

The Great Pandemics and Simulation Based Medicine

Wilfred W. Li, Ph.D.

SEAIP09/PRAGMA Institute

Taichung, Taiwan, Dec 3, 2009

National Biomedical Computation Resource
Center for Research in Biological Systems
San Diego Supercomputer Center
University of California, San Diego

PRAGMA

A Practical Collaborative Framework



**Strengthen Existing and Establish
New Collaborations**

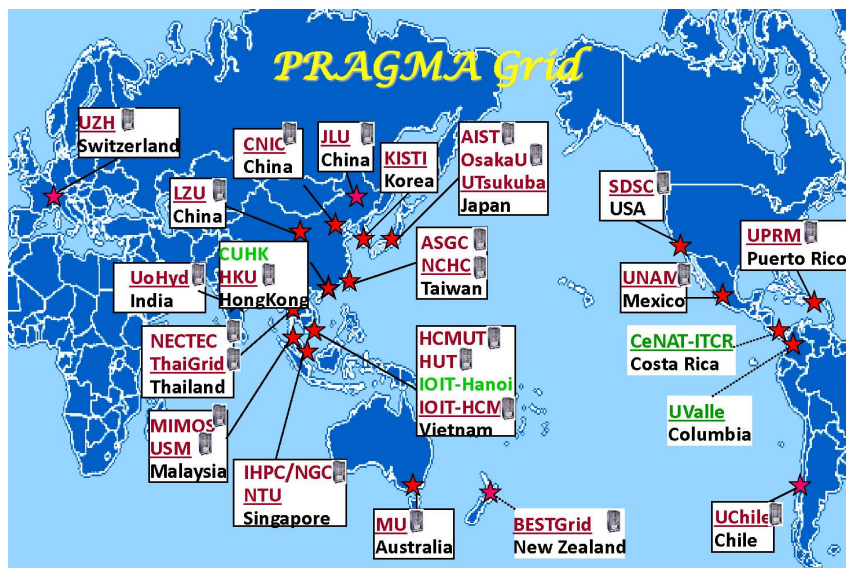
**Work with Science Teams to
Advance Grid Technologies and
Improve the Underlying
Infrastructure**

In the Pacific Rim and Globally

<http://www.pragma-grid.net>

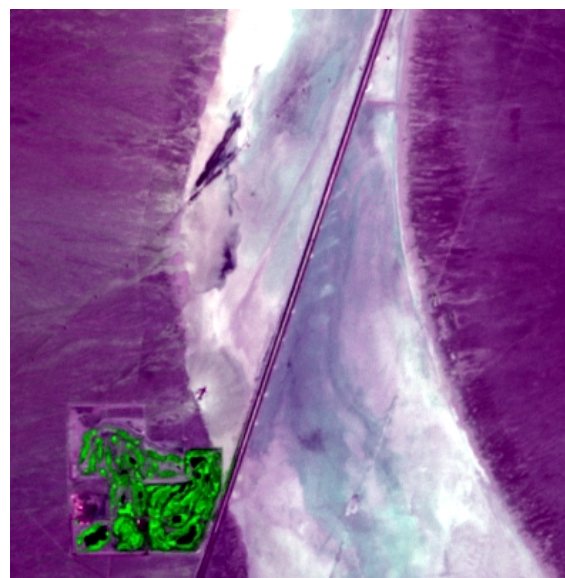
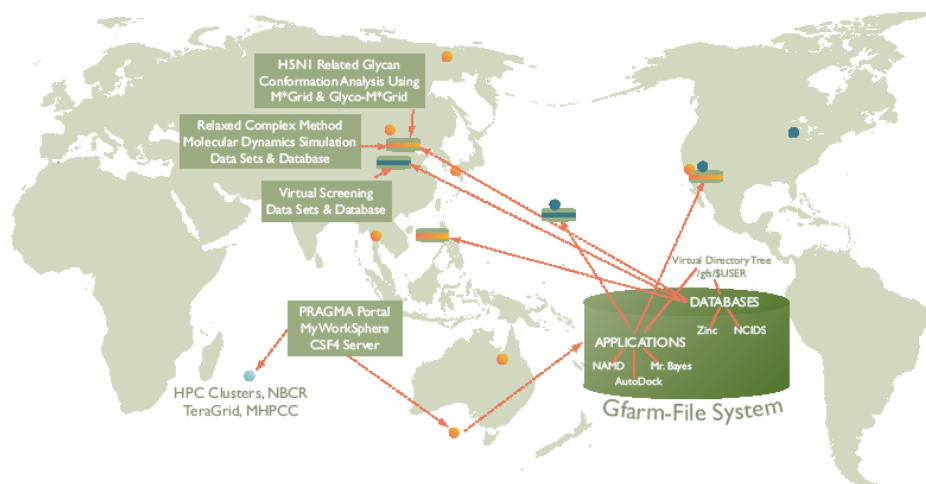
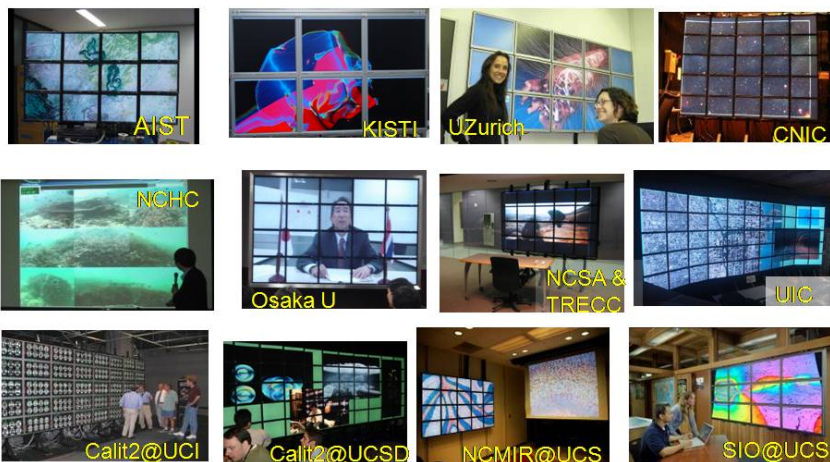
Working Groups: Organize Activities

Resources Biosciences



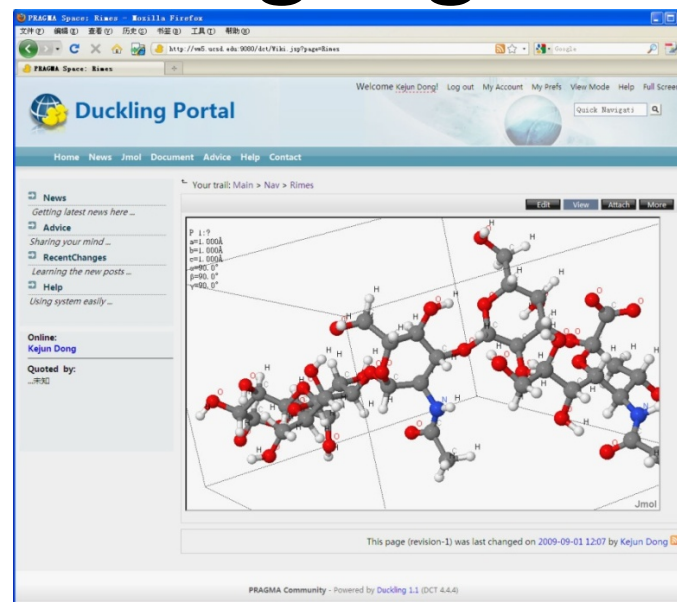
28 institutions in 16 countries/regions, 26 compute sites (+ 4 site in preparation)

October 2009

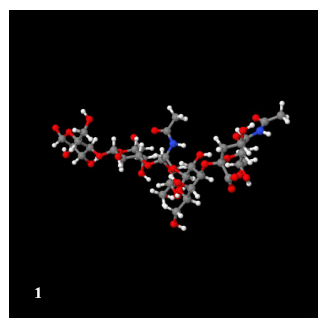
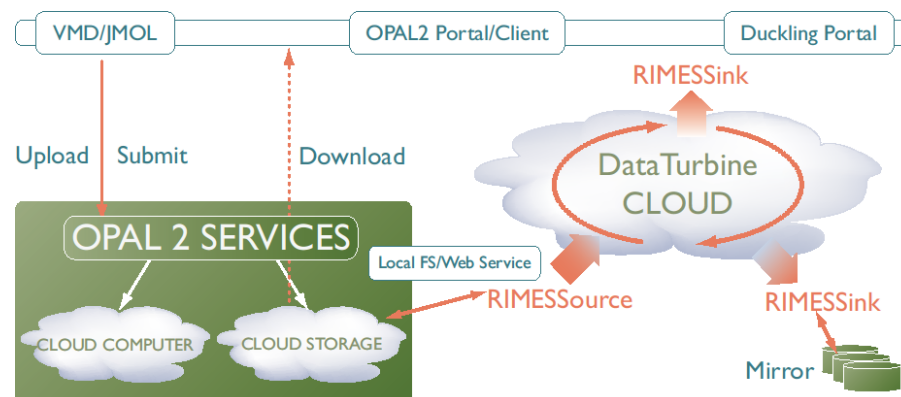


Biosciences 2008-2009 Highlights

- Streaming data from simulations to visualization
 - DataTurbine, Duckling
 - Kevin Dong
- Virtualization
 - Virtual Clusters, AMI



JMol Plug-in integrated into Duckling Portal



Transparent access of applications on Avian Flu Grid through middleware

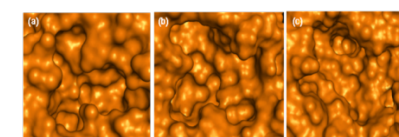
H5N1 related glycan conformation analysis using M*Grid and Glyco-M*Grid

Relaxed Complex Method Molecular Dynamics Simulation Data Sets & Database

Virtual Screening Data Sets & Database

ID	Owner	Description	Date	Exp	Result	Details
1	Asn88	DATA_200y	Apr 22, 2008 8:00:09 PM	(link)	(link)	(link)
2	Asn88	DATA_200y	Apr 22, 2008 8:00:11 PM	(link)	(link)	(link)
3	Asn88	DATA_200y	Apr 22, 2008 8:00:12 PM	(link)	(link)	(link)
4	Asn88	DATA_200y	Apr 22, 2008 8:00:13 PM	(link)	(link)	(link)
5	Asn88	DATA_200y	Apr 22, 2008 8:00:14 PM	(link)	(link)	(link)
6	Asn88	DATA_200y	Apr 22, 2008 8:00:15 PM	(link)	(link)	(link)
7	Asn88	DATA_200y	Apr 22, 2008 8:00:16 PM	(link)	(link)	(link)
8	Asn88	DATA_200y	Apr 22, 2008 8:00:17 PM	(link)	(link)	(link)
9	Asn88	DATA_200y	Apr 22, 2008 8:00:18 PM	(link)	(link)	(link)
10	Asn88	DATA_200y	Apr 22, 2008 8:00:19 PM	(link)	(link)	(link)

