Computer Aided Drug Discovery: Fishing in the Virtual Flood of *in silico* Data

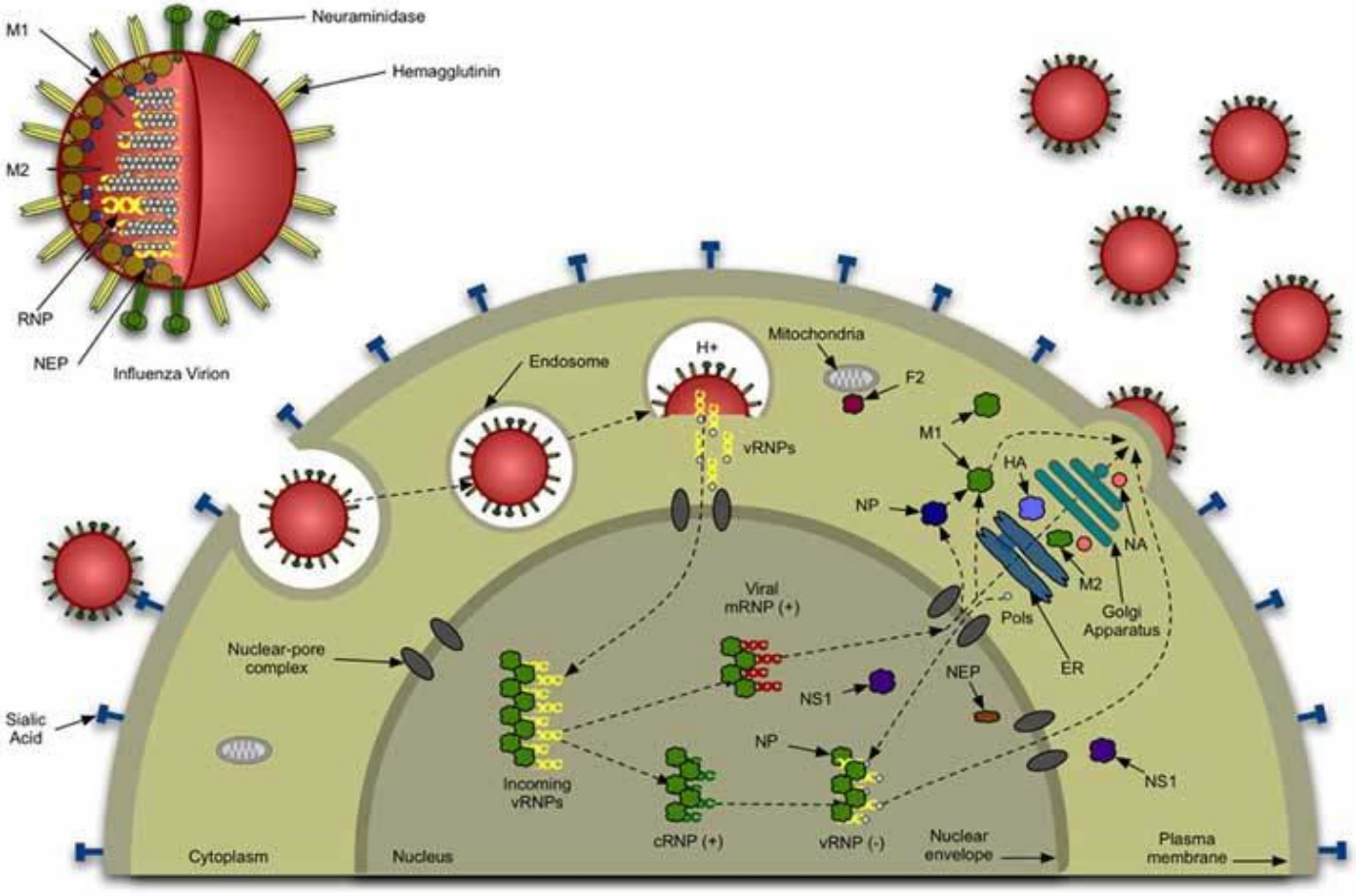
Wilfred W. Li, Ph.D.

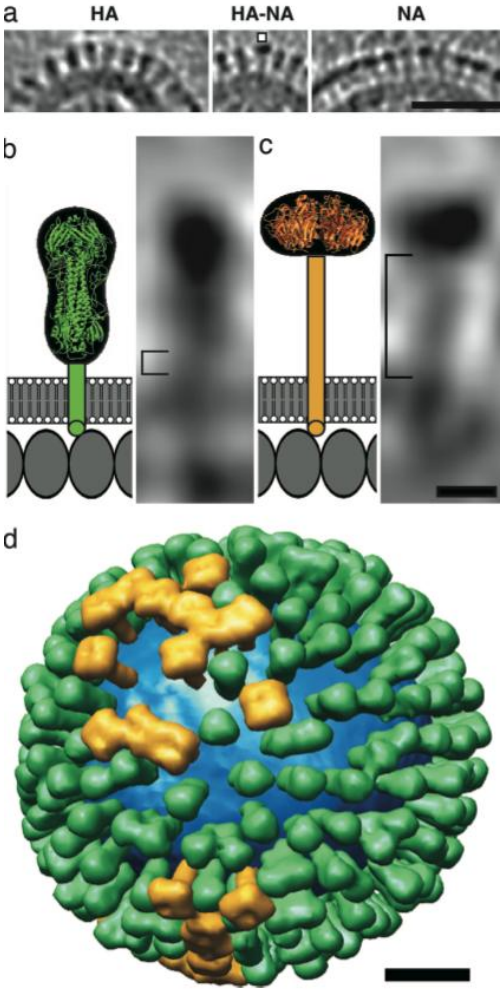
SEAIP 2010

蕙荪林场, Dec 9, 2010

National Biomedical Computation Resource
CRBS, CALIT2
San Diego Supercomputer Center
University of California, San Diego

Viral Replication Life Cycle



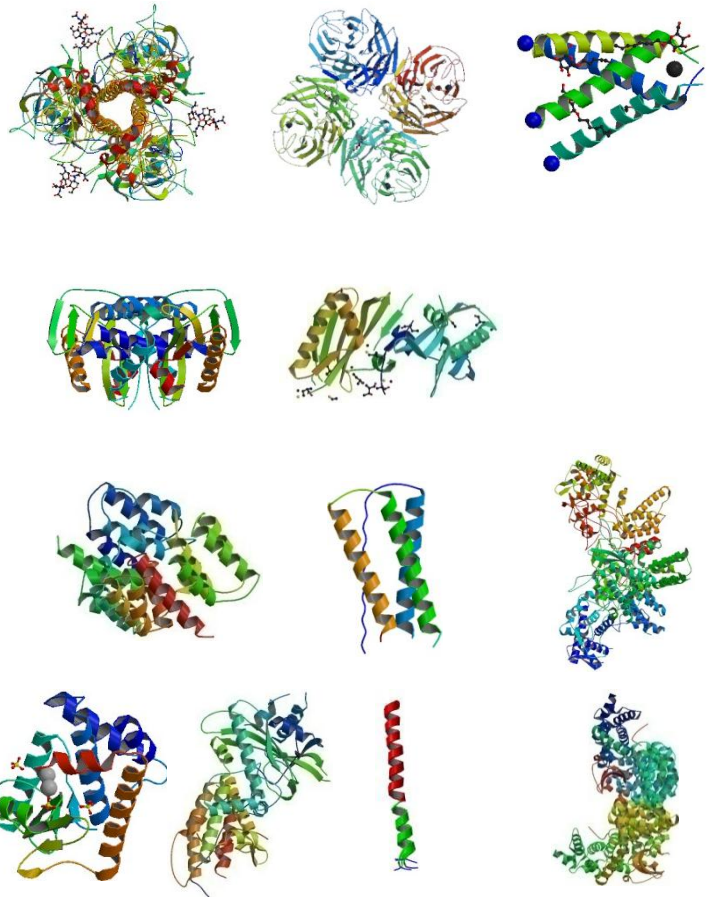


Harris et al, PNAS, 2006

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

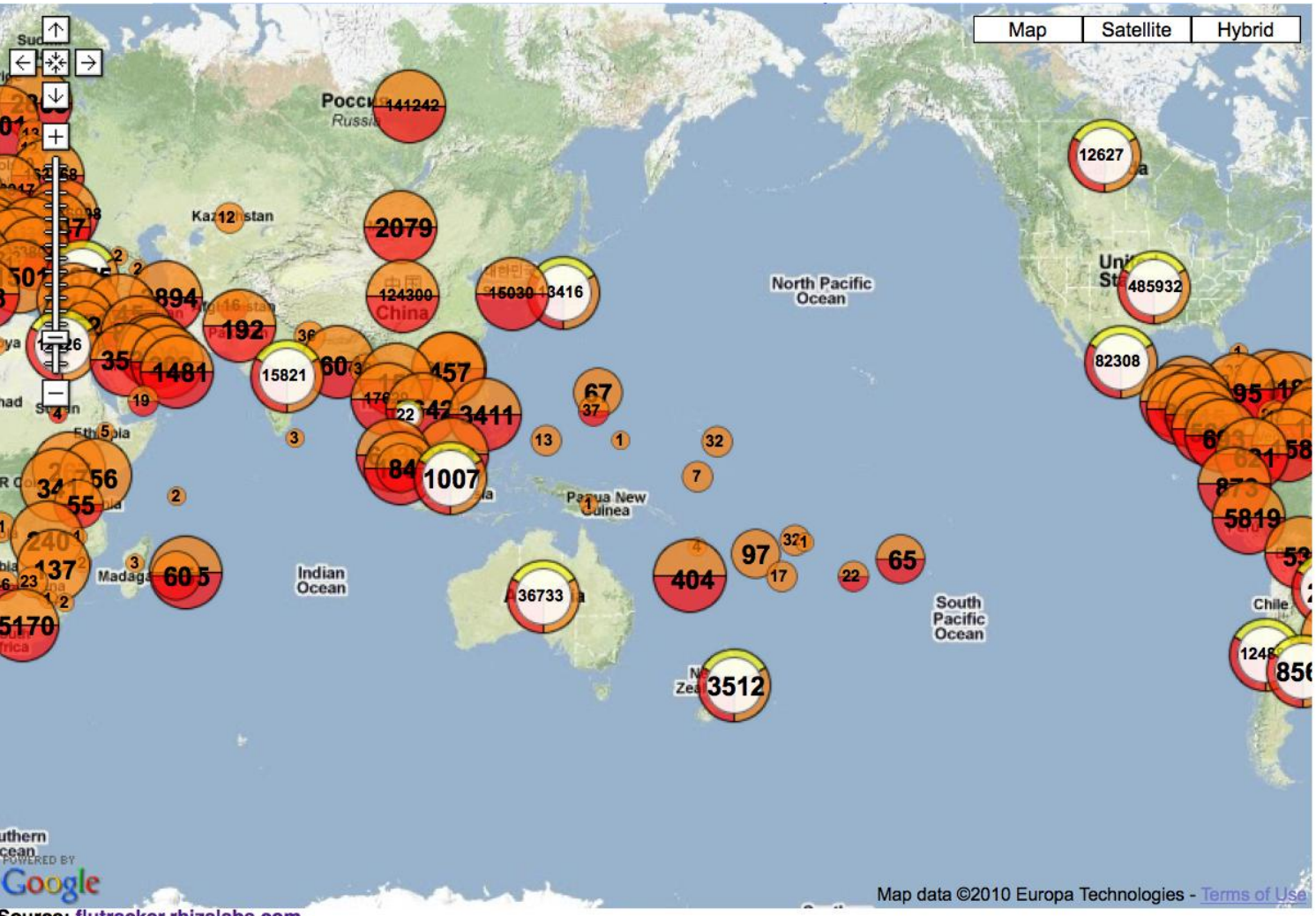
Influenza proteome and crystallome

Protein	Selected known functions	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 ribonuleoprotein 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

2009 H1N1 Pandemic Influenza



Source: flutracker.rhizalabs.com

- Suspected (9029)
- Confirmed/Probable (8753967)
- Fatal (14346)
- Suspected & Confirmed/Probable
- Suspected & Fatal
- Confirmed/Probable & Fatal
- Suspected, Confirmed/Probable & Fatal

Circle size relative to number of cases. Zoom for more detail.

