

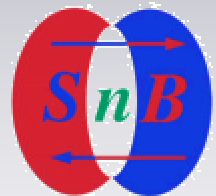
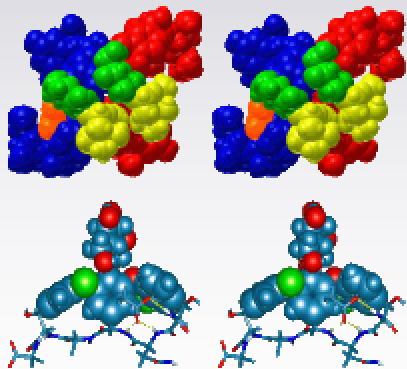
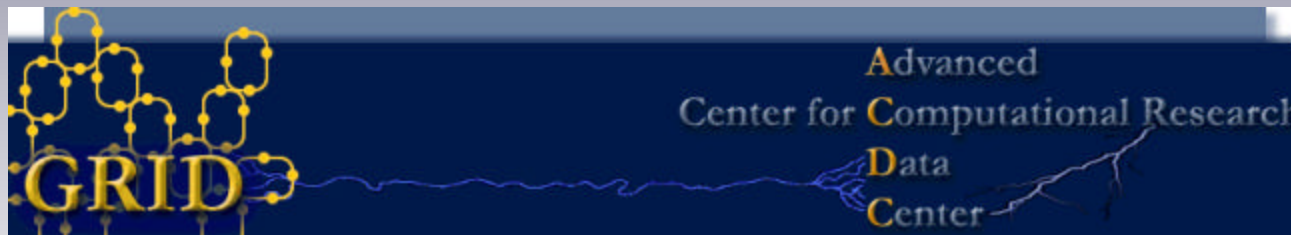
# BnP on the Grid

Russ Miller<sup>1,2,3</sup>, Mark Green<sup>1,2</sup>, Charles M. Weeks<sup>3</sup>

<sup>1</sup>Center for Computational Research, SUNY-Buffalo

<sup>2</sup>Computer Science & Engineering SUNY-Buffalo

<sup>3</sup>Hauptman-Woodward Medical Research Institute



University at Buffalo

The State University of New York

NSF, NIH, DOE, NYS

# Grid Computing

iVD gL

NEESgrid

Data GRID

GLOBAL GF

GriPhyN  
Data Intensive Science

European GRID Forum

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

DISCOM

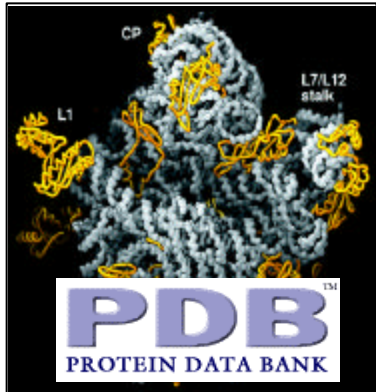
SinRG

APGrid

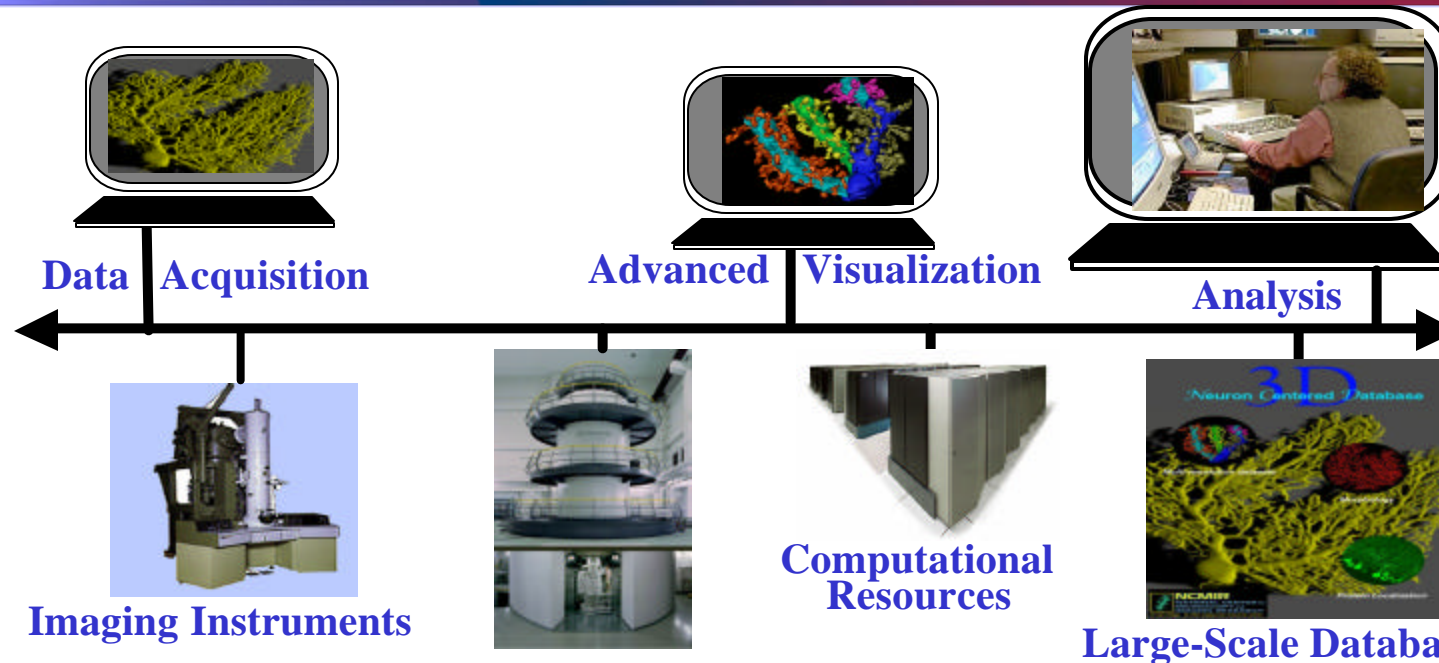
IPG ...

APAN  
Asia-Pacific Advanced Network

EUROGRID



# Grid Computing Overview



Thanks to  
Mark Ellisman

- 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

# Factors Enabling the Grid

- **Internet is Infrastructure**
  - ❑ Increased network bandwidth and advanced services
- **Advances in Storage Capacity**
  - ❑ Terabyte costs less than \$5,000
- **Internet-Aware Instruments**
- **Increased Availability of Compute Resources**
  - ❑ Clusters, supercomputers, storage, visualization devices
- **Advances in Application Concepts**
  - ❑ Computational science: simulation and modeling
  - ❑ Collaborative environments ® large and varied teams
- **Grids Today**
  - ❑ Moving towards production; Focus on middleware

