HPC, Computational Science & Engineering, *Shake-and-Bake*, and 21st Century Academia

Russ Miller

Cyberinfrastructure Lab

The State University of New York at Buffalo

NSF, NIH, DOE, NIMA, NYS, HP

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

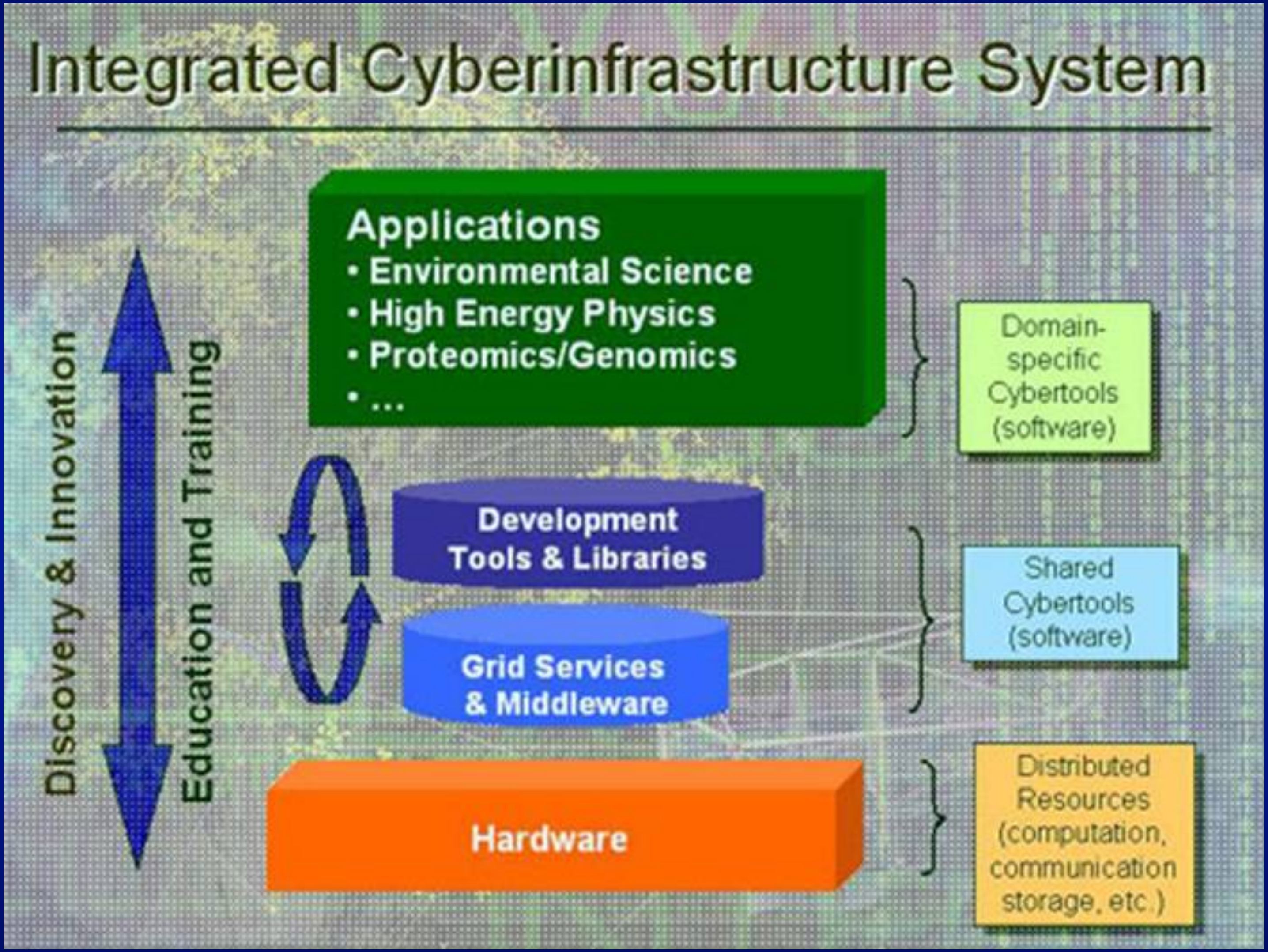
Academia in the 21st Century: High-Level View

- 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

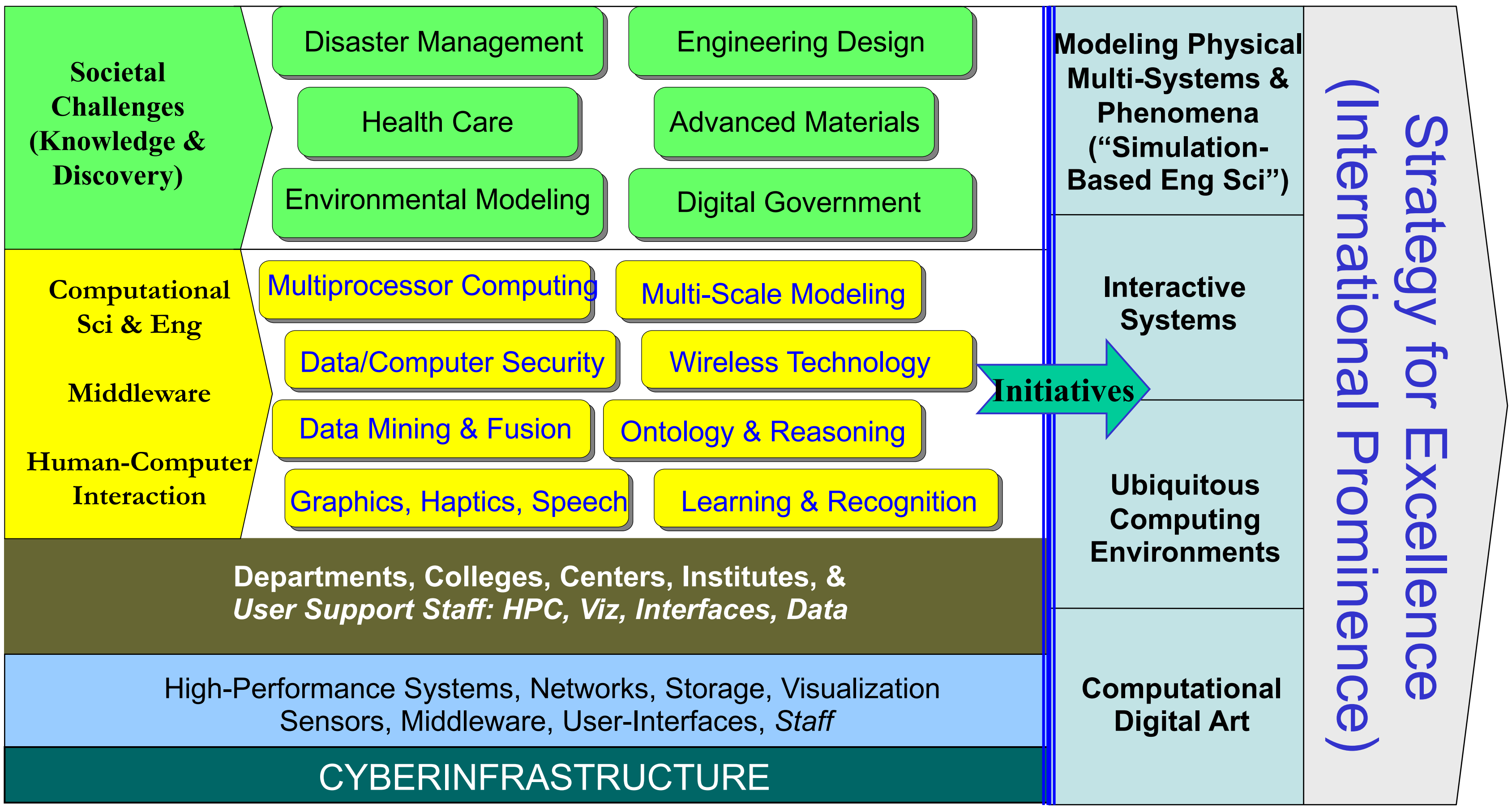
- Create links between enabling technologists and disciplinary users
- Improve efficiency of knowledge-driven applications in myriad disciplines
  - New Techniques
  - New Algorithms
  - New Interactions (people & systems)
- Support HPC infrastructure, research, and applications
- Deliver high-end cyberinfrastructure 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)

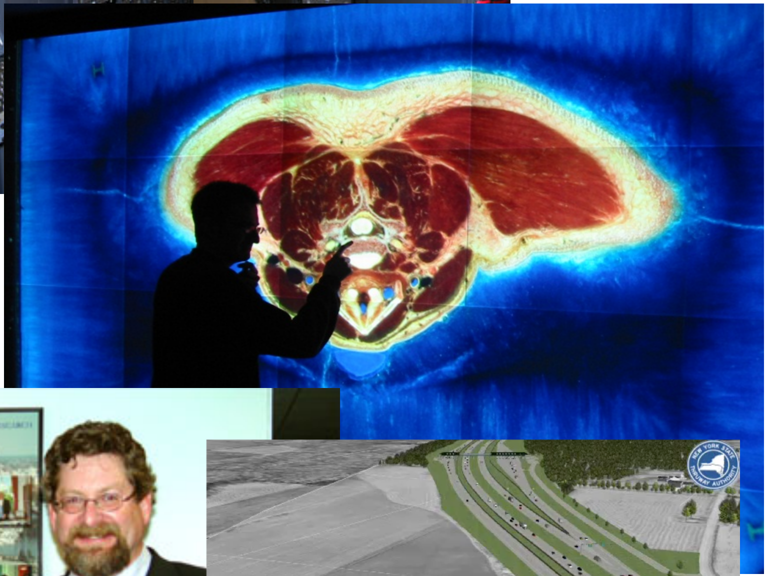


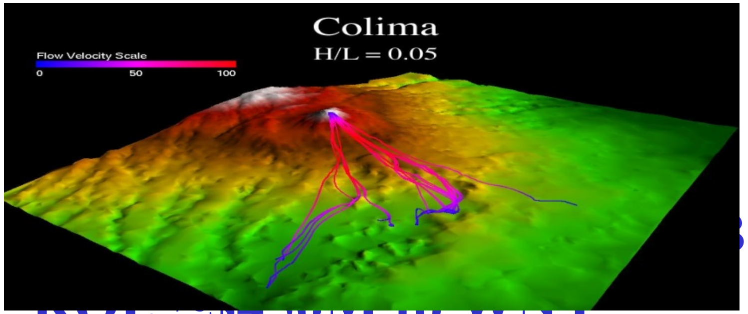
Academic Computing Initiative: Organization

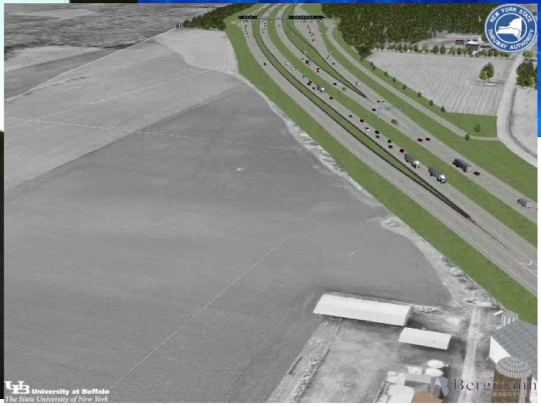
- Must be Pervasive Across the Entire University
- Must Remove Barriers
- Groups Must Interact
  - Research Groups
  - Support Staff
  - Students
  - Departments
  - Colleges
- Issues
  - Tenure & Promotion
  - University vs Colleges vs Departments vs Faculty vs Centers/Institutes vs Degrees vs Courses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

Center for Computational Research (CCR)

- Founding Director (1998-2006)
- Facts & Figures
  - Top Academic HPC Center in World
  - Top 25 HPC System
  - Massive High-End Storage
  - Significant Visualization
  - Special-Purpose Systems
  - ~30 FTEs Staff
  - 140 Projects Annually
- EOT





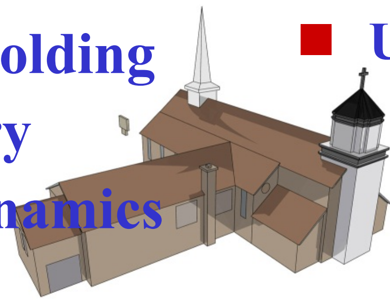


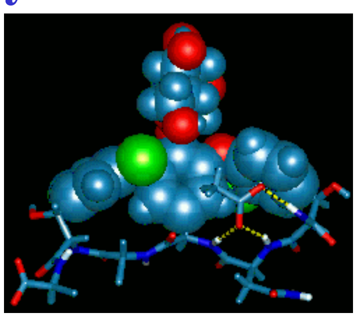
CCR Highlights (1998-2006)

- 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

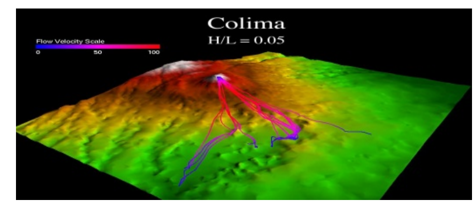
CCR Research & Projects

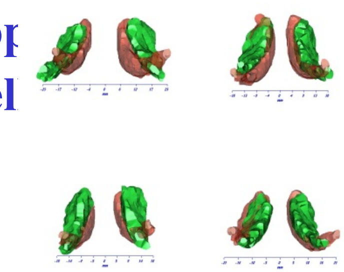
- Archaeology
- Bioinformatics/Protein Folding
- Computational Chemistry
- Computational Fluid Dynamics
- Data Mining/Database
- Earthquake Engineering
- Environ Modeling & Simulation
- Grid Computing
- Molecular Structure Determination
- Physics





- 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 Microscop
  - Crystallization Wel
  - 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

Real-time Simulation

- Key Receptor Sites
- Multiple Viewpoints
- Fully Interactive
- Aerial Photography

Animation & Simulation

Rendered Scenes

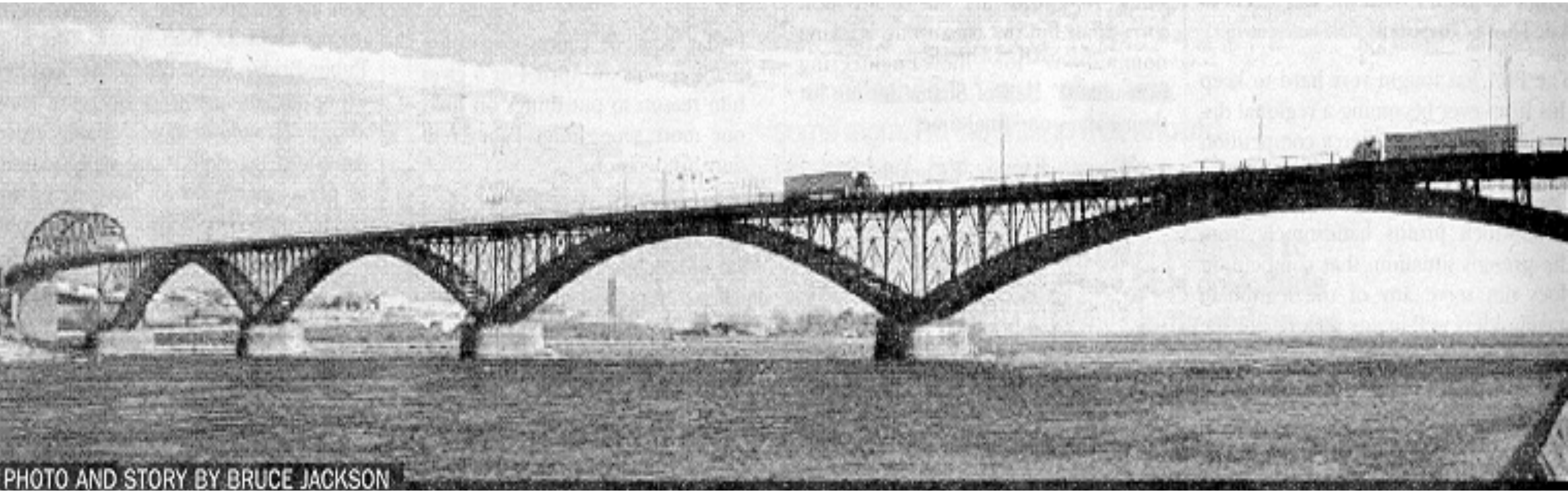
Visualization in Planning Studies

Williamsville Toll Barrier Improvement Project

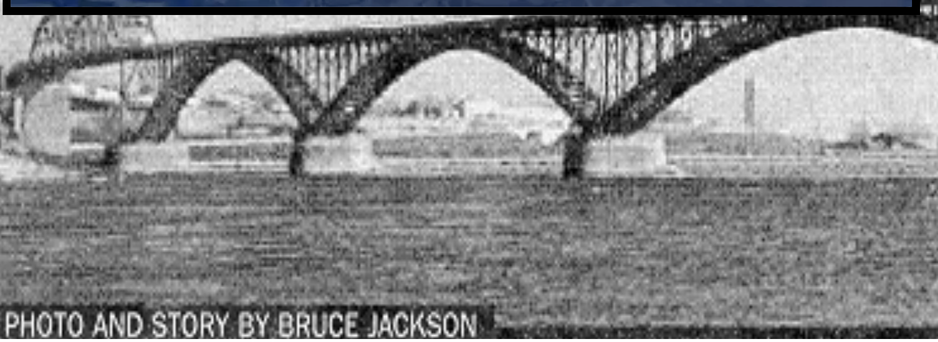
Initial Photo Match incorporating real and computer-generated components

Peace Bridge Visualization: Animation & Simulation

- International Crossing
- The Problem
  - 75 year old bridge
  - 3 lanes – poor capacity
  - Existing US plaza: small and poor design
- Proposed Options
  - Relocate US plaza
  - Build a 3-lane companion span & rehab existing bridge
  - Build a six lane signature span



Peace Bridge Visualization: Animation & Simulation

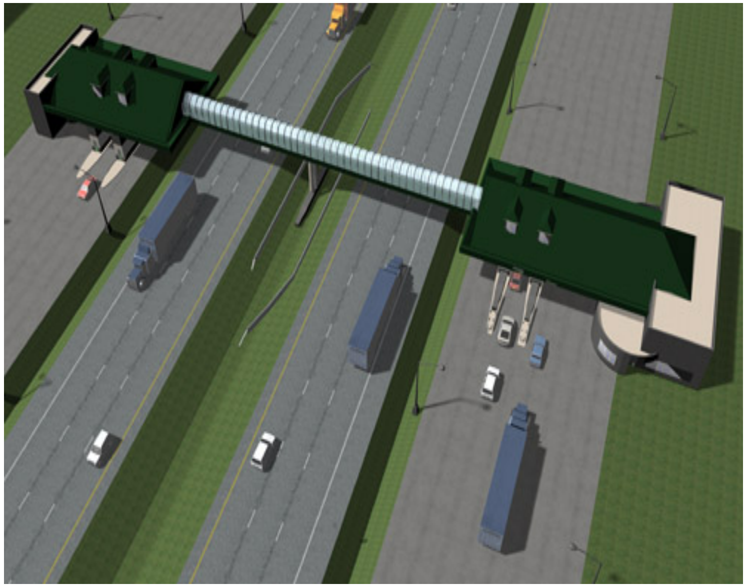


- **Proposed Options**
  - Relocate US plaza
  - Build a 3-lane companion span & rehab existing bridge

Thruway HOT Lanes Animation

Urban Modeling & Visualization

- Peace Bridge Gateway Improvement Project
- Olmsted Park Conservancy
- Williamsville Toll Barrier Relocation
- Buffalo Niagara Medical Campus



M. Innus, A. Koniak, A. Levesque, T. Furlani

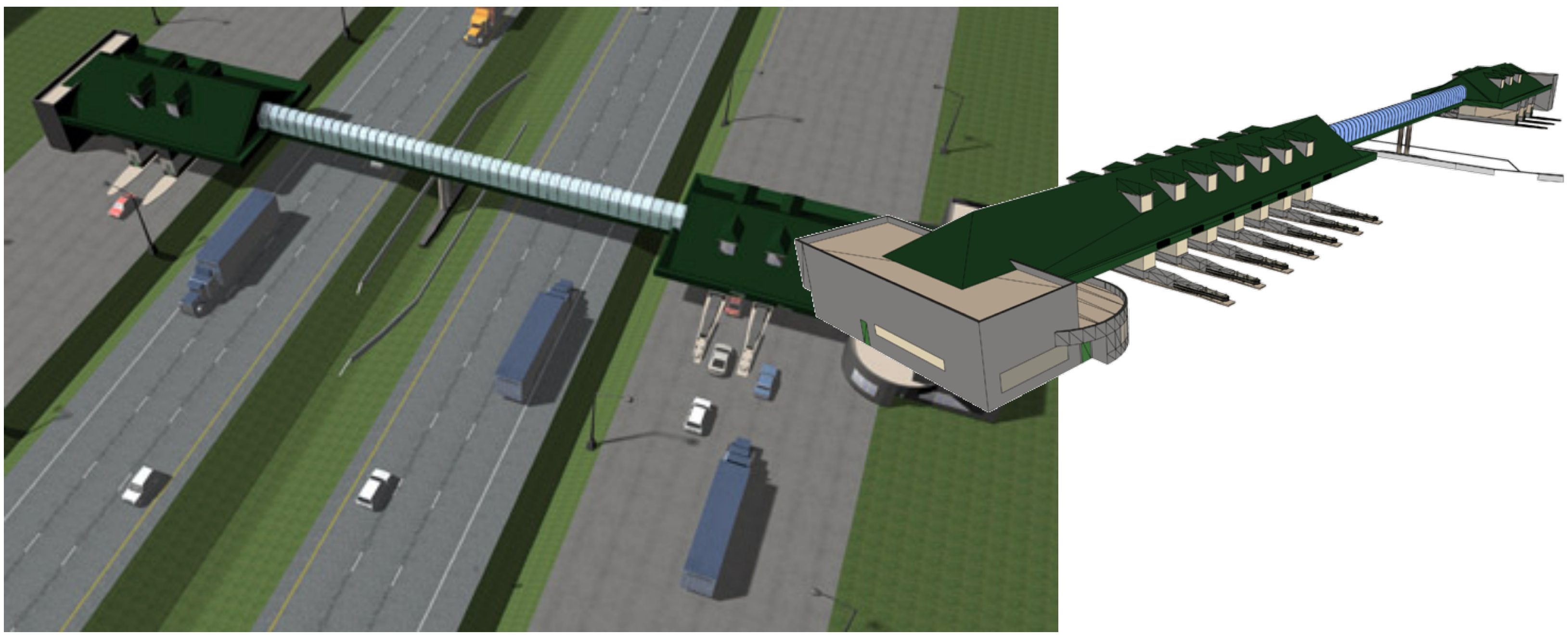
CCR Model Development

- *StreetScenes®* is a Virtual Reality (VR) software solution for 3D interactive visualization of surface traffic.
- Import data from most traffic simulation packages
  - Corsim
  - Synchro
  - Vissim

H. Bucher

Urban Modeling & Visualization

- High Speed EZPass
- Planning tool for NYS Thruway Authority
- Visualization of real traffic data
- Interactive model for public meetings and demonstrations

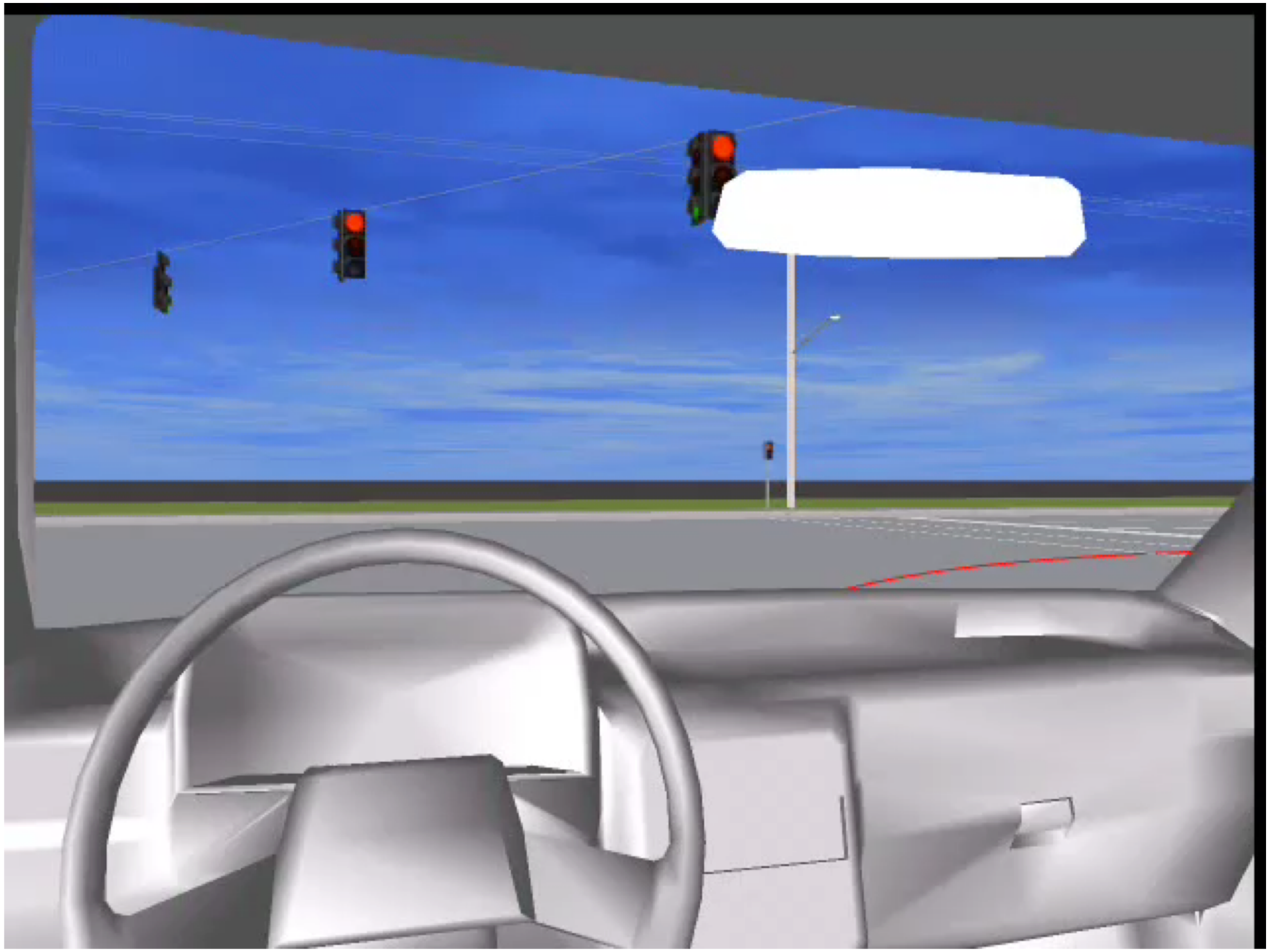


M. Innus, A. Koniak, A. Levesque, T. Furlani

Accident Reconstruction

The Accident

Accident Animation (Driver's View)



StreetScenes® 3D Traffic Simulation

- *StreetScenes®* is a Virtual Reality (VR) software solution for 3D visualization of surface traffic
- 3D model of proposed soccer stadium in Rochester
- Used *StreetScenes®* to import output file from Synchro traffic simulation

MTV

IBC Digital & CCR

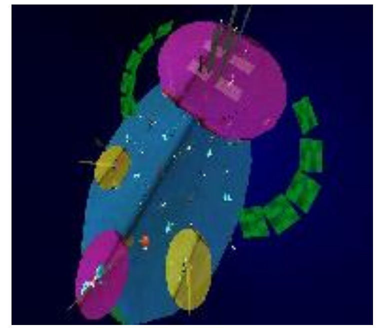
Song: I'm OK (I Promise)
Band: Chemical Romance
Gaming Environment: Death Jr.

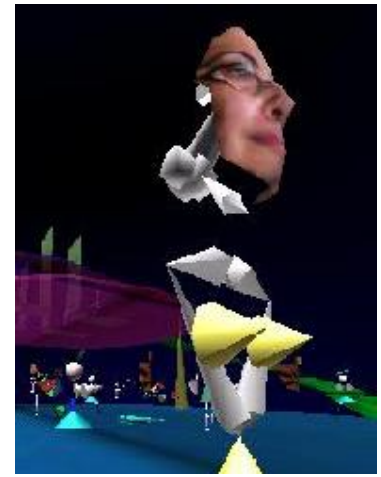
Public Forum

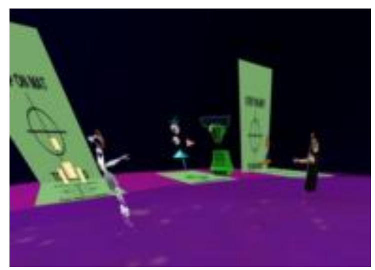
Virtual Reality

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



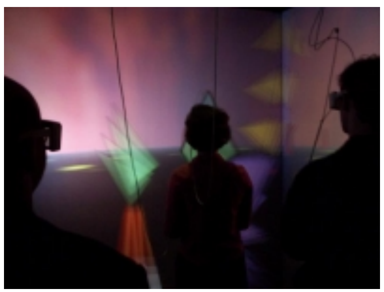




J. Anstey

The Thing Growing

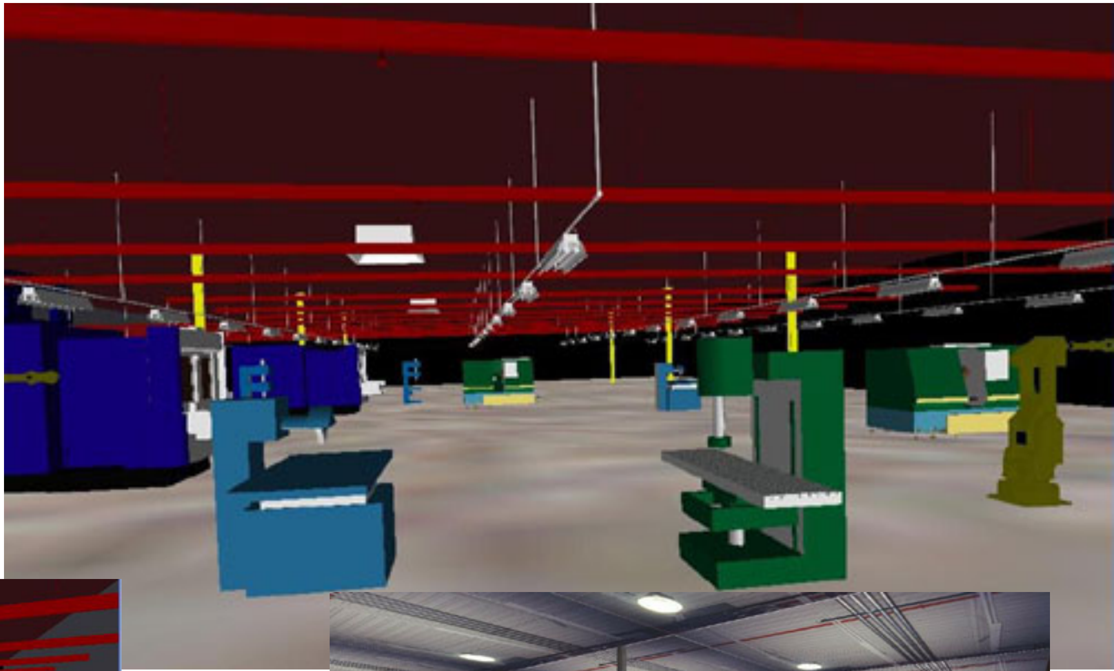
- VR work of fiction build for CAVE at EVL 1997-2000
- Users is protagonist
- User interacts with computer controlled characters
- Based on short story of J. Anstey

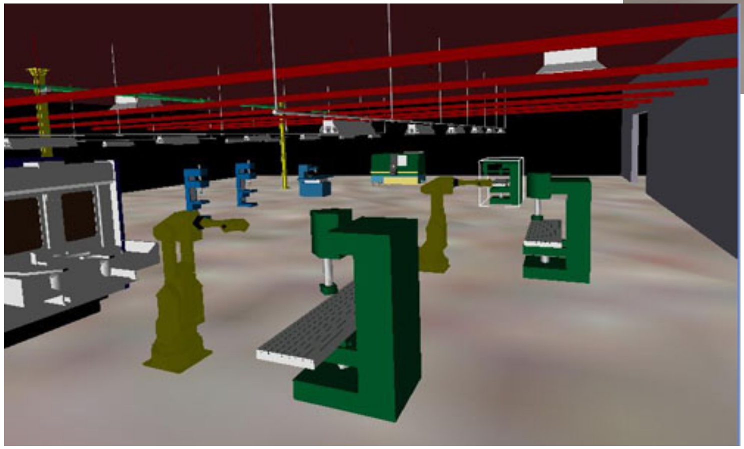


J. Anstey

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

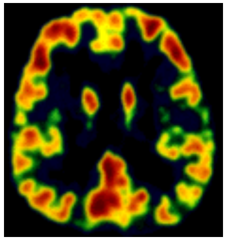
Buffalo, New York

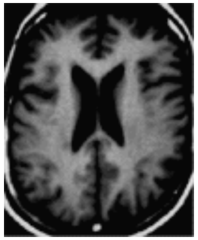
- The Queen City: 2nd Largest City in NYS
- City of Lights
  - First U.S. city to have electric street lights
  - Pan American Exposition (1901)
    - Pres. McKinley Shot
- Architecture
  - Frederick Law Olmsted
  - Frank Lloyd Wright

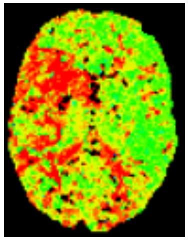
- Underground Railroad
  - Slaves escaped to freedom in Canad
- Four straight Super Bowl appearances
- Culinary Delights
  - Beef on Weck, Pizza, Fish Fries
  - (Buffalo) Wings: Anchor Bar, 1964
- Health Problems
  - Heart Disease/Stroke
  - Multiple Sclerosis

Recent Biomedical Advances (Buffalo, NY)

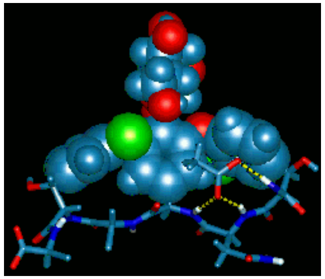
- PSA Test (screen for Prostate Cancer)
- Avonex: Interferon Treatment for Multiple Sclerosis
- Artificial Blood
- Nicorette Gum
- Fetal Viability Test
- Edible Vaccine for Hepatitis C
- Timed-Release Insulin Therapy
- Anti-Arrythmia Therapy
  - Tarantula venom







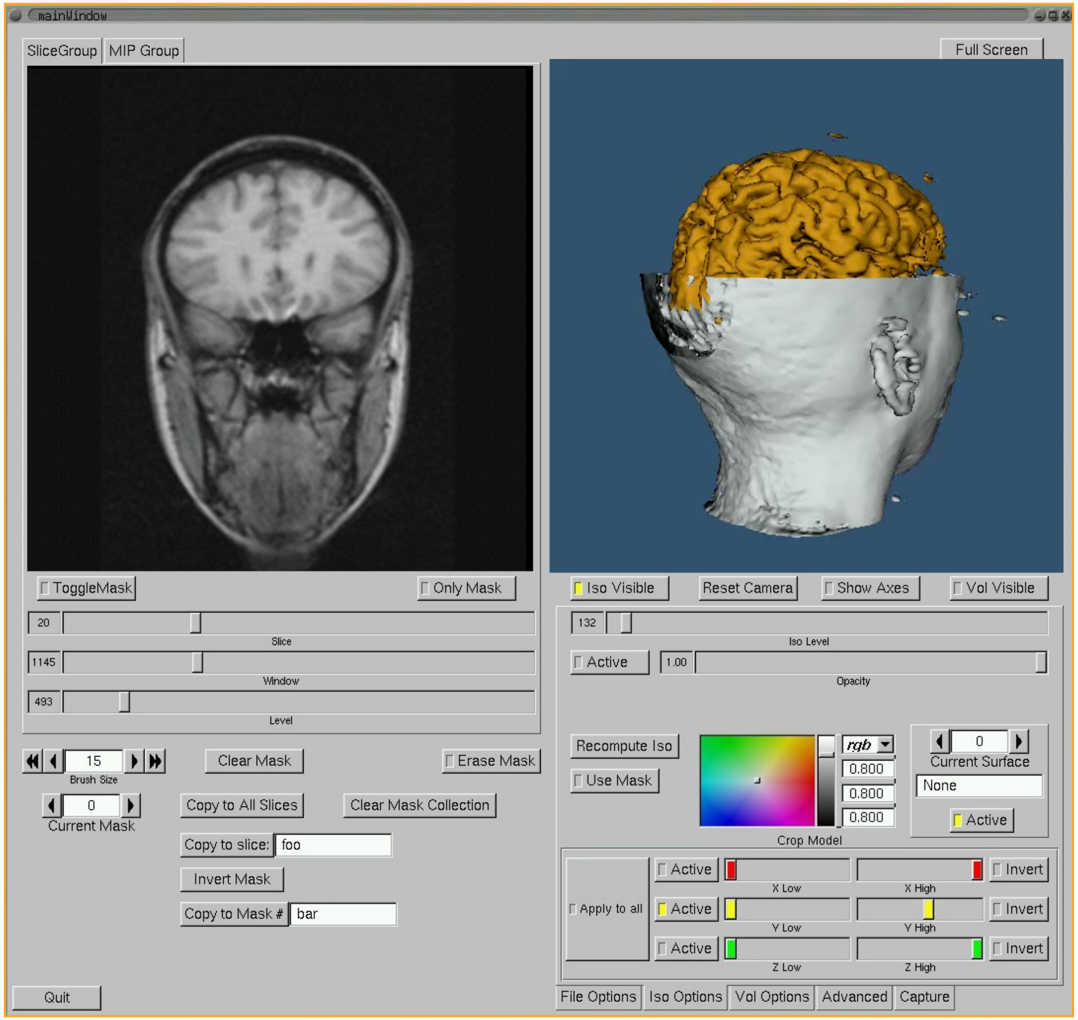
- Direct Methods Structure Determination
  - Listed on "Top Ten Algorithms of the 20th Century"
- Vancomycin
- Gramacidin A
- High Throughput Crystallization Method: Patented
- NIH National Genomics Center: Northeast Consortium
- Howard Hughes Medical Institute: Center for Genomics & Proteomics



Scientific Visualization

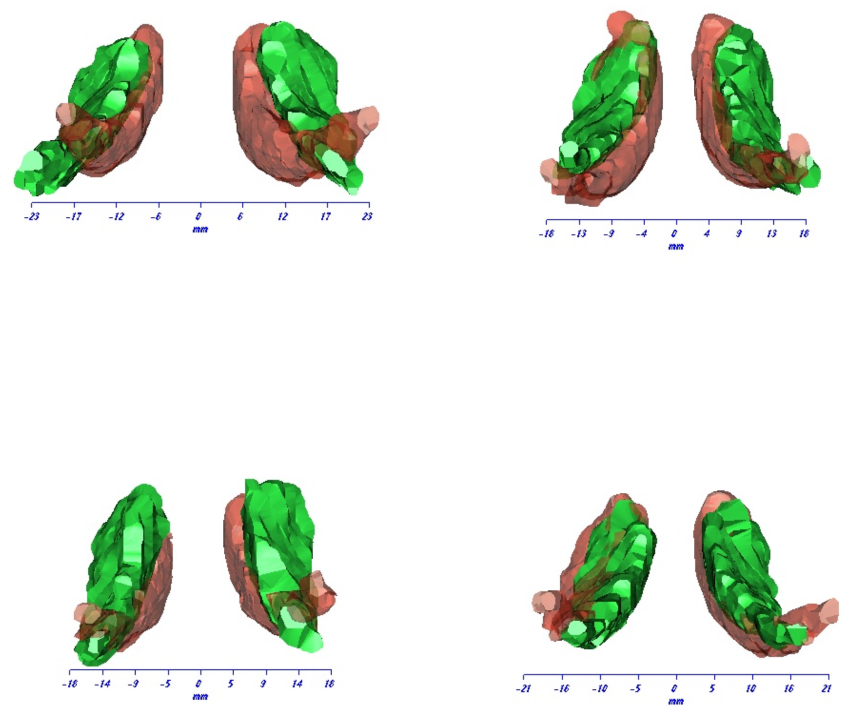
Multiple Sclerosis Project

- 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



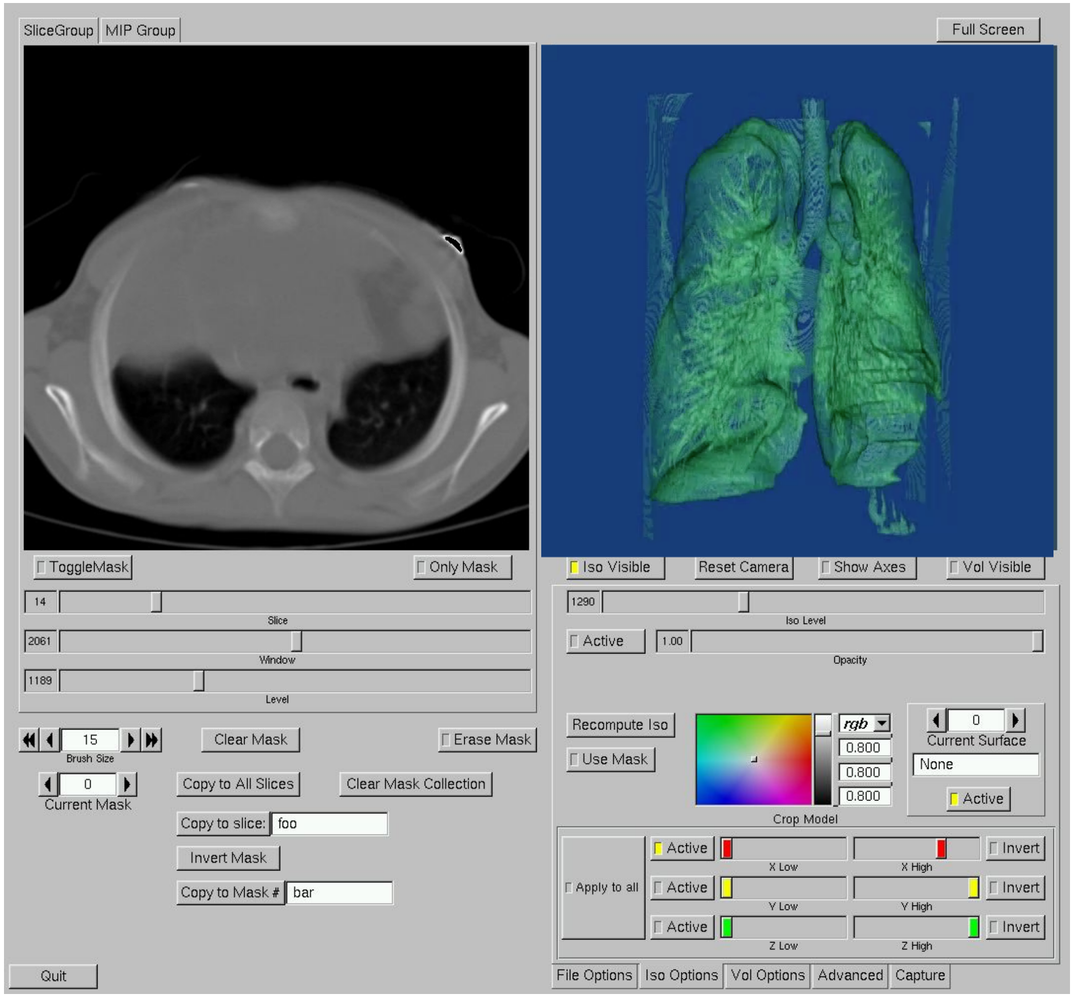
Multiple Sclerosis Project

- Compare caudate nuclei between MS patients and healthy controls
- Looking for size as well as structure changes
  - Localized deformities
  - Spacing between halves
- Able to see correlation between disease progression and physical structure changes