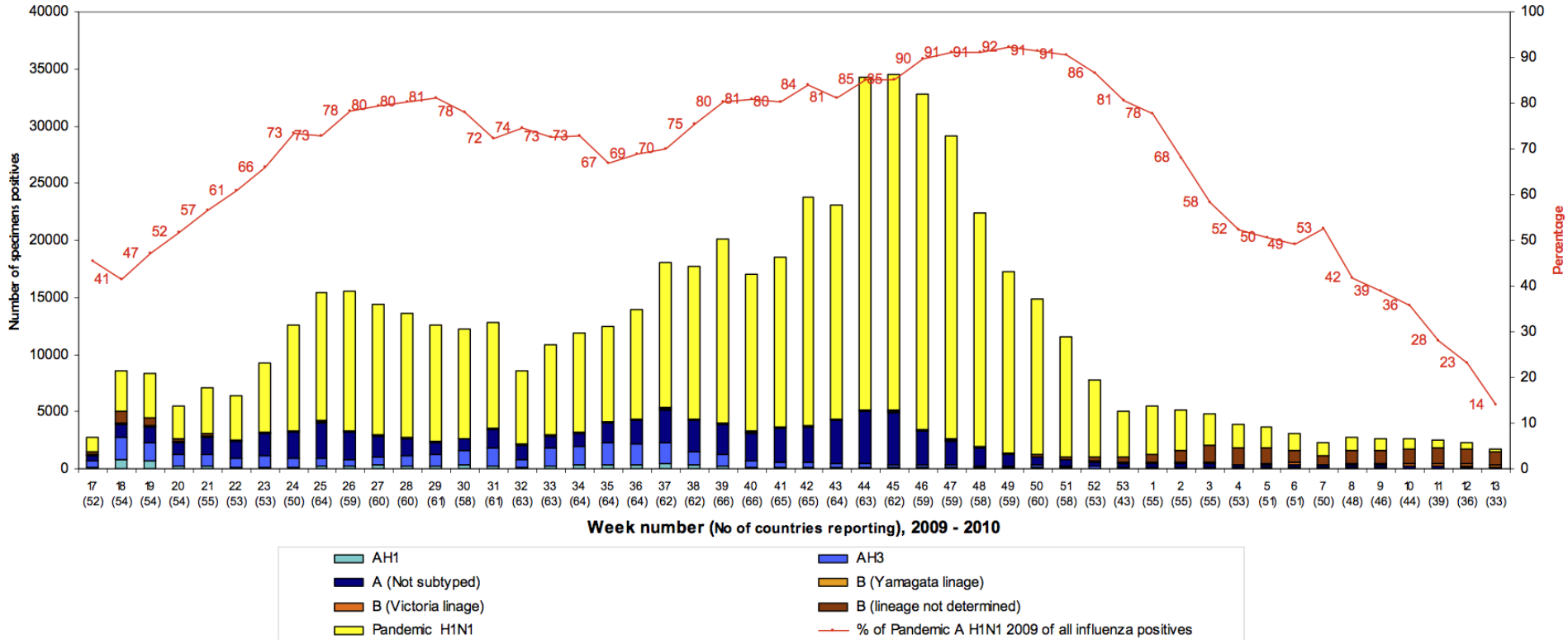
- Cumulative cases represented in Google Map as of 21 Apr, 2010
- WHO: 18769 deaths to date
- US: 4642 deaths; 480230 total cases
- Malaysia: 74 deaths, 6463 total cases
- China: 650 deaths; 124300 total cases

- Postpandemic period as of Aug 2010
- 0.5~1% death rate, similar to seasonal flu
- Targets younger and healthy individuals, different from seasonal flu (90% > 65 years older)

Source: http://flutracker.rhizalabs.com/

WHO Status Update

Week 17 (Apr 19, 2009) to Week 13 (Apr 3, 2010)



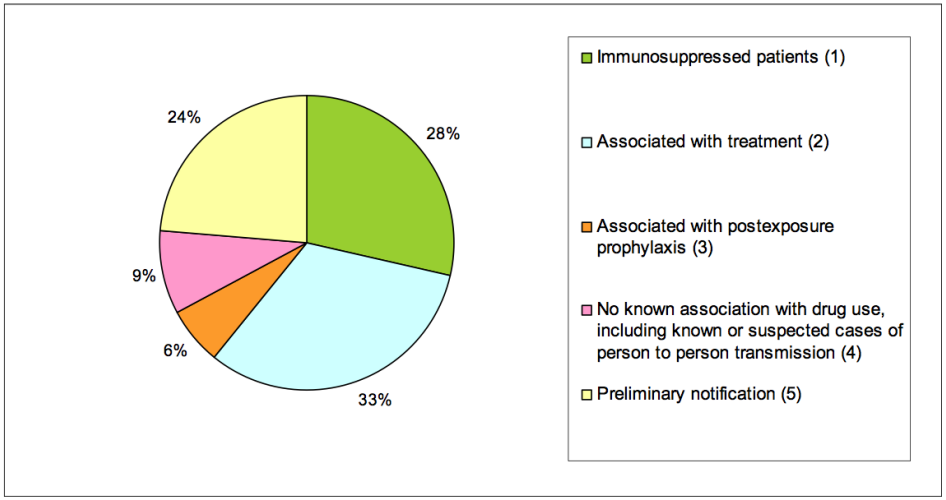
http://www.who.int/csr/disease/swineflu/laboratory23_04_2010/en/index.html

Oseltamivir (Tamiflu) Resistance in 2009 H1N1 pandemic virus

Table 1: Geographical distribution of oseltamivir resistance by the WHO regions (as of 04 August 2010)

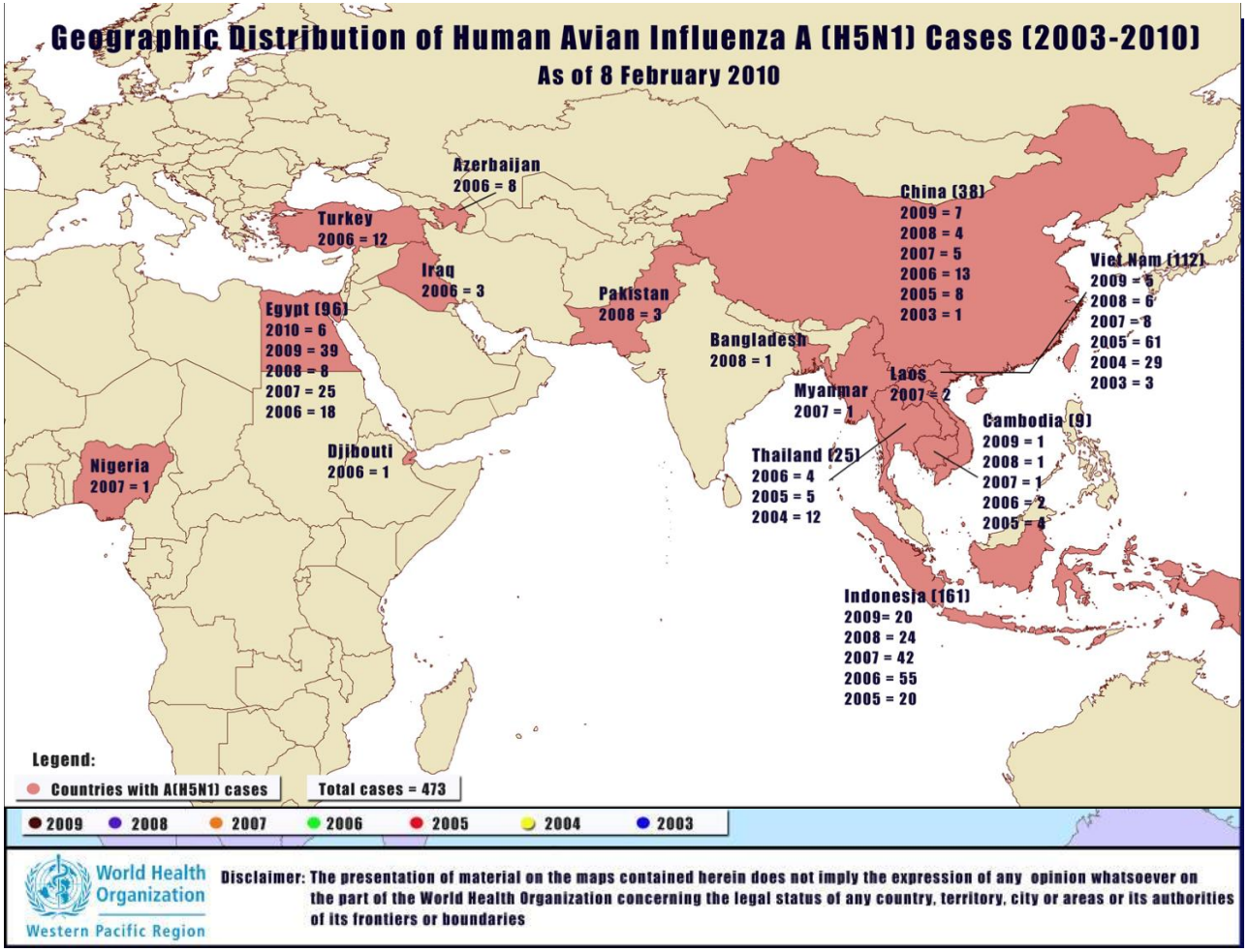
	WHO Region					
	AFRO	EMRO	EURO	PAHO	SEARO	WPRO
Number of oseltamivir resistant isolates	0	1	99	82	0	120

AFRO (Africa), AMRO (Americas), EMRO (Eastern Mediterranean), EURO (Europe), SEARO (South-East Asia) and WPRO (Western Pacific)



http://www.who.int/csr/disease/swineflu/oseltamivirresistant20100806.pdf

Reported human cases of avian influenza (H5N1) infection

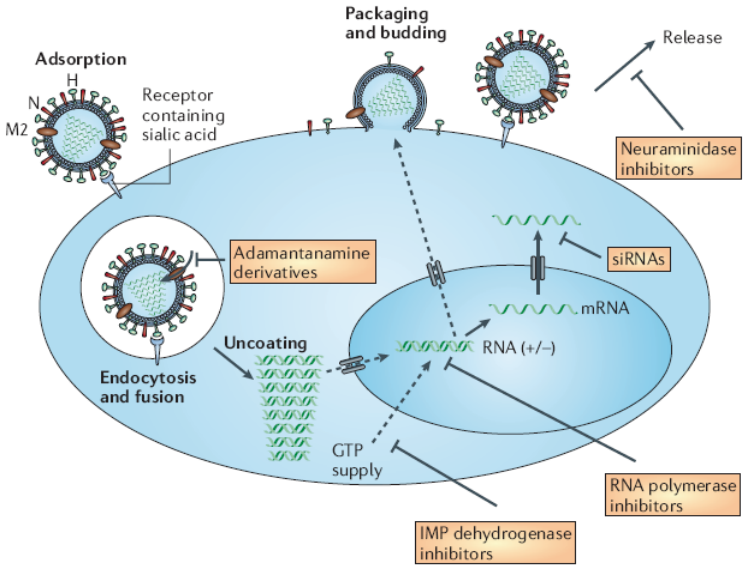


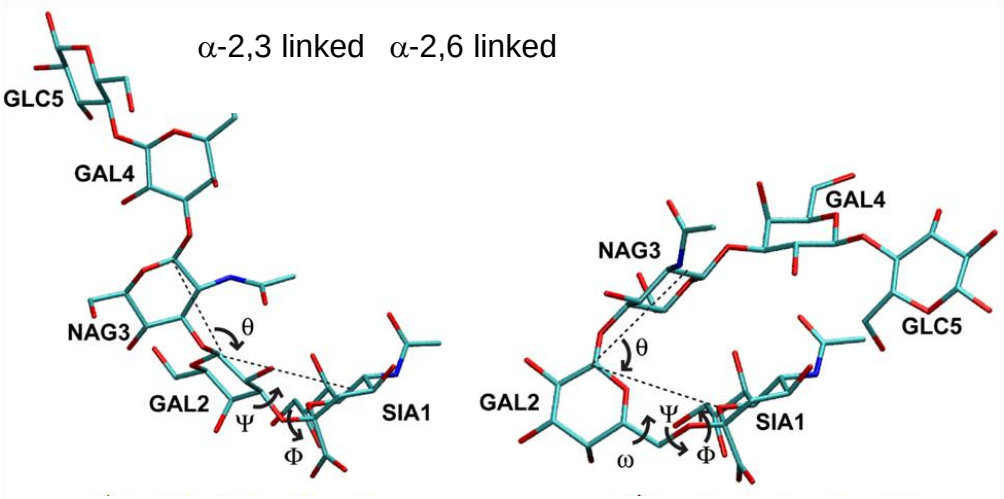
http://www.wpro.who.int/sites/csr/data/data_Maps.htm

