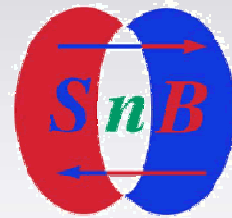# High-Performance Computing & High-End Visualization in Buffalo

Russ Miller

Center for Computational Research

**Computer Science & Engineering
SUNY-Buffalo**

Hauptman-Woodward Medical Inst

**NSF, NIH, DOE
NIMA, NYS, HP**

Grid3 2003

SnB

GRID

Advanced Center for Computational Research Data Center

**University at Buffalo**

The State University of New York

Buffalo, New York
Two Seasons: Winter and July 4

- **The Queen City: 2nd Largest City in NYS**
- **Waterfront City: on North Coast of U.S.**
- **City of Lights**
  - First U.S. city to have electric street lights
  - Pan American Exposition (1901)
    - Pres. McKinley Shot
- **Architecture**
  - Frederick Law Olmsted
  - Frank Lloyd Wright
- **Underground Railroad: slaves escaped to freedom in Canada**
- **Four straight Super Bowl appearances**
- **Culinary Delights**
  - (Buffalo) Wings: Anchor Bar, 1964
  - Beef on Weck
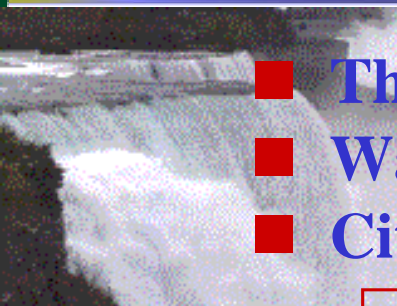
Center for Computational Research
1998-2005 Overview

- **High-End Computing, Storage, Networking, and Visualization**
  - ❑ **~140 Research Groups in 37 Depts**
    - ○ Physical Sciences
    - ○ Life Sciences
    - ○ Engineering
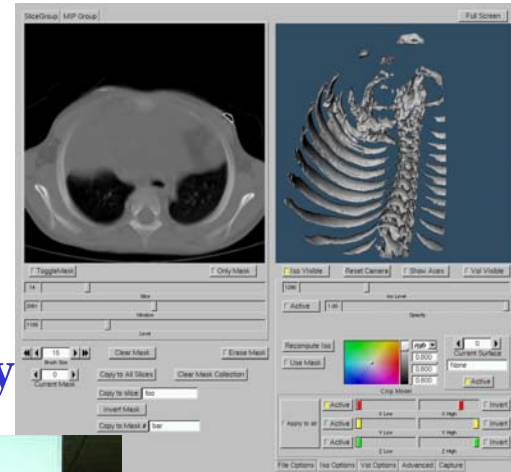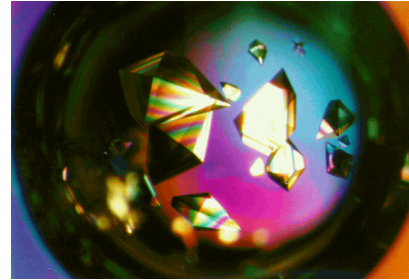   - ○ Scientific Visualization, Medical Imaging, Virtual Reality
  - ❑ **13 Local Companies**
  - ❑ **10 Local Institutions**
- **External Funding: $300M+**
- **Total Leveraged WNY: $500M+**
- **1100+ Publications**
- **EOT, Economic Development, Software, Media, Algorithms, Consulting, Training, CPU Cycles…**

CCR by the Numbers (2005)

- **Current Technical Staff: 20→13**
  - Seven (7) Contracted & Univ Tech Staff cut by Provost
  - Associate Director
  - Computational Scientist (3)
  - Database Administrator
  - Scientific Visualization
  - System Administration (5)
  - Storage Area Network Admin
  - Multimedia
- **Support Staff: 3 FTE**
  - Financial/Contracts (2)
  - Receptionist
- **Research Staff: 5 FTE**

- **Initial 7-Year Funding Model**
  - SUNY-Buffalo Contribution: $1.3M
    - Personnel: $1.2M
    - Operating: $0.1M
  - User's Contributions: $0.4M
  - Annual Expend: ~$2.4M
  - Opportunistic Funding Model
    - Equipment, Maintenance, Licenses
 - **ROI: $7M → $300M @ SUNY-B**
- **New Administration's Model**
  - Moved into Bioinformatics
  - Move 30 mins from Main Campus
  - 50% Additional Reduction in Personnel & Operating in Addition to Current Cuts
  - Maintain Opportunistic Funding
  - Provost Rationale
    - Will Increase Users Contributions
    - "Provides Stability"

Major Compute/Storage Resources

- **Dell Linux Cluster (10TF peak)**
  - 1600 Xeon EM64T Processors (3.2 GHz)
  - 2 TB RAM; 65 TB Disk
  - Myrinet / Force10
  - 30 TB EMC SAN
- **Dell Linux Cluster (2.9TF peak)**
  - 600 P4 Processors (2.4 GHz)
  - 600 GB RAM; 40 TB Disk; Myrinet
- **Dell Linux Cluster (6TF peak)**
  - 4036 Processors (PIII 1.2 GHz)
  - 2TB RAM; 160TB Disk; 16TB SAN
- **IBM BladeCenter Cluster (3TF peak)**
  - 532 P4 Processors (2.8 GHz)
  - 5TB SAN

- **SGI Altix3700 (0.4TF peak)**
  - 64 Processors (1.3GHz ITF2)
  - 256 GB RAM
  - 2.5 TB Disk
- **Apex Bioinformatics System**
  - Sun V880 (3), Sun 6800
  - Sun 280R (2)
  - Intel PIIIs
  - Sun 3960: 7 TB Disk Storage

- **HP/Compaq SAN**
  - 75 TB Disk; 190 TB Tape
  - 64 Alpha Processors (400 MHz)
  - 32 GB RAM; 400 GB Disk

CCR Visualization Resources

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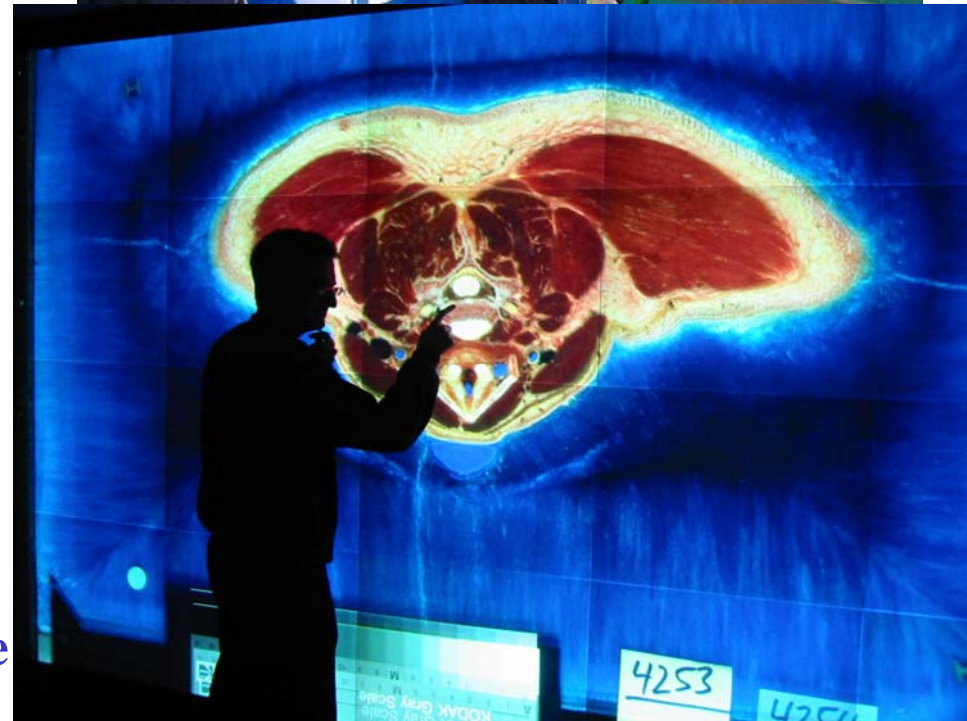
CCR Research & Projects

- **Archaeology**
- **Bioinformatics/Protein Folding**
- **Computational Chemistry**
- **Computational Fluid Dynamics**
- **Data Mining/Database**
- **Earthquake Engineering**
- **Environ Modeling & Simulation**
- **Grid Computing**
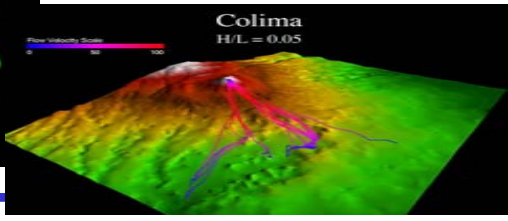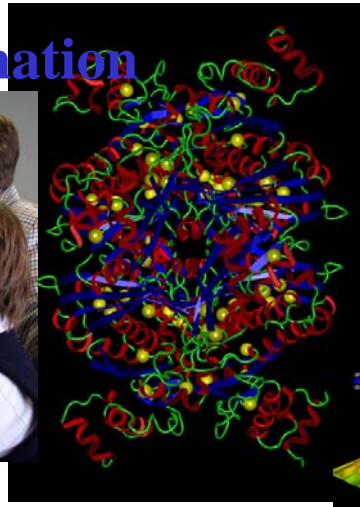- **Molecular Structure Determination**
- **Physics**
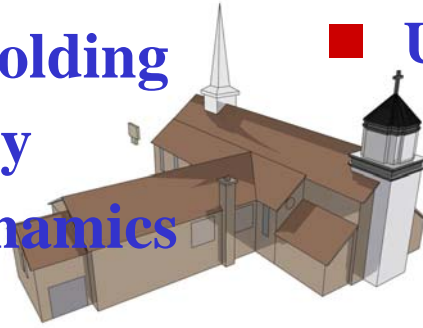
- **Videos: MTV**
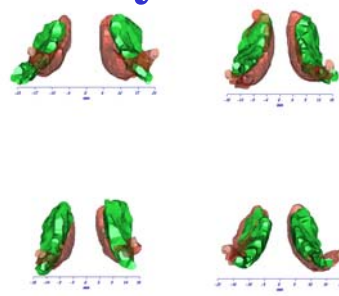- **Urban Simulation and Viz**
  - StreetScenes
  - I-90 Toll Barrier
  - Medical Campus
  - Peace Bridge
- **Accident Reconstruction**
- **Scientific Viz**
  - Dental
  - Surgery
  - MRI/CT Scan
  - Confocal Microscopy
  - Crystallization Wells
  - Collaboratories

University at Buffalo *The State University of New York* **Center for Computational Research**

CCR

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**

actctrl, Time: 07h 00m 21.4s
Camera at (-17,529) Alt 76ft Heading 180° (South)

Animation & Simulation

Rendered Scenes

Williamsville Toll Barrier Improvement Project

Williamsville Toll Barrier Improvement Project (Night)