3D Medical Visualization

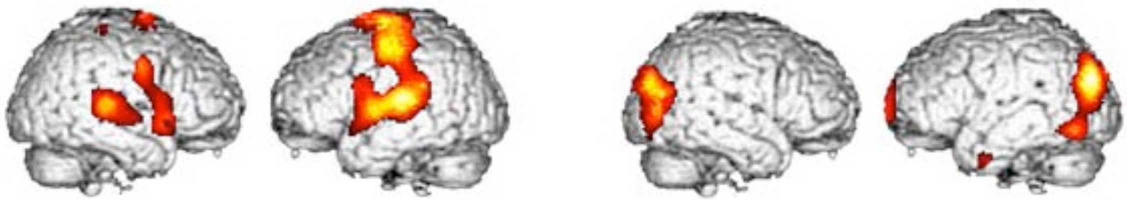
- Reads data output from a CT or MRI Scan
- Collaboration with Children's Hospital
- Visualize multiple surfaces and volumes
- Export images, movies or CAD file
- Pre-surgical planning
- Runs on a PC



M. Innus

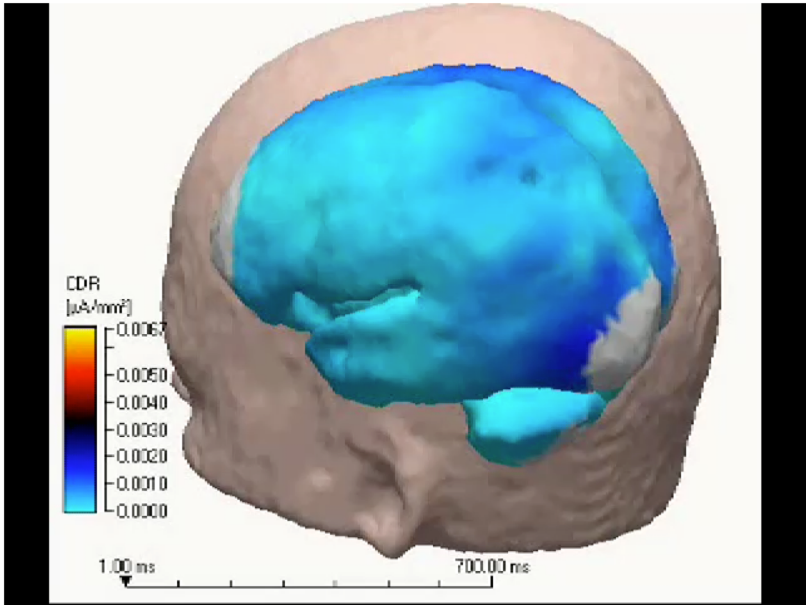
Mapping Brain Activity

Sites Activated Sites Deactivated



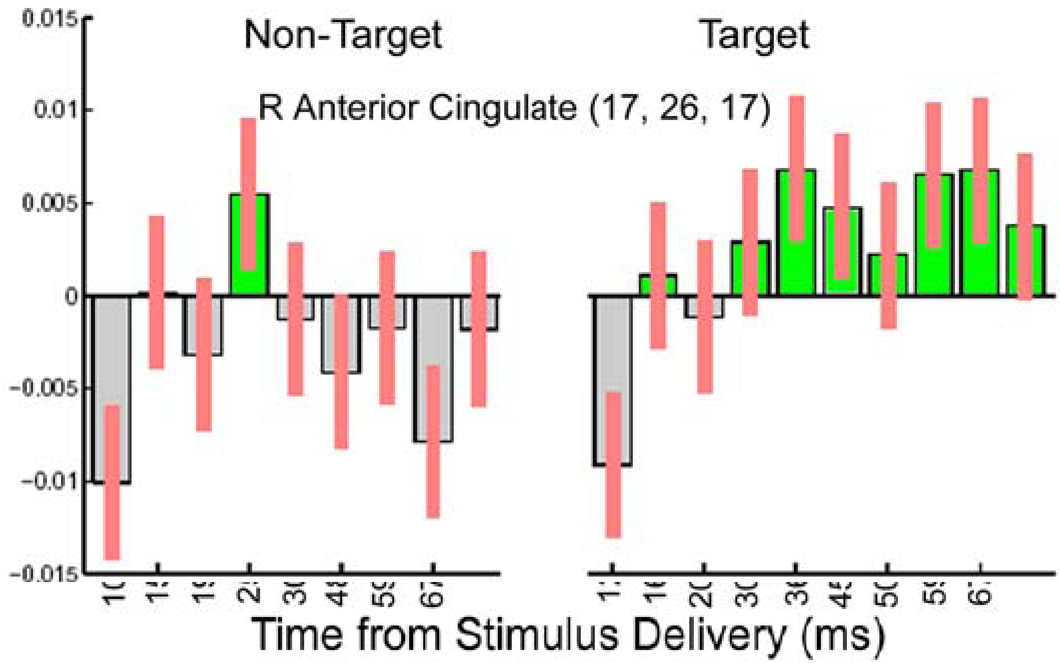
Positron emission tomography (PET), shows sites activated and deactivated as subjects decide whether a sound is a target or not.

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



A. Lockwood

Mapping Brain Activity

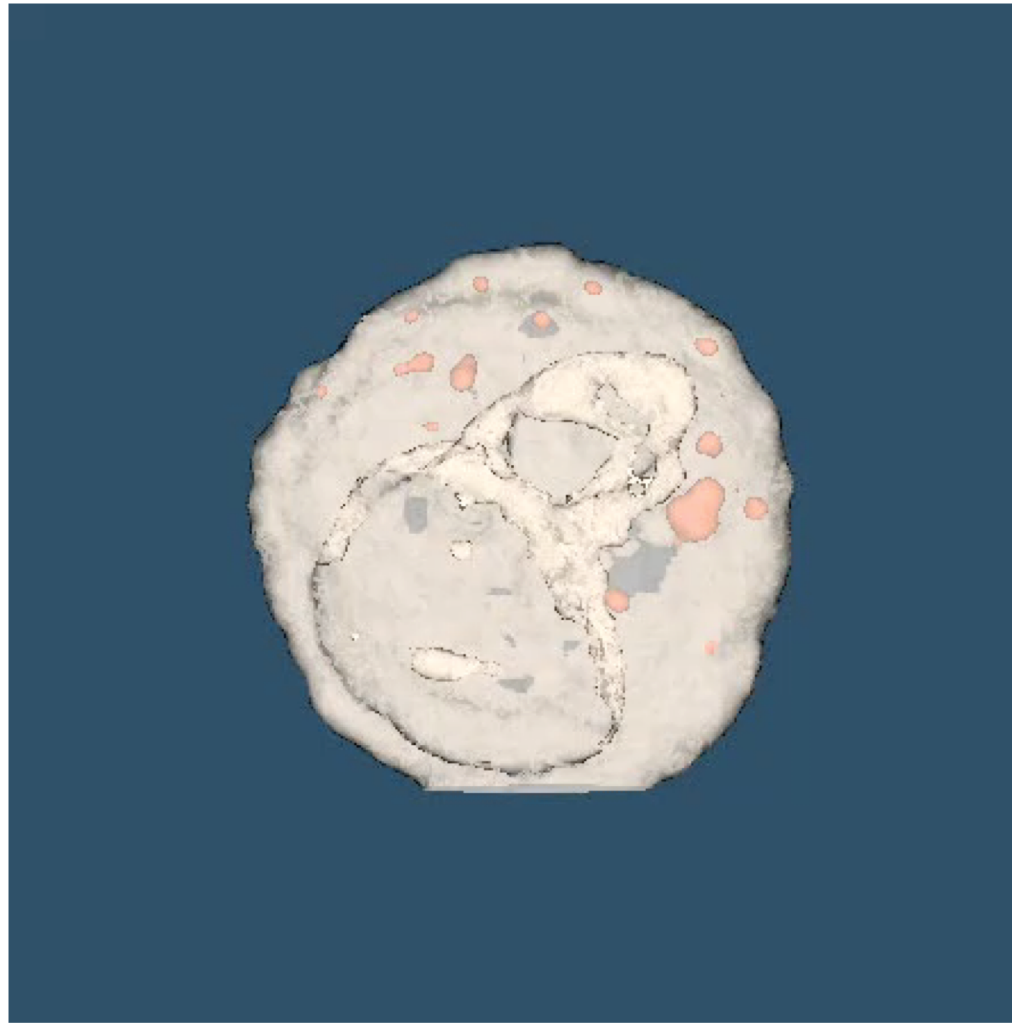


Temporal sequence of anterior cingulate cortex activation in response to targets and non-targets. This brain region controls attention-related neural activity. Green bars indicate significant differences compared to T = 0, the time of stimulus presentation.

A. Lockwood

Confocal Microscopy

- 3D Reconstruction of an Oral Epithelial Cell
- Translucent White Surface Represents the Cell Membrane
- Reddish Surface Represents Groups of Bacteria

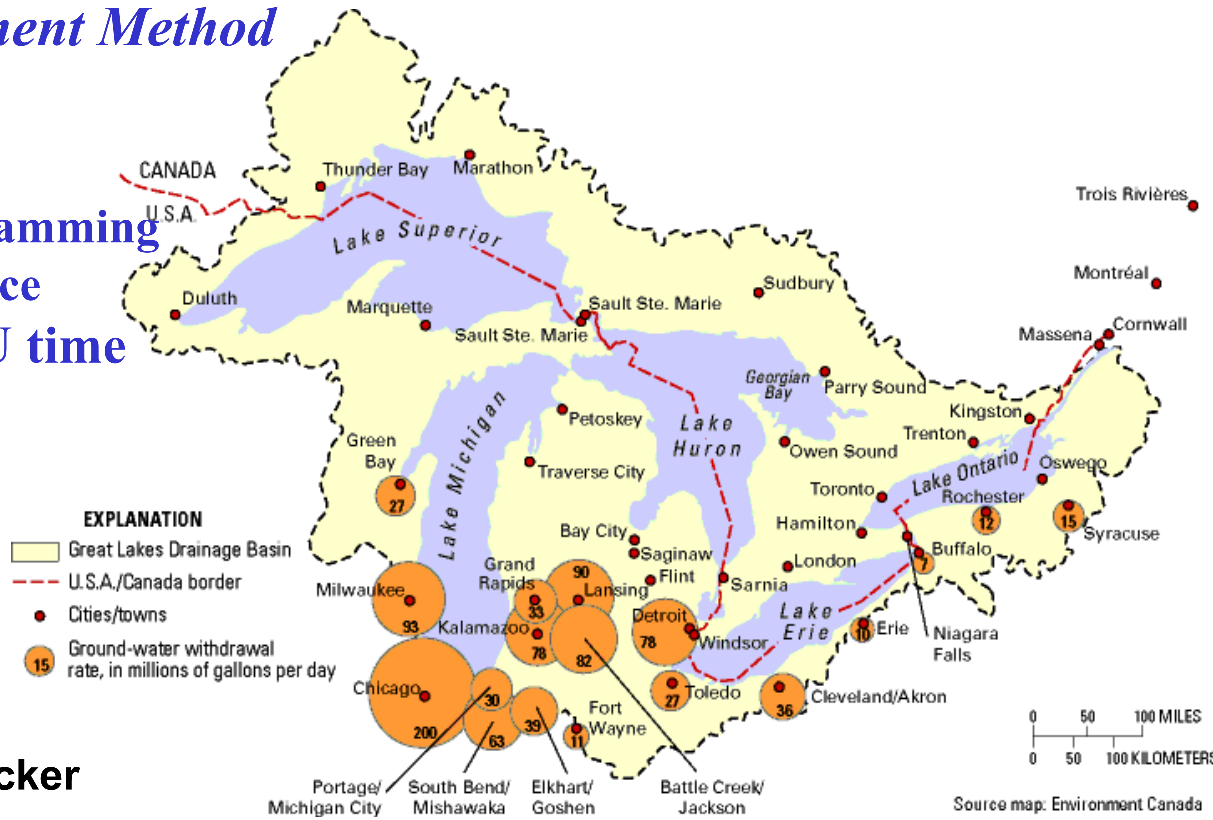


Science & Engineering

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

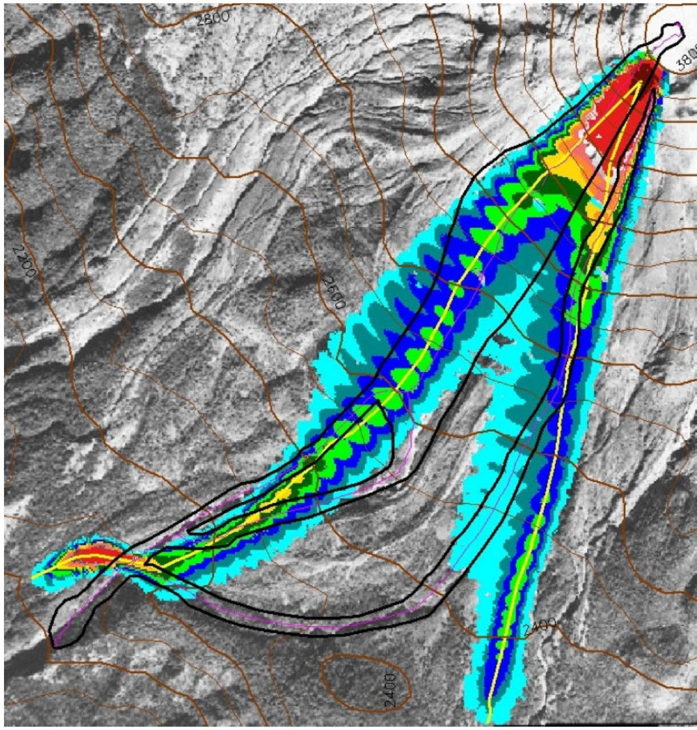


A. Rabideau, I. Jankovic, M. Becker

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

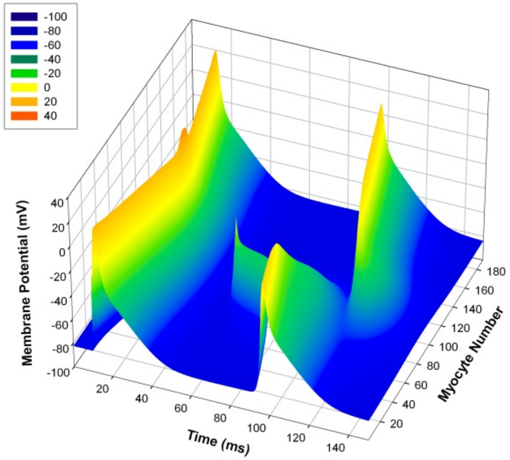


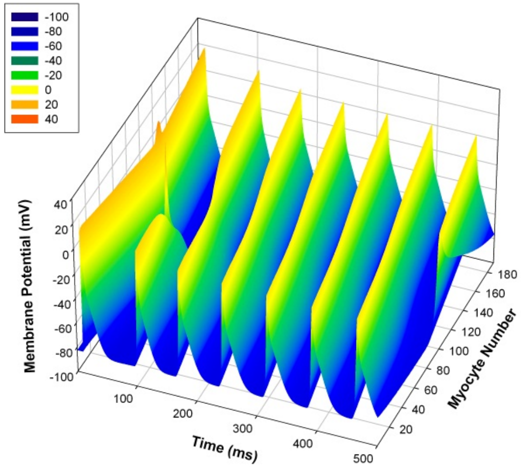
Flow models of Colima volcano
In Mexico – courtesy Rupp et. al.'06

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

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

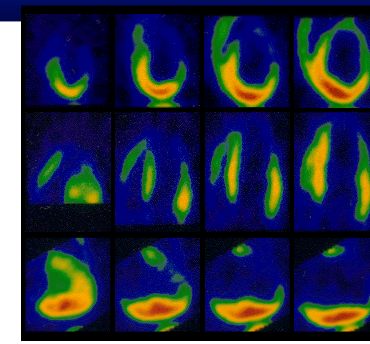




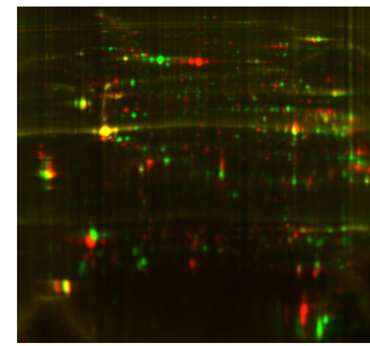
Non-sustained and sustained arrhythmia

Center for Cellular and Systems Electrophysiology

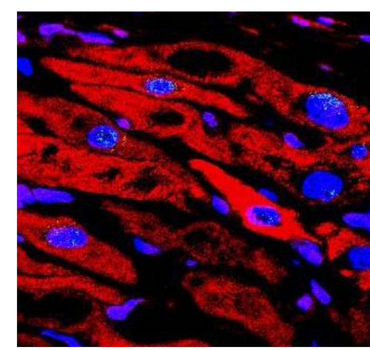
Cardiovascular Research



- **Molecular Imaging – PAREPET Clinical Study Analysis of cardiac PET** (Positron Emission Tomography) **scans aims to revolutionize assessment of an individual's risk for sudden cardiac death.**



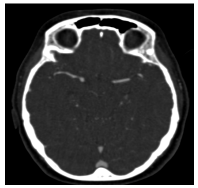
- **High-Throughput Discovery – Proteomics and Genomics**
- **Protein and gene expression profiling using differential in-gel electrophoresis and microarray technology provides a blueprint for the cellular mechanisms involved in hibernating myocardium.**

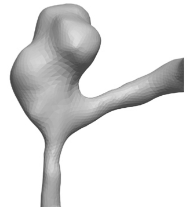


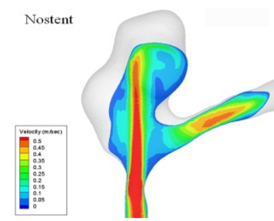
- **Translate results to identify gene and other therapeutic targets aimed at improving heart function and survival.**

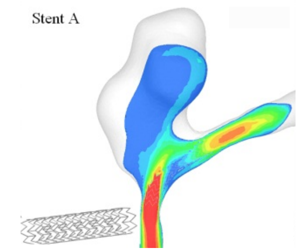
Center for Research in Cardiovascular Medicine

Cerebral Aneurysm: Virtual Intervention









CTA/ MRI/ Rotational Angiograph

3D Patient-Specific Aneurysm Geometry

CFD

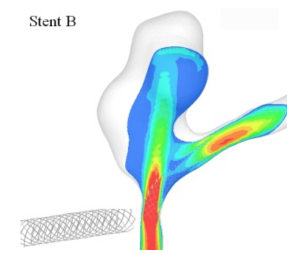
Virtual Interventions

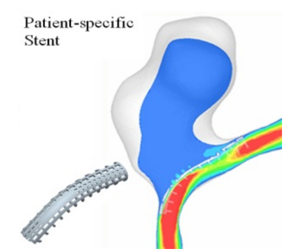
Hemodynamics

Arterial Wall Mechanics

Biological & Pathological Response

Risk Assessment & Treatment Recommendation

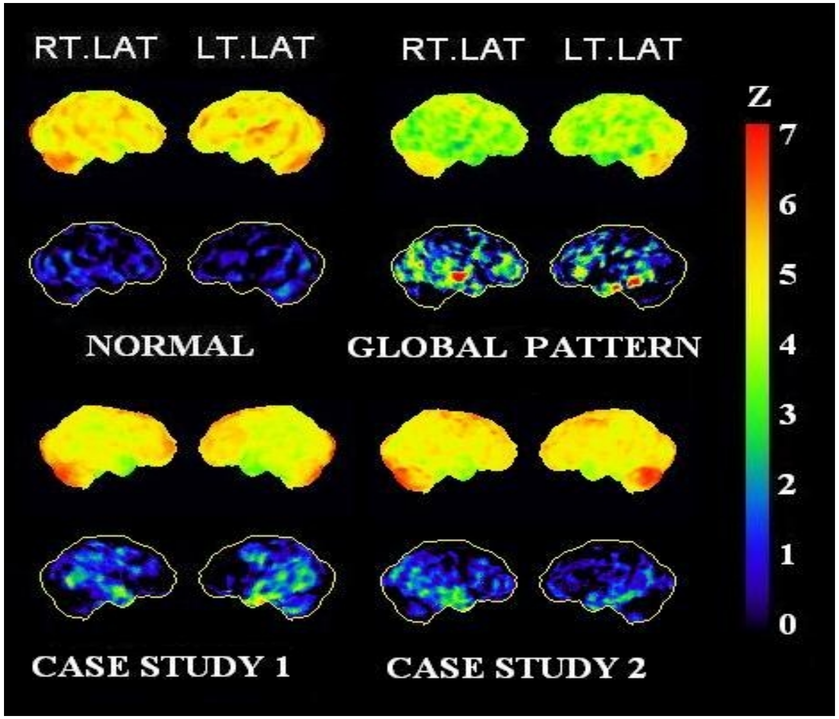




- Understand Cerebral Vascular Disease Mechanisms
- Enable Patient-Specific Quantitative Analysis
- Diagnose, Treatment Planning, R&D
- Help Improve Interventions

Vascular Dementia Imaging

- Early diagnosis of dementia from cerebral small vessel disease using computer analysis of SPECT Images
- Collaboration between Nuclear Medicine, CCR, Neurology, and Kaleida Stroke Center
- Funded by the Pfeiffer Foundation
- Fractal scores:
  - Normal 0.75
  - Global Pattern 1.13
  - Case Study 1 0.96
- Case Study 1 Moderate white matter and cortical hypo-perfusion with visual memory, speed of processing, and verbal fluency deficits



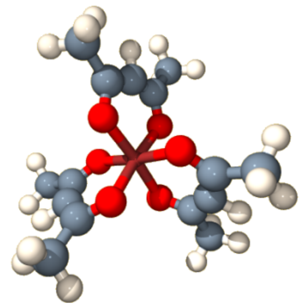
J. Baker, M. Innus

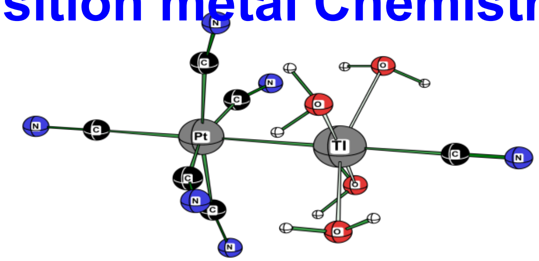
Theoretical and Computational Chemistry

Chemistry

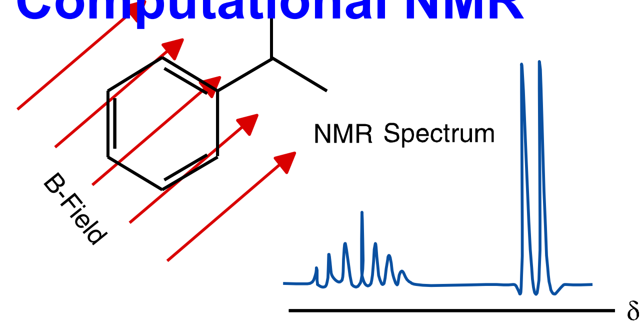
Applied to:

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



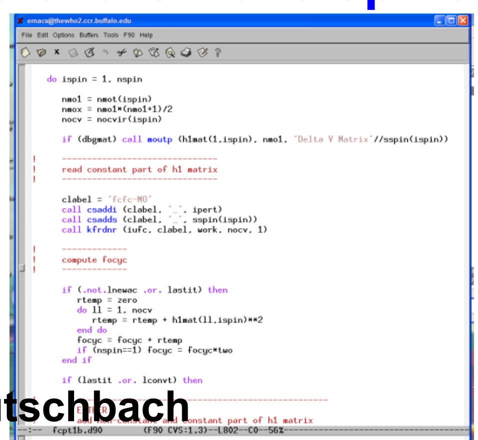


Computational NMR

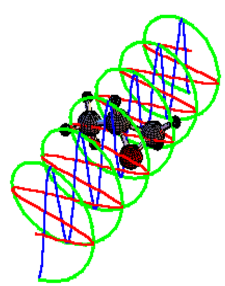


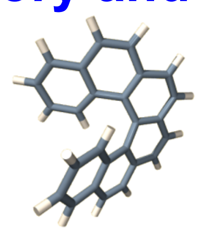
Magnetic Properties of Molecules

Software Development

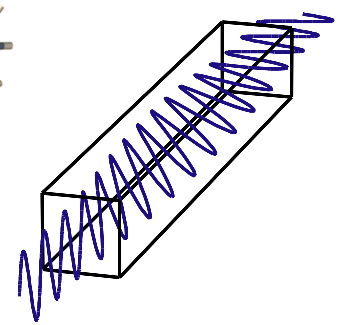


Theory and Calculation of Optical Activity

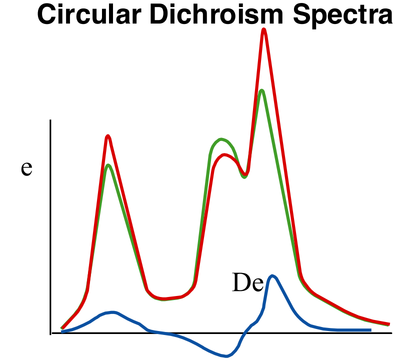




Optical Rotation



Circular Dichroism Spectra



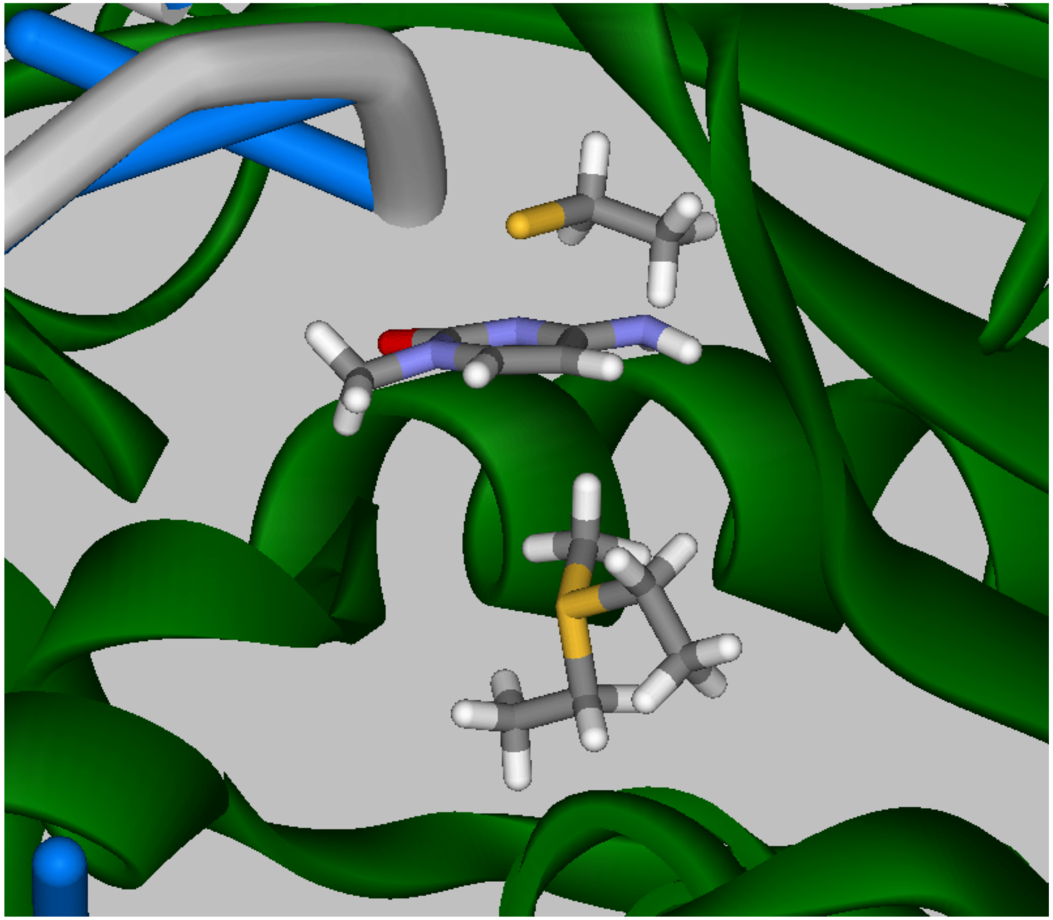
J. Autschbach

Understanding How Proteins Work

Collaboration with Merck Pharmaceutical Company

Modeling:

- DNA-Protein Interaction (understanding cancer)
- Drug-Protein Interaction (understanding blood clotting)



Movie shows a chemical reaction between a protein and DNA, which is responsible for some types of cancer.

M. Freindorf, T. Furlani

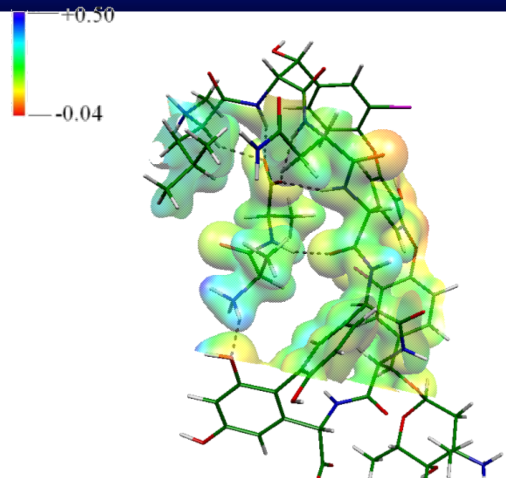
Computational Chemistry

- 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

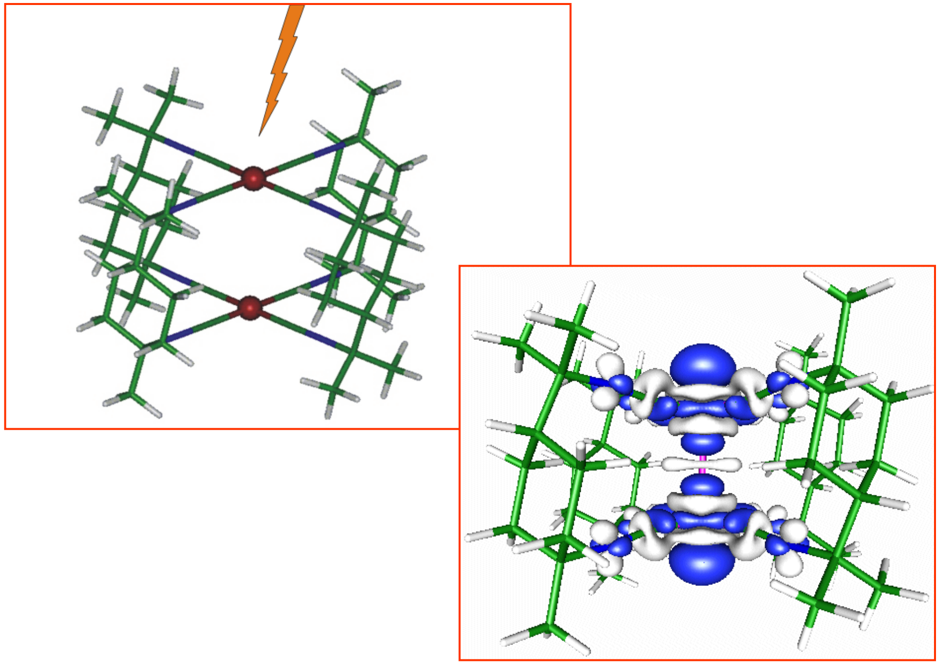
T. Furlani, J. Autschbach, M. Freindorf

Understanding Large Molecules and Fleeting Species

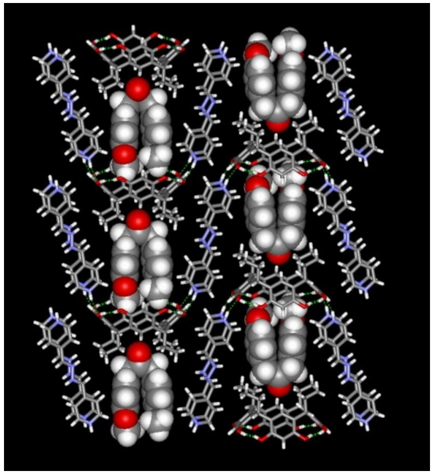
Chemistry



Calculation of antibiotic molecules: electrostatic potential of vancomycin



A molecule changes on excitation by light



A supramolecular solid

P. Coppens

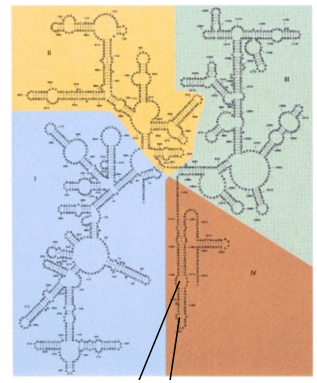
Prediction of RNA Structure to Facilitate Design of Drugs Targeting RNA

Chemistry

aggtccctcg cggatggagc tgaaatcagt cgaagatacc agctggctgc aactgtttat
taaaaacaca gcactgtgca aacacgaaag tggacgtata cggtgtgacg cctgcccggt
gccggaaggt taattgatgg ggttagcgca agcgaagctc ttgatcgaag ccccggtaa
cggcggccgt aactataacg gtcctaaggt agcgaaattc cttgtcgggt aagttccgac
ctgcacgaat ggcgtaatga tggccaggct gtctccaccc gagactcagt gaaattgaac
tcgctgtgaa gatgcagtgt acccgcggca agacggaaag accccgtgaa cctttactat
agcttgacac tgaacattga gccttgatgt gtaggatagg tgggaggctt tgaagtgtgg
acgccagtct gcatggagcc gaccttgaaa taccaccctt taatgtttga tgttctaacg
ttgacccgta atccgggttg cggacagtgt ctggtgggta gtttgactgg ggcggtctcc
tcctaaagag taacggagga gcacgaaggt tggctaatcc tggtcggaca tcaggaggtt
agtgcaatgg cataagccag cttgactgcg agcgtgacgg cgcgagcagg tgcgaaagca
ggtcatagtg atccggtggt tctgaatgga agggccatcg ctcaacggat aaaaggtact
ccggggataa caggctgata ccgcccaaga gttcatatcg acggcggtgt ttggcacctc
gatgtcggct catcacatcc tggggctgaa gtaggtccca agggtatggc tgttcgccat
ttaaagtggt acgcgagctg ggtttagaac gtcgtgagac agttcggtcc ctatctgccg
tgggcgctgg agaactgagg ggggctgctc ctagtacgag aggaccggag tggacgcatc
actggtgttc gggttgtcat gccaatggca ctgcccggta gctaaatgcg gaagagataa
gtgctgaaag catctaagca cgaaacttgc cccgagatga gttctccctg accctttaag
ggtcctgaag gaacgttgaa gacgacgacg ttgataggcc gggtgtgtaa gcgcagcgat
gcgttgagct aaccggtact aatgaaccgt gaggcttaac ctt

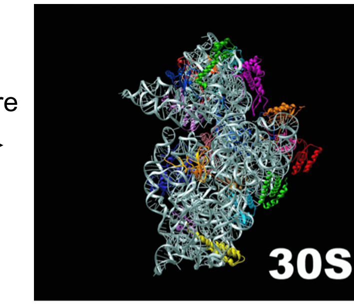
23s rRNA Sequence

RNA Secondary Structure Prediction

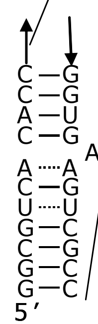


23s rRNA 2D Structure

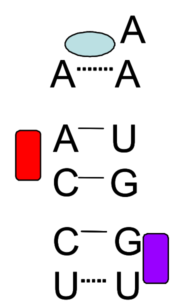
RNA 3D Structure



Search Small-Molecule-RNA Ligand Database for RNA motifs to which we have identified an organic ligand

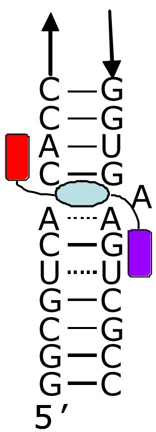


A-site
Aminoglycoside target



Results from ligand database search

Modular Assembly to increase affinity

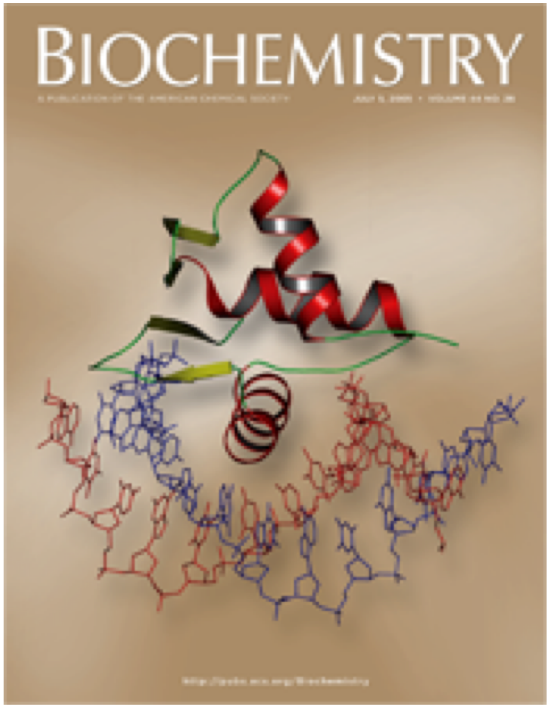


$$\Delta G_{ligand} = \Delta G_1 + \Delta G_2 + \Delta G_3 + \Delta G_{linker(s)}$$

M. Disney

3D Structure of Proteins

- Direct Methods for Crystal Structure Determination
  - Listed on “Top Ten Algorithms of the 20th Century”
- UB/HWI collaborative software development
  - SnB – determine protein heavy-atom substructures
    (http://www.hwi.buffalo.edu/SnB/)
  - BnP – determine complete protein structures
    (http://www.hwi.buffalo.edu/BnP/)
- Applications to drug design
  - AIDS
  - Arthritis
  - Cancer
  - Heart disease
  - SARS



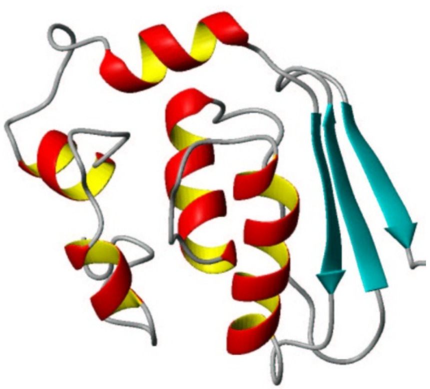
Hauptman-Woodward Institute

R. Miller, C. Weeks

Determining 3D Protein Structure

- NMR-based Structural Biology and Structural Genomics
- Bio-NMR Methodology
- NMR-based Metabonomics in Cancer Research

Structural Biology

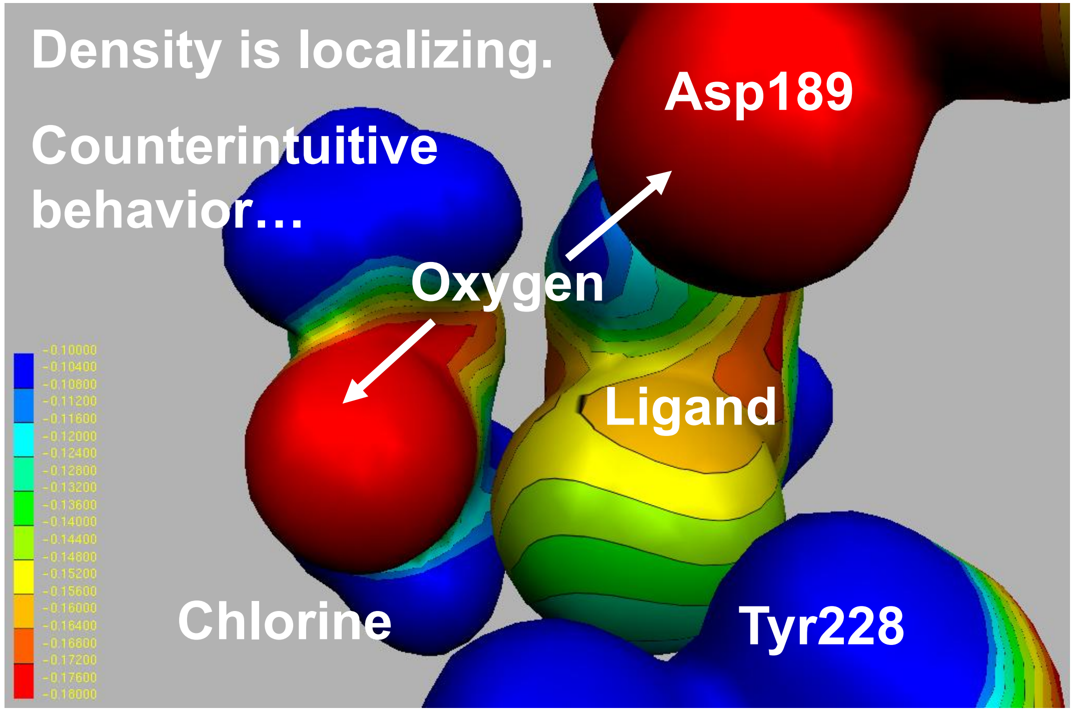


Propelled by Recent Advances, NMR Moves Into the Fast Lane

A speedy new NMR technique could finally help structural genomics groups achieve their goal of devising factory-style approaches to mapping protein structures at high speeds

T. Szyperski

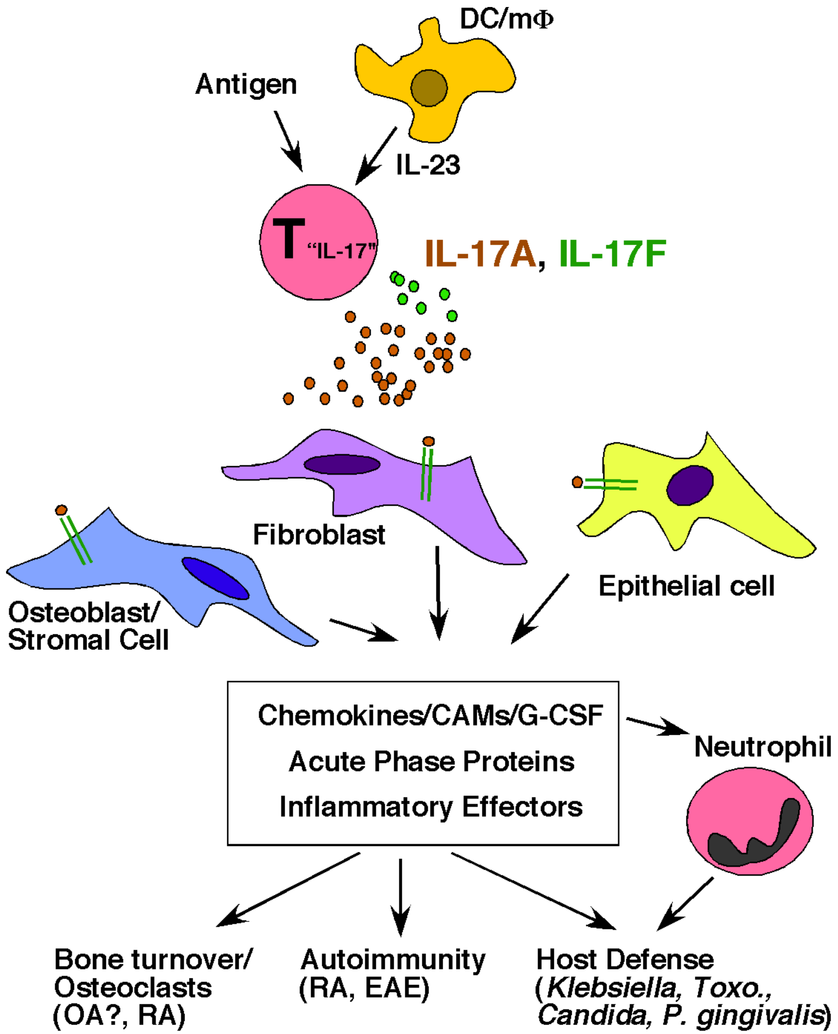
Binding in a Drug-Receptor Complex



- Ligand docked with residues of the active site of thrombin. Electrostatic potential map superimposed onto the electron density isosurface.
- The goal is to elucidate the thermodynamics of molecular recognition in binding.

