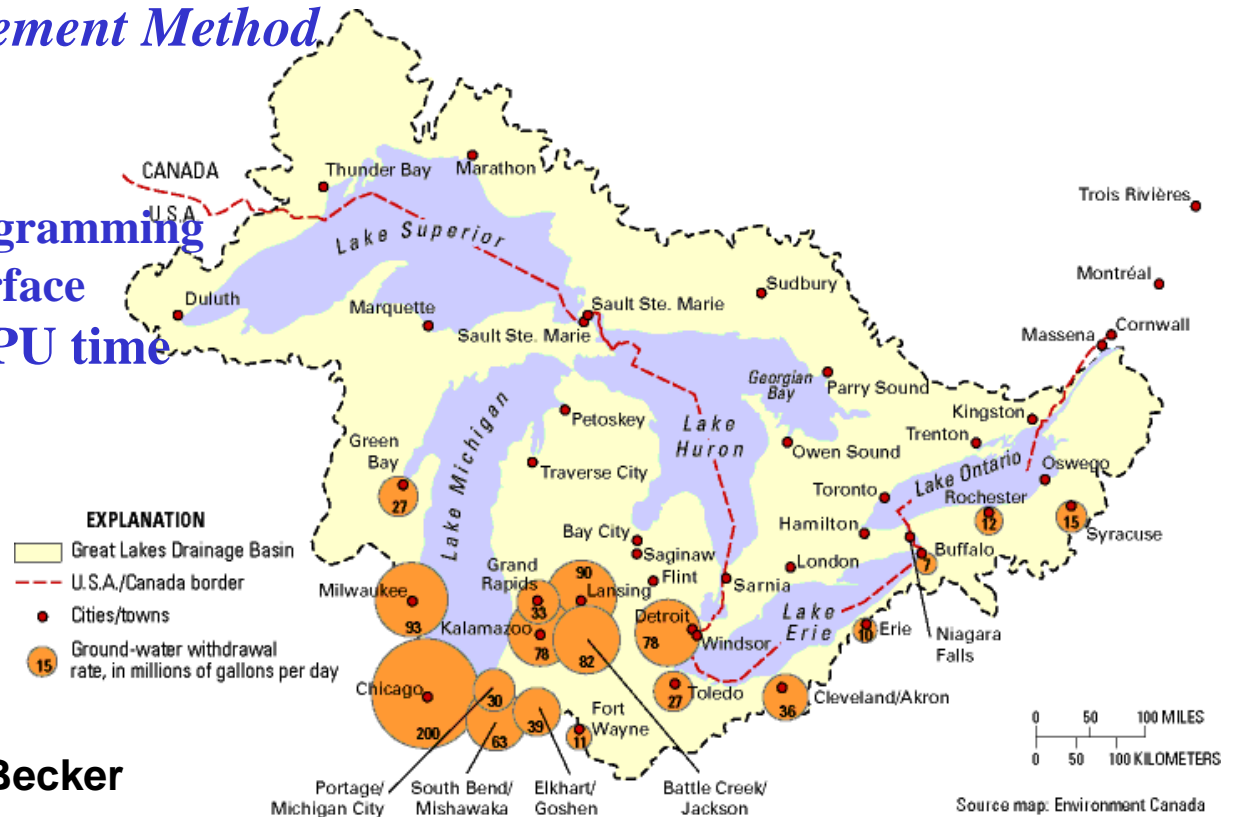


Science & Engineering

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



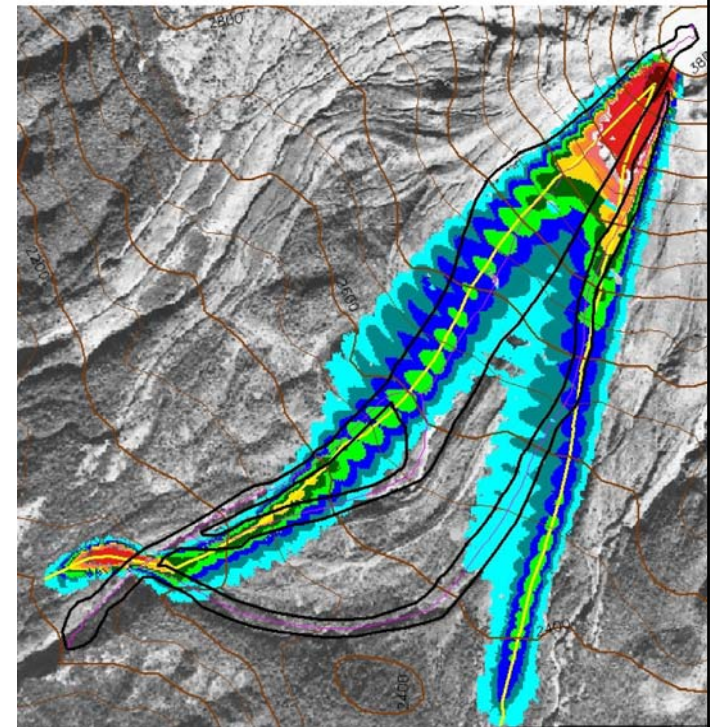
A. Rabideau, I. Jankovic, M. Becker



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



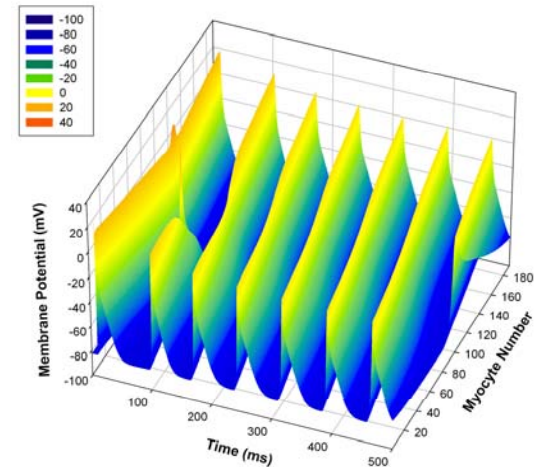
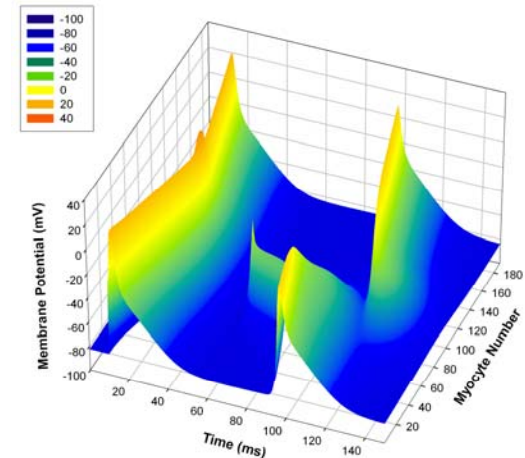
University at Buffalo The State University of New York

Cyberinfrastructure Laboratory

CI Lab

Cardiac Arrhythmia

- Comprehensive models of cardiac cells
- Modeling multicellular cardiac tissues and mechanisms of arrhythmias in the heart
- Simulation of genetic heart disease and arrhythmia suppression by drug application



**Non-sustained and
sustained arrhythmia**

Center for Cellular and Systems Electrophysiology

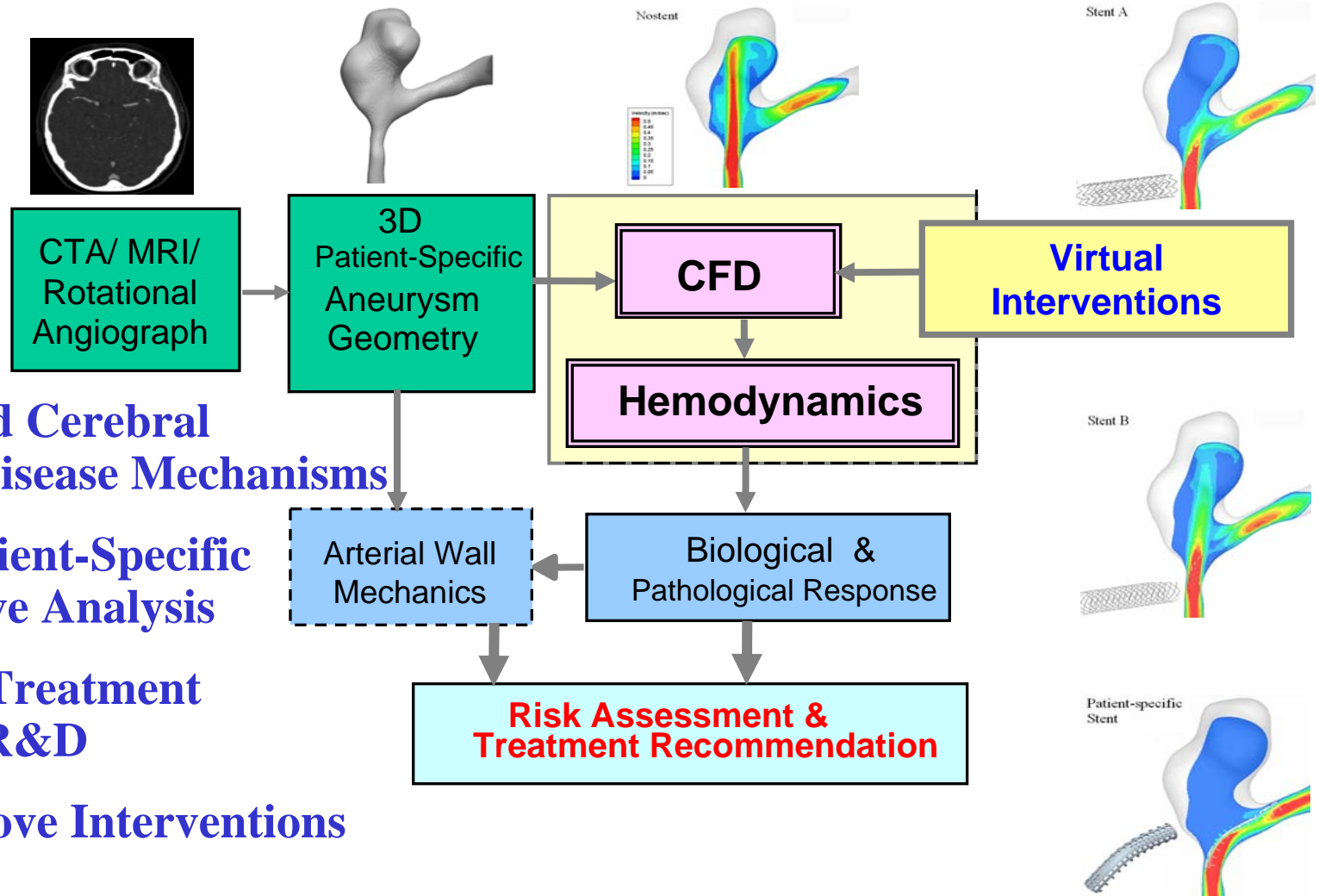


University at Buffalo *The State University of New York*

Cyberinfrastructure Laboratory

CI Lab

Cerebral Aneurysm: Virtual Intervention



- Understand Cerebral Vascular Disease Mechanisms
- Enable Patient-Specific Quantitative Analysis
- Diagnose, Treatment Planning, R&D
- Help Improve Interventions

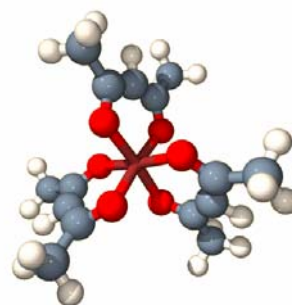
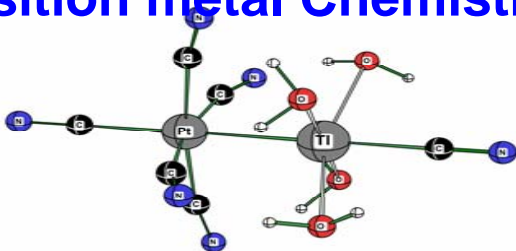


Theoretical and Computational Chemistry

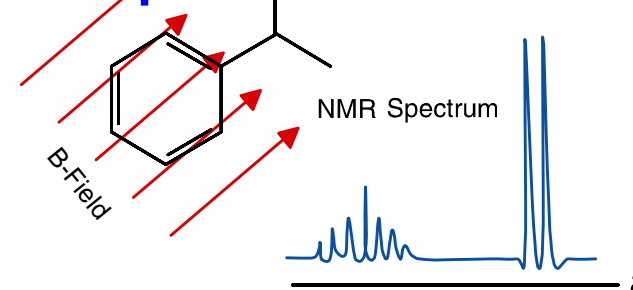
Chemistry

Applied to:

- Polypeptides
- Carbon Nanotubes
- Fullerenes
- Cluster Compounds
- Transition metal Chemistry



Computational NMR



Magnetic Properties of Molecules

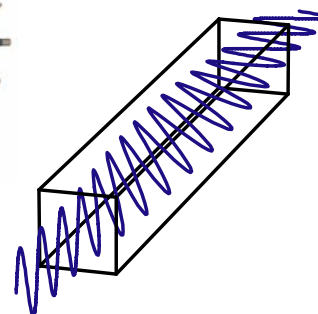
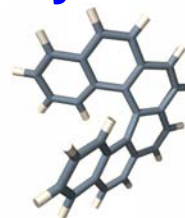
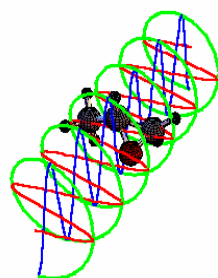


Software Development

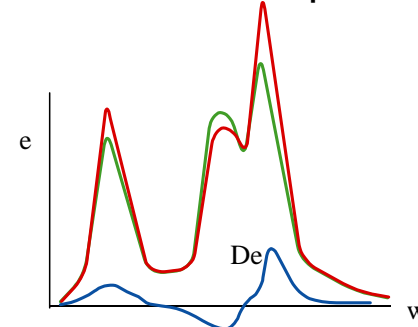
```
do ispin = 1, nspin
  ncol = ncol(ispin)
  nrow = nrow(ispin)
  ncv = ncvr(ispin)
  if (debug) call mout (lclast(1,ispin), ncol, 'Delta Y Matrix'//nspin(ispin))
  read constant part of h1 matrix
  label = 'f1f=00'
  call csadd (label, ..., ipert)
  call csadd (label, ..., nspin(ispin))
  call kfrdr (lufc, label, work, ncv, 1)
  compute f1fc
  if (.not.(lrowc .or. lcolc)) then
    rtemp = zero
    do ll = 1, ncv
      rtemp = rtemp + h1ast(ll,ispin)**2
    end do
    f1fc = f1fc + rtemp
    if (ispin==1) f1fc = f1fc*two
  end if
  if (lcolc .or. lrowc) then
    constant part of h1 matrix
  end if
end do
```

Theory and Calculation of Optical Activity

Optical Rotation



Circular Dichroism Spectra



J. Autschbach



University at Buffalo The State University of New York

Cyberinfrastructure Laboratory

CI Lab

Computational Chemistry

- **UB Software development in Quantum Chemistry**
 - **Q-Chem** – development of combined QM/MM methods for large molecular systems such as proteins
 - **ADF** – development of algorithms to calculate magnetic and optical properties of molecules
- **Used to determine**
 - **3D Molecular Structure**
 - **Electronic Spectra**
 - **Chemical Reactivity**
- **Applications**
 - **Pharmaceutical Drug Design**
 - **Industrial Catalysis**
 - **Materials Science**
 - **Nanotechnology**



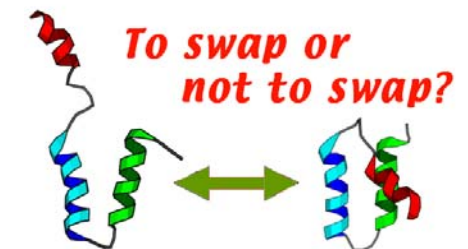
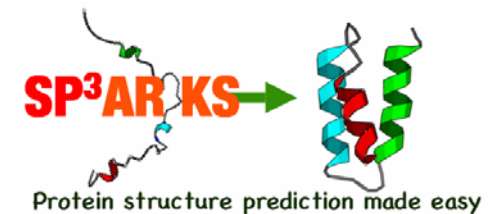
T. Furlani, J. Autschbach, M. Freindorf



Computational Biology & Bioinformatics

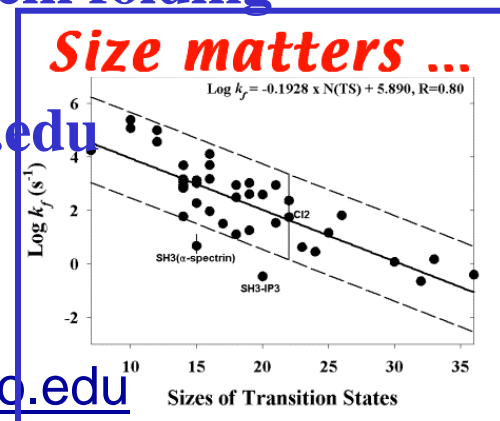
■ Development of Bioinformatic Tools

- ❑ **SPEM** – align multiple sequences for discovering hidden evolution information of genes.
- ❑ **SPARKS/SP³** – predict three-dimensional structures of proteins by matching a query sequence with known structural templates.
- ❑ **DFIRE** – predict binding affinities of protein-protein, protein-ligand, and protein-DNA complexes for structure-based drug design.



■ Mechanistic study of protein folding and binding

- <http://theory.med.buffalo.edu>

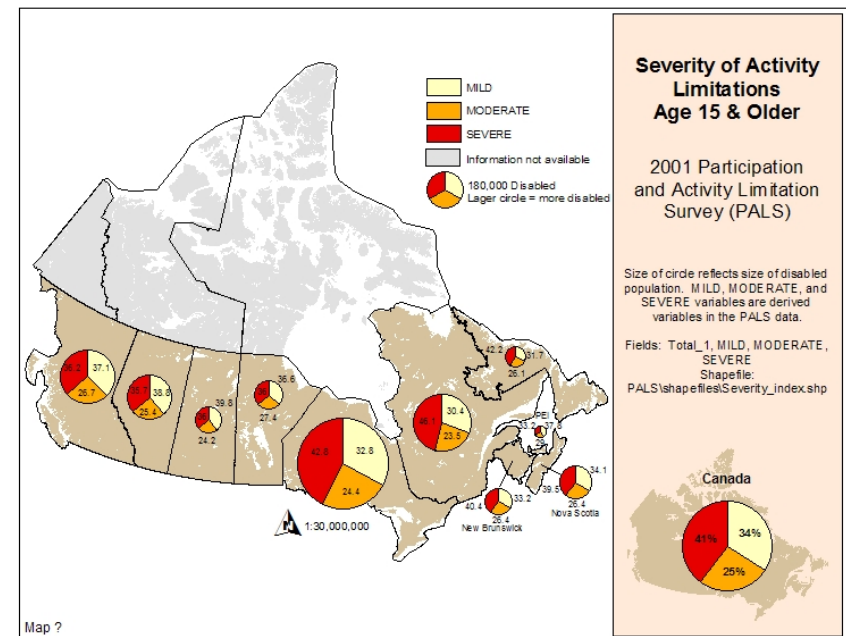
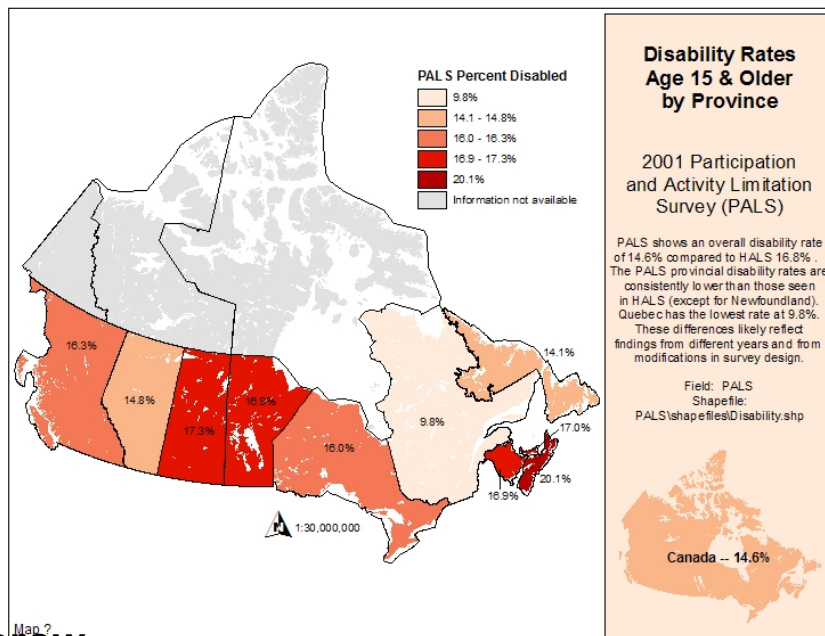


M. Halfon <http://theory.med.buffalo.edu>



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.

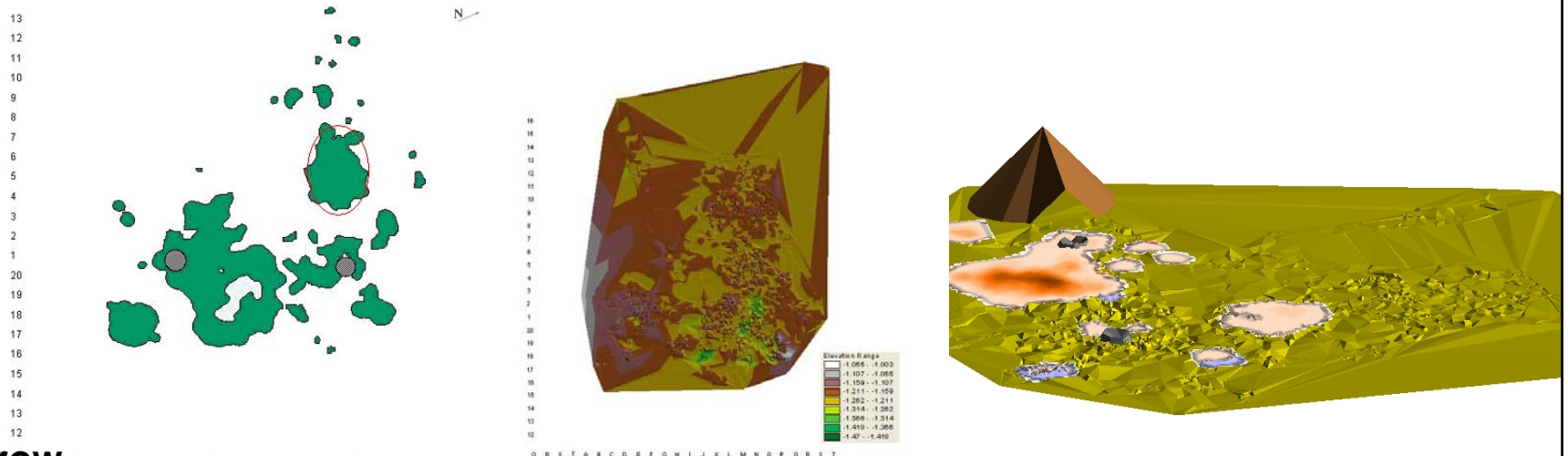


E. Zubrow



Verberie Paleolithic Site in France

- **Intrasite spatial analysis and 3D modeling of the a Late Upper Paleolithic archaeological site in the Paris Basin of France**
- **Social Systems GIS Lab in the Dept. of Anthropology is working with researchers from the CNRS in Paris**
- **Sponsored by the National Science Foundation**

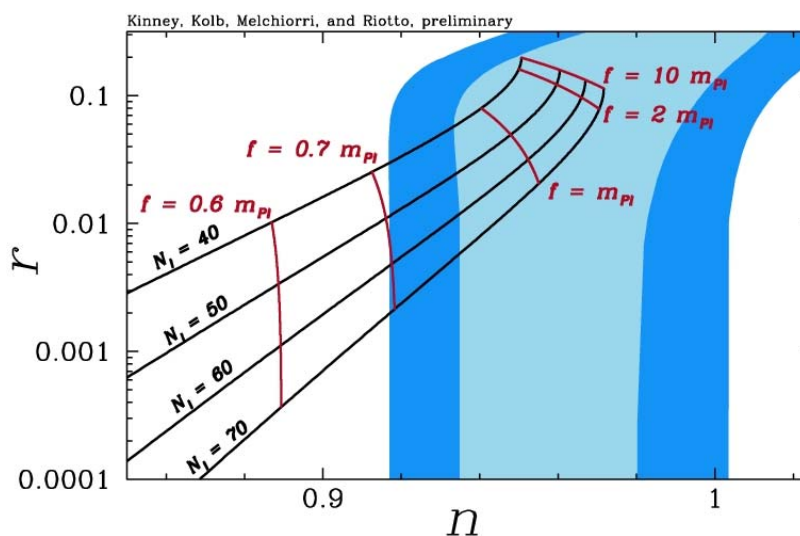
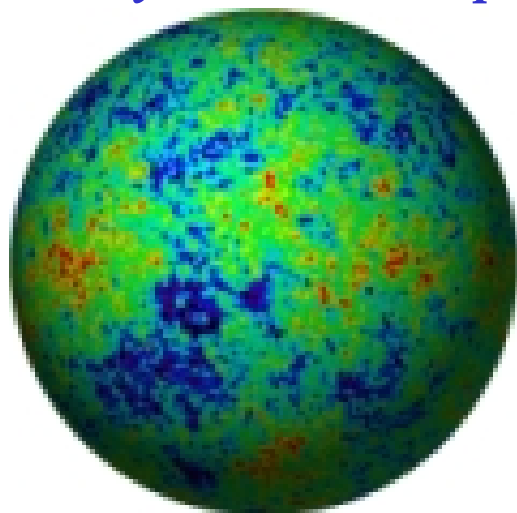