CNIC VSDB



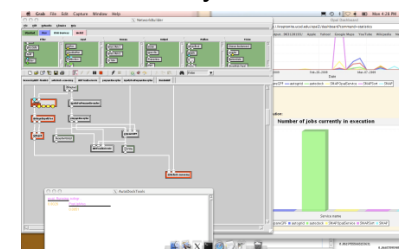
Crystal

MD-CHARMM

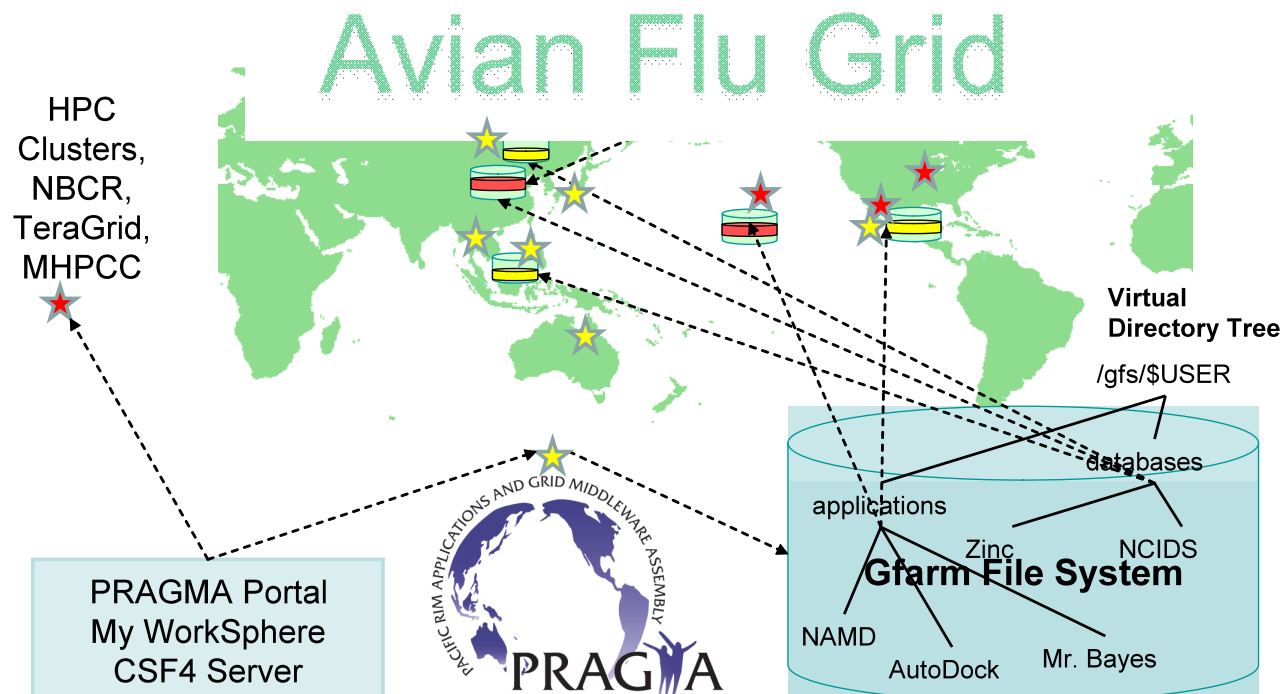
MD-NAMD



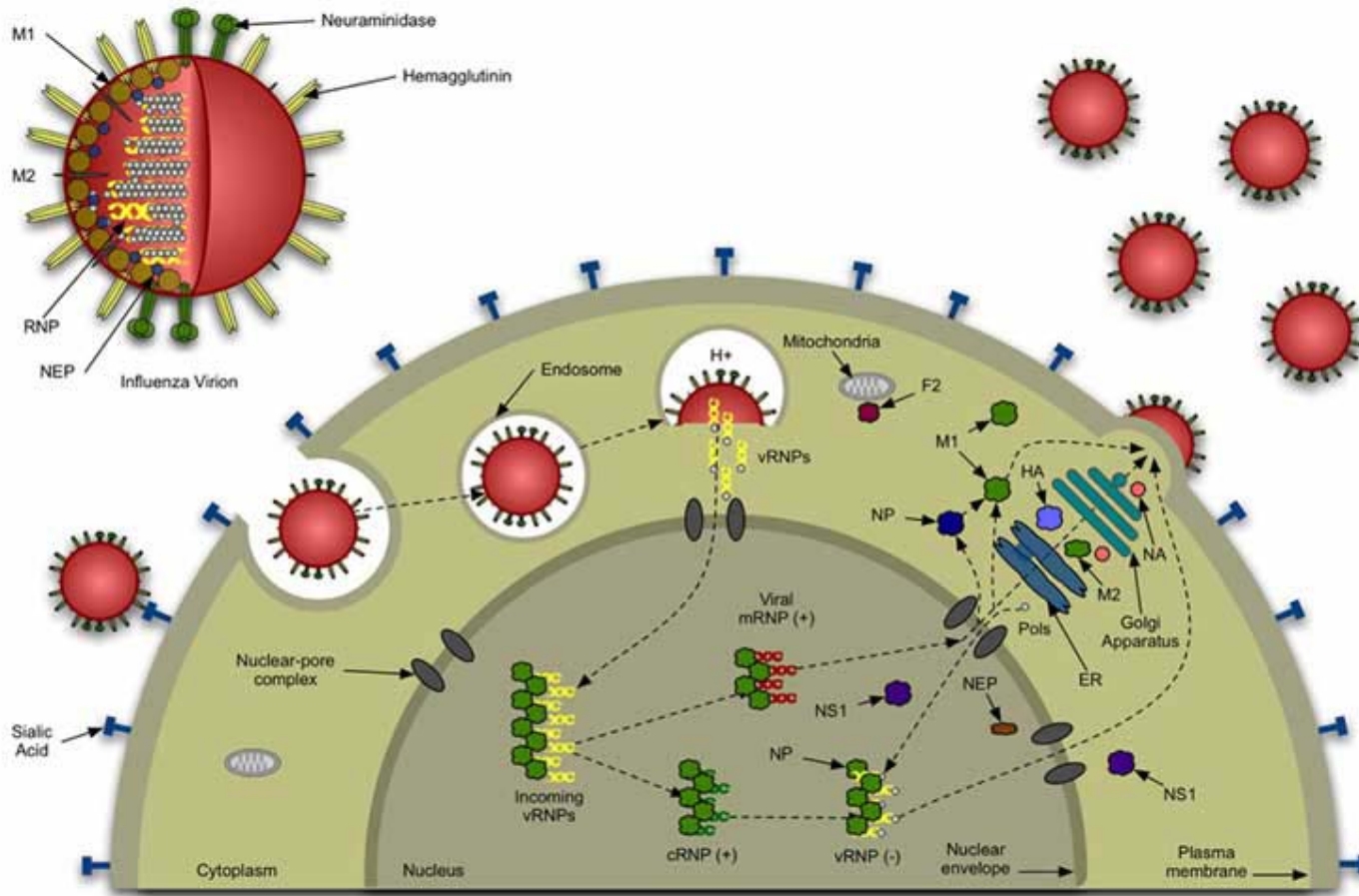
Konkuk Glyco-M*Grid



NBCR CADD



Viral Replication Life Cycle



Backtracking the Virus

H1N1	1918
H2N2	1957-58
H5N1	1961
H3N2	1968-69
H1N1	1976
H1N2	1980
H5N2	1983-84
H5N2	1992
H5N1	1997
H9N2	1999
H3N2	2001-02
H5N1	2004-05

Tracing The Reassortment

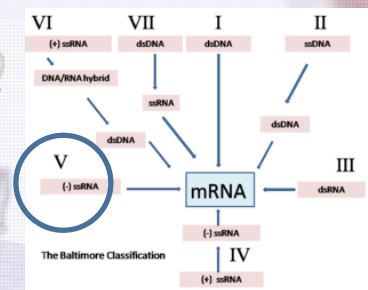
-H1N1, H2N2, H3N2 circulate extensively in humans

-H2N2 (<H1N1) 1958 subtype Asian Flu derived from the reassortment of avian and human strain.

-H3N2 (<H2N2) 1968 Hong Kong Flu antigenic shift. IS derived from the reassortment of avian and human strain. It is a H2N2 + H3N2 cross species transmission

-H1N2 (<H3N2) 1980, No major epidemic 1999-2000, 2001-2002
H1N1 (<H3N2)

-H1N1 and H3N2 cocirculating

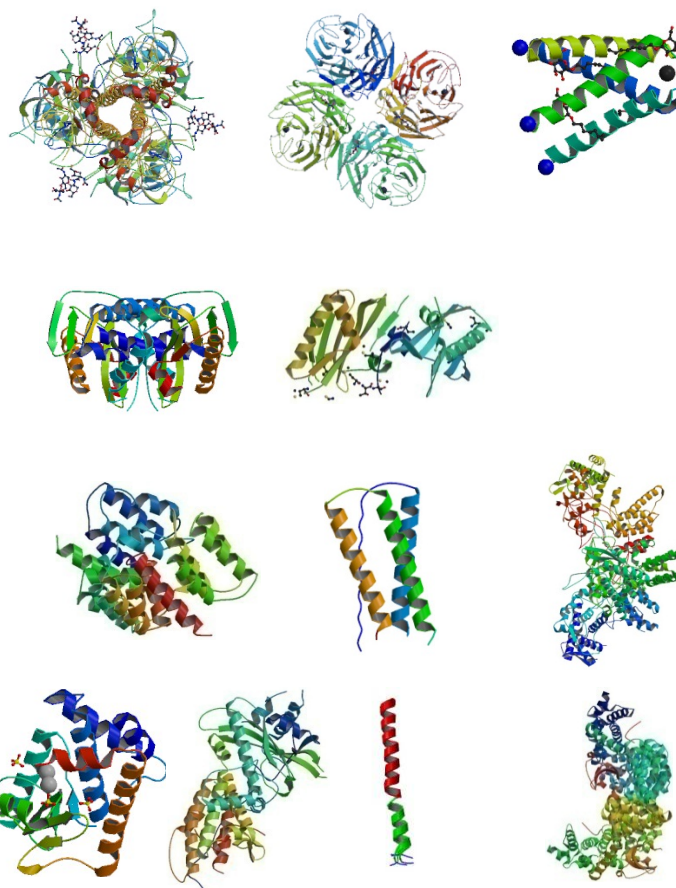


Baltimore viral classification

<http://www.reactome.org/>
<http://www.wikipedia.org>
<http://library.thinkquest.org/05aug/01479/prevention1.html>

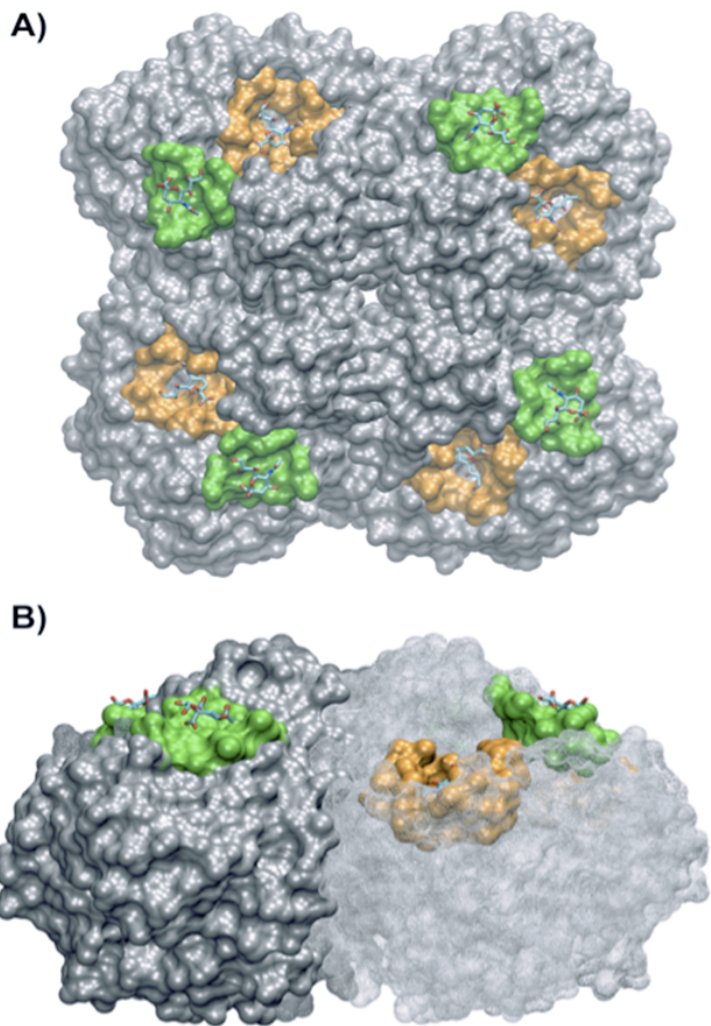
Influenza proteome and crystallome

Protein	Selected functions known	Crystal structural or NMR info	PDB ID and References
HA (Hemagglutinin)	Glycan receptor binding, membrane fusion	H5 trimeric complex	H5: 2FK0 ⁷ ; 2IBX ¹⁹ ;
NA (Neuraminidase)	Cleavage of SIA terminal residue, release of viral particles	N1 tetrameric complex	N1: 2HTY, 2HU0 ⁶ ;
M2 (proton channel)	Uncoating of viral envelope in endosome	NMR or crystal structure of transmembrane domain in complex with amantadine	M2 transmembrane domain: 2RLF ⁸ ; 3BKD ²⁰
NS1 (Nonstructural protein 1)	Host immune response modulation	RNA binding domain (RBD) and effector domain (ED) crystallized separately or in complex	NS1 RBD: 1AIL ²¹ ; NS1 ED: 2GX9, ²² H5N1 NS1: 3F5T ²³
M1 (matrix protein)	Formation of ribonucleoprotein complexes with viral RNA	Two domains separated by linker region	M1 N-terminal domain: 1AA7 ²⁴
NS2, NEP (Nonstructural protein 2, nuclear export protein)	Nuclear export of viral ribonucleoproteins	ED available	NS2 ED: 1PD3 ²⁵ ,
NP (nucleoprotein)	Formation of viral capsid and packaging of RNA	Full length	NP: 2IQH ²⁶
PA (acidic protein)	Endonuclease and cap snatching	N-terminal domain; C-terminal domain bound to PB2	PA N-terminal domain: 2W69 ²⁷ . PA C-terminal domain: 2ZNL ²⁸
PB1 (basic protein 1)	Polymerase catalytic subunit	In complex with PA	PB1 N-terminal domain: 2ZNL ²⁸
PB1-F2 (basic protein 1 frame 2)	Pro-apoptosis	NMR structure	PB1-F2: 2HN8 ²⁹
PB2 (basic protein 2)	Nuclear import of RNA; capped RNA recognition	NMR and crystal structure in complex with importin	PB2 C-terminal domain: 2JDQ ⁹ ; 3CW4 ³⁰