D. Hangauer, M. Freindorf

Defining Cytokine Signaling Mechanisms

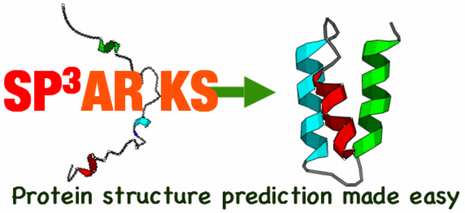


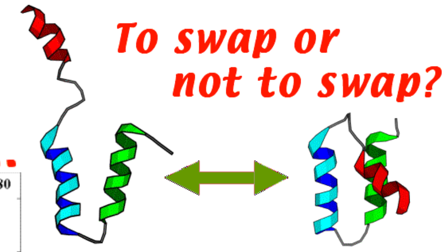
S. Gaffen, Z. Hu

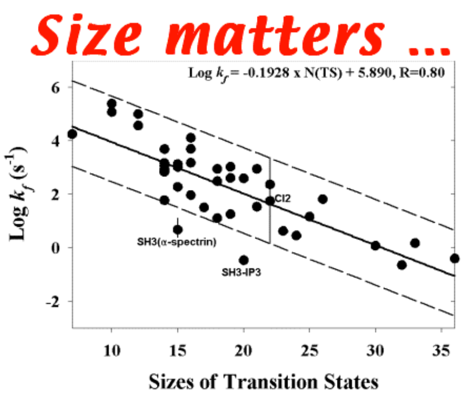
- T cells secrete cytokines such as IL-17 to promote host defense and/or autoimmunity
- Microarrays used to define IL-17 gene targets in various cell types
- Computational and statistical approaches used to compare the promoters of IL-17 target genes in mouse and human genomes to identify conserved transcription factor binding sites (TFBS), with the ultimate goal of understanding how IL-17 mediates molecular signals
- IL-17 target promoters contain conserved TFBSs, including NF-kB and C/EBP

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**





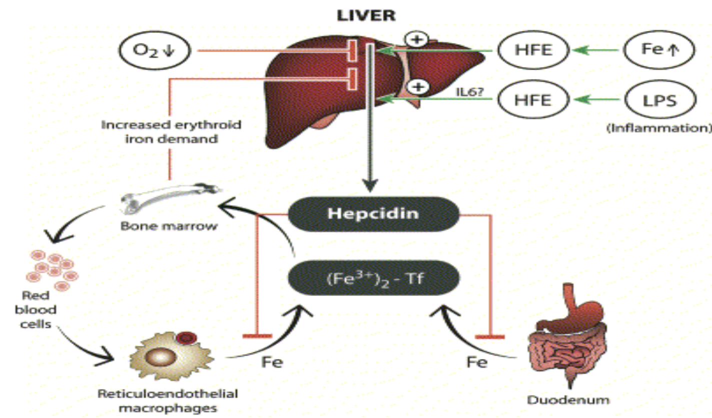


M. Halfon http://theory.med.buffalo.edu

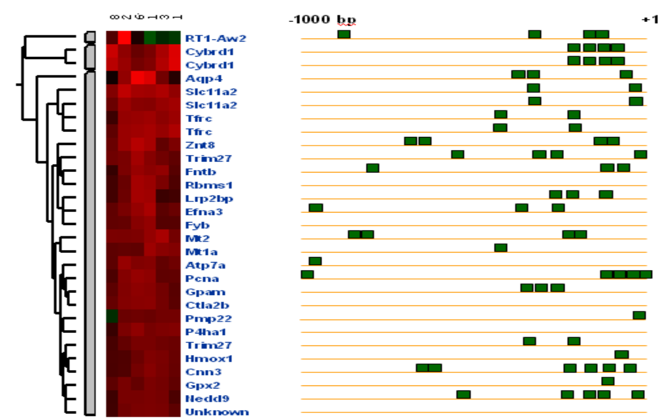
Genome-Wide Study of Iron Homeostasis

Bioinformatics

- Physiological and Hepcidin-hormonal regulators are involved in iron homeostasis
- Intestinal iron transport controls overall body iron homeostasis
- Computational biology to discover new genes involved in iron homeostasis
- Systems biology to reveal regulatory pathway responding to iron status
- Known and novel genes are regulated according to iron status
- Sp1 or related TFs may be involved in regulating expression of some genes during iron-deficiency



Hentze M.W., *et. al.* Cell 117(3):285-97, 2004

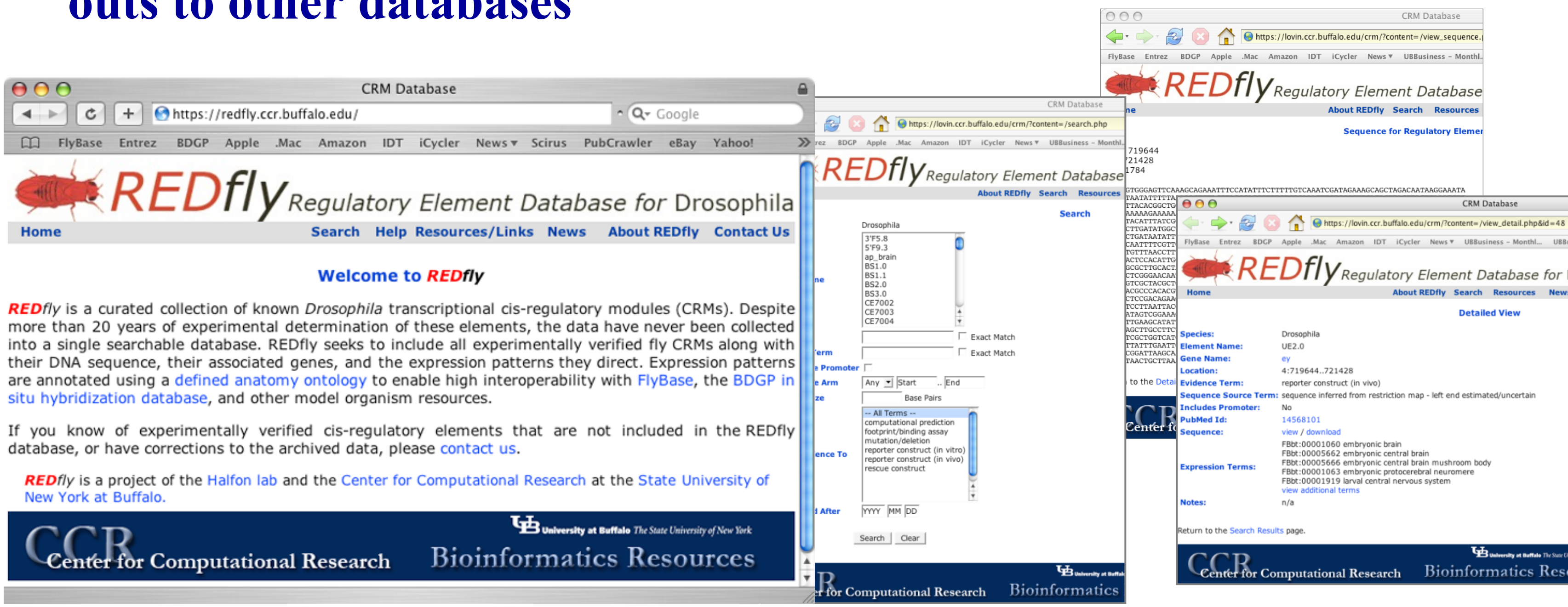


Distribution of SP1 on promoters of genes induced during iron-deficiency

J. Collins, Z. Hu

Regulation of Gene Expression

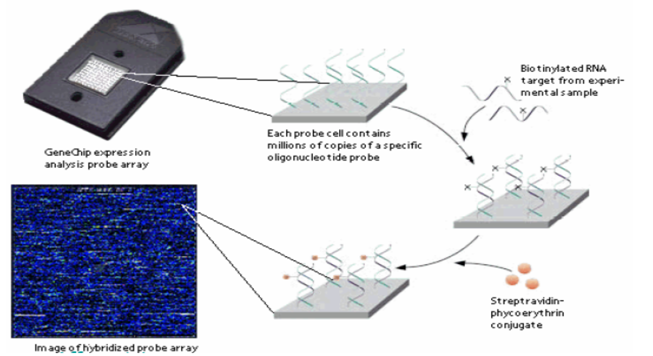
- ***REDfly*** (Regulatory Element Database for Fly) Database of verified transcriptional regulatory elements
- Over 650 entries
- Most comprehensive resource of animal regulatory elements
- Fully searchable, has DNA sequence and gene expression data, link-outs to other databases

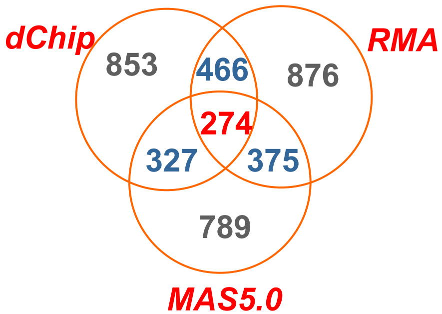


M. Halfon, S. Gallo

Data Mining and Analysis

- GeneChips hybridized with cRNAs of biological interest
- Many probe set algorithms for summarizing expression intensity
- Significant differences of differentially expressed genes generated by different algorithms from the same dataset

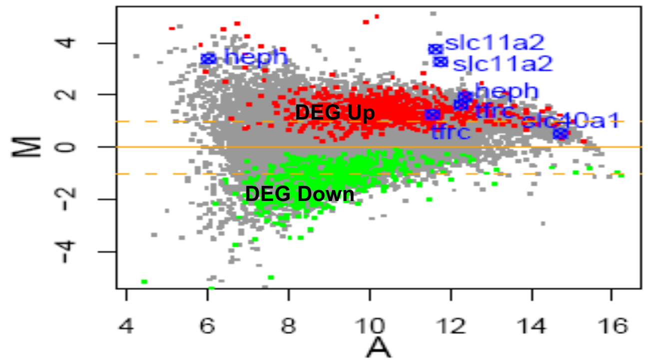




31% - 55% overlap

Which algorithm is best?

Great impact on subsequent expression data analysis



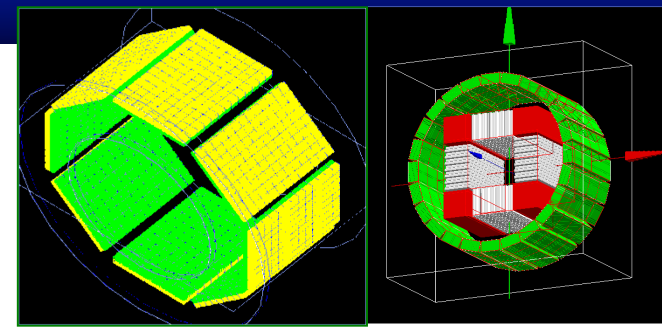
- Novel statistical approach for data variance and result bias analyses
- No external reference data needed
- Algorithm evaluation with direct applications to experimental datasets of interest


BMC Bioinformatics
BioMed Central
Methodology article
Open Access
Utilization of two sample *t*-test statistics from redundant probe sets to evaluate different probe set algorithms in GeneChip studies

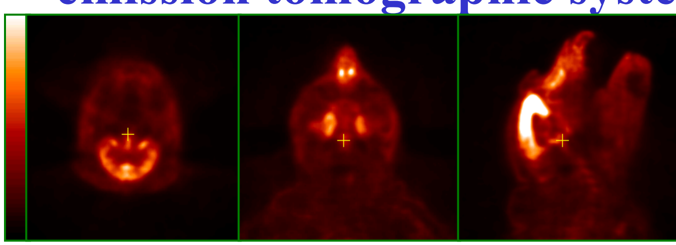
Z. Hu, G. Willsky

Nuclear Medicine

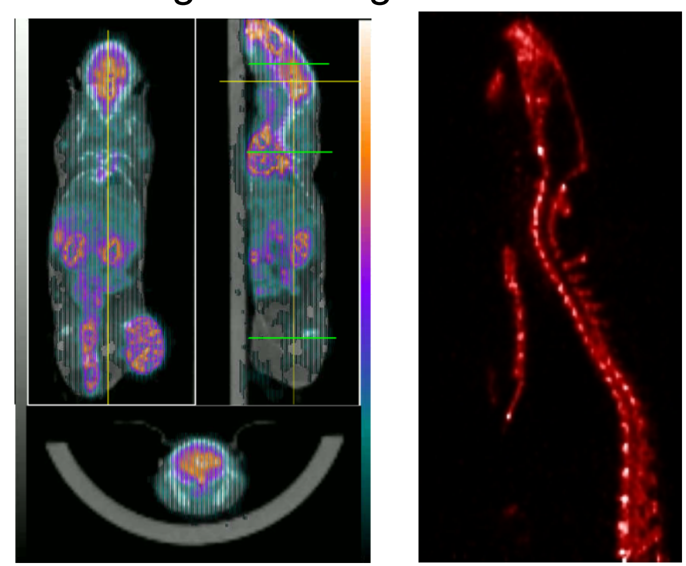
- **Monte Carlo simulation**
  - **for modeling imaging system characteristics, optimizing system design, and validating data correction algorithms.**
- **Image reconstruction**
  - **for development of high resolution image reconstruction algorithm and software for both human and animal nuclear emission tomographic systems.**



Two virtual imagers simulated for system modeling and design evaluation.



Transverse, coronal and sagittal views of a monkey brain scanned on a dedicated brain PET using the radioligand ^{18}F-FCWAY.

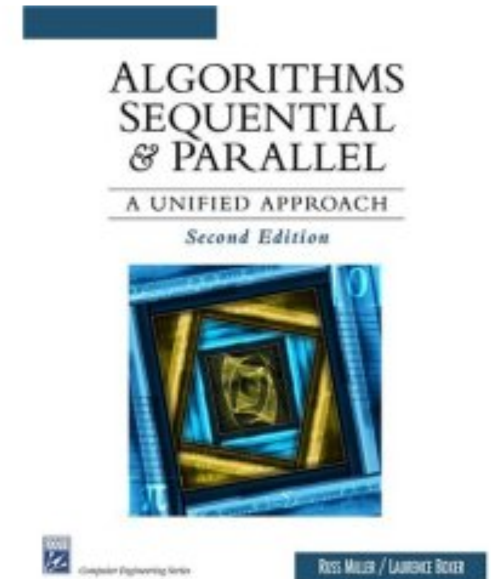


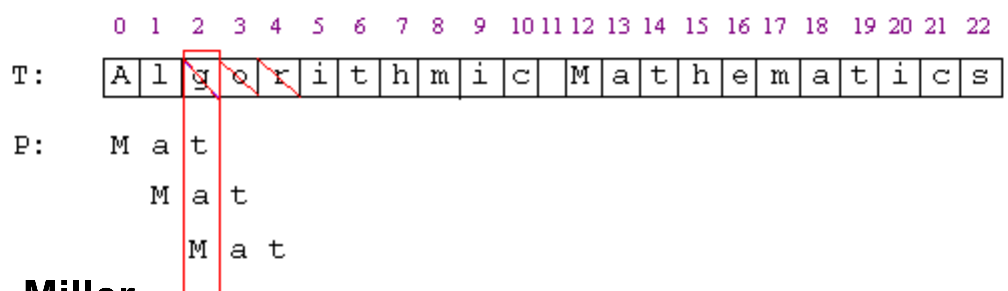
Co-registered ^{18}F-FDG PET and CT mouse images (left) and a ^{18}F-Fluorine bone image of a 250 gram rat.

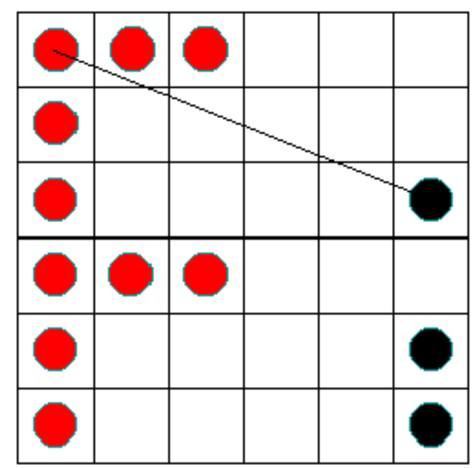
Y. Yao, M. Jones

Parallel Algorithms

- String pattern matching searche for word processors, Web, molecular biology
- Image processing
- Computational geometry
- Fundamental operations



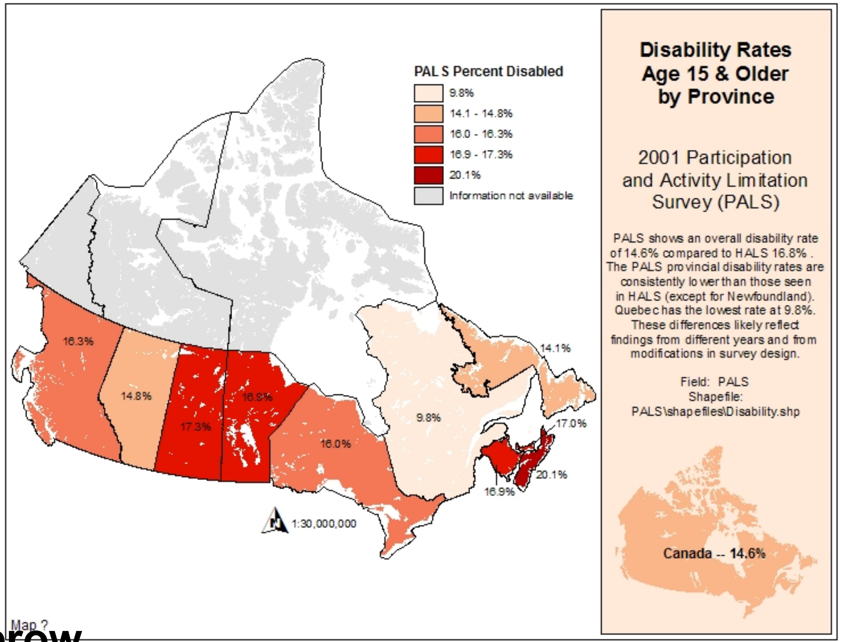


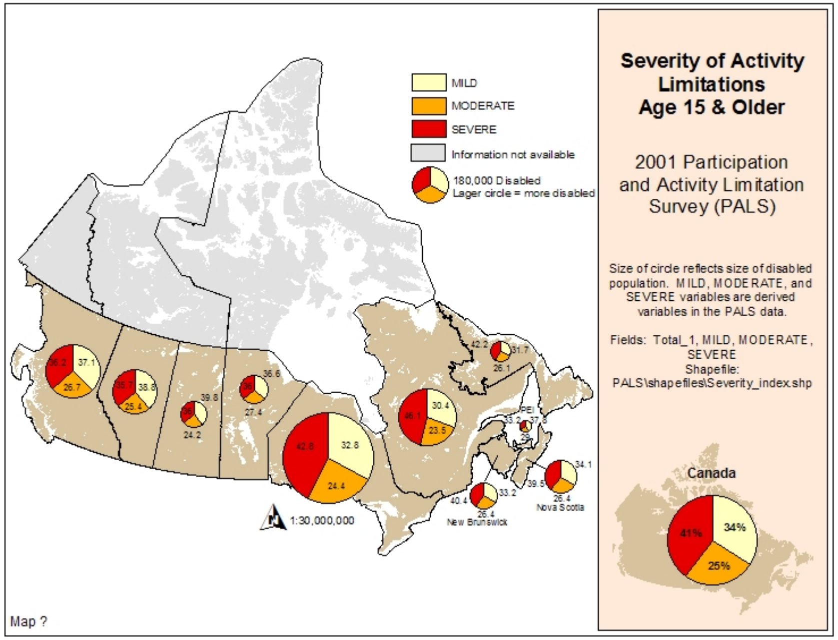


L. Boxer, R. Miller

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.

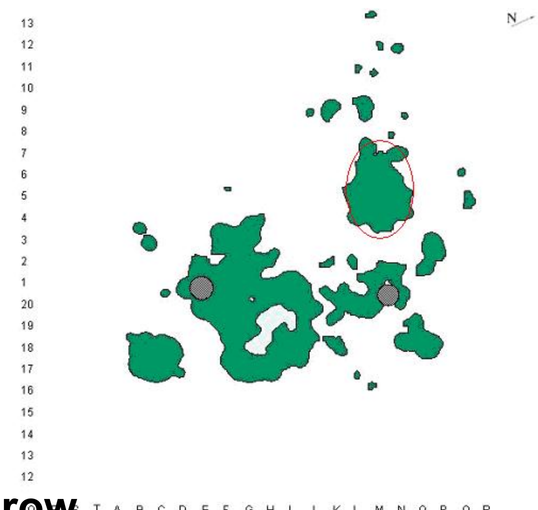


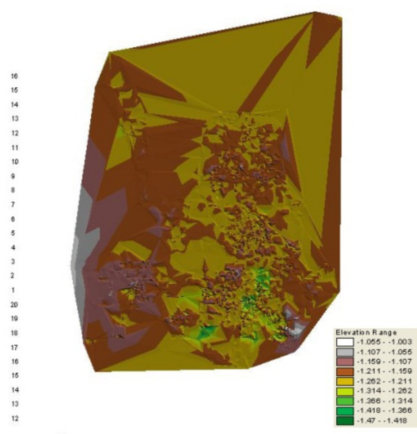


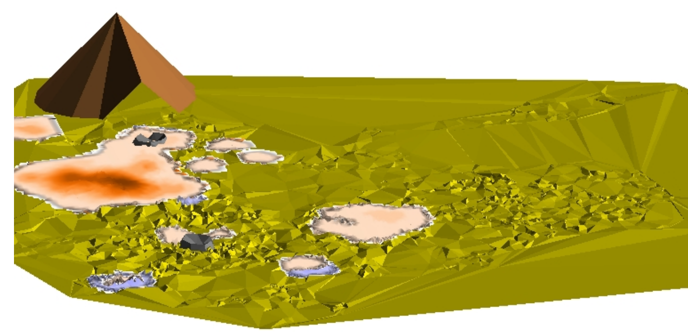
E. Zubrow

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



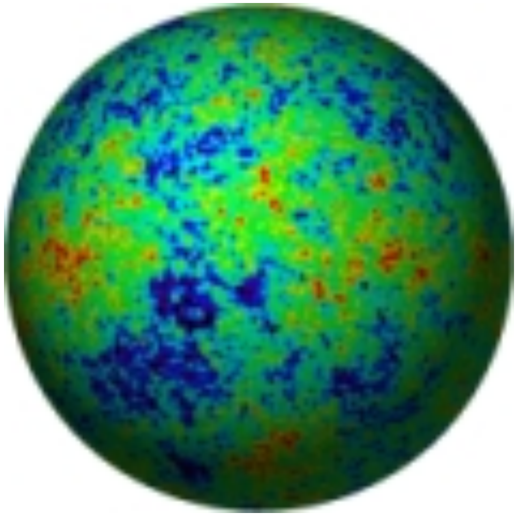


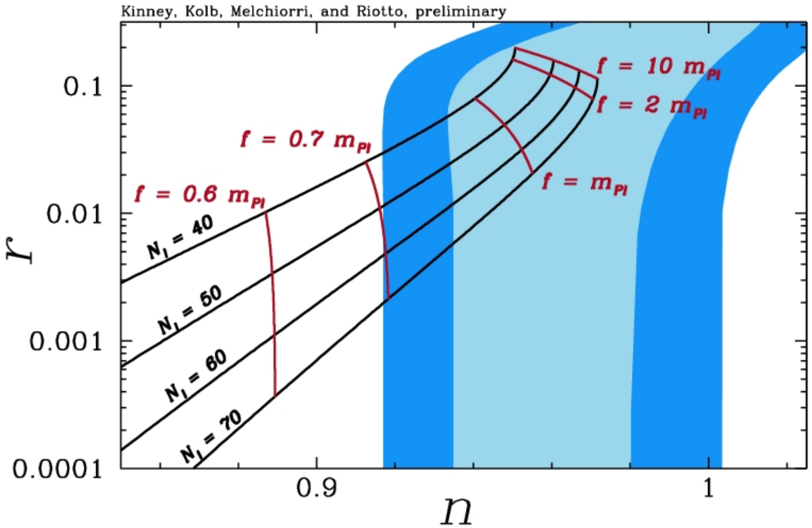


E. Zubrow

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*



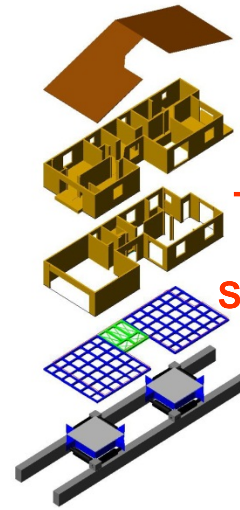


W. Kinney

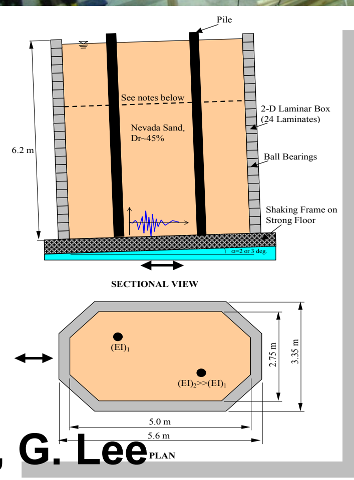
UB's Structural Engineering and Earthquake Simulation Laboratory (SEESL)

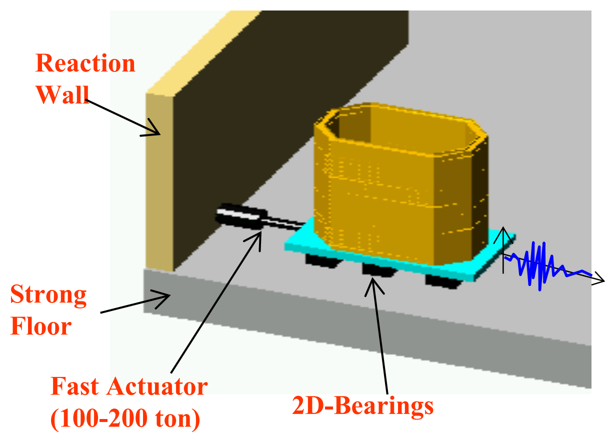
Structural Engineering

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



Two-story Townhouse on Twin Shake Tables



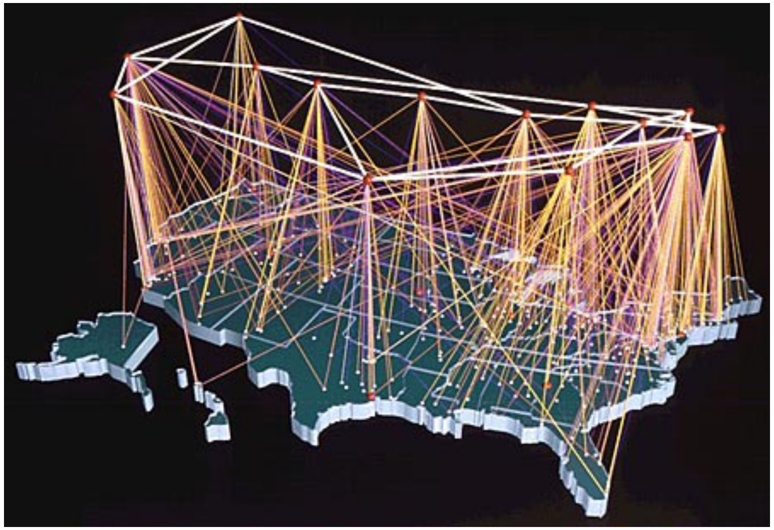


2-D Geotechnical Laminar Box Tests of Pile Foundations Subjected to Soil Liquefaction

M. Bruneau, A. Reinhorn, G. Lee

Cyberinfrastructure in STEM Education

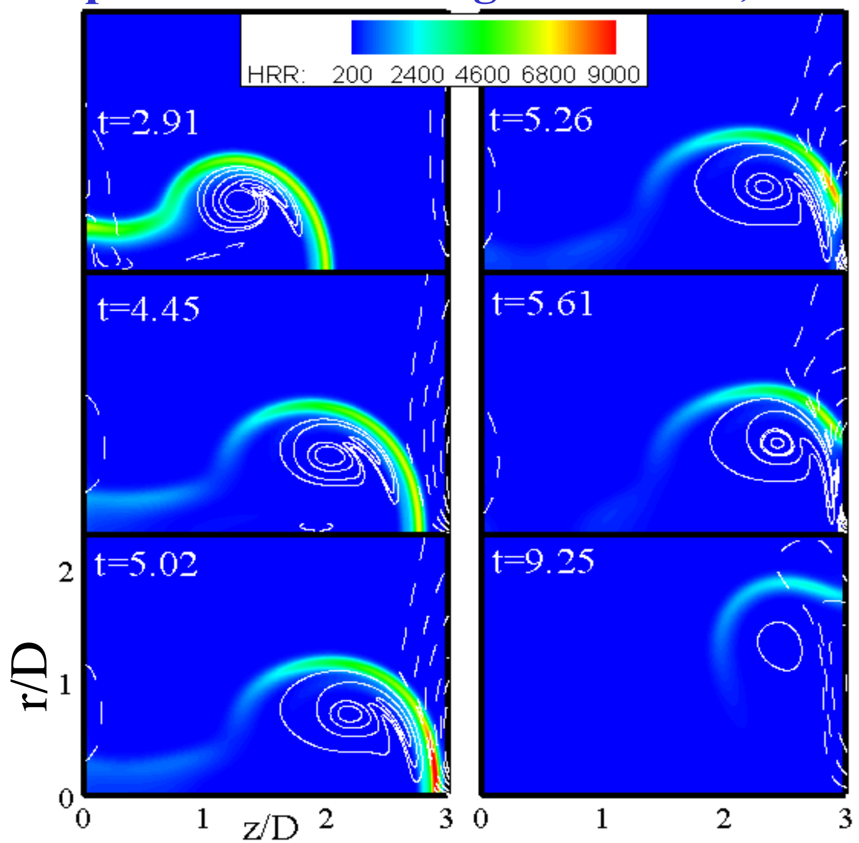
- **Developing a scalable, multi-site cyber-infrastructure for Science, Technology, Engineering and Mathematics (STEM) education and training called *MyDesignSpace*.**
- **Implementing a digital design repository to enhance instruction and learning in STEM education.**
- ***MyDesignSpace* will also help bridge existing gaps between secondary and collegiate STEM education.**



NYSCEDII
New York State Center for Engineering Design and Industrial Innovation
www.nyscedii.buffalo.edu

Understanding Combustion

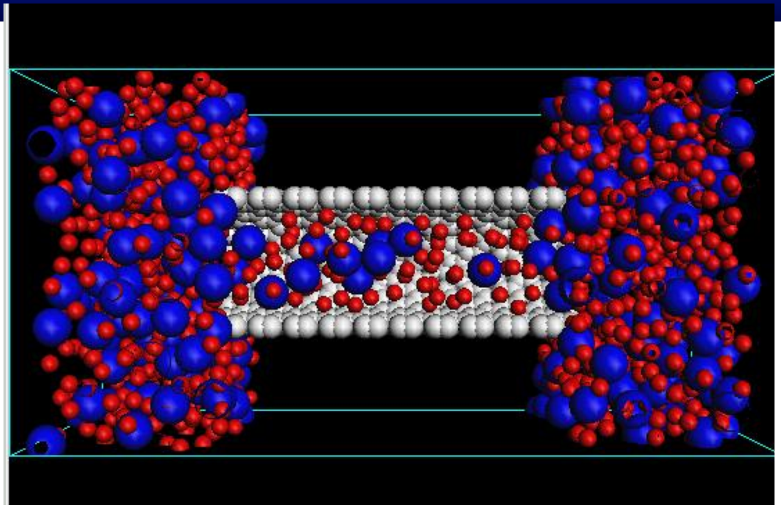
- Flame-wall interaction modeling for a non-premixed flame propelled by a vortex ring.
- In this figure different time instants are shown during the interaction. White line contours and color contours represent vortex ring and flame, respectively.
- Key Features:
  - Modeling of Detailed GRI3. Mechanism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

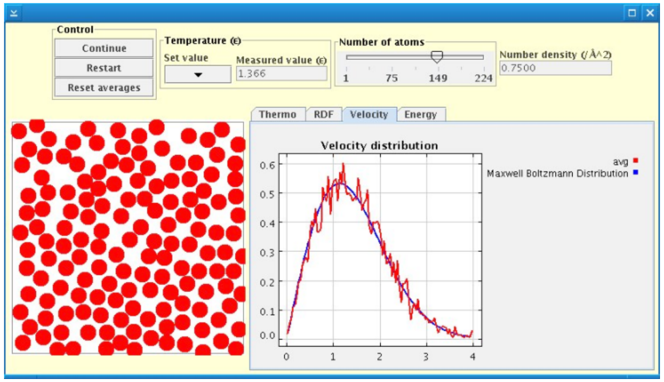


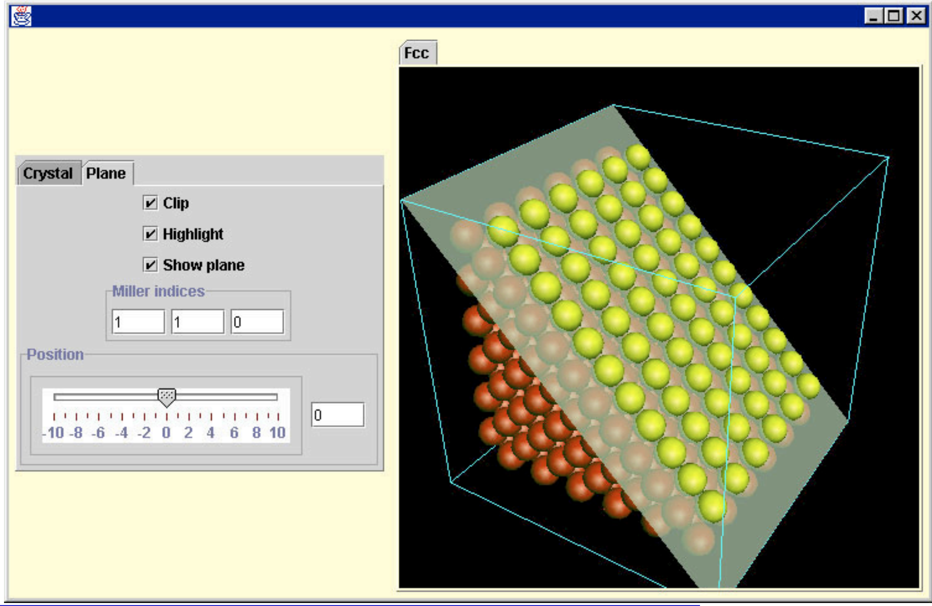
C. Madnia

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





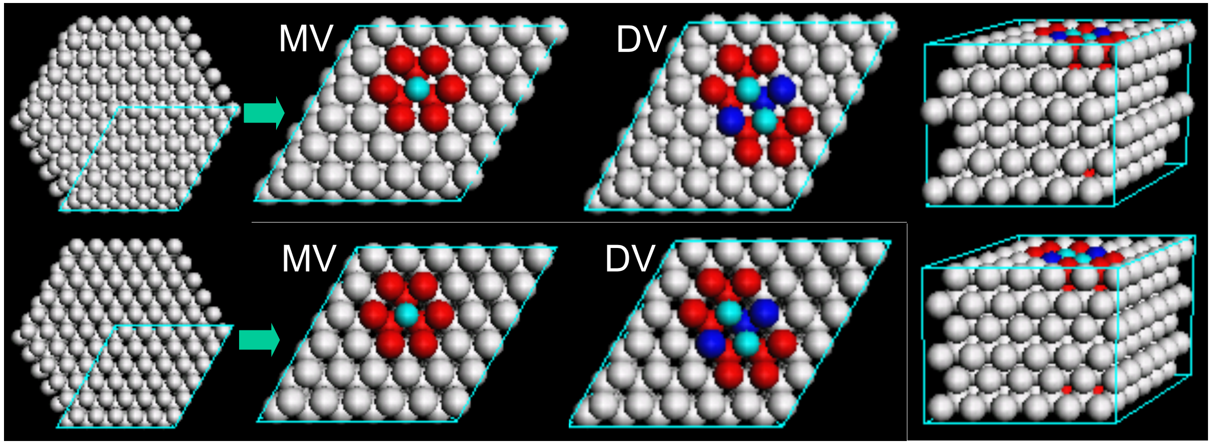


D. Kofke

Computational Materials Science

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

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

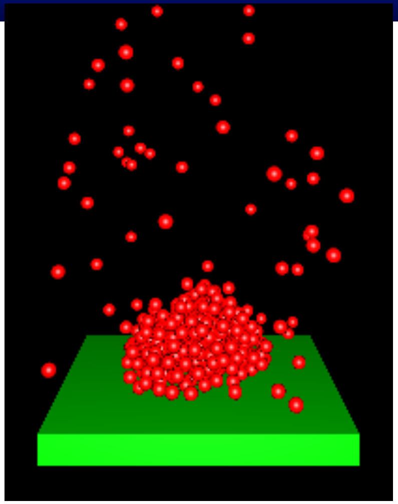


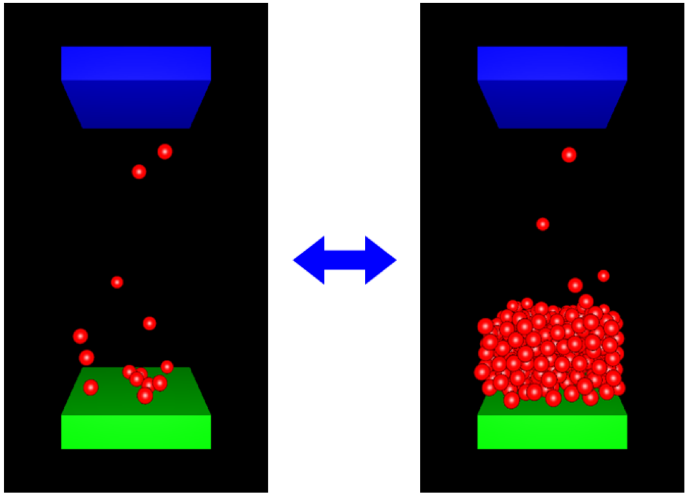
Simulation cells of solids with mono- and di-vacancies (light blue spheres), highlighting atoms neighboring the defects.