~~Jan 13, 2009~~ Feb 2, 2010 statistics

- ~~6 countries with continuous cases totals 356~~ Egypt reported 6 cases in 2010
- 15 countries with ~~412~~471 cases to date; 282 deaths
- Age ranged 3 months to 81 yr, 90% <39 yr
- Death rate varies by country, average of 63%, highest 83% (Indonesia)
- People 10-19 years affected most; people > 50yr least affected.
- Gender neutral

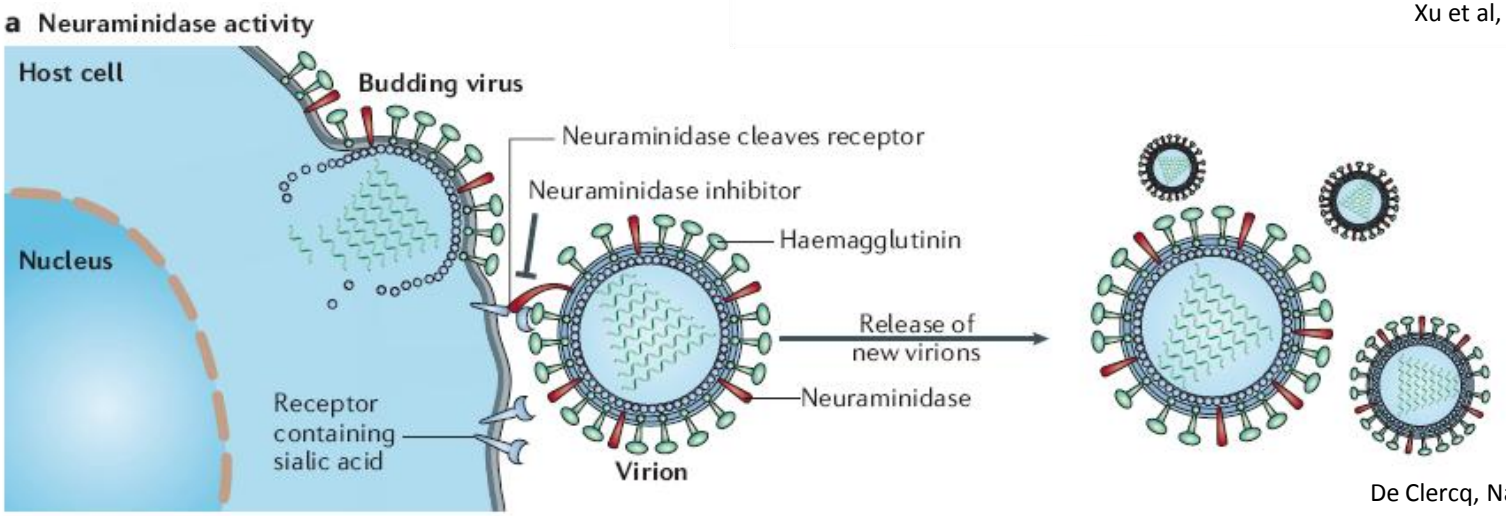
Points of Intervention in Viral Infectious Cycle





HA Binds to naturally occurring glycan receptors

Xu et al, JMB, 2009



De Clercq, Nat Rev, 2006

Vaccines against IV

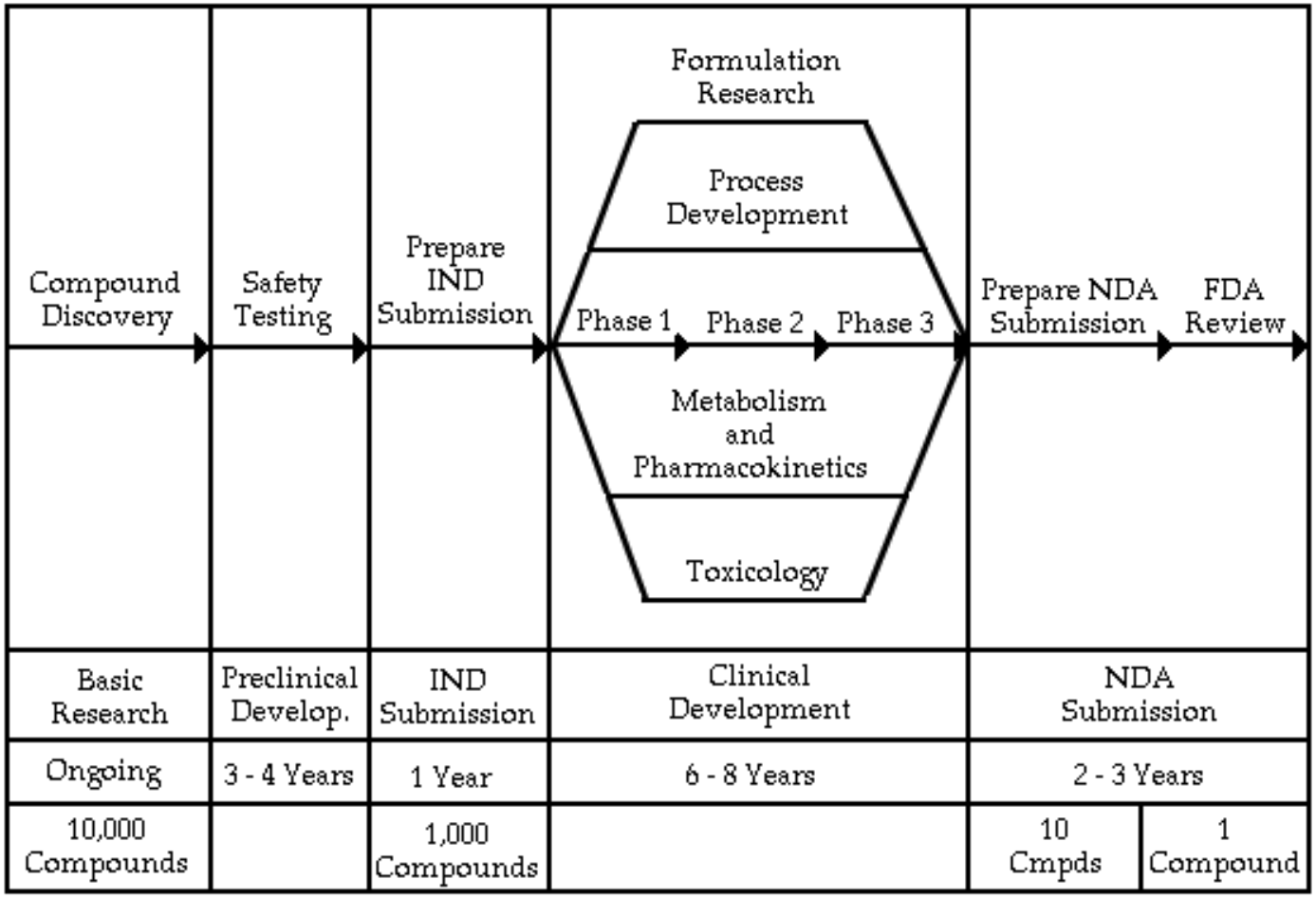
- The best way to protect yourself is get vaccinated annually.
  - **The "flu shot"** — an inactivated vaccine (containing killed virus) that is given with a needle, usually in the arm.
  - **The nasal-spray flu vaccine** —a vaccine made with live, weakened flu viruses that do not cause the flu (sometimes called LAIV for "live attenuated influenza vaccine" or FluMist®).

- Trivalent Vaccine for 2010-2011 flu season:
  - A/California/7/2009 (H1N1);
  - A/Perth/16/2009 (H3N2);
  - and B/Brisbane/60/2008.

- Recommended strongly for high risk population
  - Pregnant women
  - Children younger than 5, but especially children younger than 2 years old
  - People 50 years of age and older
  - People of any age with certain chronic medical conditions
  - People who live in nursing homes and other long-term care facilities
  - People who live with or care for those at high risk for complications from flu, including Health care workers
- Mild side Effects for most, though potentially fatal allergic reactions for some people
  - http://www.cdc.gov/vaccinesafety/Concerns/H1N1/index.html
  - Our of 82 million vaccinated, 1 in 10,000 reported adverse effects, and 10% of those are severe.
- New cell based production methods may provide doses of vaccines on demand within weeks, instead of months when using chicken eggs

http://www.flu.gov/individualfamily/vaccination/index.html

Typical Drug Discovery Process



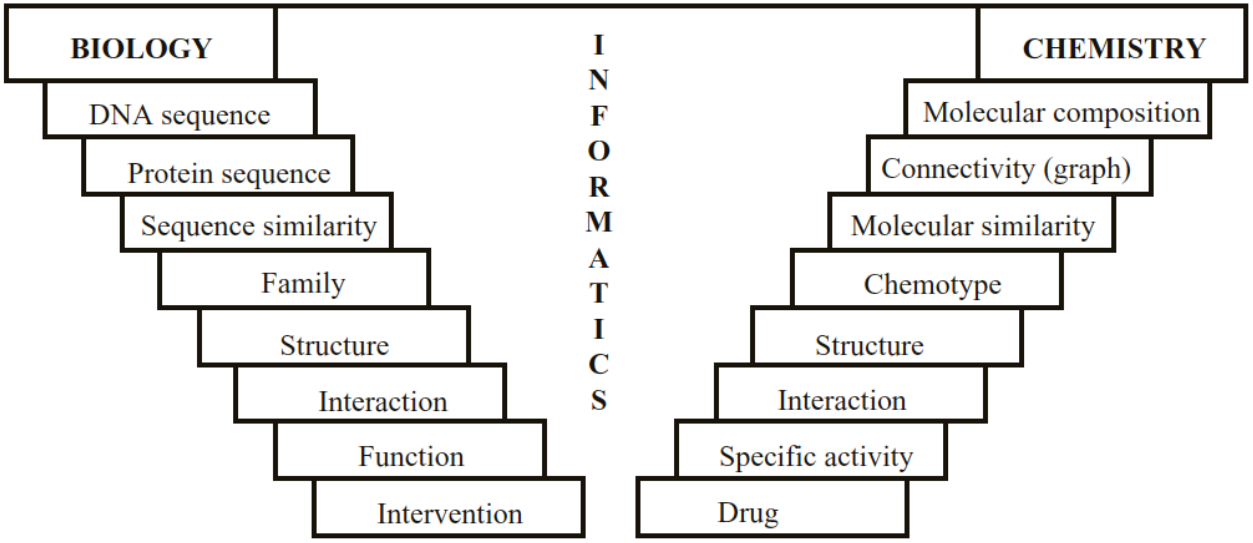
IND: investigational new drug
Phase I: pharmacokinetics (ADME)
and safety
Phase II: controlled trials for safety and efficacy
Phase III: Expanded trials involving hundreds to thousands
ADME: adsorption, metabolism, distribution and excretion
NDA: new drug application
FDA: federal drug administration

Reform: fast track and priority review – less than 1 year for FDA approval

http://www.netsci.org/Courseware/Drugs/Intro/

Informatics for Biology and Chemistry: bioinformatics and cheminformatics

Drug-like	Lead-like
MW < 500	MW < 350
ClogP < 5	ClogP < 3.0
Hydrogen bond donors < 5	Chemically stable
Hydrogen bond acceptors < 10	
Number of rotatable bonds ≤ 10	
PSA ≤ 140Å^2	
Peptides not suitable	Non-substrate peptides suitable
Eliminate reactive functional groups, promiscuous inhibitors, and metabolically unstable compounds	

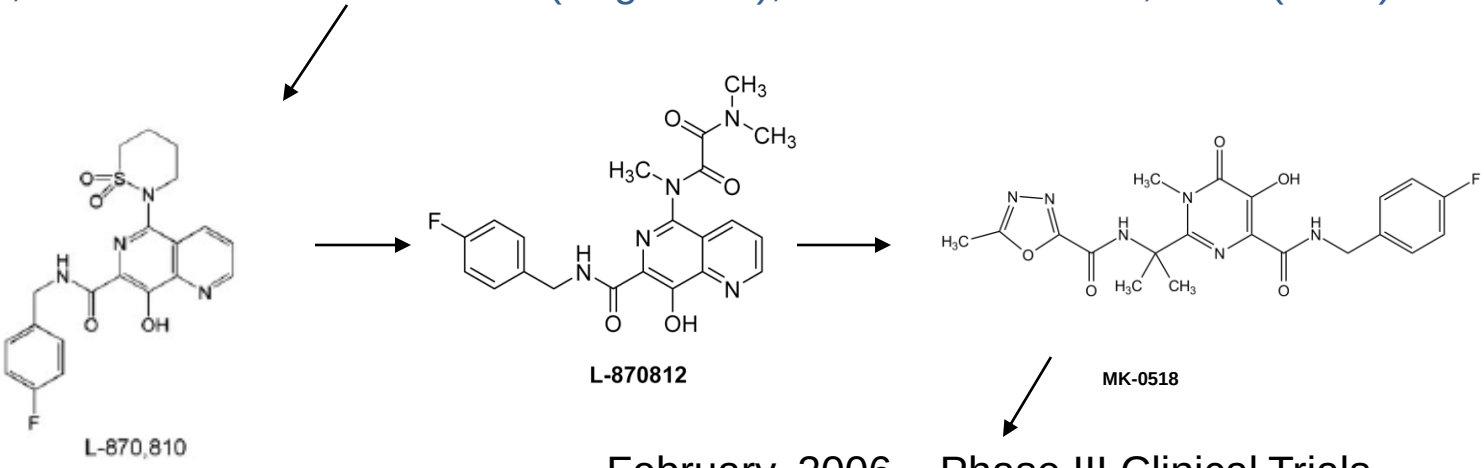


Bunin et al, Chemoinformatics, 2007

Relaxed Complex Scheme and Ensemble based Virtual Screening Contributed to HIV Integrase Inhibitor Development

Discovery of unexpected binding site in HIV-1 Integrase using MD and AutoDock:
Schames, … & McCammon, *J. Med. Chem.* (released on web, early 2004)

“ Exploration of the structural basis for this unexpected result … suggests an approach to the development of integrase inhibitors with unique resistance profiles.”
D. Hazuda et al., *Proc. Natl. Acad. Sci. USA* (Aug. 2004), refers to Schames, et al. (2004).