Eastbound Toll Barrier

Westbound Toll Barrier

Initial Photo Match incorporating real and computer-generated components

Peace Bridge Visualization: Animation & Simulation

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

PHOTO AND STORY BY BRUCE JACKSON
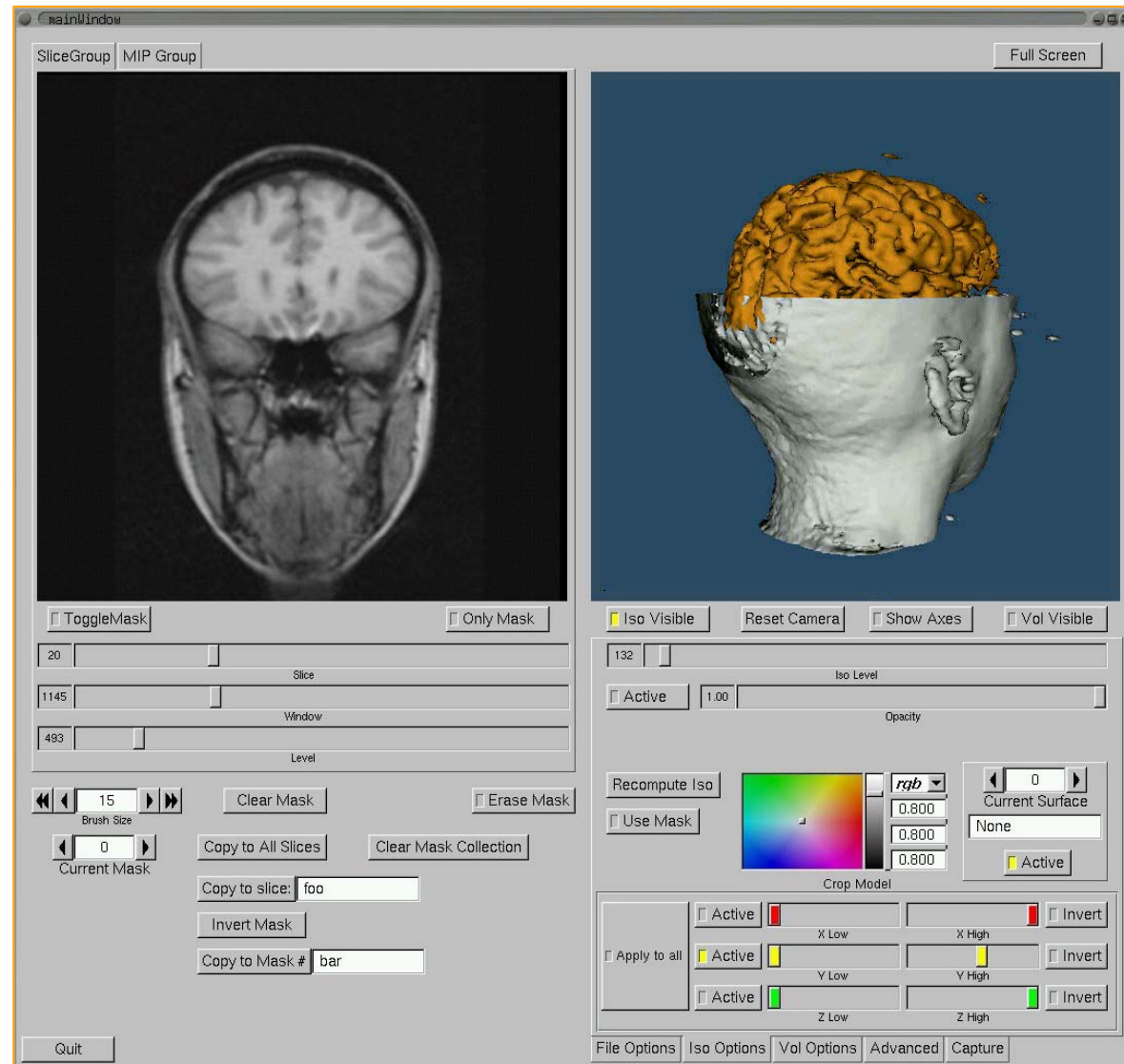
Public Forum

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**
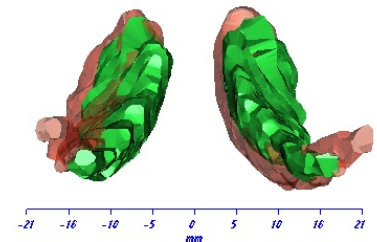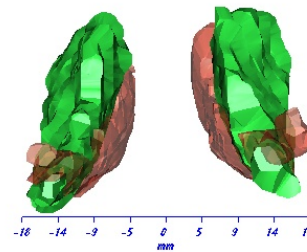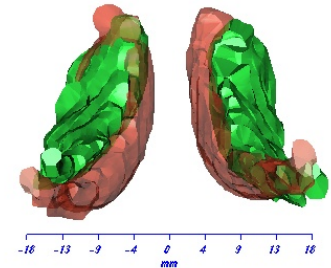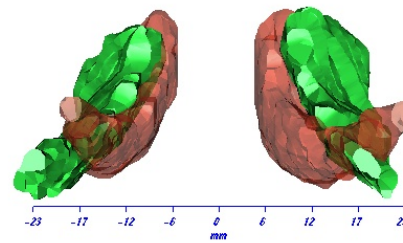
3D Medical Visualization App

- **Collaboration with Children's Hospital**
  - ❑ **Leading miniature access surgery center**
- **Application reads data output from a CT Scan**
- **Visualize multiple surfaces and volumes**
- **Export images, movies or CAD representation of model**
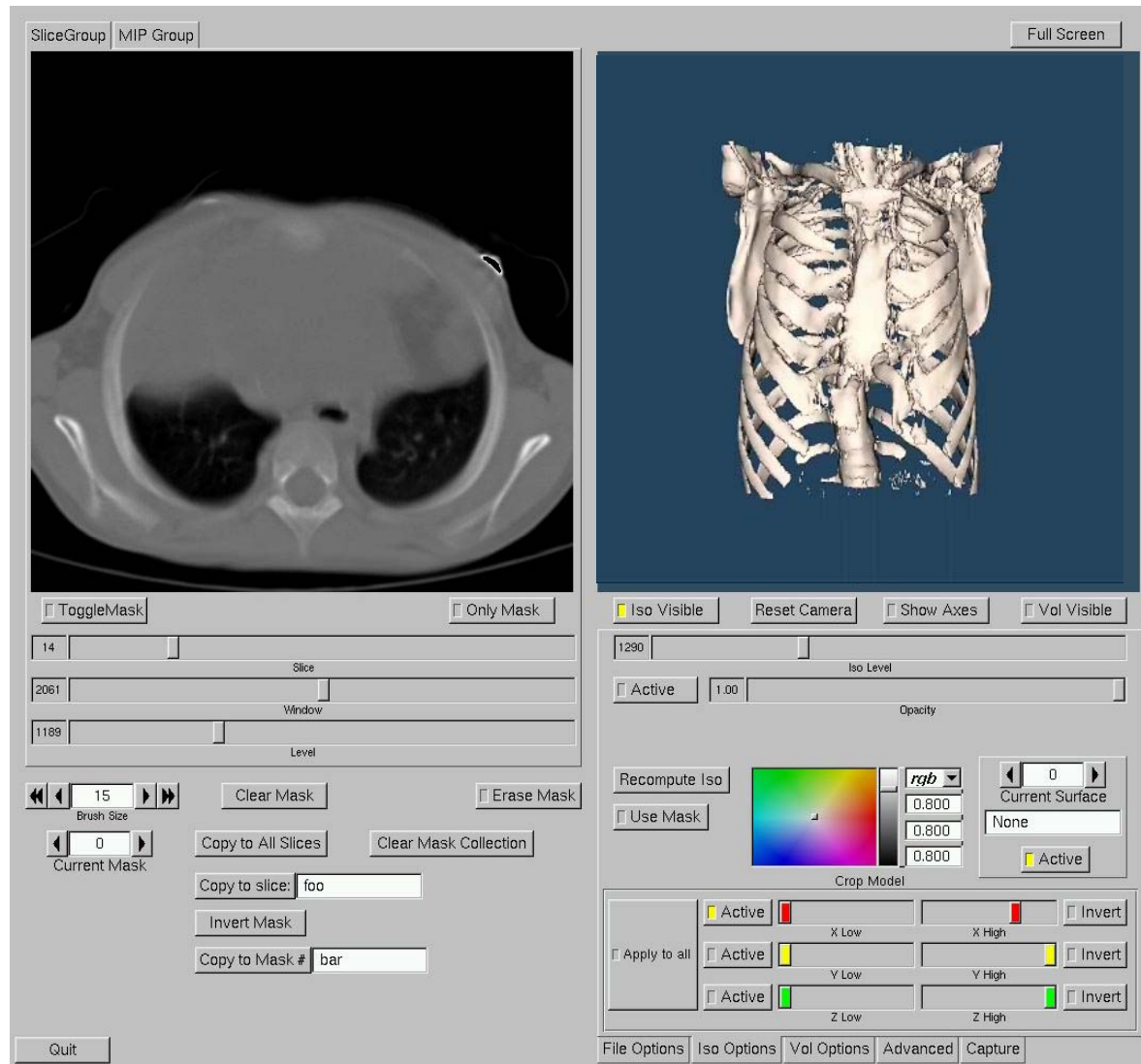
Science & Engineering

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**
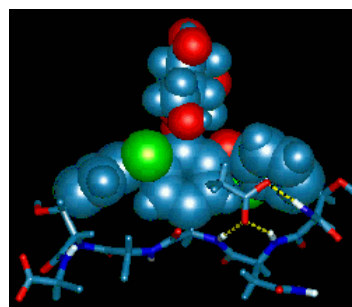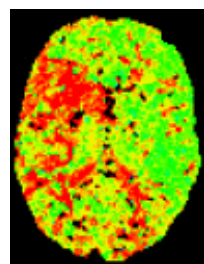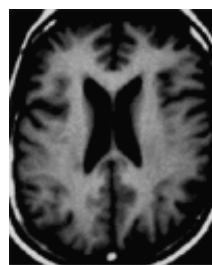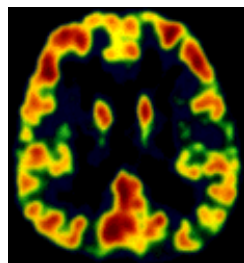
- **Direct Methods Structure Determination**
  - **Listed on "Top Ten Algorithms of the 20th Century"**
  - **Vancomycin & Gramacidin A**
- **High Throughput Crystallization Method**
- **NIH National Genomics Center: Northeast Consortium**
- **HHMI: Center for Genomics & Proteomics**
- **NYSCOEBLS: $360M with RPCI & HWI**

Groundwater Flow Modeling

- **Regional-scale modeling of groundwater flow and contaminant transport (Great Lakes Region)**

- **Ability to include all hydrogeologic features as independent objects**

- **Current work is based on *Analytic Element Method***

- **Key features:**
  - ❑ **High precision**
  - ❑ **Highly parallel**
  - ❑ **Object-oriented programming**
  - ❑ **Intelligent user interface**
  - ❑ **GIS facilitates large-scale regional applications**

- **Utilized 10,661 CPU days (32 CPU years) of computing in past year on CCR's commodity clusters**
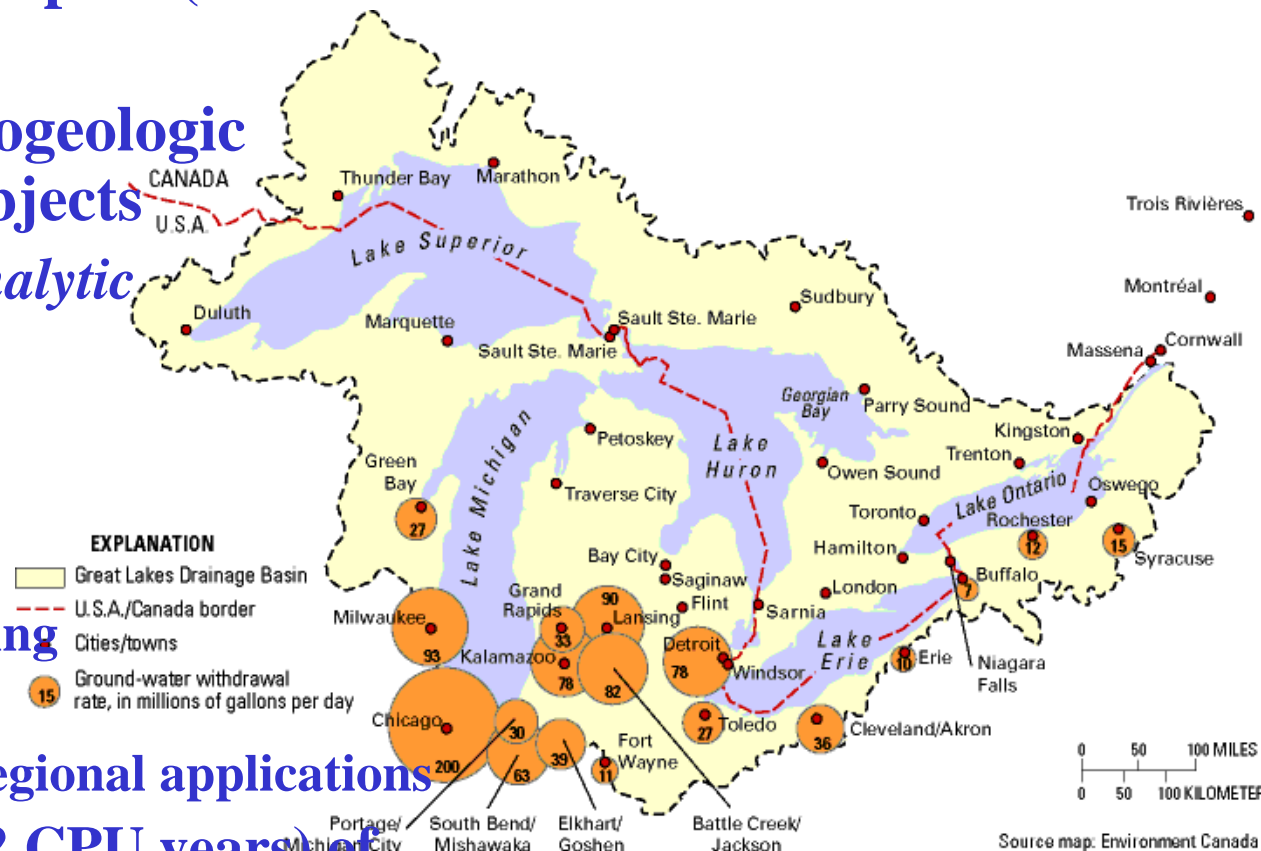
Geophysical Mass Flow Modeling

- **Modeling of Volcanic Flows, Mud flows (flash flooding), and Avalanches**
- **Integrate information from several sources**
  - **Simulation results**
  - **Remote sensing**
  - **GIS data**
- **Develop realistic 3D models of mass flows**
- **Present information at appropriate level**