E. Zubrow T A B C D E F G H I J K L M N O P Q R . . .



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*



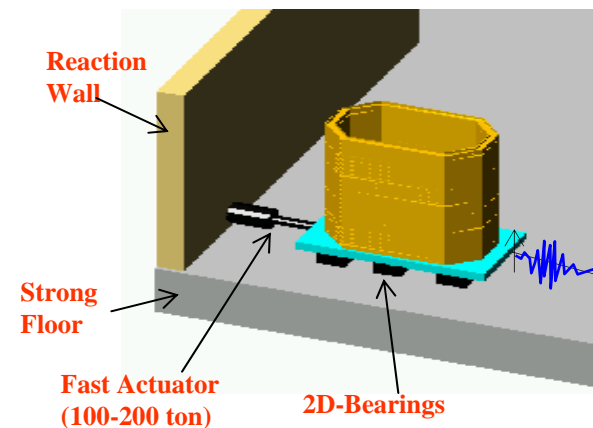
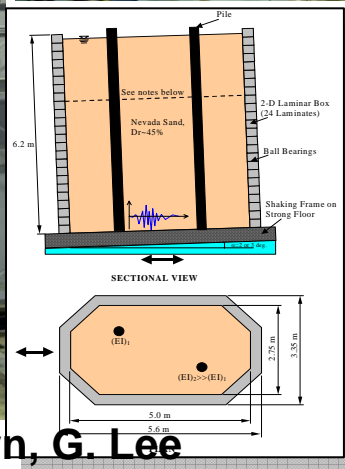
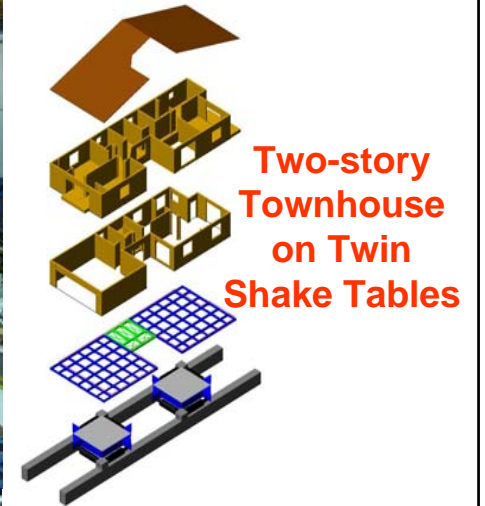
W. Kinney



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

Structural Engineering

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



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

M. Bruneau, A. Reinhorn, G. Lee



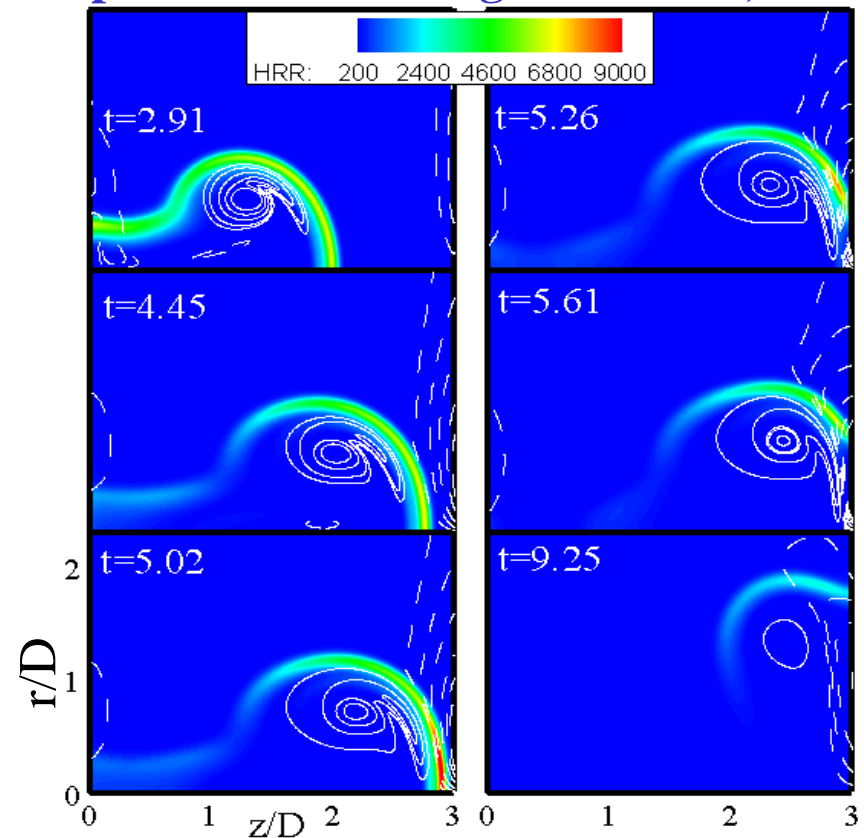
University at Buffalo The State University of New York

Cyberinfrastructure Laboratory

CI Lab

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



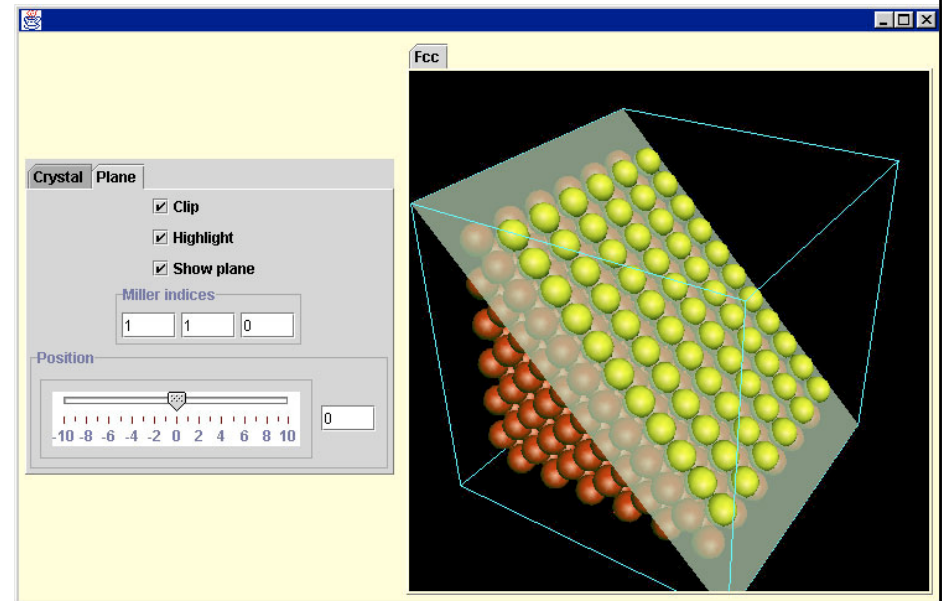
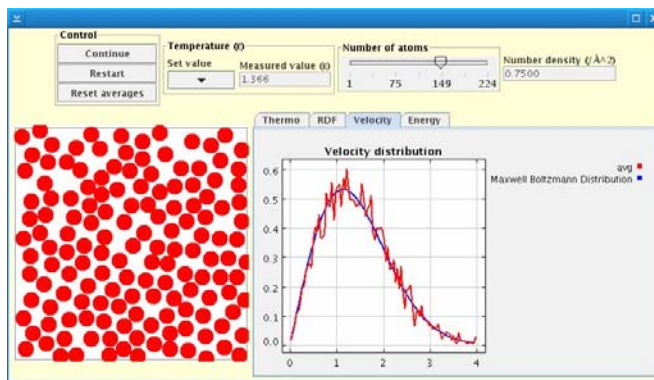
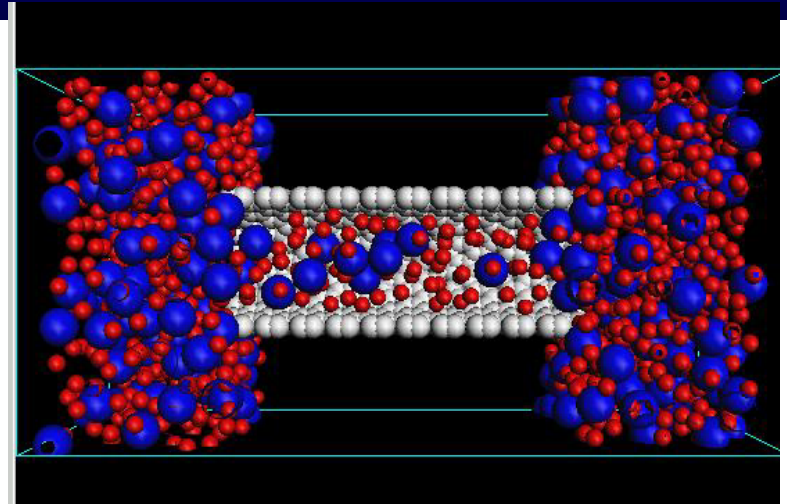
University at Buffalo The State University of New York

Cyberinfrastructure Laboratory

CI Lab

Molecular Simulation Software

- Molecular simulation has wide application in existing and emerging technologies
- Recent advances in information technology make simulation more broadly accessible
- *Etomica* development environment permits easy construction of simulations
- Object-oriented, Extensible, Interactive, Portable and Adaptable
- Stand-alone simulations can be constructed as a teaching tools

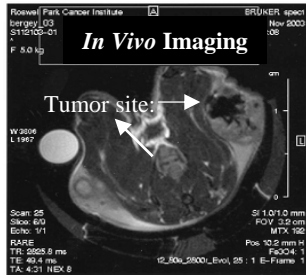


D. Kofke

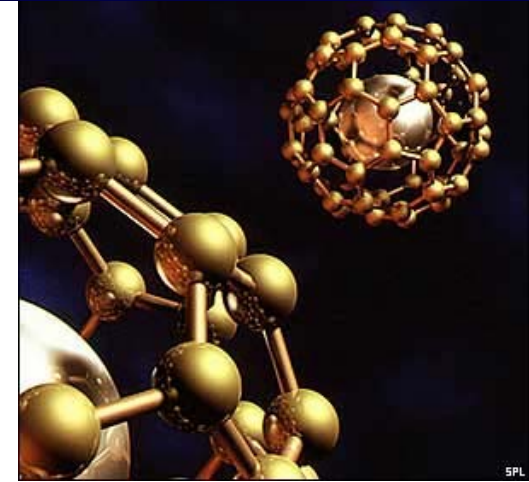
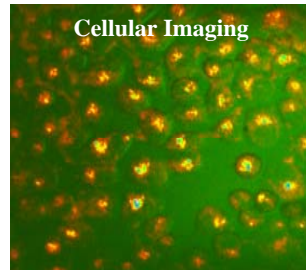


Nanomedicine Program

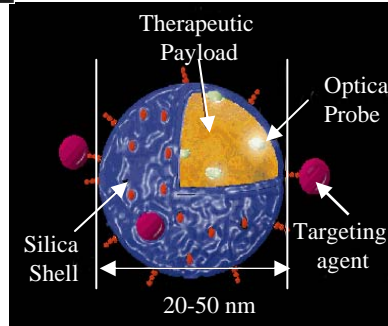
World class Research Program Melding Nanotechnology with Biomedical Sciences



Multi-Modal Imaging



BioCompatibility/ Distribution



In Vivo Sensing

Building from the Bottom Up



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”