D. Kofke

Nano Confinement of Fluids

- Fluids in the presence of one or more surfaces exhibit rich phase behavior that can be strikingly different than that observed for bulk fluids
- A fundamental understanding of the relationship between a system's microscopic interactions and the phase behavior of a system is essential for the development of novel materials
- Molecular simulation is a useful tool for developing these relationships through the use of model systems that mimic the behavior of real fluids

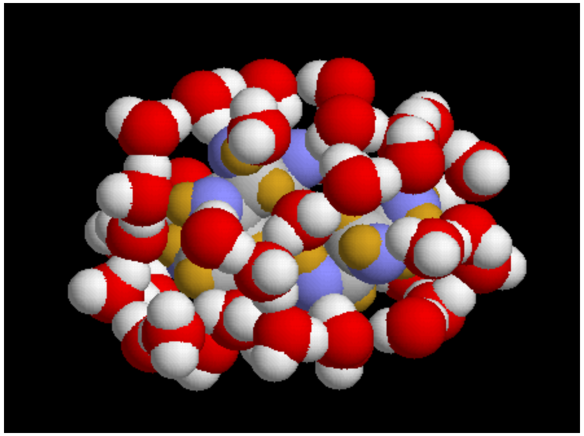




J. Errington

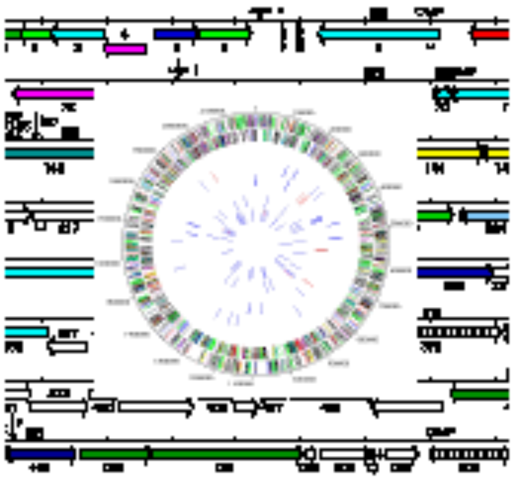
Aqueous Solutions

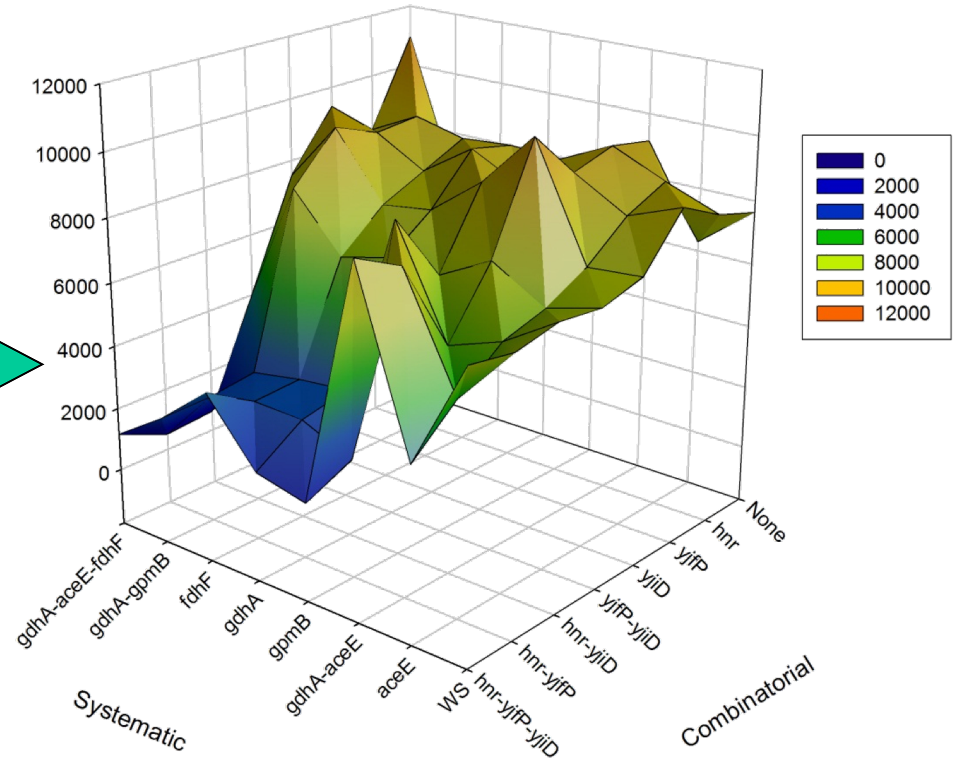
- The behavior of water and aqueous mixtures plays a key role in biology, chemistry, physics, and the design of many chemical and biological processes
- To gain a fundamental understanding of aqueous solutions, one must consider the effect the microscopic hydrogen-bond network has on the macroscopic properties of the system
- The goal of our program is to obtain a more complete understanding of aqueous systems using this molecular approach
- The diagram is a snapshot from a molecular dynamics simulation that depicts the organization of water molecules within 3.5 Å of a trehalose molecule



J. Errington

Designing Cellular Phenotypes



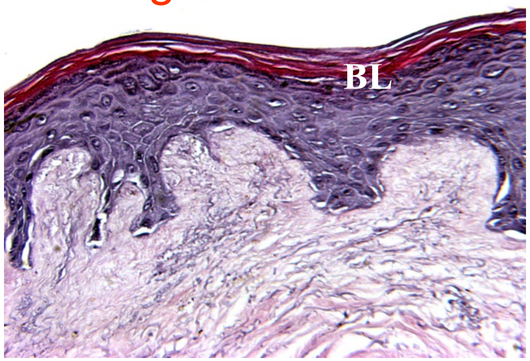


- Genome-wide metabolic models of sequenced microorganisms
- Optimization of metabolic and cellular phenotypes
- Goal is to design biocatalysts for the production of pharmaceuticals and high-value chemicals

M. Koffas

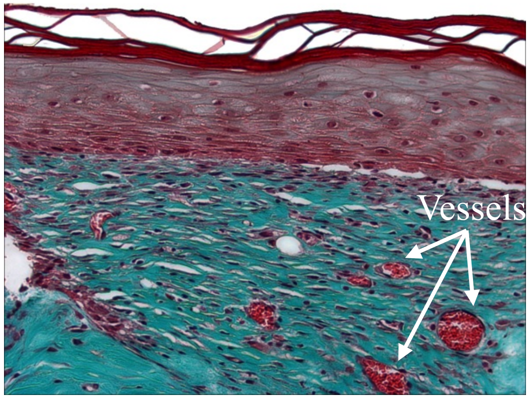
Stem Cells and Tissue Engineering

Bioengineered Skin

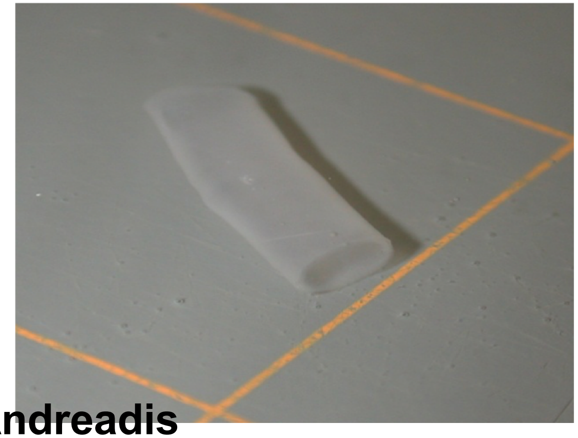


Genetically modified tissue engineered skin:
1. Wound healing e.g. burns, chronic wounds
2. Insulin delivery for treatment of diabetes
3. Development of a model to study tumor invasion

Bioengineered Skin Transplanted onto mouse

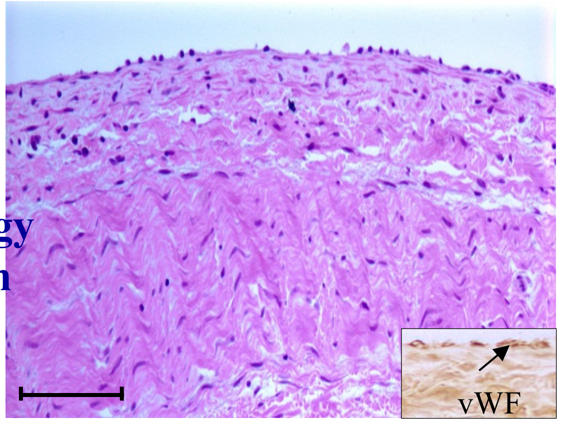


Bioengineered Blood Vessel



Tissue engineered blood vessels (TEV):
1. Bypass surgeries
2. Model to study mechano transduction and vascular biology
3. TEVs from bone marrow stem cells

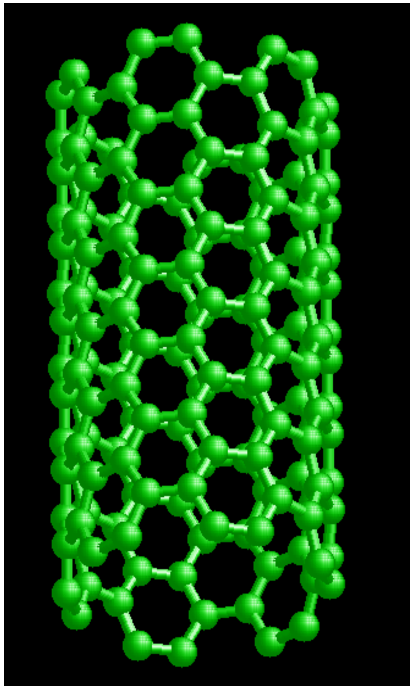
TEV Transplanted in jugular vein of lambs

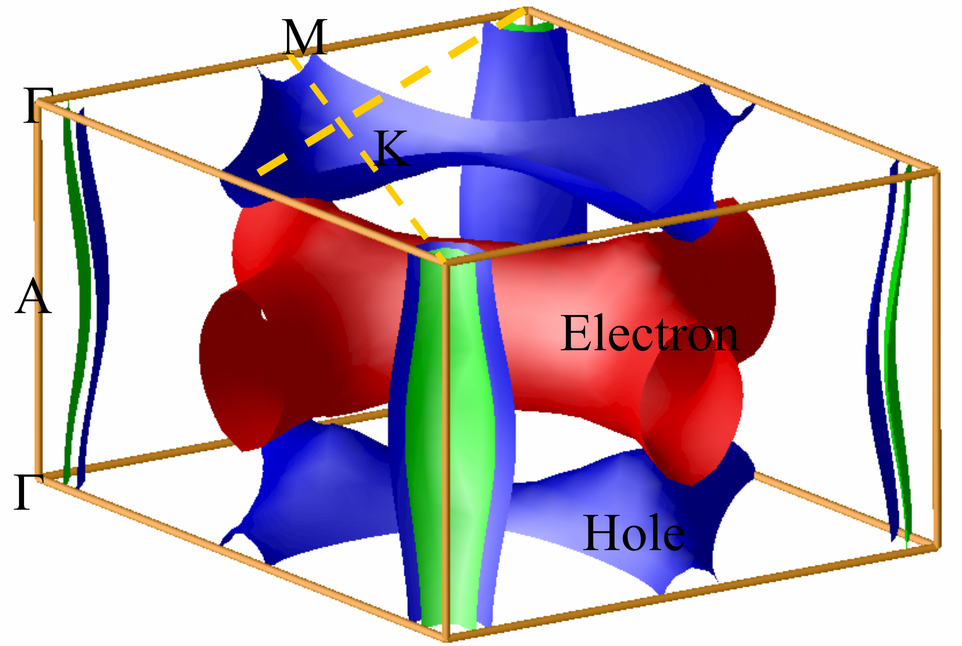


S. Andreadis

Designing New Materials

- Understand and predict materials properties
- Materials design from first-principles
- Development of new theoretical and computational techniques

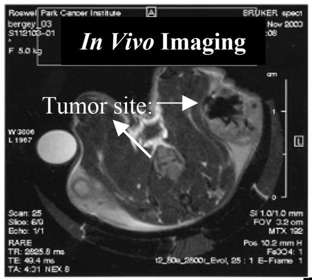




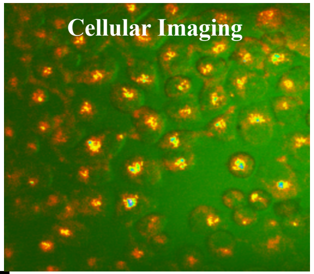
P. Zhang

Nanomedicine Program

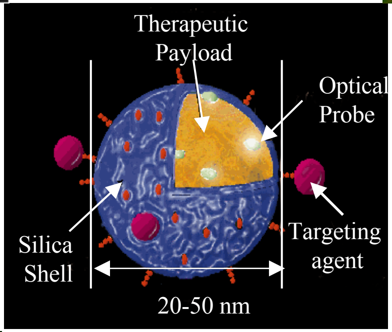
World class Research Program Melding Nanotechnology with Biomedical Sciences



Multi-Modal Imaging

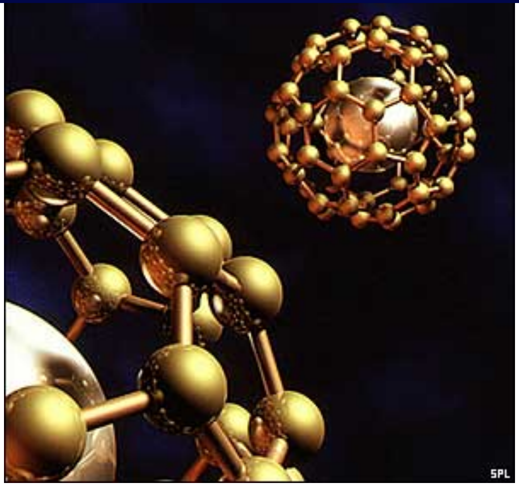


BioCompatibility/ Distribution



***In Vivo* Sensing**

Targeted Therapy



Building from the Bottom Up

State of the Art Molecular Imaging and Nanocharacterization Facilities

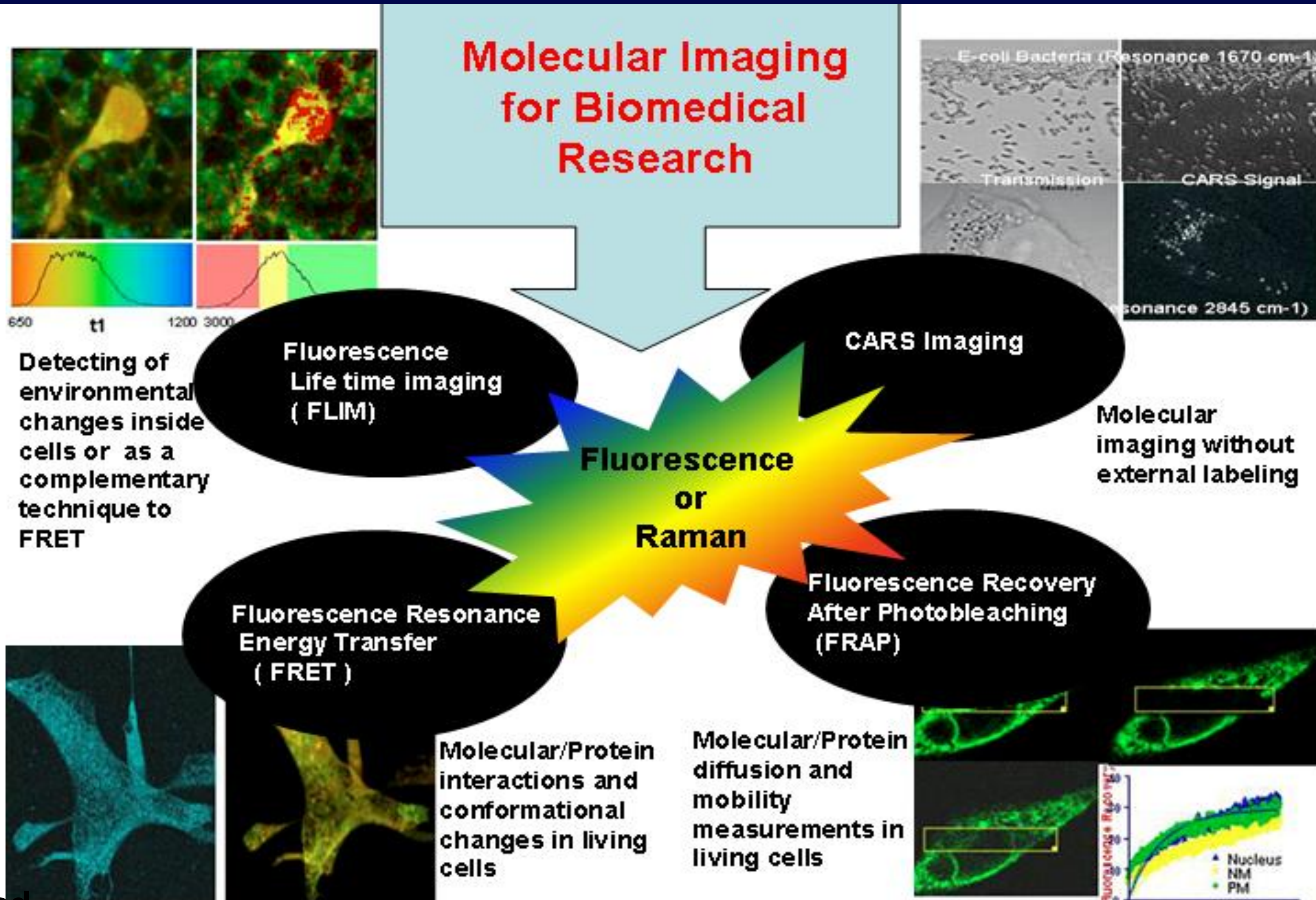
- Multiphoton Laser Scanning System
- Confocal Imaging including FRET, FLIM & FRAP analysis
- Coherent Anti-Stokes Raman Imaging
- Optical Trapping/Dissection
- Advanced Laser Systems

P. Prasad

www.biophotonics.buffalo.edu

"Leading the Way to Technology through Innovation"

Molecular Imaging

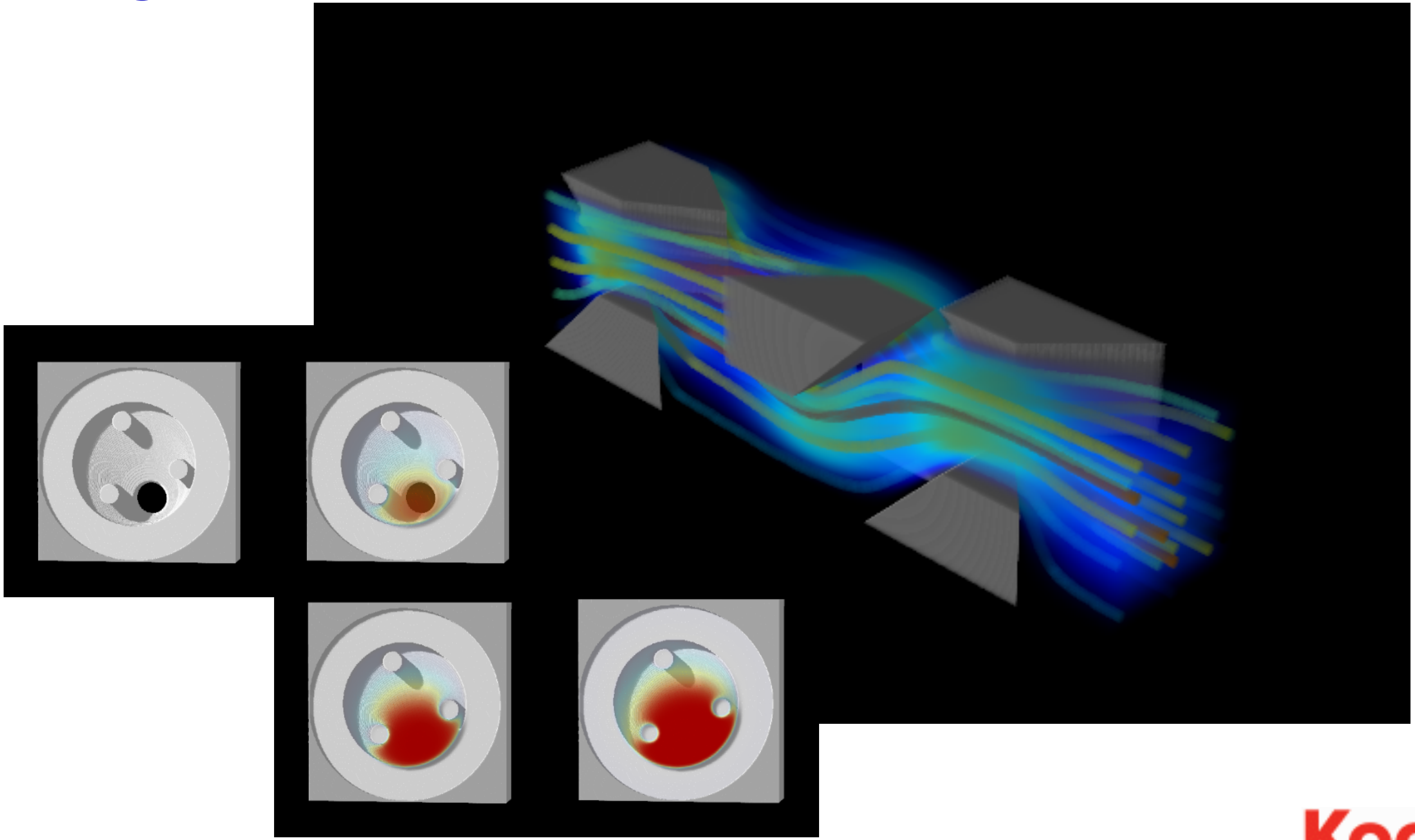


P. Prasad

"Lighting the Way to Technology through Innovation"

Industrial 3D Flow Analysis

- Modeling of Complex 3D and Mixing Flows for Part Analysis and Design



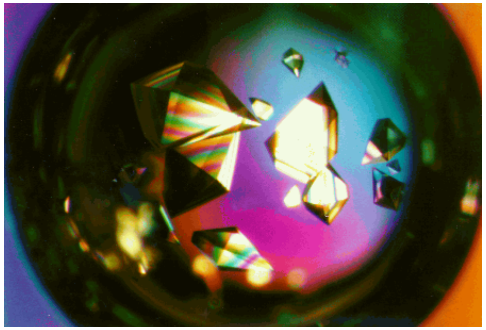
Kodak

Shake-and-Bake

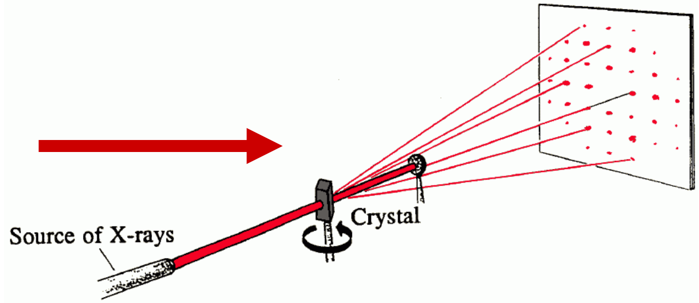
Molecular Structure Determination from X-Ray Crystallographic Data

Molecular Structure Determination via *Shake-and-Bake*

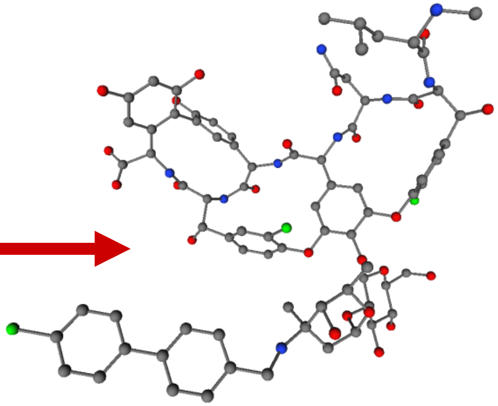
- *SnB* Software by UB/HWI
  - IEEE “Top Algorithms of the Century”
- Worldwide Utilization
- Critical Step
  - Rational Drug Design
  - Structural Biology
  - Systems Biology
- Vancomycin
  - “Antibiotic of Last Resort”
- Current Efforts
  - Grid
  - Collaboratory
  - Intelligent Learning



1. Isolate a single crystal



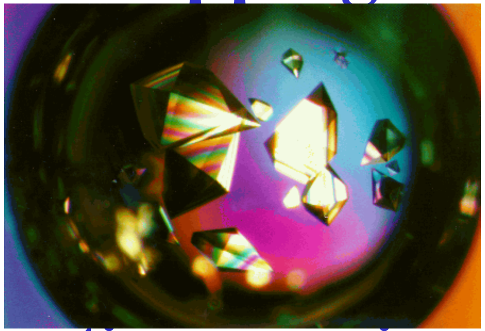
2. Perform the X-Ray diffraction experiment



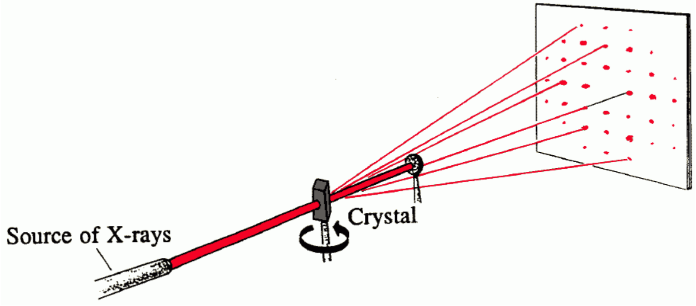
3. Determine the crystal structure

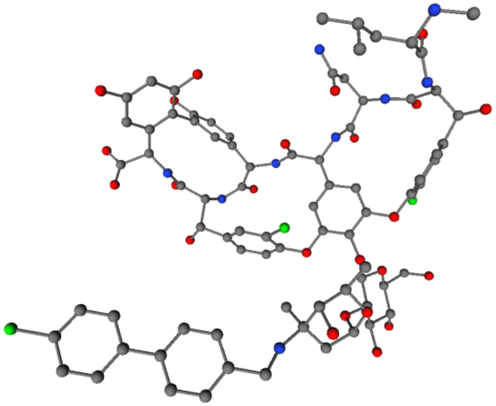
X-Ray Crystallography

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



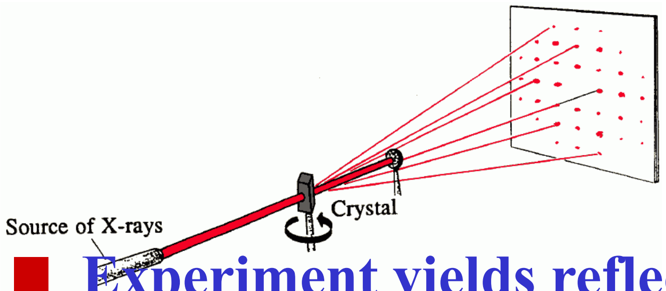
1. Isolate a single crystal.
2. Perform the X-Ray diffraction experiment.



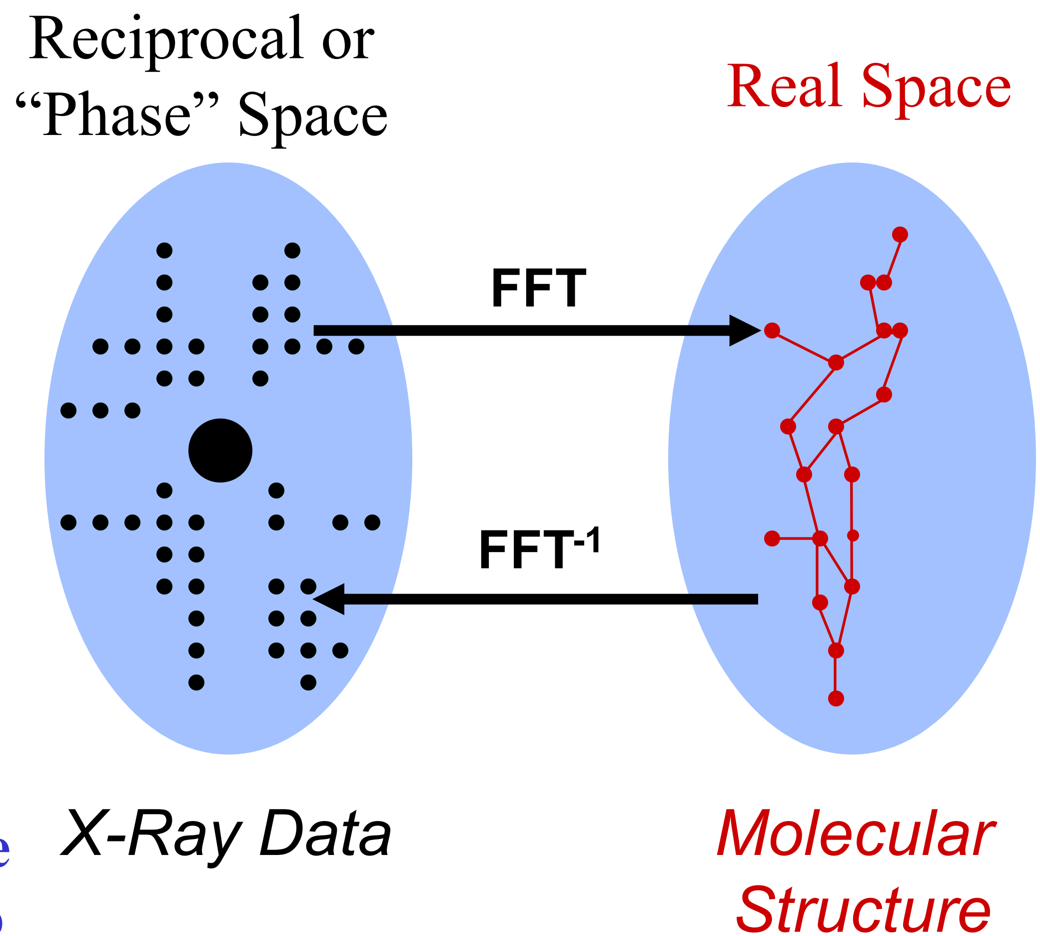


3. Determine molecular structure that agrees with diffration data.

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.



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.

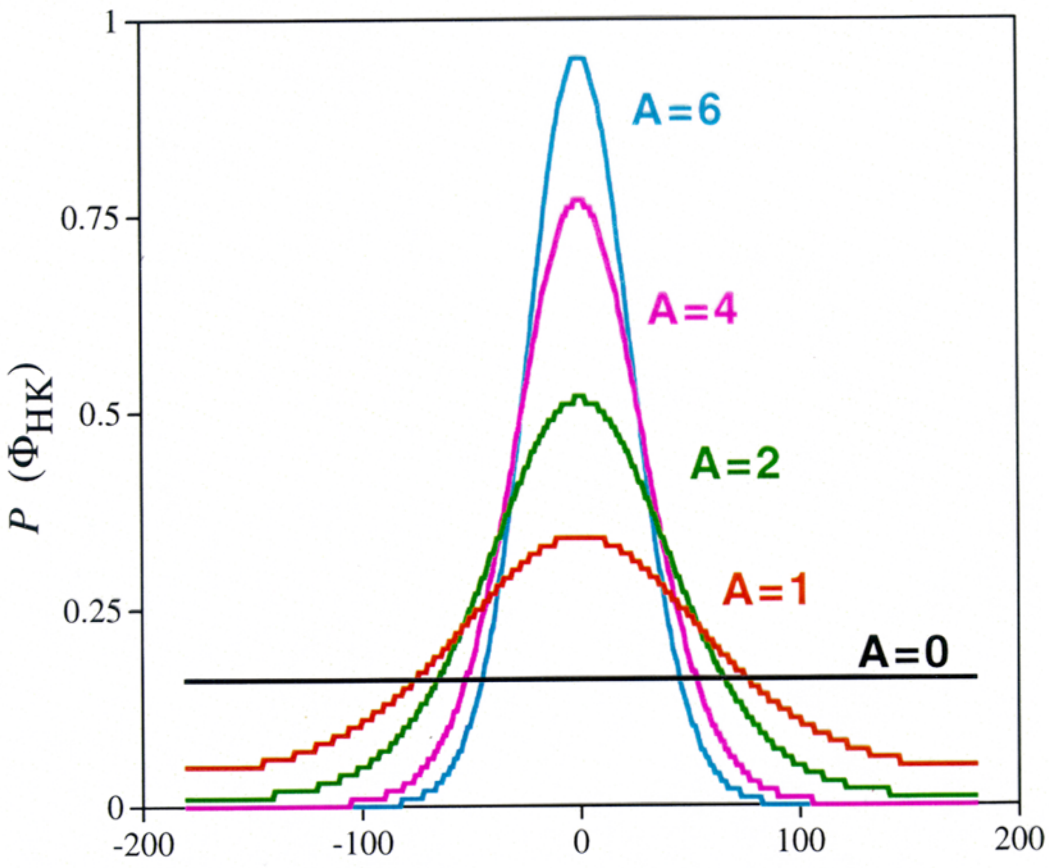
Normalized Structure-Factor Magnitudes: $|E_H|$

$$E_H = |E_H| \exp(i\phi_H)$$

$$|E_H| = \frac{|F_H|}{\left\langle |F_H|^2 \right\rangle^{1/2}} = \frac{k\left\langle \exp[-B_{iso}(\sin\theta)^2/\lambda^2 \right\rangle^{-1} |F_H|_{meas}}{\left(\varepsilon_H \sum_{j=1}^{N} f_j^{\,2}\right)^{1/2}}$$

- $\langle|E|\rangle$ constant for concentric resolution shells.
- $\langle|E|\rangle$ constant regardless of reflection class (ε_H correction factor).
- The *renormalization* condition, $\langle|E|^2\rangle=1$ is always imposed

Cochran Distribution



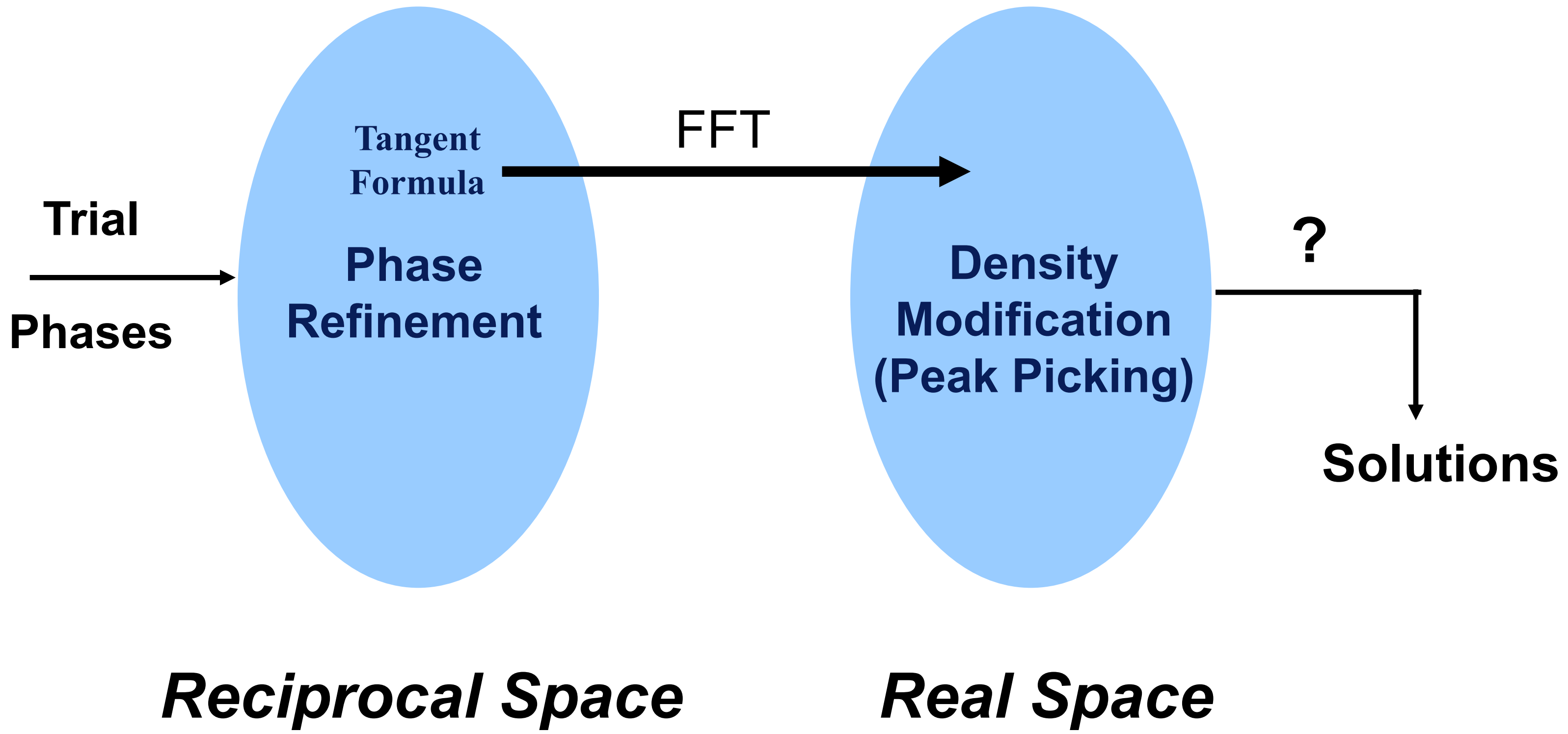
$\Phi_{HK}=\phi_H + \phi_K + \phi_{-H-K}$

- *N*=non-H atoms in unit cell
- Each triplet of phases or structure invariant, Φ_{HK}, has an associated parameter

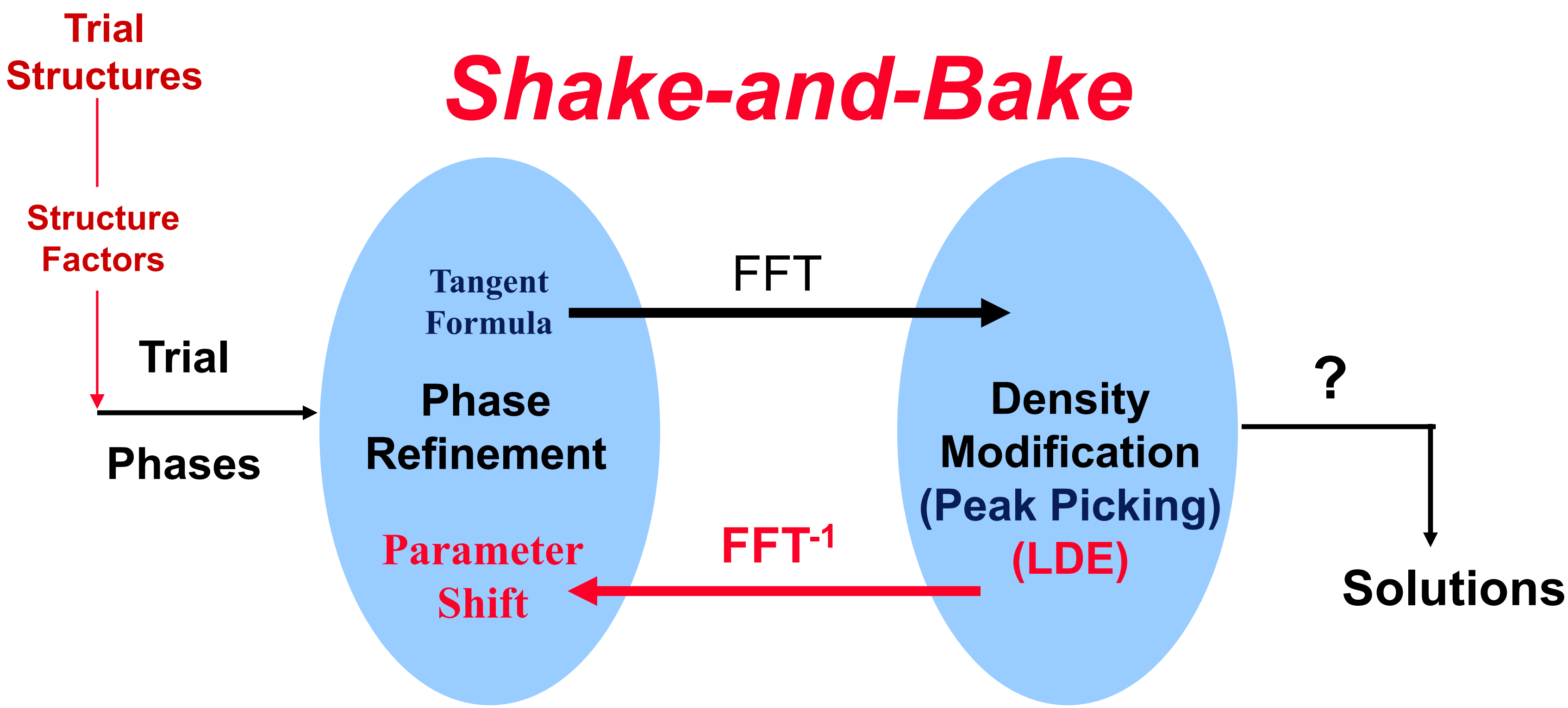
$$A_{HK}=2|E_H E_K E_{-H-K}|/N^{1/2}$$

- A_{HK} is large if
  - $|E_H|$, $|E_K|$, $|E_{-H-K}|$ are large
  - *N* is small
- If A_{HK} is large, $\Phi_{HK} \approx 0$

Conventional Direct Methods



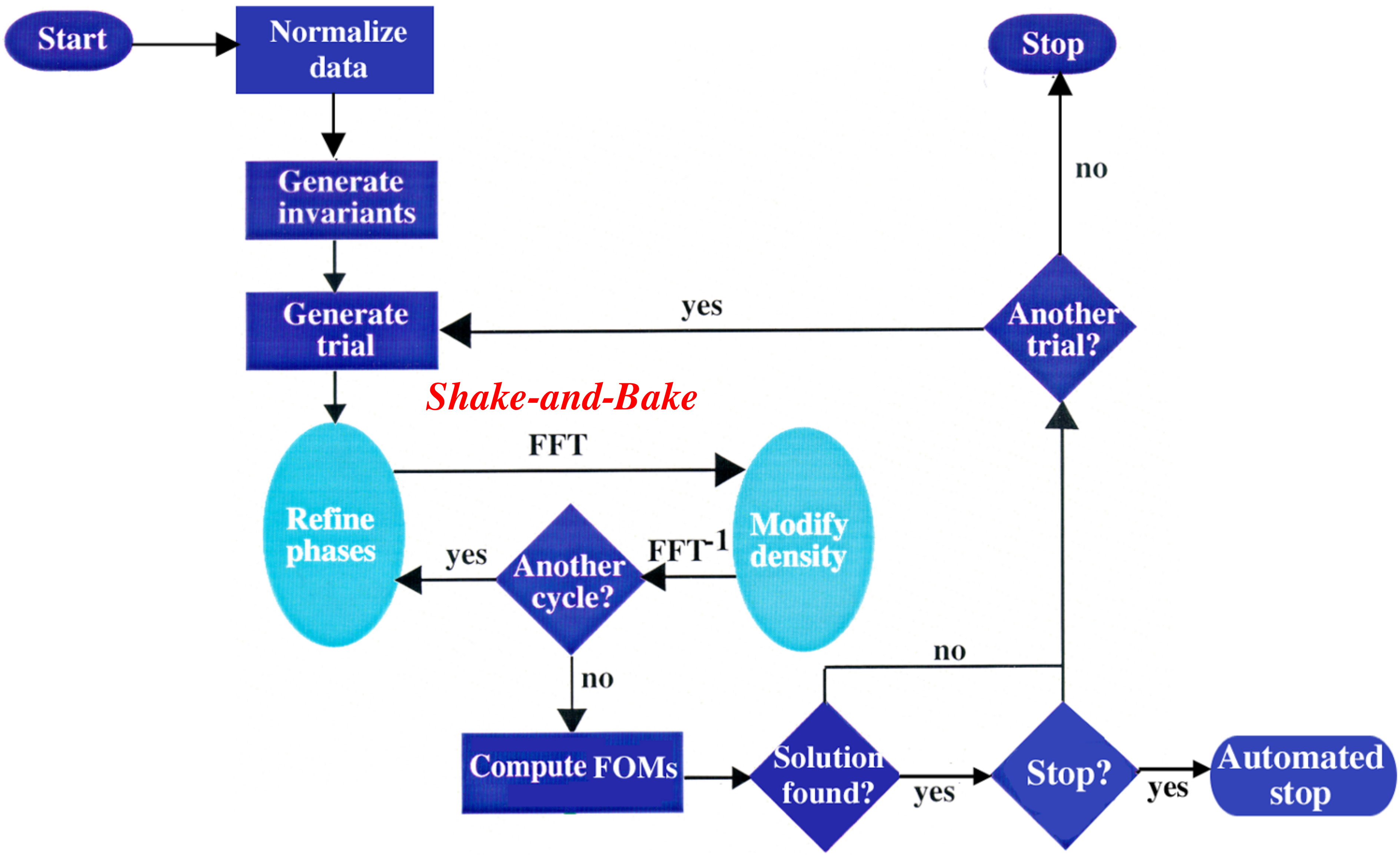
Shake-and-Bake Method: Dual-Space Refinement