Colima
H/L = 0.05

Flow Velocity Scale
0   50   100

SCALABLE IT

Scientists

Hazard Managers

Public

High End Simulation and Visualization

Remote Real time visualization server

Web Servers

Web based visualization
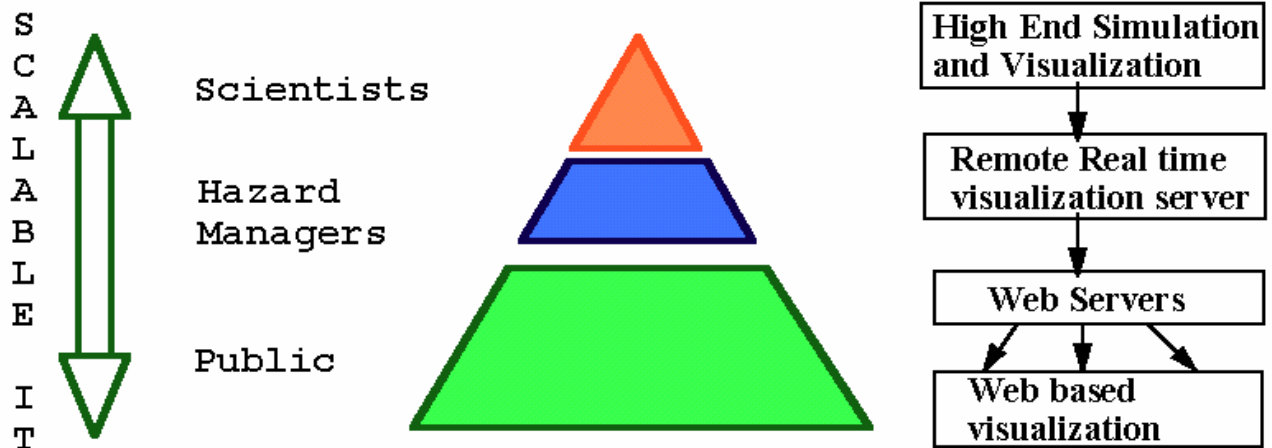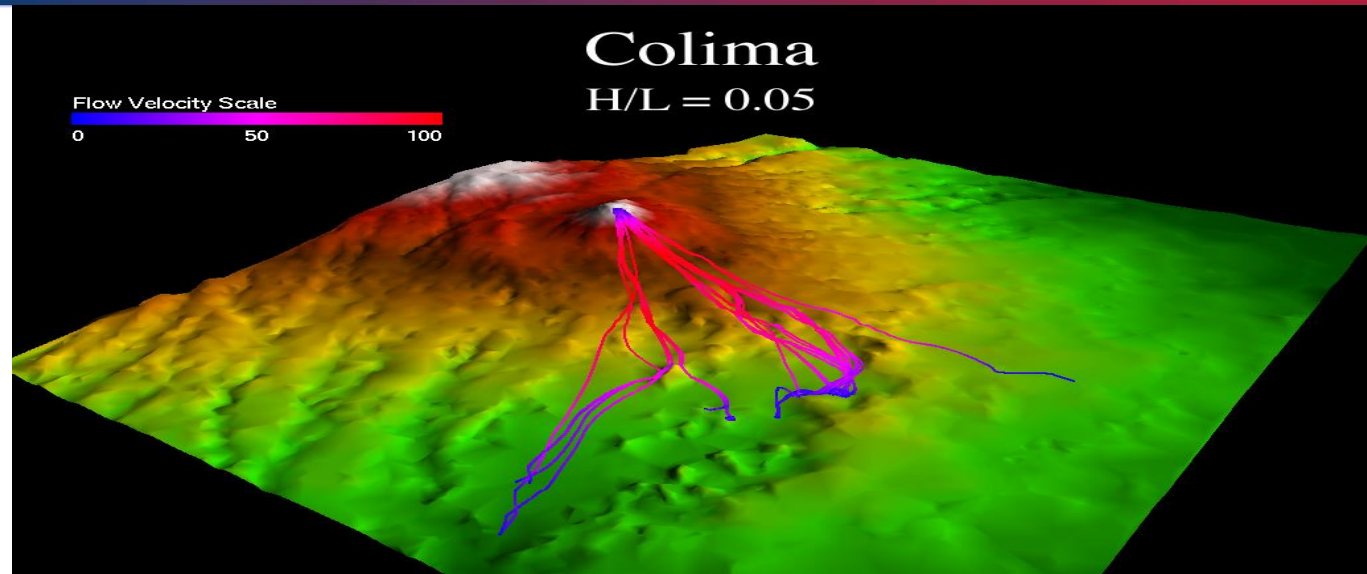
Shake-and-Bake

Molecular Structure Determination

from X-Ray Crystallographic Data

X-Ray Crystallography

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

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

Source of X-rays   Crystal
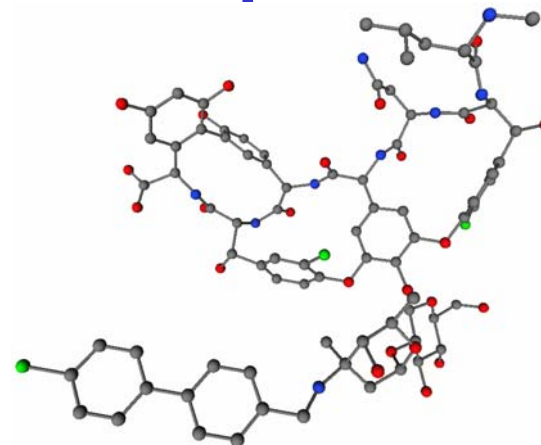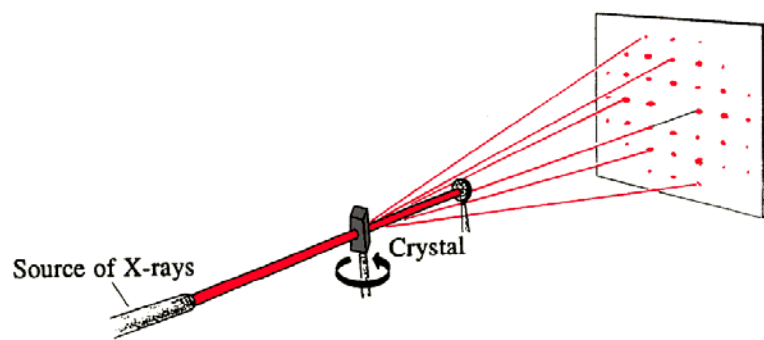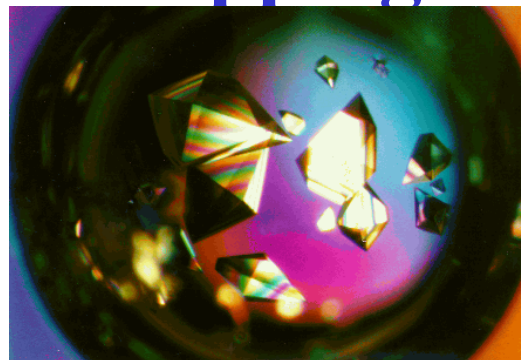
  3. **Determine molecular structure that agrees with diffration data.**

X-Ray Data & Corresponding Molecular Structure

Reciprocal or "Phase" Space

Real Space
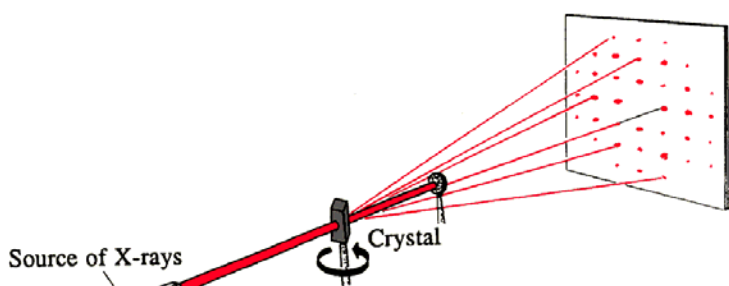
- **Experiment yields reflections and associated intensities.**

- **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.**

FFT

FFT^{-1}
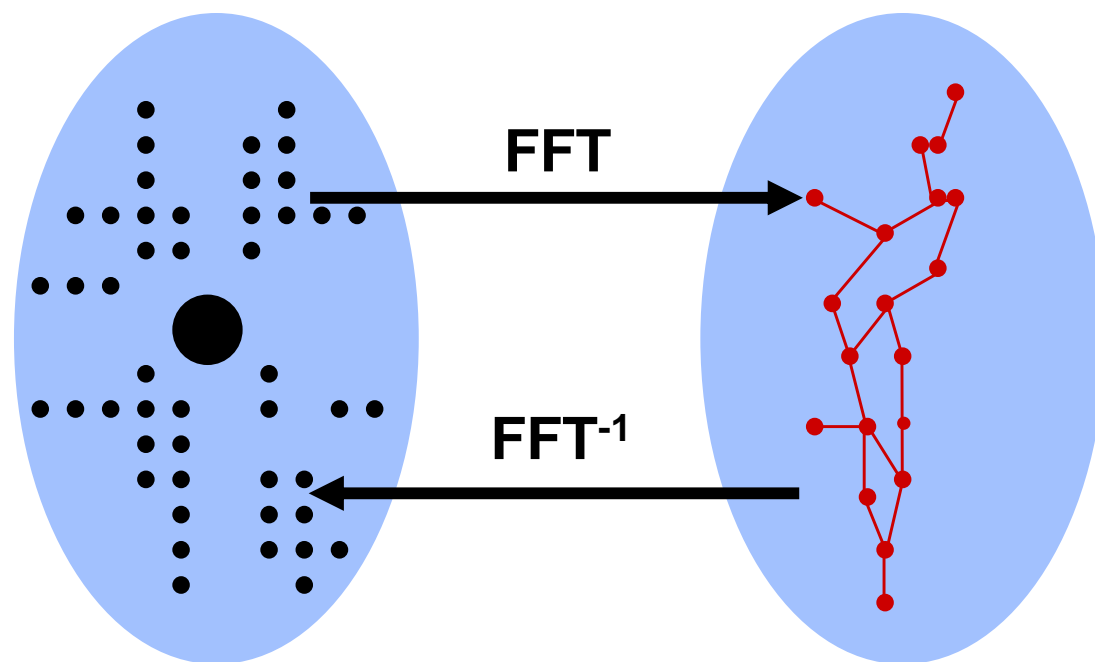
X-Ray Data

Molecular Structure

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.**

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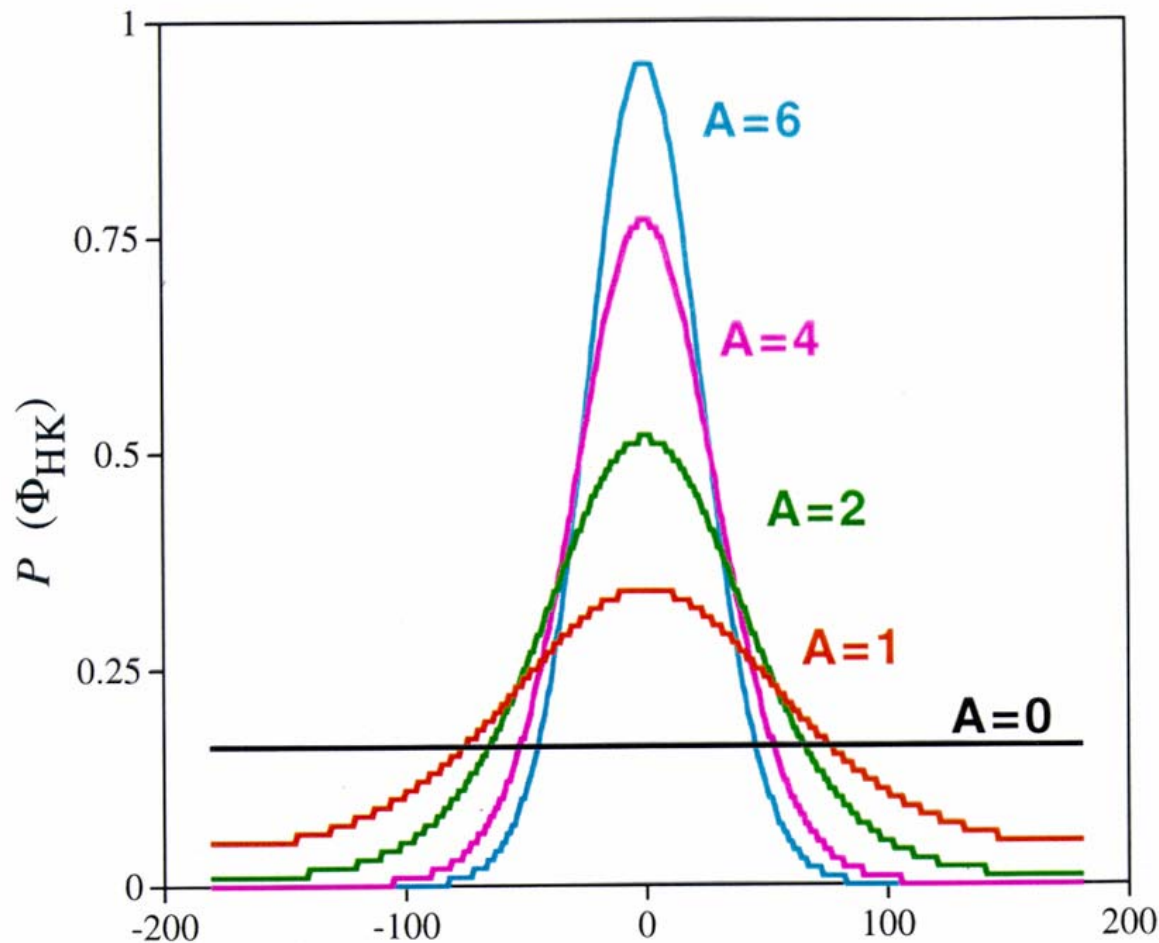
- N=non-H atoms in unit cell
- Each triplet of phases or structure invariant, Φ_{HK}, has an associated parameter

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

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

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

Conventional Direct Methods

Trial

Phases

Tangent
Formula

**Phase
Refinement**

FFT

**Density
Modification
(Peak Picking)**

?