# SnB on Grids

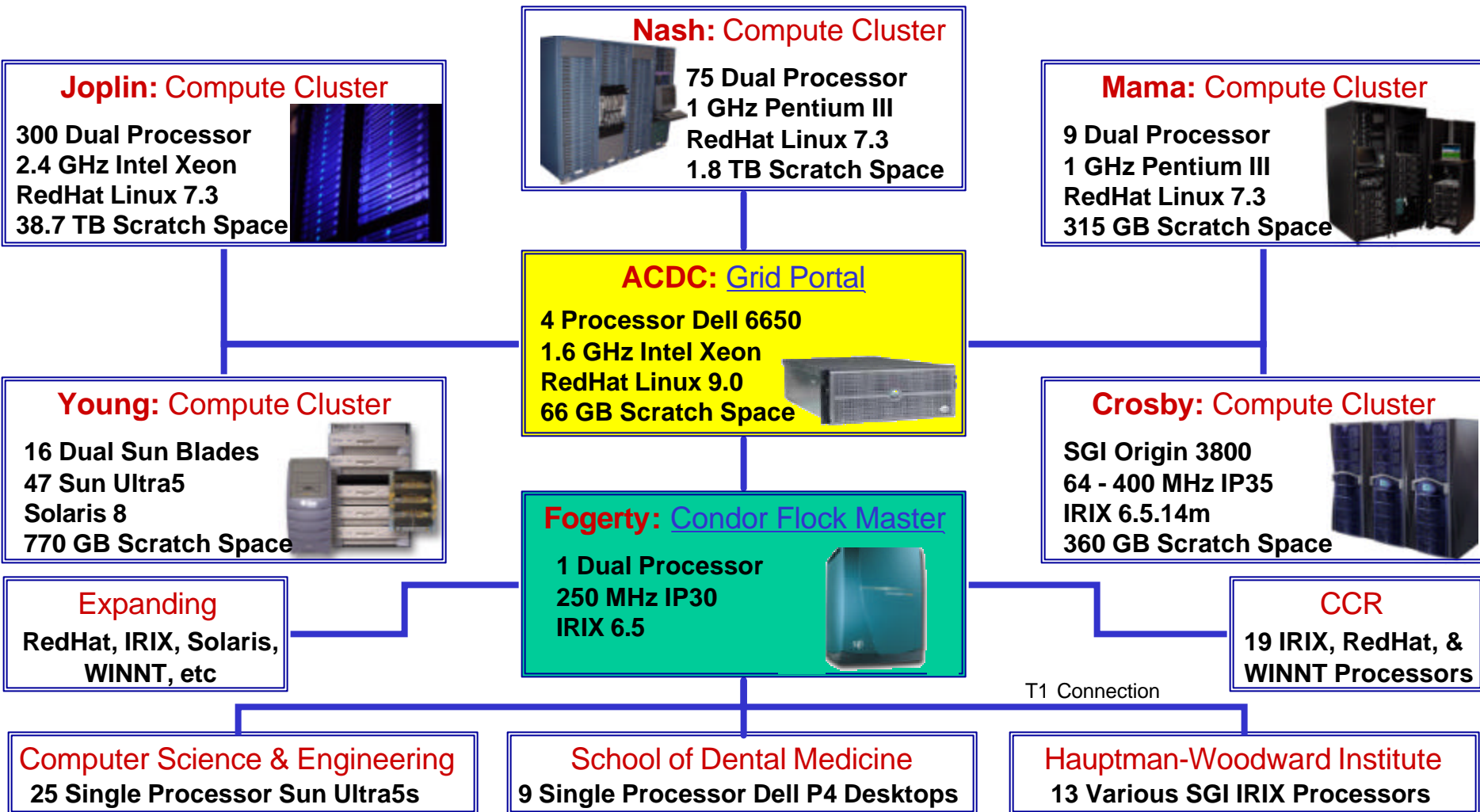
- **ACDC-Grid (Western New York)**
  - ❑ CIT (UB), CCR (UB), CSE (UB), Dental (UB), HWI
  - ❑ Linux, Windows, IRIX, AIX, Solaris
  - ❑ Pentium, Itanium, Power, MIPS
- **Grid3+ (International): GriPhyN, PPDG**
  - ❑ 29 Sites: ANL, SMU, BNL, BU, CalTech-Grid3, CalTech-PG, FIU, HU, IU, JHU, KNU, OU-HEP, OU-OSCER, PDSF, PSU, Rice, UB, UCSD, UCSD-Prod, UIC, UFL-Grid3, UFL-PG, UMICH, UNM, FNAL, UTA, UWMad, UWMil, Vanderbilt
  - ❑ GriPhyN, PPDG, iVDGL, LIGO, ATLAS/CMS/LHC@CERN
  - ❑ VOs: iVDGL, LIGO, SDSS, USATLAS, USCMS and BTeV
  - ❑ Linux/Pentium, VDT, Globus, ACDC Monitoring, MonaLisa, Ganglia, Condor, PBS, LSF, FBS, PyGlobus, Perl, Pacman
- **IBM NE BioGrid (Northeast USA)**
  - ❑ MIT, Harvard, MGH
  - ❑ Regatta, Pentium, Linux

# Major CCR Resources (12TF & 290TB)

- **Dell Linux Cluster: #22® #25® #38® #95**
    - ❑ 600 P4 Processors (2.4 GHz)
    - ❑ 600 GB RAM; 40 TB Disk; Myrinet
  - **Dell Linux Cluster: #187® #368® off**
    - ❑ 4036 Processors (PIII 1.2 GHz)
    - ❑ 2TB RAM; 160TB Disk; 16TB SAN
  - **IBM BladeCenter Cluster: #106**
    - ❑ 532 P4 Processors (2.8 GHz)
    - ❑ 5TB SAN
  - **SGI Origin3700 (Altix)**
    - ❑ 64 Processors (1.3GHz ITF2)
    - ❑ 256 GB RAM
    - ❑ 2.5 TB Disk
  - **SGI Origin3800**
    - ❑ 64 Processors (400 MHz)
    - ❑ 32 GB RAM; 400 GB 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
- 
- **IBM RS/6000 SP: 78 Processors**
  - **Sun Cluster: 80 Processors**
  - **SGI Intel Linux Cluster**
    - ❑ 150 PIII Processors (1 GHz)
    - ❑ Myrinet

# Advanced Computational Data Center

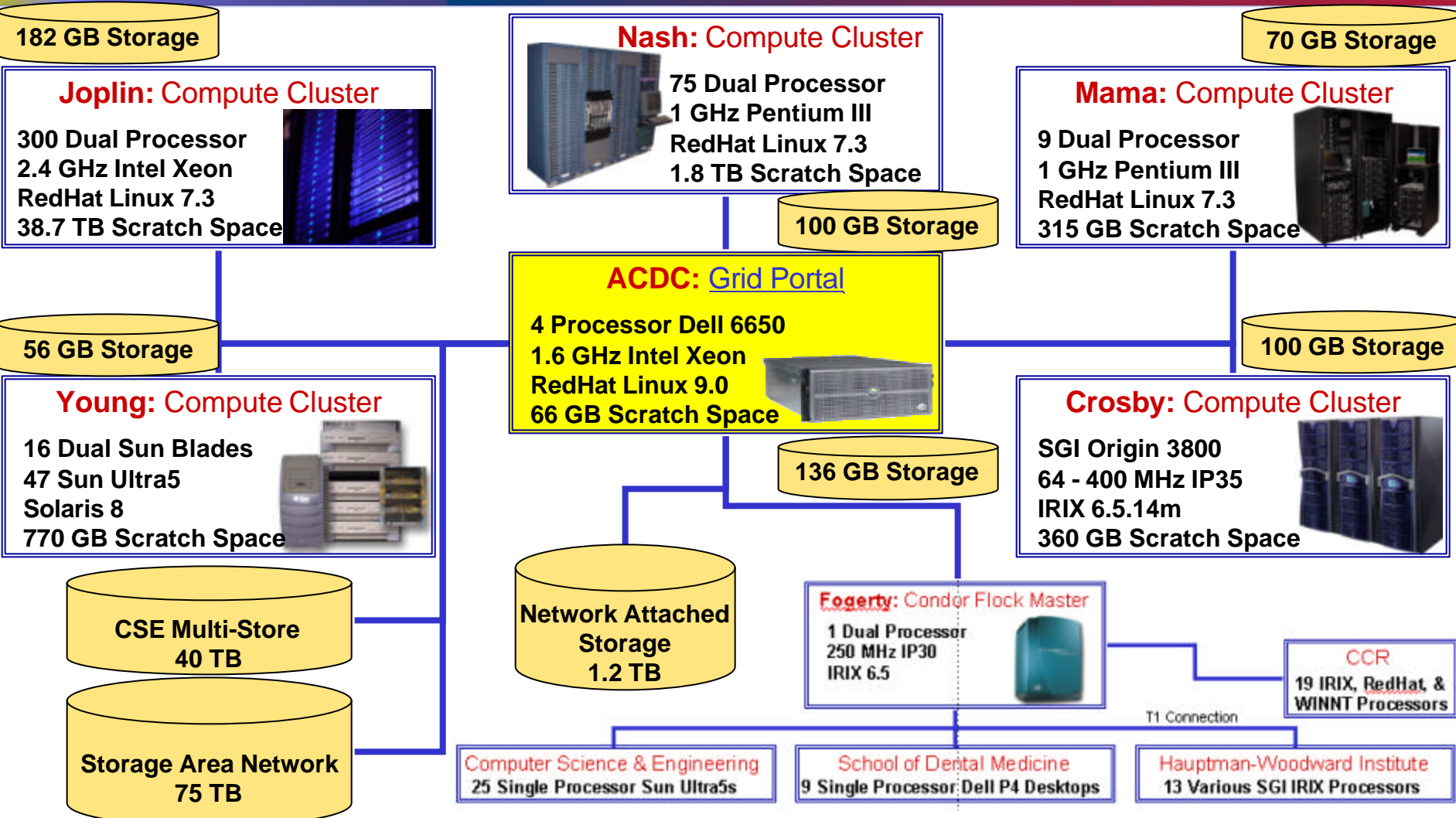
## ACDC: Grid Overview



Note: Network connections are 100 Mbps unless otherwise noted.