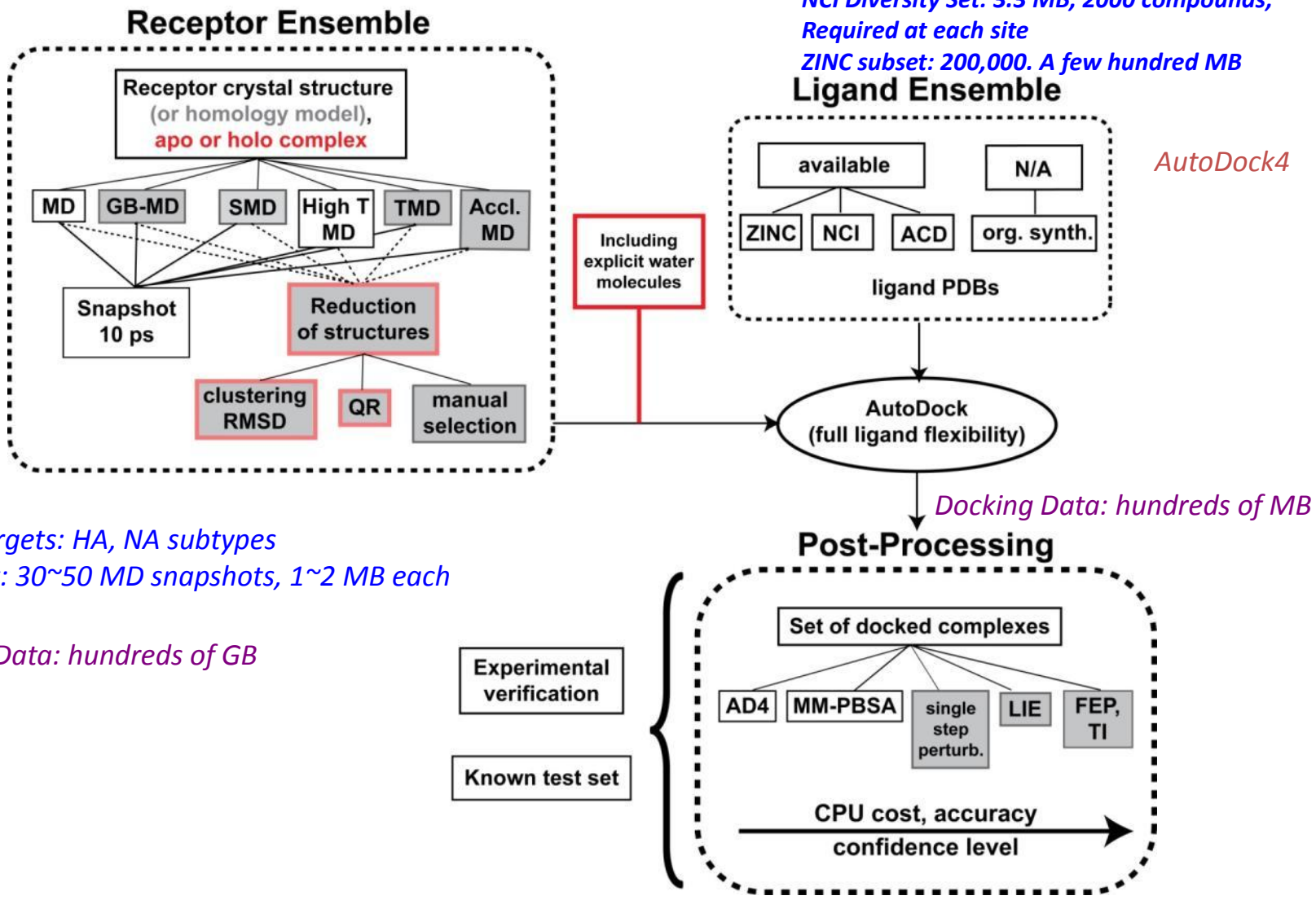
New Class of HIV Drugs: Merck & Co.

February, 2006 – Phase III Clinical Trials
February, 2007 – Name announced:
Isentress (raltegravir)
October, 2007 – FDA “fast track” approval

Source: A. McCammon

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

AutoDock4

Docking Data: hundreds of MB

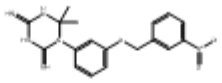
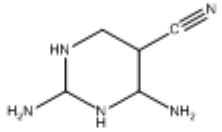
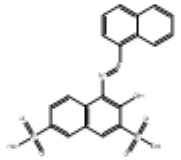
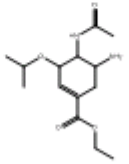
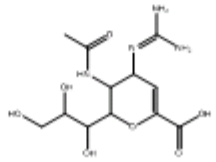
Multiple targets: HA, NA subtypes
Each target: 30~50 MD snapshots, 1~2 MB each

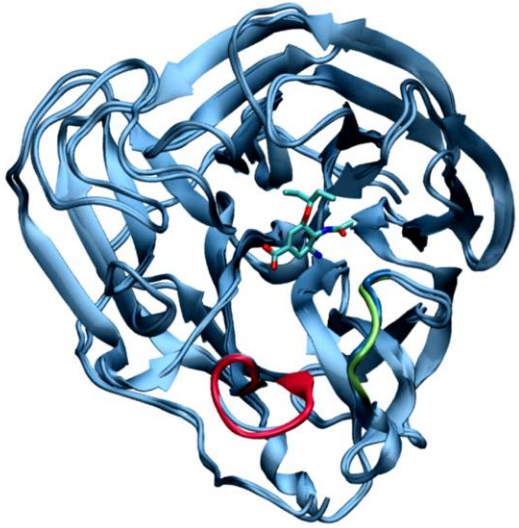
Simulation Data: hundreds of GB

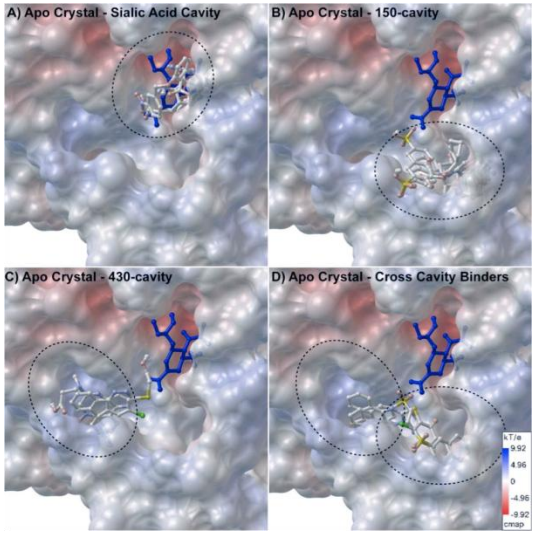
Total data to date: ~5 TB in long term storage.
Each experiment is about 1 Petaflops accumulative in computation cost.

Source: Amaro

Ensemble based Virtual Screening of NCIDS1 Top Hits against NA

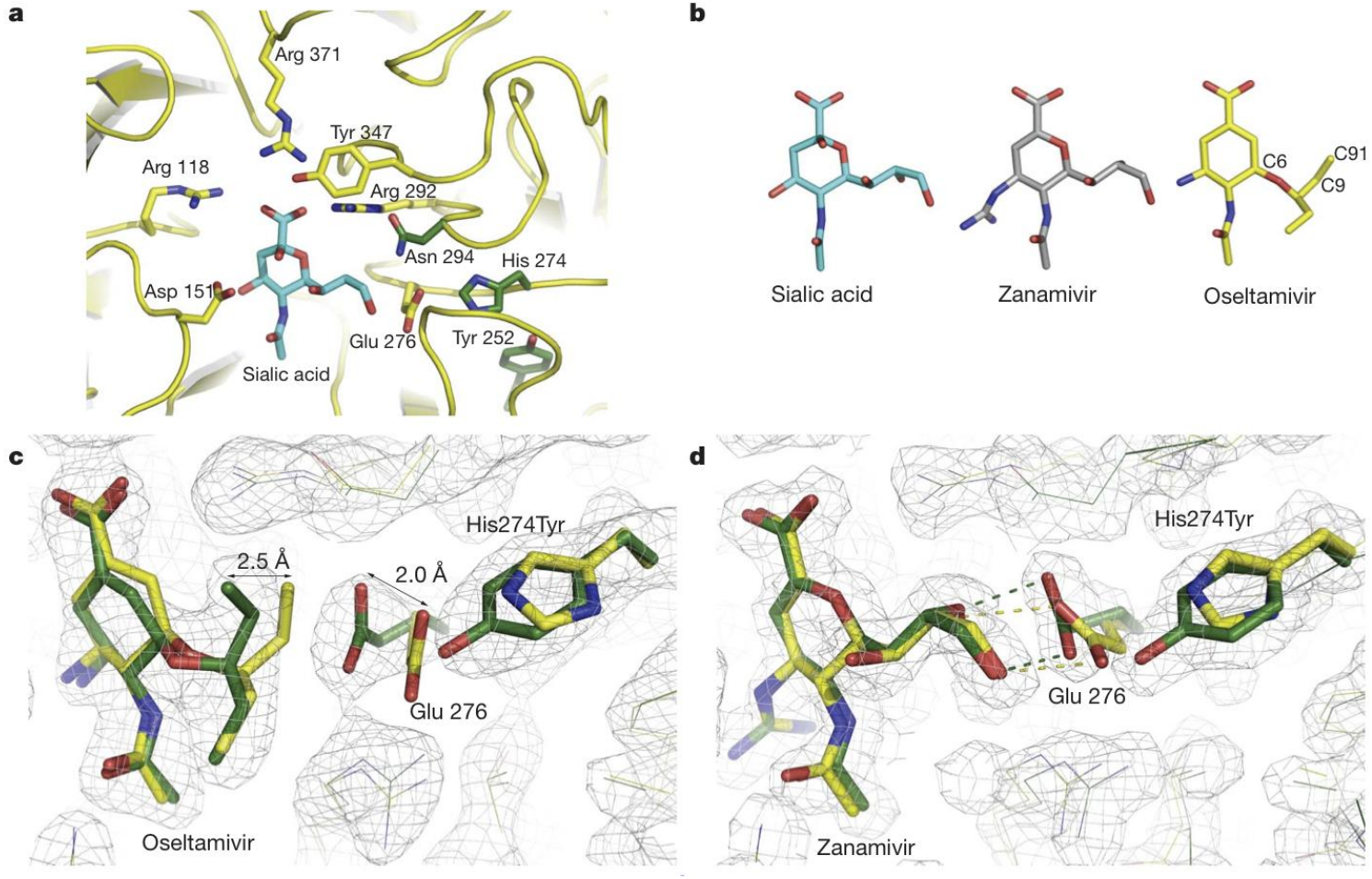
Rank	NSC	Mean Energy	Predicted K_i (μM)	Chemical Structure	Binding Site	Apo Crystal Rank	Holo Crystal Rank
1	109836	-10.63	0.016		SA-cavity	15	1
2	211332	-10.34	0.026		SA-cavity	212	10
3	45583	-10.09	0.040		SA-cavity 150-cavity 430-cavity	6	18
-	Oseltamivir	-9.82	0.063 (0.3 – 1.0)		SA-cavity	238	5
-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





- Cheng et al, JMC, 2008

Mechanism of H275Y Resistance Mutation of N1



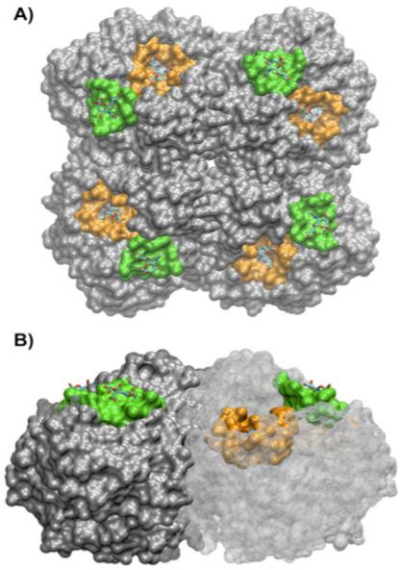
Collins PJ et al. Nature: 2008.