University at Buffalo The State University of New York

Cyberinfrastructure Laboratory

CI Lab

Shake-and-Bake

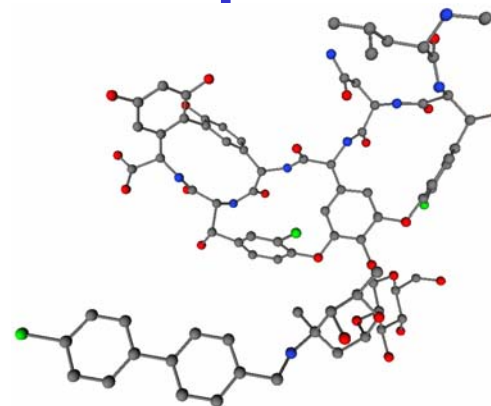
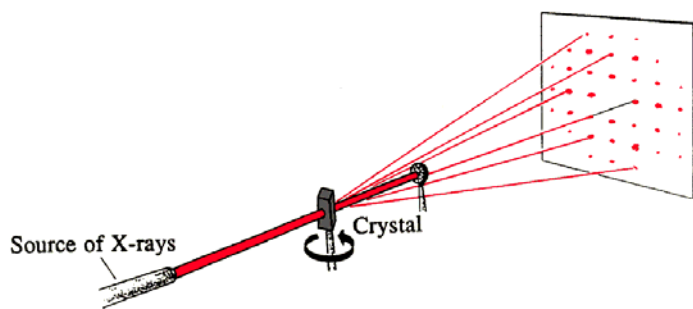
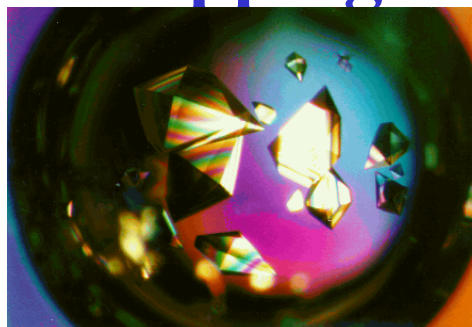
**Molecular Structure Determination
from X-Ray Crystallographic Data**

X-Ray Crystallography

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

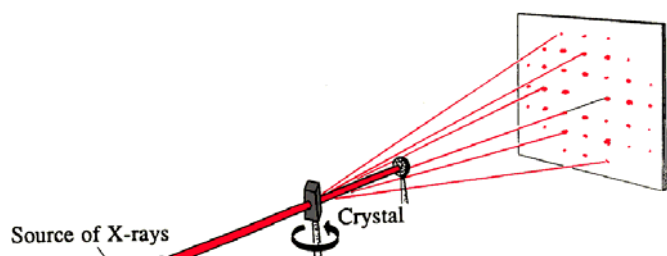
- **Procedure:**

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



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

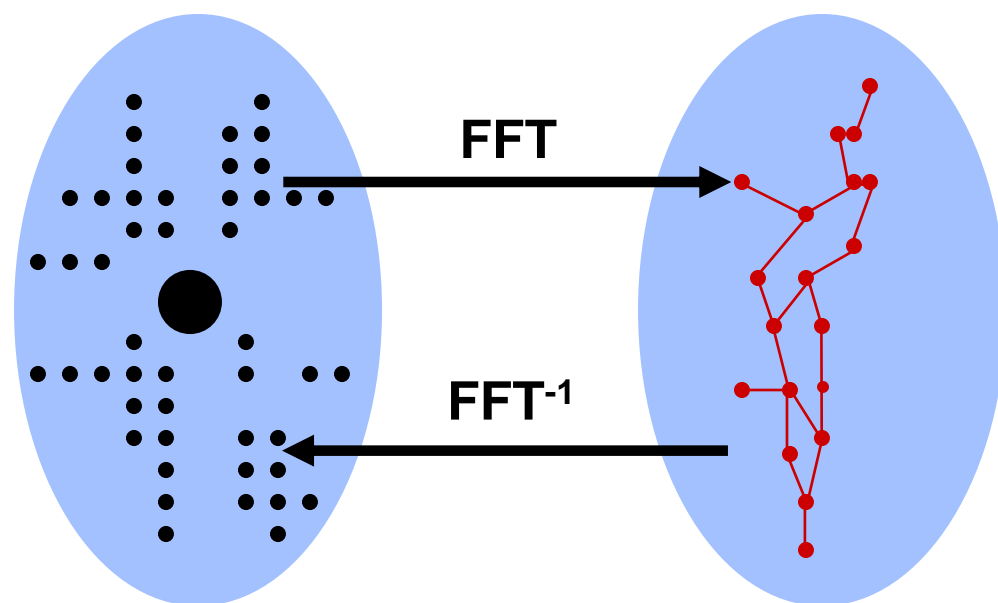
X-Ray Data & Corresponding Molecular Structure