# ACDC Data Grid Overview

## (Grid-Available Data Repositories)



Note: Network connections are 100 Mbps unless otherwise noted.



# ACDC-Grid Cyber-Infrastructure

## ■ Predictive Scheduler

- Define quality of service estimates of job completion, by better estimating job runtimes by profiling users.

## ■ Data Grid

- Automated Data File Migration based on profiling users.

## ■ High-performance Grid-enabled Data Repositories

- Develop automated procedures for dynamic data repository creation and deletion.

## ■ Dynamic Resource Allocation

- Develop automated procedures for dynamic computational resource allocation.

# ACDC-Grid

CCR Grid Computing Services - Microsoft Internet Explorer

CCR University at Buffalo The State University of New York

## 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 Biomimetics, 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.

Get your "Grid Computing Guide"?

CCR Grid Computing Services Data Management - Microsoft Internet Explorer

CCR University at Buffalo The State University of New York

## Center for Computational Research GRID PORTAL

High Performance Grid Computing

PORTAL LOGOUT

VIEW: Group  UserList:

- replaye
  - KeyMaster
  - Morpheus
    - Tank
    - Agent
    - Rabbit
    - Tank
    - Morpheus
    - Oracle.m
    - Neo

Browser view of "miller" group files published by user

CCR Grid Computing Services Grid Admin - Microsoft Internet Explorer

CCR University at Buffalo The State University of New York

## Center for Computational Research GRID PORTAL

High Performance Grid Computing

PORTAL LOGOUT

View statistics for:

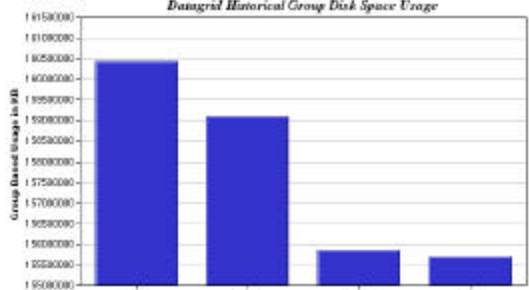
Data based on:

from starting date:

to ending date:    inclusive

for:  resources:

**Baagrid Historical Group Disk Space Usage**



Group	Disk Space Usage (KB)
miller	~1,050,000
griddev	~1,000,000
ccstaff	~1,000,000
mlgreen	~1,000,000

CCR Grid Computing Services: Grid Admin - Microsoft Internet Explorer

CCR University at Buffalo The State University of New York

## Center for Computational Research GRID PORTAL

High Performance Grid Computing

PORTAL LOGOUT

View statistics for:

Data based on:

from starting date:

to ending date:    inclusive

for:  resources:

File_num	File_ID	Filename	Dir_ID	Resource_ID	Owner	Groupname	Type
1	56033	Cypher.txt	52831	10	mlgreen	griddev	txt
2	56034	Cypher.sh	52858	10	mlgreen	griddev	sh
3	56035	Oracle.asc	52958	10	mlgreen	griddev	asc
4	56036	Cypher.sh	52634	10	mlgreen	miller	sh
5	56037	Rabbit.dat	52830	10	mlgreen	ccstaff	dat
6	56038	Agent.exe	53064	10	mlgreen	griddev	exe
7	56039	Dozer.sh	52852	10	mlgreen	griddev	sh
8	56040	Neo.asc	52187	10	mlgreen	mlgreen	asc
9	56041	Agent.mpg	52833	10	mlgreen	mlgreen	mpg
10	56042	Tank.txt	52188	10	mlgreen	mlgreen	txt
11	56043	Smith.xls	52258	10	mlgreen	ccstaff	xls
12	56044	KeyMaster.csh	52186	10	mlgreen	miller	csh
13	56045	Oracle.csh	52622	10	mlgreen	griddev	csh
14	56046	Dozer.xls	52808	10	mlgreen	mlgreen	xls
15	56047	Cypher.exe	52204	10	mlgreen	griddev	exe
16	56048	Rabbit.ppt	52861	10	mlgreen	miller	ppt
17	56049	Neo.dat	52217	10	mlgreen	ccstaff	dat
18	56050	Cypher.asc	53086	10	mlgreen	griddev	asc



# ACDC-Grid Administration

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

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  
Center Fleck 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.)

**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:

Optional grid-mapfile path:

Do not stage this file to the grid nodes

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

Center for Computational Research **GRID PORTAL**  
High Performance Grid Computing

**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:

Run Script:

Run As User:

[Return to the Database Job Admin menu.](#)  
[Return to the Grid Admin menu.](#)

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

Center for Computational Research **GRID PORTAL**  
High Performance Grid Computing

**MDS Resource Update Status**

Current Time: 16-September-2003 10:59:12

Resource	Last Updated	Next Update	Status
crasby.ccr.buffalo.edu	16-September-2003 09:15:30	2 minutes	OK
rogerty.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
neius.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



# Molecular Structure Determination via *Shake-and-Bake*

## ■ *SnB* Software by UB/HWI

- ❑ “Top Algorithms of the Century”

## ■ Worldwide Utilization

## ■ Critical Step

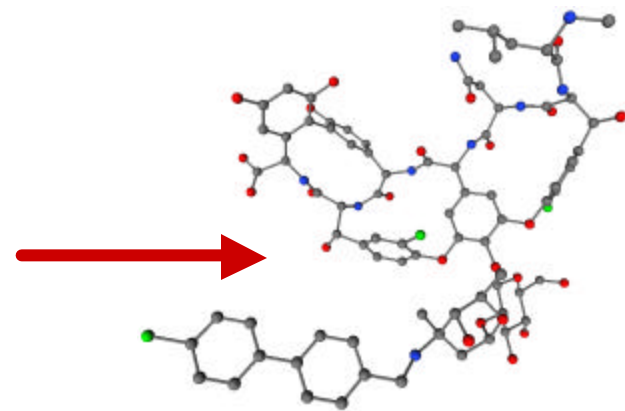
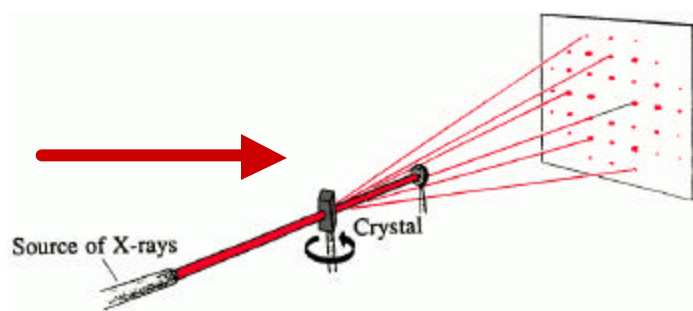
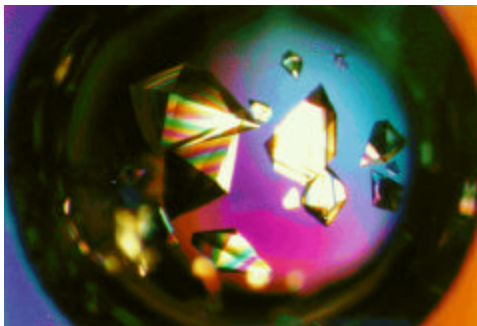
- ❑ Rational Drug Design
- ❑ Structural Biology
- ❑ Systems Biology

## ■ Vancomycin

- ❑ “Antibiotic of Last Resort”

## ■ Current Efforts

- ❑ Grid
- ❑ Collaboratory
- ❑ {Intelligent Learning}



Title:

Structure ID (1.6 char. max):  Space Group:  ...