Electrostatic Steering and 2nd Sia binding site in Pandemic Influenza Virus

Table 1. Association Rates (k_{on}) of Ligands to the Active Site and Secondary Site of Neuraminidases from Three Influenza Strains[a]

	$k_{on,active\ site}$ ($\mu M^{-1} \cdot s^{-1}$)	$k_{on,secondary\ site}$ ($\mu M^{-1} \cdot s^{-1}$)	$k_{on,secondary\ site}/k_{on,active\ site}$
Oseltamivir-N1-a	5.17 (8×10^{-4})	9.73 (1×10^{-3})	1.88
Sialic Acid-N1-a	9.41×10^{-2} (1×10^{-4})	208 (5×10^{-3})	2210
Oseltamivir-N2-h	12.1 (2×10^{-3})	84.2 (6×10^{-3})	6.96
Sialic Acid-N2-h	0.503 (5×10^{-4})	1.78 (1×10^{-3})	3.54
Sialic Acid-H1N1-h	0.168 (8×10^{-4})	12.6 (1×10^{-3})	75.0

a "-a" and "-h" denote avian- or human-derived strains, respectively. Standard errors of the mean are shown in parentheses.

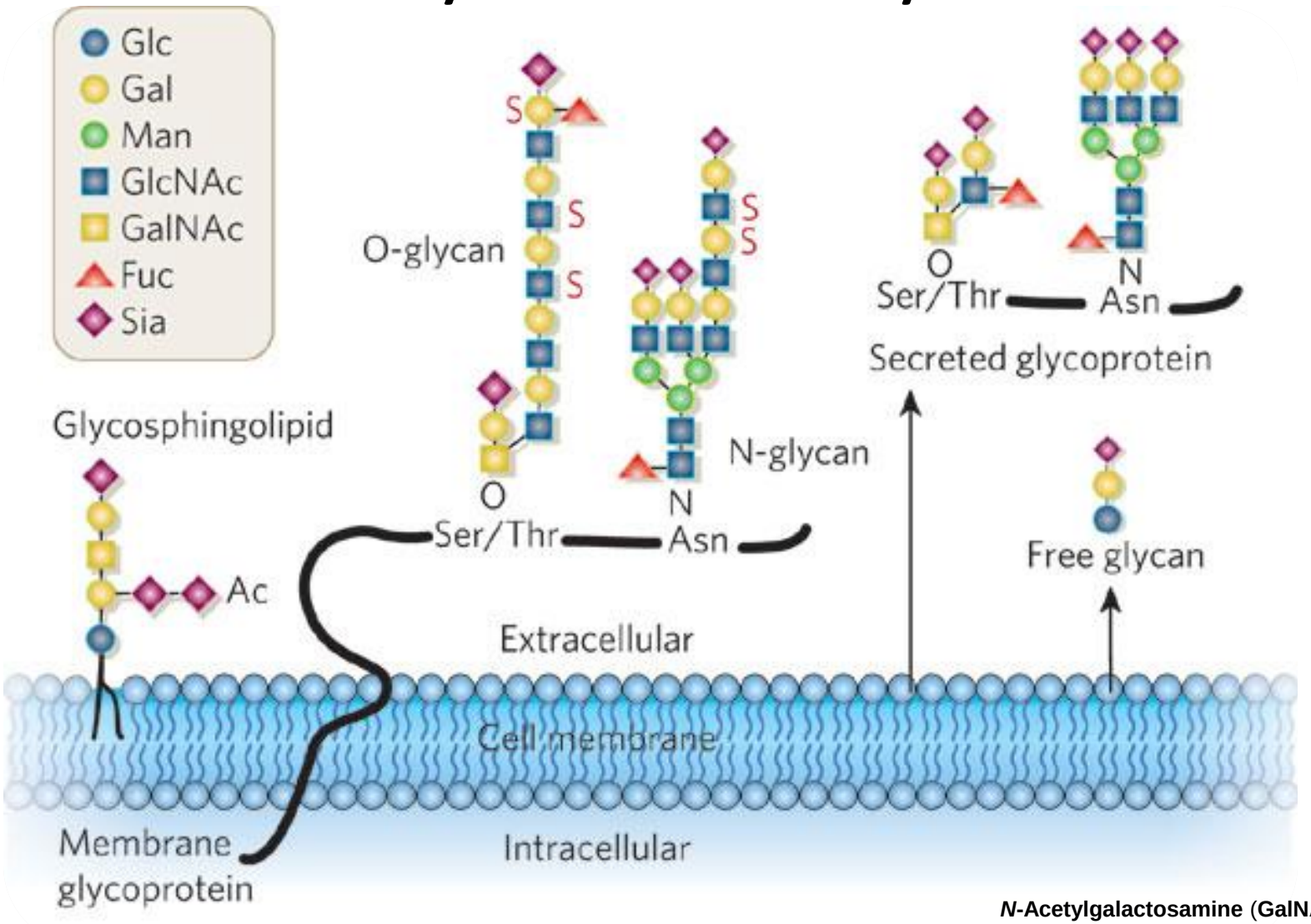


Green: 2nd site

	366-373	399-403	430-433
Avian N1	KSTNSRSG	AITDWS	RPKE
Human N2	ISKDLRSG	DSDNRS	RKQE
H1N1	KSISSRNG	GINEWS	RPKE

Sung et al, JACS, 2010

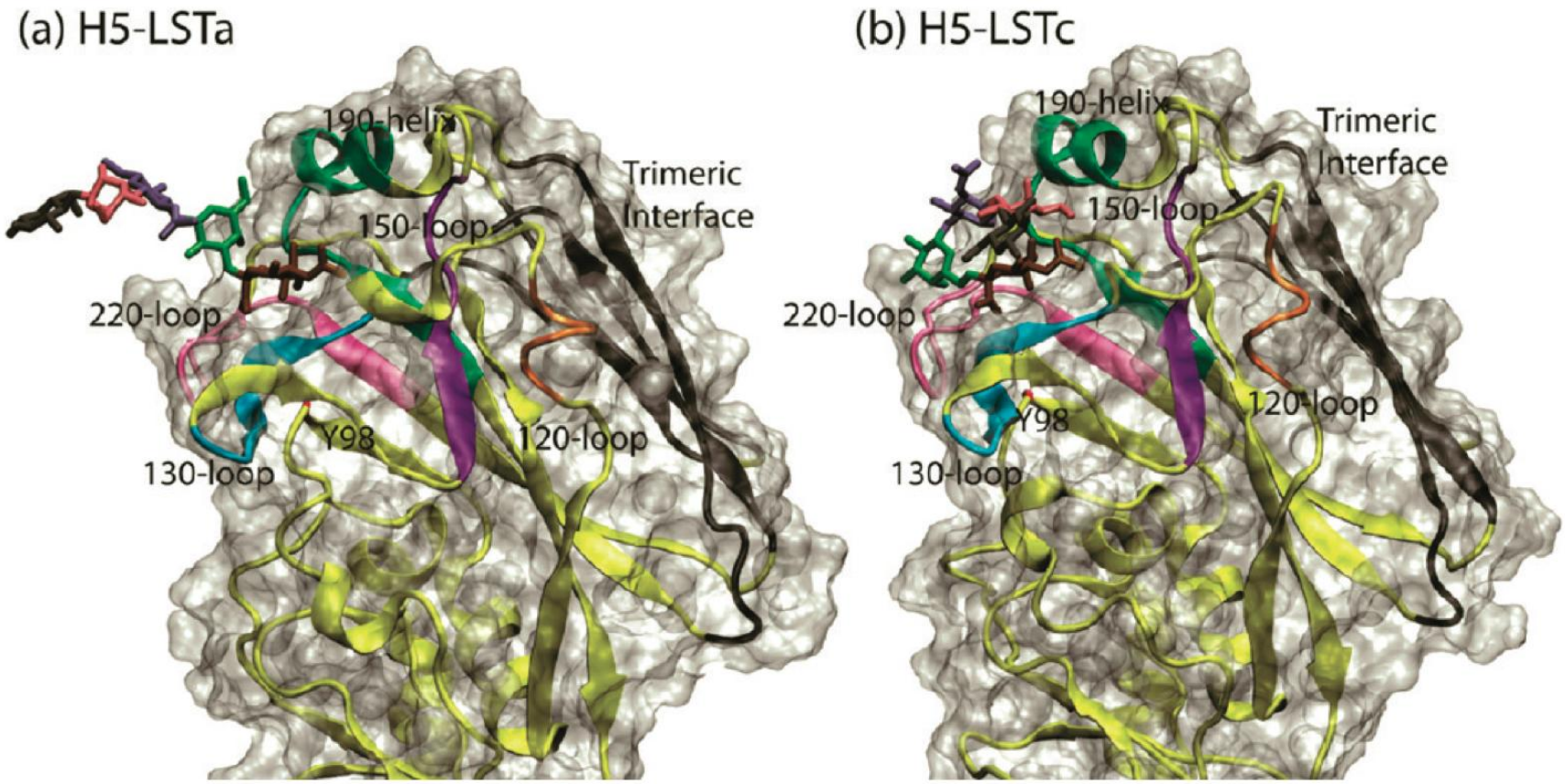
Glycan Diversity



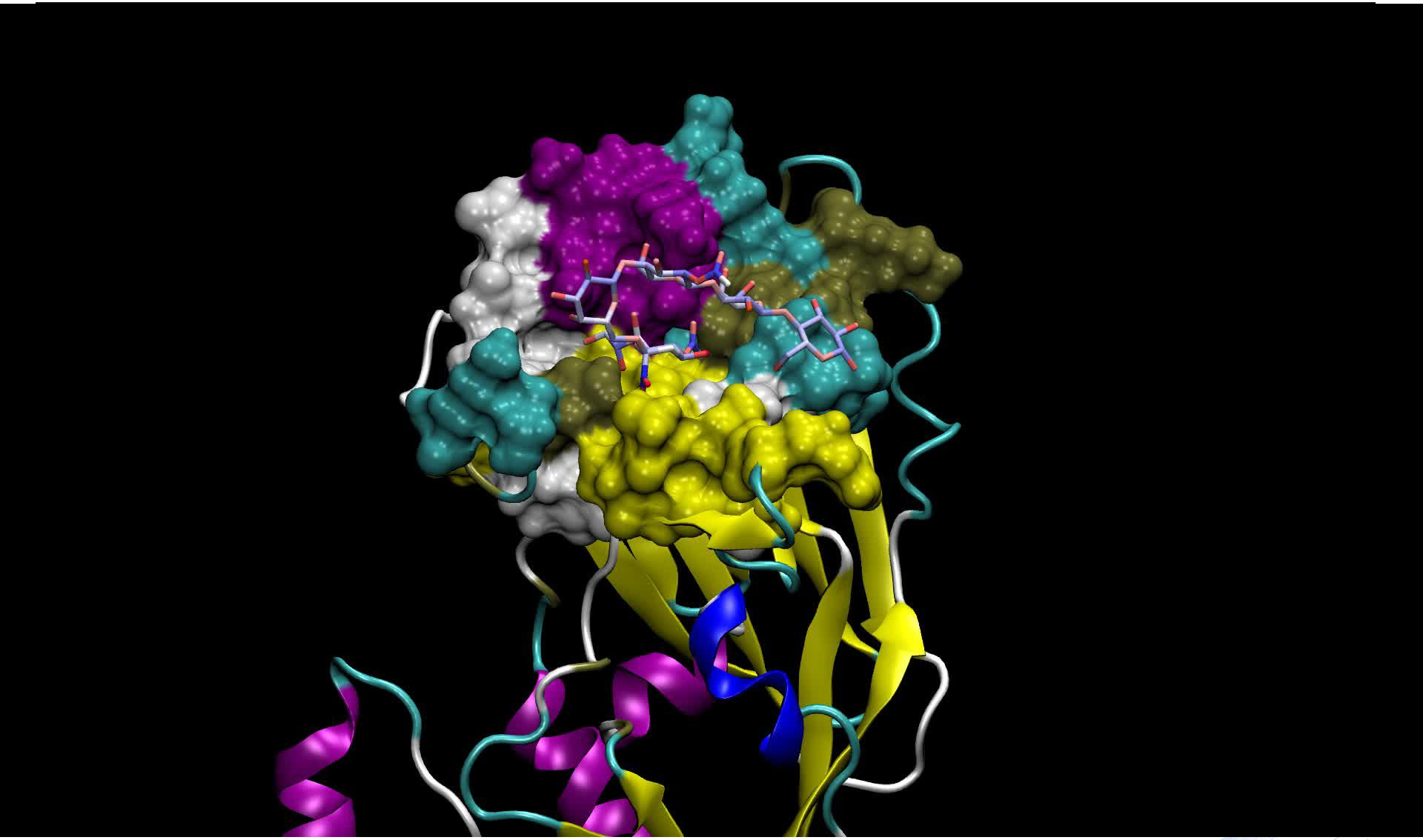
***N*-Acetylgalactosamine (GalNAc)**

N.M.Varki, A.Varki. (2007) *Pathobiology in Focus.* Diversity in Sialic Acid Presentations: implications for biology and disease. *Laboratory Investigation*, 1-7

