Grid-Based Research, Development, and Deployment in New York State

R. Miller, J.J. Bednasz, S. M. Gallo, C.L. Ruby: SUNY-Buffalo
K. Chiu, M. Govindaraju, M. Lewis: SUNY-Binghamton
C.M. Weeks: Hauptman-Woodward Institute

Russ Miller
Director, CI Laboratory
Dept of Comp Sci & Eng, SUNY-Buffalo
Hauptman-Woodward Inst.

NSF, NIH, DOE, NIMA, NYS, HP
www.cse.buffalo.edu/faculty/miller/CI/

GRID
Advanced
Center for Computational Research
Data
Center

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 (1.25 TF; 140 Cores; 22TB)
  - Dell Clusters (Intel Head Nodes; Workers: Intel 8×2 ×4 and AMD 8×2×2)
  - Virtual Memory Machines (Intel 2×4×4)
  - GigE, Managed Switches, InfiniBand
  - 22 TB Dell Storage (2)
  - Condor Flock (35 Intel/AMD)
- Production Equipment
  - Dell Workstations
  - Dell 15 TB Storage
  - Access to CCR equipment (13TF Dell/Intel clusters)

Evolution of CI Lab Projects

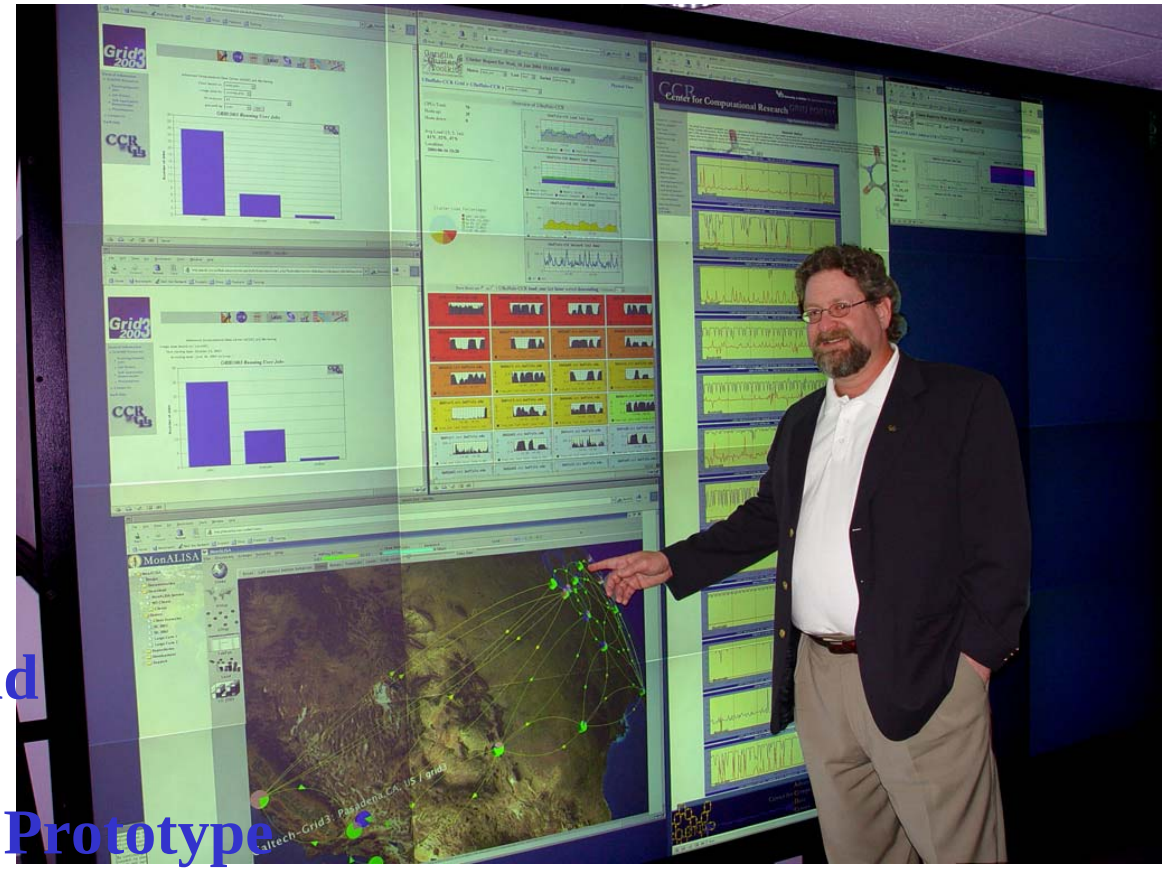
- **ACDC-Grid**
  - **Experimental Grid: Globus & Condor**
  - **Integrate Data & Compute, Monitor, Portal, Node Swapping, Predictive Scheduling/Resource Management**
  - **GRASE VO: Structural Biology, Groundwater Modeling, Earthquake Eng, Comp Chemistry, GIS/BioHazards**
  - **Buffalo, Buffalo State, 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