Cell Constants and Cell Errors (Cell Errors optional)

A:  +/-  B:  +/-  C:  +/-

Alpha:  +/-  Beta:  +/-  Gamma:  +/-

Native Asymmetric Unit Contents

Given R residues, then by default use C5R, N1.2R, O1.5R, H8R. No. Residues (optional):

Contents:

Examples: C6H12O6 or C6 H12 O6 or C6,H12,O6

Initial Data Sets

Add	Delete
	1
Name (8 char. max.)	vancohkl
File name	trivanco.hkl
File type	F, Sig(F) ▼
Data set type	Native ▼
Wavelength	1.5418
Max. resolution	0.97
Anomalous dispersion	Not measured ▼
Heavy element type	
Nat. element replaced	
No. expected sites	
f'	
f''	

# General Information Screen for TriVanco

Normalize reflections and generate invariants for SnB jobs

Output Data Set	Job Type	Input Data Sets		Normalization Method	Select ?
		Native	Derivative		
vancohkl	Basic	vancohkl		Wilson (anisotropic)	<input checked="" type="checkbox"/>

Prepare data for SnB runs.

1. Select one or more data sets.
2. If desired, choose a non-default normalization method.
3. Click "Normalize" to compute E values.
4. Click "Generate" to create a structure invariant file.

Normalize...

Generate...

Merge and scale reflections for use in PHASES

Output File	File Type	Input Data Sets		Relative Wilson Scaling Method	Select ?
		Native	Derivative		

Prepare data for protein phasing.

1. Select data sets to use.
2. If desired, choose a non-default scaling method.
3. Click "Merge & Scale".

For each selected set, the appropriate merged & scaled file (\*.scl) will be created.

Merge & Scale

Review Results

Normalize	Generate Invariants	Merge & Scale
vancohkl	vancohkl	

- Update List
- View Results
- Clean Files \*

\* for Normalize only

# Normalize Reflections and Generate Invariants

**Invariants**

Input invariant file:

Number of triplet invariants to use:

**Trials to Process**

Starting phases from:  Random seed (prime):

Number of trials:  Input phase file:

Starting at trial:  Input atom file:

Keep complete (every trial) peak file?  Yes  No

**Cycles**

Number of Shake-and-Bake cycles:  Keep complete (every cycle) trace file?  Yes  No

Terminate trials failing the R-Ratio test?  Yes  No R-Ratio cutoff:

**Phase Refinement Method**

Parameter Shift  Tangent Formula Number of passes through phase set:

Phase shift:  Number of shifts:

**Real-Space Constraints**

Number of peaks to select:  Minimum interpeak distance:  Fourier grid size:

Minimum distance between symmetry-related peaks:  (defines special position excluded volume)

Number of special position peaks to keep:

Perform extra cycles with more peaks?  Yes  No Number of extra cycles:  Number of peaks:

**Twice Baking**

Trials for E-Fourier filtering (fourier refinement)?  None  All  Best Only

Number of cycles:  Number of peaks:  Minimum |E|:

**Automatic solution identification criteria for substructures**

Rmin Improvement  Rcryst Improvement

Define *SnB* Parameters: 10,000 trials; 404 cycles; 40400 triplets

Required Information

Queueing system:  Local (no batch system)  Grid

File ID for results:

Grid Job Submission

Authenticate:

Queue:  (Default: grid)

Number of Processors:

Grid Job Status

Job status for currently submitted job

Queue Time: 7/14/04 10:56:32  
User: nmshah  
Grid Job ID: 2243  
Session ID:1693285786  
  
Current Status:  
  
Grid Job 2243 Accepted and Queued

Run SnB job on Grid



Review Results

\* vancohl/job-vanco

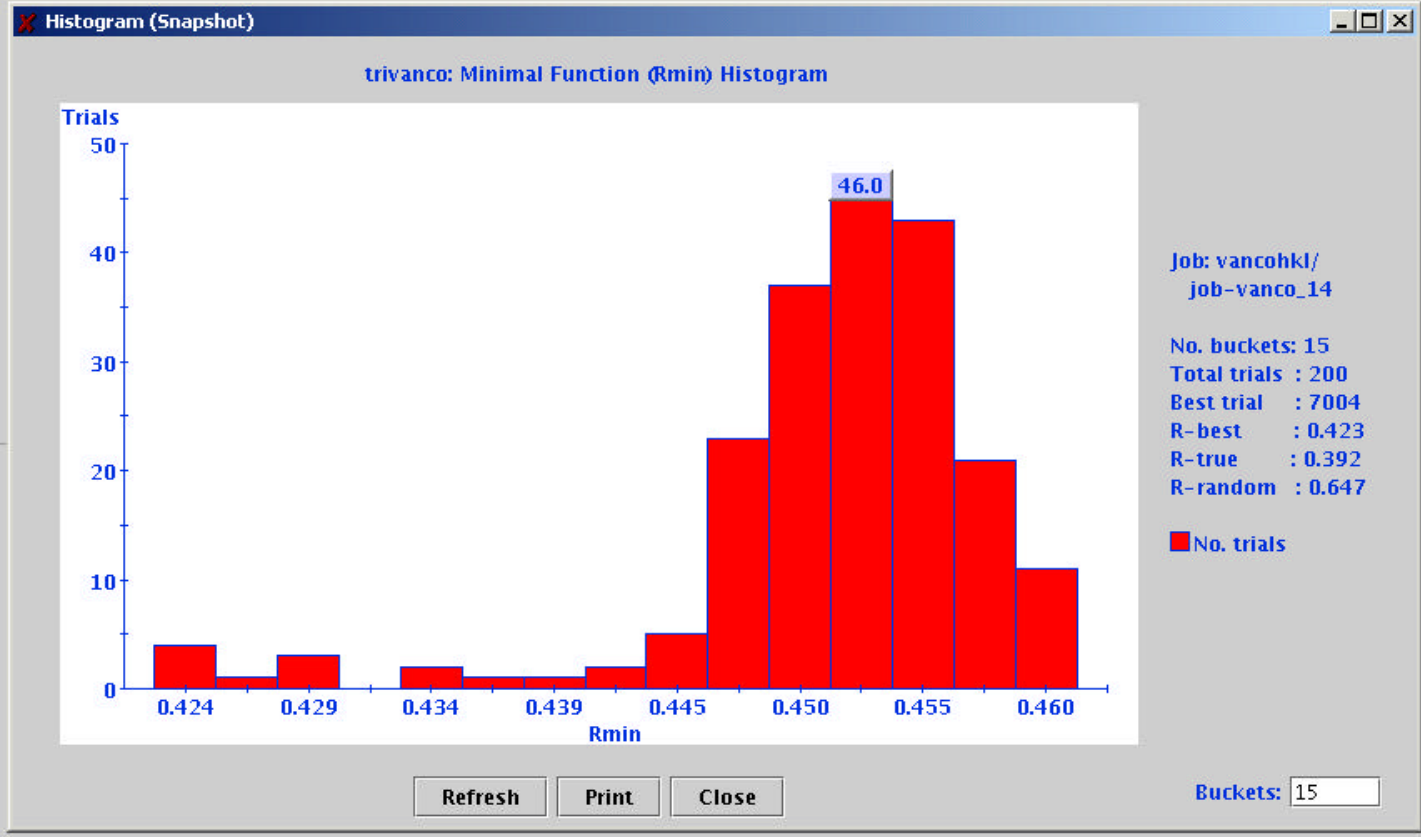
\* active job

Steps in Substructure Solution

1. Select an SnB job from the list at the left.
2. Determine whether solution(s) have been found.
3. Select peaks to be saved in a PDB file as a substructure model.
4. Optionally, peak selection may involve comparison of two or more probable solutions.
5. If desired, improve the model by adding or deleting peaks.