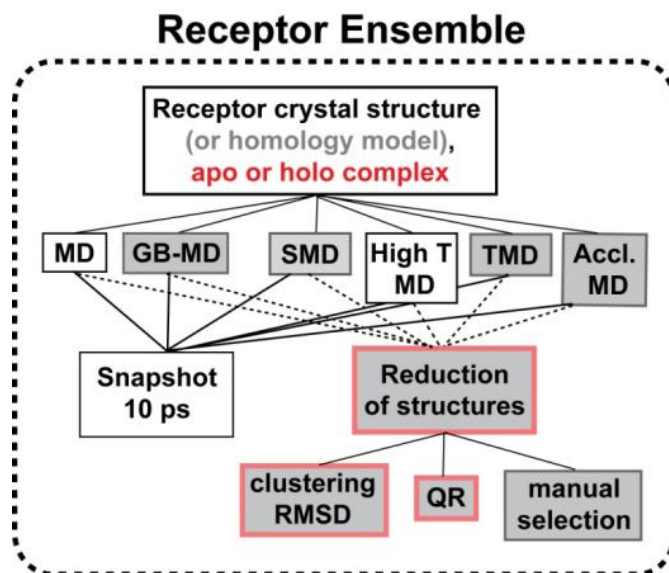
<http://www.pdb.org>

Known Drug Targets – Neuraminidase



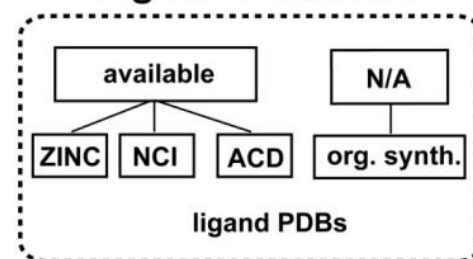
Ensemble-based Virtual Screening with Relaxed Complex Scheme

NAMD2
Amber



NCI Diversity Set: 3.3 MB, 2000 compounds;
Required at each site

ZINC subset: 200,000. A few hundred MB
Ligand Ensemble

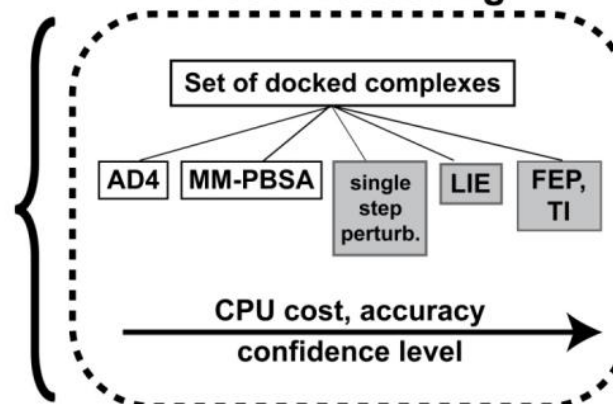


AutoDock4

AutoDock
(full ligand flexibility)

Docking Data: hundreds of MB

Post-Processing



Multiple targets: HA, NA subtypes
Each target: 30~50 MD snapshots, 1~2 MB each

Simulation Data: hundreds of GB

Experimental
verification

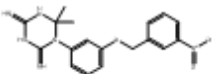
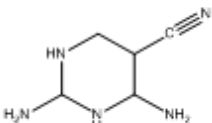
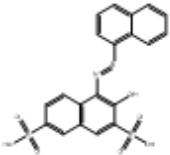
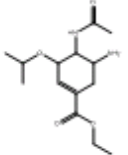
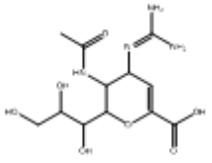
Known test set

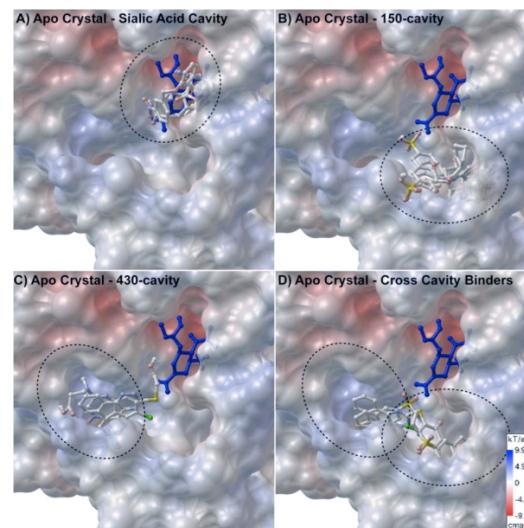
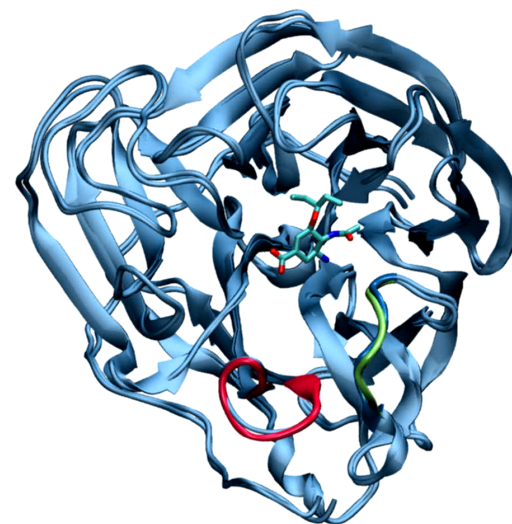
Total data to date: ~5 TB in long term storage.

Each experiment is about 1 Petaflops accumulative in computation cost.

Source: Amaro

NA Ensemble based Virtual Screening Top Hits

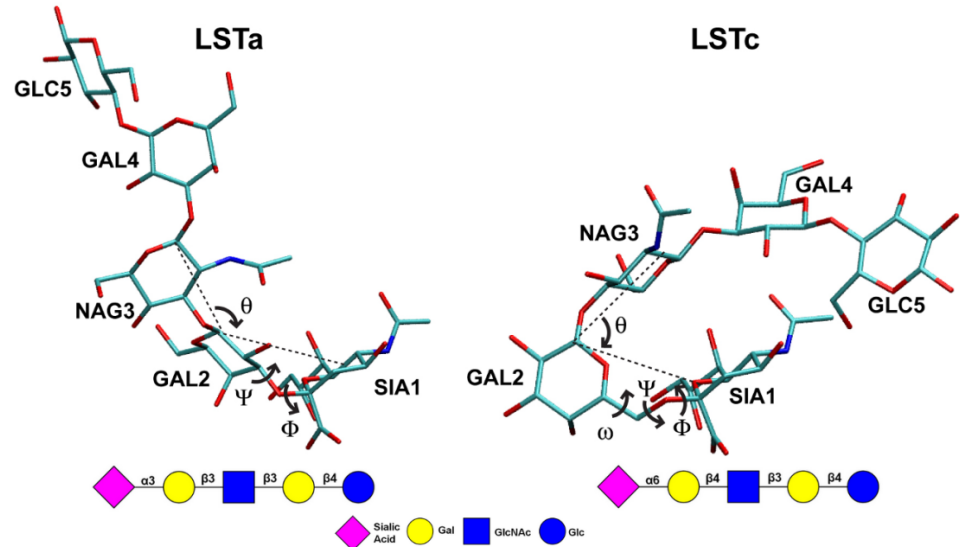
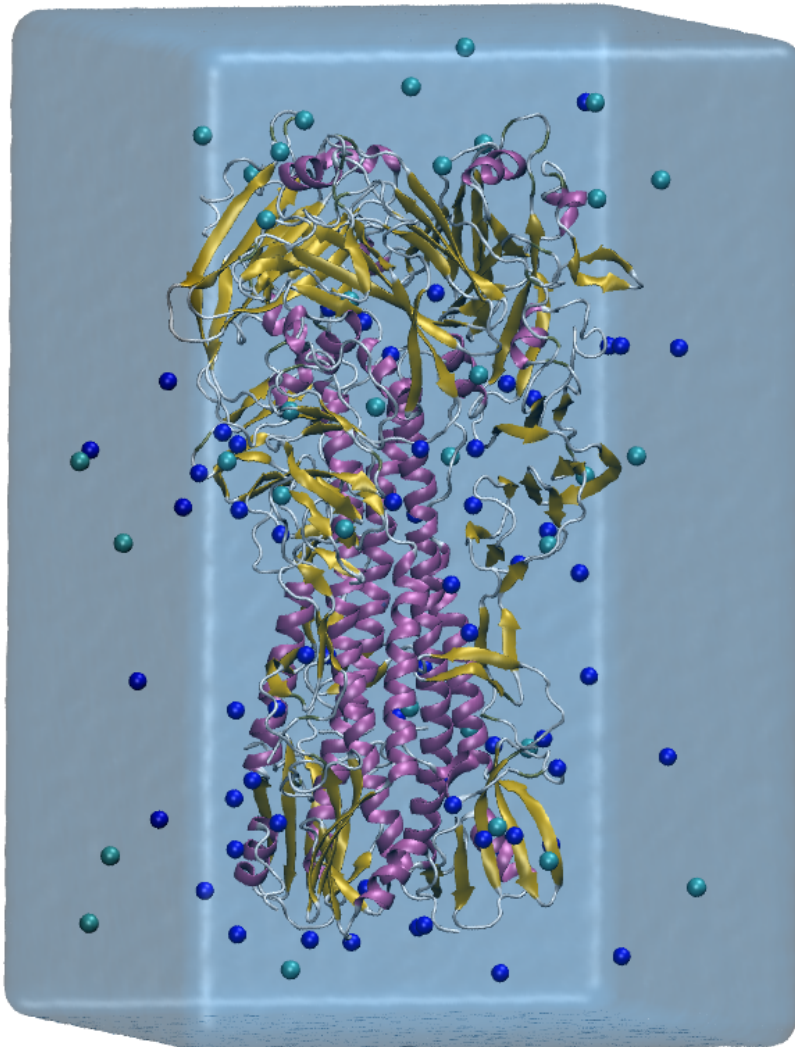
Rank	NSC	Mean Energy	Predicted K_i (μ M)	Chemical Structure	Binding Site	Apo Crystal Rank	Holo Crystal Rank
1	109836	-10.63	0.016		SA-cavity	15	1
2	211332	-10.34	0.026		SA-cavity	212	10
3	45583	-10.09	0.040		SA-cavity 150-cavity 430-cavity	6	18
-	Oseltamivir	-9.82	0.063 (0.3 – 1.0)		SA-cavity	238	5
-	Zanamivir	-9.38	0.133 (0.5 – 2.5)		SA-cavity	230	12



- Patent filed
- Cheng et al, JMC, 2008

How avian or swine viruses cross over to humans?