Solutions

Reciprocal Space   *Real Space*
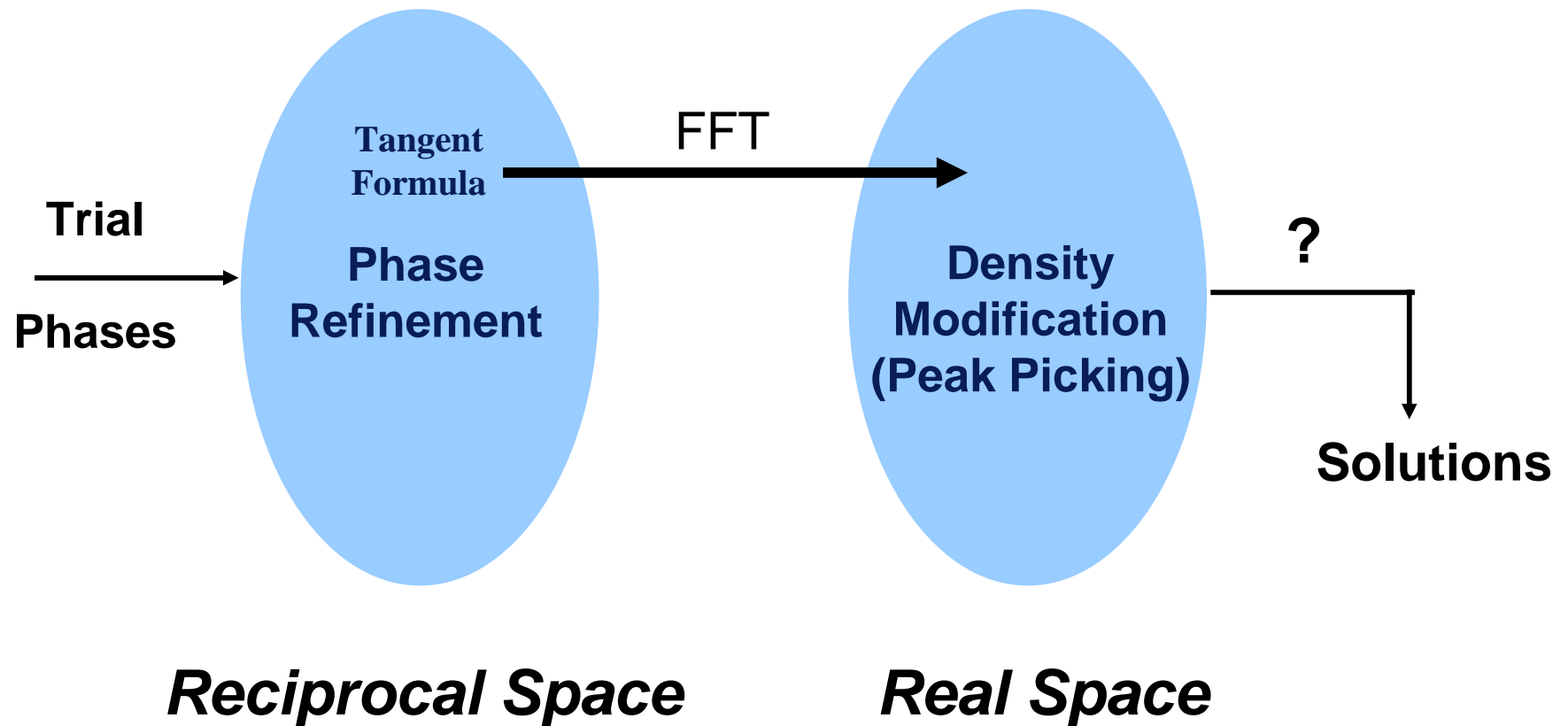
Shake-and-Bake Method: Dual-Space Refinement

Shake-and-Bake

Trial Structures

↓

Structure Factors

↓

Trial

→

Phases

Phase Refinement

Tangent Formula — **FFT** →

Density Modification (Peak Picking) (LDE)

?

Parameter Shift ← **FFT⁻¹**

Solutions

Reciprocal Space "Shake"

Real Space "Bake"
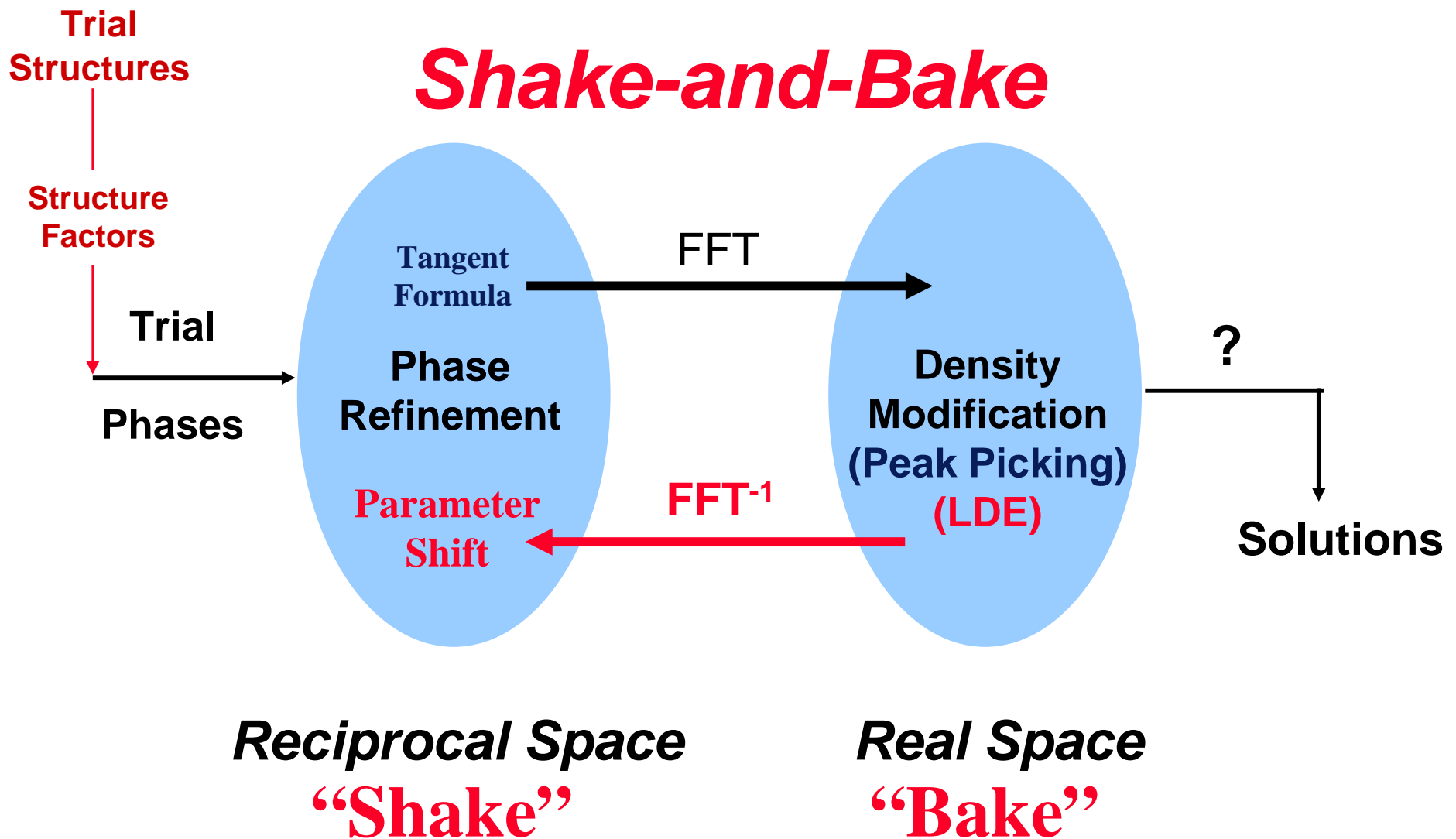
A Direct Methods Flowchart
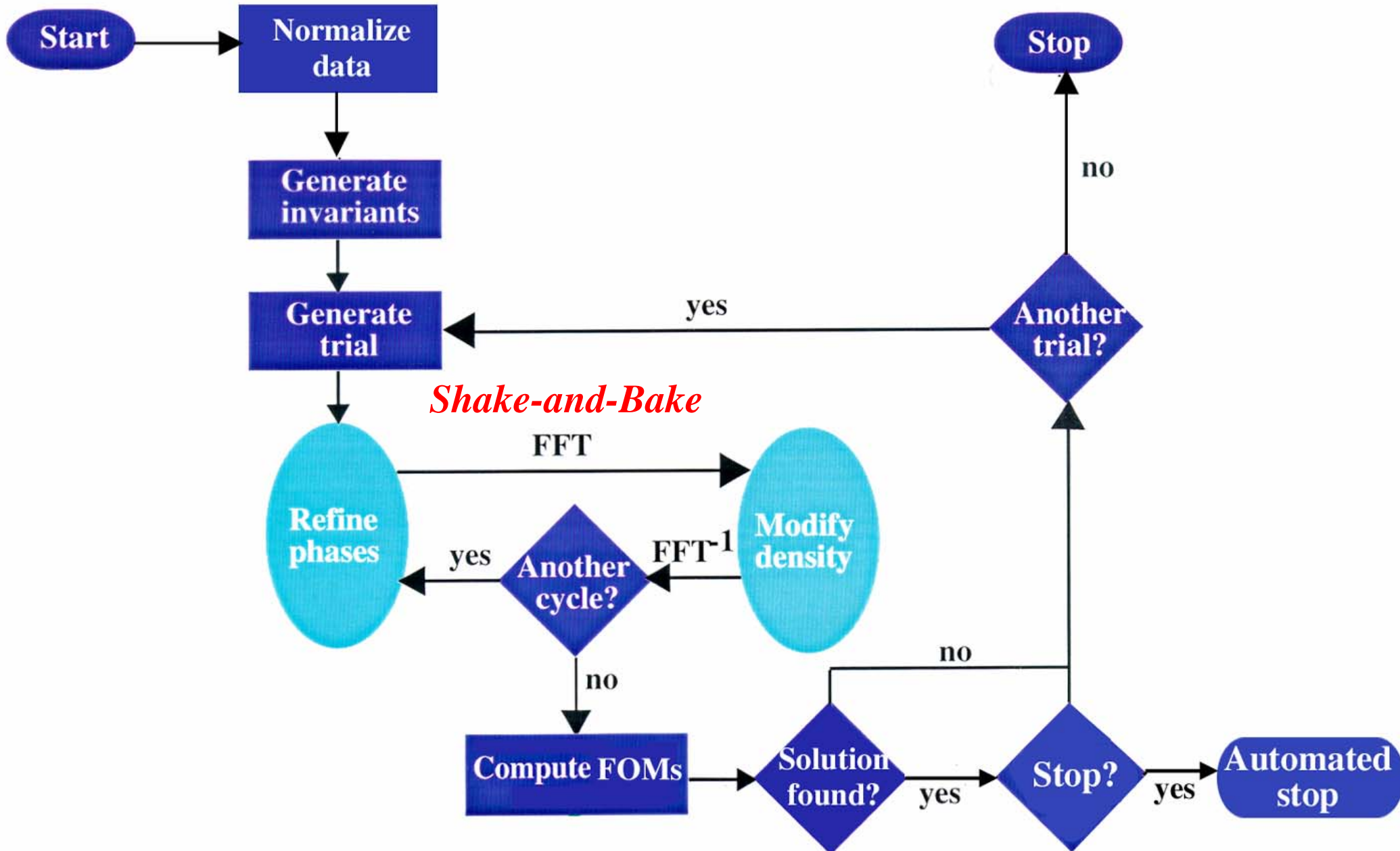
Useful Relationships for Multiple Trial Phasing