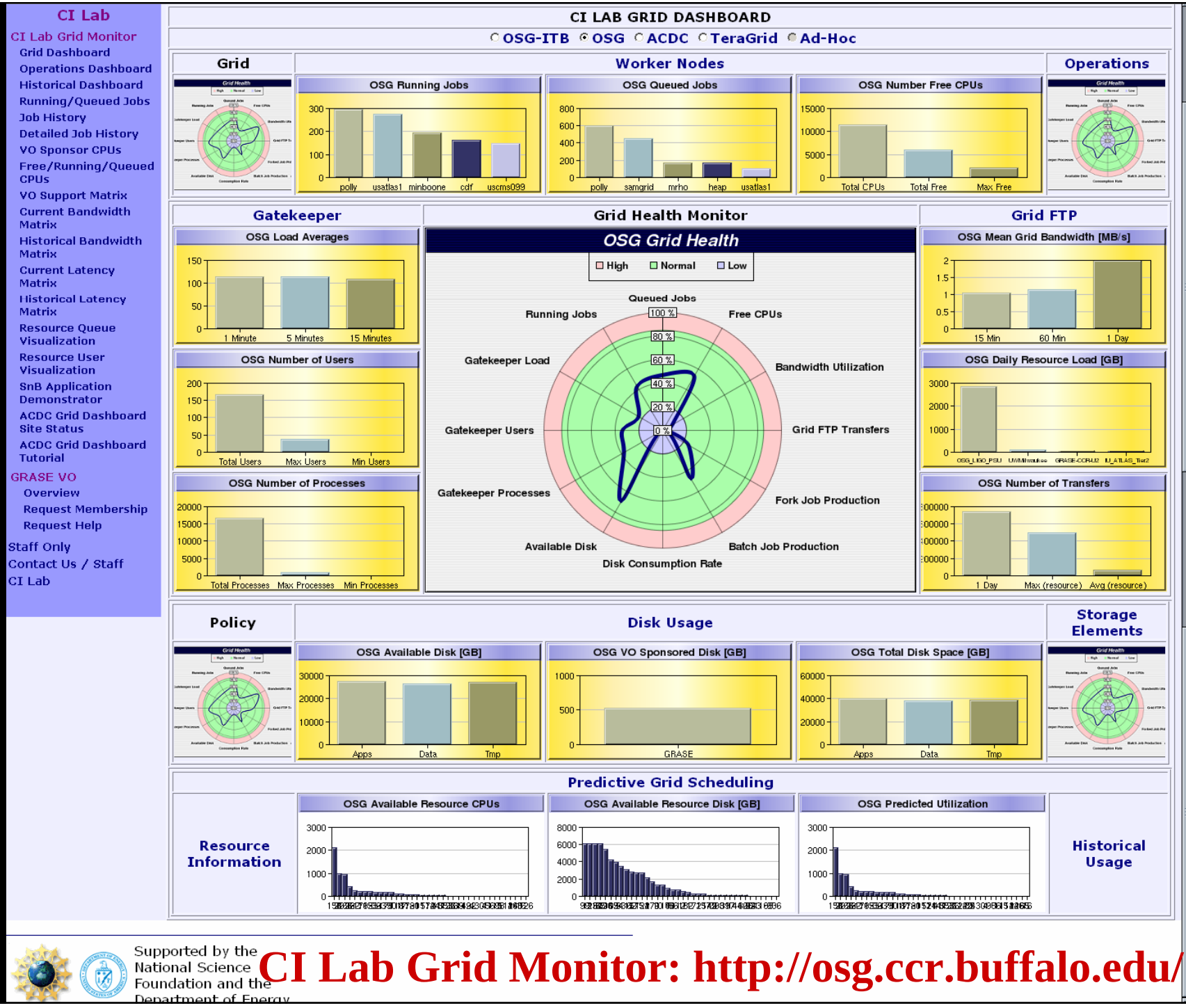
CI Lab Collaborations

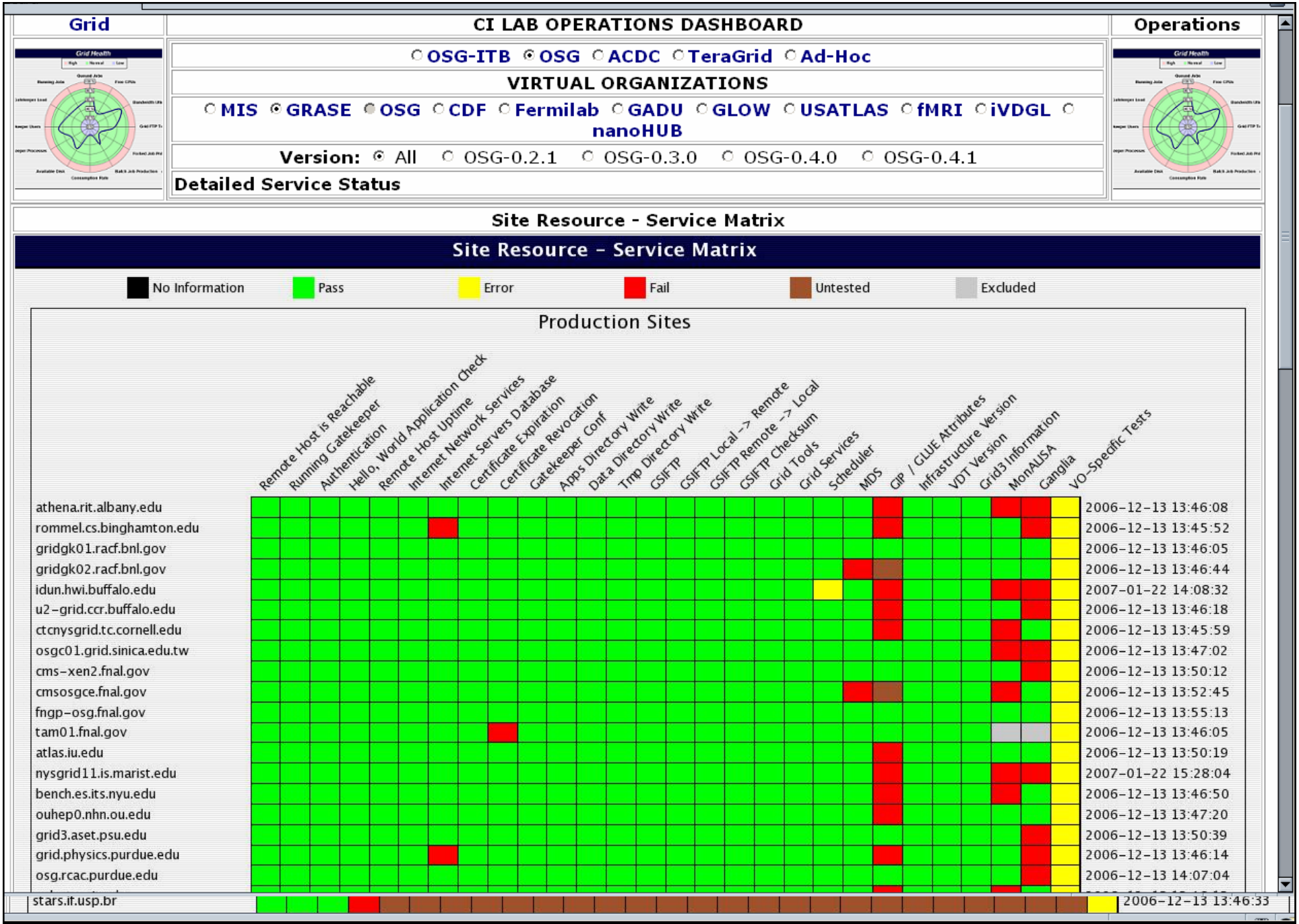
- High-Performance Networking Infrastructure
- Grid3+ Collaboration
- iVDGL Member
  - Only External Member
- Open Science Grid
  - GRASE VO
- NYS CI Initiative
  - Executive Director
  - Various WGs
- Grid-Lite: Campus Grid
  - HP Labs Collaboration
- Innovative Laboratory Prototype
  - Dell Collaboration