HA-Glycan Receptor Binding Domain

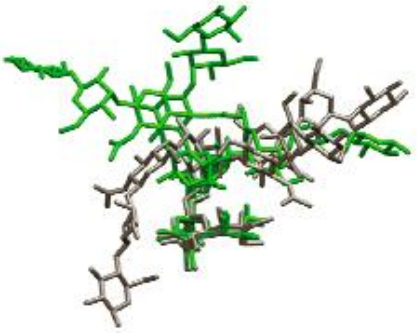


Visualization of HA-Glycan Interactions

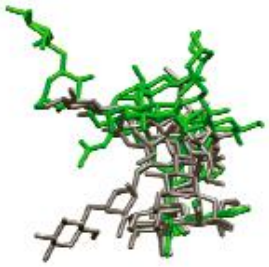


Glycan topology and induced conformational changes

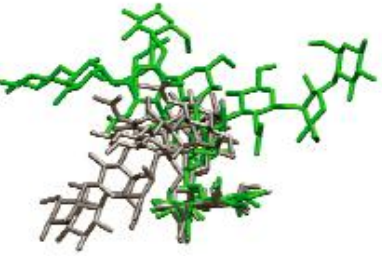
Unbound LSTa/LSTc



(b) LSTa-H3/LSTc-H3



LSTa-H5/LSTc-H5



(d) LSTa-H9/LSTc-H9

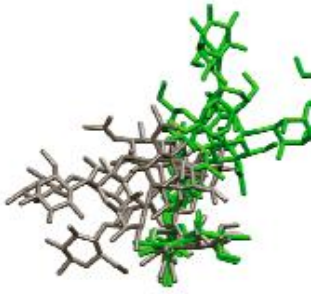


Table 4. Glycan conformational entropic changes between bound and free states.

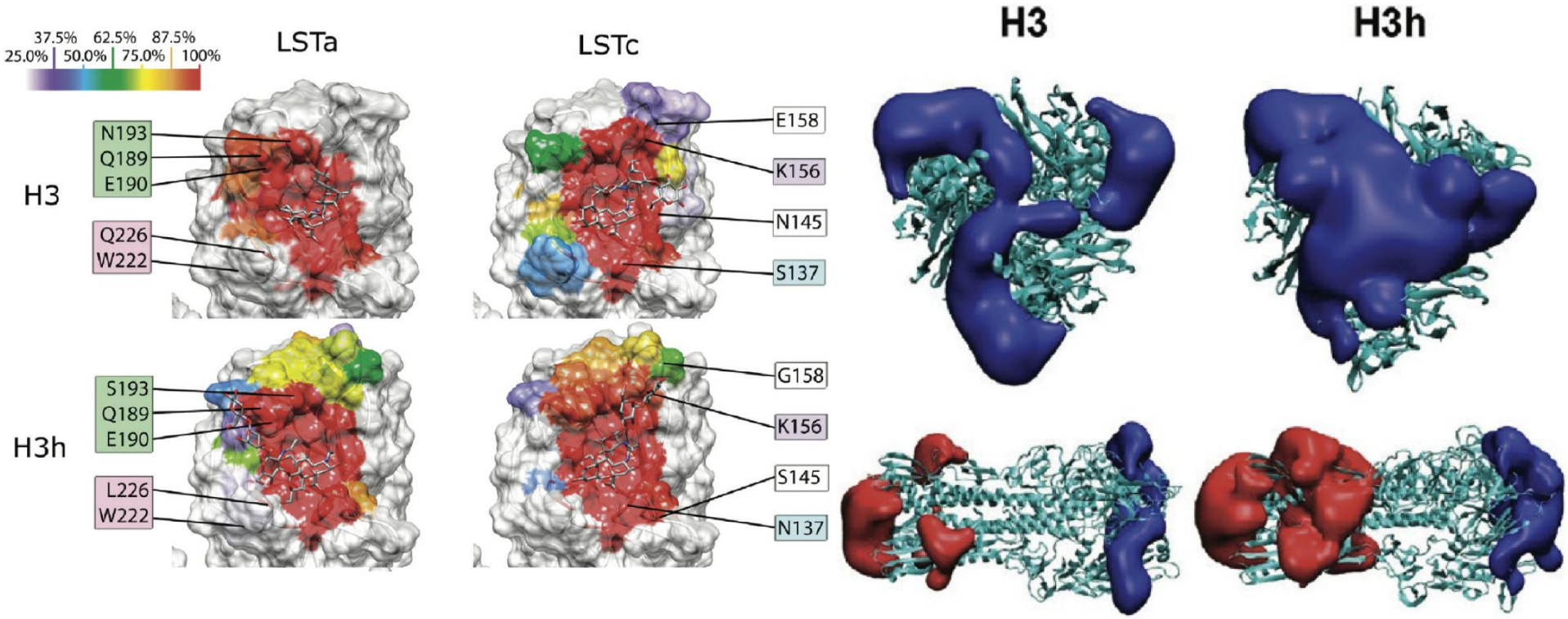
		H3	H5	H9
LSTa	$-T\Delta S_{Conf}$	1.59 (0.34)	0.28 (0.01)	2.46 (0.02)
	$-T\Delta S_{\alpha-2,3}$	0.79 (0.18)	0.13 (0.03)	1.89 (0.03)
	% Contribution	50%	46%	77%
LSTc	$-T\Delta S_{Conf}$	2.47 (0.80)	5.52 (0.11)	4.85 (1.64)
	$-T\Delta S_{\alpha-2,6}$	1.39 (0.45)	3.55 (0.06)	2.49 (0.82)
	% Contribution	56%	64%	38%
	$\Delta(-T\Delta S_{Conf})$	0.88 (0.43)	5.24 (0.06)	2.39 (0.82)

The percent contribution is $-T\Delta S_{\alpha-2,3}$ or $-T\Delta S_{\alpha-2,6}$ relative to $-T\Delta S_{Conf}$. Standard errors are given in parenthesis. All units are in kcal/mol. Temperature is 310K. $\Delta(-T\Delta S_{Conf})$ is the difference between $-T\Delta S_{Conf}$ of LSTc and LSTa bound to the same HA. Standard errors are given in parenthesis.

Free: green; Bound: Gray

Xu et al, JMB, 2009

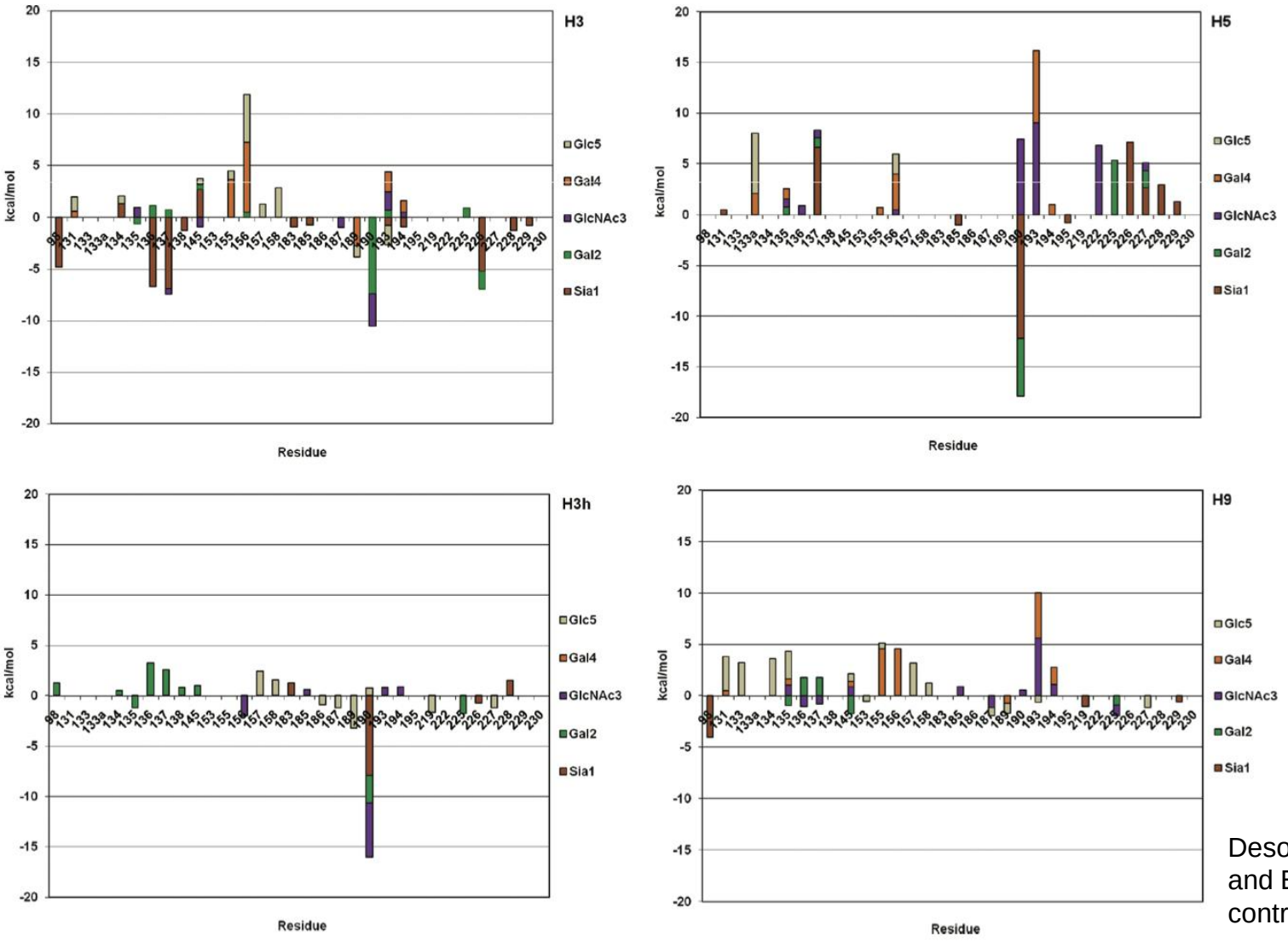
Role of Electrostatic Steering in Receptor Specificity Switch



Newhouse et al, JACS, 2009

Newhouse et al, JACS 2009

Propensity Index from Interaction Energy Analysis for HA Receptor Domain Residues