Tangent Formula

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

Parameter Shift Optimization

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

$$\text{where} \ |E_H| \propto |F_H| \ \text{normalized in resolution shells}$$

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

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

Ph8755: *SnB* Histogram
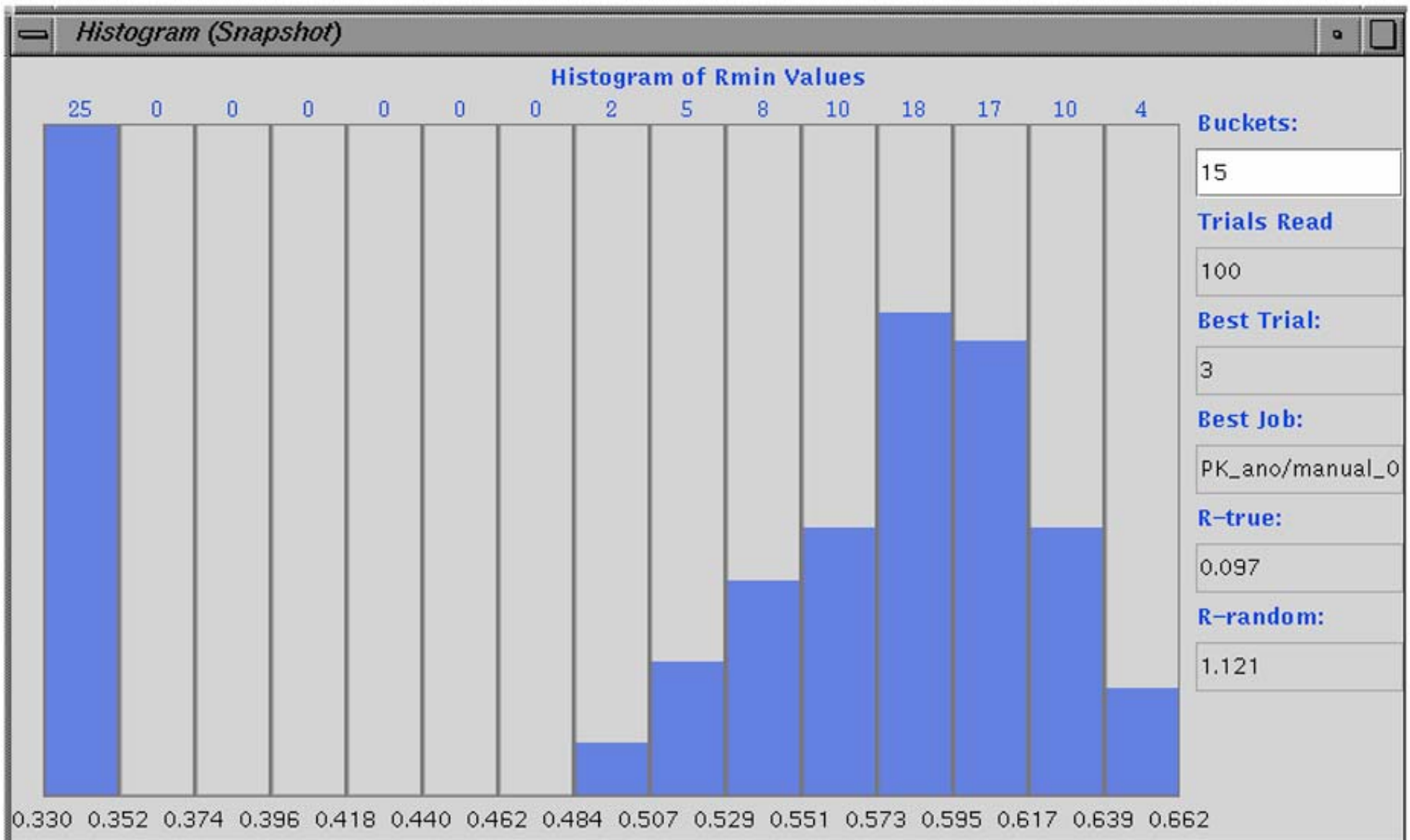
Phasing and Structure Size

Se-Met with *Shake-and-Bake* ······ ?

567 kDa (160 Se)

Se-Met

Multiple Isomorphous Replacement ······ ?

Shake-and-Bake

Conventional Direct Methods

Vancomycin

| 0 | 100 | 1,000 | 10,000 | 100,000 |
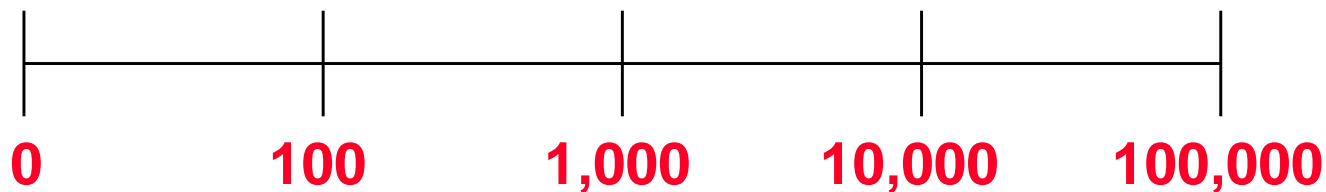
Number of Atoms in Structure

Grid Computing

Grid Computing
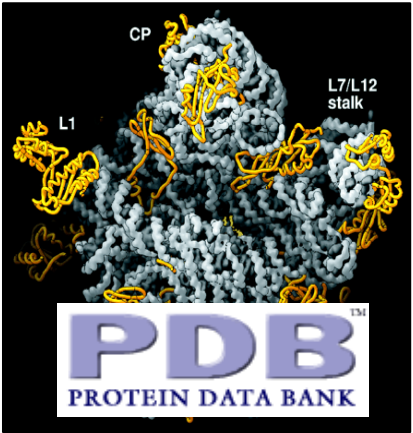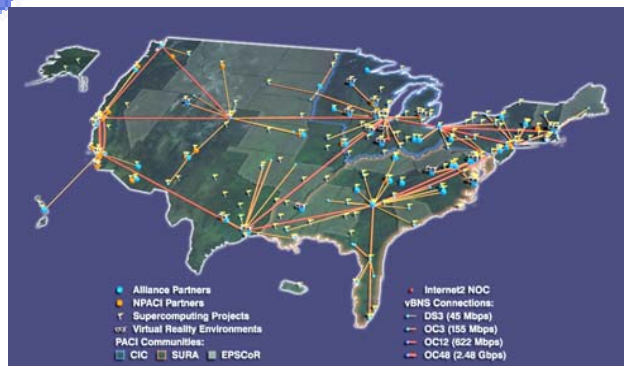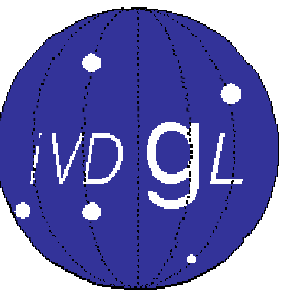
Grid Computing Overview
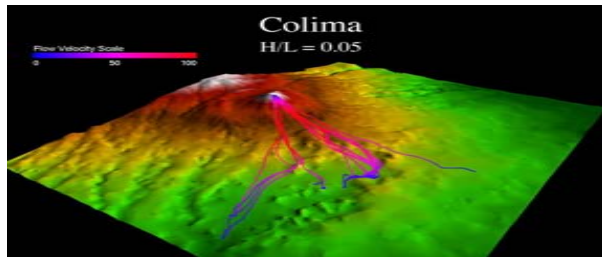
Data **Acquisition**
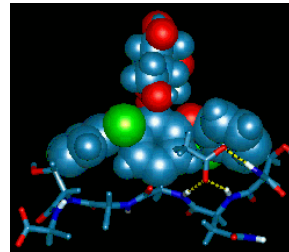
Advanced **Visualization**

Analysis

Imaging Instruments
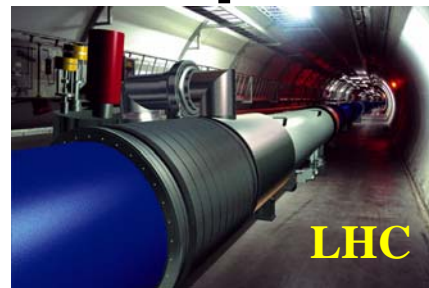
LHC

Computational Resources
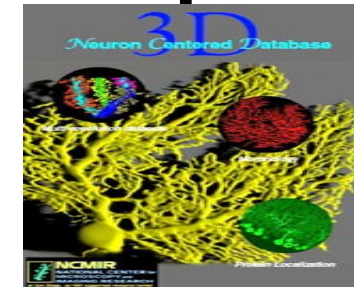
Large-Scale Databases

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

ACDC-Grid Collaborations

- **High-Performance Networking Infrastructure**
- **WNY/NYS Grid Initiative**
- **Grid3+ Collaboration**
- **iVDGL Member**
  - ❑ **Only External Member**
- **Open Science Grid Member**
  - ❑ **Organizational Committee**
  - ❑ **Blueprint Committee**
  - ❑ **Security Working Group**
  - ❑ **Data Working Group**
  - ❑ **GRASE VO**
- **Grid-Lite: Campus Grid**
  - ❑ **HP Labs Collaboration**
- **Innovative Laboratory Prototype**
  - ❑ **Dell Collaboration**

ACDC-Grid Cyber-Infrastructure

- **Integrated Data Grid**
  - ❏ **Automated Data File Migration based on profiling users.**
- **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.**

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 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

Browser view of "miller" group files published by user "rappleye"

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
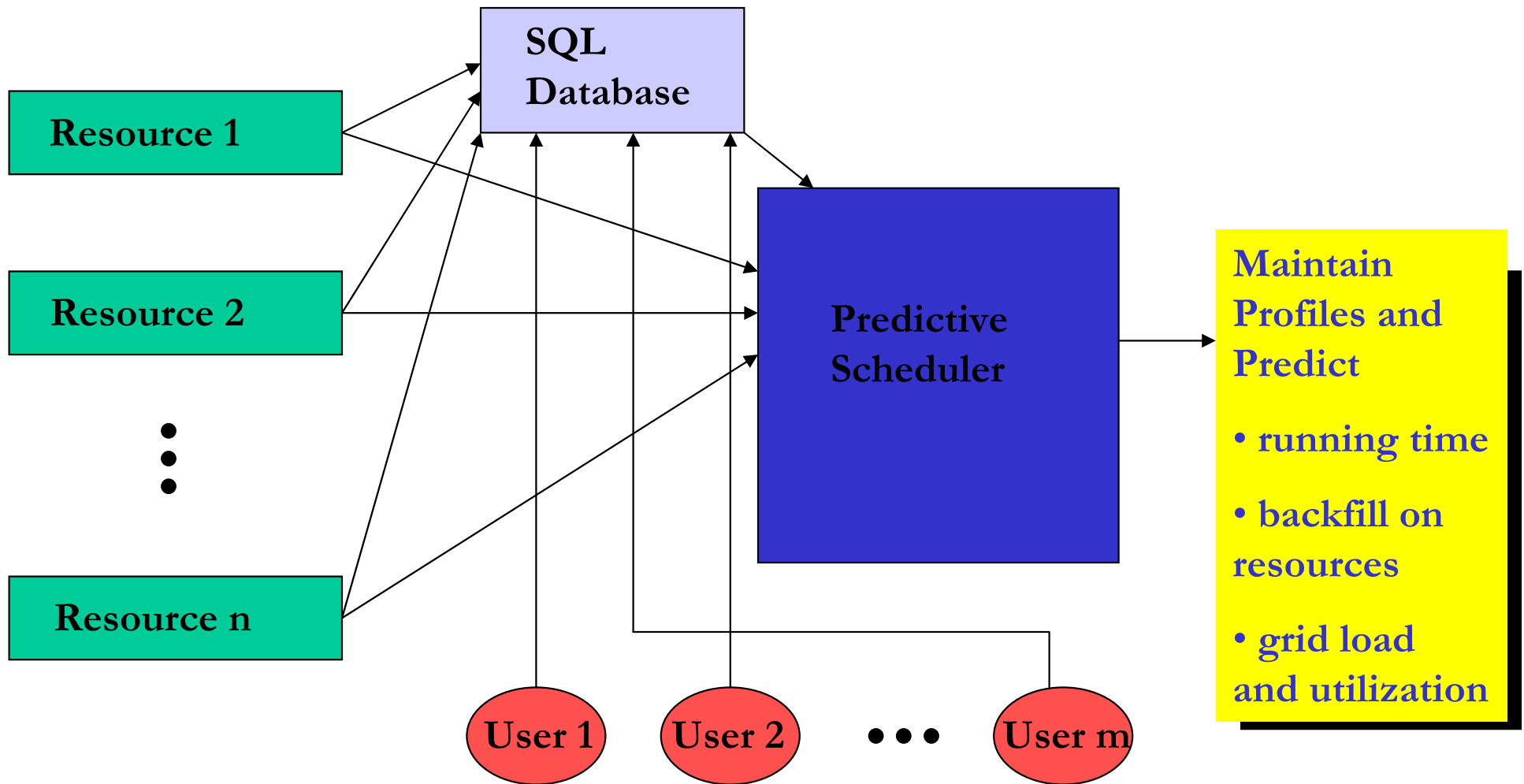
ACDC-Grid Dynamic Resource Allocation at SC03 with Grid3