CI Lab Projects

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


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



CI Lab Operations Dashboard

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**

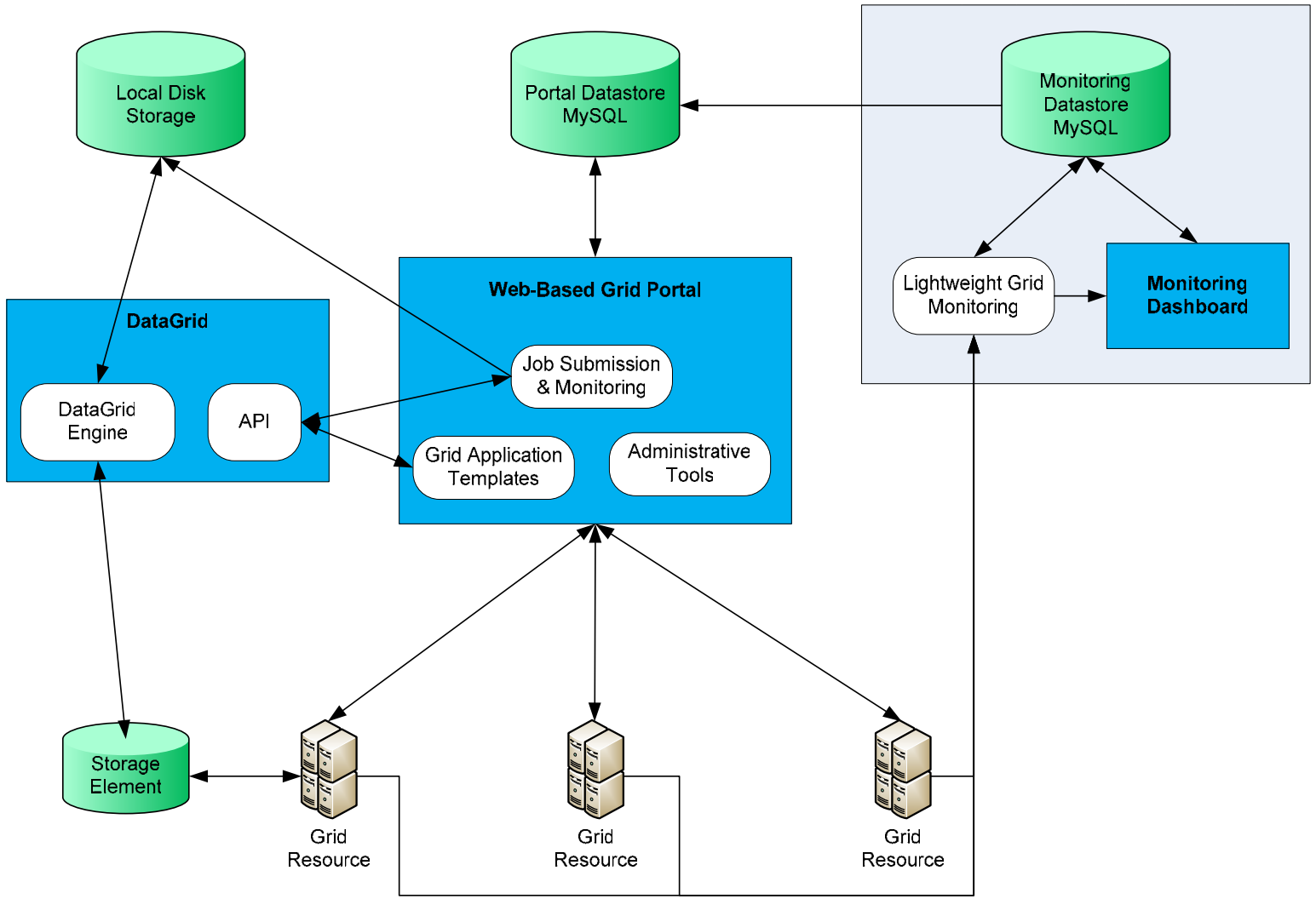
ACDC-Grid Data Grid Functionality

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

Grid-Enabling Application Templates (GATs)

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

NYS Grid Portal



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

Mail Home Netscape

New Tab | CCR Grid Computing Services:

Cyberinfrastructure Laboratory

Grid Portal

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

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

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

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

Contact Us / Staff
CI Lab
Staff Only

Welcome to the Cyberinfrastructure Laboratory Grid Portal

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

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

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

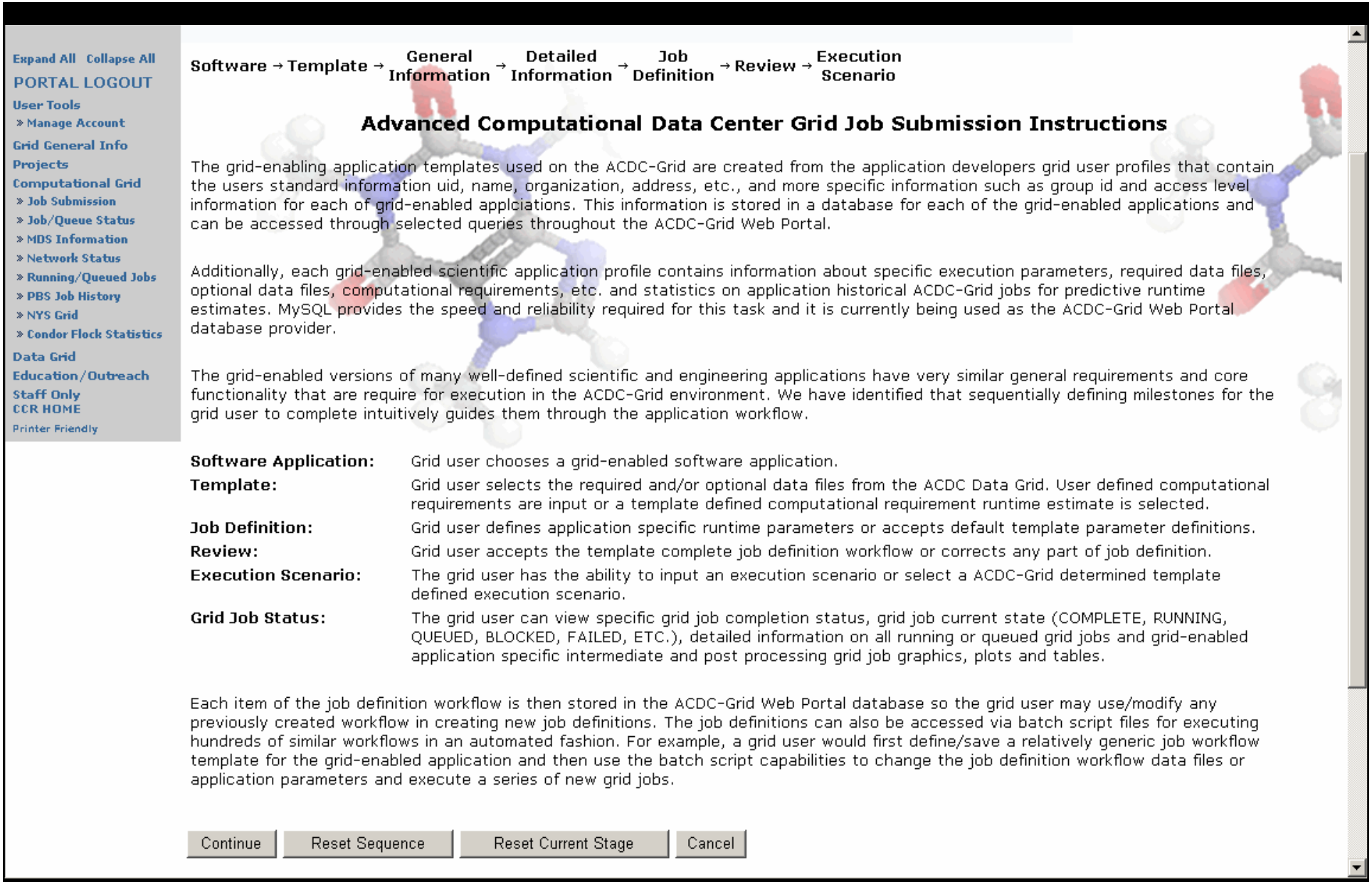
The grids developed by the CI Lab support teaching and research activities, as well as providing infrastructure that includes high-end data, computing, imaging, grid-enabled software, all of which relies on the New York State Research Network (**NYSERNet**).