Job Selection & Control

- Update List
- Select None
- Select Another Trial
- Kill Job
- Delete Job



Histogram of Final Rmin Values after 200 Trials Completed

Review Results

- \* vancohl/job-vanco

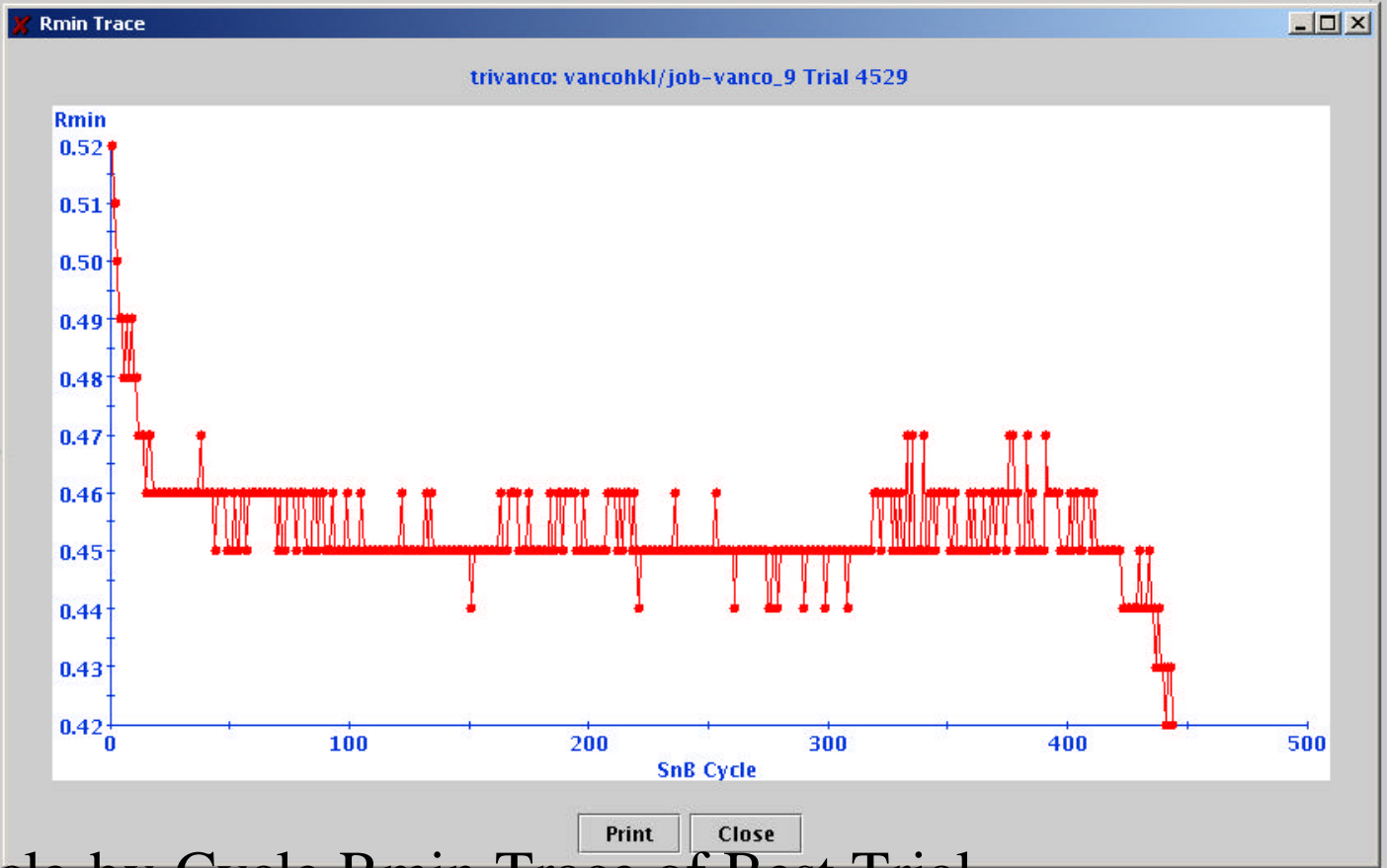
\* active job

Steps in Substructure Solution

1. Select an SnB job from the list at the left.
2. Determine whether solution(s) have been found.
3. Select peaks to be saved in a PDB file as a substructure model.
4. Optionally, peak selection may involve comparison of two or more probable solutions.
5. If desired, improve the model by adding or deleting peaks.