- **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
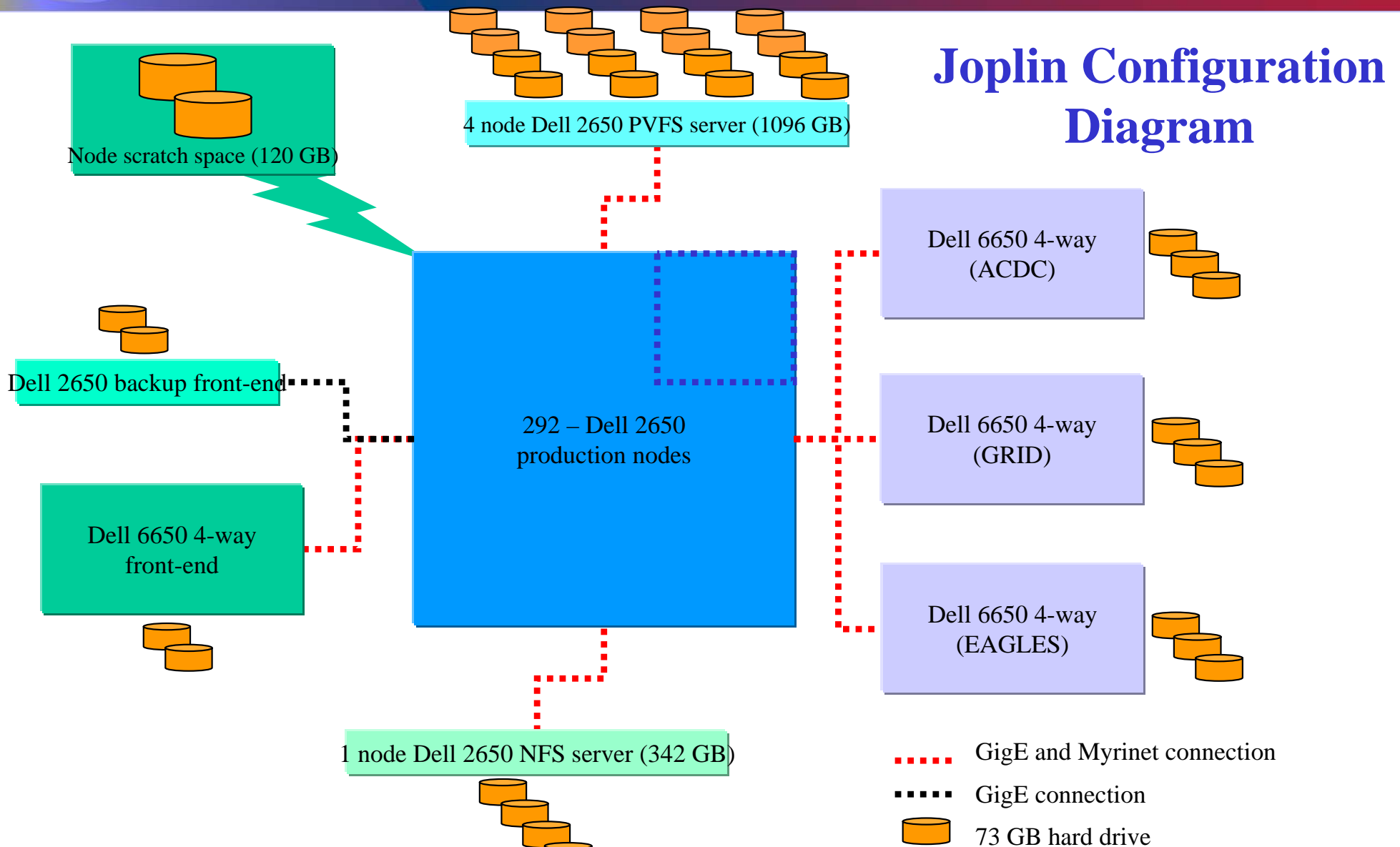
Joplin Configuration Diagram

Node scratch space (120 GB)

4 node Dell 2650 PVFS server (1096 GB)

Dell 6650 4-way (ACDC)

292 – Dell 2650 production nodes

Dell 2650 backup front-end

Dell 6650 4-way front-end

Dell 6650 4-way (GRID)

Dell 6650 4-way (EAGLES)

1 node Dell 2650 NFS server (342 GB)

····· GigE and Myrinet connection

····· GigE connection

73 GB hard drive

University at Buffalo *The State University of New York* **Center for Computational Research** CCR

ACDC-Grid Monitoring: The ACDC-Grid DASHBOARD

http://osg.ccr.buffalo.edu
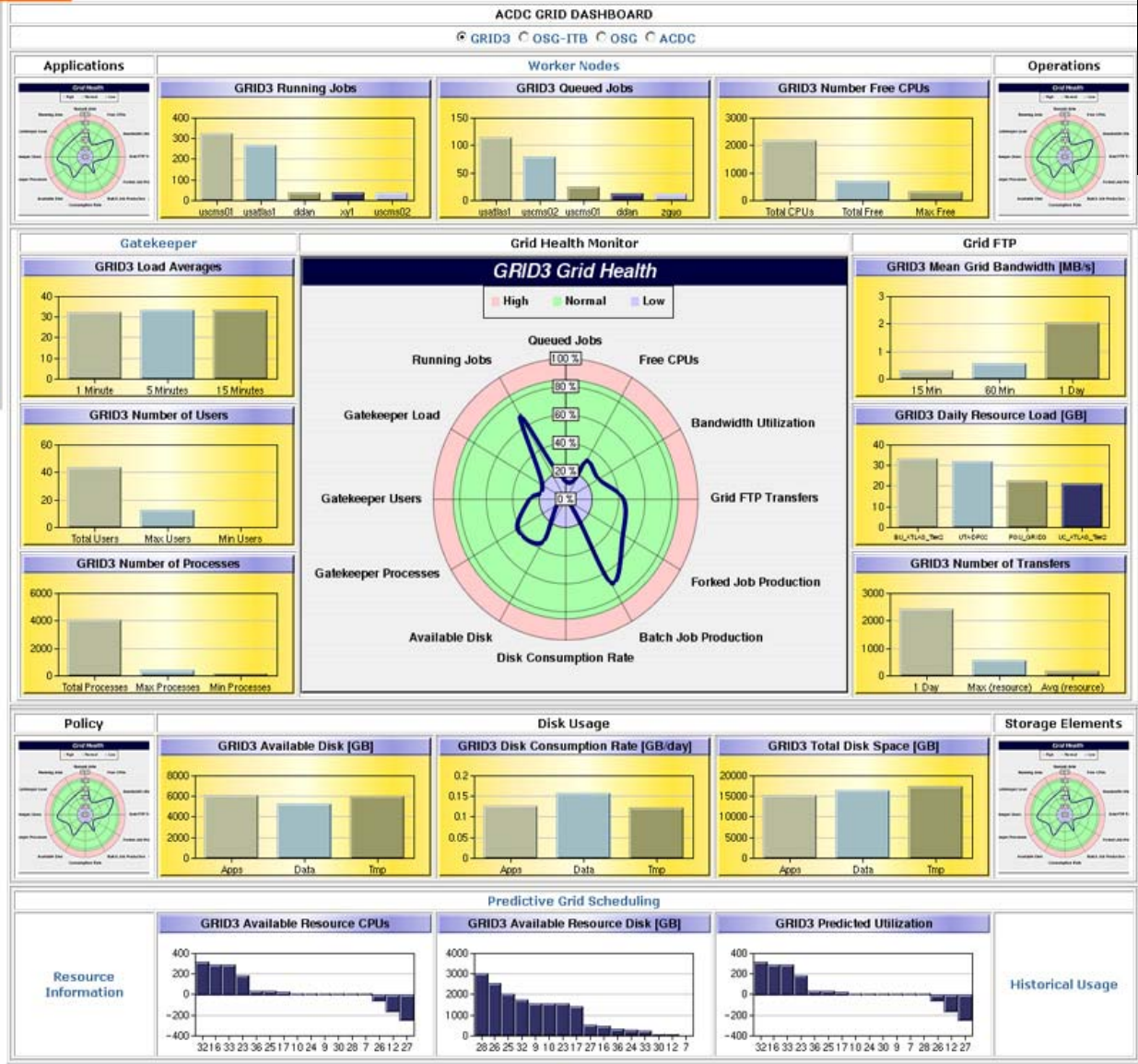
ACDC-Grid Administration

CCR Grid Computing Services: Grid Admin - Microsoft Internet Explorer

File   Edit   View   Favorites   Tools   Help

UB University at Buffalo *The State University of New York*

CCR **Center for Computational Research** GRID PORTAL

High Performance Grid Computing

Grid Site Administration

PORTAL LOGOUT
User Tools
» Manage Account
Grid General Info
Projects
Resources
» Computational Grid
» Job Submission
» Job/Queue Status
» Data Grid
» Data Grid Statistics
» Network Status
» Running/Queued Jobs
» PBS Job History
» Grid Portal Statistics
» Condor Flock Statistics
» User Information
Education/Outreach
Staff Only
CCR HOME

Users
Groups
Portal Event Log
Database Job List

Organizations (add, edit, delete)
Resources (view, refresh, ping, delete, create host certificate)

Globus Administration
Reports (machine usage, user access to machines, etc.)

CCR Grid Computing Services: Grid Admin: Globus - Microsoft Internet Explorer

File   Edit   View   Favorites   Tools   Help

UB University at Buffalo *The State University of New York*

CCR **Center for Computational Research** GRID PORTAL

High Performance Grid Computing

Generate Globus grid-mapfile

Specifying an optional include file will cause the contents of this file to be included at the top of the generated grid-mapfile. If a grid-mapfile path is specified a copy of the generated file will be saved into this location. The generated file will be staged to the grid nodes unless the box is checked.

Optional include file: `/home/griddev/www/grid-mapfile.inc`
Optional grid-mapfile path:

☐ **Do not stage the file to the grid nodes**

Generate   Reset

RTAL LOGOUT
Tools
nage Account
General Info
cts
urces
mputational Grid
b Submission
b/Queue Status
ta Grid
ta Grid Statistics
twork Status
nning/Queued Jobs

CCR Grid Computing Services: Database Job Admin - Microsoft Internet Explorer

File   Edit   View   Favorites   Tools   Help

UB University at Buffalo

CCR **Center for Computational Research** GRID

High Performance

PORTAL LOGOUT
User Tools
» Manage Account
Grid General Info
Projects
Resources
» Computational Grid
» Job Submission
» Job/Queue Status
» Data Grid
» Data Grid Statistics
» Network Status
» Running/Queued Jobs
» PBS Job History
» Grid Portal Statistics
» Condor Flock Statistics
» User Information
Education/Outreach
Staff Only
CCR HOME

**Create New Database Job**

Create a new database job that can be run by the portal. Job scripts must reside in `/home/griddev/www/jobscripts` prior to creating the database job entry.

Job Name:
Full Path To Script:
Accepts Arguments: No
Run Script: No
Run As User: admin (Admin, Admin)

Create Job   Reset

Return to the **Database Job Admin** menu.
Return to the **Grid Admin** menu.

CCR Grid Computing Services: Grid Admin - Resources - Microsoft Internet Explorer

File   Edit   View   Favorites   Tools   Help

UB University at Buffalo *The State University of New York*

CCR **Center for Computational Research** GRID PORTAL

High Performance Grid Computing

MDS Resource Update Status

Current Time: 16-September-2003 10:58:12

PORTAL LOGOUT
User Tools
» Manage Account
Grid General Info
Projects
Resources
» Computational Grid
» Job Submission
» Job/Queue Status
» Data Grid
» Data Grid Statistics
» Network Status
» Running/Queued Jobs
» PBS Job History
» Grid Portal Statistics
» Condor Flock Statistics
» User Information
Education/Outreach
Staff Only
CCR HOME

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

Return to the **Grid Resource Admin** menu.
Return to the **Grid Admin** menu.

Advanced Center for Computational Research Data Center

UB **University at Buffalo**   *The State University of New York*   **Center for Computational Research**   CCR

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**

University at Buffalo *The State University of New York*

CCR
Center for Computational Research GRID PORTAL

High Performance Grid Computing

Advanced Computational Data Center Grid Jobs

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 Submission: This section contains forms for the selection of a grid-enabled application, modification of a application template, grid job definition review and grid job submission.

Grid Job Status: This section contains grid user based 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.