Comparative Molecular Dynamics Simulations of HA's



HA Trimers: avian H3 (1968 HK flu), H5 (Pandemic threat) and swine H9 (1999 HK flu)

apo and LSTa/LSTc bound/unbound

Explicit solvent, 0.15M NaCl, pH 7.4

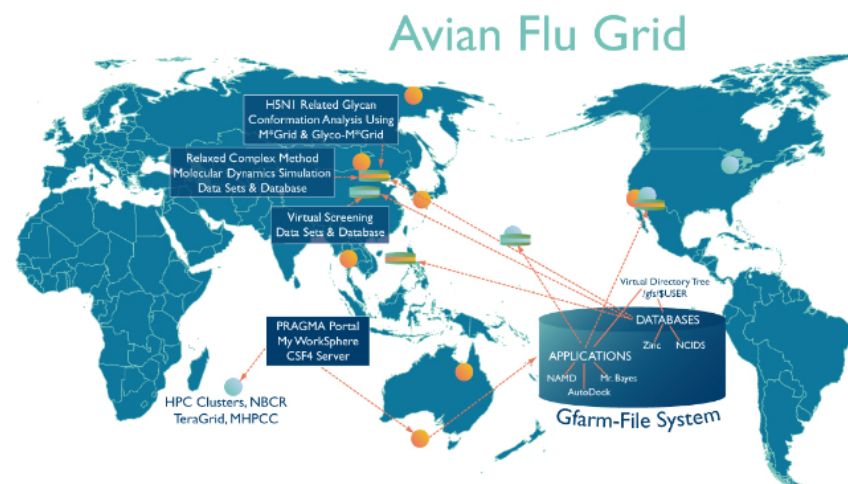
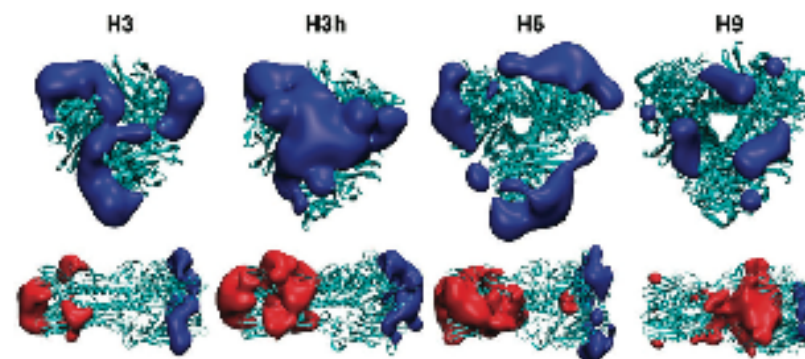
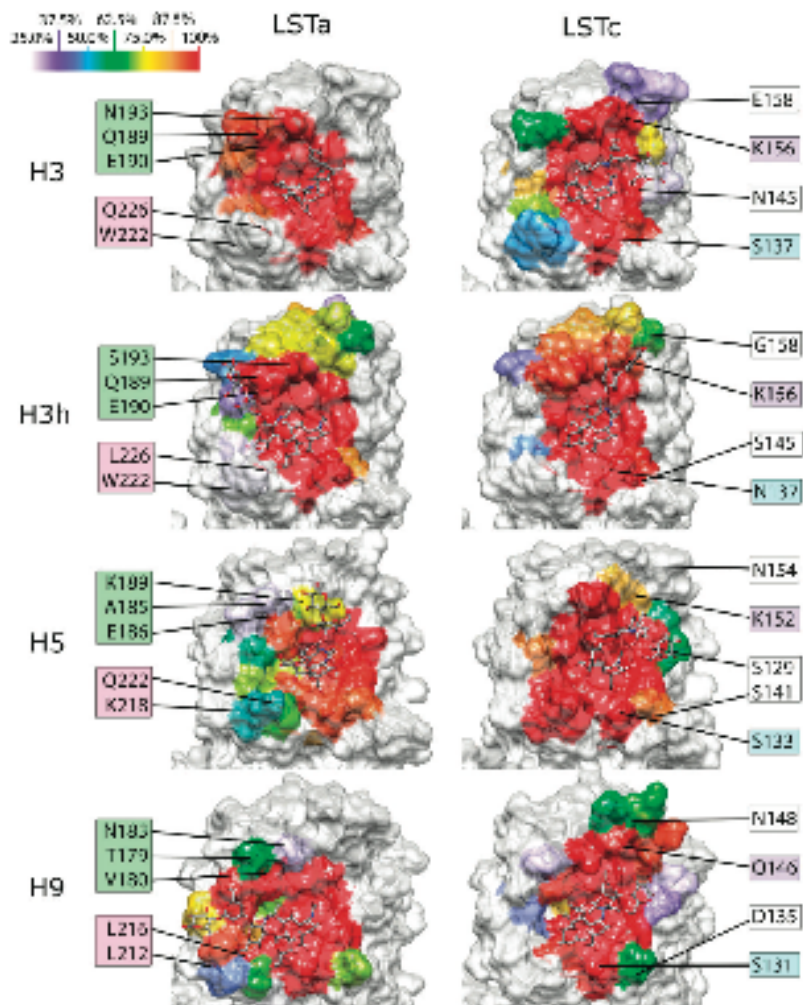
~350,000 atoms

NAMD2, Amber99SB/Glycam06 force field

1 fs timestep, PBC, PME

Xu, et al. JMB, 2009.

Scientific Results



Transparent access of applications on Avian Flu Grid through middleware

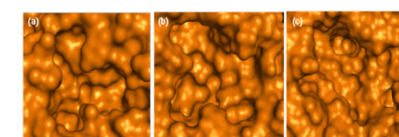
H5N1 related glycan conformation analysis using M*Grid and Glyco-M*Grid

Relaxed Complex Method Molecular Dynamics Simulation Data Sets & Database

Virtual Screening Data Sets & Database

ID	Owner	Description	Date	Exp	Result	Details
1	Asn88	DATA_200y	Apr 22, 2008 8:00:09 PM	(link)	(link)	(link)
2	Asn88	DATA_200y	Apr 22, 2008 8:00:11 PM	(link)	(link)	(link)
3	Asn88	DATA_200y	Apr 22, 2008 8:00:12 PM	(link)	(link)	(link)
4	Asn88	DATA_200y	Apr 22, 2008 8:00:13 PM	(link)	(link)	(link)
5	Asn88	DATA_200y	Apr 22, 2008 8:00:14 PM	(link)	(link)	(link)
6	Asn88	DATA_200y	Apr 22, 2008 8:00:15 PM	(link)	(link)	(link)
7	Asn88	DATA_200y	Apr 22, 2008 8:00:16 PM	(link)	(link)	(link)
8	Asn88	DATA_200y	Apr 22, 2008 8:00:17 PM	(link)	(link)	(link)
9	Asn88	DATA_200y	Apr 22, 2008 8:00:18 PM	(link)	(link)	(link)
10	Asn88	DATA_200y	Apr 22, 2008 8:00:19 PM	(link)	(link)	(link)

CNIC VSDB



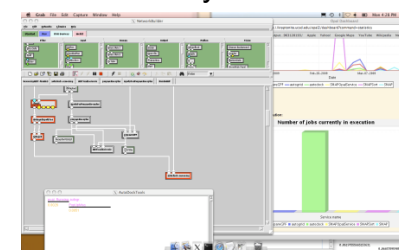
Crystal

MD-CHARMM

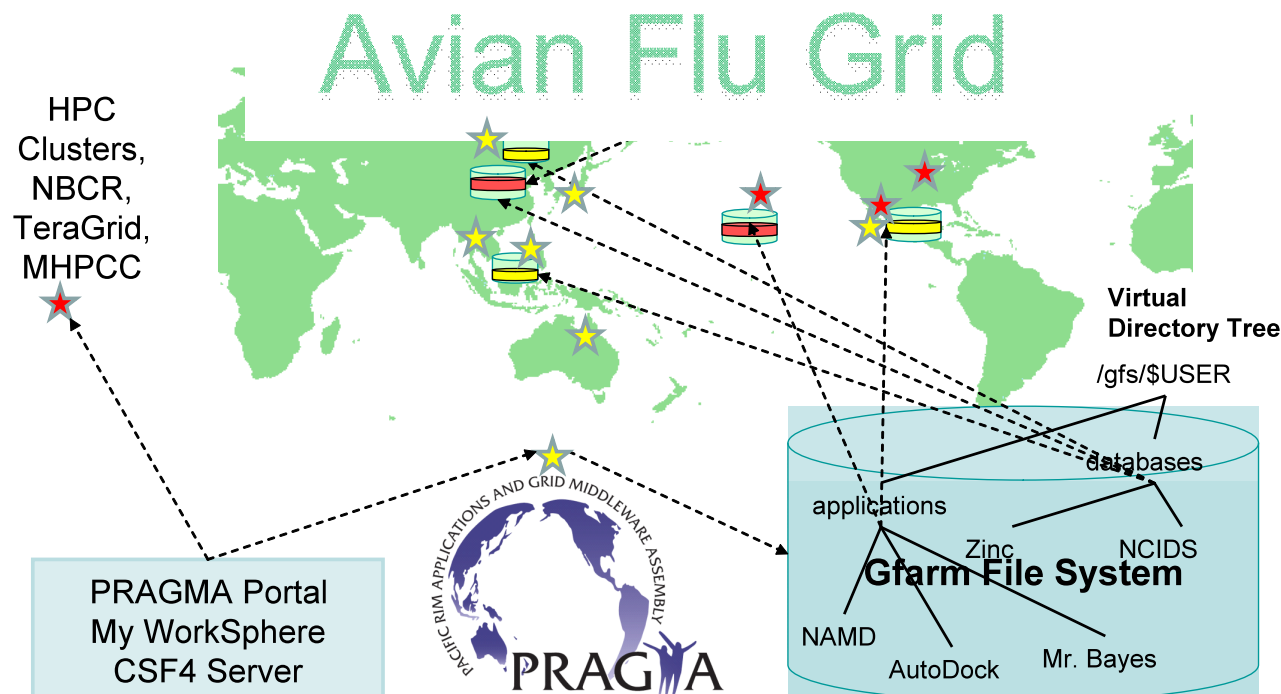
MD-NAMD



Konkuk Glyco-M*Grid



NCBR CADD



Recap of PRAGMA 16

– Daejeon, Korea, Mar 2009

- Infectious Diseases Research in Cyberinfrastructure (iDRiC Workshop)--KISTI
- Drugscreeener-G – KISTI, Korea
- Grid Enabled Virtual Screening Service (GVS) – ASGC, Taiwan
- CADD Pipeline – NBCR, USA
- WISDOM project – CNRS, EU
- Glyco-M*Grid – Kookmin & Konkuk U, Korea

Follow-up activities

- Continued development of application specific services – AutoDock, NAMD as Opal services
- Continued development of workflows – NBCR Summer Institute training of CADD pipeline
- Glyco-M*Grid – Suntae Hwang
- Release of Opal 2.1 with condor and CSF4 4.0.5.1 plugins
- Ability to use Condor web service interface for virtual screening using TeraGrid

Recap of PRAGMA 17

--Hanoi, Vietnam, Oct 2009

- Real-time bioMEdical data Streaming and visualization (RIMES) – CNIC, China
- PRIME Student Host Sites
 - USM, NTU, CNIC
 - Osaka U., Monash U.
 - New in 2010: JLU, host of PRAGMA 19
- MURPA Students
 - Cheminformatics and Nimrod/K based workflows
- Calit2 Summer Internship
 - Kevin Wu

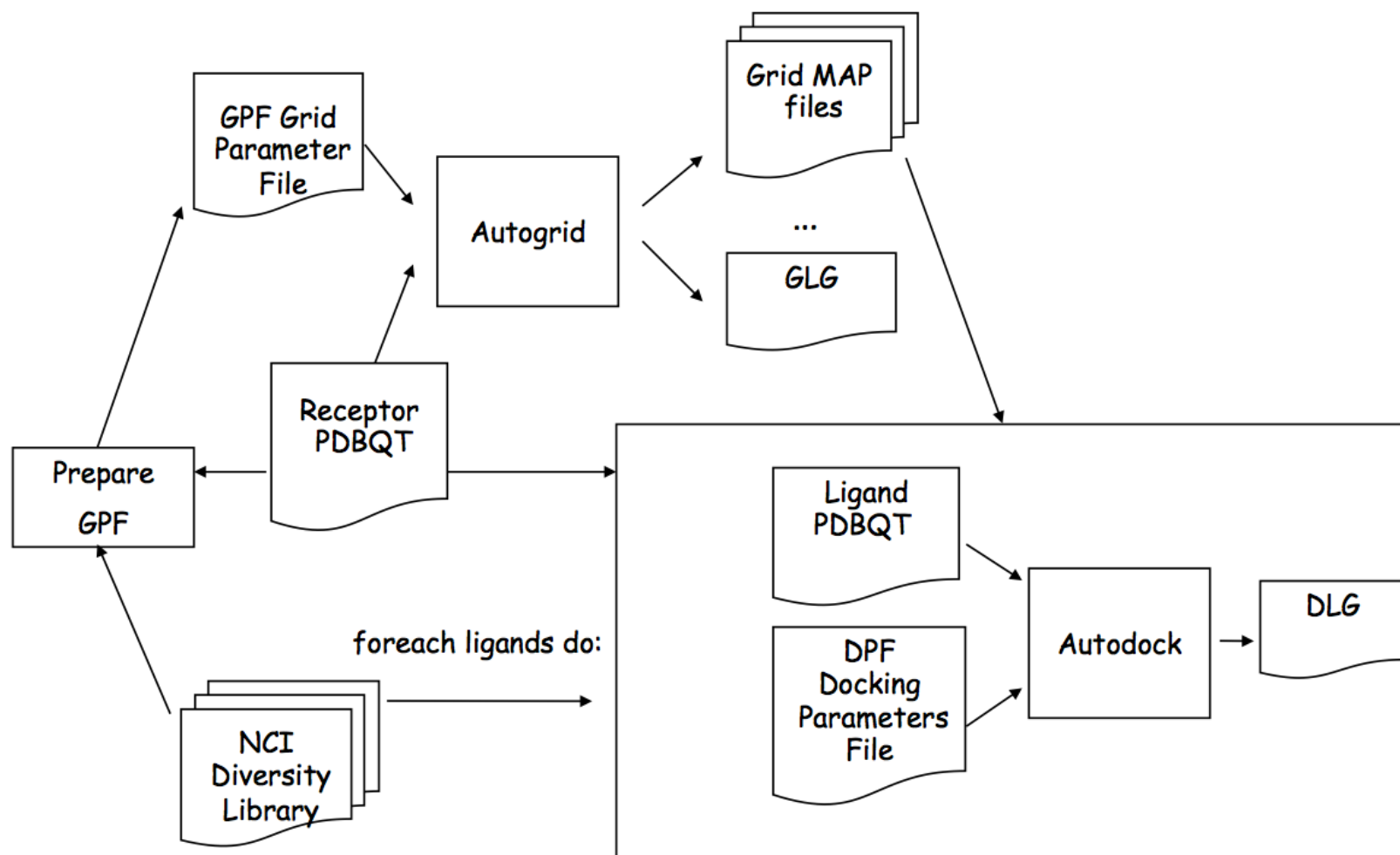
Next Milestones

- Production use of Gfarm for sharing simulation data
 - Production use by PRAGMA 18
- Virtual machine scheduling using CSF4
 - Demonstration by PRAGMA 19
- Expanded Education and Training Opportunities
 - MURPA; PRIME
 - ISGC workshops; PRAGMA Institutes