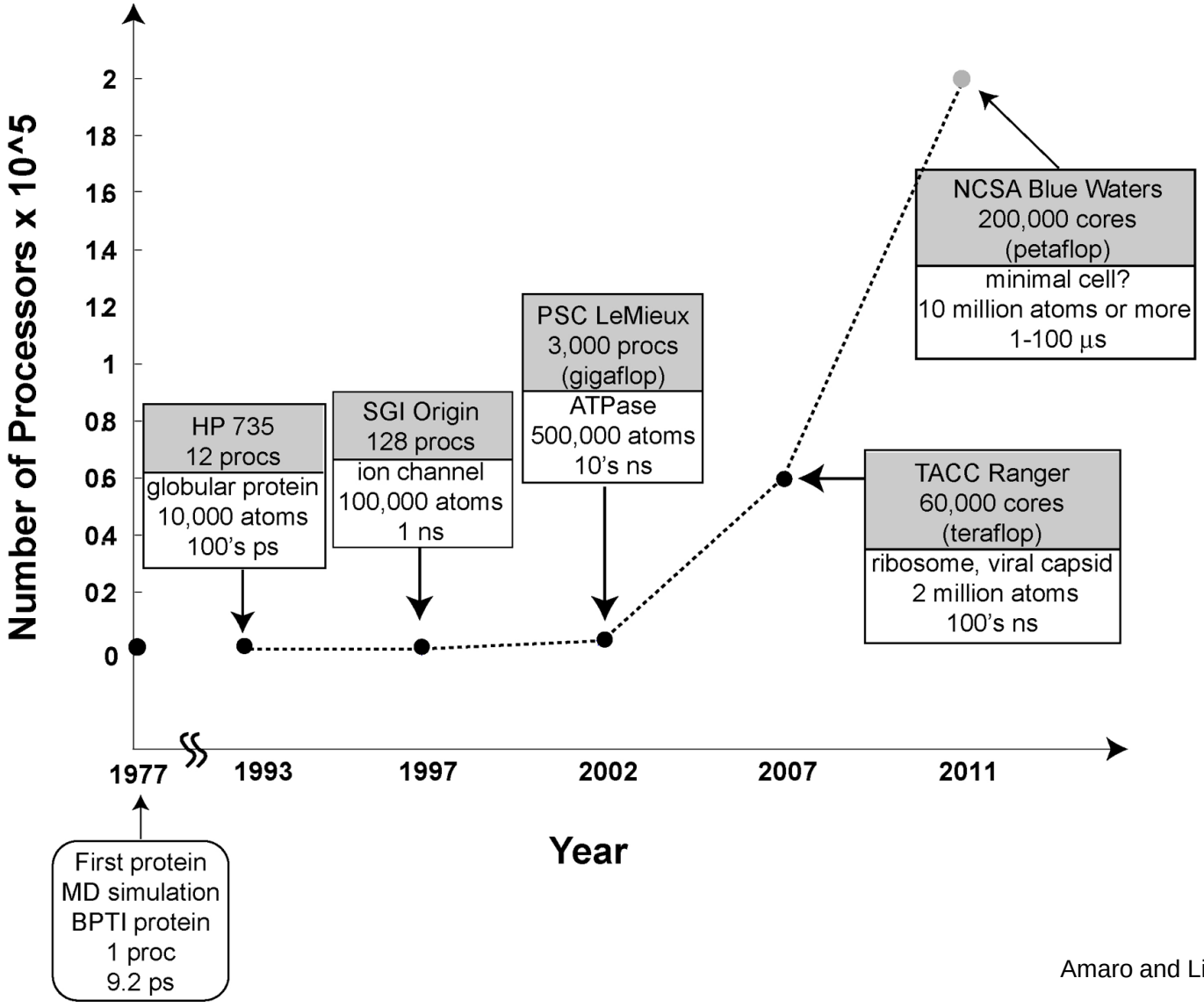


Desolvation Energy and Entropic contributions ignored

Figure 7. Propensity index for RBD residues to favor LSTa or LSTc in bound state. Negative values favor bound LSTa, and vice versa for bound LSTc.

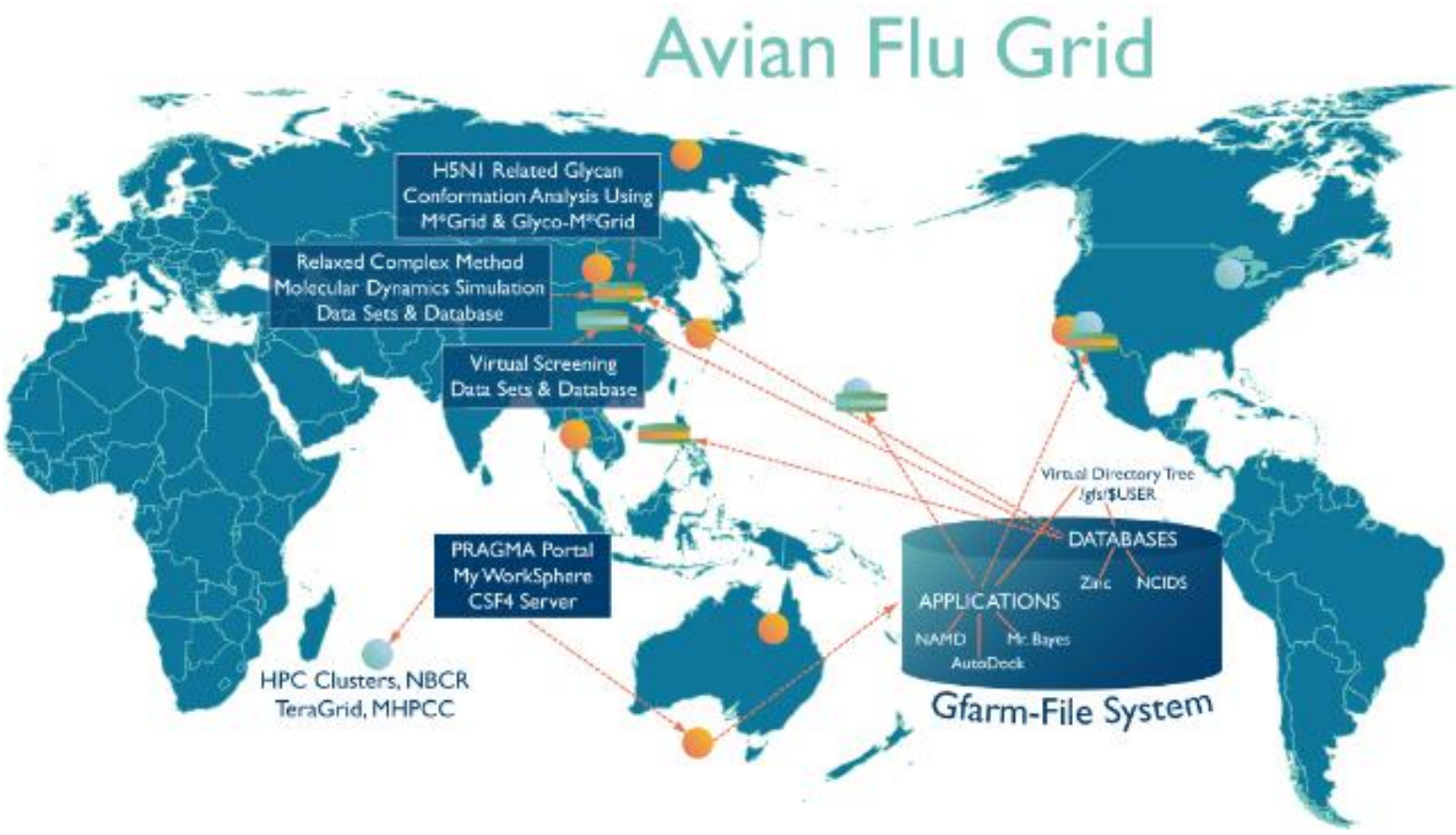
Newhouse et al, JACS, 2009

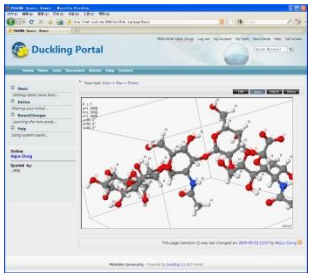
Computing Power Readily Available at Supercomputer Centers