Reciprocal Space
"Shake"

Real Space
"Bake"

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

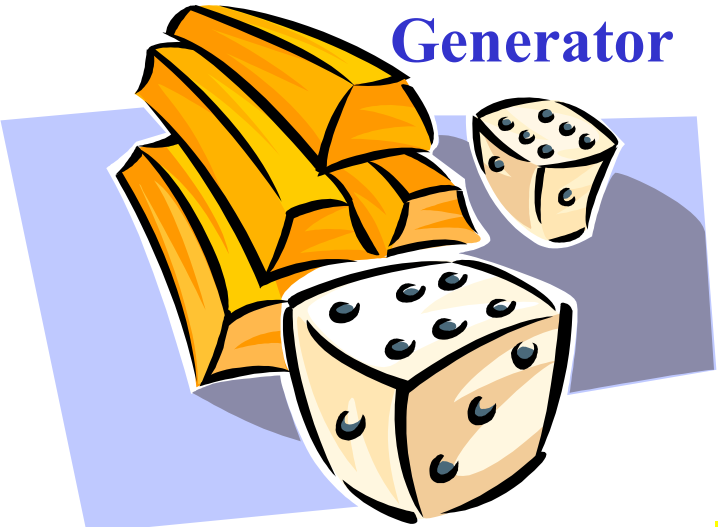
Triplets

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

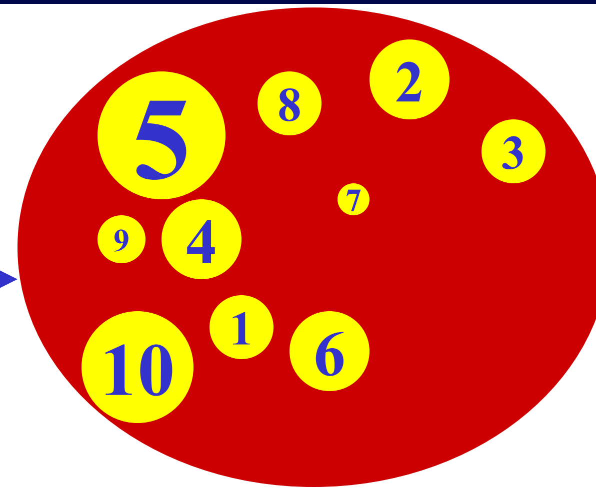
Getting Started: Random Atoms

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

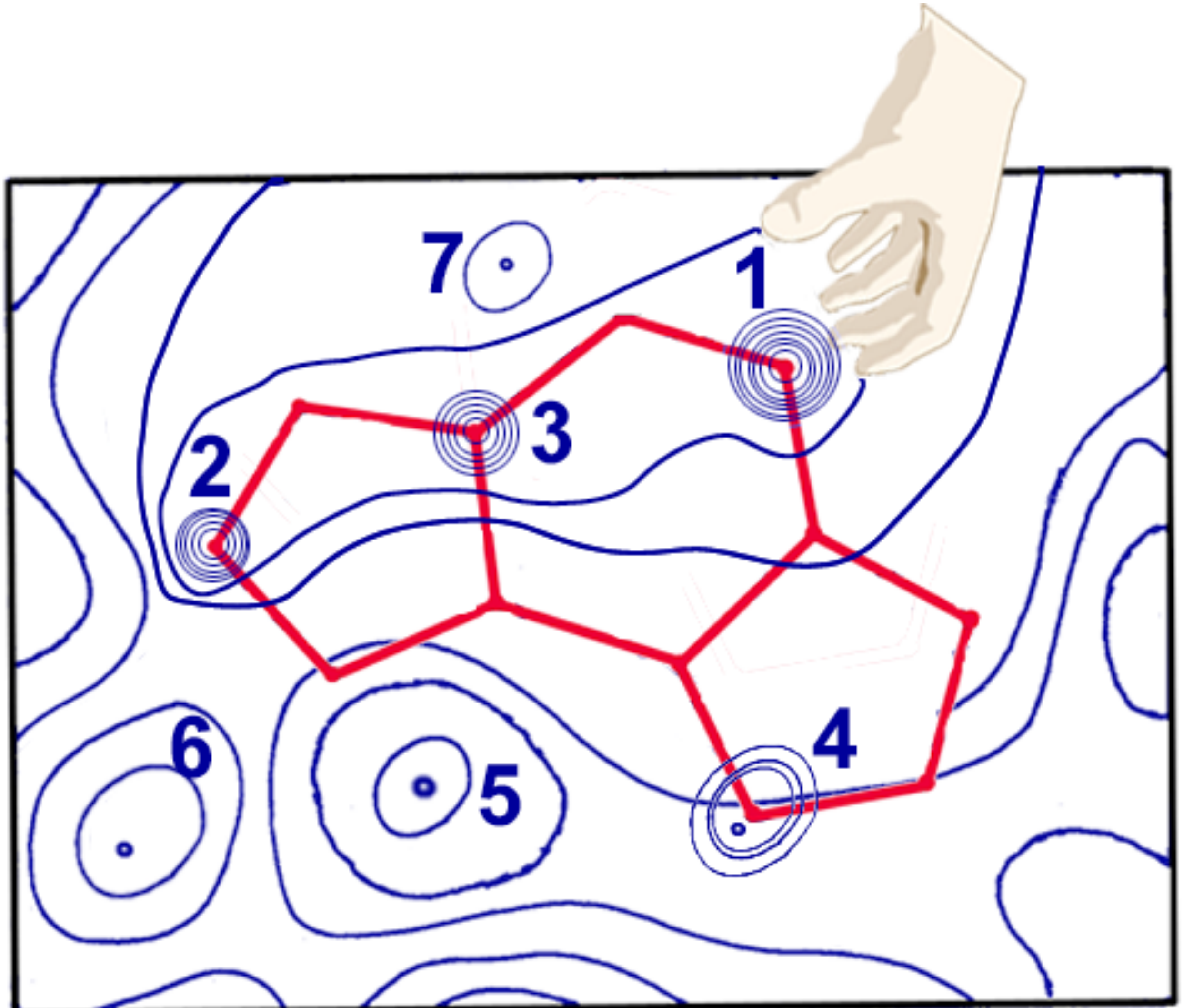
$$R(\phi) = \frac{1}{\sum_{H,K} W_{HK}} \sum_{H,K} W_{HK} \left(\cos\Phi_{HK} - \frac{I_1(W_{HK})}{I_0(W_{HK})} \right)^2$$

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



Sorted Trials

Sorted Trial Data

Trial	Cycle	Refl Phased	Rmin	R Cryst.	CC	R Ratio	Peak Ratio
97	56	836	0.349	0.27	0.45	0.05	1.2
51	56	836	0.350	0.26	0.43	0.03	1.1
82	56	836	0.350	0.26	0.44	0.03	1.1
30	56	836	0.351	0.26	0.45	0.03	1.0
56	56	836	**0.351**	**0.27**	**0.48**	0.03	1.1
----	----	----	----	----	----	----	----
93	56	836	**0.506**	**0.36**	**0.36**	0.08	1.0
81	56	836	0.515	0.38	0.37	0.18	2.3
69	56	836	0.522	0.37	0.39	0.21	2.6
63	56	836	0.523	0.37	0.39	0.21	2.5
16	56	836	0.525	0.39	0.43	0.21	2.7

Solutions

Nonsolutions

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 $k = \sum |E_o| / \sum |E_c|$

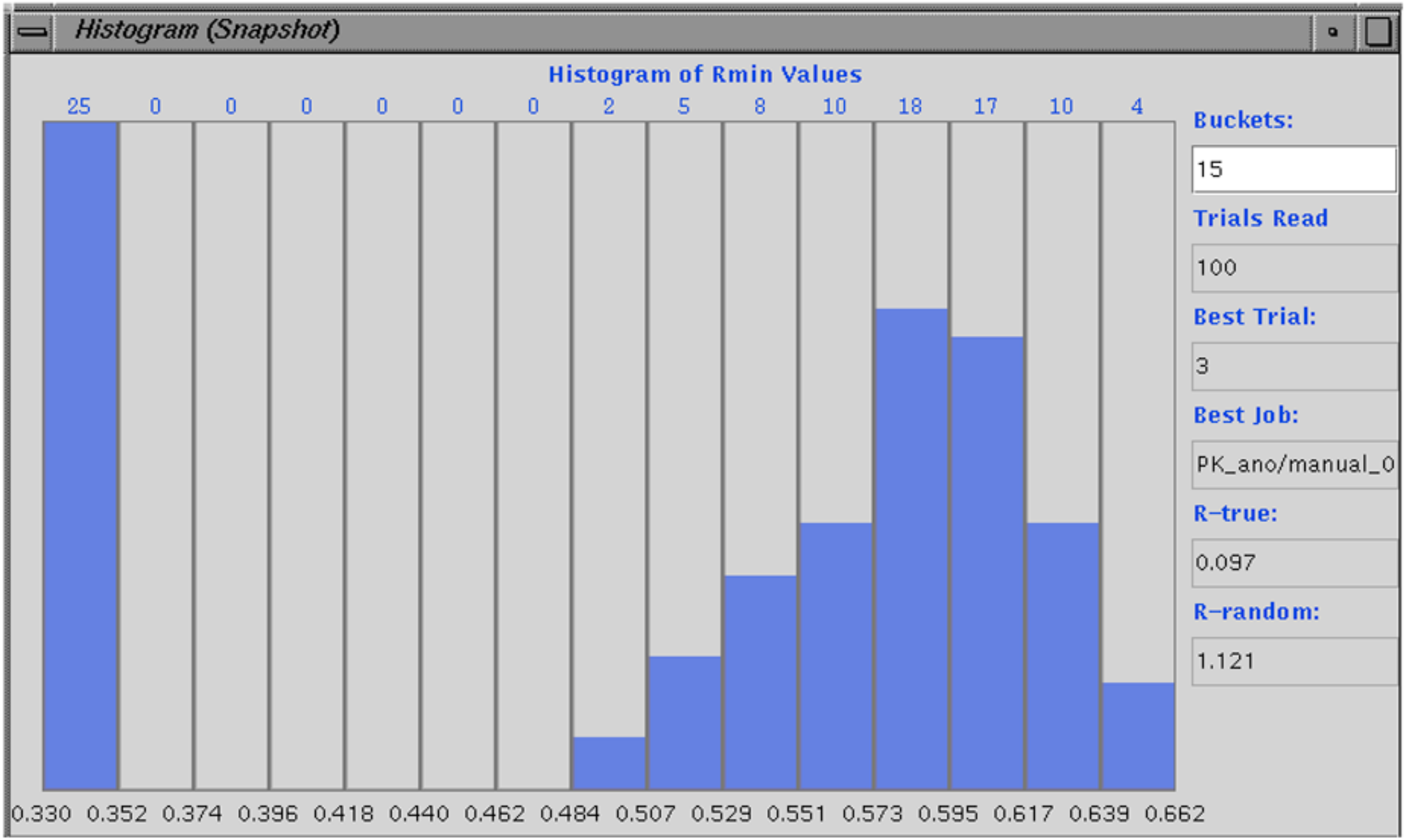
3. Correlation Coefficient (CC)

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

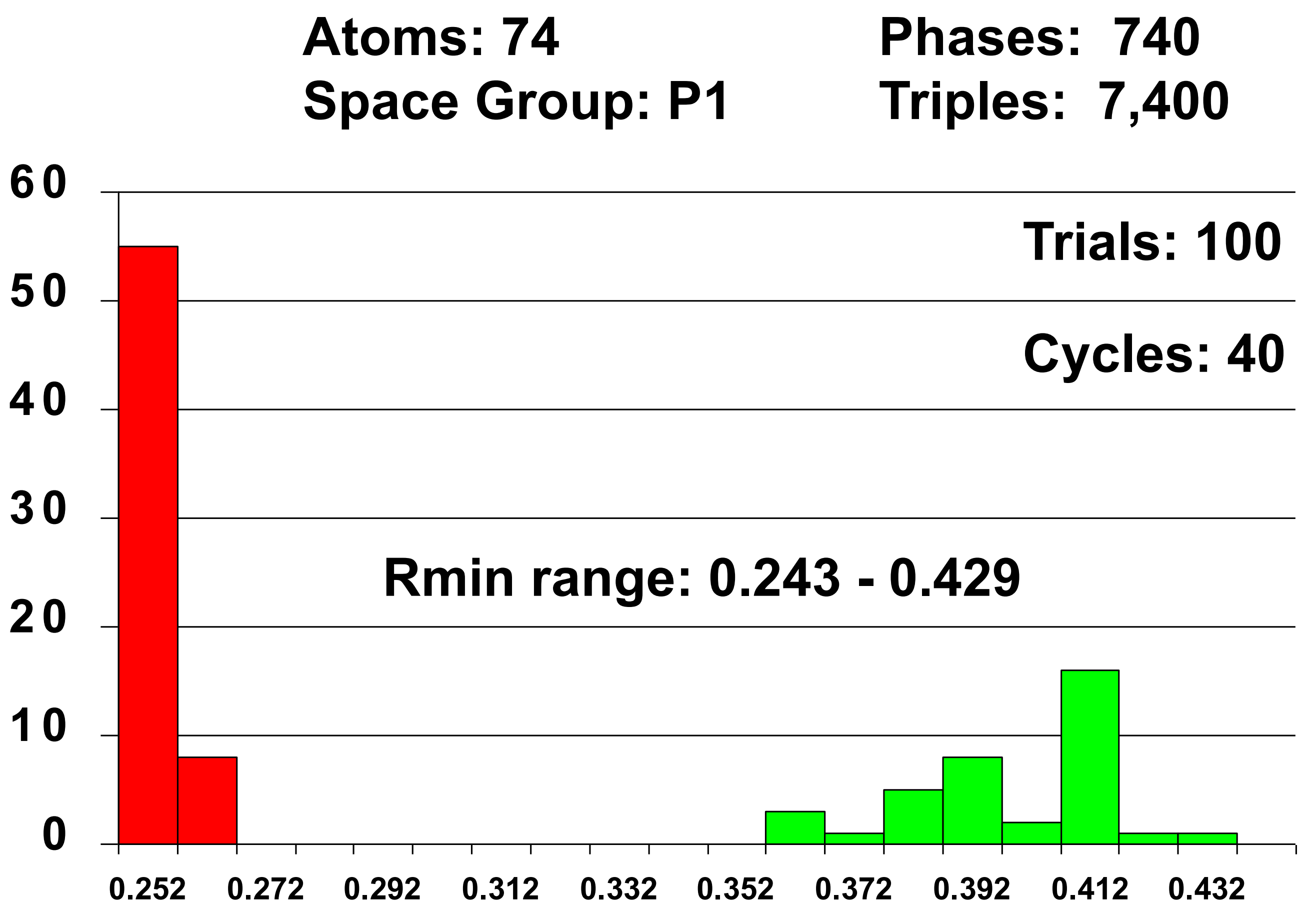
$$\{[\sum wE_o^4 \sum w - (\sum wE_o^2)^2][\sum wE_c^4 \sum w - (\sum wE_c^2)^2]\}^{1/2}$$

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

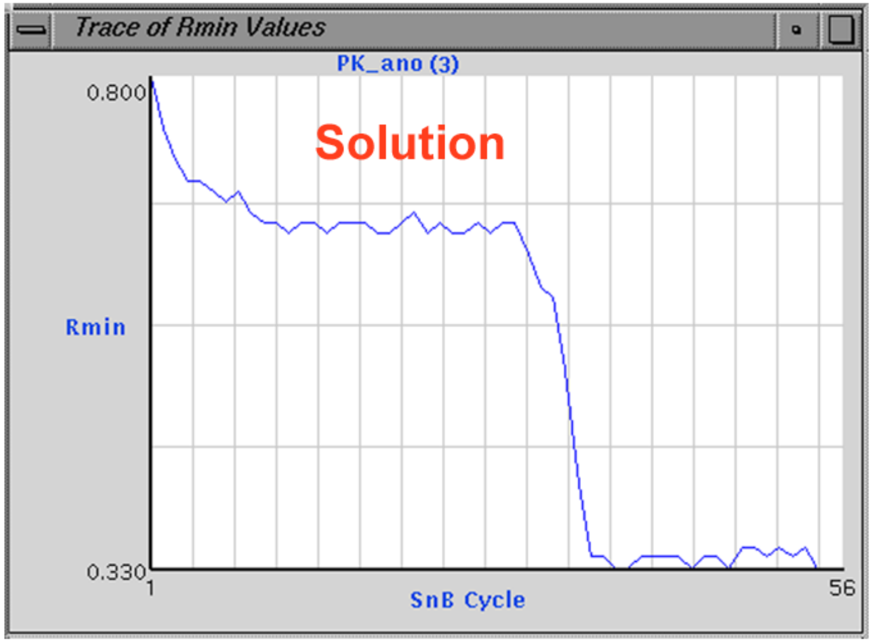
Ph8755: *SnB* Histogram

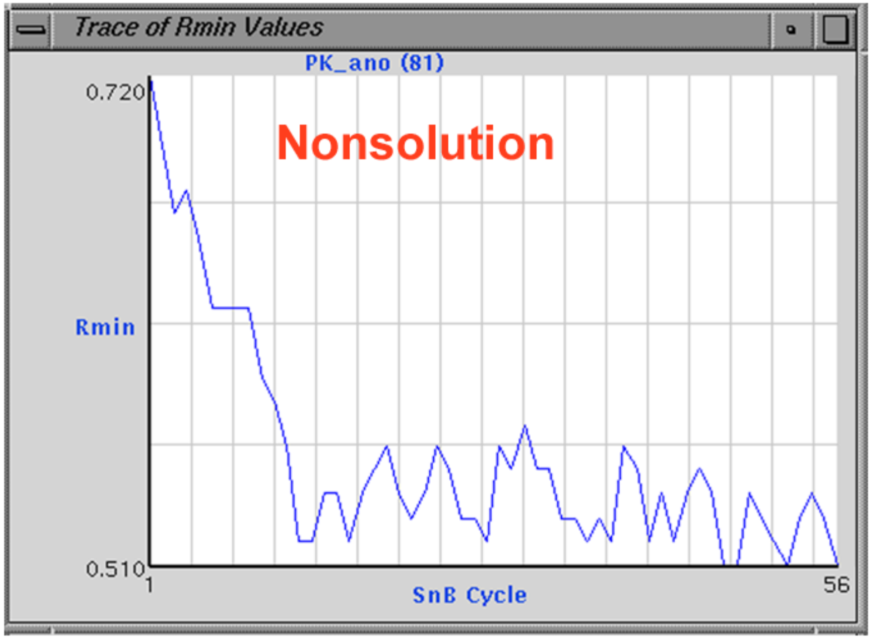


Ph8755: *SnB* Histogram

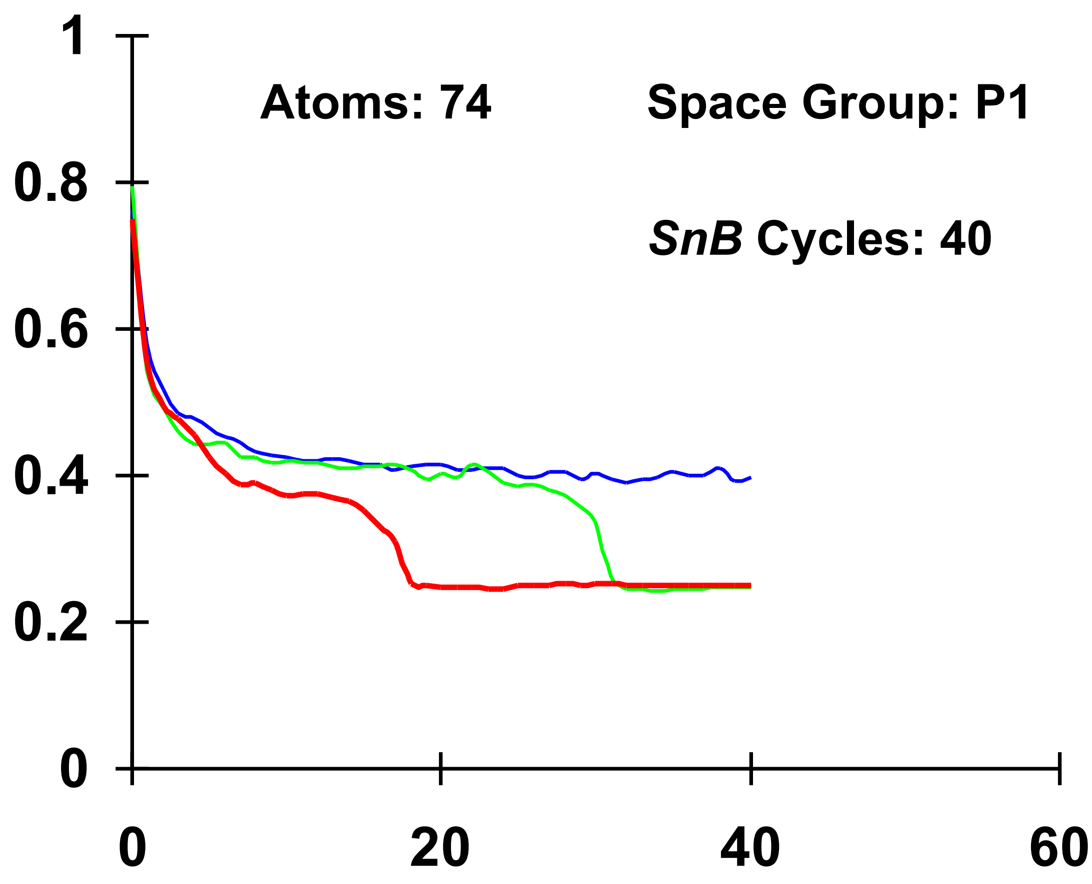


Minimal Function Traces





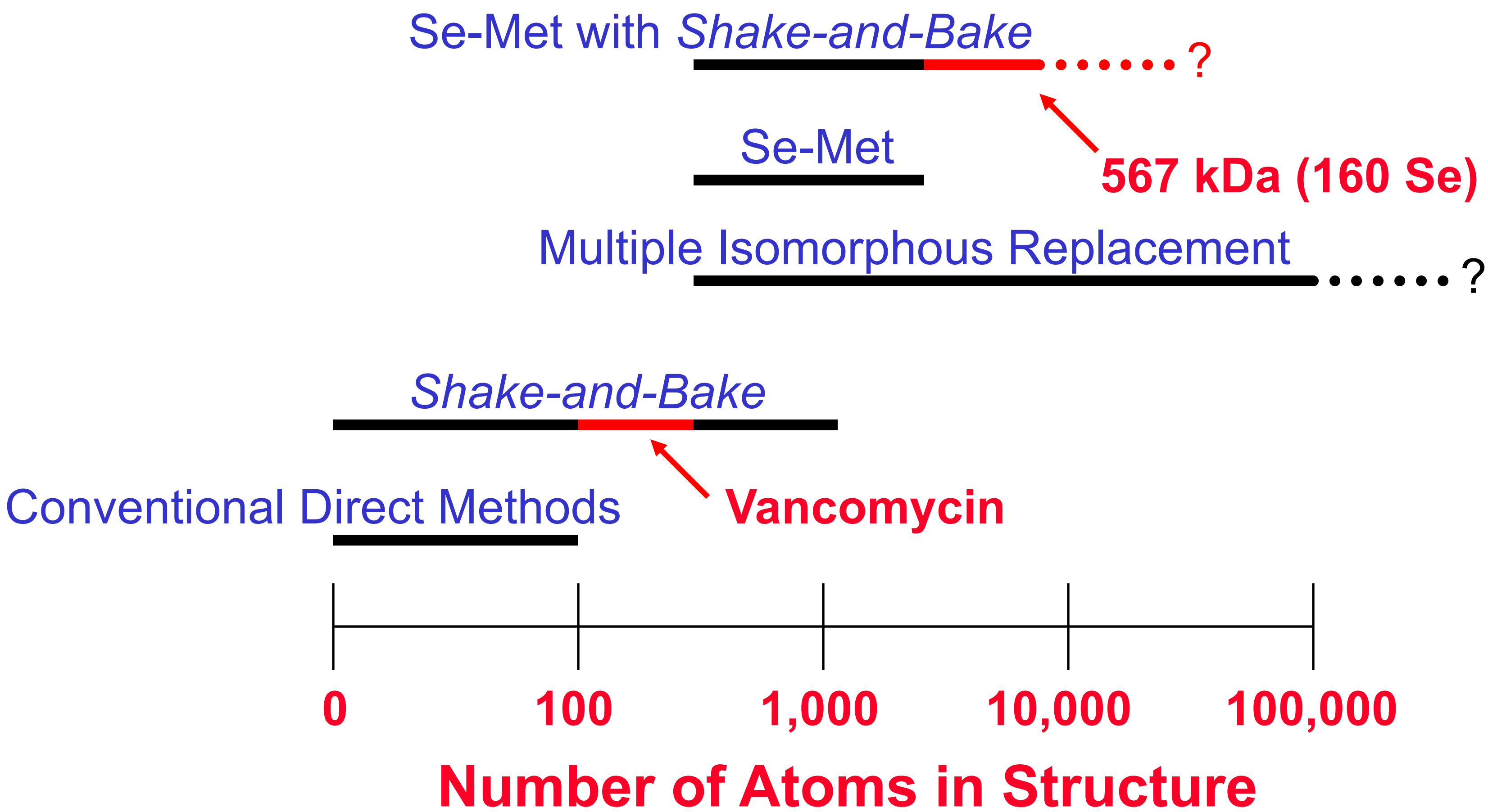
Ph8755: Trace of *SnB* Solution



Default *SnB* Parameters (given *n* atoms)

Parameter	Full Structures	Substructures
Phases	$10n$	$30n$
Triplet Invariants	$100n$	$300n$
Cycles		
$n<100$	$n/2$	$2n$
$n>100$	n	$2n$
Peaks		
$n<100$	n	n
$n>100$	$0.8n$	$0.8n$

Phasing and Structure Size

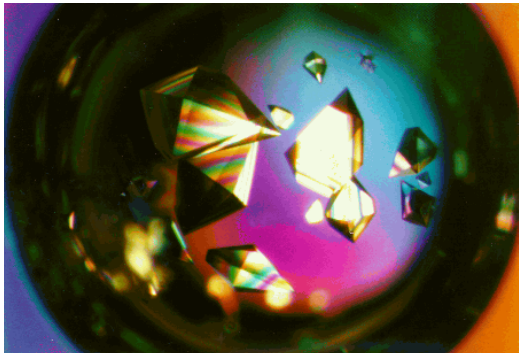


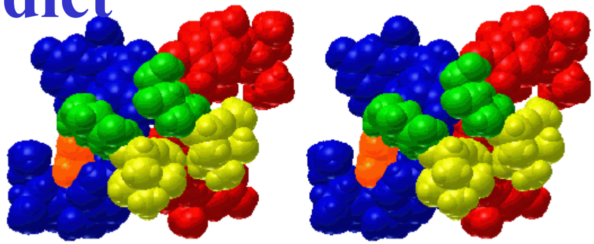
Shake-and-Bake Applications: Structure Size and Data Resolution

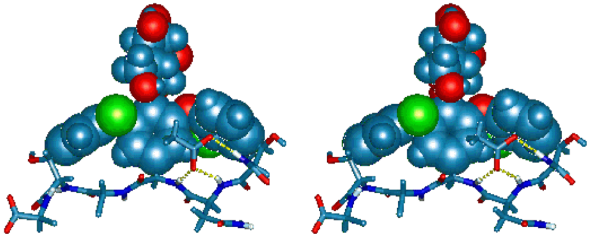
- Basic Data (Full Structure)
  - ~750 unique non-H atoms (equal)
  - ~2000 such atoms including 8 Fe's
  - 1.1-1.2Å data (equal atom)
  - 1.3-1.4Å data (unequal atoms, sometimes)
- SAS or SIR Difference Data (substructures)
  - 160 Se (567 kDa / ASU)
  - 3-4Å data
  - 5Å truncated data have also worked

Vancomycin

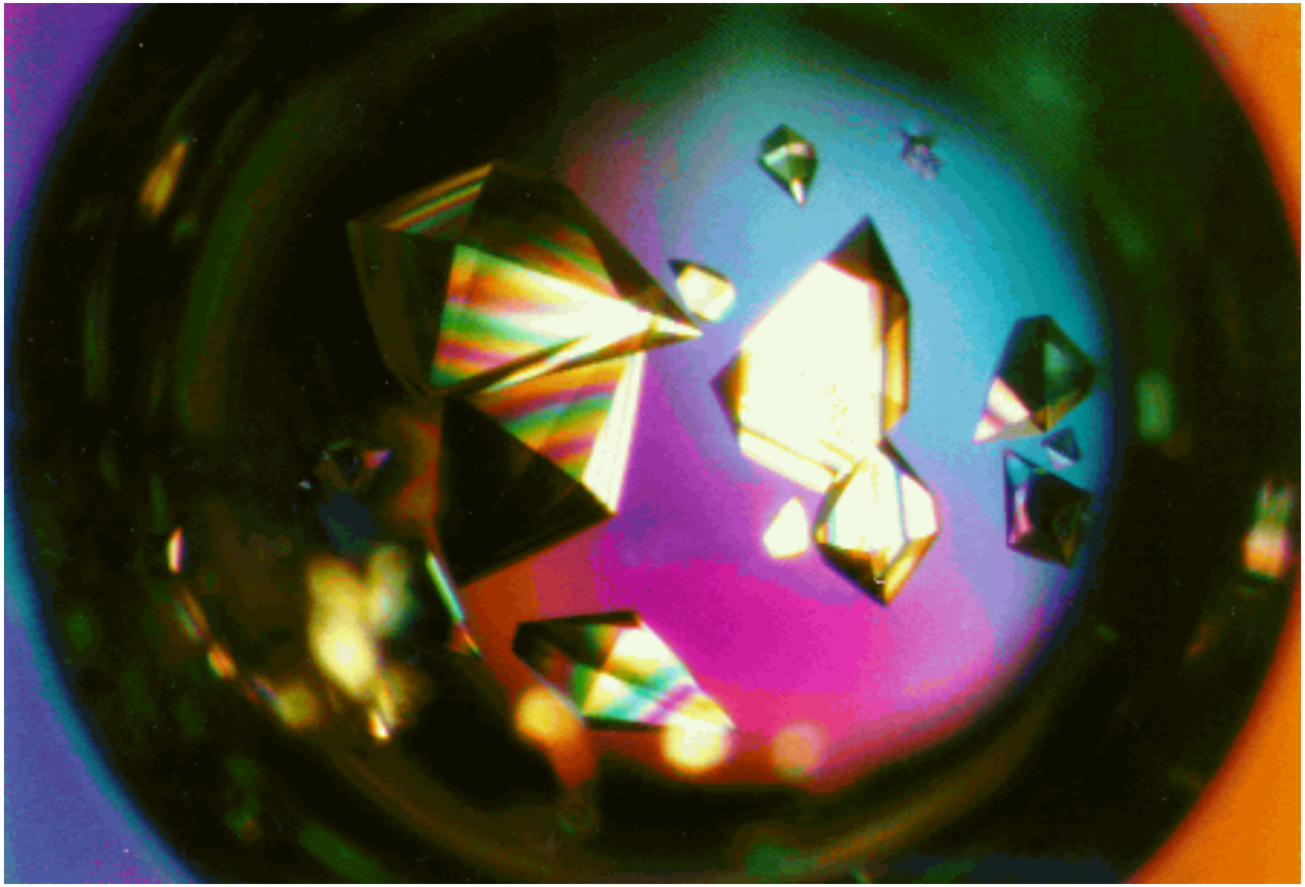
- Interferes with formation of bacterial walls
- *Last line of defense* against deadly
  - streptococcal and staphylococcal bacteria strains
- Vancomycin resistance exists (Michigan)
- Can't just synthesize variants and test
- Need structure-based approach to predict
- Solution with *SnB* (*Shake-and-Bake*)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



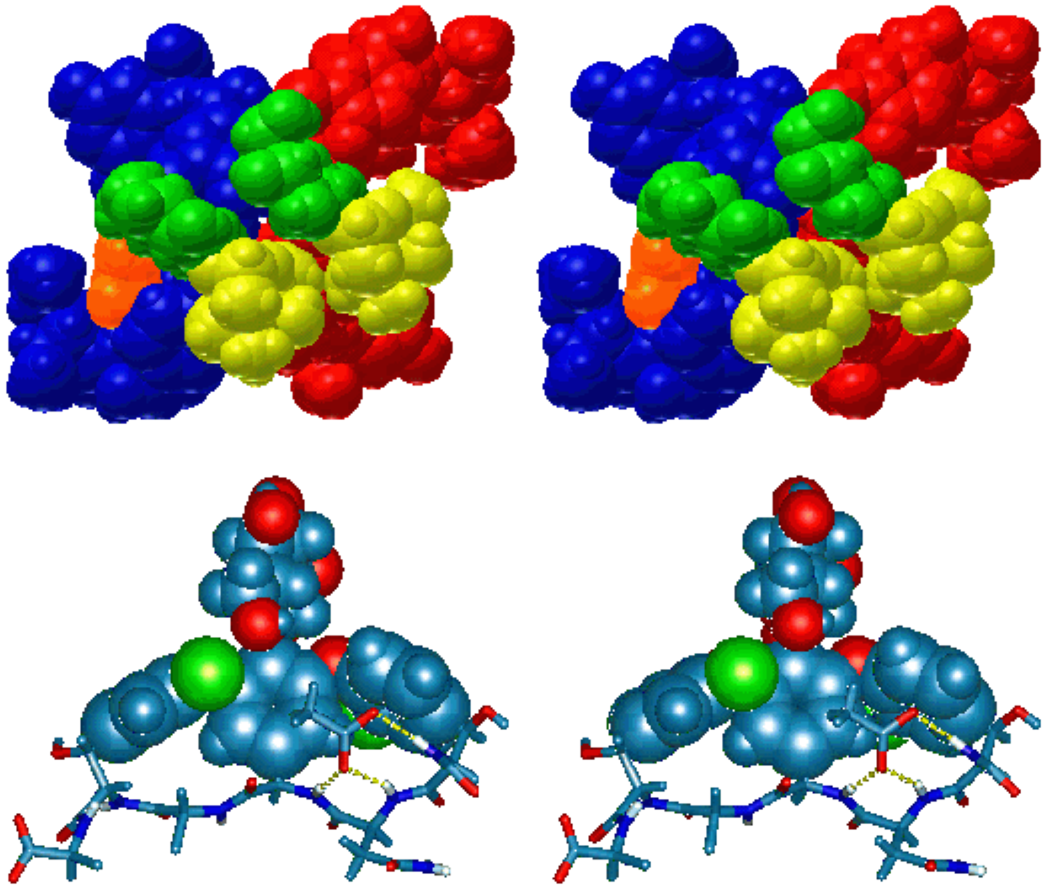




Vancomycin Crystal (courtesy of P. Loll)



Vancomycin Crystal Structure Views (courtesy of P. Loll & P. Axelsen)



Grid Computing

iVDgL

NEESgrid

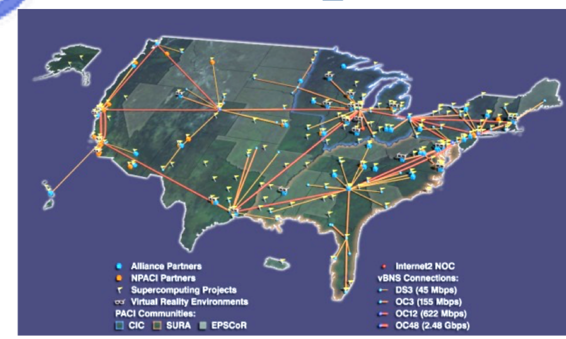
Data GRID

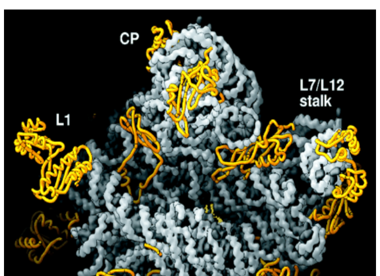
GLOBAL GF

GriPhyN
Data Intensive Science

European GRID Forum

SDSC/UCSD • NCSA/UIUC • Caltech • ANL
TERAGRID
NSF PACI





PDB
PROTEIN DATA BANK

United States virtual observatory

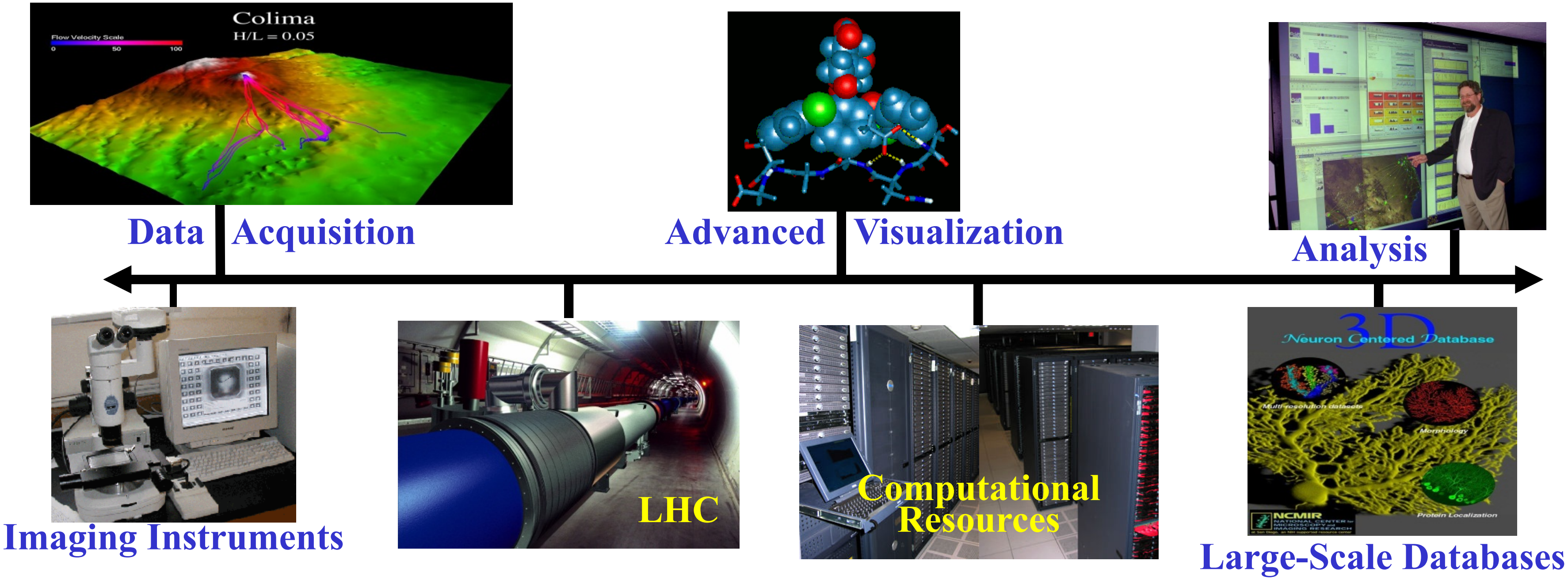
EUROGRID

APAN
Asia-Pacific Advanced Network

Open Science Grid

GRID
Advanced Center for Computational Research Data Center

Grid Computing Overview



- 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

"Middleware"

- Intermediate Software Layer between Application Codes and Grid Resources
- Required for applications, users, and resource providers to operate effectively in a manner transparent to the user
- Security; Resource Management; Data Access; Policies; Accounting;
- Globus; Condor
- Checks availability of Resources
  - CPUs; Storage; Networking; Render Farms; etc.
- Scheduling / Workload Management System
- Resource Broker
  - Evaluates Job and Breaks Up/Submits

NSF Middleware Initiative (NMI)

- Develop, improve, and deploy a suite of reusable software components for use in national-scale "cyberinfrastructure".
- APST, Condor, CPM, DataCutter, DataCutter STORM, Globus Toolkit, GPT, Gridconfig, GridPort, GridSolve, GSI OpenSSH, Inca, KX.509/KCA, Look, MPICH-G2, MyProxy, Network Weather Service, OpenSAML, PERMIS, PyGlobus, Shibboleth, SRB Client, UberFTP, and WebISO (Web Initial Sign-on).