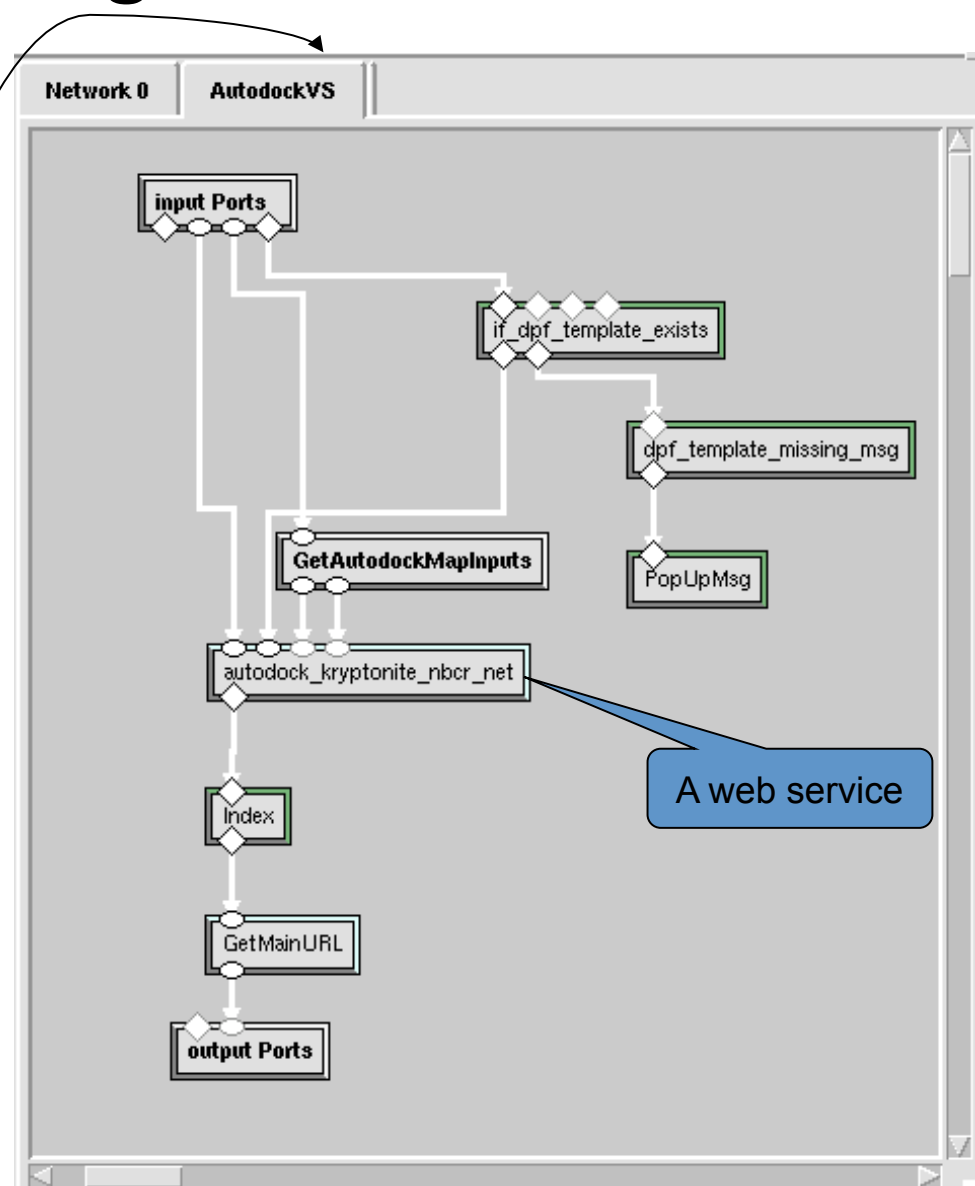
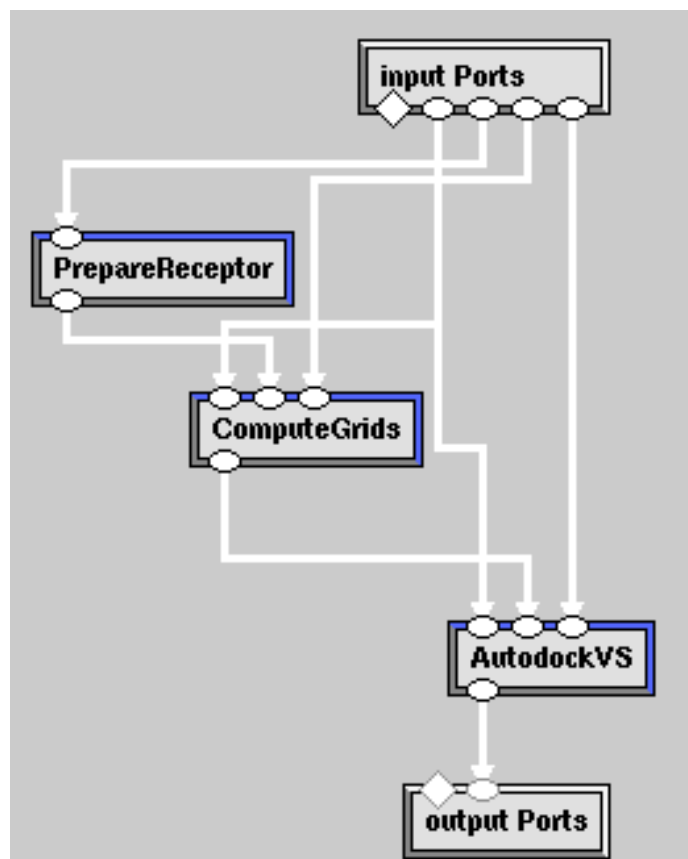
New Challenges

- Virtualization – What does it mean to us?
- Production environment – Where is it? What form should it take?
- The Connection between Productivity and Grid/Cloud – Most work is still done on local clusters, the desire to use the grid/Cloud is there, but the infrastructure is still evolving
- Sociology of software engineering – Can we collaborate using Facebook?

AutoDock Workflow

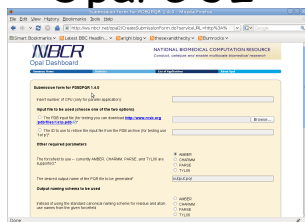


A Virtual Screening Vision Workflow

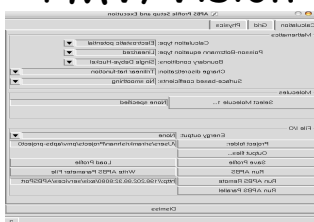


Transparent Access Layer for Applications

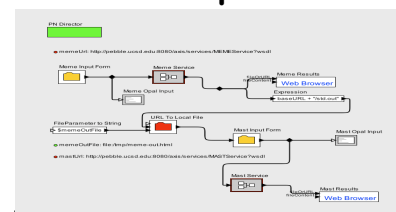
Opal GUI



PMV/Vision



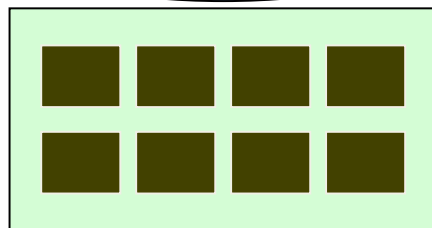
Kepler



Application Services

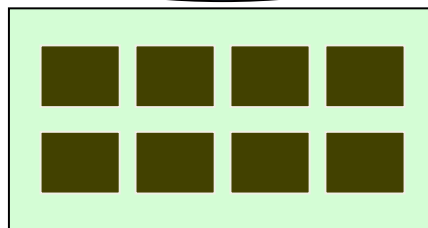
Grid/Cloud Resources

Globus



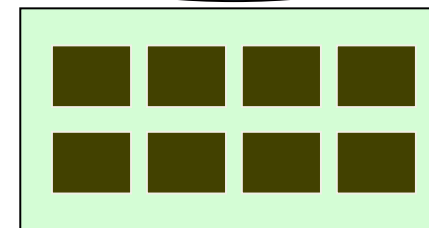
Condor pool

Globus



SGE Cluster

Globus



PBS Cluster

NBCR PDB2PQR Advanced Submission



Submission form for PDB2PQR 1.4.0 - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://ws.nbcrc.net/opal2/CreateSubmissionForm.do?serviceURL=http%3A%2F%2Fws.nbcrc.net/opal2/CreateSubmissionForm.do

Smart Bookmarks Latest BBC Headlin... arighi blog thesexandthecity Bumrocks

NBCR
Opal Dashboard

NATIONAL BIOMEDICAL COMPUTATION RESOURCE
Conduct, catalyze and enable multiscale biomedical research

Summary Home Statistics **Submissions** About Opal

Submission form for PDB2PQR 1.4.0

Insert number of CPU (only for parallel application):

Input file to be used (choose one of the two options)

☐ The PDB input file (for testing you can download <http://www.rcsb.org/pdb/files/1a1p.pdb>)*

☐ The ID to use to retrieve the input file from the PDB archive (for testing use '1a1p')*

Other required parameters

The forcefield to use -- currently AMBER, CHARMM, PARSE, and TYL06 are supported.*

The desired output name of the PQR file to be generated*

Output naming schema to be used

Instead of using the standard canonical naming scheme for residue and atom, use names from the given forcefield

☒ AMBER
☐ CHARMM
☐ PARSE
☐ TYL06

Done

Annotations:

- Exclusive group (points to the radio buttons for input file selection)
- String (points to the input field for the number of CPUs)
- Input file (points to the input field for the PDB file ID)
- Exclusive enumeration (points to the radio buttons for forcefield selection)

Group 1 (points to the input file selection options)

Group 2 (points to the forcefield selection options)

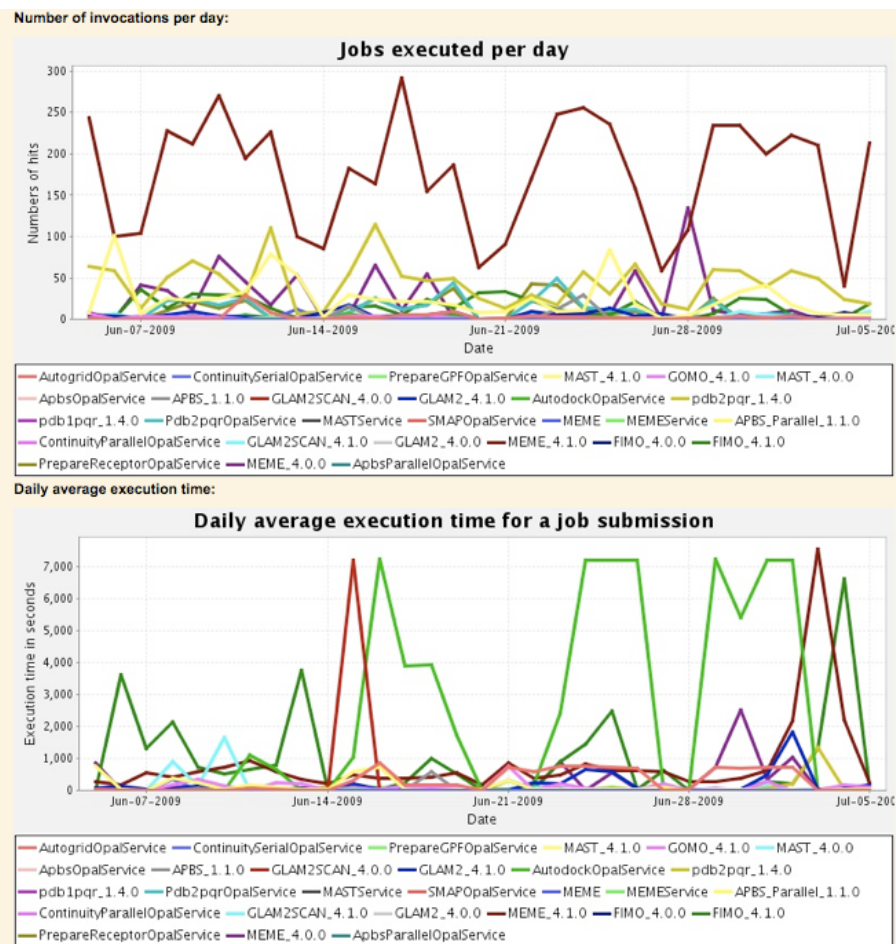
Group 3 (points to the output naming schema options)