GRID

Advanced
Center for Computational Research
Data
Center

** Development Portal **
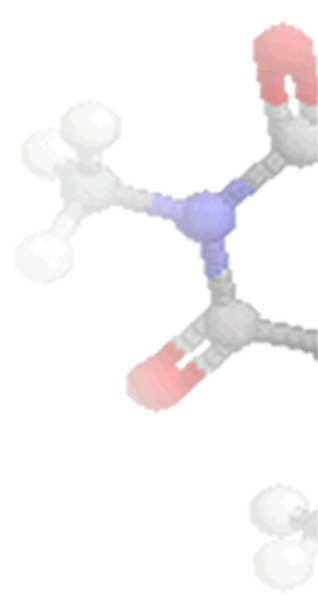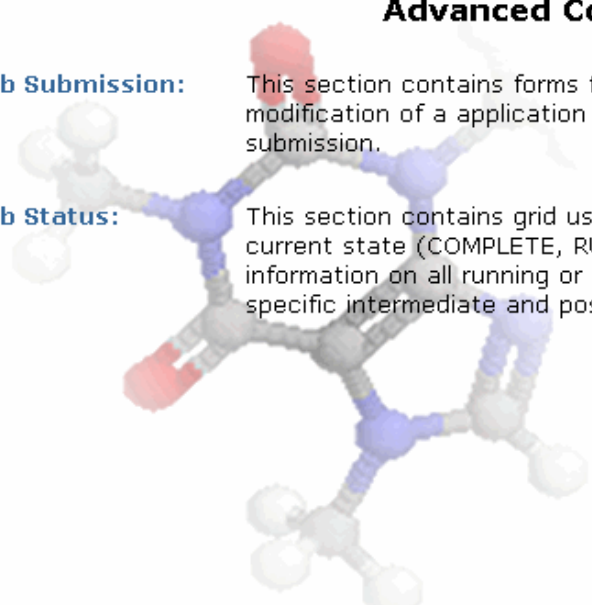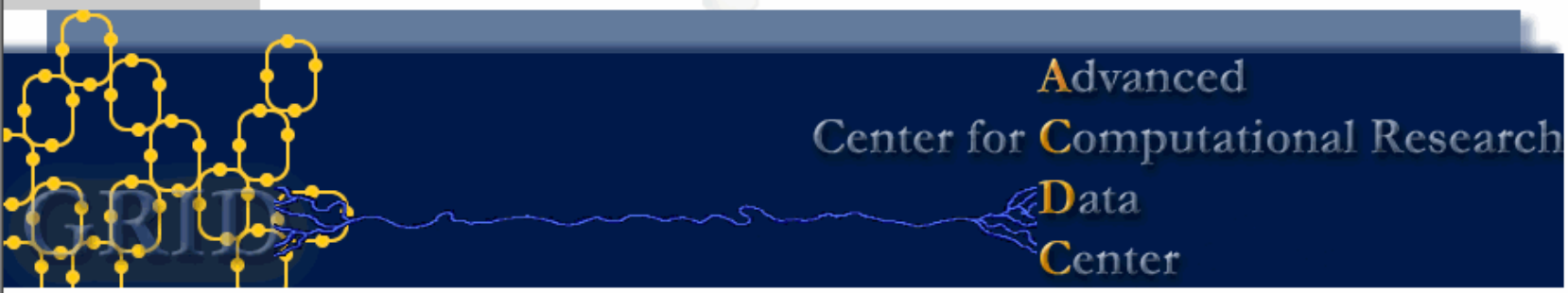
Startup Screen for ACDC-Grid Job Submission
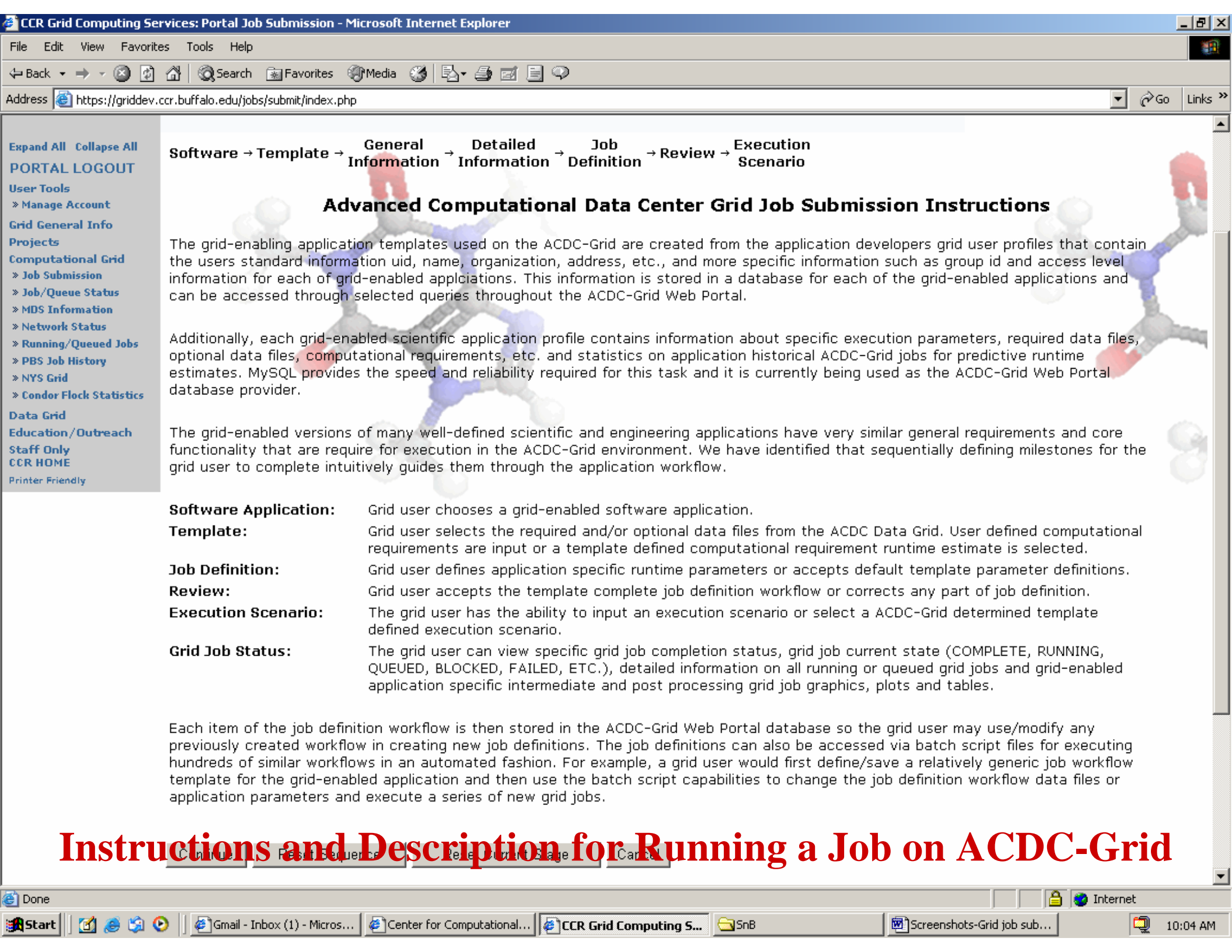
Software → Template → General Information → Detailed Information → Job Definition → Review → Execution Scenario

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

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.

Instructions and Description for Running a Job on ACDC-Grid

File   Edit   View   Favorites   Tools   Help

⇐ Back ▾ → ▾ ⊗ ② ⌂ | ◎ Search ☒ Favorites ⑨ Media ③ | ⬅▾ ☐ ☑ ☐ ♀

Address ☐ https://griddev.ccr.buffalo.edu/jobs/submit/index.php ▾ | ⟳ Go | Links »

UB University at Buffalo *The State University of New York*

CCR
Center for Computational Research GRID PORTAL

High Performance Grid Computing

Expand All Collapse All

PORTAL LOGOUT

User Tools
» Manage Account

Grid General Info

Projects

Computational Grid
» Job Submission
» Job/Queue Status
» MDS Information
» Network Status
» Running/Queued Jobs
» PBS Job History
» NYS Grid
» Condor Flock Statistics

Data Grid

Education/Outreach

Staff Only
CCR HOME

Printer Friendly

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

Select a GAT: [BnP Auto Run ▾]

| BnP Auto Run |
| EADR |
| Ostrich |
| POM |
| Q-Chem |
| SnB |
| SnB DREAR |
| Split |
| snb-dev |

[Continue]   [Reset Current Stage]   [Cancel]

Return to the

Advanced
Center for Computational Research
Data
Center
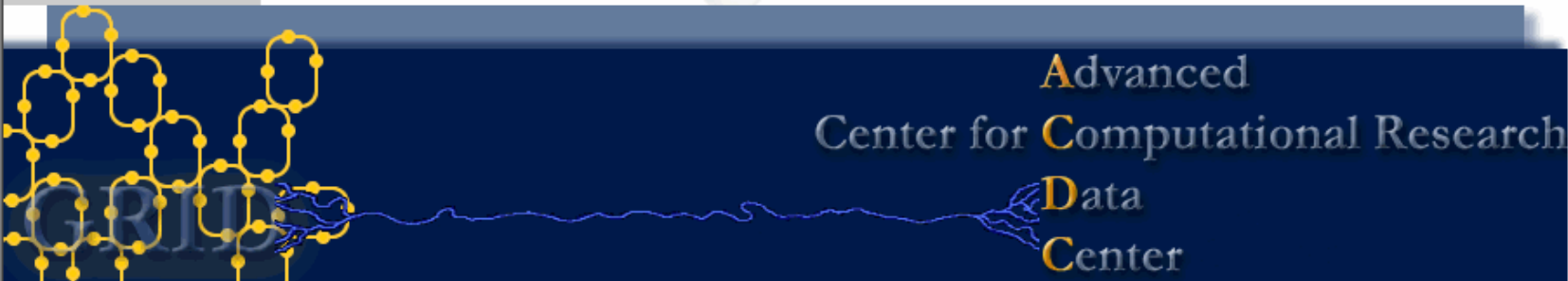
GRID

** Development Portal **

Software Package Selection

Done   🔒 ⊙ Internet

🏁Start | ☐ ⊜ ☺ ◑ | ☐ Gmail - Inbox (1) - Micros... | ☐ Center for Computational... | ☐ CCR Grid Computing S... | ☐ SnB | ☐ Screenshots-Grid job sub... | 🕙 10:05 AM

File   Edit   View   Favorites   Tools   Help

← Back → ⊗ ② ⌂ | ⊘Search ⌾Favorites ⊘Media ③ | ▤▾ ④ ☑ ▤ ◎

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

UB University at Buffalo *The State University of New York*

CCR
Center for Computational Research GRID PORTAL

High Performance Grid Computing

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

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

○ Enter structure definition manually

◉ Select structure from Data Grid: [Select Config File]

[Continue]   [Reset Sequence]   [Reset Current Stage]   [Cancel]

Return to the Grid Job Menu

Advanced
Center for Computational Research
Data
Center

GRID

** Development Portal **

Full Structure / Substructure Template Selection

🔿 Done   🔒 🌐 Internet

🏁Start | 🖼 🅴 🅶 ⊙ | 🅴Gmail - Inbox (1) - Micros... | 🅴Center for Computational... | 🅴CCR Grid Computing S... | 🗀SnB | 🔲Screenshots-Grid job sub... | 🕐 10:05 AM

User Tools
» Manage Account

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

General Information

Structure Information

Title: lled

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)	

Default Parameters Based on Template

🔗 Done   🔒 ⊕ Internet

🏁Start | 🖥 🖉 😊 🗔 ⊙ | 🔗Gmail - Inbox - Mi... | 🔗Center for Comp... | 🔗CCR Grid Comp... | 📁SnB | 🔟Screenshots-Grid ... | 🔗clearwater.ccr.bu... | 🔟 10:16 AM

File   Edit   View   Favorites   Tools   Help

Back   Search   Favorites   Media

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

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]

Default Parameters (cont'd)

Return to the Grid Job Menu

Advanced
Center for Computational Research

Done
Internet

Start   Gmail - Inbox - Mi...   Center for Comp...   CCR Grid Comp...   SnB   Screenshots-Grid ...   clearwater.ccr.bu...   10:16 AM

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

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

Run Normalization

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

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

Number of reflections to use: [   ]

Number of invariants to save: 0

Generate Invariants

[Continue]   [Reset Sequence]   [Reset Current Stage]   [Cancel]

Generating Reflections (Drear)

Done   🔒 ⬤ Internet

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

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 $|F| / \mathrm{sig}(|F|)$ for local scaling: 3.0

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

Run Normalization

Generate Invariants

Data resolution cutoffs ?   Low: 999.0   High: 0.94

Minimum allowed $|E| / \mathrm{sig}(|E|)$: 3.0   Maximum $|E|$: 5.0

Minimum allowed invariants / reflection ratio: 5.0

Initial values for adjustable parameters

Minimum $|E| / \mathrm{sig}(|E|) = \mathrm{ZMin}$: 3.0

Number of reflections to use: 840

Number of invariants to save: 8400

Generate Invariants

Continue   Reset Sequence   Reset Current Stage   Cancel

Invariant Generation

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 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**

- *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)

UB University at Buffalo *The State University of New York*

CCR Center for Computational Research   GRID PORTAL

High Performance Grid Computing

Details for Grid Job 447 - iledhkl

Job Detail Information

Status: RUNNING

Rmin Min: 0.344 Rmin Max: 0.56

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

Total Trials: 1000

Complete Trials: 285

Resource: clearwater.ccr.buffalo.edu

Best Trial Number: 34

Best Trial Rmin: 0.344

Processors: 5

Trial Summary

Grid Job 447 Trial Summary
Number of Trials Complete: 285 (28.5 %)

Walltime Summary

Grid Job 447 Walltime Summary
Walltime Consumed: 2 (0.3%)

Grid Job Trial Histogram

Grid Job 447 Histogram - 285 Trials Complete - Status RUNNING
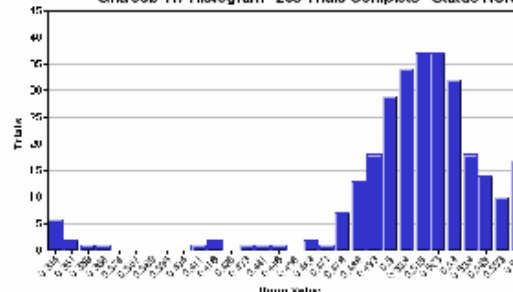
Click on image for enlarged view.