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

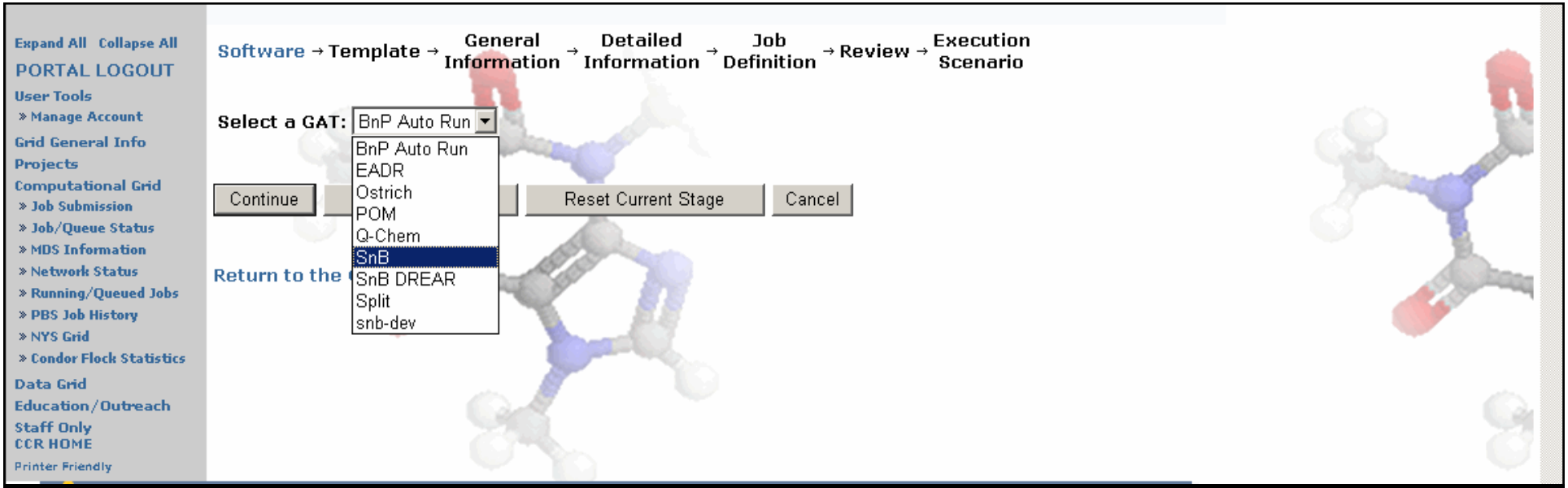
BnP

Software : BnP
Field : Protein crystal structure determination

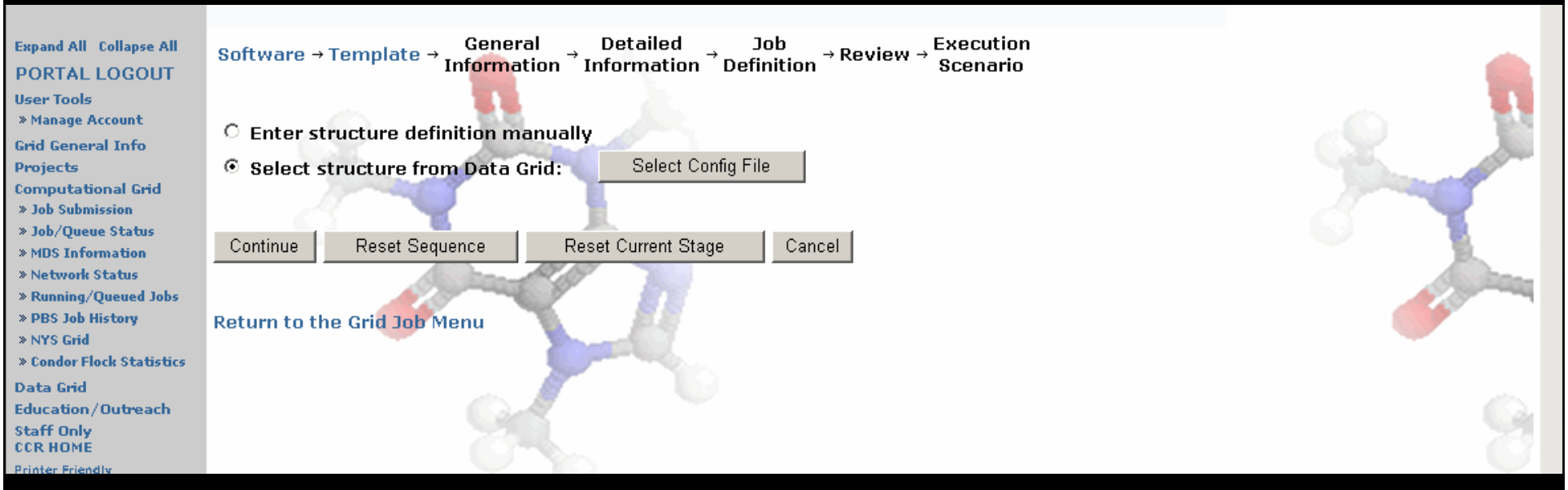
Startup Screen for CI Lab Grid Job Submission