- 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.*

Reciprocal or
“Phase” Space

Real Space

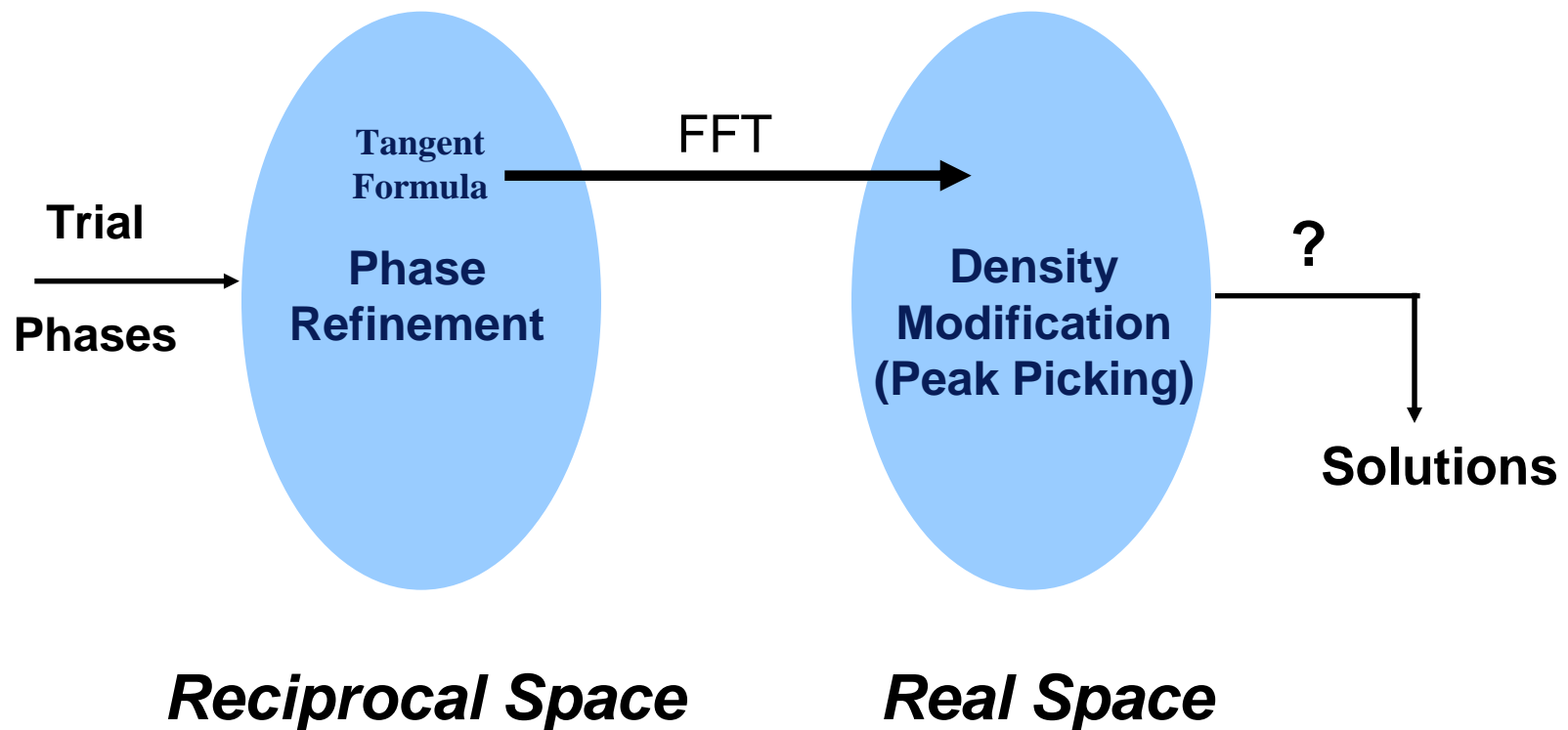


X-Ray Data

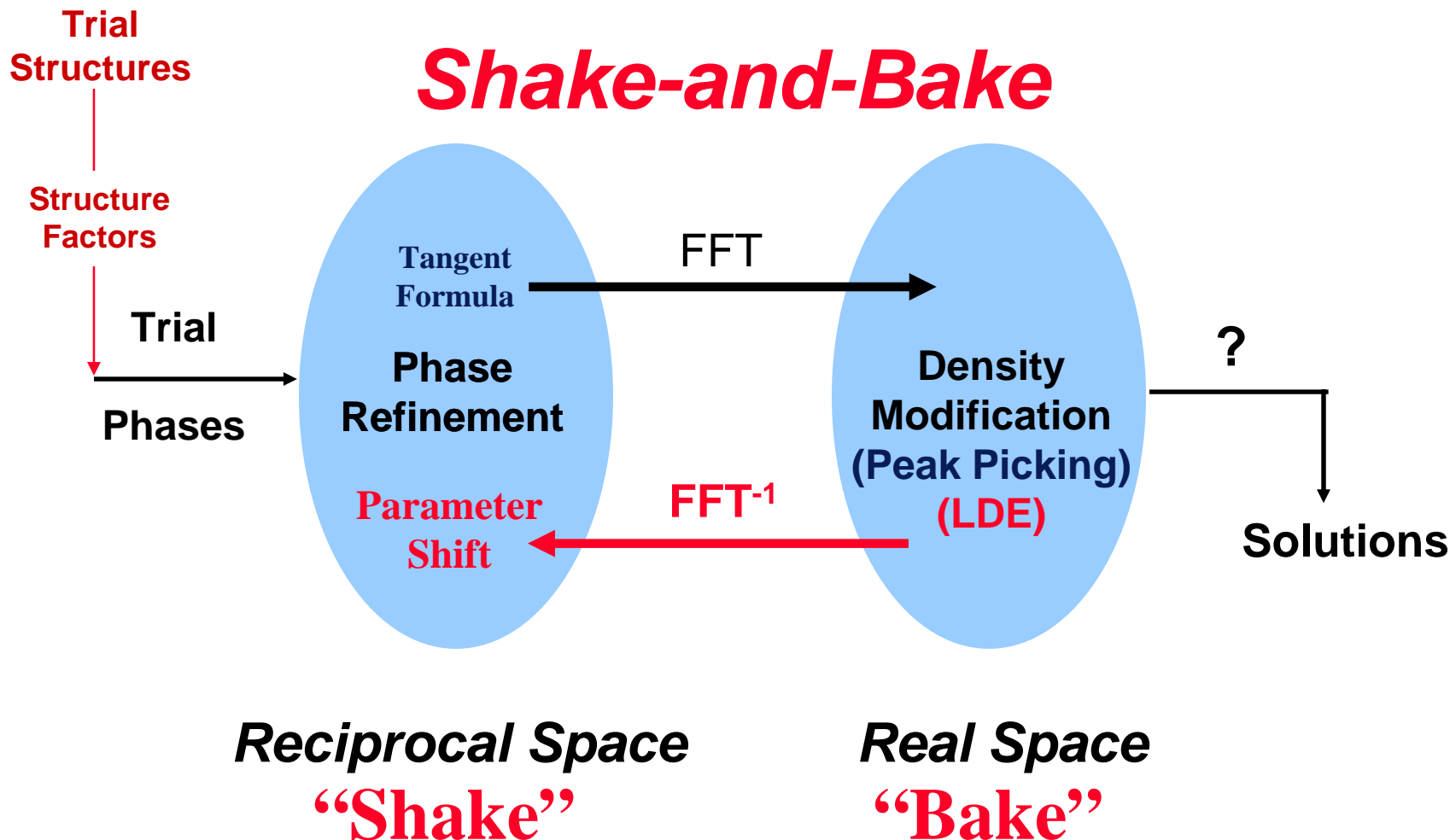
Molecular
Structure



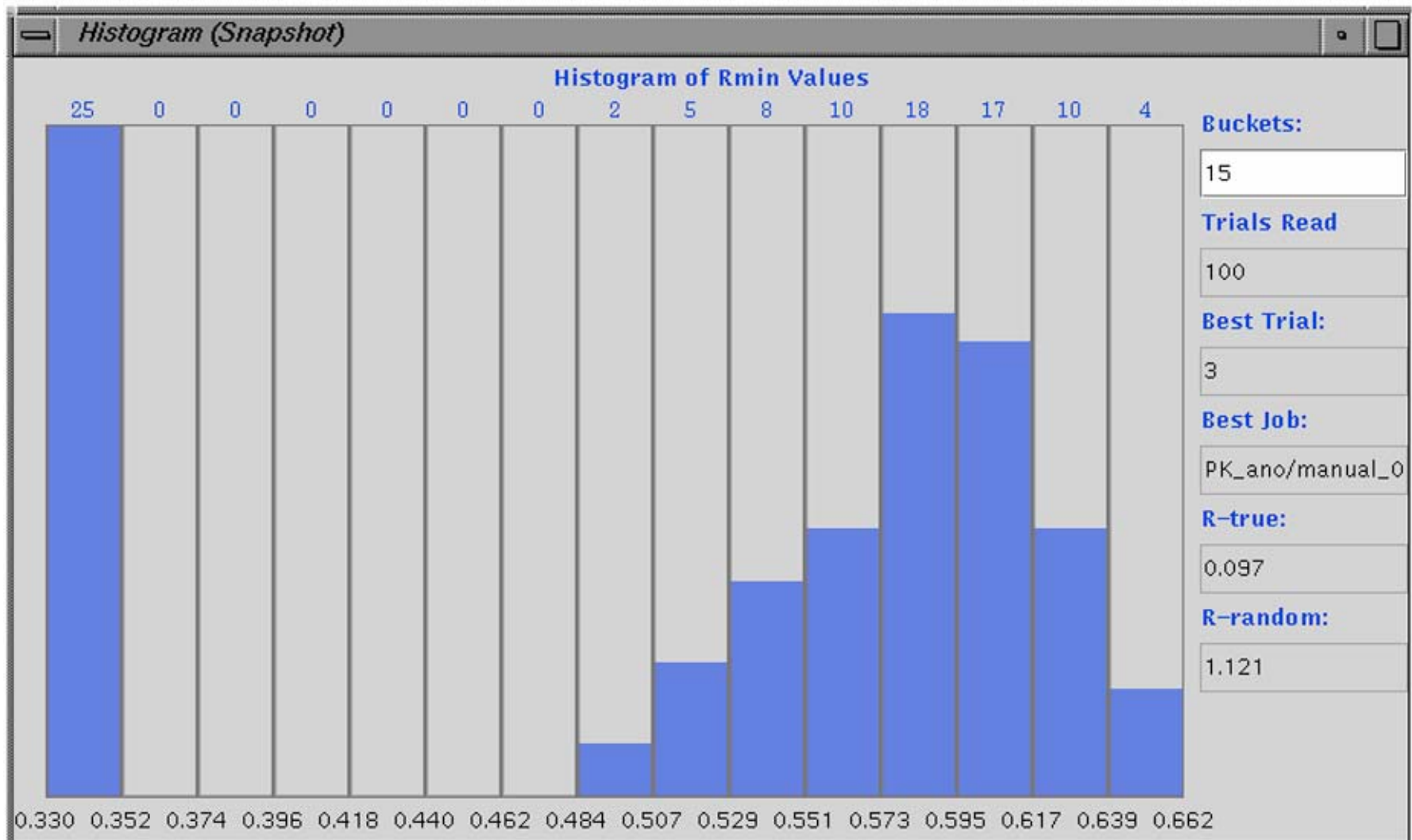
Conventional Direct Methods



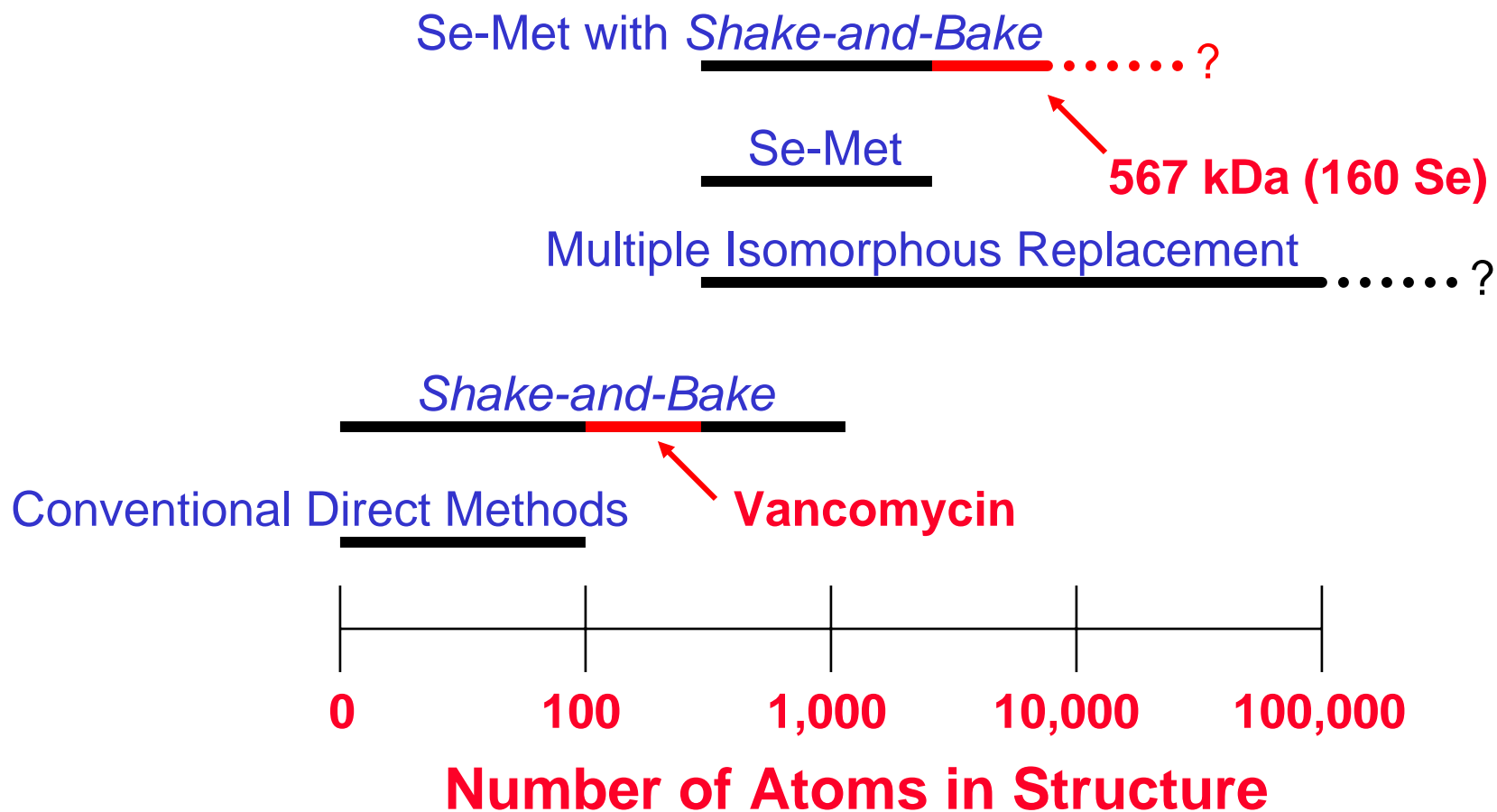
Shake-and-Bake Method: Dual-Space Refinement



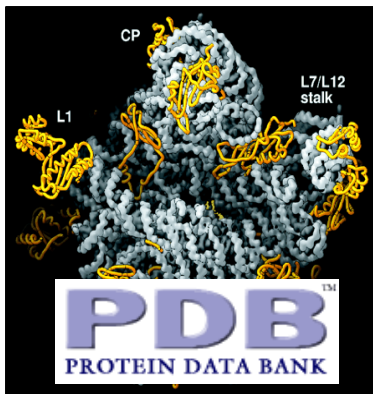
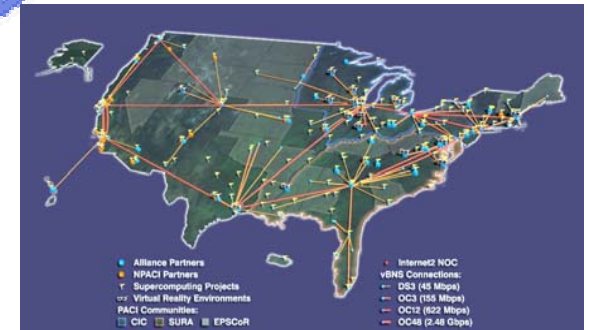
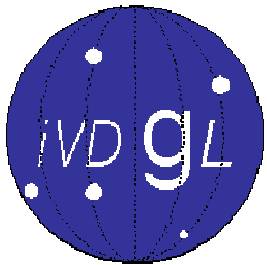
Ph8755: *SnB* Histogram



Phasing and Structure Size



Grid Computing



Asia-Pacific Advanced Network

Advanced
Center for Computational Research
Data
Center

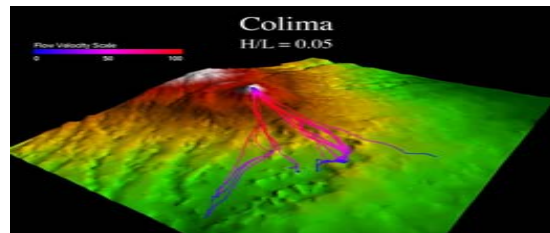


University at Buffalo The State University of New York

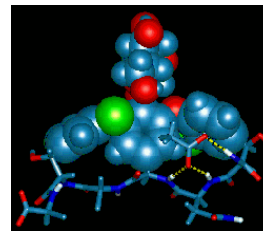
Cyberinfrastructure Laboratory

CI Lab

Grid Computing Overview



Data Acquisition



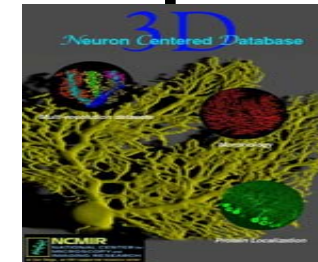
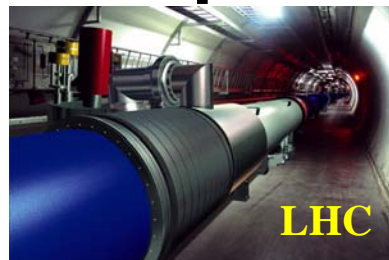
Advanced Visualization



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 Commission)**
 - ❑ 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!



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**



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}



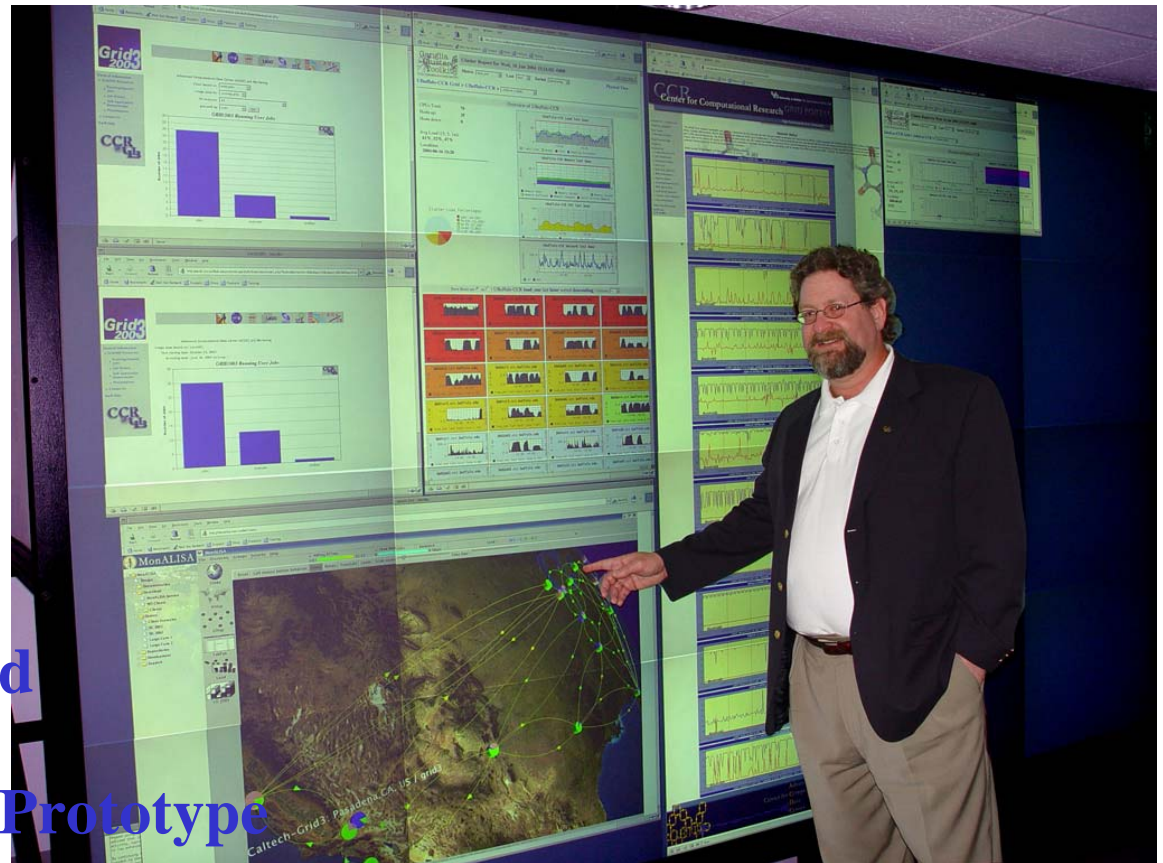
NYS Grid Resources

- **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 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**

