Grid Computing in New York State

J. J. Bednasz, S. M. Gallo, R. Miller, C.L. Ruby, and C.M. Weeks

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

NSF, NIH, DOE, NIMA, NYS, HP

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



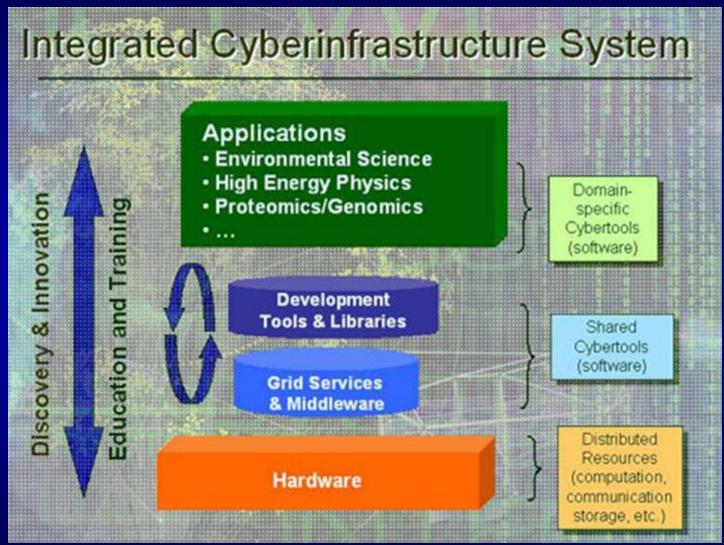


Cyberinfrastructure

- 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



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."

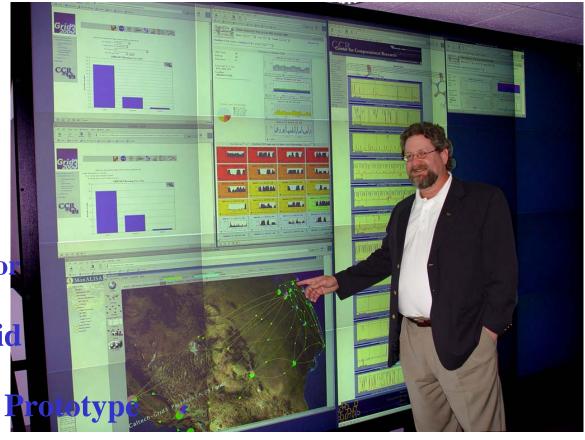
Cyberinfrastructure Lab

- Premise
 - ☐ CI sits at core of modern simulation & modeling
 - ☐ CI allows for new methods of investigation to address previously unsolvable problems
- Focus
 - **□** Algorithms
 - **□** Portals
 - **☐** Interfaces
 - **■** Middleware
- **■** Goal
 - ☐ Free end-users to do disciplinary work
- Funding (2001-pres): NSF ITR, NSF CRI, NSF MRI, NYS, Fed



CI Lab Collaborations

- **High-Performance Networking Infrastructure**
- Grid3+ Collaboration
- iVDGL Member
 - **□** Only External Member
- Open Science Grid
 - **□** GRASE VO
- NYSGrid.org
 - **NYS CI Initiative**
 - **☐** Fndg Executive Director
 - **□** Various WGs
- Grid-Lite: Campus Grid
 - **☐** HP Labs Collaboration
- Innovative Laboratory 1
 - **□** Dell Collaboration





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

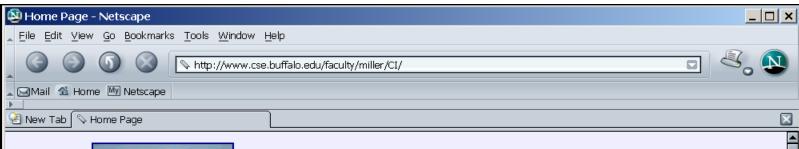
- **■** Extension to Hardened Production-Level System State-Wide
- ☐ Albany, Binghamton, Buffalo, Geneseo, Canisius, Columbia, HWI, Niagara, [Cornell, NYU, RIT, Rochester, Syracuse, Marist], {Stony Brook, RPI, Iona}