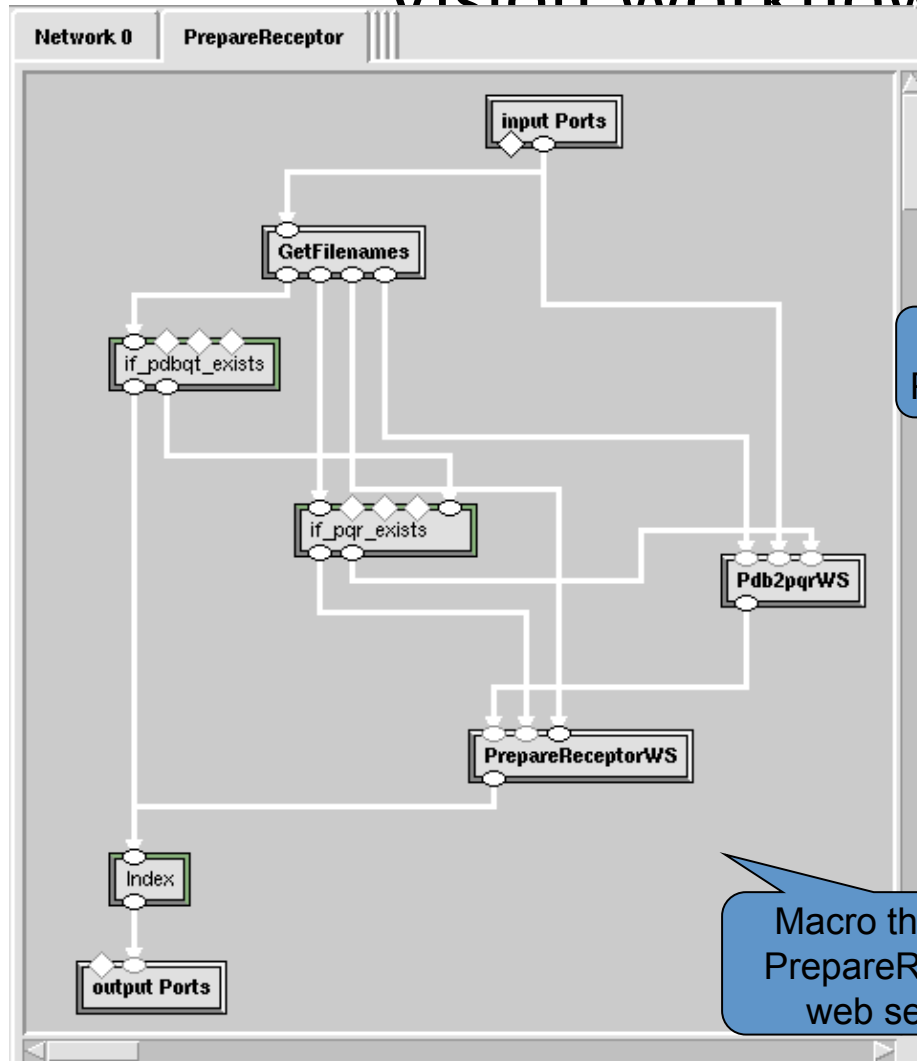
National Center for
Research Resources



Opal2 Dashboard: Usage Statistics



Vision Workflow Snippet Using Opal

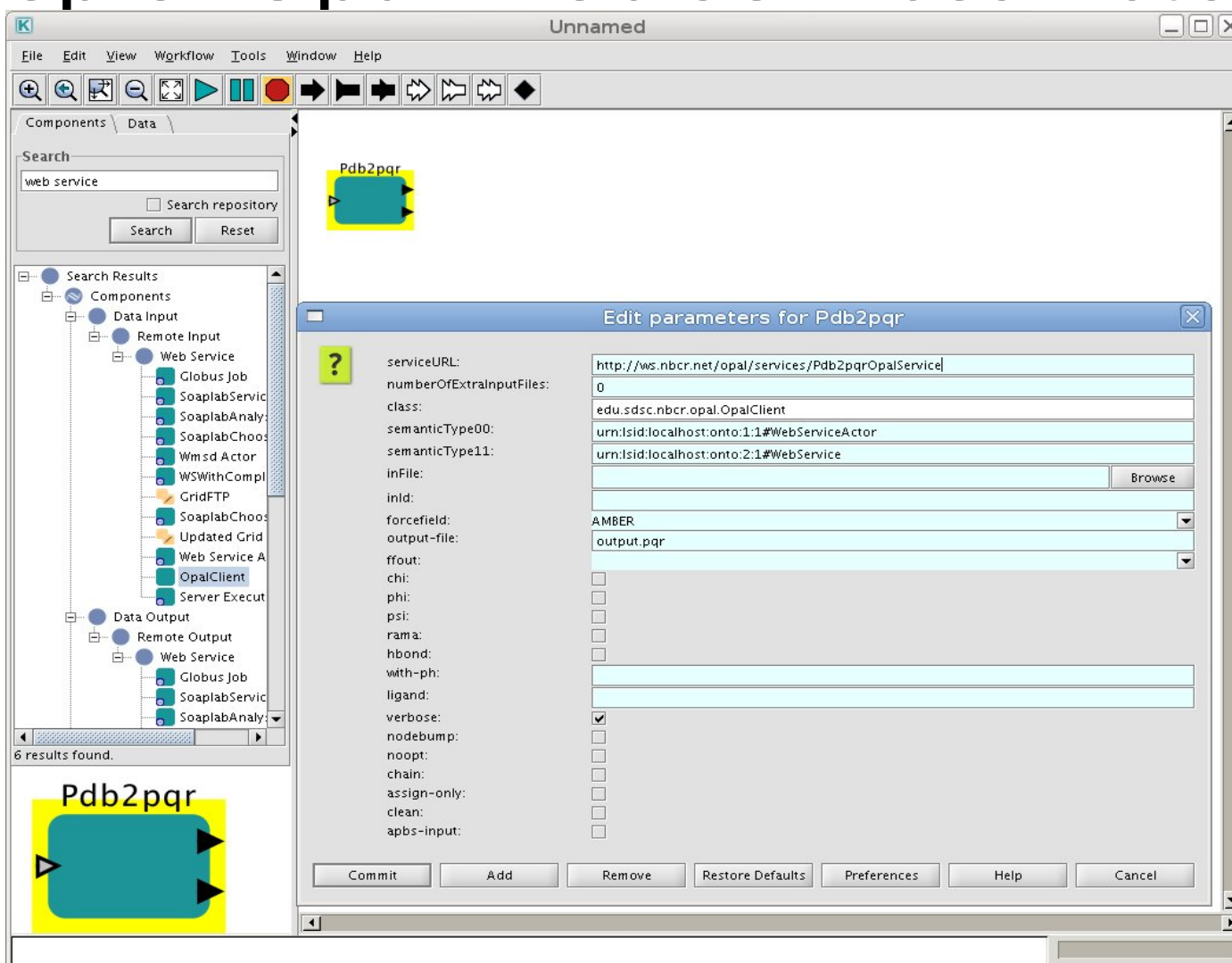


Macro that runs PDB2PQR web service.

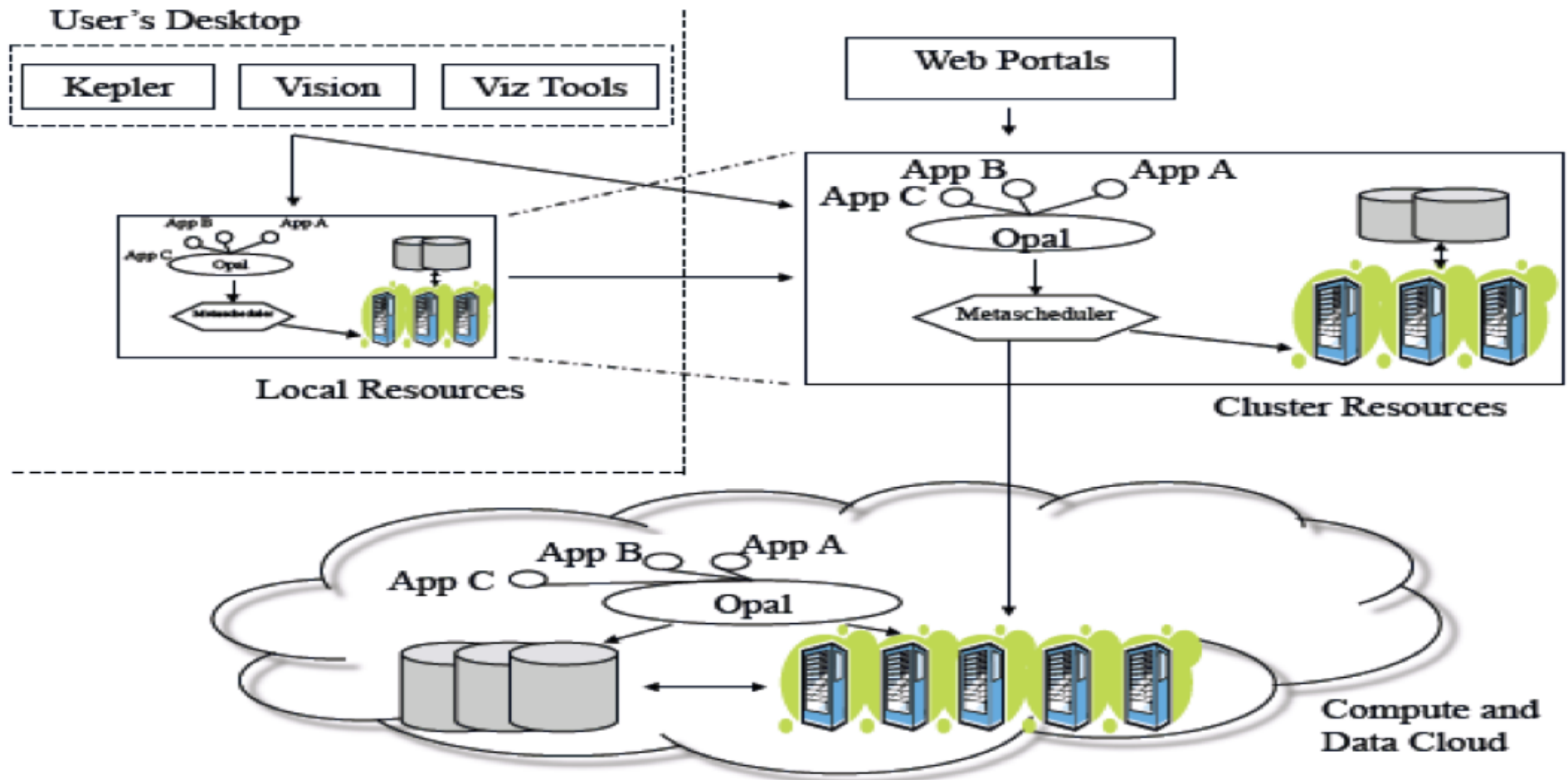
Macro that runs PrepareReceptor web service

- Two Major Steps
 1. Run PDB2PQR web service.
 - ✱ This step is skipped if an appropriate PQR file exists on the local machine.
 2. Run PrepareReceptor web service.
 - ✱ Output is URL to PDBQT
- PDB2PQR and PrepareReceptor are skipped if an appropriate PDBQT file exists on the local machine.
 - Output is PDBQT path on local machine.