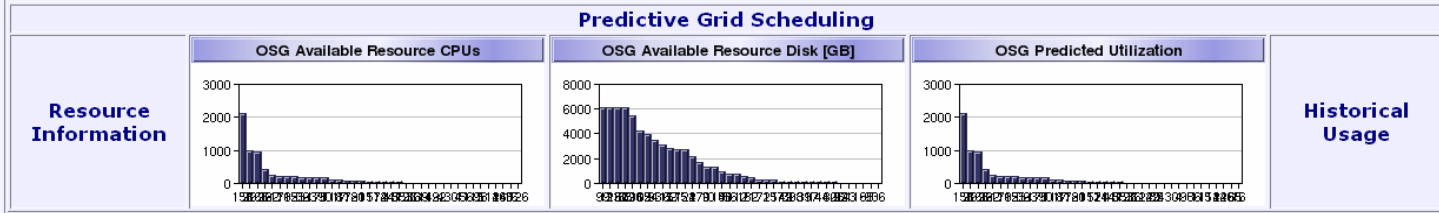
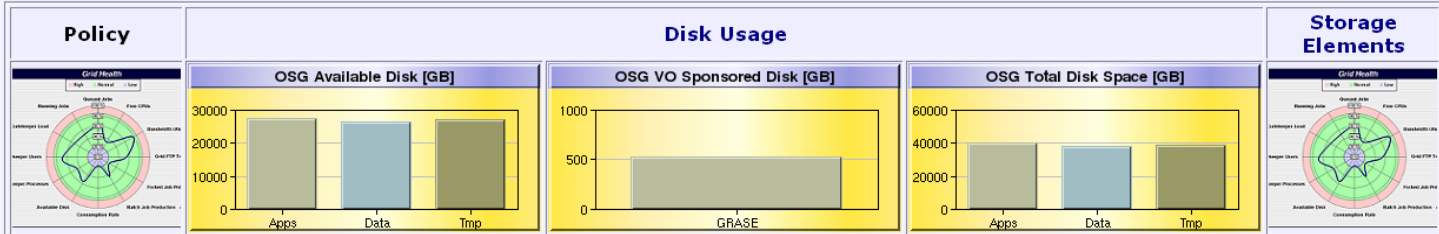
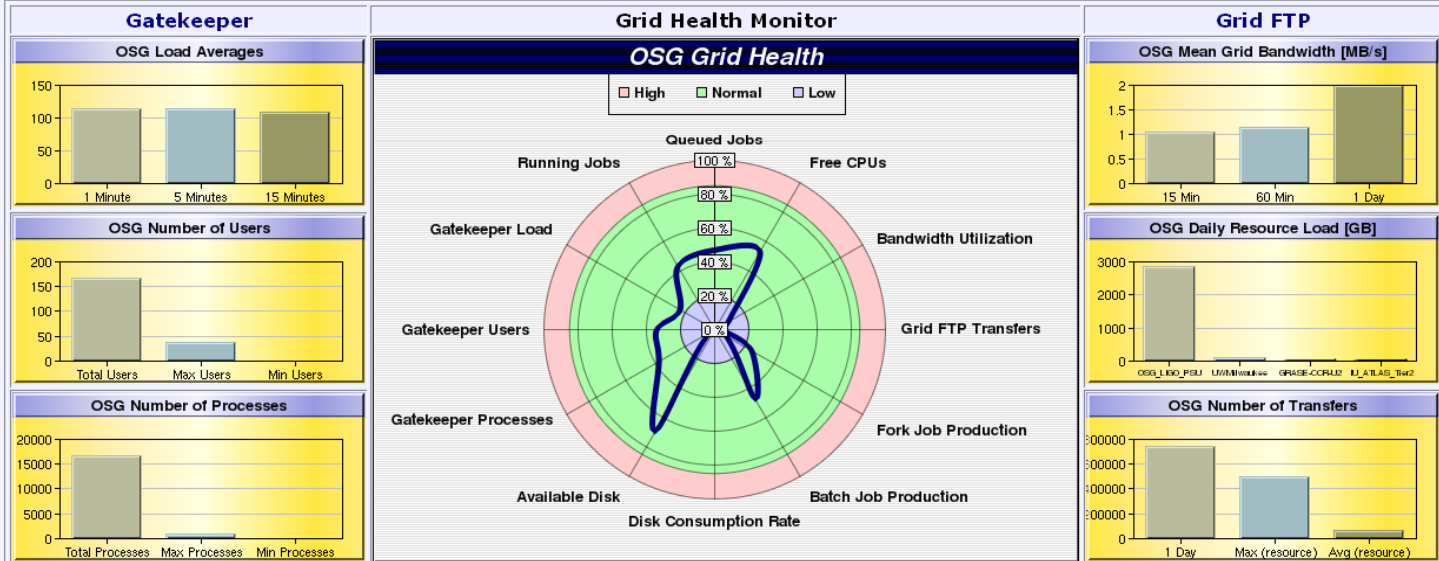
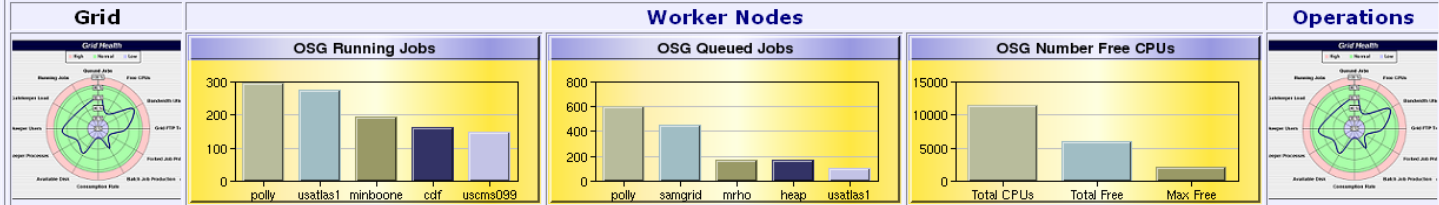


CI Lab

- 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

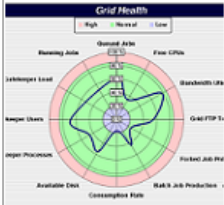
OSG-ITB OSG ACDC TeraGrid Ad-Hoc



Supported by the National Science Foundation and the Department of Energy

CI Lab Grid Monitor: <http://osg.ccr.buffalo.edu/>

Grid



CI LAB OPERATIONS DASHBOARD

OSG-ITB
 OSG
 ACDC
 TeraGrid
 Ad-Hoc

VIRTUAL ORGANIZATIONS

MIS
 GRASE
 OSG
 CDF
 Fermilab
 GADU
 GLOW
 USATLAS
 fMRI
 iVDGL
 nanoHUB

Version:
 All
 OSG-0.2.1
 OSG-0.3.0
 OSG-0.4.0
 OSG-0.4.1

Detailed Service Status

Operations



Site Resource - Service Matrix

Site Resource - Service Matrix

■ No Information
 ■ Pass
 ■ Error
 ■ Fail
 ■ Untested
 ■ Excluded

Production Sites

	Remote Host is Reachable	Running Gatekeeper	Authentication	Hello, World Application Check	Remote Host Uptime	Internet Network Services	Internet Servers Database	Certificate Expiration	Certificate Revocation	Gatekeeper Conf	Apps Directory Conf	Data Directory Write	Tmp Directory Write	CSIF-TP	CSIF-TP Local -> Remote	CSIF-TP Remote -> Local	Grid Tools	Grid Services	Scheduler	MDS	Grid / GLUE Attributes	Infrastructure Version	VDT Version	Grid3 Information	MonAJSA	Ganglia	VO-Specific Tests	
athena.rit.albany.edu	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:46:08	
rommel.cs.binghamton.edu	Pass	Pass	Pass	Pass	Fail	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:45:52	
gridgk01.racf.bnl.gov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:46:05	
gridgk02.racf.bnl.gov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:46:44	
idun.hwi.buffalo.edu	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2007-01-22 14:08:32	
u2-grid.ccr.buffalo.edu	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:46:18	
ctcnysgrid.tc.cornell.edu	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:45:59	
osgc01.grid.sinica.edu.tw	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:47:02	
cms-xen2.fnal.gov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:50:12	
cmsosgce.fnal.gov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:52:45	
fngp-osg.fnal.gov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:55:13	
tam01.fnal.gov	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:46:05	
atlas.iu.edu	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:50:19	
nysgrid11.is.marist.edu	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2007-01-22 15:28:04	
bench.es.its.nyu.edu	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:46:50	
ouhep0.nhn.ou.edu	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:47:20	
grid3.aset.psu.edu	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:50:39	
grid.physics.purdue.edu	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:46:14	
osg.rcac.purdue.edu	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 14:07:04	
stars.if.usp.br	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	Pass	2006-12-13 13:46:33	

CI Lab Operations Dashboard

Predictive Scheduler

- **Build profiles based on statistical analysis of logs of past jobs**
 - Per User/Group
 - Per Resource
- **Use these profiles to predict runtimes of new jobs**
- **Make use of these predictions to determine**
 - Resources to be utilized
 - Availability of Backfill



Grid Administration

The collage displays four screenshots of the CCR Grid Portal interface:

- Grid Site Administration:** Shows a sidebar with navigation options like 'Manage Account', 'Grid General Info', and 'Resources'. The main content area includes sections for 'Organizations (add, edit, delete)', 'Resources (view, refresh, ping, delete, create host certificate)', and 'Globus Administration Reports (machine usage, user access to machines, etc.)'.
- Generate Globus grid-mapfile:** A form for generating a grid-mapfile. It includes fields for 'Optional include file:' (set to /home/griddev/www/grid-mapfile.inc) and 'Optional grid-mapfile path:'. A checkbox for 'Do not stage the file to the grid nodes' is present.
- Create New Database Job:** A form to create a new database job. It includes fields for 'Job Name:', 'Full Path To Script:', 'Accepts Arguments:' (set to No), 'Run Script:' (set to No), and 'Run As User:' (set to admin (Admin, Admin)).
- MDS Resource Update Status:** A table showing the status of various resources. The current time is 16-September-2003 10:58:12.

Resource	Last Updated	Next Update	Status
crosby.ccr.buffalo.edu	16-September-2003 09:15:30	2 minutes	OK
fogerty.ccr.buffalo.edu	16-September-2003 10:45:30	2 minutes	OK
joplin.ccr.buffalo.edu	16-September-2003 10:45:15	2 minutes	OK
mama.ccr.buffalo.edu	16-September-2003 10:45:15	2 minutes	OK
nash.ccr.buffalo.edu	16-September-2003 10:45:15	2 minutes	OK
nexus.hwi.buffalo.edu	16-September-2003 10:45:20	2 minutes	OK
yardbirds.ccr.buffalo.edu	16-September-2003 10:45:13	2 minutes	OK
young.ccr.buffalo.edu	16-September-2003 10:45:27	2 minutes	OK



ACDC-Grid Data Grid Functionality

- **Basic file management functions are accessible via a platform-independent web interface.**
- **User-friendly menus/interface.**
- **File Upload/Download to/from the Data Grid Portal.**
- **Simple Web-based file editor.**
- **Efficient search utility.**
- **Logical display of files (user/ group/ public).**
- **Ability to logically display files based on metadata (file name, size, modification date, etc.)**



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 My 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

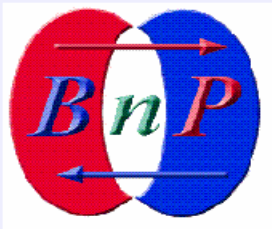
The **Cyberinfrastructure Laboratory**, in conjunction with the **Center for Computational Research**, has created an integrated Data and Computational Grid. This site is devoted to a Grid Portal that provides access to applications that can be run on a variety of grids. A related site contains a **Grid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



Software : BnP
Field : Protein crystal structure determination

Startup Screen for CI Lab Grid Job Submission

Expand All Collapse All

PORTAL LOGOUT

User Tools

» Manage Account

Grid General Info

Projects

Computational Grid

» Job Submission

» Job/Queue Status

» MDS Information

» Network Status

» Running/Queued Jobs

» PBS Job History

» NYS Grid

» Condor Flock Statistics

Data Grid

Education/Outreach

Staff Only

CCR HOME

Printer Friendly

Software → Template → General Information → Detailed Information → Job Definition → Review → Execution Scenario

Advanced Computational Data Center Grid Job Submission Instructions

The grid-enabling application templates used on the ACDC-Grid are created from the application developers grid user profiles that contain the users standard information uid, name, organization, address, etc., and more specific information such as group id and access level information for each of grid-enabled applications. This information is stored in a database for each of the grid-enabled applications and can be accessed through selected queries throughout the ACDC-Grid Web Portal.