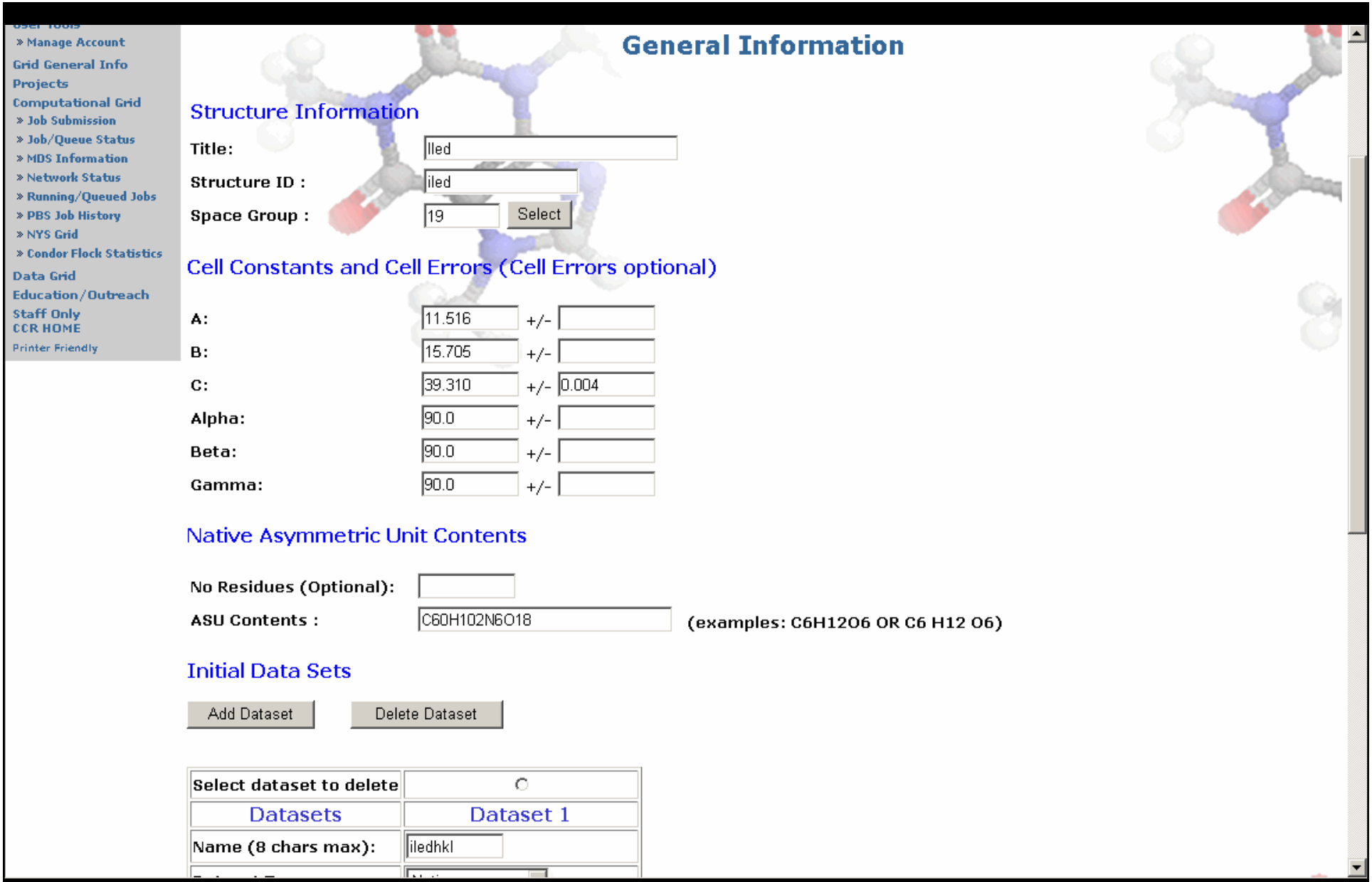
Instructions and Description for Running a Job on ACDC-Grid