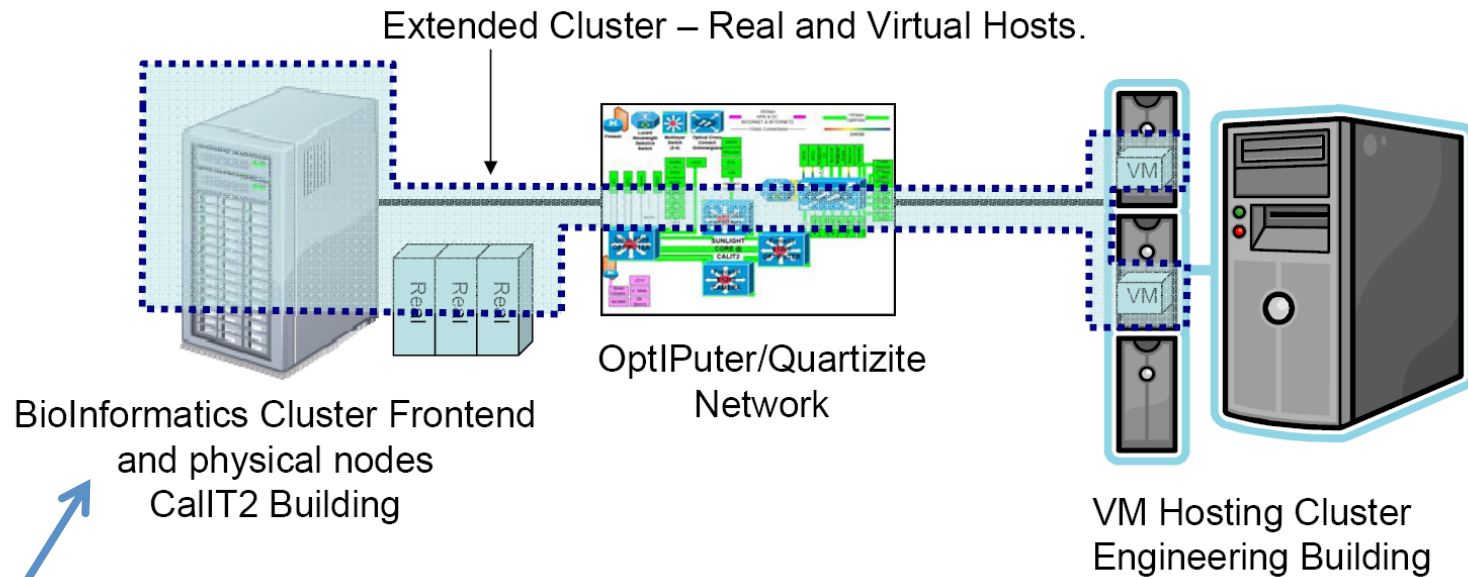
Kepler Opal Web Services Actor



Opal 2 for SaaS



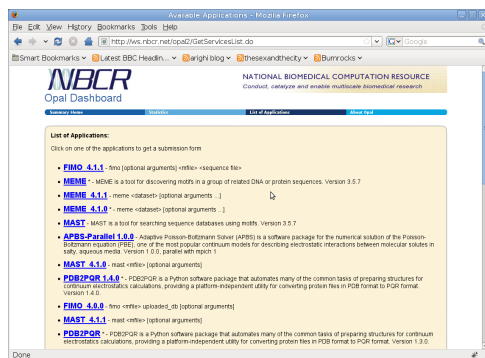
Campus Cloud: Cluster Extension



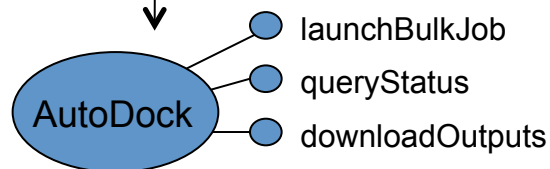
- VMs: Software/OS defined by the frontend:
 - Users, file system mount, queuing system, software versions, etc

Source: Papadopoulos

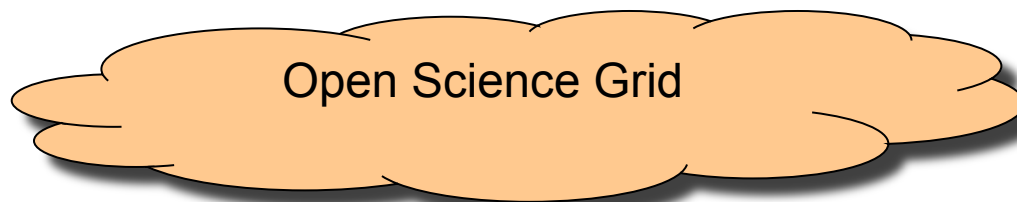
Use of OSG Resources – Collaboration with RENCI



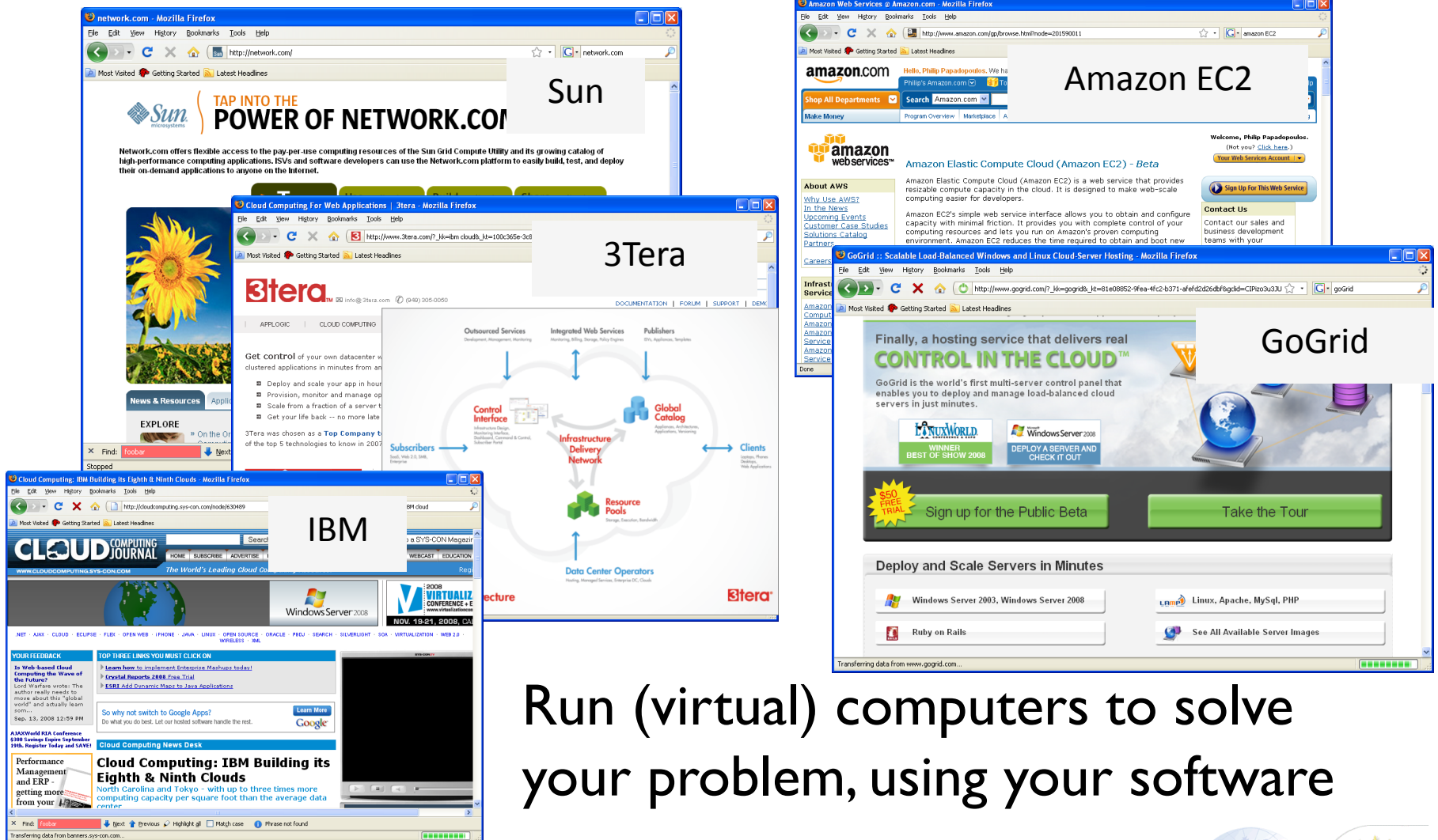
NBCR (UCSD)



RENCI (UNC)



IaaS – Infrastructure as a Service – Cloud with the most promise



The collage illustrates various IaaS providers and their offerings. The central diagram depicts a cloud architecture with the following components:

- Control Interface**: Manages design, provisioning, and control.
- Global Catalog**: Manages billing, storage, and policy.
- Infrastructure Delivery Network**: Manages design, provisioning, and control.
- Resource Pools**: Manages design, provisioning, and control.
- Data Center Operators**: Manage physical infrastructure.
- Outsourced Services**: Design, management, monitoring.
- Integrated Web Services**: Monitoring, billing, storage, policy, design.
- Publishers**: APIs, applications, templates.
- Subscribers**: Interact with the cloud.
- Clients**: Interact with the cloud.

Run (virtual) computers to solve your problem, using your software

Web Form for Virtual Screening Service

Submission form for AutoDock NCI Library Screening

Insert number of CPU (only for parallel application):

Scheduler

Scheduler to be used*

- ☐ SGE
☐ CSF

Required Autogrid Files (choose one of the following)

URL from where the autogrid files will be downloaded

(Ex 1. use <http://kryptonite.nbc.net/app1256780764236> with NCI_DS1)

(Ex 2. use <http://kryptonite.nbc.net/app1258313717575> with sample - only 6 ligands) *

Upload autogrid file tarball (in tar.gz format)*

Browse...

DPF

Upload your Template DPF

Browse...

Library (choose one of the following)

- ☐ sample
☐ NCIDS_SC
☐ NCI_DS1
☐ NCI_DS2
☐ human_metabolome

Existing library to be used for screening

Amazon EC2 Dashboard

Navigation

Region:  **US-East** ▼

- > **EC2 Dashboard**
- INSTANCES
 - > **Instances**
- IMAGES
 - > **AMIs**
 - > **Bundle Tasks**

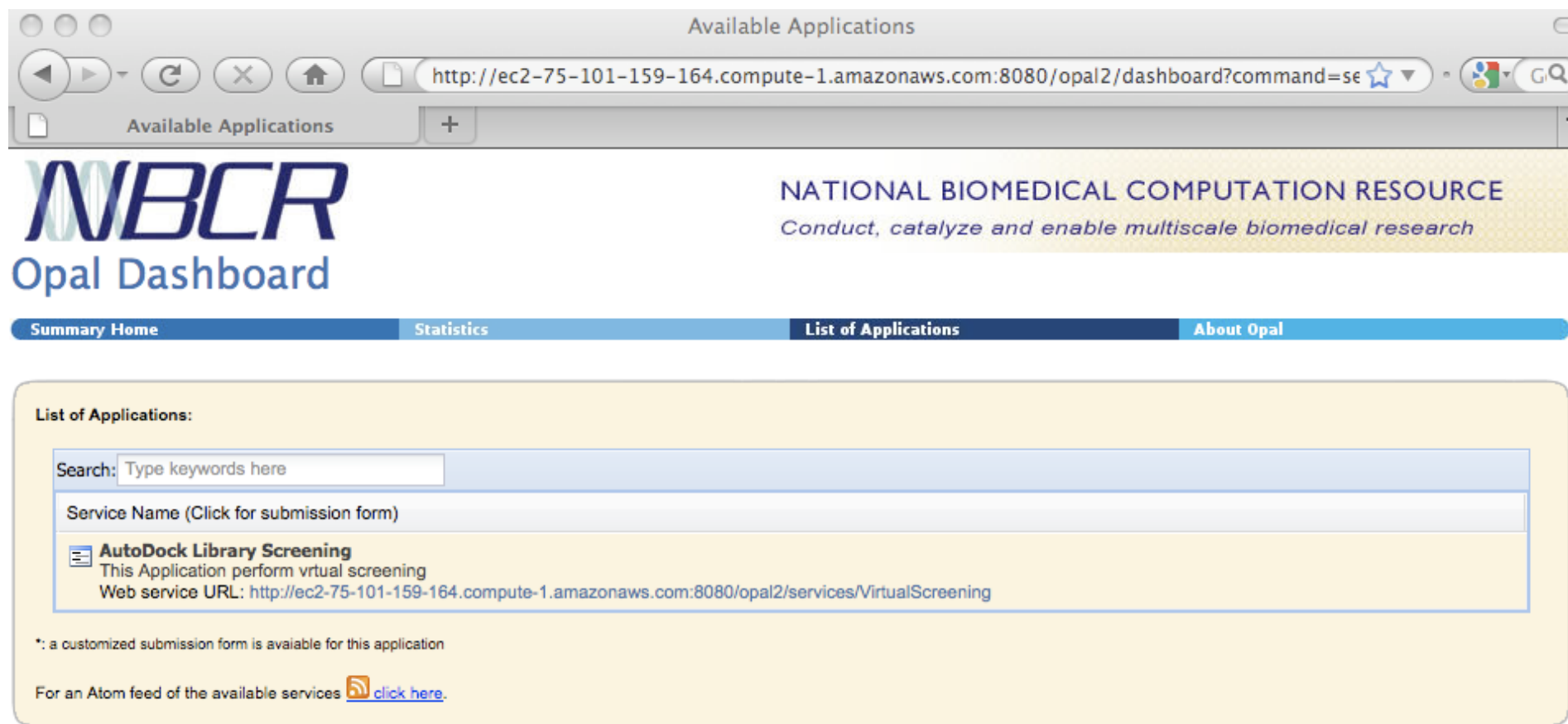
My Instances

 Launch Instance
 Instance Actions ▼
Reserved Instances ▼
Show/Hide
Refresh
Help

Viewing: All Instances 1 to 5 of 5 Instances

	Instance	AMI ID	Security Groups	Type	Status ▲	Public DNS
☐	 i-2b9c0643	ami-d90cedb0	open	m1.large	 running	ec2-75-101-
☐	 i-4b92f723	ami-d90cedb0	open	m1.large	 running	ec2-75-101-
☐	 i-bdea8fd5	ami-d90cedb0	open	m1.large	 running	ec2-75-101-
☐	 i-f383e49b	ami-d90cedb0	open	m1.large	 running	ec2-174-129
☐	 i-01fc9b69	ami-d90cedb0	open	m1.large	 running	ec2-174-129

Cloud Computing with Amazon EC2



The screenshot shows a web browser window titled "Available Applications" with the address bar displaying the URL: `http://ec2-75-101-159-164.compute-1.amazonaws.com:8080/opal2/dashboard?command=se`. The page content includes the NBCR logo and the text "Opal Dashboard". A navigation bar contains links for "Summary Home", "Statistics", "List of Applications" (which is highlighted), and "About Opal".