Additionally, each grid-enabled scientific application profile contains information about specific execution parameters, required data files, optional data files, computational requirements, etc. and statistics on application historical ACDC-Grid jobs for predictive runtime estimates. MySQL provides the speed and reliability required for this task and it is currently being used as the ACDC-Grid Web Portal database provider.

The grid-enabled versions of many well-defined scientific and engineering applications have very similar general requirements and core functionality that are require for execution in the ACDC-Grid environment. We have identified that sequentially defining milestones for the grid user to complete intuitively guides them through the application workflow.

- Software Application:** Grid user chooses a grid-enabled software application.
- Template:** Grid user selects the required and/or optional data files from the ACDC Data Grid. User defined computational requirements are input or a template defined computational requirement runtime estimate is selected.
- Job Definition:** Grid user defines application specific runtime parameters or accepts default template parameter definitions.
- Review:** Grid user accepts the template complete job definition workflow or corrects any part of job definition.
- Execution Scenario:** The grid user has the ability to input an execution scenario or select a ACDC-Grid determined template defined execution scenario.
- Grid Job Status:** The grid user can view specific grid job completion status, grid job current state (COMPLETE, RUNNING, QUEUED, BLOCKED, FAILED, ETC.), detailed information on all running or queued grid jobs and grid-enabled application specific intermediate and post processing grid job graphics, plots and tables.

Each item of the job definition workflow is then stored in the ACDC-Grid Web Portal database so the grid user may use/modify any previously created workflow in creating new job definitions. The job definitions can also be accessed via batch script files for executing hundreds of similar workflows in an automated fashion. For example, a grid user would first define/save a relatively generic job workflow template for the grid-enabled application and then use the batch script capabilities to change the job definition workflow data files or application parameters and execute a series of new grid jobs.

Continue

Reset Sequence

Reset Current Stage

Cancel

Instructions and Description for Running a Job on ACDC-Grid

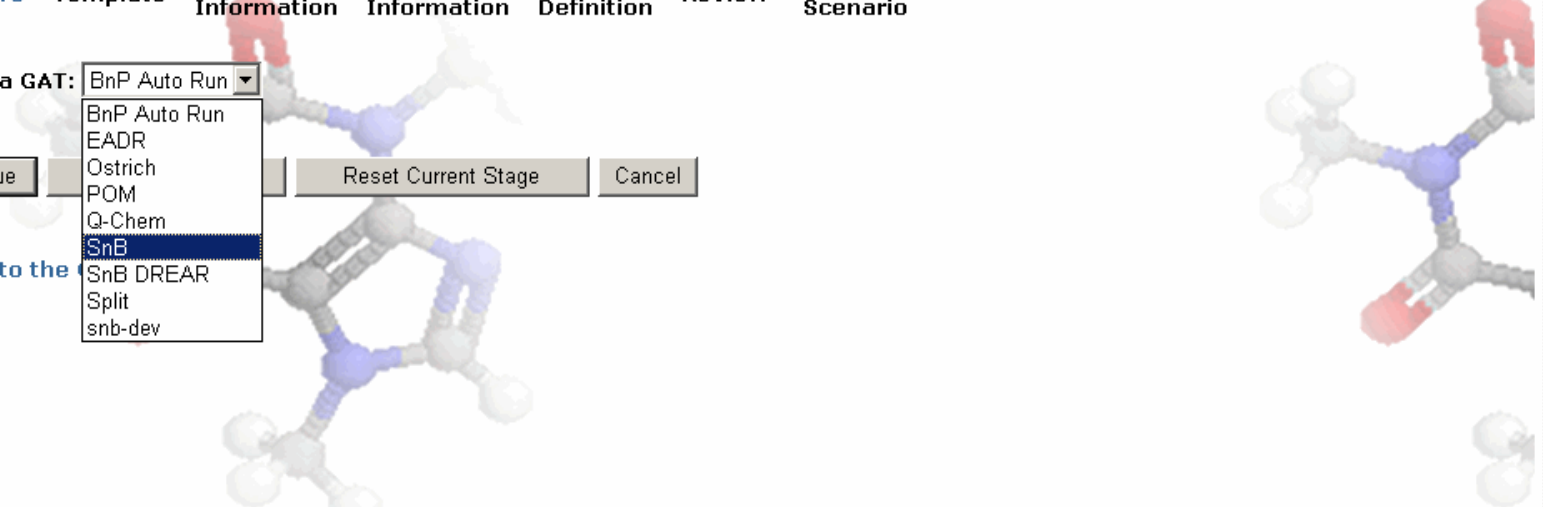
Expand All Collapse All
PORTAL LOGOUT
User Tools
» Manage Account
Grid General Info
Projects
Computational Grid
» Job Submission
» Job/Queue Status
» MDS Information
» Network Status
» Running/Queued Jobs
» PBS Job History
» NYS Grid
» Condor Flock Statistics
Data Grid
Education/Outreach
Staff Only
CCR HOME
Printer Friendly

Software → Template → General Information → Detailed Information → Job Definition → Review → Execution Scenario

Select a GAT:
BnP Auto Run
EADR
Ostrich
POM
Q-Chem
SnB
SnB DREAR
Split
snb-dev

Continue Reset Current Stage Cancel

Return to the



Software Package Selection

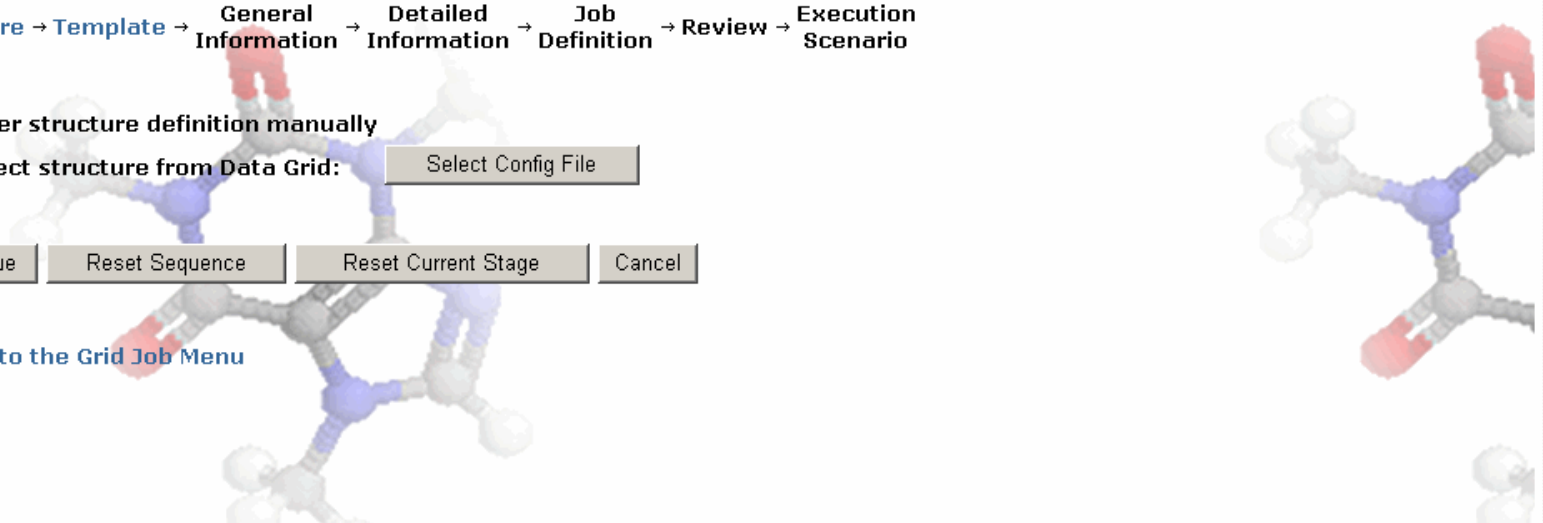
Expand All Collapse All
PORTAL LOGOUT
User Tools
» Manage Account
Grid General Info
Projects
Computational Grid
» Job Submission
» Job/Queue Status
» MDS Information
» Network Status
» Running/Queued Jobs
» PBS Job History
» NYS Grid
» Condor Flock Statistics
Data Grid
Education/Outreach
Staff Only
CCR HOME
Printer Friendly

Software → Template → General Information → Detailed Information → Job Definition → Review → Execution Scenario

Enter structure definition manually
 Select structure from Data Grid:

Continue Reset Sequence Reset Current Stage Cancel

Return to the Grid Job Menu

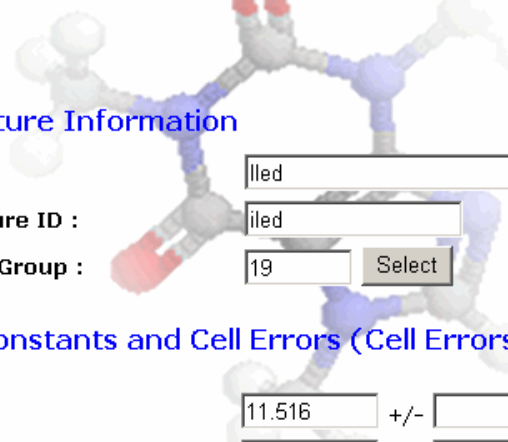
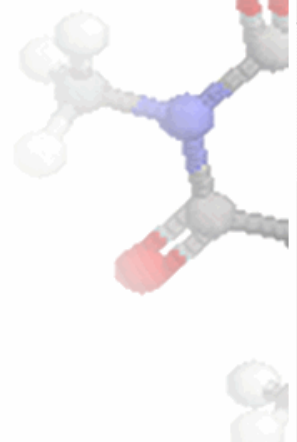


Full Structure / Substructure Template Selection

USER TOOLS

- » Manage Account
- Grid General Info
- Projects
- Computational Grid
- » Job Submission
- » Job/Queue Status
- » MDS Information
- » Network Status
- » Running/Queued Jobs
- » PBS Job History
- » NYS Grid
- » Condor Flock Statistics
- Data Grid
- Education/Outreach
- Staff Only
- CCR HOME
- Printer Friendly

General Information

Structure Information

Title:

Structure ID :

Space Group :

Cell Constants and Cell Errors (Cell Errors optional)

A: +/-

B: +/-

C: +/-

Alpha: +/-

Beta: +/-

Gamma: +/-

Native Asymmetric Unit Contents

No Residues (Optional):

ASU Contents : (examples: C6H12O6 OR C6 H12 O6)

Initial Data Sets

Select dataset to delete	○
Datasets	Dataset 1
Name (8 chars max):	<input type="text" value="iledhkl"/>

Default Parameters Based on Template

User Tools

» Manage Account

Grid General Info

Projects

Computational Grid

» Job Submission

» Job/Queue Status

» MDS Information

» Network Status

» Running/Queued Jobs

» PBS Job History

» NYS Grid

» Condor Flock Statistics

Data Grid

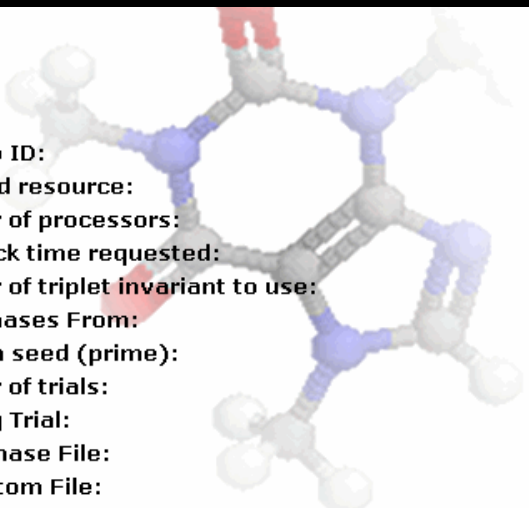
Education/Outreach

Staff Only

CCR HOME

Printer Friendly

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)**

Details for Grid Job 447 - iledhkl

Job Detail Information

Status: **RUNNING**

Rmin Min: 0.344 Rmin Max: 0.56

Last Updated: 15-Mar-2005 10:22:00

Total Trials: 1000

Complete Trials: 285

Resource: clearwater.ccr.buffalo.edu Processors: 5

Best Trial Number: 34

Best Trial Rmin: 0.344

Trial Summary

Grid Job 447 Trial Summary

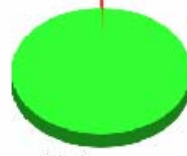
Number of Trials Complete: 285 (28.5%)



Walltime Summary

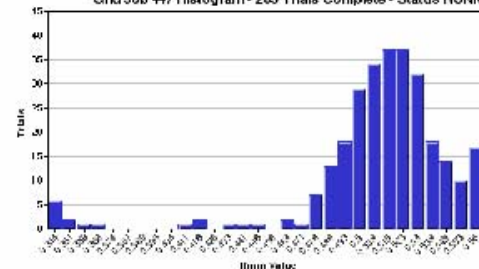
Grid Job 447 Walltime Summary

Walltime Consumed: 2 (0.3%)



Grid Job Trial Histogram

Grid Job 447 Histogram - 285 Trials Complete - Status RUNNING



Click on image for enlarged view.