Grid Issues

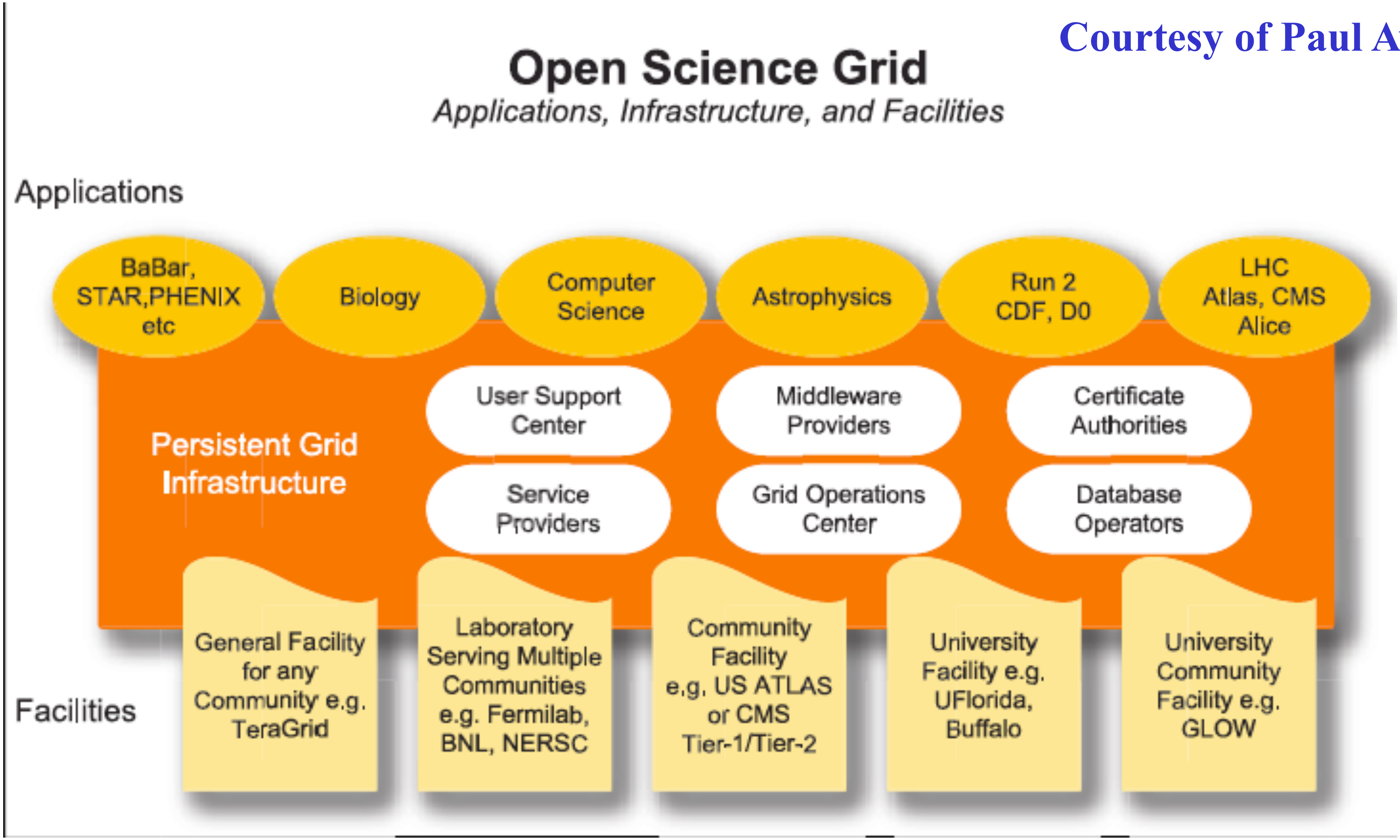
- High-Throughput Computing
- Transparent Integration of Data, Computing, Sensors/Devices, Networking
- Heterogeneous Resources
- Standards (Grid, Data)
- Major User Communities
  - High-Energy Physics and Astrophysics
  - Medicine and Biological Sciences
  - Earth Sciences
- Public Funding Still Critical
- Grids are in their Infancy

Major Grid Initiatives

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

Open Science Grid

Courtesy of Paul Avery



Cyberinfrastructure

- Foster & Kesselman: “a domain-independent computational infrastructure designed to support science.”
- NSF: “comprehensive phenomenon that involves creation, dissemination, preservation, and application of knowledge”
- Generic: transparent and ubiquitous application of technologies central to contemporary engineering and science
- NSF Cyberinfrastructure (OCI)
  - HPC Hardware and Software
  - Data Collections
  - Science Gateways/Virtual Organizations
  - Support of Next Generation Observing Systems

Miller’s Cyberinfrastructure Lab

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

Evolution of CI Lab Projects

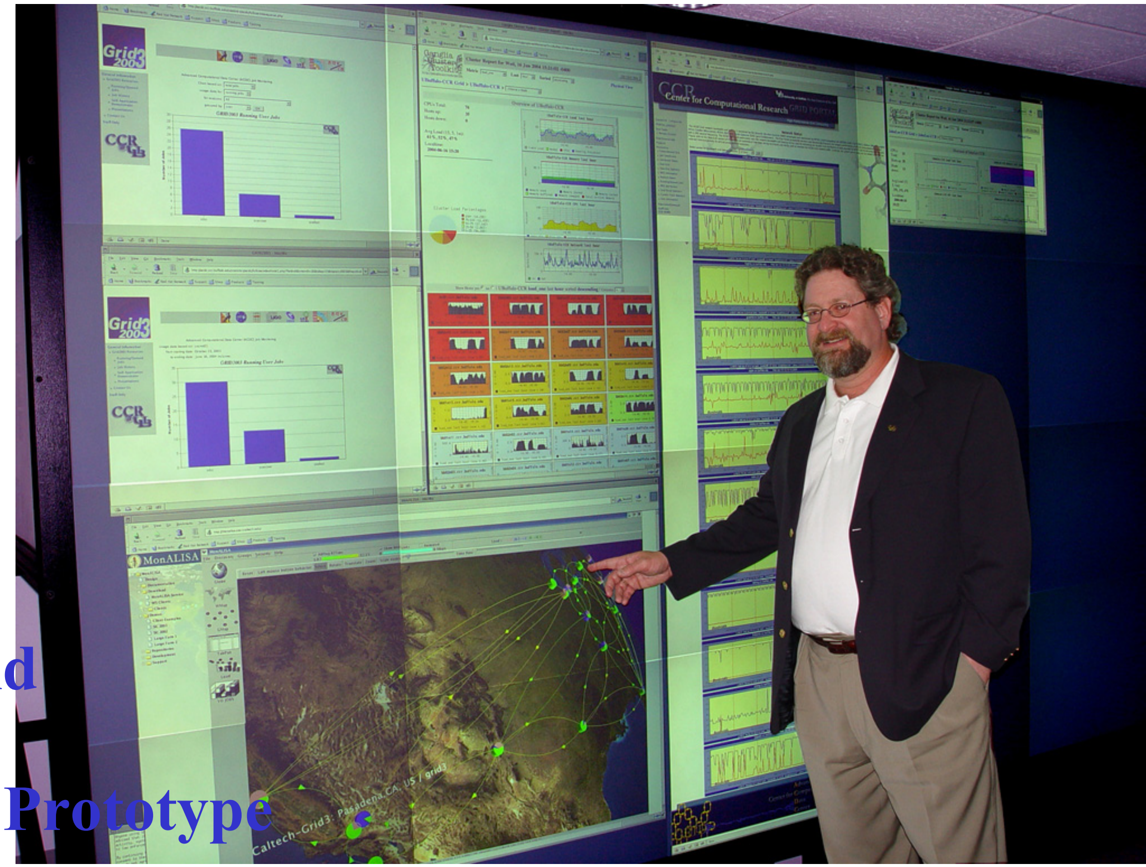
- ACDC-Grid
  - Experimental Grid: Globus & Condor
  - Integrate Data & Compute, Monitor, Portal, Node Swapping, Predictive Scheduling/Resource Management
  - GRASE VO: Structural Biology, Groundwater Modeling, Earthquake Eng, Comp Chemistry, GIS/BioHazards
  - Buffalo, Buffalo State, Canisius, Hauptman-Woodward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

NYS Grid Resources

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

CI Lab Collaborations

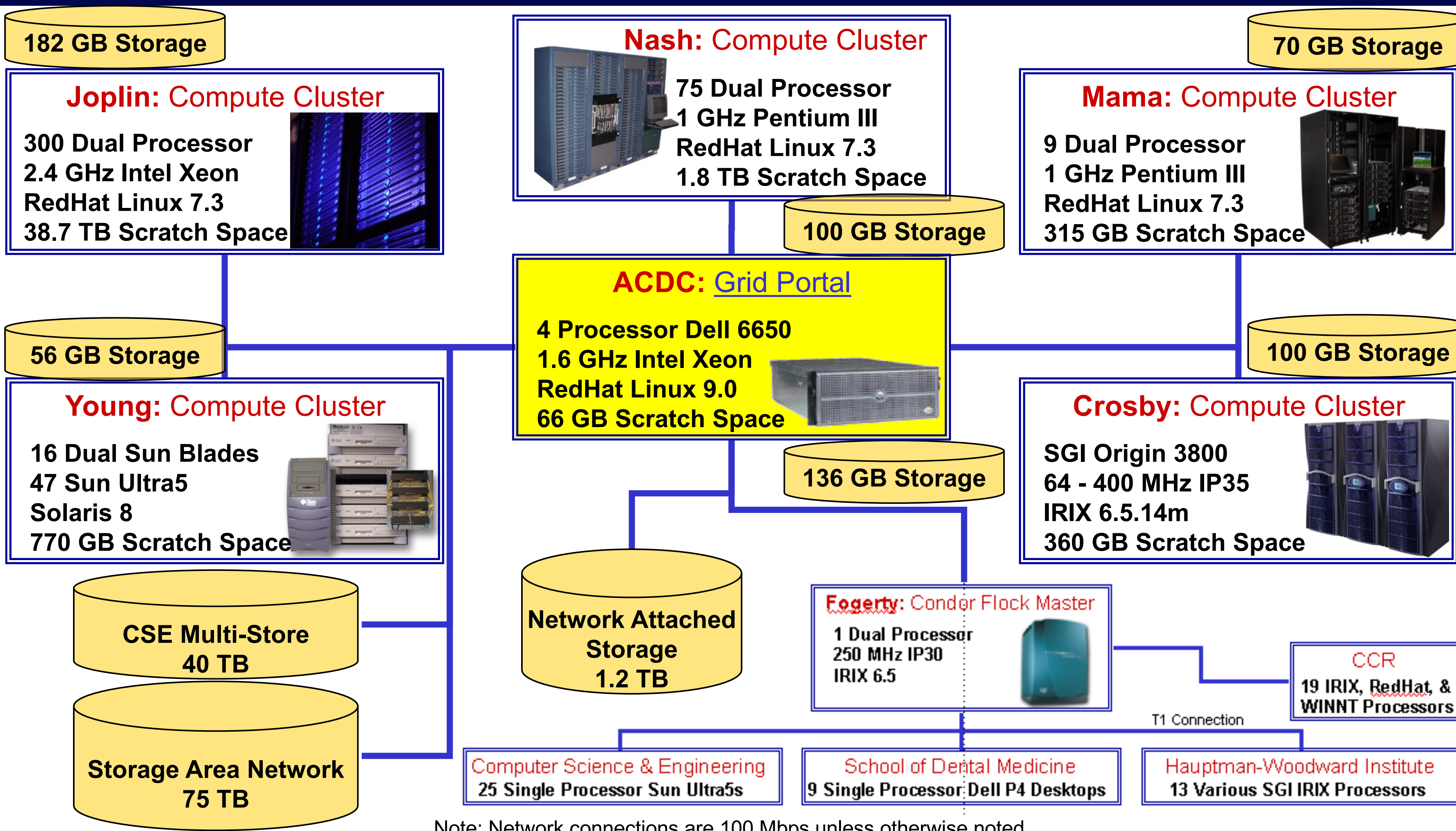
- High-Performance Networking Infrastructure
- Grid3+ Collaboration
- iVDGL Member
  - Only External Member
- Open Science Grid
  - GRASE VO
- NYSGrid.org
  - NYS CI Initiative
  - Executive Director
  - Various WGs
- Grid-Lite: Campus Grid
  - HP Labs Collaboration
- Innovative Laboratory Prototype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



ACDC-Grid Collaborations II

- Grass Roots NYS Grid (pre-NYSGrid.org)
  - SUNY-Albany
  - SUNY-Binghamton
  - SUNY-Buffalo
  - SUNY-Geneseo
  - Canisius College
  - Columbia
  - Hauptman-Woodward Inst.
  - Niagara University
- GRASE VO: Grid Resources for Advanced Science and Engineering Virtual Organization
  - (Non-Physics Research)
  - Structural Biology
  - Groundwater Modeling
  - Earthquake Engineering
  - Computational Chemistry
  - GIS/BioHazards

ACDC Data Grid Overview
(Grid-Available Data Repositories)

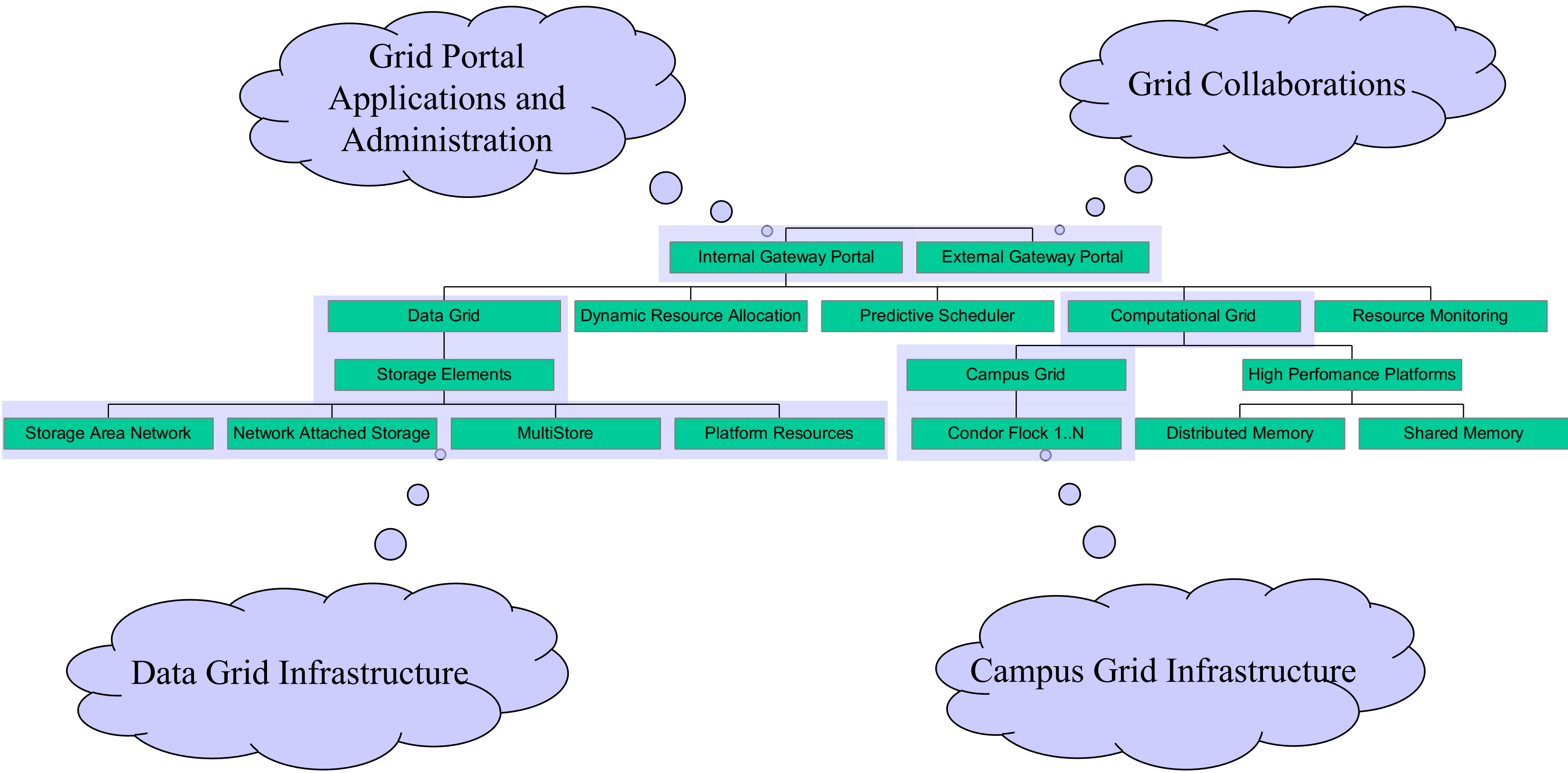


Note: Network connections are 100 Mbps unless otherwise noted.

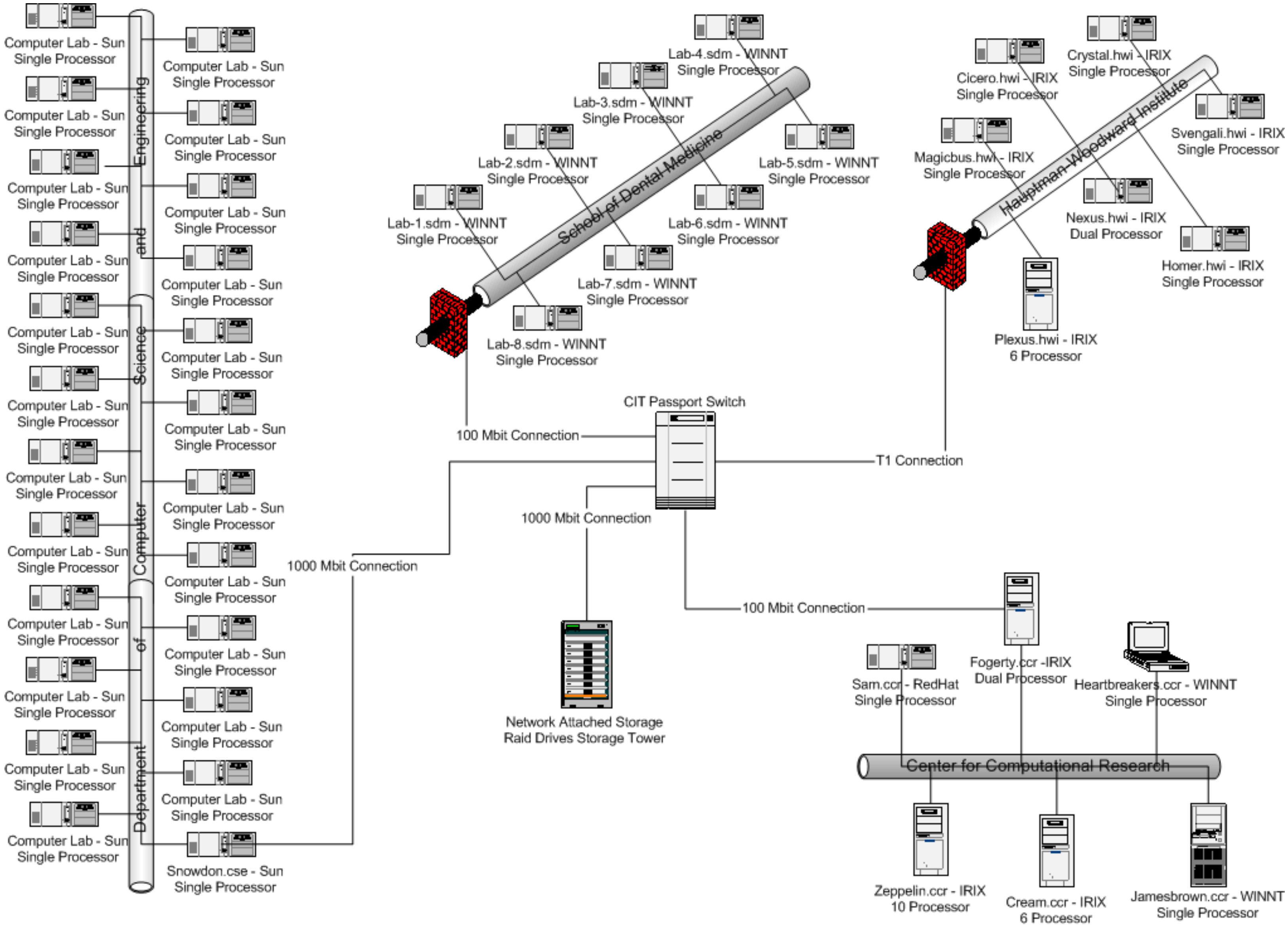
CI Lab Projects

- 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

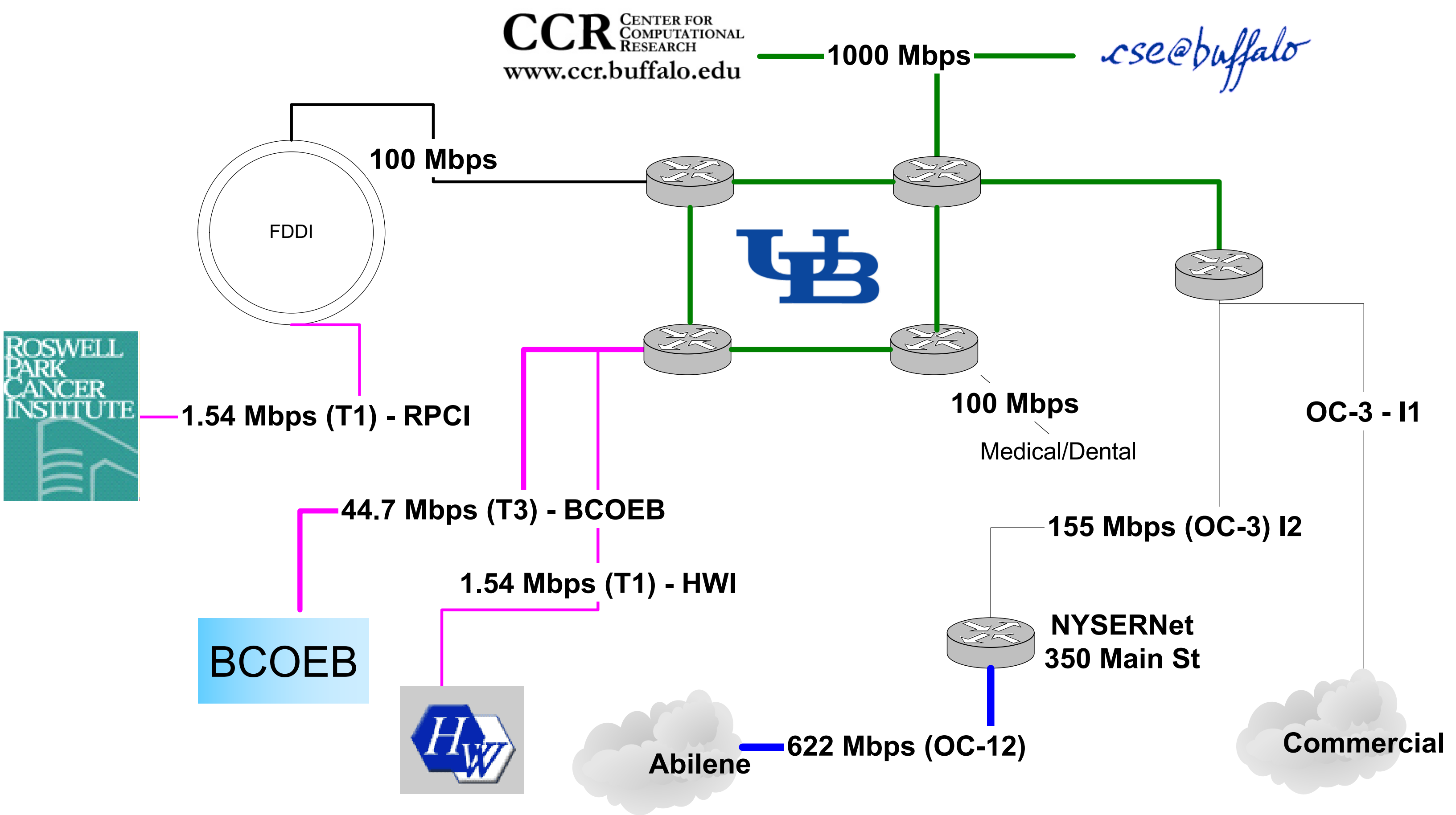
ACDC-Grid System Architecture



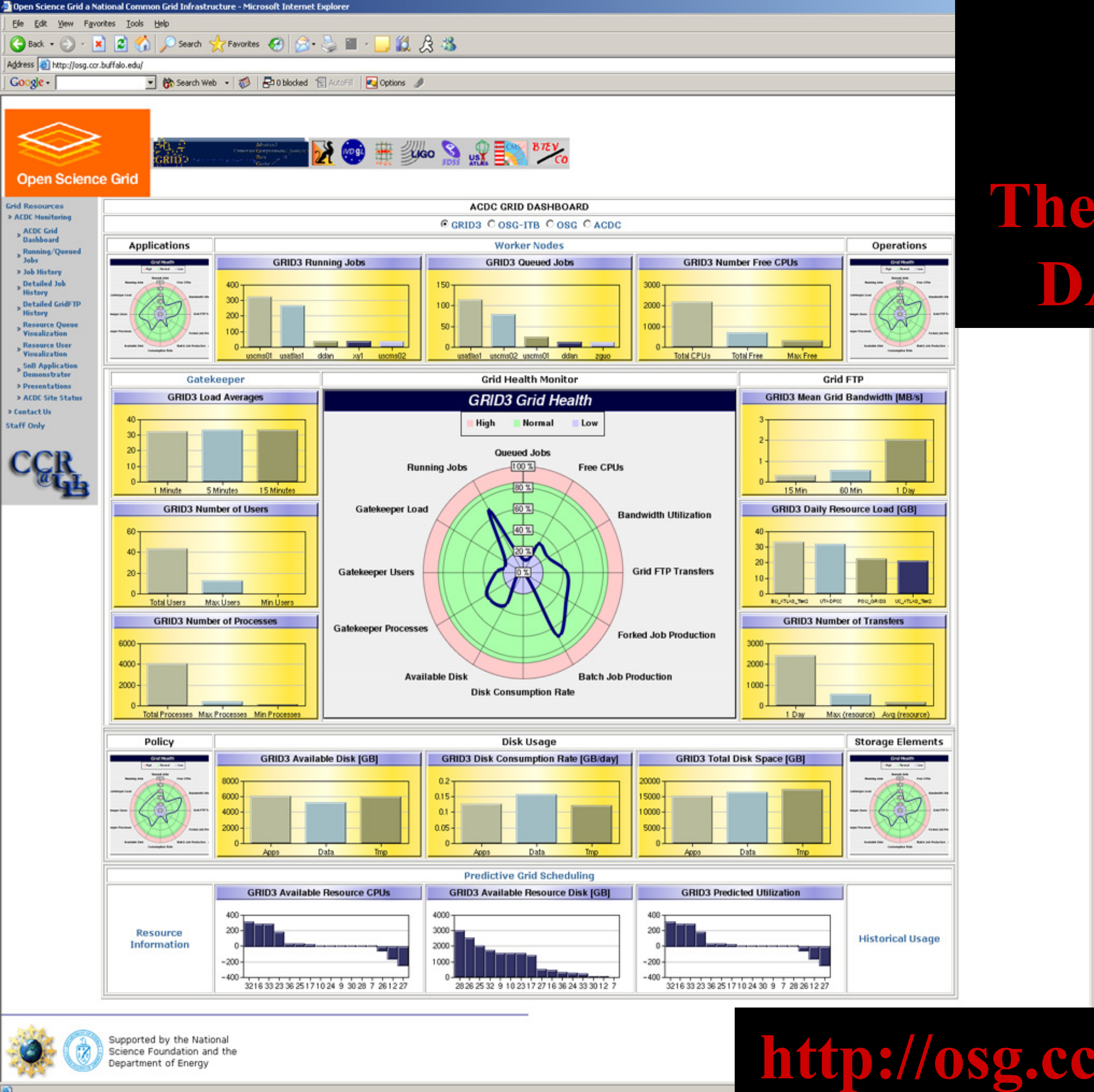
Initial ACDC Campus Grid



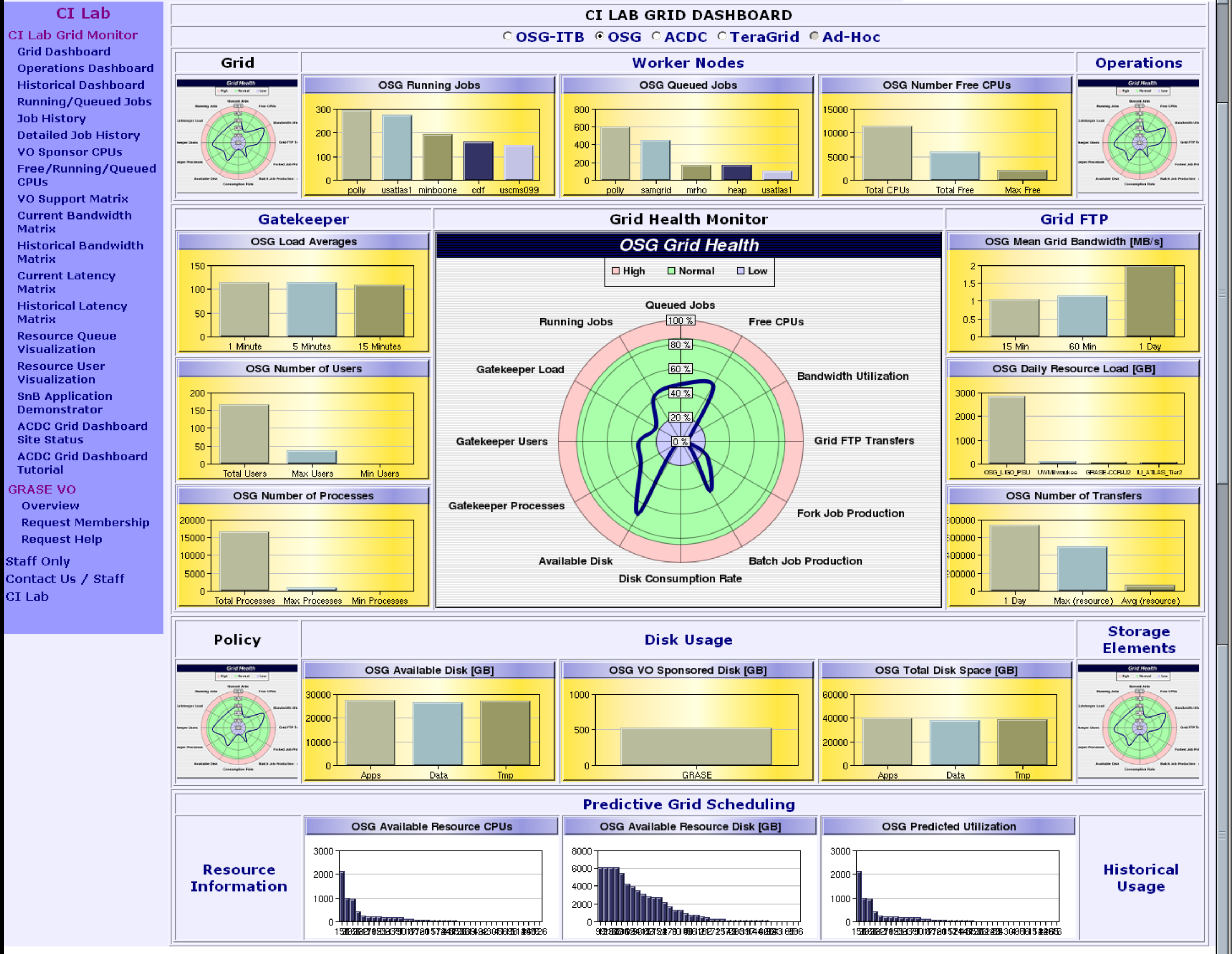
Network Connections



ACDC-Grid Monitoring: The ACDC-Grid DASHBOARD



http://osg.ccr.buffalo.edu

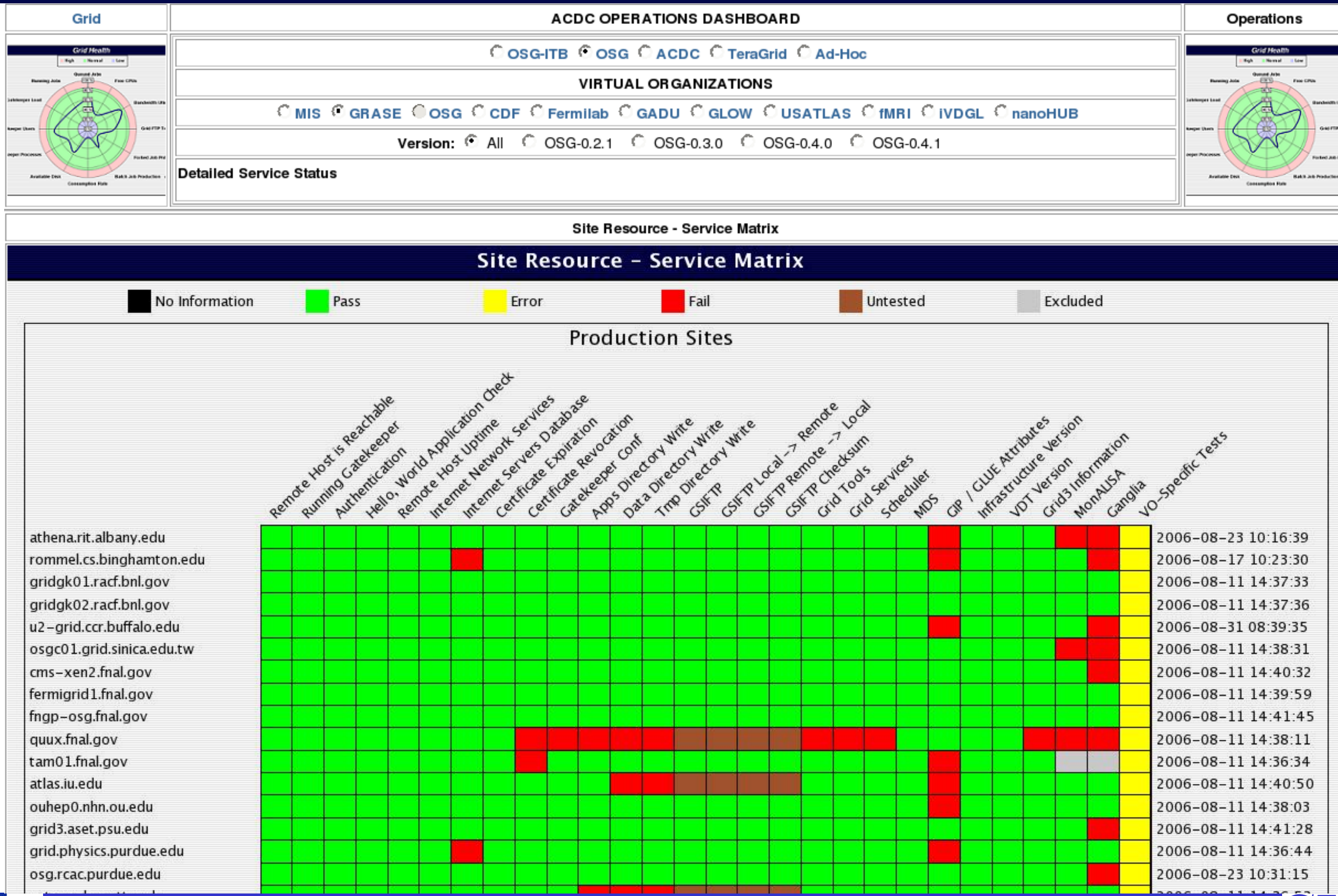


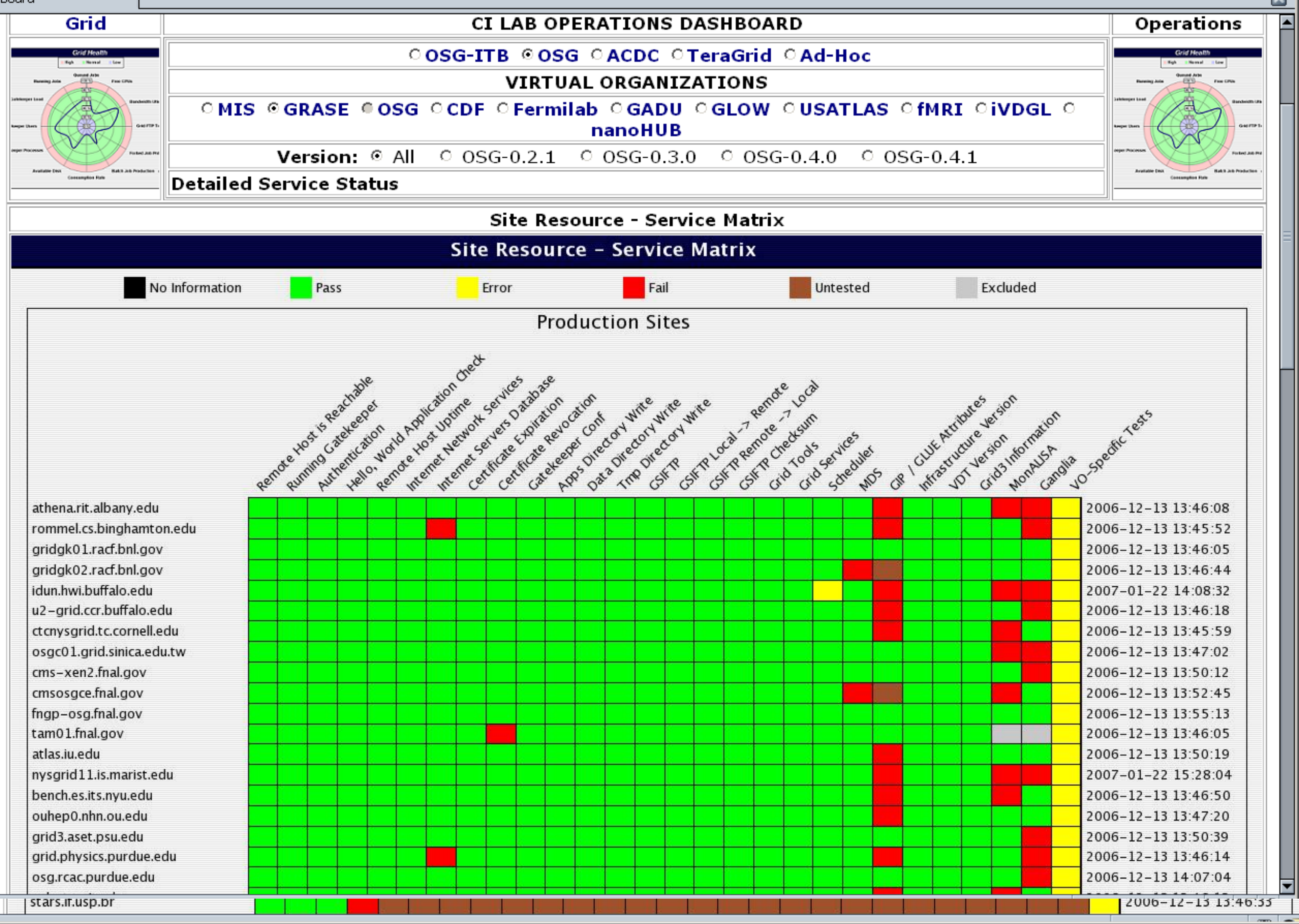
Supported by the National Science Foundation and the Department of Energy

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

ACDC Monitor

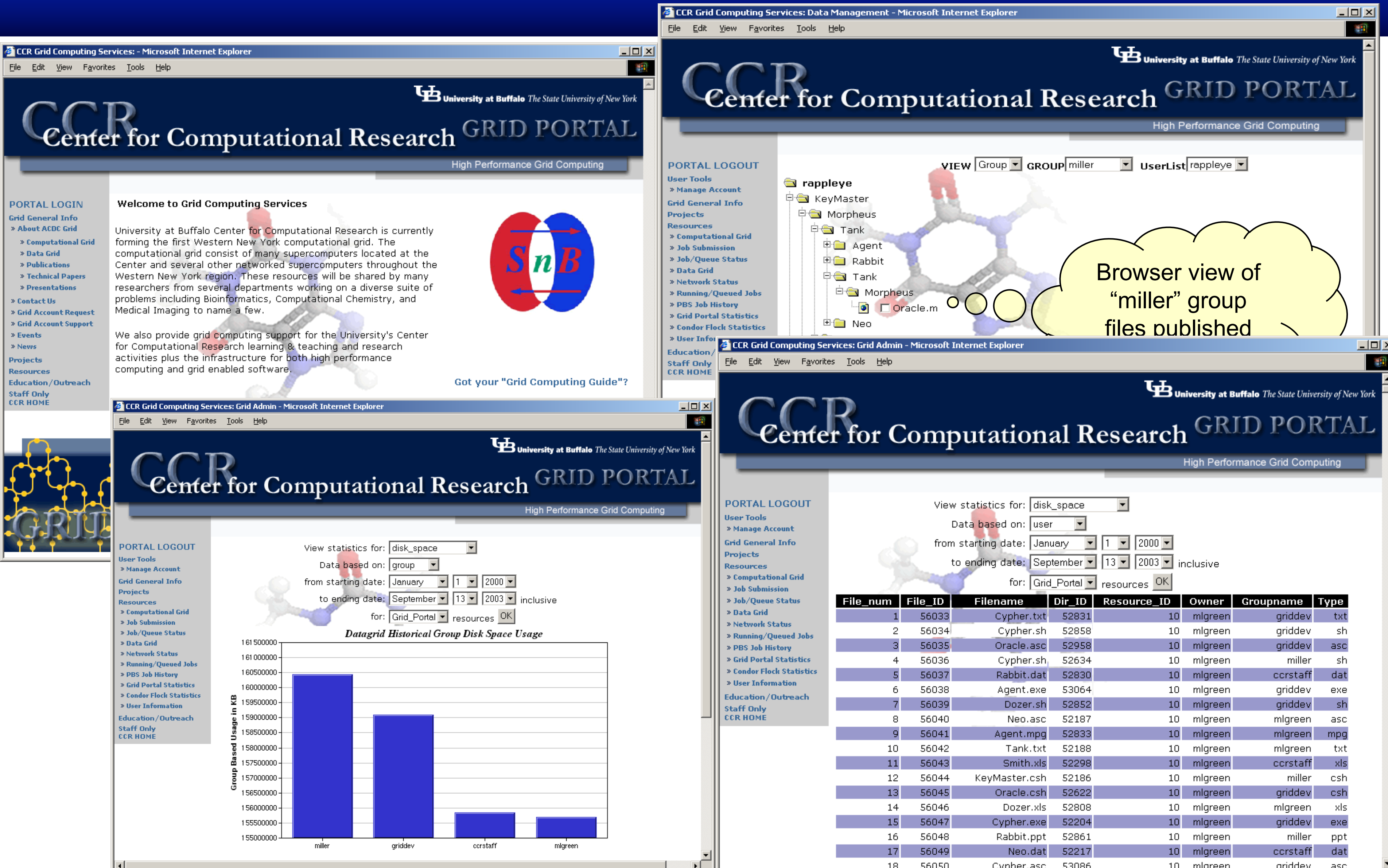
http://osg.ccr.buffalo.edu/operations-dashboard.php?grids=3&vos=10