Job Selection & Control

- Update List
- Select None
- Select Another Trial
- Kill Job
- Delete Job

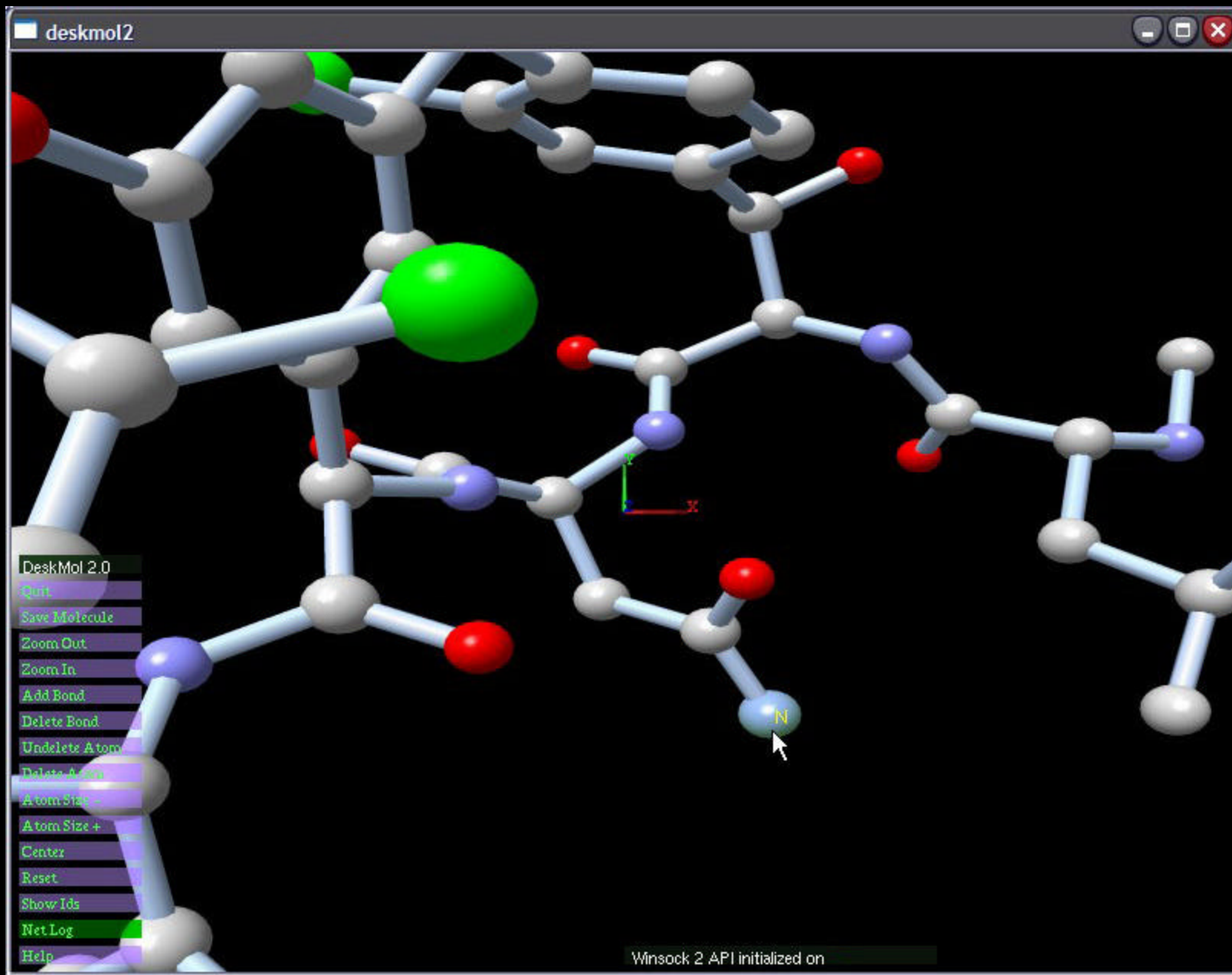


Cycle-by-Cycle Rmin Trace of Best Trial

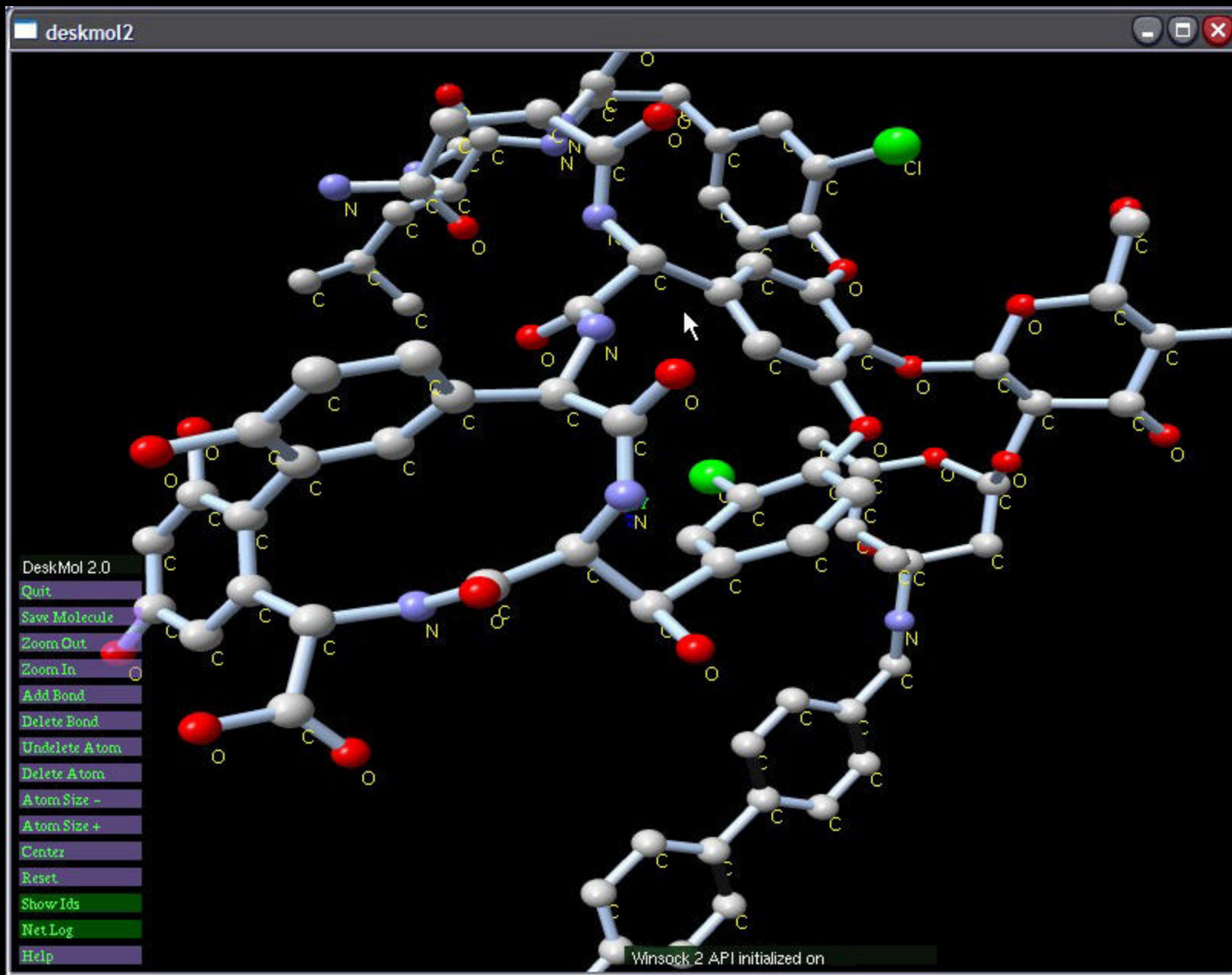


# Live Demo

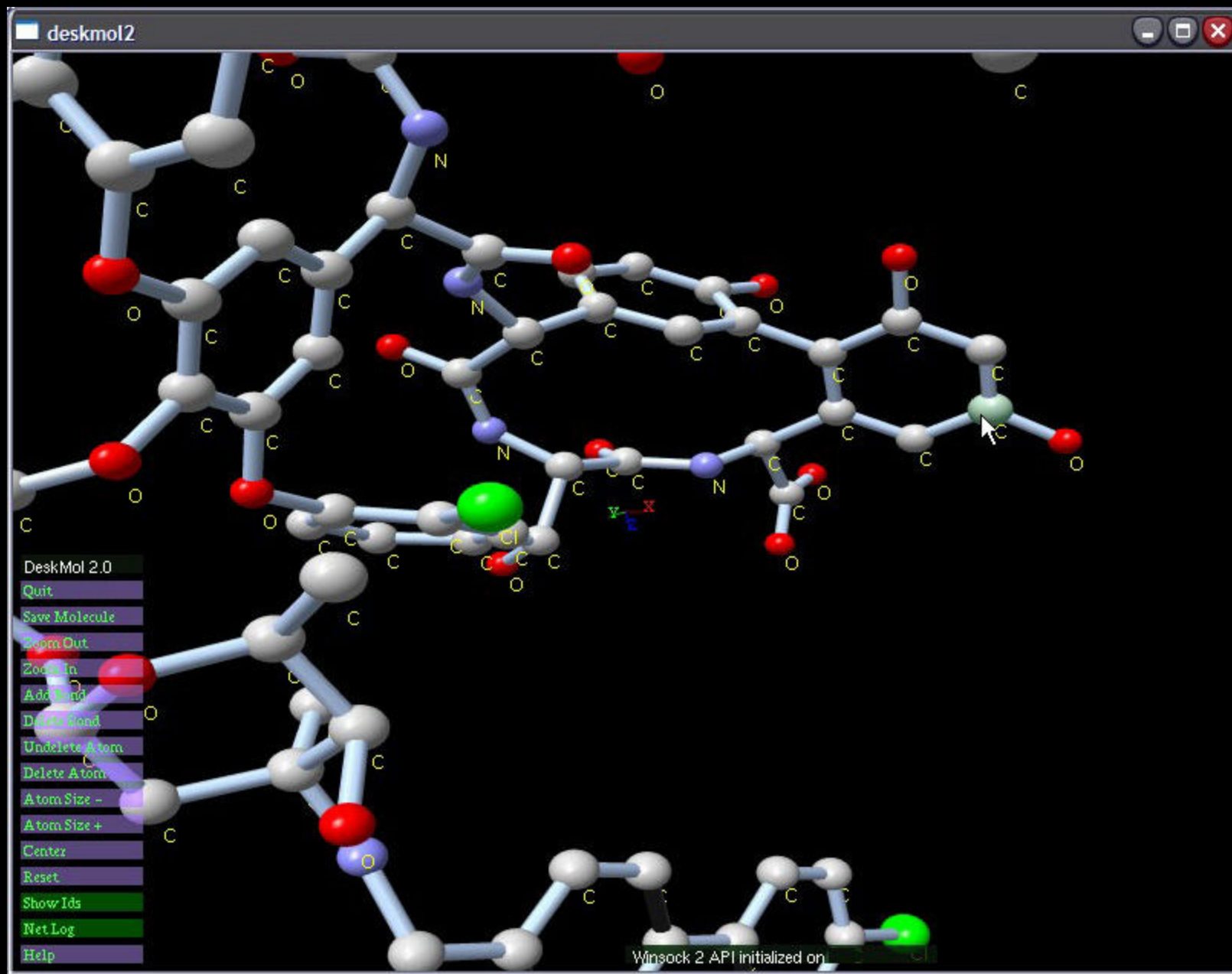
- **Time to Gamble**
- **Demonstration using ACDC Grid in Buffalo**