Under the "List of Applications" section, there is a search bar with the placeholder text "Search: Type keywords here". Below the search bar, a table lists available services:

Service Name (Click for submission form)
AutoDock Library Screening This Application perform virtual screening Web service URL: http://ec2-75-101-159-164.compute-1.amazonaws.com:8080/opal2/services/VirtualScreening

Below the table, a note states: "*: a customized submission form is available for this application". At the bottom, a link is provided: "For an Atom feed of the available services [click here](#)."

Virtual Screening with CSF

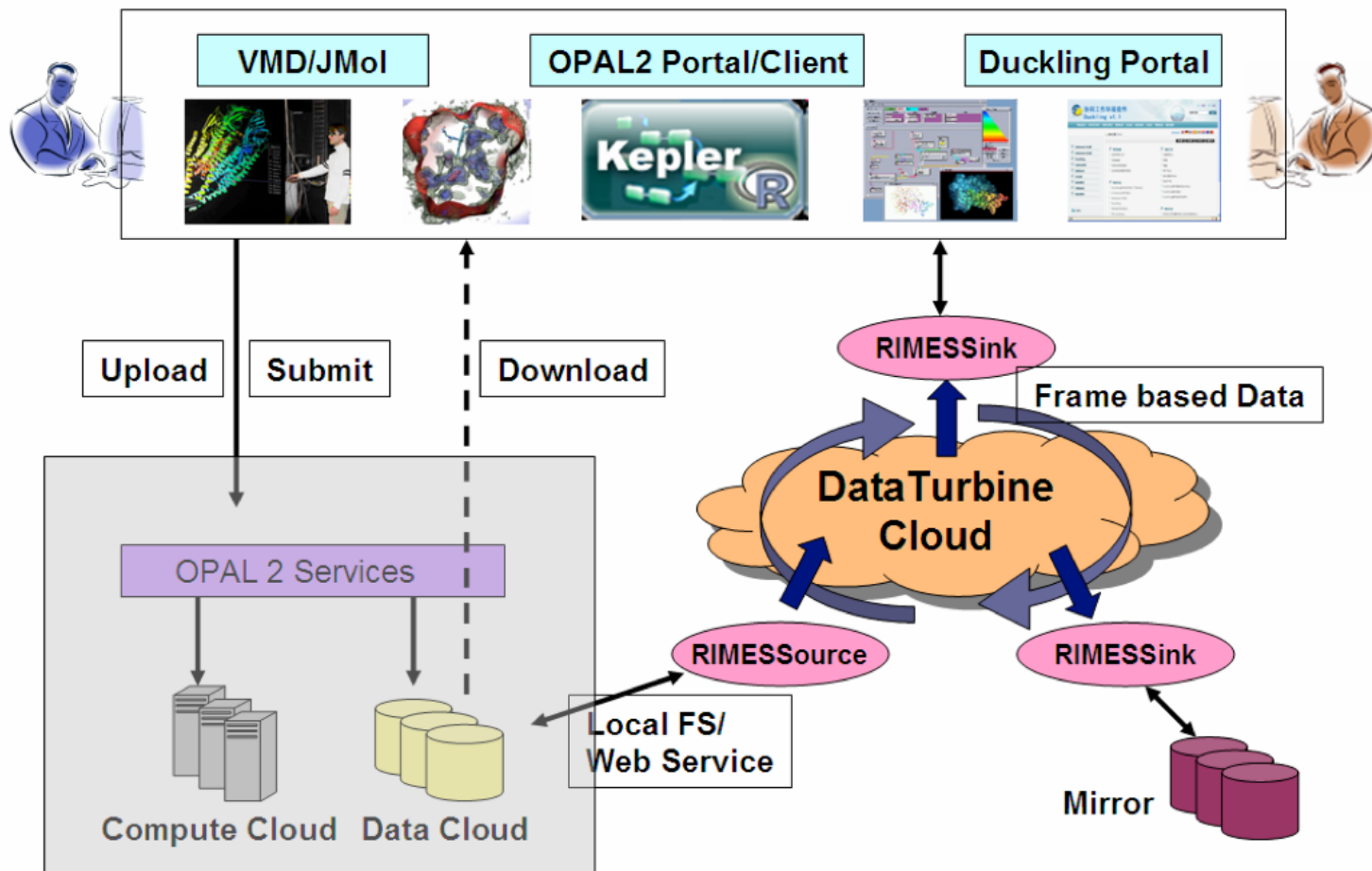
- We have set up the virtual screening web services with 5 remote hosts.

Virtual cluster at SDSC

AMAZON EC2

```
[nbcruser@rocks-32 ~]$ csf-application-list
getApplist() =>
APP NAME          INSTALLED ON          APP PATH
AUTODOCK_SLICE_PRE rocks-176.sdsc.edu     /home/nbcruser/screening/bin
AUTODOCK_SLICE_PRE rocks-153.sdsc.edu     /home/g-avian/screening/bin
AUTODOCK_SLICE_PRE sakura.hpcc.jp         /home/g-avian/screening/bin
AUTODOCK_SLICE_PRE rocks-52.sdsc.edu       /home/afguser/screening/bin
AUTODOCK_SLICE_PRE 75.101.159.164    /home/nbcruser/screening/bin
AUTODOCK_SLICE_POST rocks-176.sdsc.edu     /home/nbcruser/screening/bin
AUTODOCK_SLICE_POST rocks-153.sdsc.edu     /home/g-avian/screening/bin
AUTODOCK_SLICE_POST sakura.hpcc.jp         /home/g-avian/screening/bin
AUTODOCK_SLICE_POST rocks-52.sdsc.edu       /home/afguser/screening/bin
AUTODOCK_SLICE_POST 75.101.159.164    /home/nbcruser/screening/bin
AUTODOCK_SLICE     rocks-176.sdsc.edu     /home/nbcruser/screening/bin
AUTODOCK_SLICE     rocks-153.sdsc.edu     /home/g-avian/screening/bin
AUTODOCK_SLICE     sakura.hpcc.jp         /home/g-avian/screening/bin
AUTODOCK_SLICE     rocks-52.sdsc.edu     /home/afguser/screening/bin
AUTODOCK_SLICE     75.101.159.164     /home/nbcruser/screening/bin
```

Integrating Visualization Workflows using Real-time bioMEdical data Streaming and visualization (RIMES)

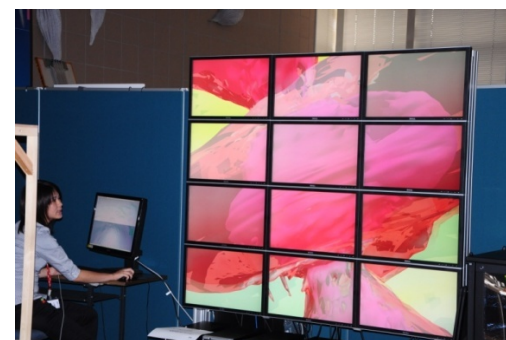


PRIME 2009

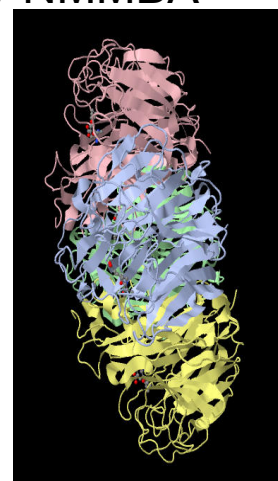
- 13 host sites
- Expanded to five sites
 - Doshisha University, Kyoto
 - National Institute of Information and Communication Technology (NiCT)
 - National Museum of Marine Biology and Aquarium (NMMBA), Kenting, Taiwan,
 - National Taiwan University (NTU),
 - University of Hyderabad
- Continued at eight sites
 - Monash U; NCHC; Osaka U; CNIC; NCREE; USM; U Auckland, U Waikato



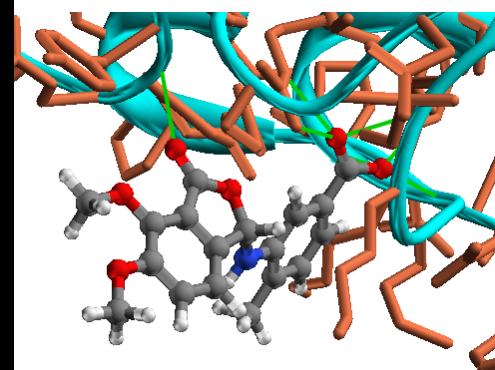
M.Nekrasov NMMBA and NCHC



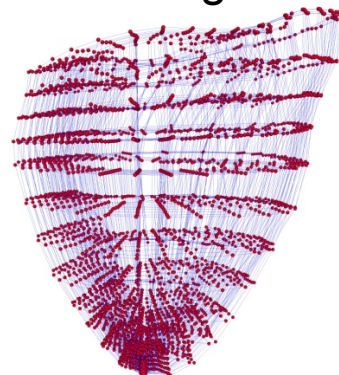
Jade Kwan, NiCT



C. Wong CNIC



M. Mui, U Hyderabad



C Lau, Osaka

Broadening Impact of Technology Engaging Future Generations



PRIME Student 2009: Jessica Hsieh, USM



1st Inaugural PRAGMA Workshop in San Diego

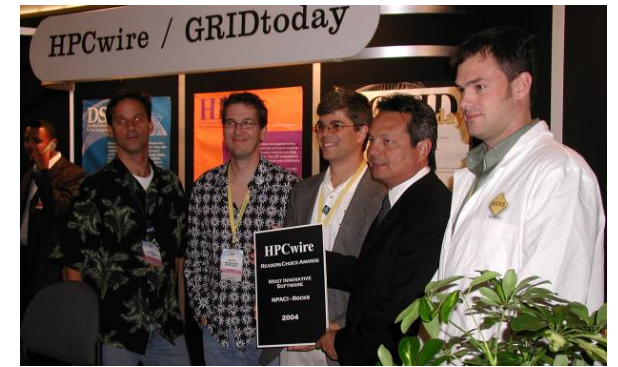
Welcome back to San Diego for PRAGMA 18!



Last time we were in
San Diego, CA, USA



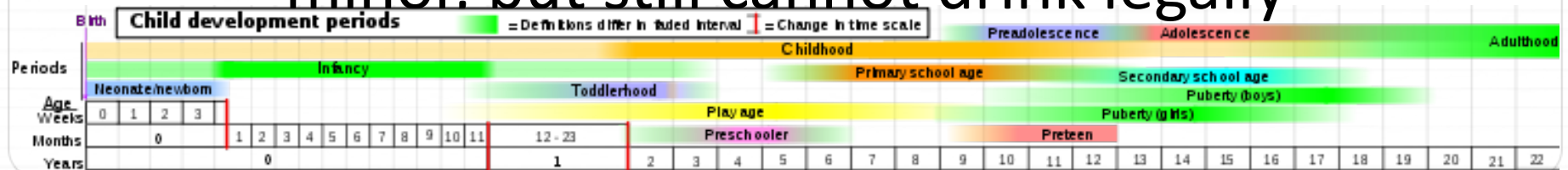
PRAGMA 7, Sept. 15-17, 2004





What does it mean to be 18?

It's about being able to vote 😊, no longer a minor, but still cannot drink legally





PRAGMA 18

Like the Grid

Like the Cloud



La Jolla, California – UC San Diego
March 2-5, 2010

March 2 – PRAGMA Institute on Implementation
& Welcome Reception

March 3-4 – PRAGMA Workshop

March 5 – Technology for Coral Reef Observatory