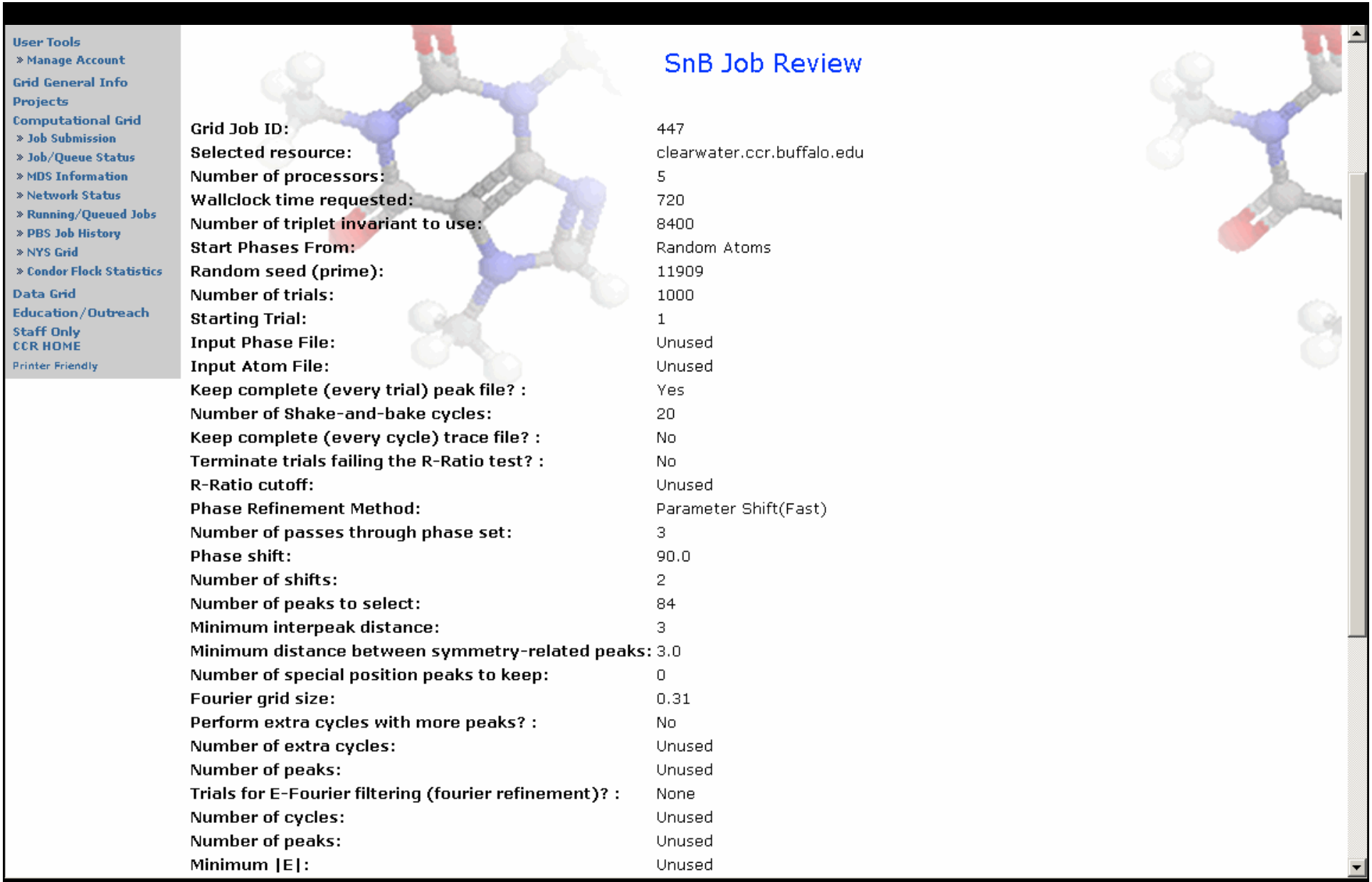
Software Package Selection



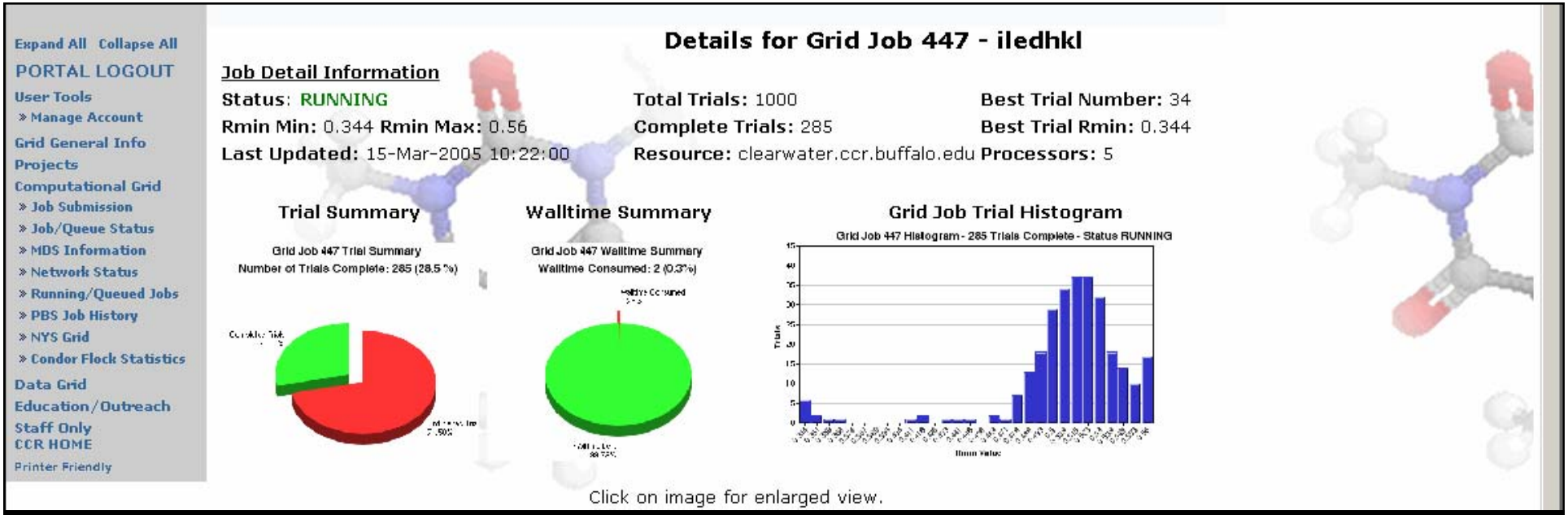
Full Structure / Substructure Template Selection