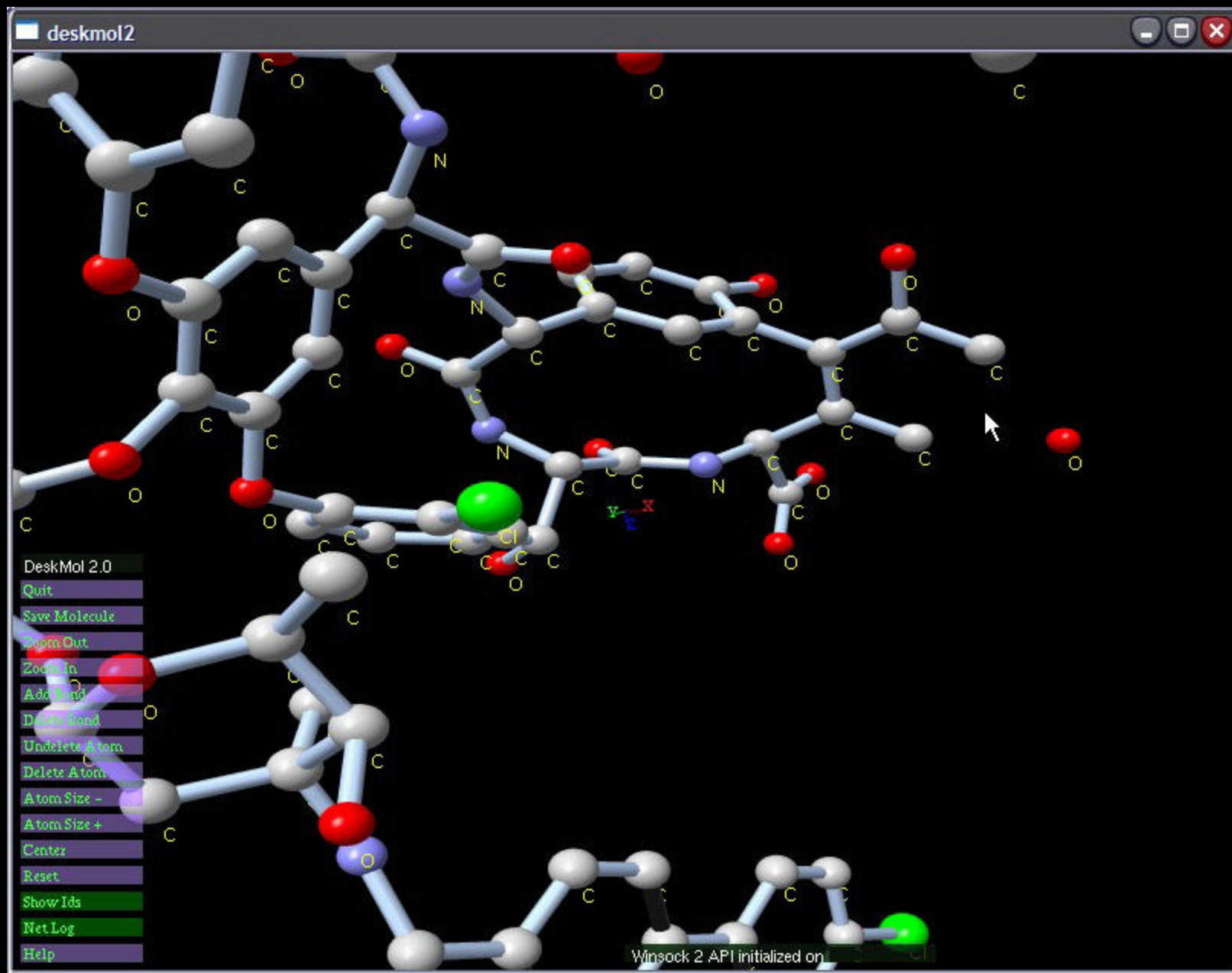
User starts up – default image of structure.



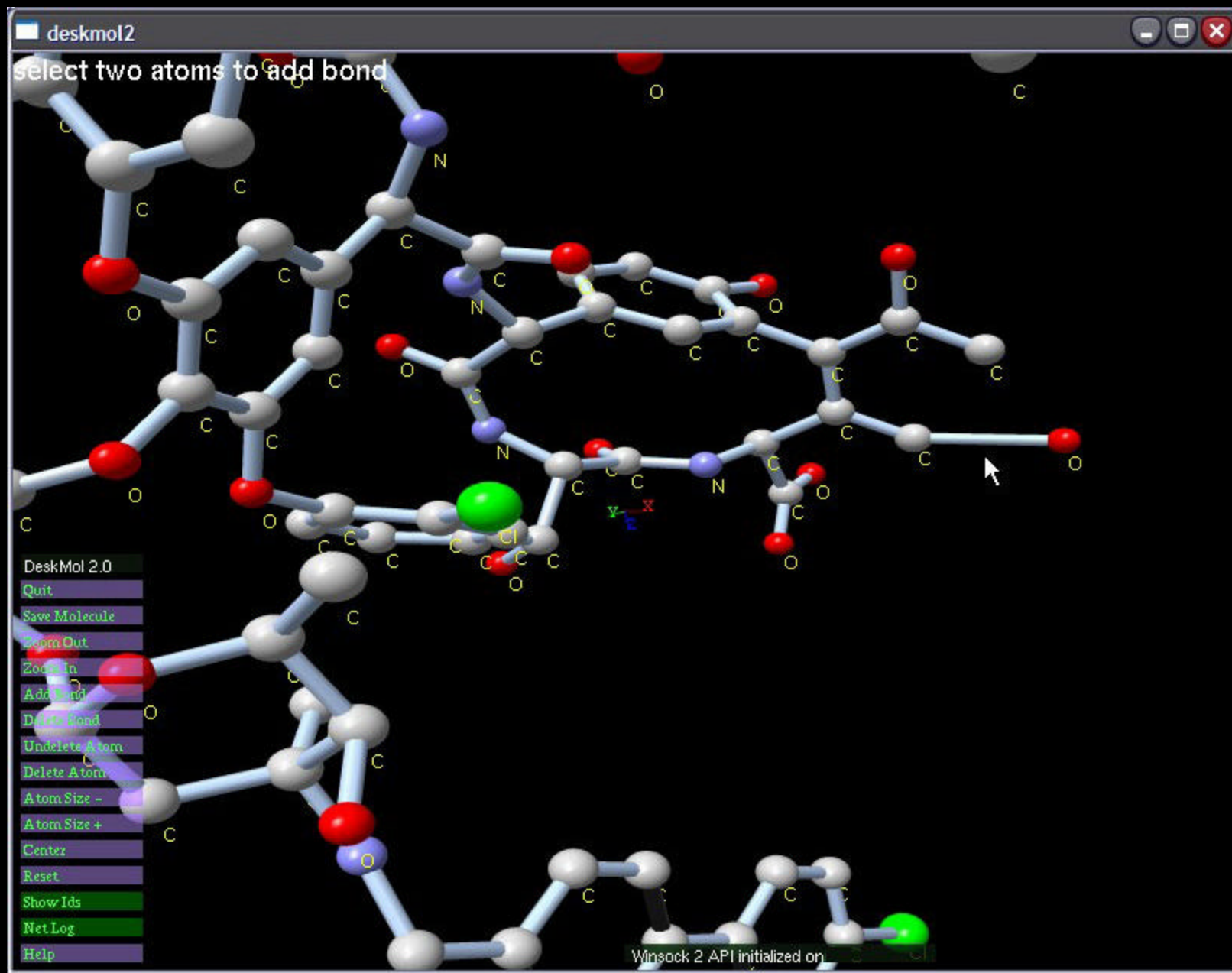
Molecule scaled, rotated, and labeled.