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







Cyberinfrastructure Laboratory

Ubiquitous High-End Computing, Data, Networking, & Visualization

Dr. Russ Miller

UB Distinguished Professor of Computer Science & Engineering

CI Lab

Biography

Research

- Overview
- Papers
- Presentations
- CI Lab
 - Overview
 - · Projects
 - Trojec
 - ∘ <u>News</u>
 - Personnel
 - · NYS Grid
 - Collaborations
 - Monitor
 - Portal
- SnB
- NYSGrid.org
- CCR

Teaching

Personal Info

<u>Utilities</u>

Contact Info

Home

Introduction

The Cyberinstitute of the State of New York (CSNY) was approved in June of 2006 by the Center of Excellence in Bioinformatics and Life Sciences. Delays in the announcement of CSNY and its establishment have led to the creation of the CyberInfrastructure lab (CI). CI projects are currently supported by an NSF ITR grant, an NSF CRI grant, and the Center for Computational Research.

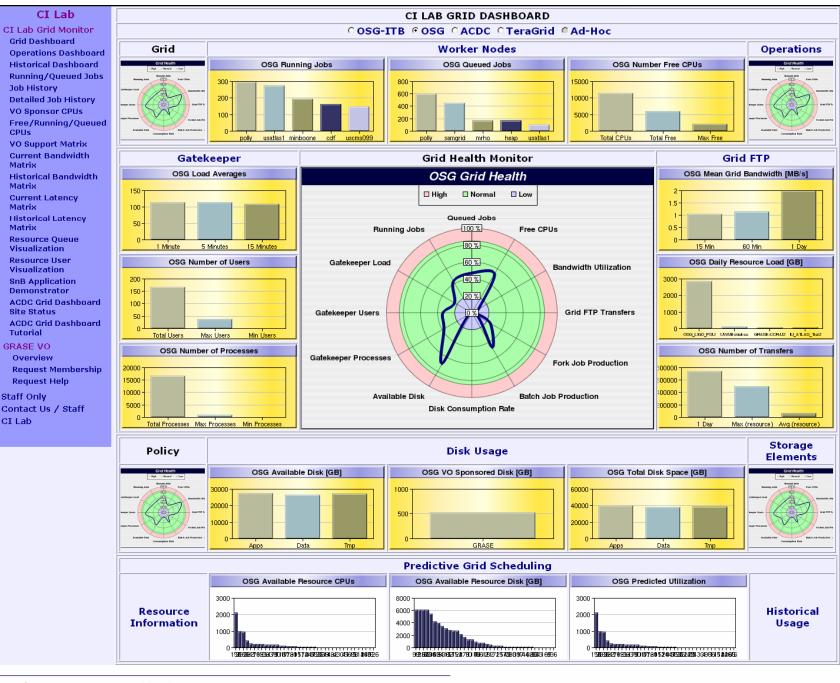
In the 21st century, leading academic institutions will embrace our digital data-driven society and empower students to compete in this knowledge-based economy. In order to support research, scholarship, education, and community outreach, CI has been established to integrate research in disciplinary domains, including science, engineering, and biomedicine, with research in enabling technologies and interfaces. This 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.

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 coftware algorithms, postals, and interfaces that will enable research and