Return to the Grid Job Menu

Graphical Representation of Intermediate Job Status

Advanced

Center for Computational Research

Data

University at Buffalo *The State University of New York*

CCR
Center for Computational Research   GRID PORTAL

High Performance Grid Computing

Expand All   Collapse All

PORTAL LOGOUT

User Tools
» Manage Account

Grid General Info
Projects

Computational Grid
» Job Submission
» Job/Queue Status
» MDS Information
» Network Status
» Running/Queued Jobs
» PBS Job History
» NYS Grid
» Condor Flock Statistics

Data Grid
Education/Outreach
Staff Only
CCR HOME
Printer Friendly

Grid Job 447 Histogram - 285 Trials Complete - Status RUNNING
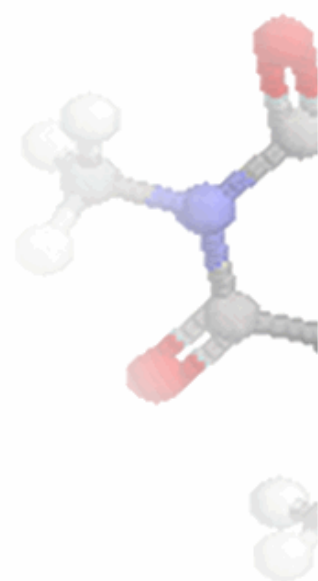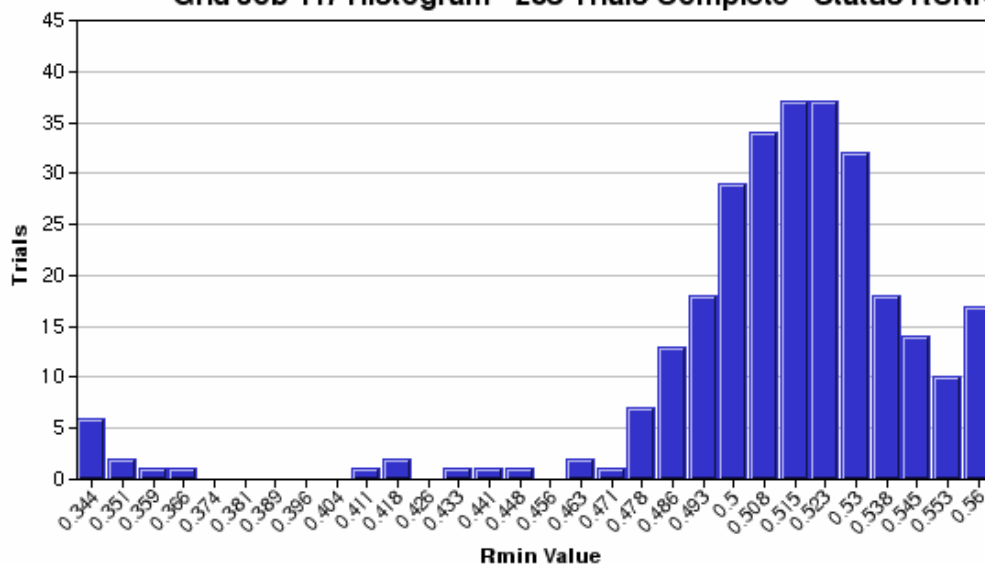
Return to Job Details
Return to the Grid Job Summary
Return to the Grid Job Menu

Histogram of Completed Trial Structures

Advanced
Center for Computational Research

Status of Jobs

Heterogeneous Back-End Interactive Collaboratory
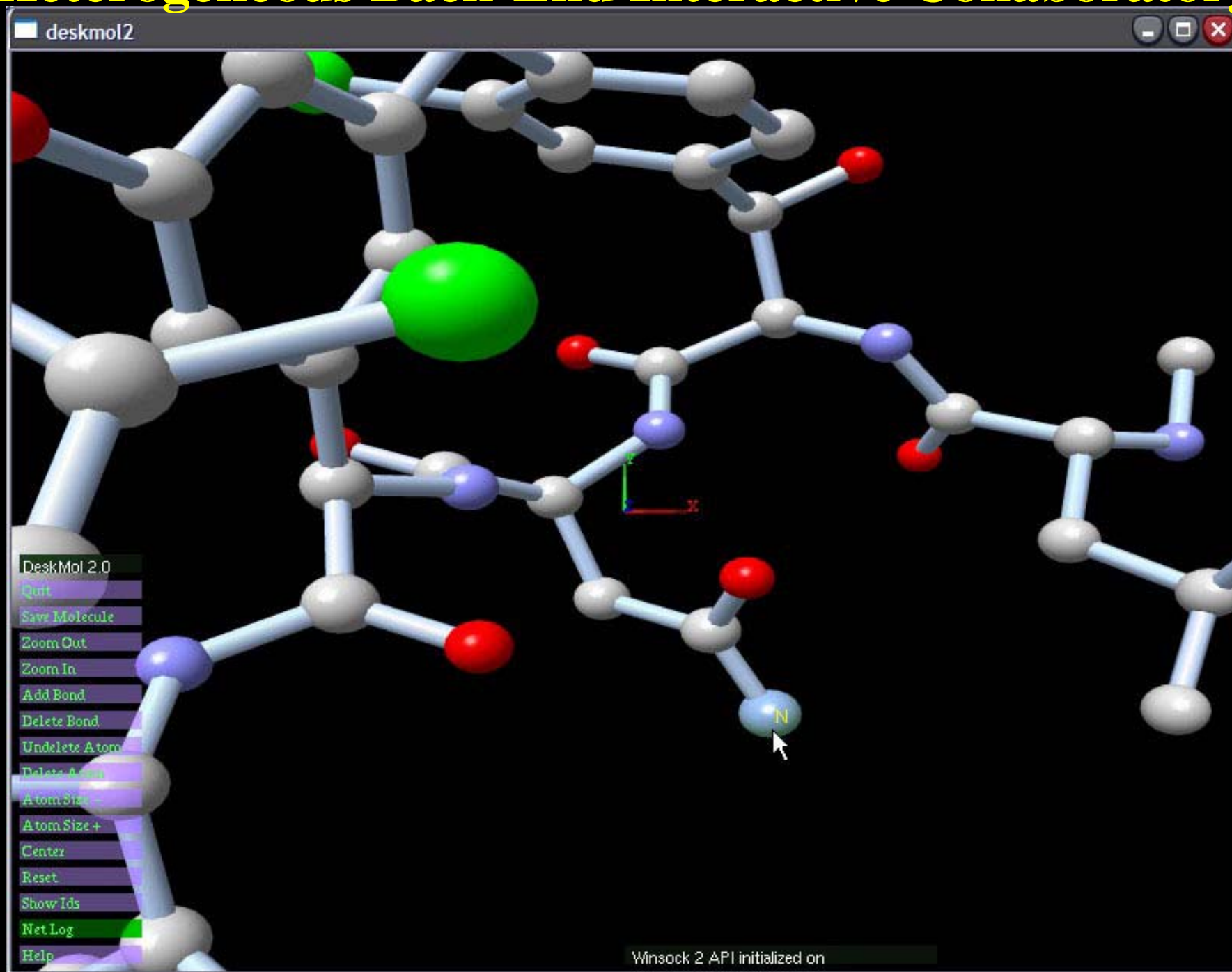
User starts up – default image of structure.
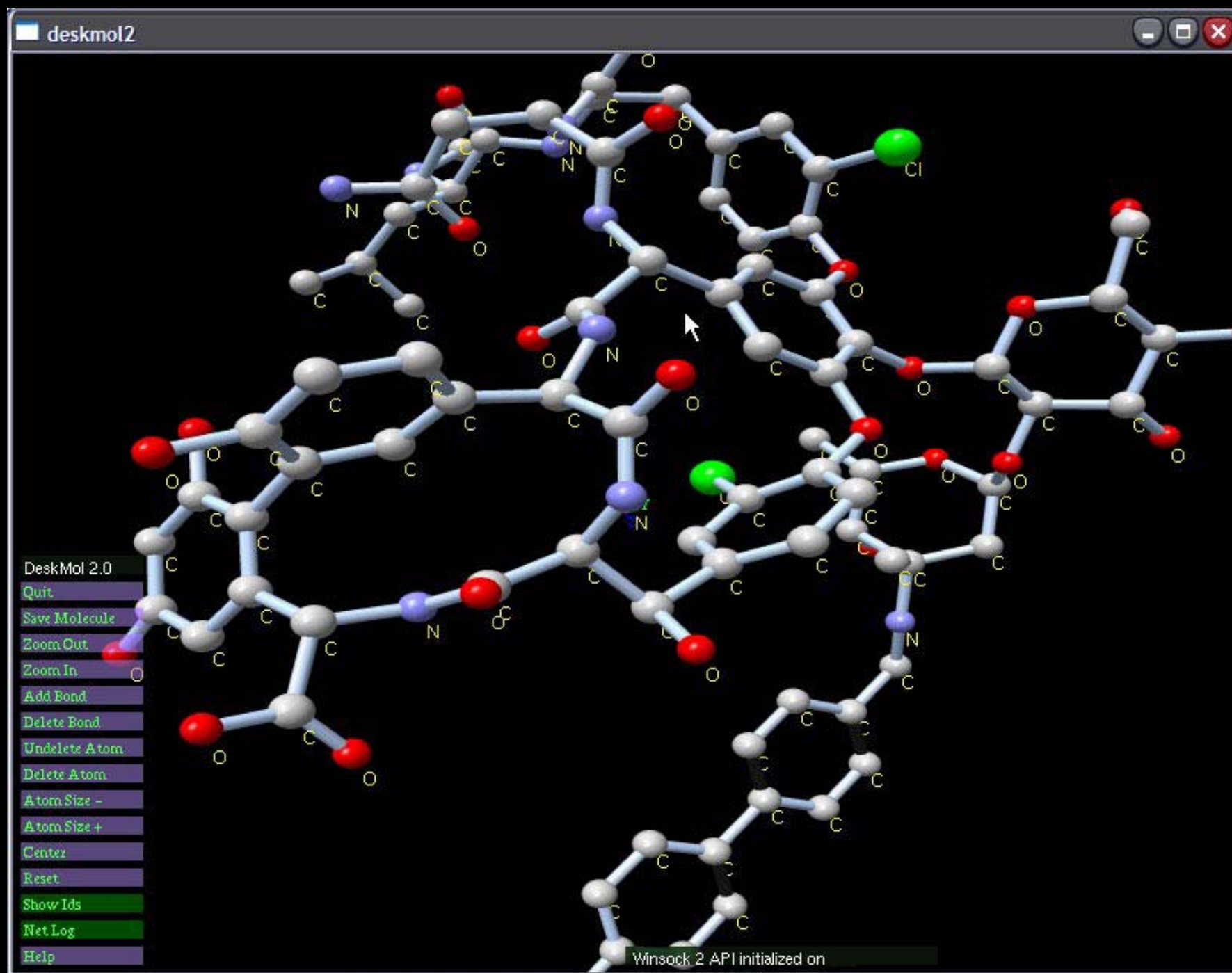
Molecule scaled, rotated, and labeled.

Current Grid Efforts

- **Grass Roots NYS Grid**
  - SUNY-Buffalo
  - SUNY-Binghamton
  - SUNY-Albany
  - SUNY-Geneseo
  - Hauptman-Woodward Inst.
  - Columbia
  - Niagara University
  - Canisius College

- **Harden**
  - Dashboard
  - Predictive Scheduler

- **GRASE VO: Grid Resources for Advanced Science and Engineering Virtual Organization**
  - (Non-Physics Research)
  - Structural Biology
  - Groundwater Modeling
  - Earthquake Engineering
  - Computational Chemistry
  - GIS/BioHazards

Outreach

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

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

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

- Bruce Holm
- Janet Penksa

- NSF, NIH, NYS, NIMA, NTA, Oishei, Wendt, DOE

www.ccr.buffalo.edu