User Mouse to Select Carbon Atoms

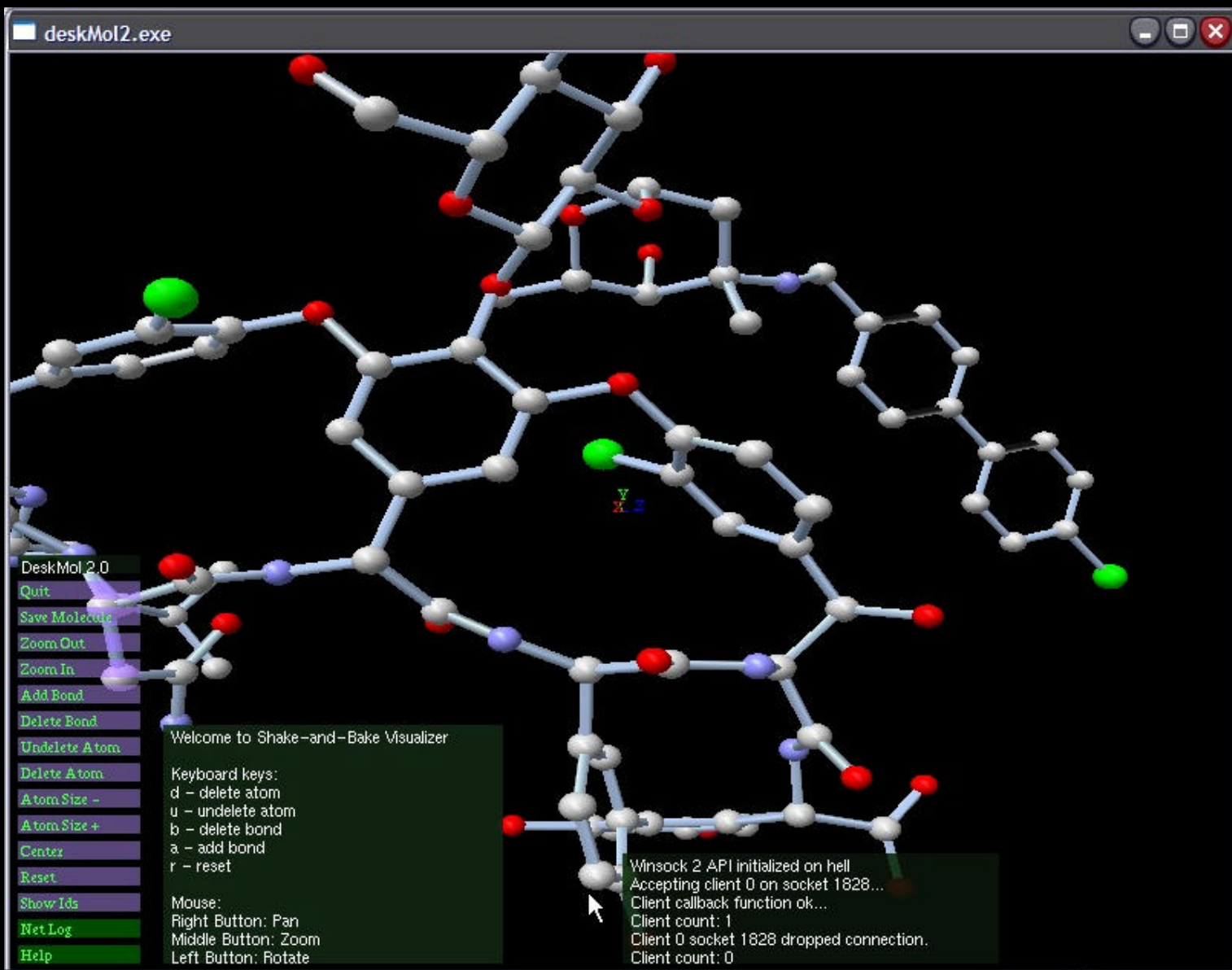


Remove Carbon Atoms (and Links)



User Adds Bond Between Atoms





## Scale Radius of Atoms

deskMol2.exe

DeskMol 2.0

- Quit
- Save Molecule
- Zoom Out
- Zoom In
- Add Bond
- Delete Bond
- Undelete Atom
- Delete Atom
- Atom Size -
- Atom Size +
- Center
- Reset
- Show Ids
- Net Log
- Help

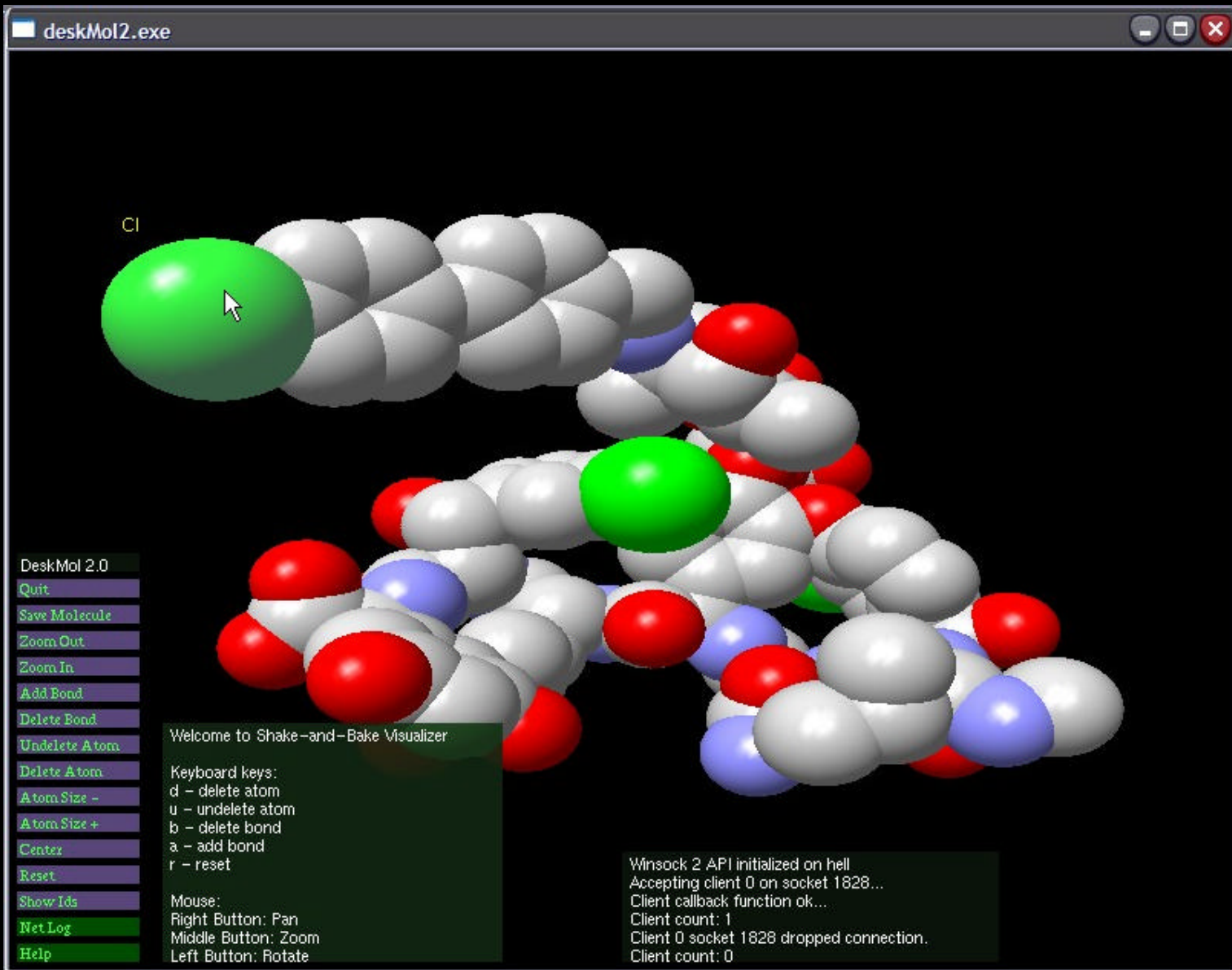
Welcome to Shake-and-Bake Visualizer

Keyboard keys:  
d - delete atom  
u - undelete atom  
b - delete bond  
a - add bond  
r - reset

Mouse:  
Right Button: Pan  
Middle Button: Zoom  
Left Button: Rotate

Winsock 2 API initialized on hell  
Accepting client 0 on socket 1828...  
Client callback function ok...  
Client count: 1

Continue Scaling Atoms



Continue Scaling Atoms

# 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

# ACDC-Grid Collaborations

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



# Acknowledgments

- **Amin Ghadersohi**
- **Naimesh Shah**
- **Stephen Potter**
- **Cathy Ruby**
- **Steve Gallo**
- **Jason Rappleye**
- **Martins Innus**
- **Jon Bednasz**
- **Sam Guercio**
- **Dori Macchioni**



[www.ccr.buffalo.edu](http://www.ccr.buffalo.edu)