CI Lab Operations Dashboard

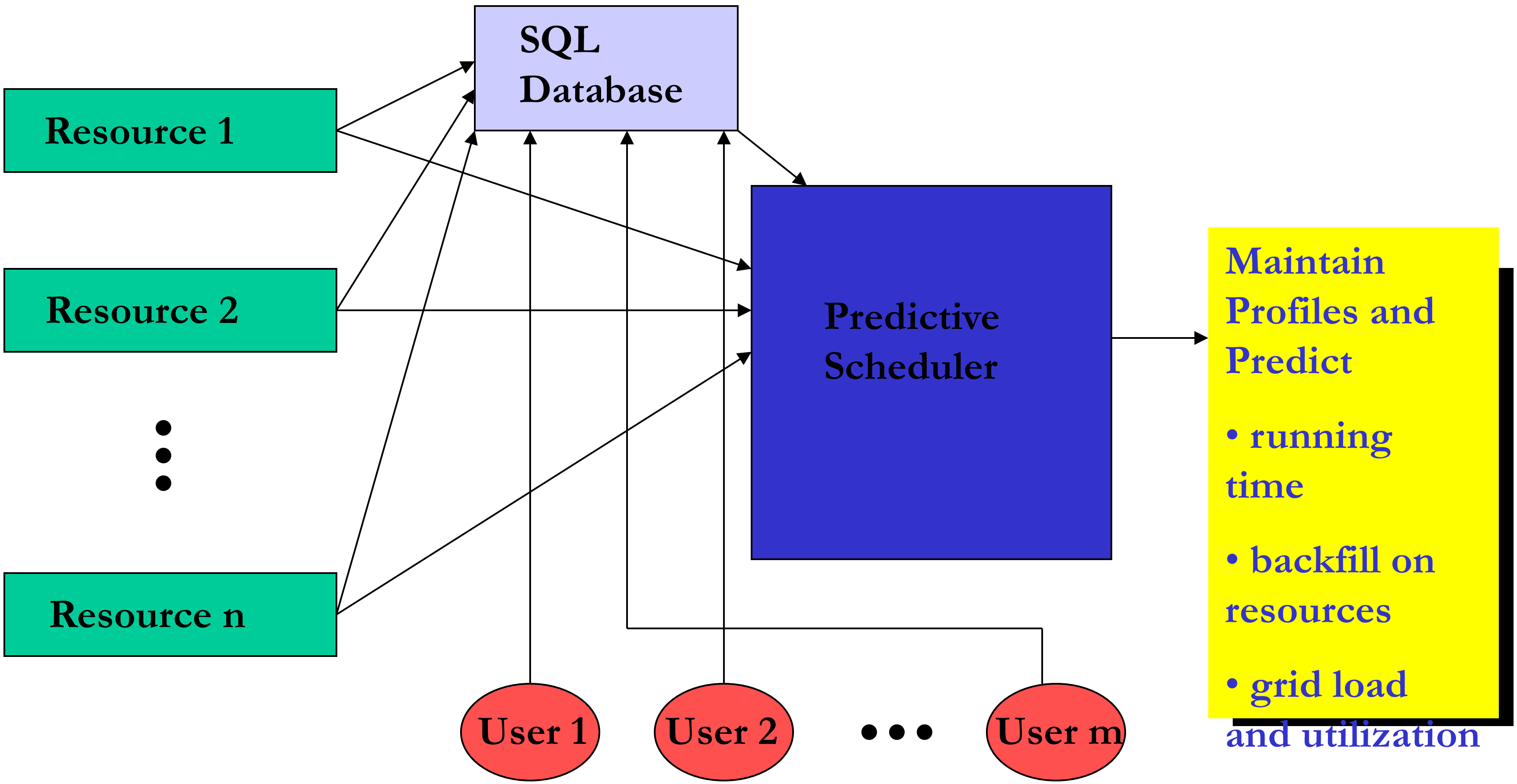
ACDC-Grid


CCR Grid Computing Services: - Microsoft Internet Explorer
Center for Computational Research
GRID PORTAL
High Performance Grid Computing
Welcome to Grid Computing Services
University at Buffalo Center for Computational Research is currently forming the first Western New York computational grid. The computational grid consist of many supercomputers located at the Center and several other networked supercomputers throughout the Western New York region. These resources will be shared by many researchers from several departments working on a diverse suite of problems including Bioinformatics, Computational Chemistry, and Medical Imaging to name a few.
We also provide grid computing support for the University's Center for Computational Research learning & teaching and research activities plus the infrastructure for both high performance computing and grid enabled software.
Got your "Grid Computing Guide"?
CCR Grid Computing Services: Data Management - Microsoft Internet Explorer
Browser view of
"miller" group
files published
CCR Grid Computing Services: Grid Admin - Microsoft Internet Explorer
Datagrid Historical Group Disk Space Usage
Group Based Usage in KB
miller
griddev
ccrstaff
mlgreen

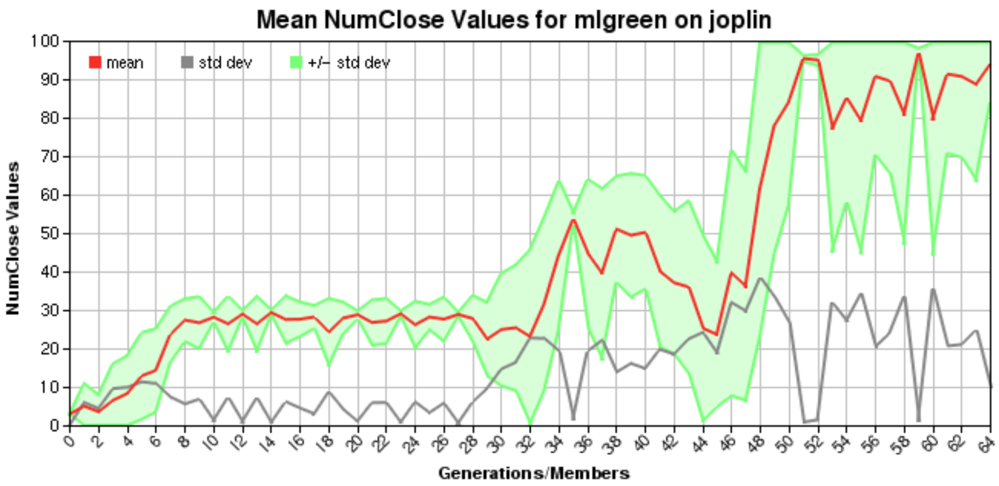
Predictive Scheduler

- 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

System Diagram

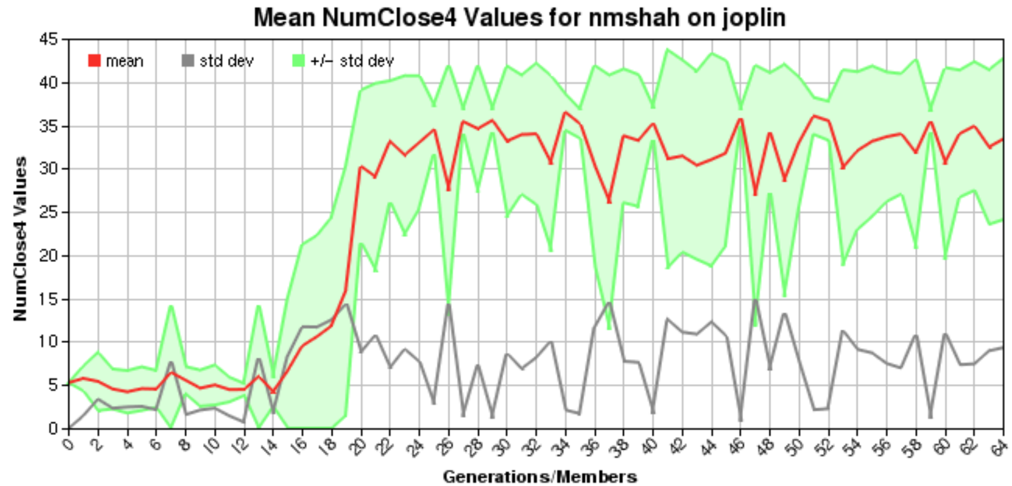


Preliminary GA results



Percent of estimates within 5% of actual values

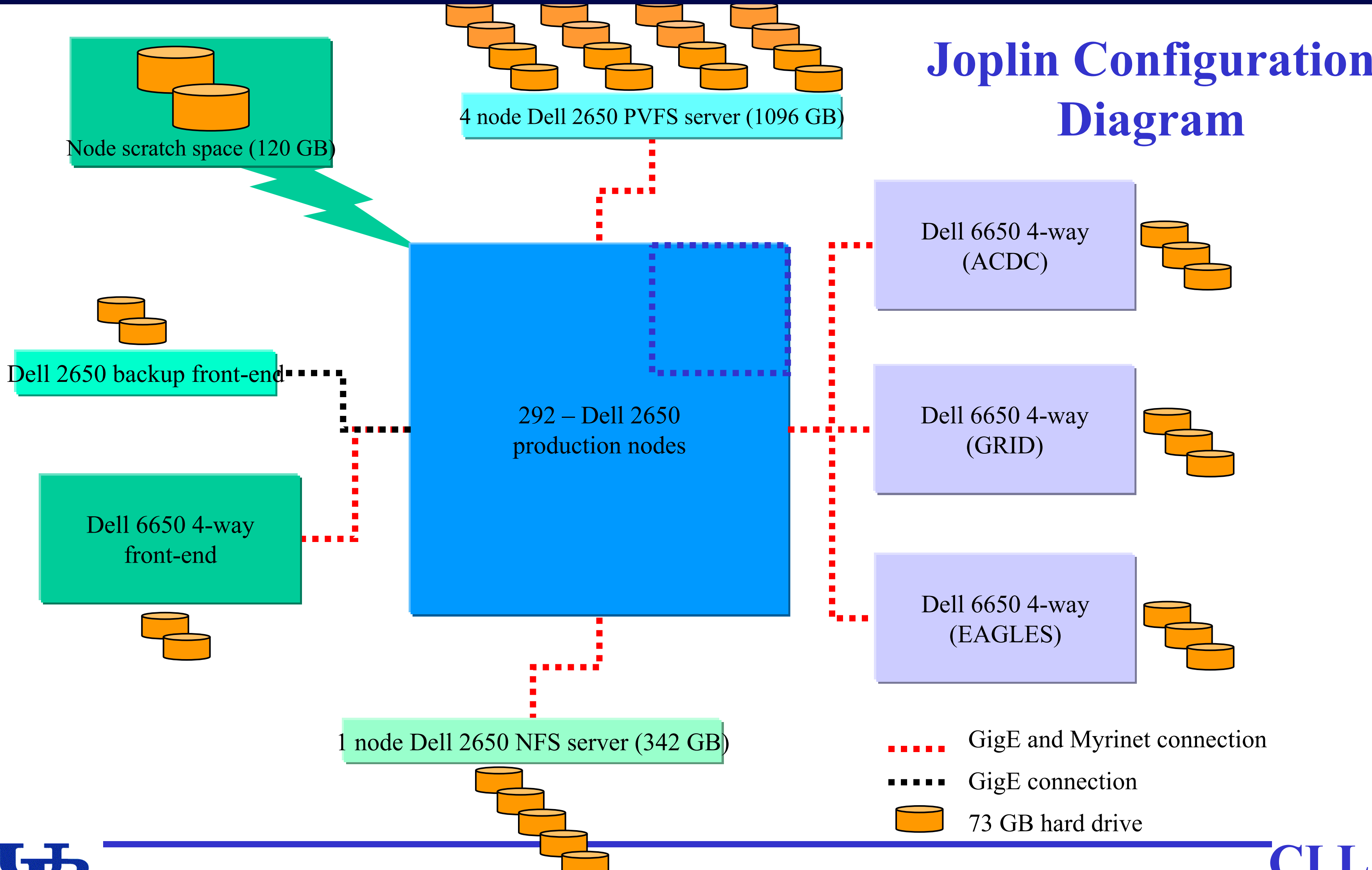
Percent of estimates within 20% of actual values



ACDC-Grid Dynamic Resource Allocation at SC03 with Grid3

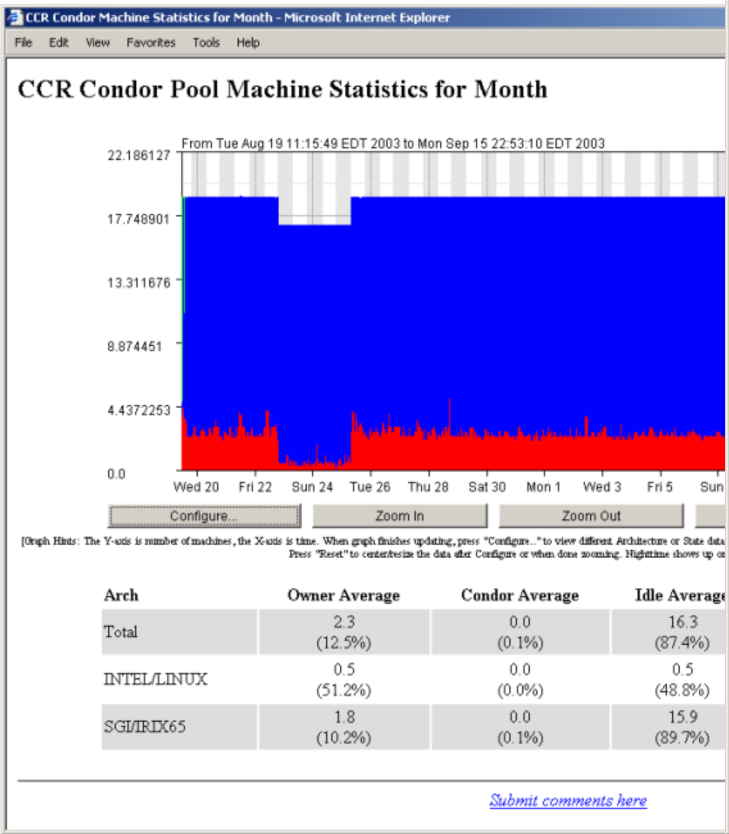
- Small number (40) of CPUs were dedicated at night
- An additional 400 CPUs were dynamically allocated during the day
- No human intervention was required
- Grid applications were able to utilize the resources and surpassed the Grid3 goals

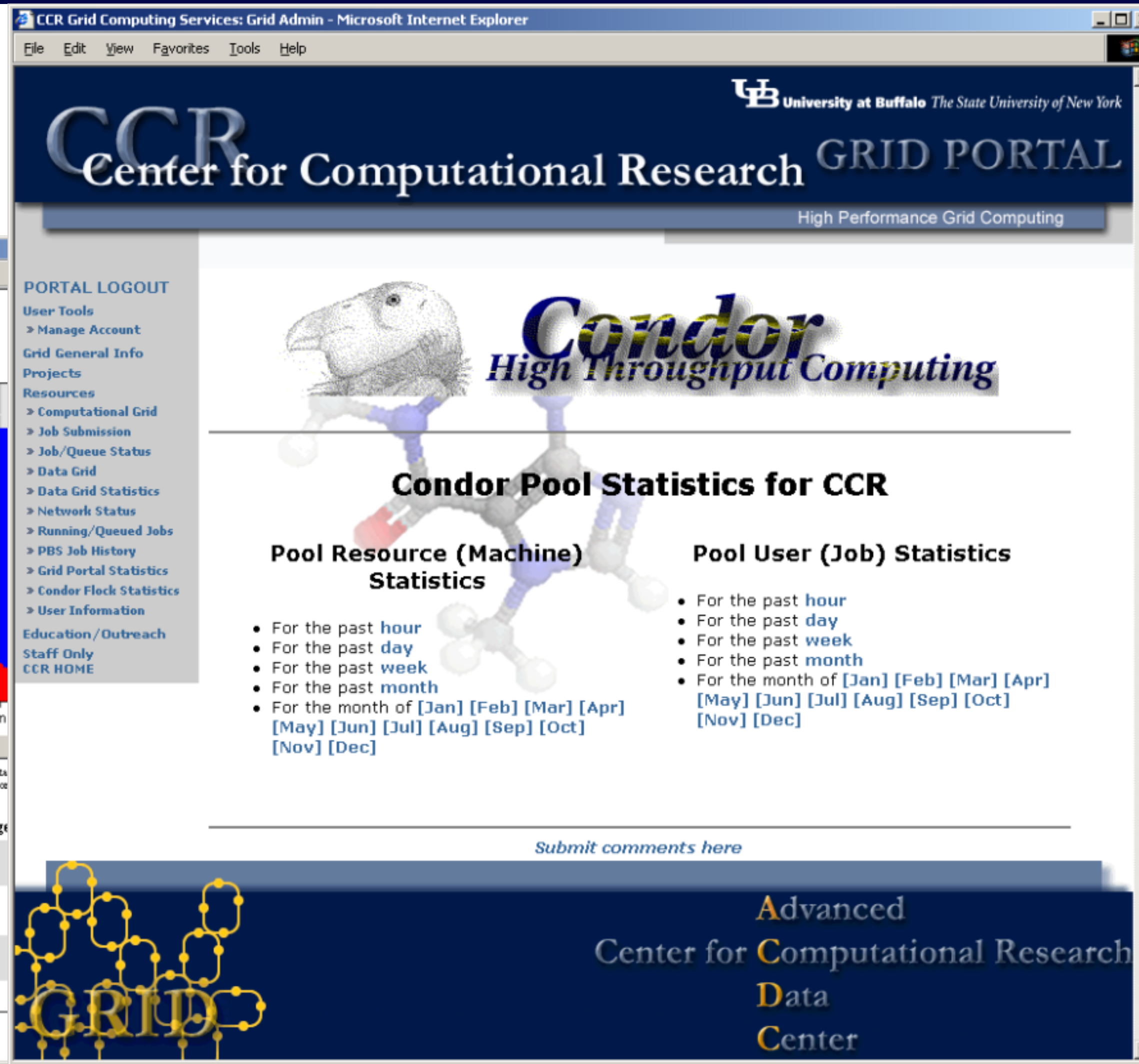
ACDC-Grid Dynamic Resource Allocation



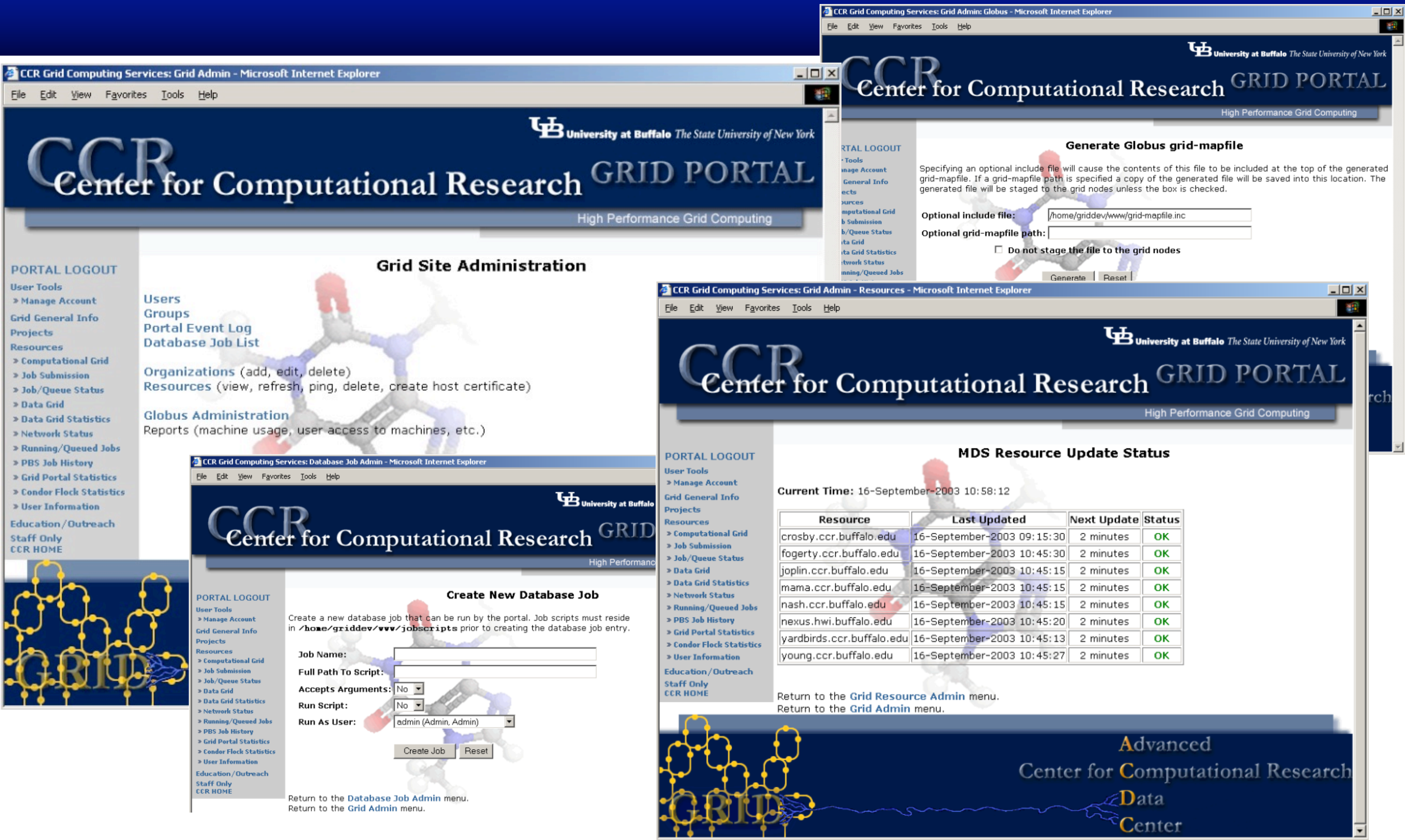
ACDC-Grid Portal Condor Flock

- CondorView integrated into ACDC-Grid Portal





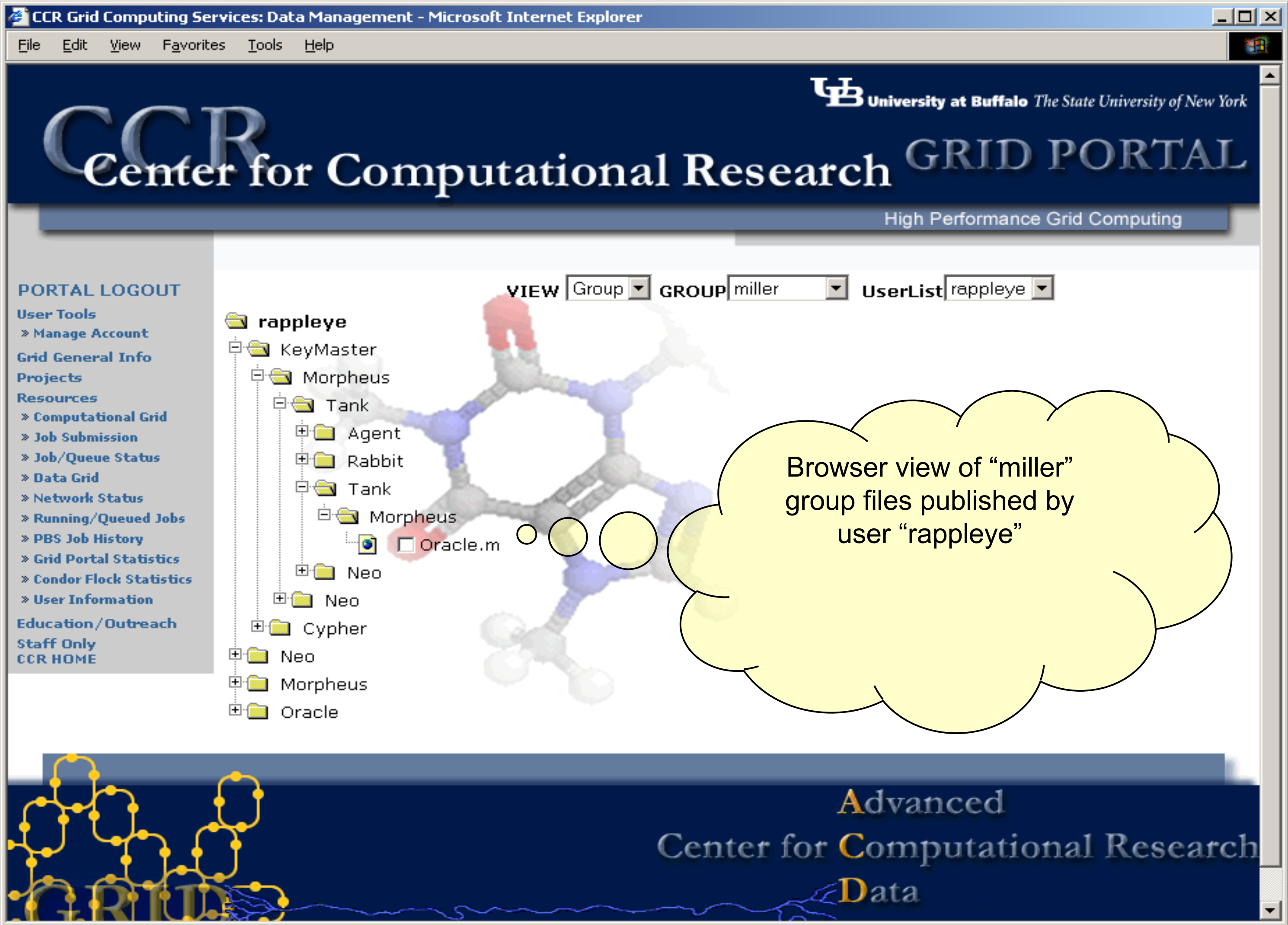
Grid Administration



Data Grid

- **Motivation:**
  - **Large data collections are emerging as important community resources.**
  - **Data Grids complement Computational Grids.**
- **Definition:** ***A data grid is a network of distributed storage resources, including archival systems, caches, and databases, which are linked logically to create a sense of global persistence.***
- **Goal: Design and implement transparent management of data distributed across heterogeneous resources.**

ACDC-Grid Data Grid



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.)

ACDC-Grid
Data Grid File Migration

- Migration Algorithm dependent on
  - User access time
  - Network capacity at time of migration
  - User profile
  - User disk quotas on various resources

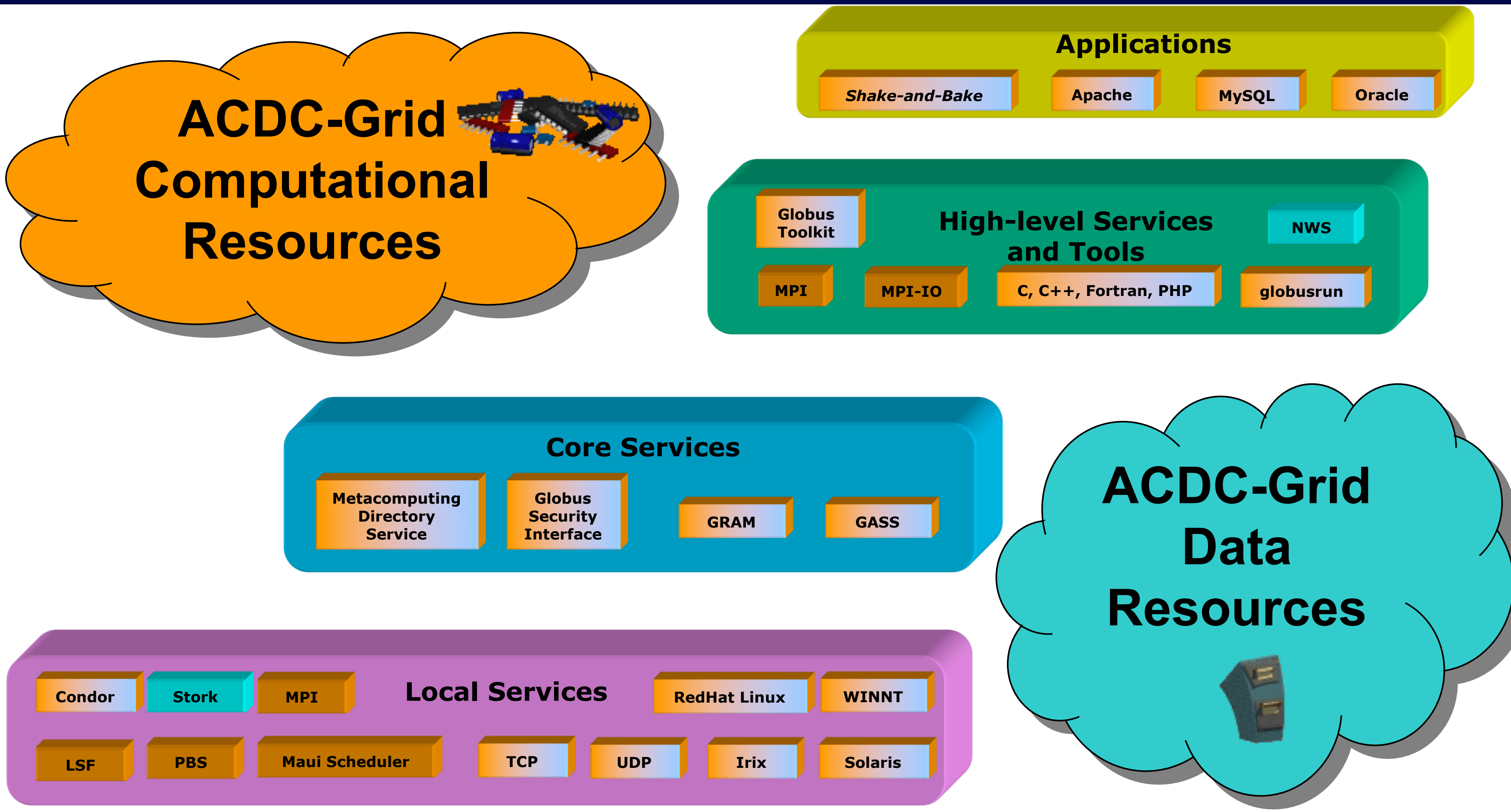
Data Grid File Aging

- For a given user, the average of the file_aging_local_param attributes of all files should be close to 1.
  - Operating tolerance before action is taken is within the range of 0.9 – 1.1.
- In this way, the user file_aging_global_param can be a function of this average.
  - If the average file_aging_local_param attribute > 1.1, then files of the user are being held to long before being migrated.
    - The file_aging_global_param value should be decreased.
  - If the average file_aging_local_param attribute < 0.9, then files of the user are being accessed at a higher frequency than the file_aging_global_param value.
    - The file_aging_global_param value should be increased.

Data Grid Resource Info

Both platforms have reduced bandwidth available for additional transfers

Grid Services and Applications



Adapted from Ian Foster and Carl Kesselman

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

Middleware

- Grid (Computational and Data)
  - Globus Toolkit 2.2.4 → direct upgrade WSRF
  - Condor 6.6.0
  - Network Weather Service 2.6
  - Apache2 HTTP Server
  - PHP 4.3.0
  - MySQL 3.23
  - phpMyAdmin 2.5.1
- Collaboratory
  - OpenGL (LibDMS, DevIL, GLUT)
  - Windows, IRIX, Mac OS X, Linux
  - CAVE, Desktop

Grid Enabled *SnB*

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

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

Mail Home Netscape

New Tab | CCR Grid Computing Services:

Cyberinfrastructure Laboratory

Grid Portal

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

CI Lab

Grid Portal Info
- Overview
- Portal Login
- Grid Account Info

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

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

Contact Us / Staff
CI Lab
Staff Only

Welcome to the Cyberinfrastructure Laboratory Grid Portal

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

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

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

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

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

Software : BnP
Field : Protein crystal structure determination

Startup Screen for CI Lab Grid Job Submission

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 applciations. 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.

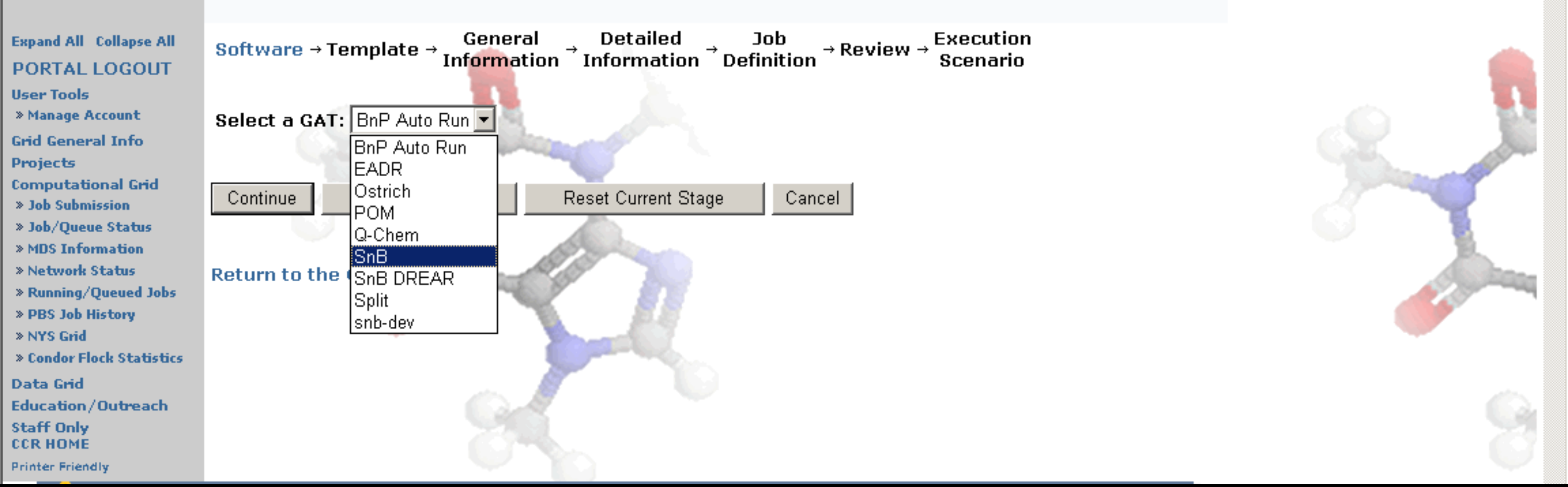
The grid-enabled versions of many well-defined scientific and engineering applications have very similar general requirements and core functionality that are require for execution in the ACDC-Grid environment. We have identified that sequentially defining milestones for the grid user to complete intuitively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

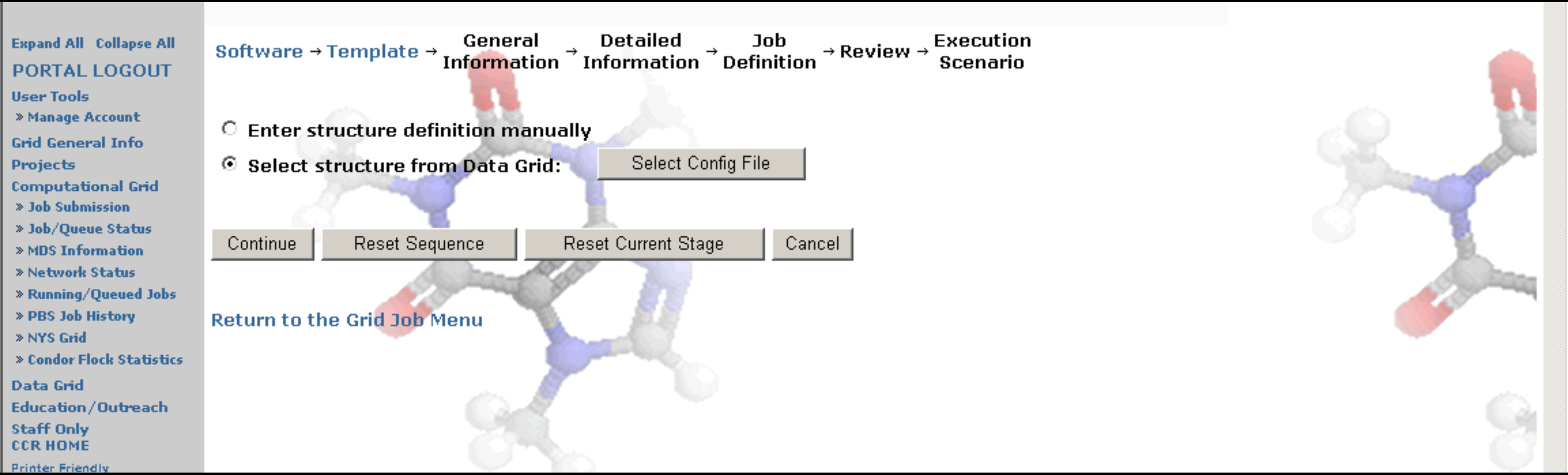
Each item of the job definition workflow is then stored in the ACDC-Grid Web Portal database so the grid user may use/modify any previously created workflow in creating new job definitions. The job definitions can also be accessed via batch script files for executing hundreds of similar workflows in 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

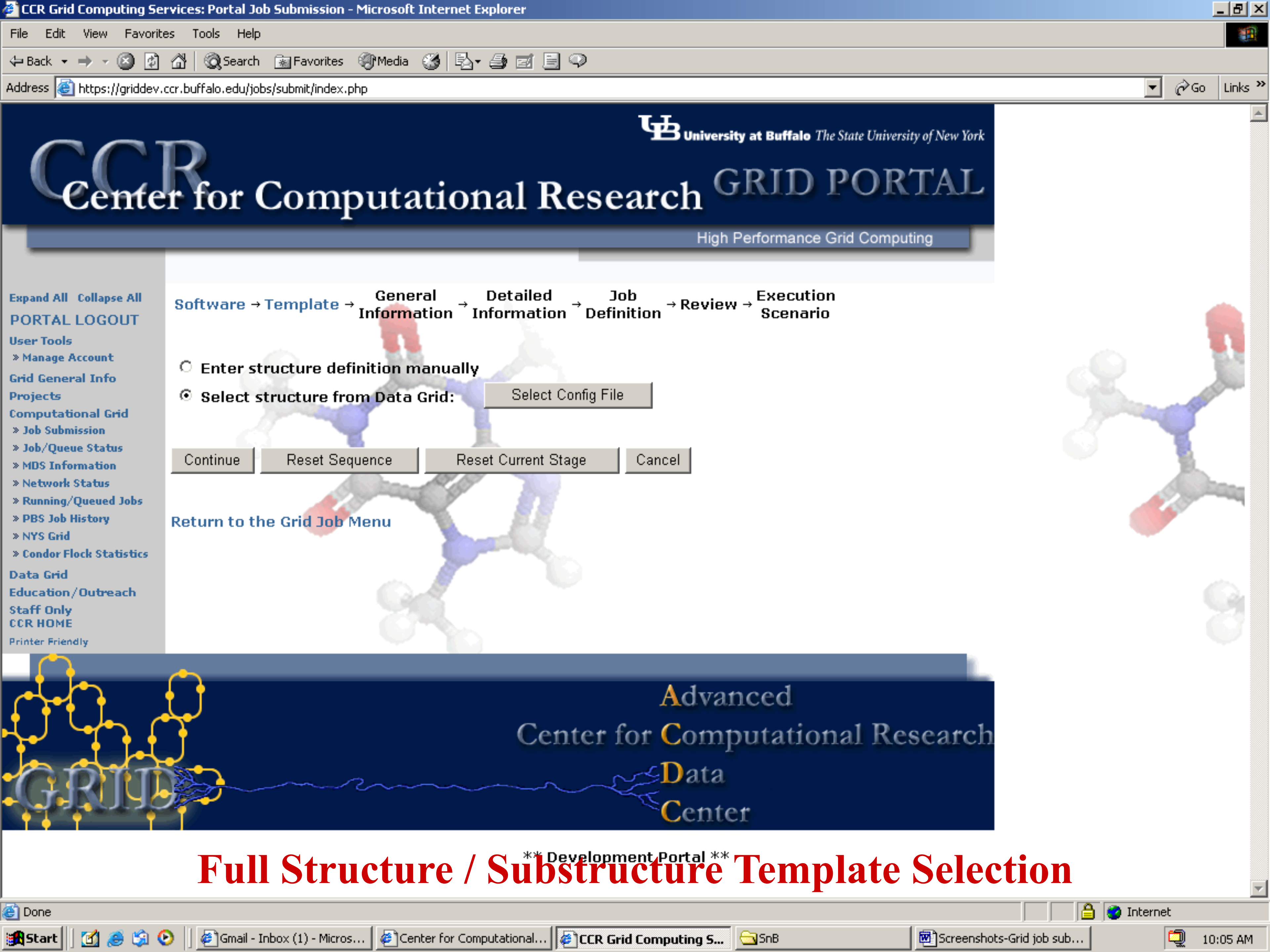
Instructions and Description for Running a Job on ACDC-Grid



Software Package Selection



Full Structure / Substructure Template Selection



Full Structure / Substructure Template Selection

General Information

Structure Information

Title: Iled

Structure ID : iled

Space Group : 19 [Select]

Cell Constants and Cell Errors (Cell Errors optional)

A:	11.516	+/-	
B:	15.705	+/-	
C:	39.310	+/-	0.004
Alpha:	90.0	+/-	
Beta:	90.0	+/-	
Gamma:	90.0	+/-	

Native Asymmetric Unit Contents

No Residues (Optional):