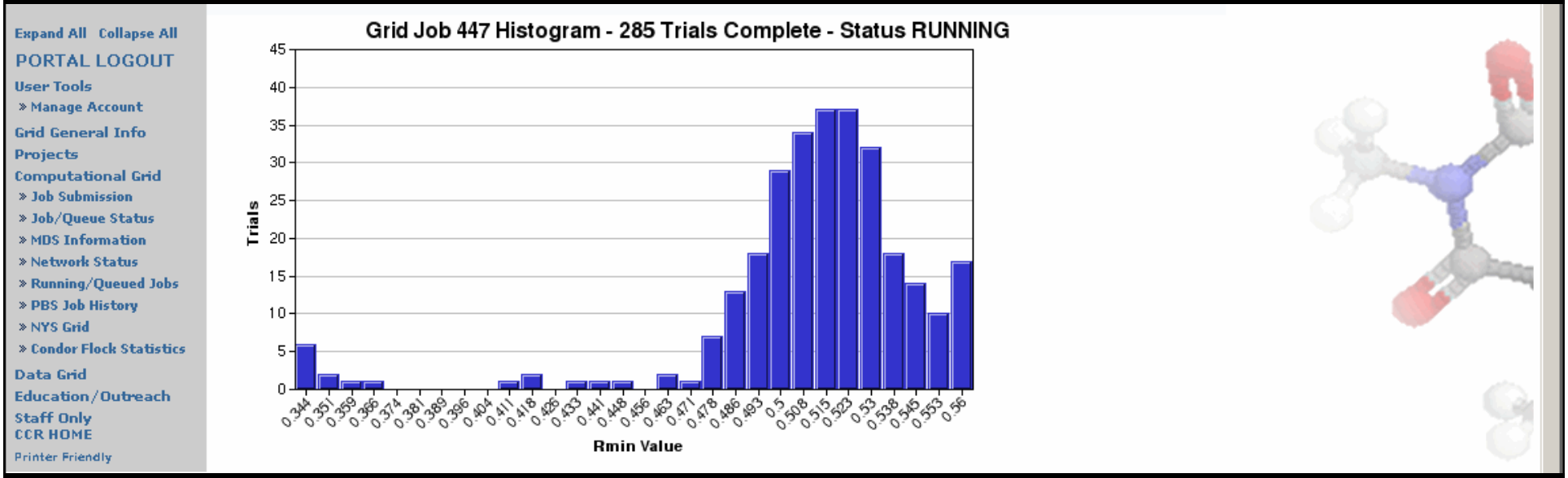
Default Parameters Based on Template



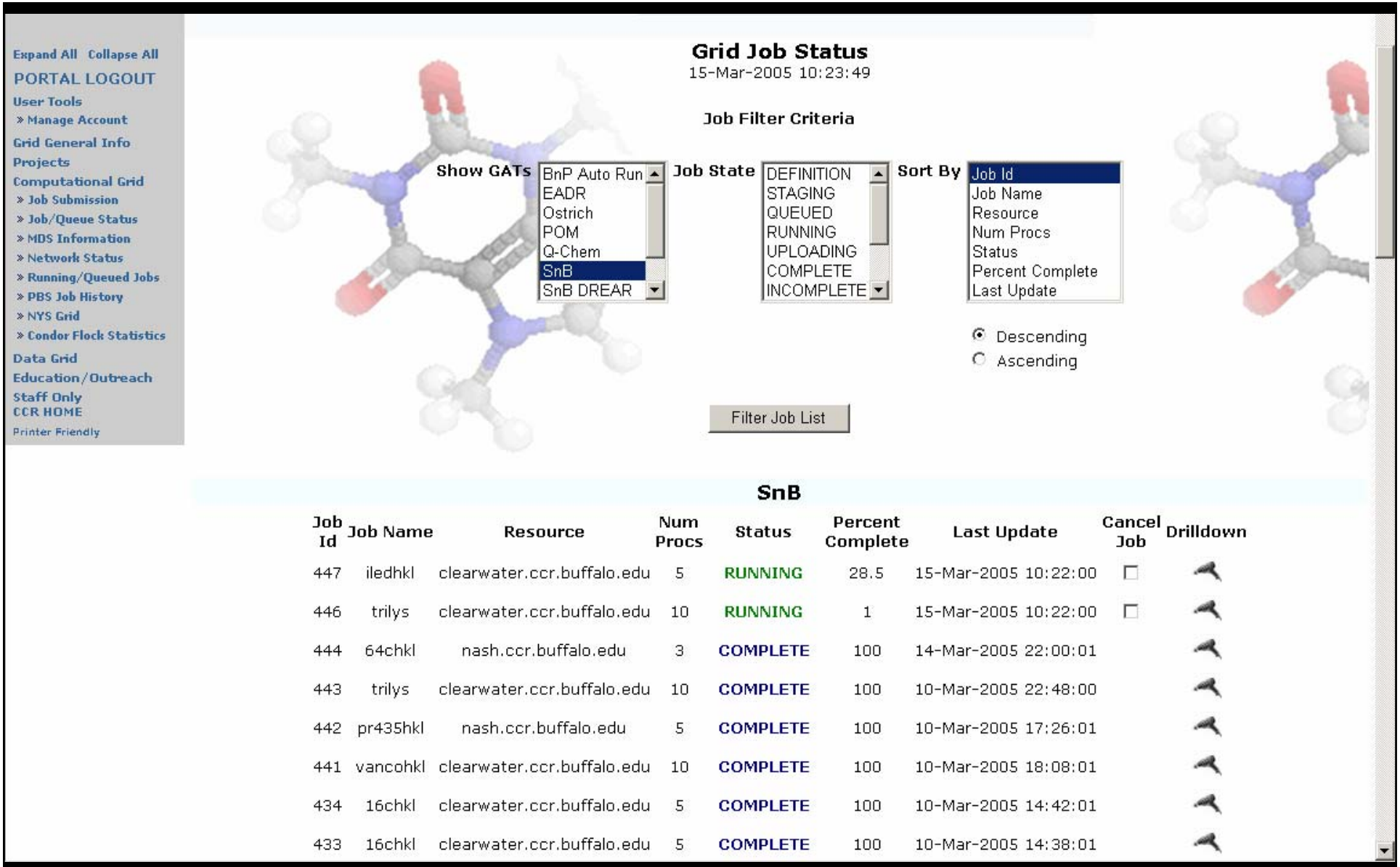
SnB Review (Grid job ID: 447)