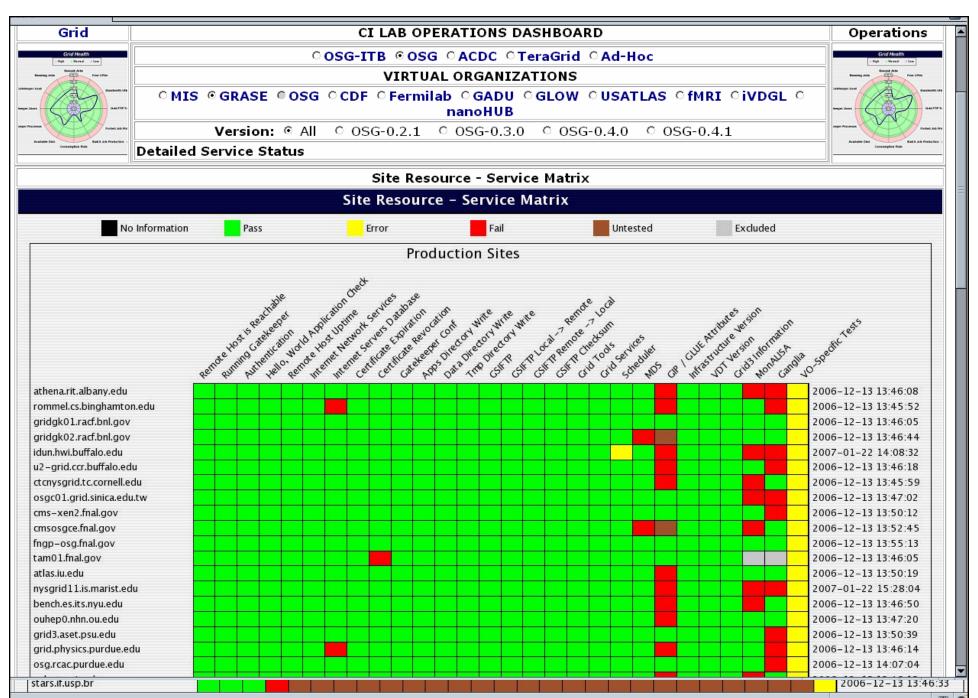
CI Lab Projects

- 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









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 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 Data Grid Functionality

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



Grid-Enabling Application Templates (GATs)

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



Grid Enabled SnB

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



Required Grid Services

- **■** Application Layer
 - □ Shake-and-Bake
 - **□** Apache web server
 - **■** MySQL database
- **High-level Services**
 - ☐ Globus, PHP, Fortran, C
- Core Services
 - Metacomputing Director Service, Globus Security Interface, GRAM, GASS
- Local Services
 - ☐ Condor, MPI, PBS, Maui, WINNT, IRIX, Solaris, RedHat Linux

Grid Implementation as a Layered Set of Services

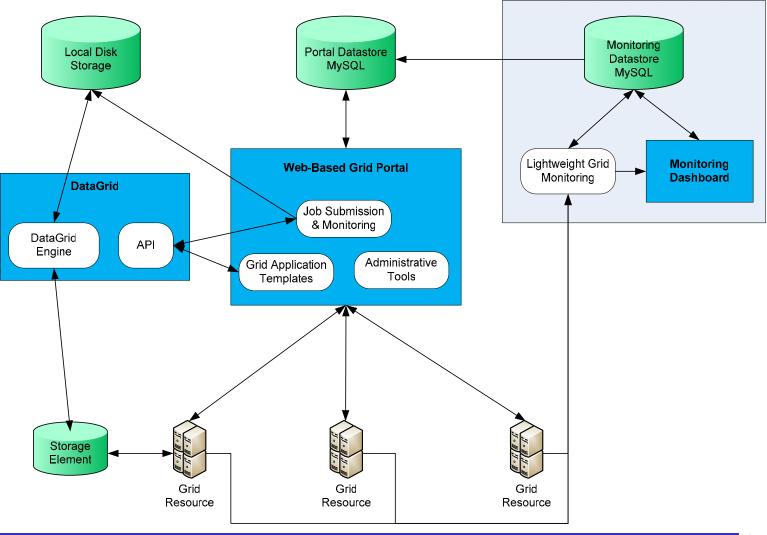








NYS Grid Portal





CI Lab



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

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

GAT/Resource Matrix

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

Tools

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 Detailed Job → Review → Execution Scenario

Advanced Computational Data Center Grid Job Submission Instructions

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

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

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

Software Application: Grid user chooses a grid-enabled software application.

Template: Grid user selects the required and/or optional data files from the ACDC Data Grid. User defined computational

requirements are input or a template defined computational requirement runtime estimate is selected.

Job Definition: Grid user defines application specific runtime parameters or accepts default template parameter definitions.

Review: Grid user accepts the template complete job definition workflow or corrects any part of job definition.

Execution Scenario: The grid user has the ability to input an execution scenario or select a ACDC-Grid determined template

defined execution scenario.