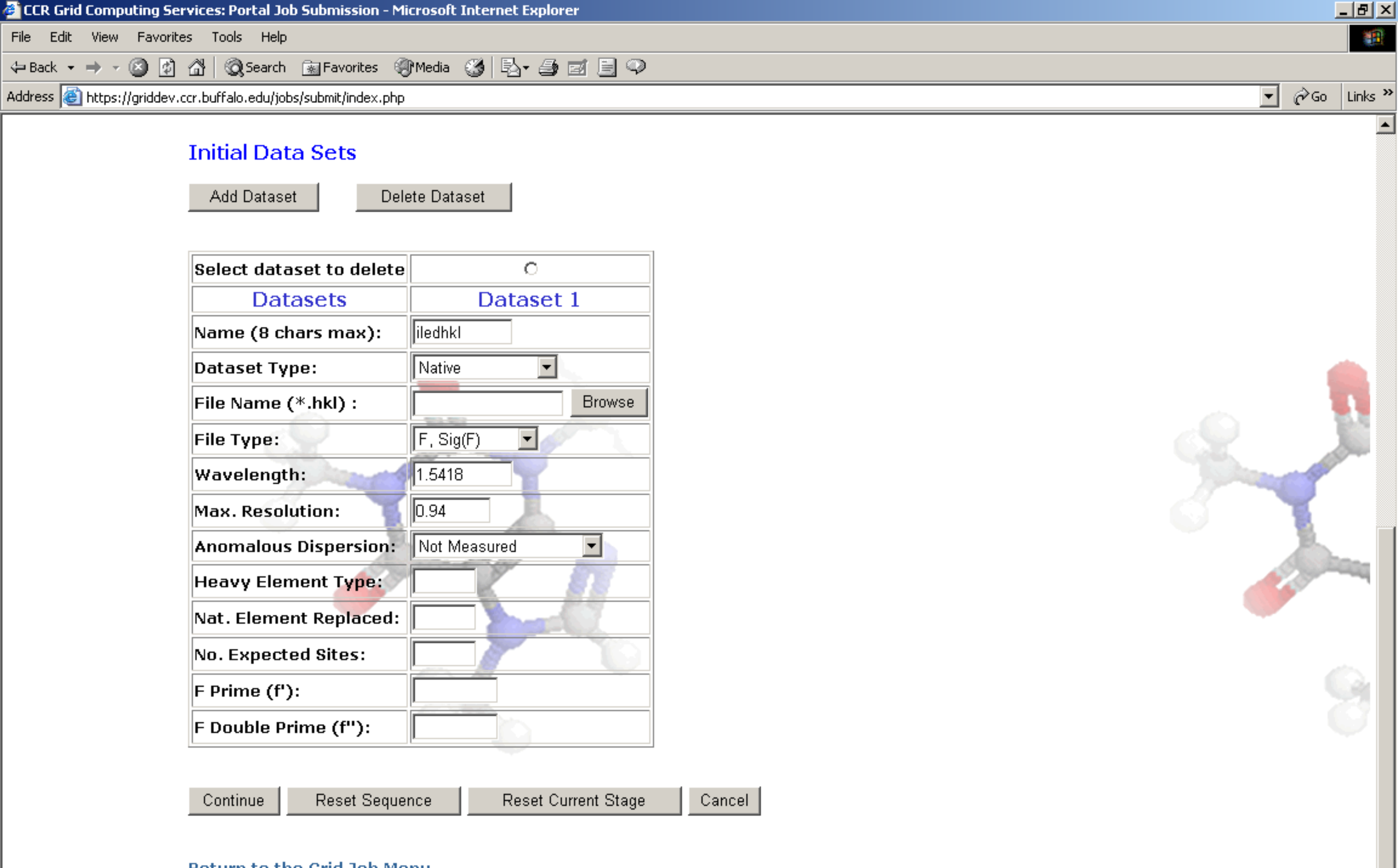
ASU Contents : C60H102N6O18 (examples: C6H12O6 OR C6 H12 O6)

Initial Data Sets

[Add Dataset] [Delete Dataset]

Select dataset to delete	○
Datasets	Dataset 1
Name (8 chars max):	iledhkl

Default Parameters Based on Template


CCR Grid Computing Services: Portal Job Submission - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address https://griddev.ccr.buffalo.edu/jobs/submit/index.php

Initial Data Sets

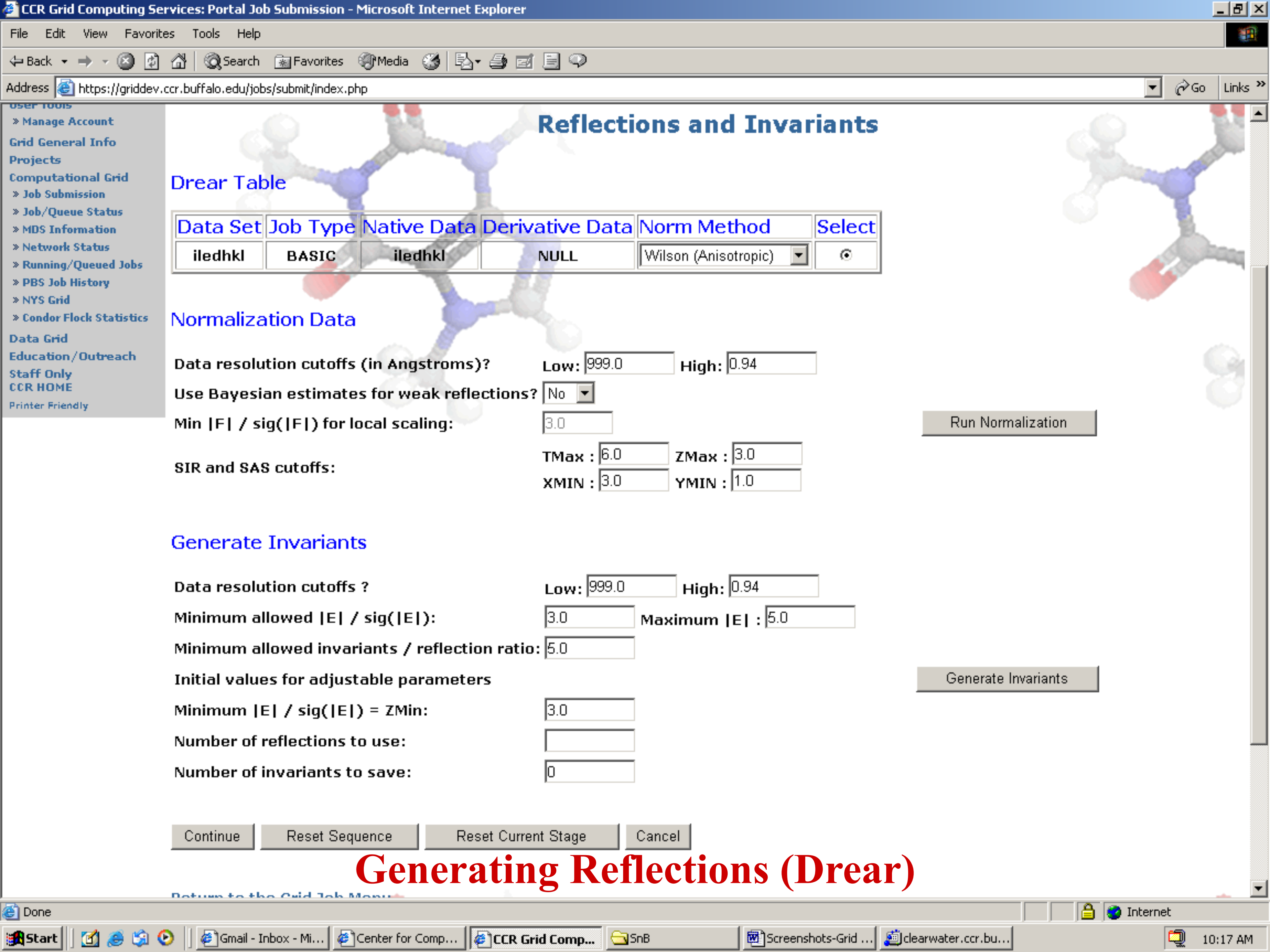
Add Dataset | Delete Dataset

Select dataset to delete	◯
Datasets	**Dataset 1**
Name (8 chars max):	iledhkl
Dataset Type:	Native
File Name (*.hkl) :	Browse
File Type:	F, Sig(F)
Wavelength:	1.5418
Max. Resolution:	0.94
Anomalous Dispersion:	Not Measured
Heavy Element Type:	
Nat. Element Replaced:	
No. Expected Sites:	
F Prime (f'):	
F Double Prime (f"):	

Continue | Reset Sequence | Reset Current Stage | Cancel

Return to the Grid Job Menu

Default Parameters (cont’d)



Generating Reflections (Drear)

CCR Grid Computing Services: Portal Job Submission - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address https://griddev.ccr.buffalo.edu/jobs/submit/index.php

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

Reflections and Invariants

Drear Table

Data Set	Job Type	Native Data	Derivative Data	Norm Method	Select
iledhkl	BASIC	iledhkl	NULL	Wilson (Anisotropic)	⦿

Normalization Data

Data resolution cutoffs (in Angstroms)? Low: 999.0 High: 0.94

Use Bayesian estimates for weak reflections? No

Min |F| / sig(|F|) for local scaling: 3.0

Run Normalization

SIR and SAS cutoffs: TMax : 6.0 ZMax : 3.0 XMIN : 3.0 YMIN : 1.0

Generate Invariants

Data resolution cutoffs ? Low: 999.0 High: 0.94

Minimum allowed |E| / sig(|E|): 3.0 Maximum |E| : 5.0

Minimum allowed invariants / reflection ratio: 5.0

Initial values for adjustable parameters

Generate Invariants

Minimum |E| / sig(|E|) = ZMin: 3.0

Number of reflections to use: 840

Number of invariants to save: 8400

Continue | Reset Sequence | Reset Current Stage | Cancel

Invariant Generation

Done | Internet

Start | Gmail - Inbox - Mi... | Center for Comp... | CCR Grid Comp... | SnB | Screenshots-Grid ... | clearwater.ccr.bu... | 10:17 AM

CCR Grid Computing Services: Portal Job Submission - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Search Favorites Media

Address https://griddev.ccr.buffalo.edu/jobs/submit/index.php Go Links

SnB Setup

Grid Parameters

Preferred resource name:	Grid Scheduler
Number of processors:	5
Wallclock time requested: (mins)	720
Job Prefix for results:	job0
Queue:	grid

SnB Run Parameters

- *Invariants*

Number of triplet invariants to use:	8400

- *Trials To Process*

Starting phases from:	Random Atoms
Random seed (prime):	11909
Number of Trials:	1000
Starting Trial:	1
Input Phase File:	none
Input Atom File:	none
Keep complete (every trial) peak file? :	Yes

- *Cycles Information*

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:	0.20

- *Phase Refinement Method*

SnB Setup

Done Internet

Start Gmail - Inbox - Mi... Center for Comp... CCR Grid Comp... SnB Screenshots-Grid ... clearwater.ccr.bu... 10:17 AM

CCR Grid Computing Services: Portal Job Submission - Microsoft Internet Explorer

https://griddev.ccr.buffalo.edu/jobs/submit/index.php

- *Phase Refinement Method*

Phase Refinement Method :	Parameter Shift (Fast)
Number of passes through phase set:	3
Phase shift:	90.0
Number of shifts:	2

- *Real-Space Constraints*

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 :	4
Number of peaks :	84

- *Twice Baking*

Trials for E-Fourier filtering (fourier refinement)? :	None
Number of cycles :	8
Number of peaks :	84
Minimum |E| :	0.75

- *Automatic solution identification criteria*

Rmin Improvement (%):	45.0
Rcryst Imporvement (%):	25.0

Continue | Reset Sequence | Reset Current Stage | Cancel

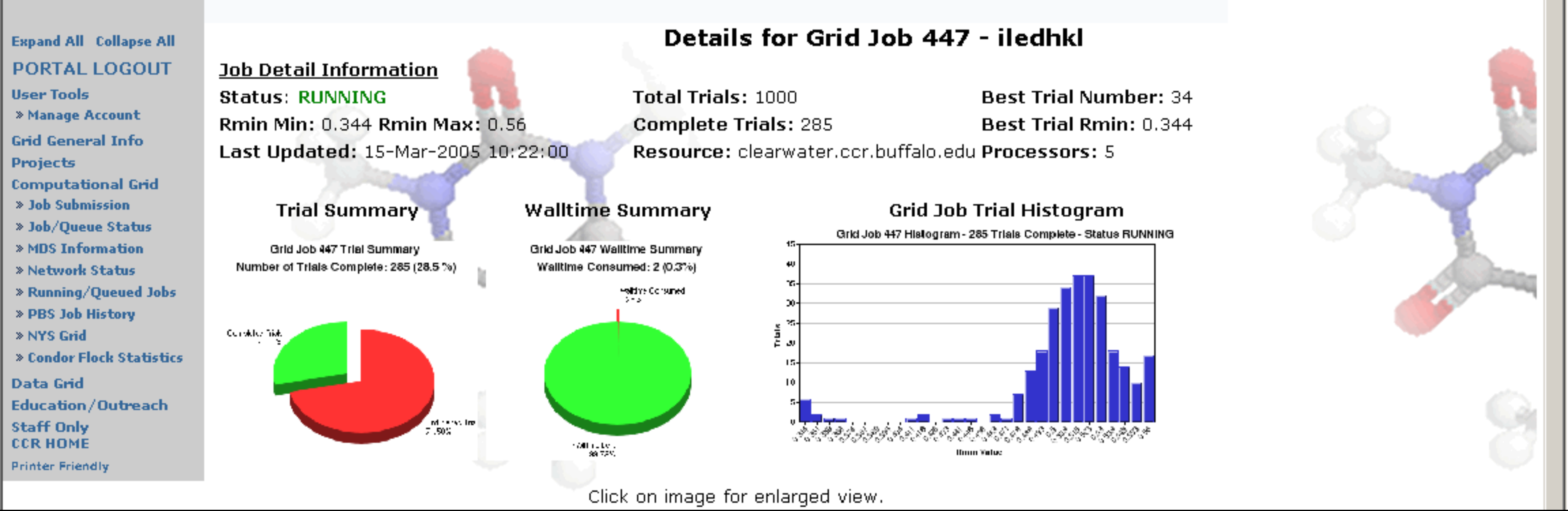
SnB Setup (cont'd)

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

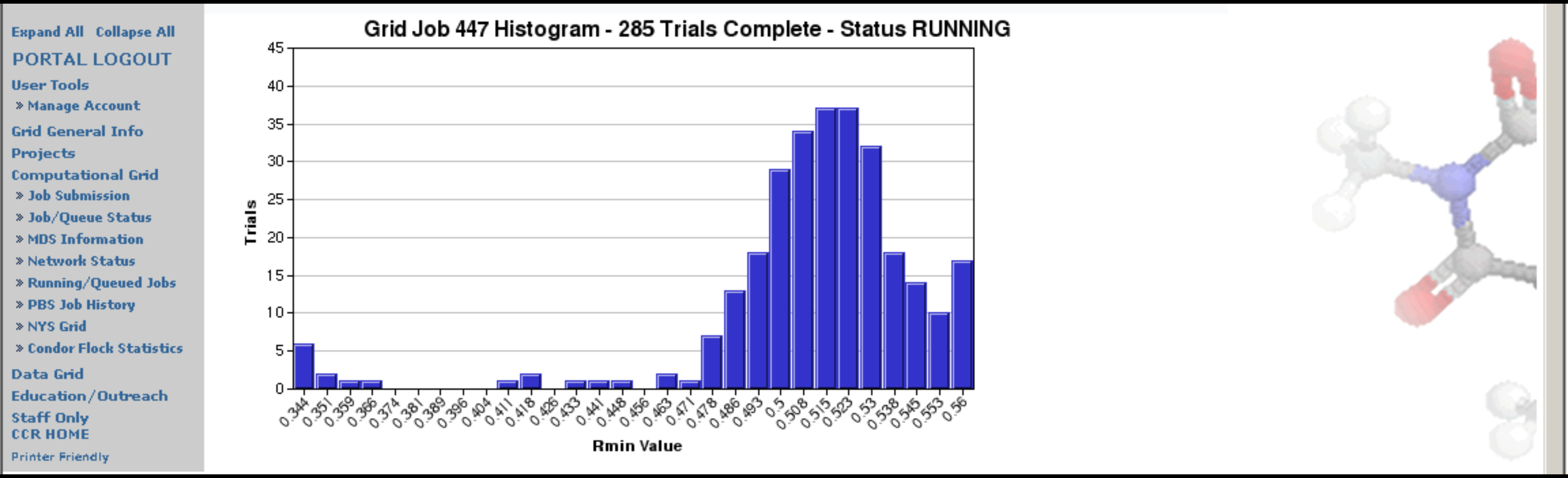
SnB Job Review

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

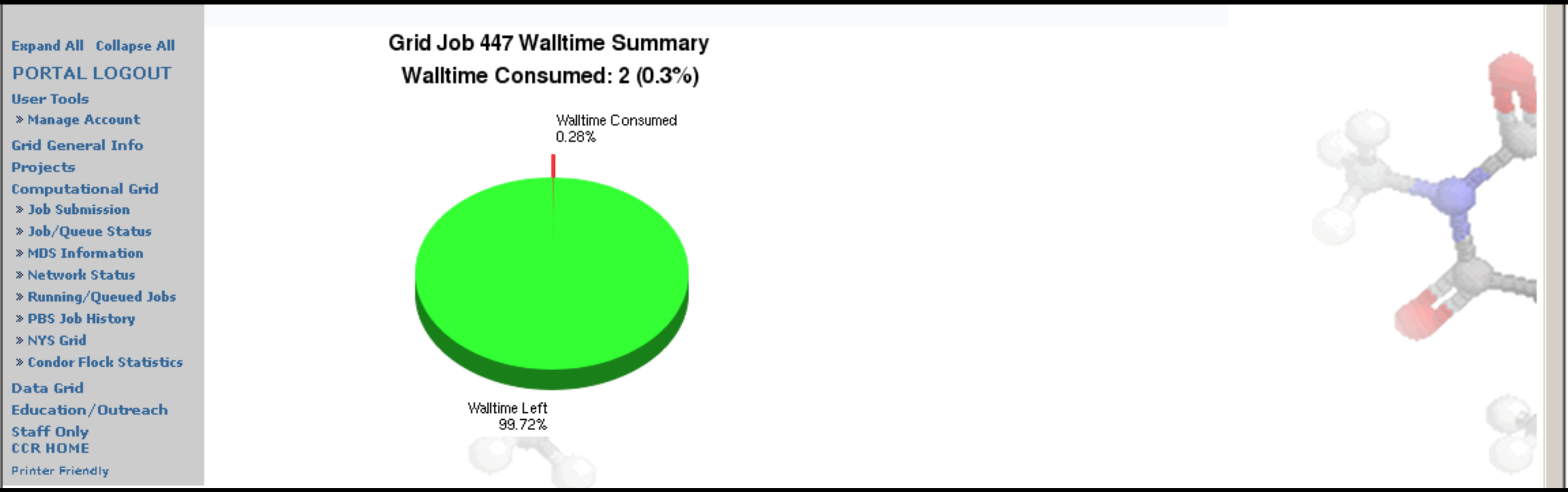
SnB Review (Grid job ID: 447)



Graphical Representation of Intermediate Job Status



Histogram of Completed Trial Structures



Walltime Summary Chart

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

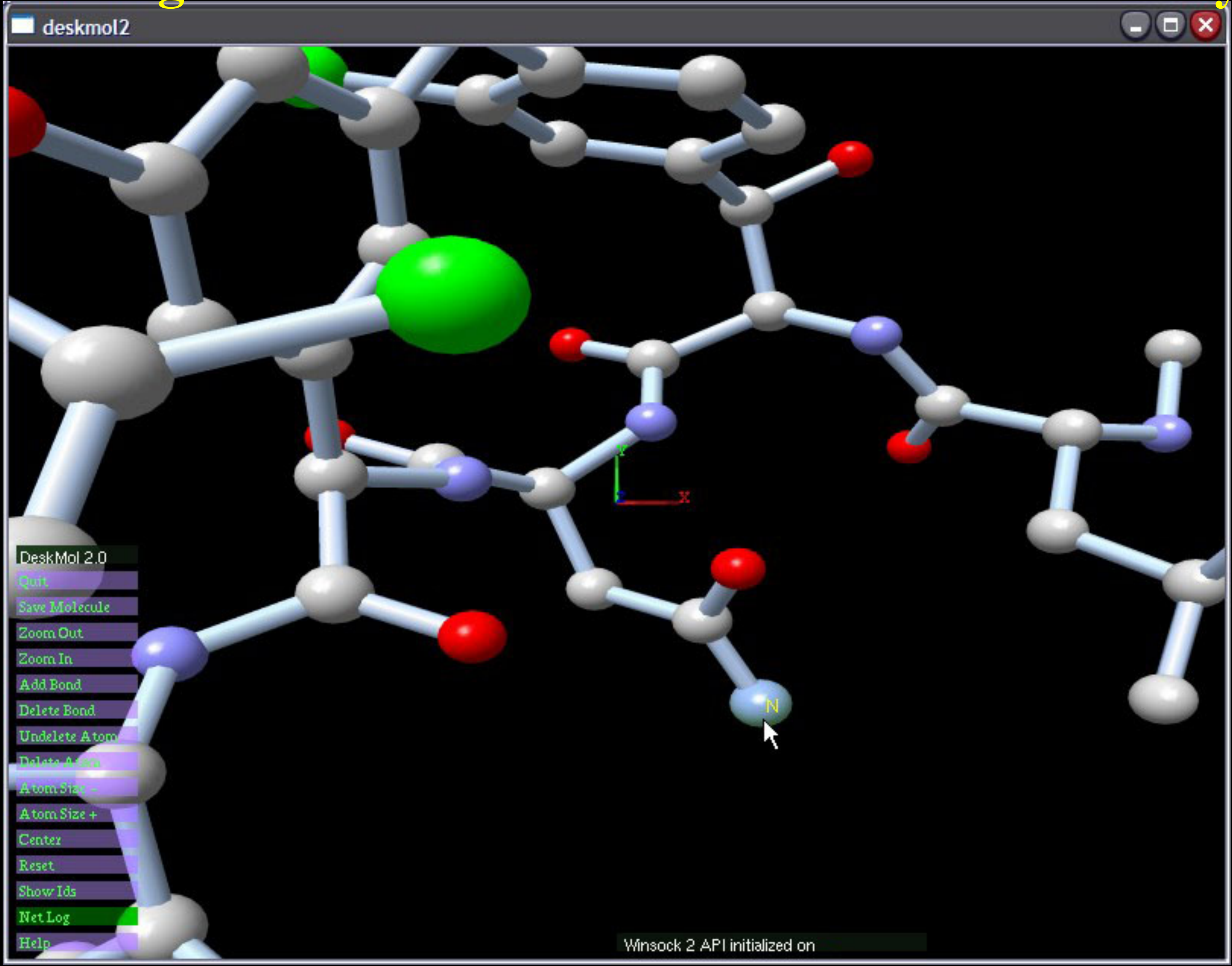
Filter Job List

SnB

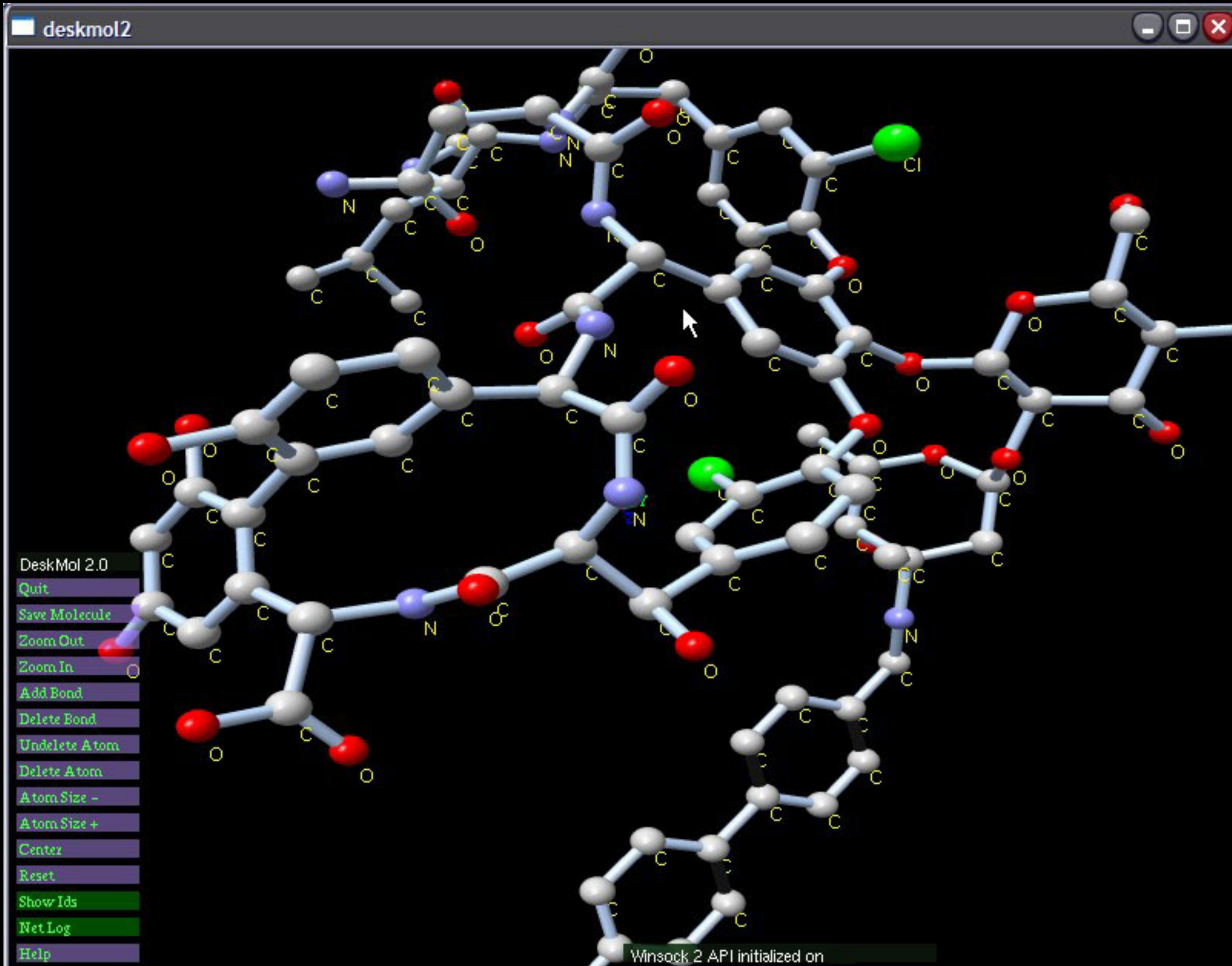
Job Id	Job Name	Resource	Num Procs	Status	Percent Complete	Last Update	Cancel Job	Drilldown
447	iledhkl	clearwater.ccr.buffalo.edu	5	RUNNING	28.5	15-Mar-2005 10:22:00		
446	trilys	clearwater.ccr.buffalo.edu	10	RUNNING	1	15-Mar-2005 10:22:00		
444	64chkl	nash.ccr.buffalo.edu	3	COMPLETE	100	14-Mar-2005 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



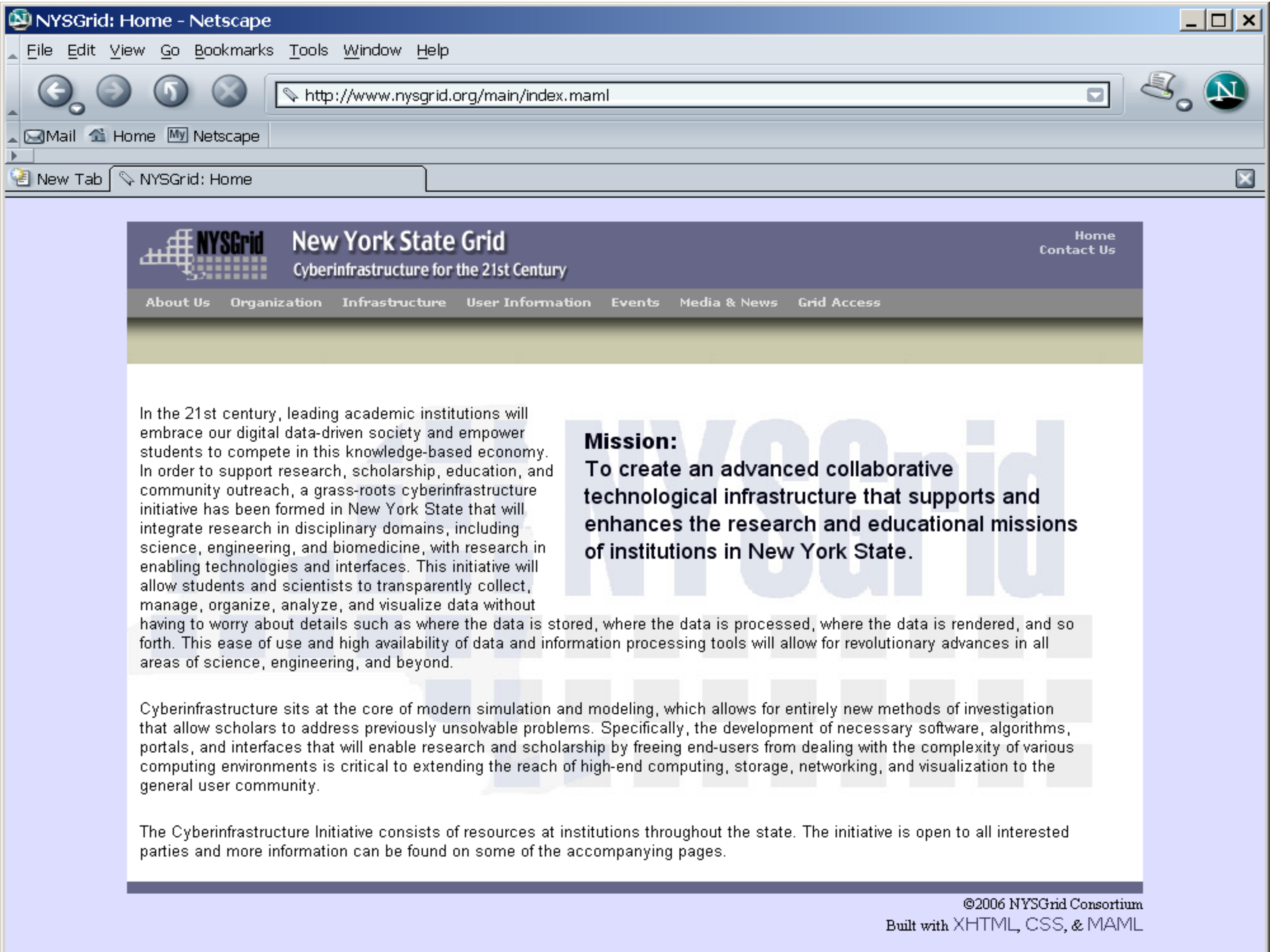
User starts up – default image of structure.



Molecule scaled, rotated, and labeled.

NYSGrid.org

- Grass-Roots Cyberinfrastructure Initiative in NYS
- Open to academic, research, government, and industrial organizations.
- Goal is to allow transparent collection, management, organization, analysis, and visualization of data, while ignoring location.
- Enable Research, Scholarship, and Economic Development in NYS.
- Mission Stmt: To create and advance collaborative technological infrastructure that supports and enhances the research and educational missions of institutions in NYS.



In the 21st century, leading academic institutions will embrace our digital data-driven society and empower students to compete in this knowledge-based economy. In order to support research, scholarship, education, and community outreach, a grass-roots cyberinfrastructure initiative has been formed in New York State that will integrate research in disciplinary domains, including science, engineering, and biomedicine, with research in enabling technologies and interfaces. This initiative will allow students and scientists to transparently collect, manage, organize, analyze, and visualize data without having to worry about details such as where the data is stored, where the data is processed, where the data is rendered, and so forth. This ease of use and high availability of data and information processing tools will allow for revolutionary advances in all areas of science, engineering, and beyond.

Mission:
To create an advanced collaborative technological infrastructure that supports and enhances the research and educational missions of institutions in New York State.

Cyberinfrastructure sits at the core of modern simulation and modeling, which allows for entirely new methods of investigation that allow scholars to address previously unsolvable problems. Specifically, the development of necessary software, algorithms, portals, and interfaces that will enable research and scholarship by freeing end-users from dealing with the complexity of various computing environments is critical to extending the reach of high-end computing, storage, networking, and visualization to the general user community.

The Cyberinfrastructure Initiative consists of resources at institutions throughout the state. The initiative is open to all interested parties and more information can be found on some of the accompanying pages.

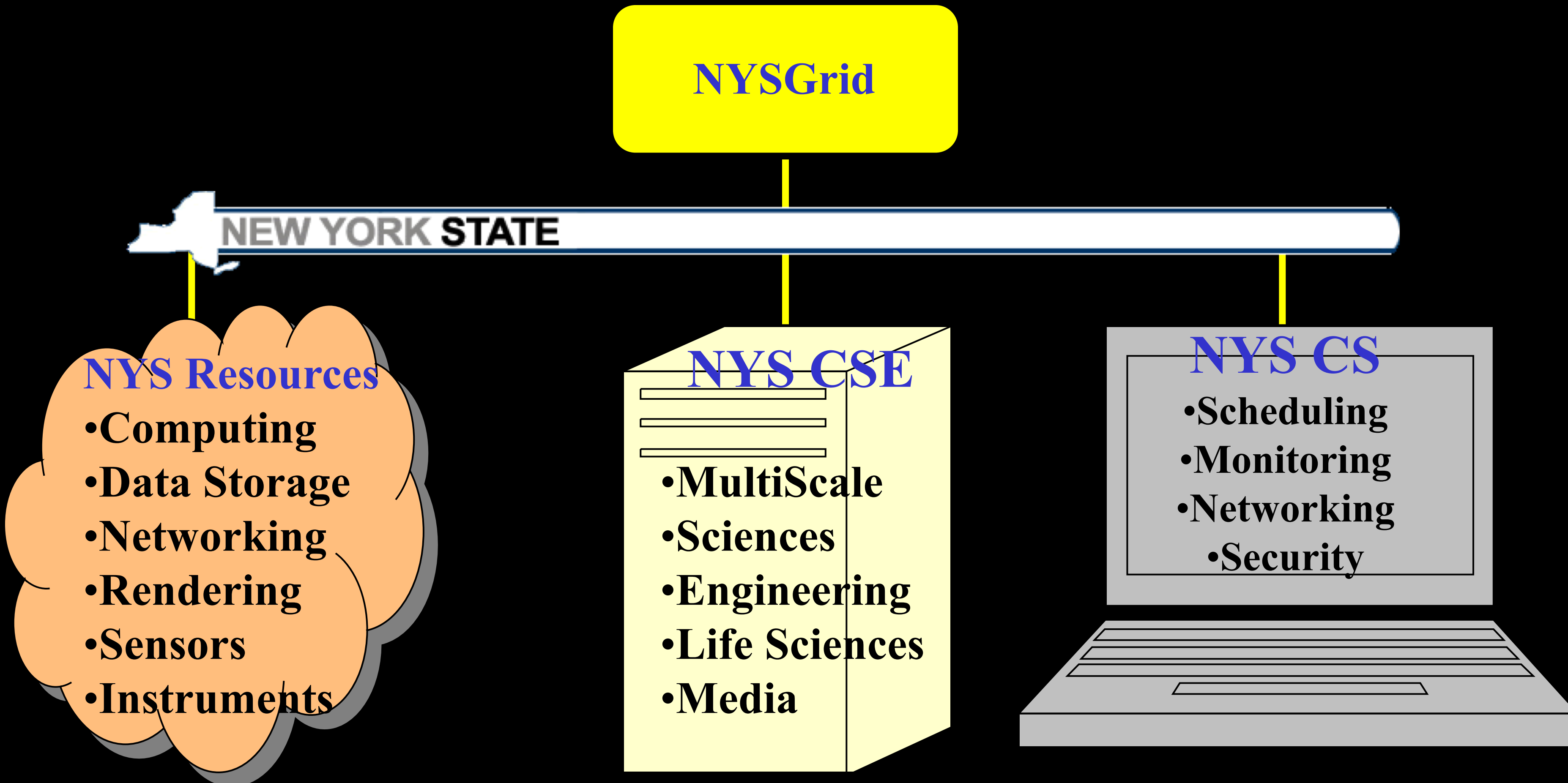
Current NYS Grid Participation

- Albany • √
- Alfred
- Binghamton • √
- Brookhaven
- Buffalo • √
- Columbia •
- Cornell • √
- Geneseo • √
- Hauptman-Woodward • √
- Iona •
- Marist • √
- Memorial Sloan-Kettering
- NYU • √
- Niagara • √
- RIT • √
- Rochester • √
- RPI •
- Stony Brook • √
- Syracuse • √
- NYSERNet

• - expressed interest in NYS Grid

√ - on NYS Grid

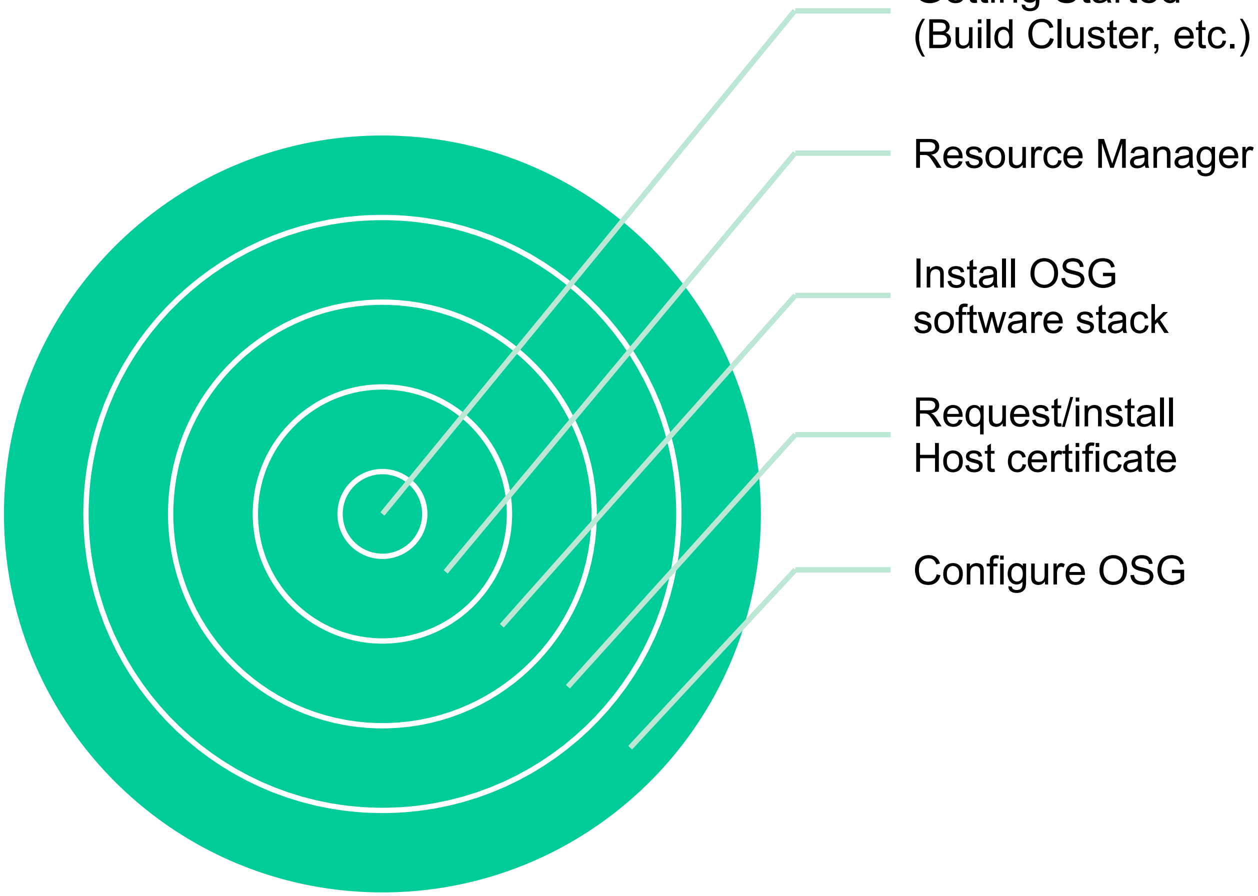
NYSGrid.org Organization



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."

NYS Grid Implementation Details

Figure Courtesy of Jon Bednasz, CCR/UB



Getting Started
(Courtesy of Jon Bednasz & Steve Gallo, CCR/UB)

- Physically build a cluster
  - 1 head node
  - 4+ compute nodes
- Install Cluster Software
  - Operating System (Red Hat)
  - Drivers for Interconnect (Myrinet, Infiniband, etc.)
  - Resource Manager (PBS, LSF, Condor, SGE)
- Identify Gatekeeper Node for OSG Software
  - Either stand alone machine or co-resident on Head Node
  - 5GB of space in /opt/grid
  - 5GB of space in /grid-tmp
- Need to have ability to adjust firewalls
- Need to have ability to add users

Installing OSG Stack on Gatekeeper

- Installs are done via PACMAN
  - wget http://physics.bu.edu/pacman/sample_cache/tarballs/pacman-3.16.1.tar.gz
- Install OSG software
  - pacman -get OSG:ce
- Install (1) Package for your Resource Manager
  - pacman -get OSG:Globus-Condor-Setup
  - pacman -get OSG:Globus-PBS-Setup
  - pacman -get OSG:Globus-LSF-Setup
  - pacman -get OSG:Globus-SGE-Setup

NYSGrid.org Technical Group

- Jon Bednasz, Buffalo, Chair
- Steve Gallo, Buffalo
- Eric Warnke, Albany
- Steaphan Greene, Binghamton
- Ken Smith, Columbia
- Resa Alvord, Cornell
- Kirk Anne, Geneseo
- Steve Potter, Hauptman-Woodward
- Robert Schiaffino, Iona
- Earle Nietzel, Marist
- Ann Rensel, Niagara
- Chris Grim, NYU
- Rick Bohn, RIT
- Bill Webster, Rochester
- Lindsay Todd, RPI
- Ajay Gupta, Stony Brook
- Jorge González Outeiriño, Syracuse

NYSGrid.org Activities & Board

Activities

- Technical Working Group
- Middleware
- User Support and Services / EOT
- Communications
- Infrastructure
- Resource Providers
- Funding

Board

- Russ Miller
- Gurcharan Khanna
- Linda Callahan
- Mark Shephard
- Tim Lance
- (Heather Stewart)
- Jim Davenport
- Chris Haile

Technical WG Current Efforts
(Led by Steve Gallo and Jon Bednasz)

- NYS Grid is Available
- OSG Jobs Running on NYS Grid
- CCR/UB & CTC/Cornell
  - Streamline bringing users onto NYS Grid
    - Documentation
    - Recommendations
- Need Early Adopters
  1. Current Grid Users
  2. New Users to Grid with HPC Needs

Middleware WG Current Efforts

- Discussions on current state of Middleware at Buffalo, Binghamton, & RPI
  - Scheduling
  - Portals
  - Monitoring
  - Fault Tolerance
  - Checkpoint/Restart

CCR Outreach

- HS Summer Workshops in Computational Science
  - Chemistry, Visualization, Bioinformatics
  - 10-14 HS Students Participate Each Summer for 2 weeks
  - Project-Based Program

CCR 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
UB