Graphical Representation of Intermediate Job Status

Expand All Collapse All

PORTAL LOGOUT

User Tools

» Manage Account

Grid General Info

Projects

Computational Grid

» Job Submission

» Job/Queue Status

» MDS Information

» Network Status

» Running/Queued Jobs

» PBS Job History

» NYS Grid

» Condor Flock Statistics

Data Grid

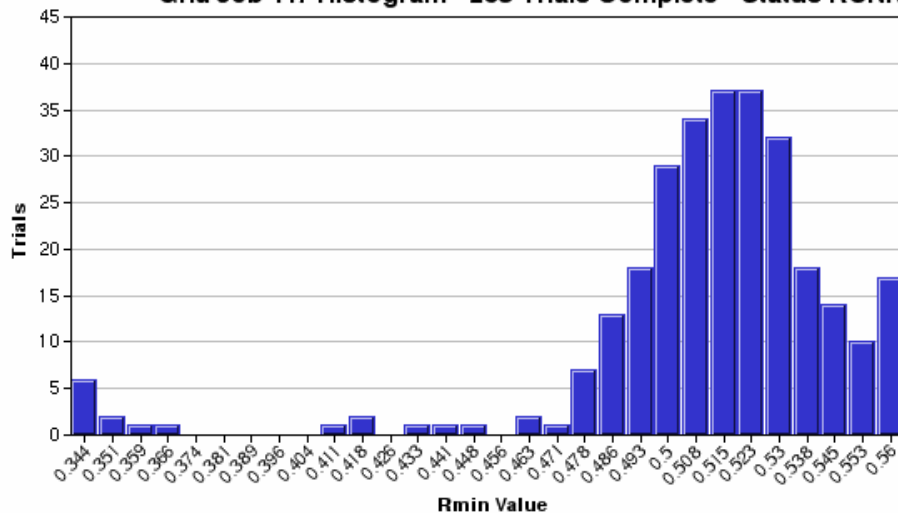
Education/Outreach

Staff Only

CCR HOME

Printer Friendly

Grid Job 447 Histogram - 285 Trials Complete - Status RUNNING



Histogram of Completed Trial Structures

Expand All Collapse All

PORTAL LOGOUT

User Tools

» Manage Account

Grid General Info

Projects

Computational Grid

» Job Submission

» Job/Queue Status

» MDS Information

» Network Status

» Running/Queued Jobs

» PBS Job History

» NYS Grid

» Condor Flock Statistics

Data Grid

Education/Outreach

Staff Only

CCR HOME

Printer Friendly

Grid Job Status

15-Mar-2005 10:23:49

Job Filter Criteria

Show GATs

BnP Auto Run
EADR
Ostrich
POM
Q-Chem
SnB
SnB DREAR

Job State

DEFINITION
STAGING
QUEUED
RUNNING
UPLOADING
COMPLETE
INCOMPLETE

Sort By

Job Id
Job Name
Resource
Num Procs
Status
Percent Complete
Last Update

Descending

Ascending

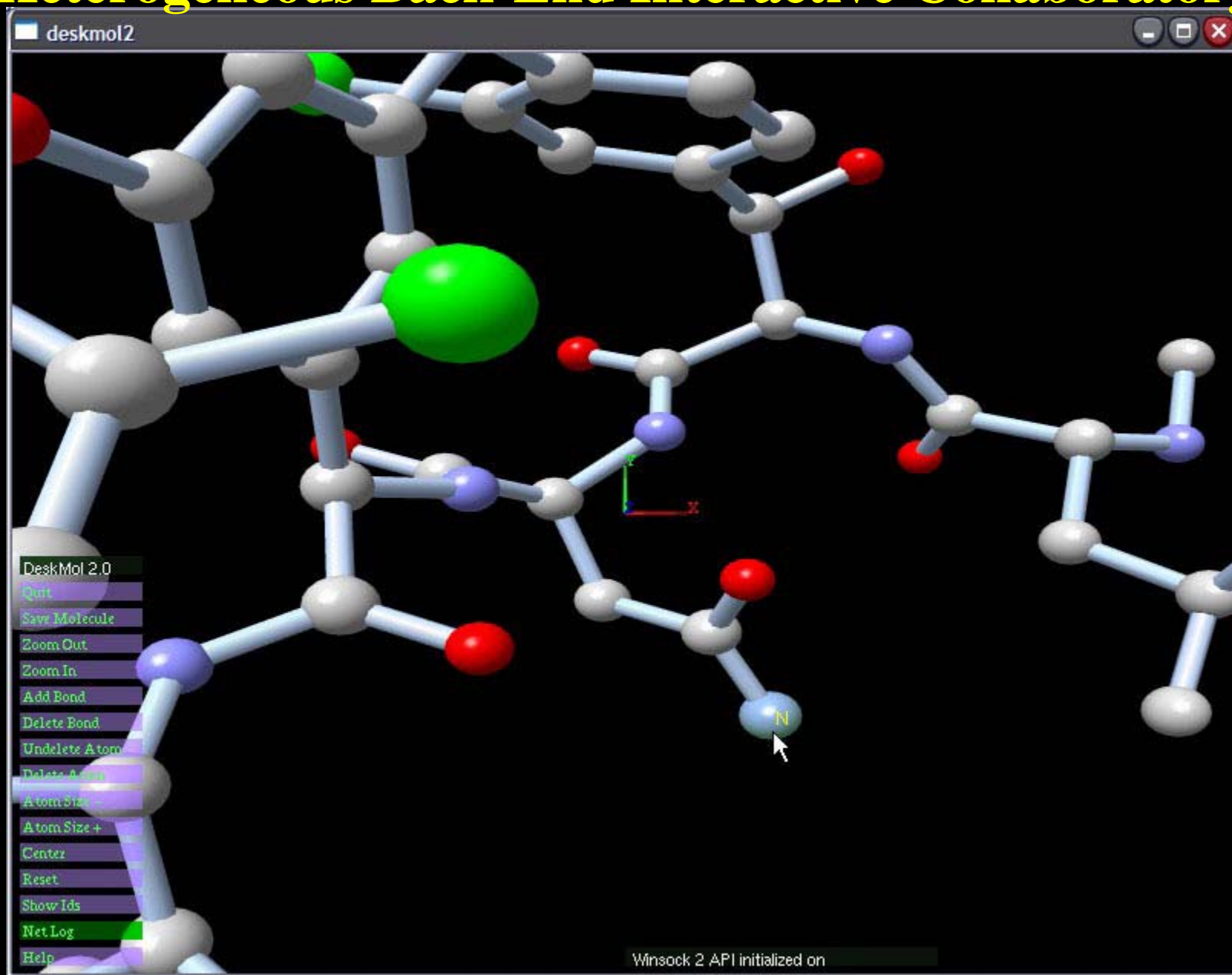
Filter Job List

SnB

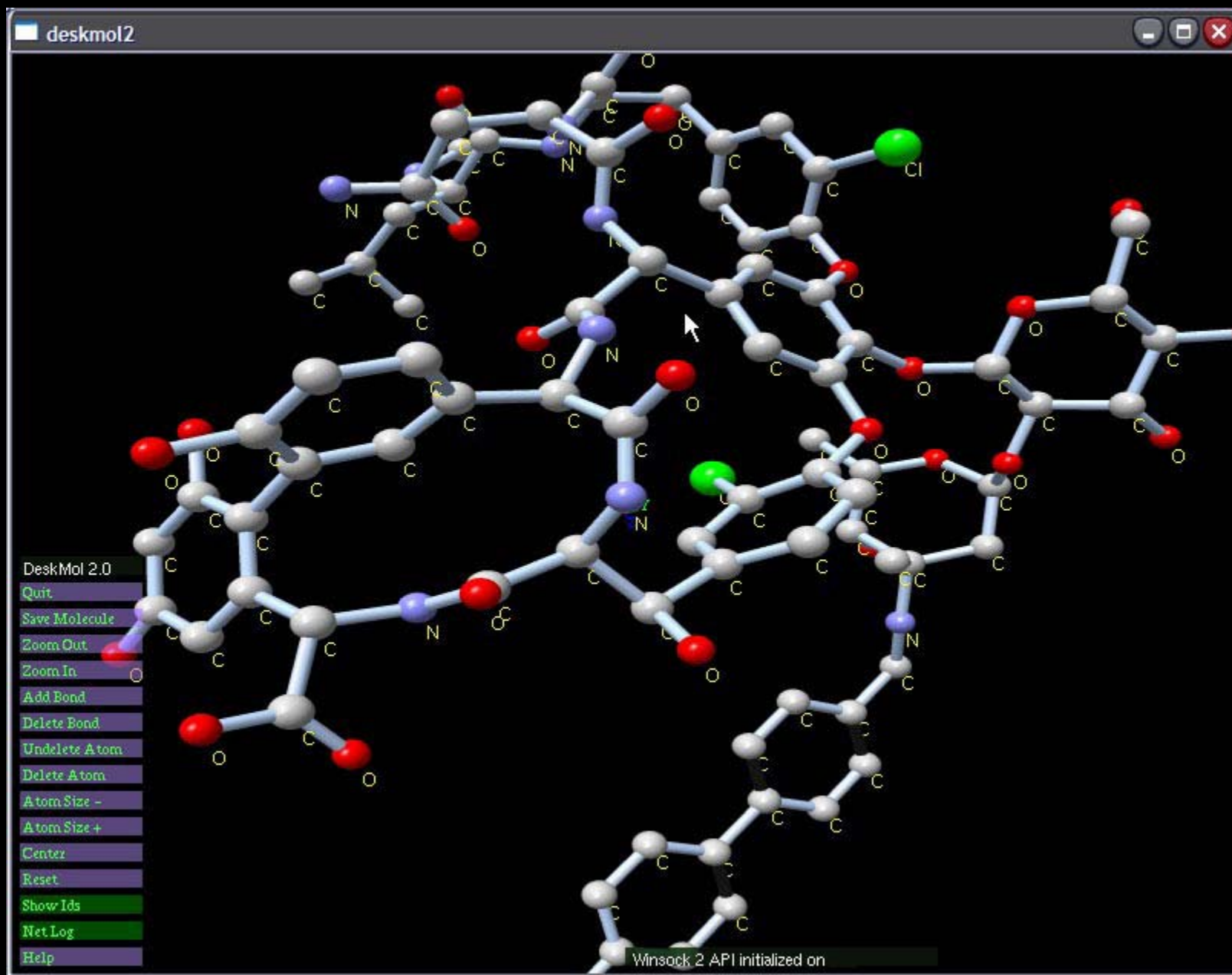
Job Id	Job Name	Resource	Num Procs	Status	Percent Complete	Last Update	Cancel Job	Drilldown
447	iledhkl	clearwater.ccr.buffalo.edu	5	RUNNING	28.5	15-Mar-2005 10:22:00	<input type="checkbox"/>	
446	trilys	clearwater.ccr.buffalo.edu	10	RUNNING	1	15-Mar-2005 10:22:00	<input type="checkbox"/>	
444	64chkl	nash.ccr.buffalo.edu	3	COMPLETE	100	14-Mar-2005 22:00:01		
443	trilys	clearwater.ccr.buffalo.edu	10	COMPLETE	100	10-Mar-2005 22:48:00		
442	pr435hkl	nash.ccr.buffalo.edu	5	COMPLETE	100	10-Mar-2005 17:26:01		
441	vancohkl	clearwater.ccr.buffalo.edu	10	COMPLETE	100	10-Mar-2005 18:08:01		
434	16chkl	clearwater.ccr.buffalo.edu	5	COMPLETE	100	10-Mar-2005 14:42:01		
433	16chkl	clearwater.ccr.buffalo.edu	5	COMPLETE	100	10-Mar-2005 14:38:01		

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