Grid Job Status: The grid user can view specific grid job completion status, grid job current state (COMPLETE, RUNNING,

QUEUED, BLOCKED, FAILED, ETC.), detailed information on all running or queued grid jobs and grid-enabled

application specific intermediate and post processing grid job graphics, plots and tables.

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

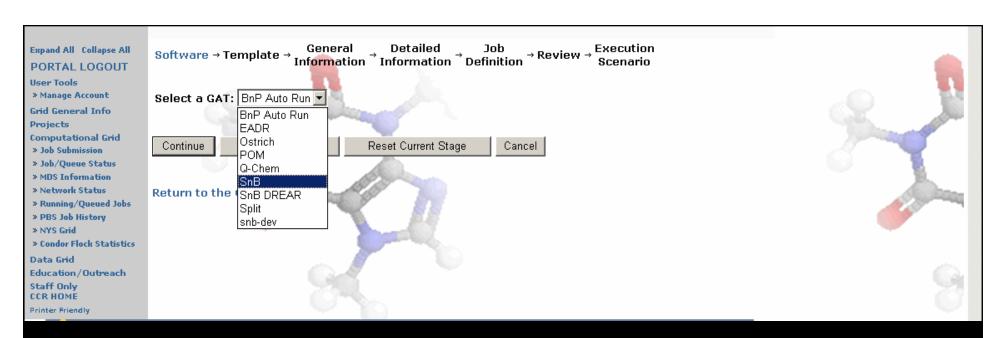