Graphical Representation of Intermediate Job Status

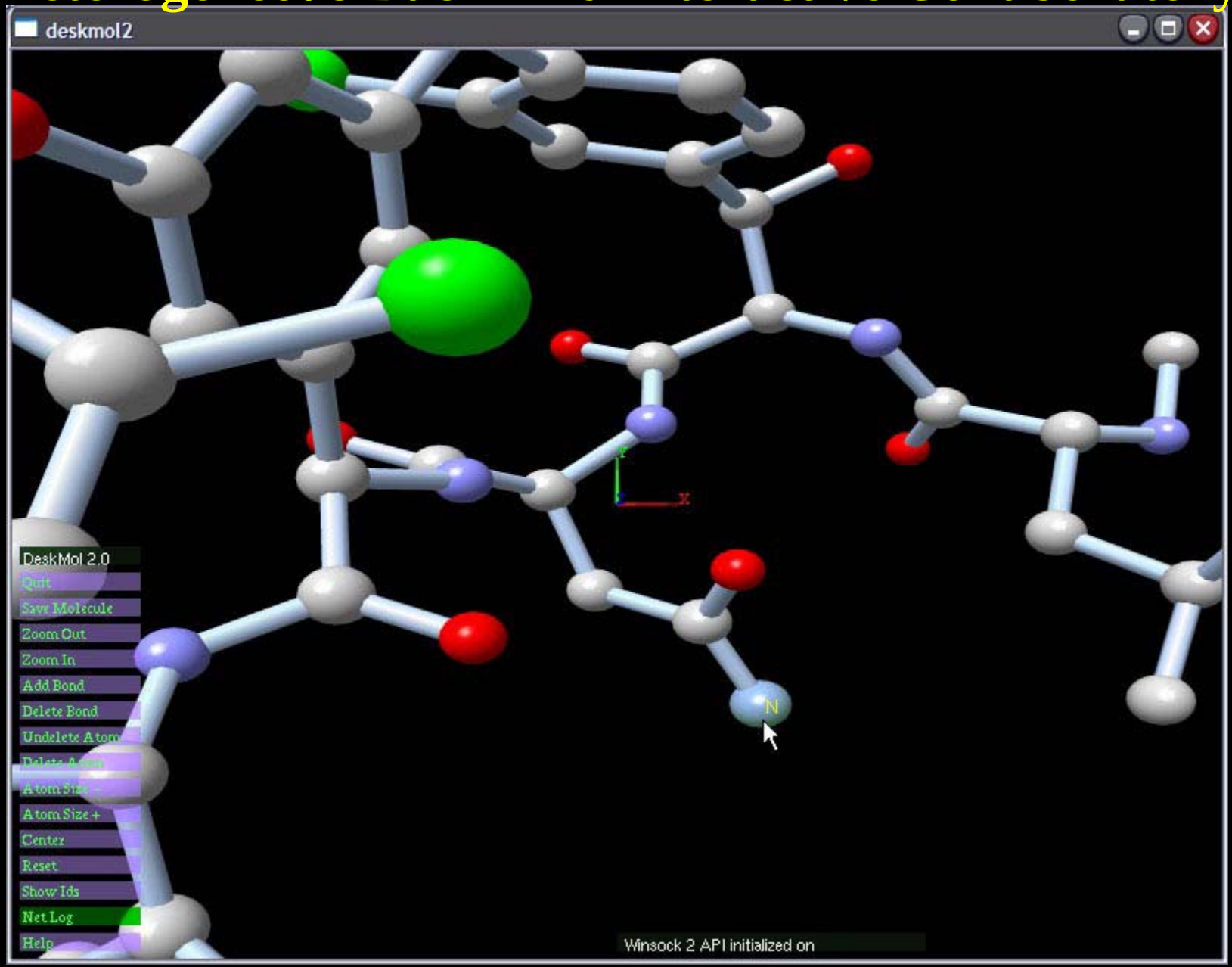


Histogram of Completed Trial Structures

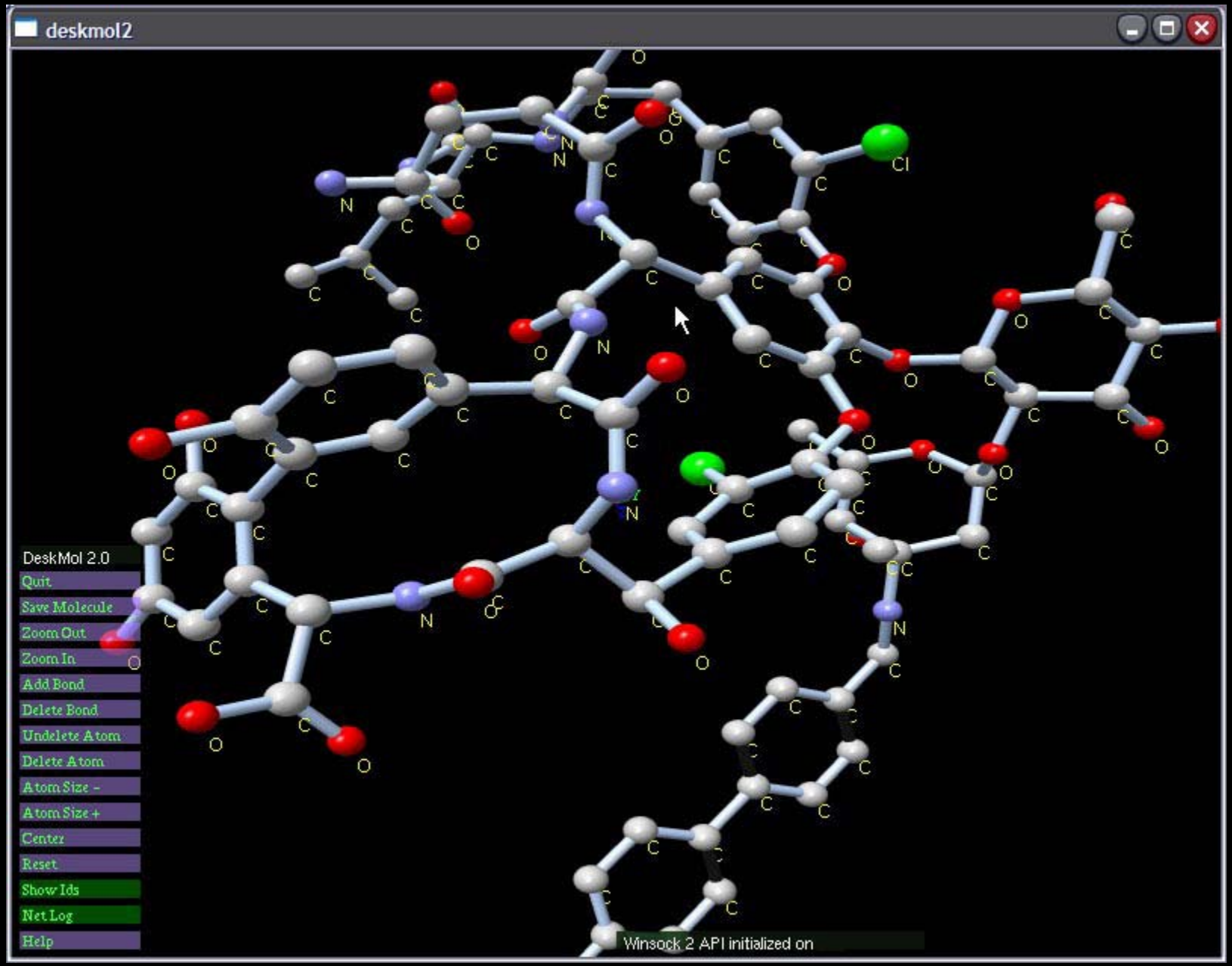


Status of Jobs

Heterogeneous Back-End Interactive Collaboratory



User starts up – default image of structure.



Molecule scaled, rotated, and labeled.

Binghamton University
Grid Computing Research Laboratory

- **Drs. Kenneth Chiu, Madhu Govindaraju, and Michael Lewis.**
- **Techniques for Web and grid service performance optimization**
- **Component frameworks for grids**
- **Instruments and sensors for grid environments**
- **Adaptive information dissemination protocols across grid overlays**
- **Emulation framework for grid computation on multi-core processors**
- **Secure grid data transfer**
- **www.grid.cs.binghamton.edu/**

New York State Grass Roots Cyberinfrastructure Initiative

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

TRUN: Transborder Research University Network

- Ontario: York, Toronto, Western Ontario, McMaster, Queen's, Waterloo, Guelph
- NYS: Buffalo, Rochester, Syracuse, Cornell, Albany, RIT
- Mission Statement: Expand and support cooperation among research universities in the border region of Province of Ontario and NYS:
  - Collaborative/consortial research
  - Joint applications for external funding
  - Cooperative academic programs
  - Faculty and student exchanges
  - Shared facilities
  - Joint conferences, symposia, workshops
- Current Focus
  - Great Lakes Sustainable Energy
  - IT-Supported Disciplinary Research
  - High Performance Computing
  - Canada-U.S. Policy and Standardization of Binational Data
- General Issues
  - Public Policy Issues, Regional Governance
  - Border Security and Mobility
  - Economic and Worforce Development
  - University Partnerships with Government and Industry
  - Health Care and Policy
  - Basic Research and Technology Transfer

www.trun.ca

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