Continue

Reset Sequence

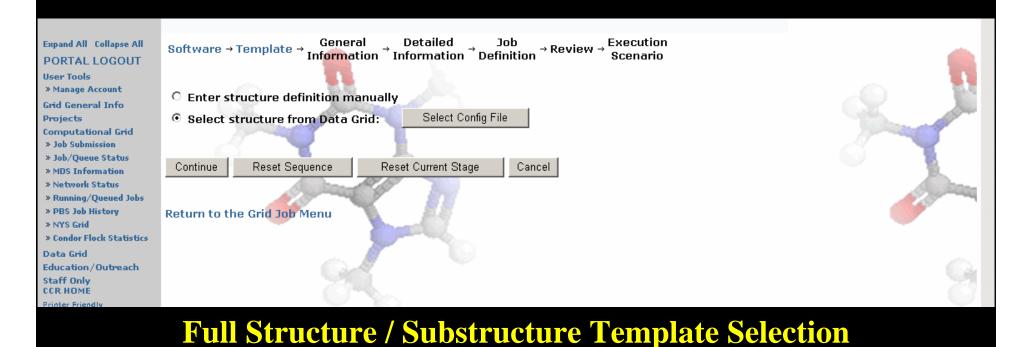
Reset Current Stage

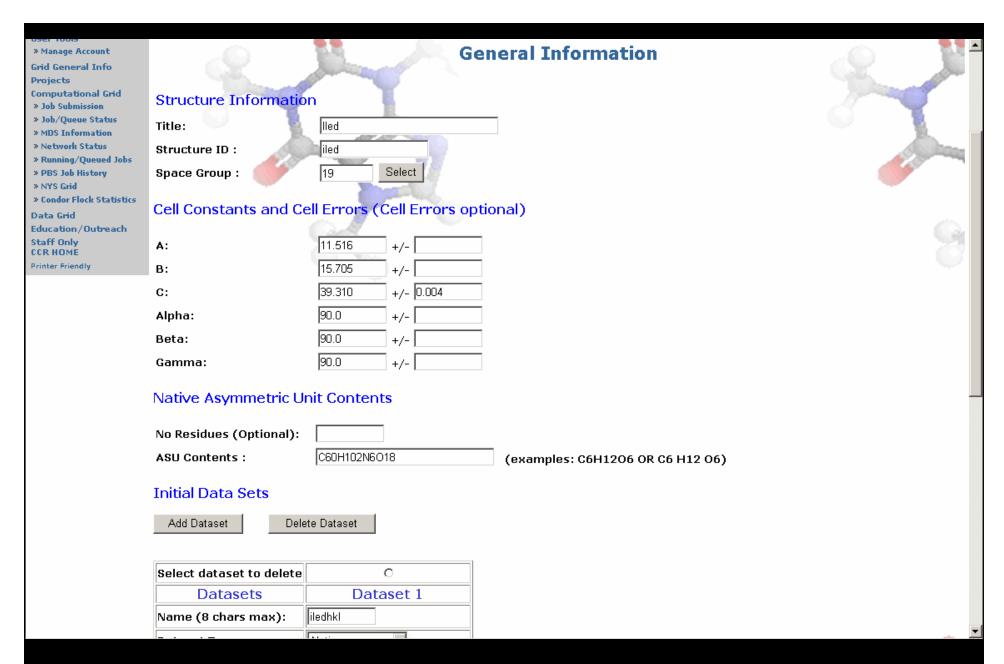
Cancel

Instructions and Description for Running a Job on ACDC-Grid



Software Package Selection



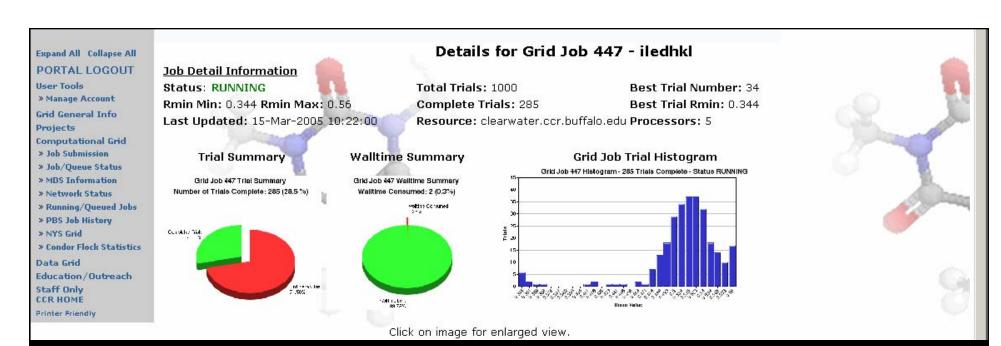


Default Parameters Based on Template

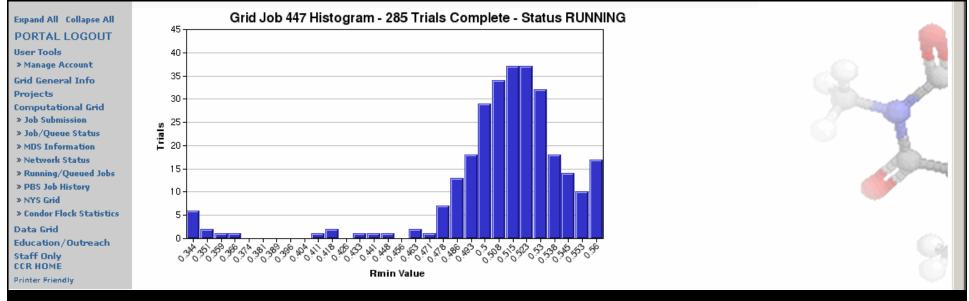
SnB Review (Grid job ID: 447)

Unused

Minimum |E|:

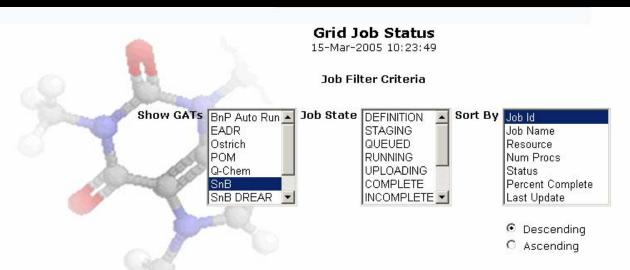


Graphical Representation of Intermediate Job Status



Histogram of Completed Trial Structures



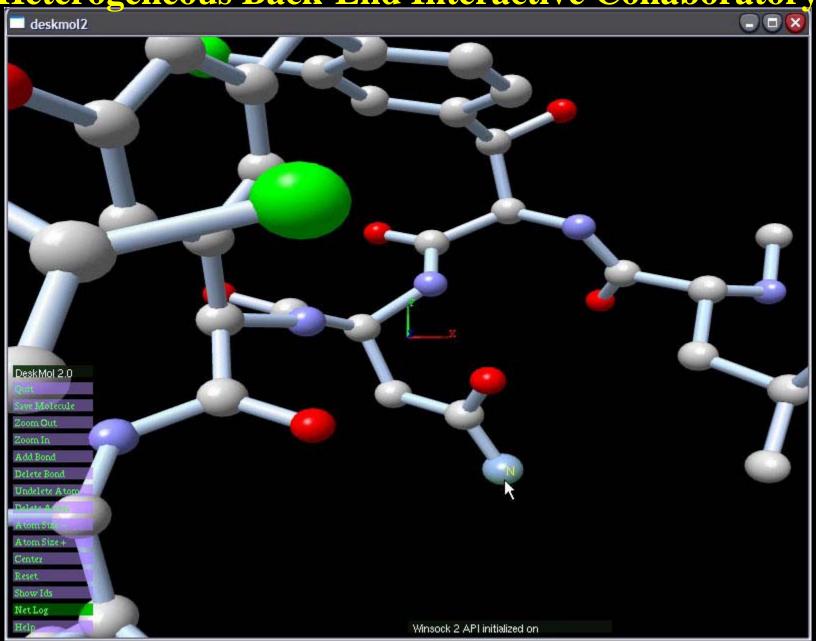


				SnB				
Job Id	Job Name	Resource	Num Procs	Status	Percent Complete	Last Update	Cancel Job	Drilldown
447	iledhkl	clearwater.ccr.buffalo.edu	5	RUNNING	28.5	15-Mar-2005 10:22:00		~
446	trilys	clearwater.ccr.buffalo.edu	10	RUNNING	1	15-Mar-2005 10:22:00		~
444	64chkl	nash.ccr.buffalo.edu	3	COMPLETE	100	14-Mar-2005 22:00:01		4
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		4
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		4
433	16chkl	clearwater.ccr.buffalo.edu	5	COMPLETE	100	10-Mar-2005 14:38:01		4

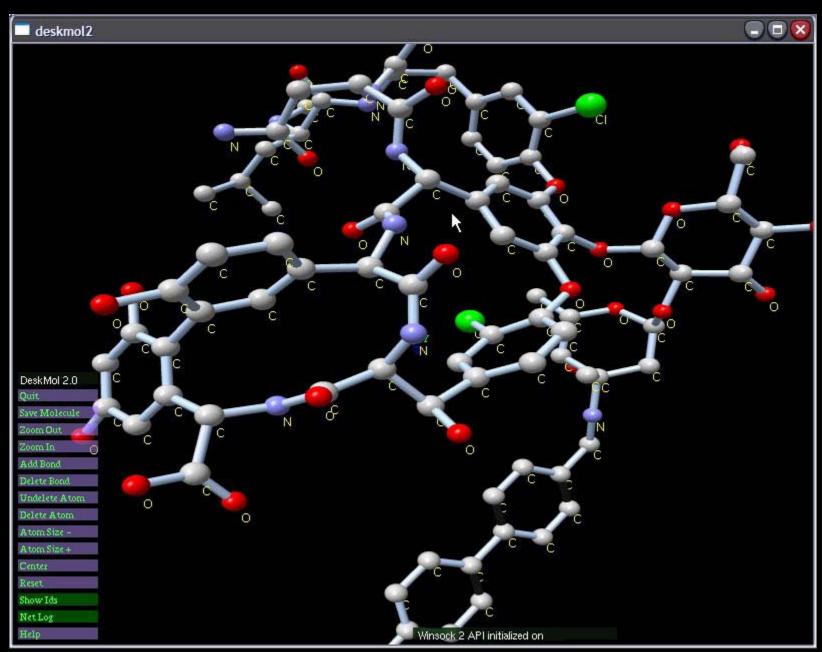
Filter Job List

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/



NYSGrid.org

- Grass-Roots Cyberinfrastructure Initiative in NYS.
- Open to academic and research institutions.
- Mission Stmt: 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.
- www.nysgrid.org



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