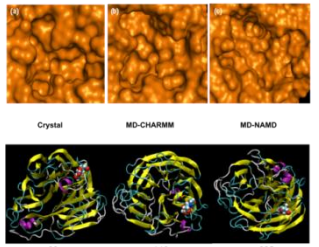
Amaro and Li, CTMC, 2010

Transparent access of applications on Avian Flu Grid through middleware

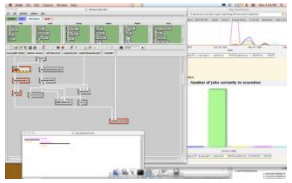




CNIC Duckling Portal



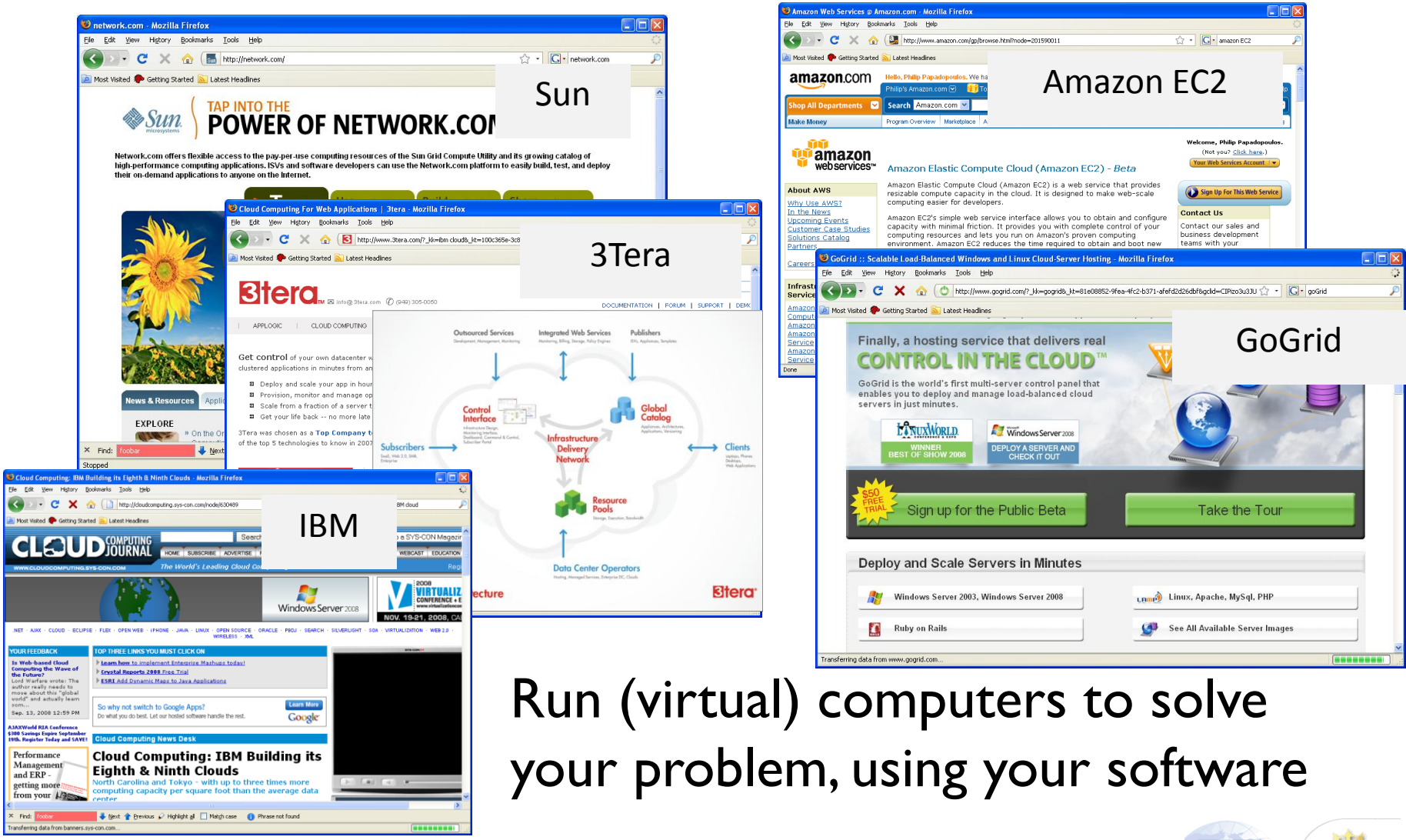
Konkuk Glyco-M*Grid



NBCR CADD

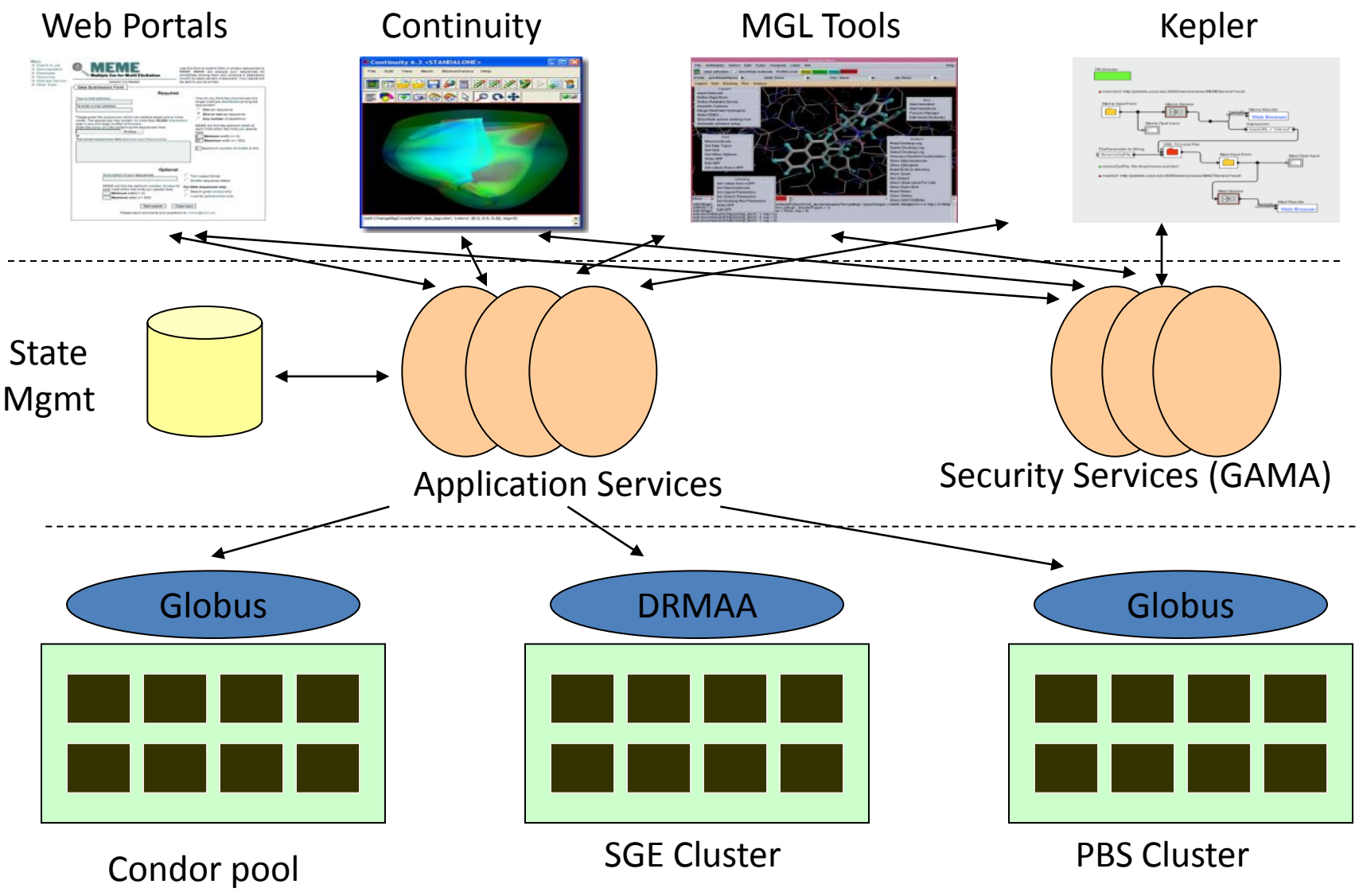
National Center for Research Resources

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

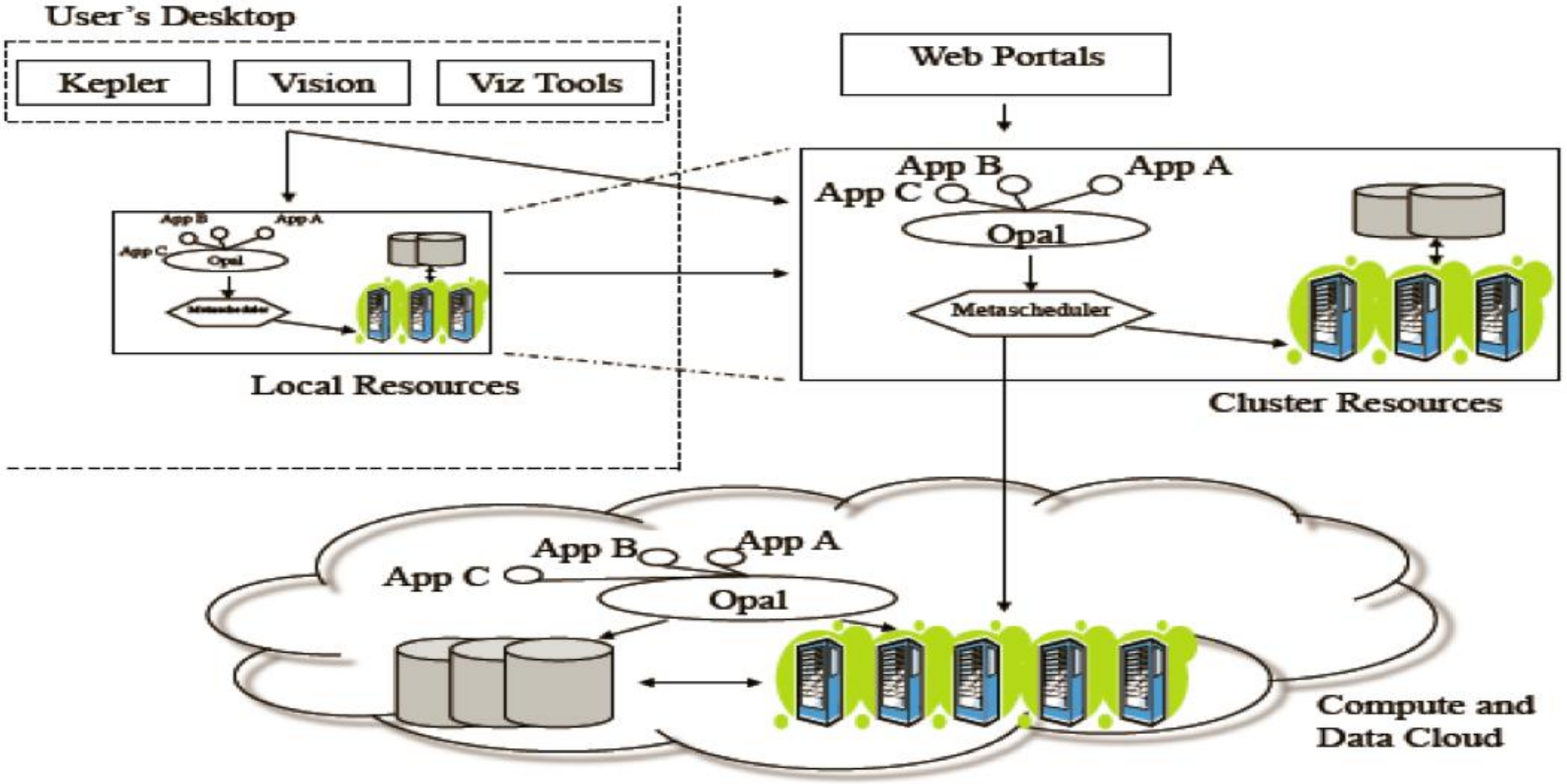


Run (virtual) computers to solve your problem, using your software

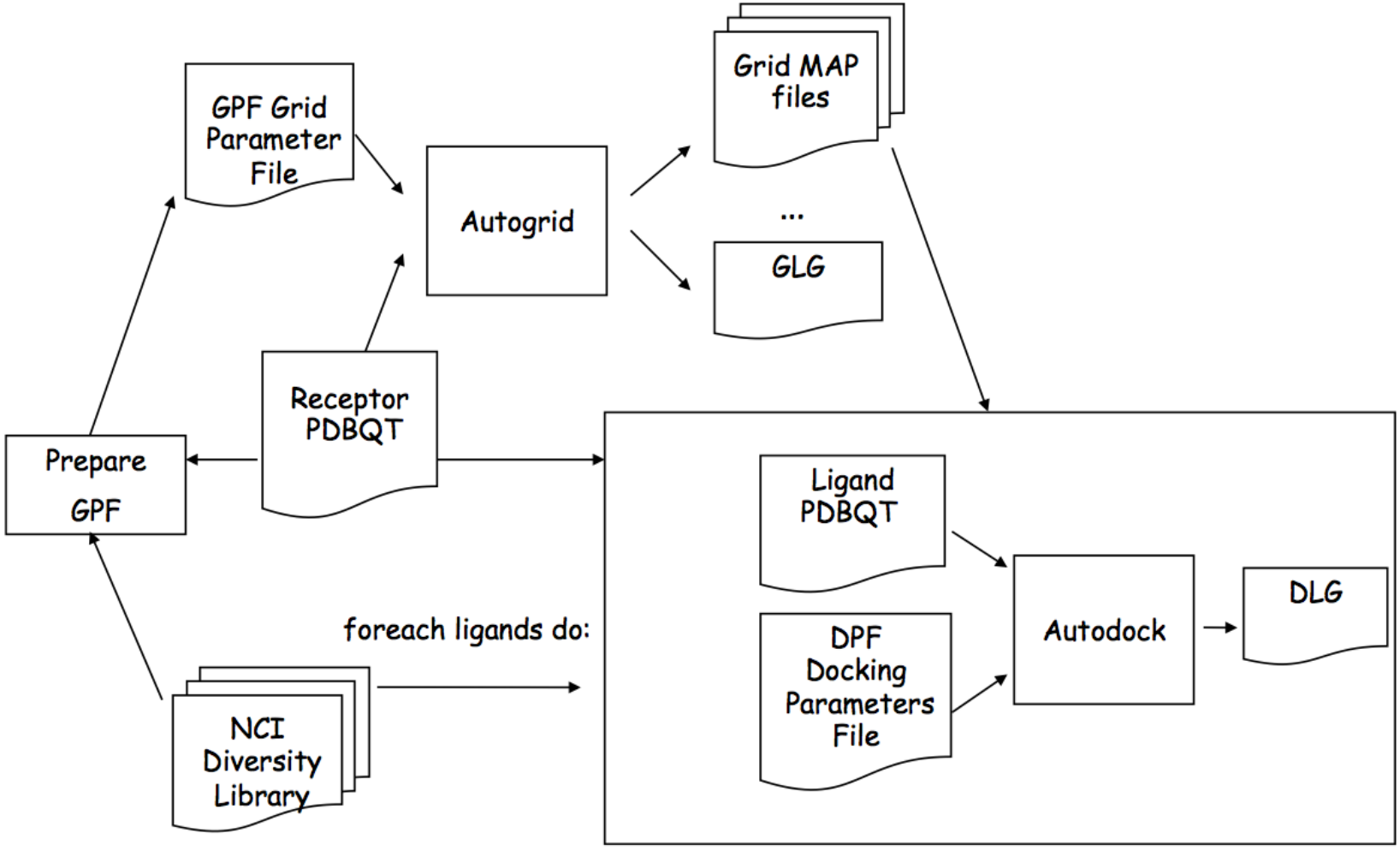
NBCR Opal Web Service Toolkit v1



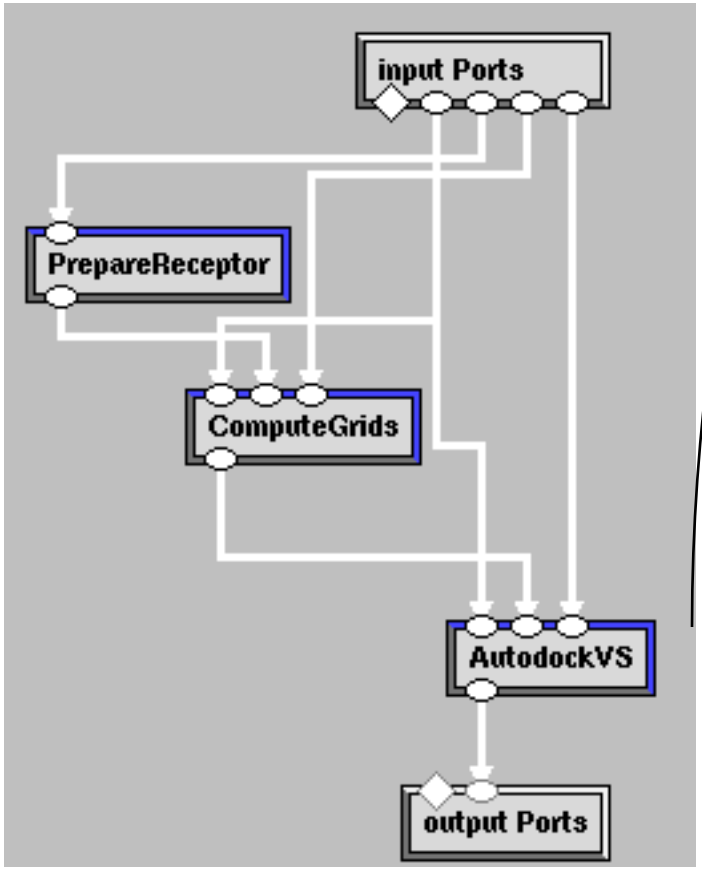
Opal 2 for SaaS



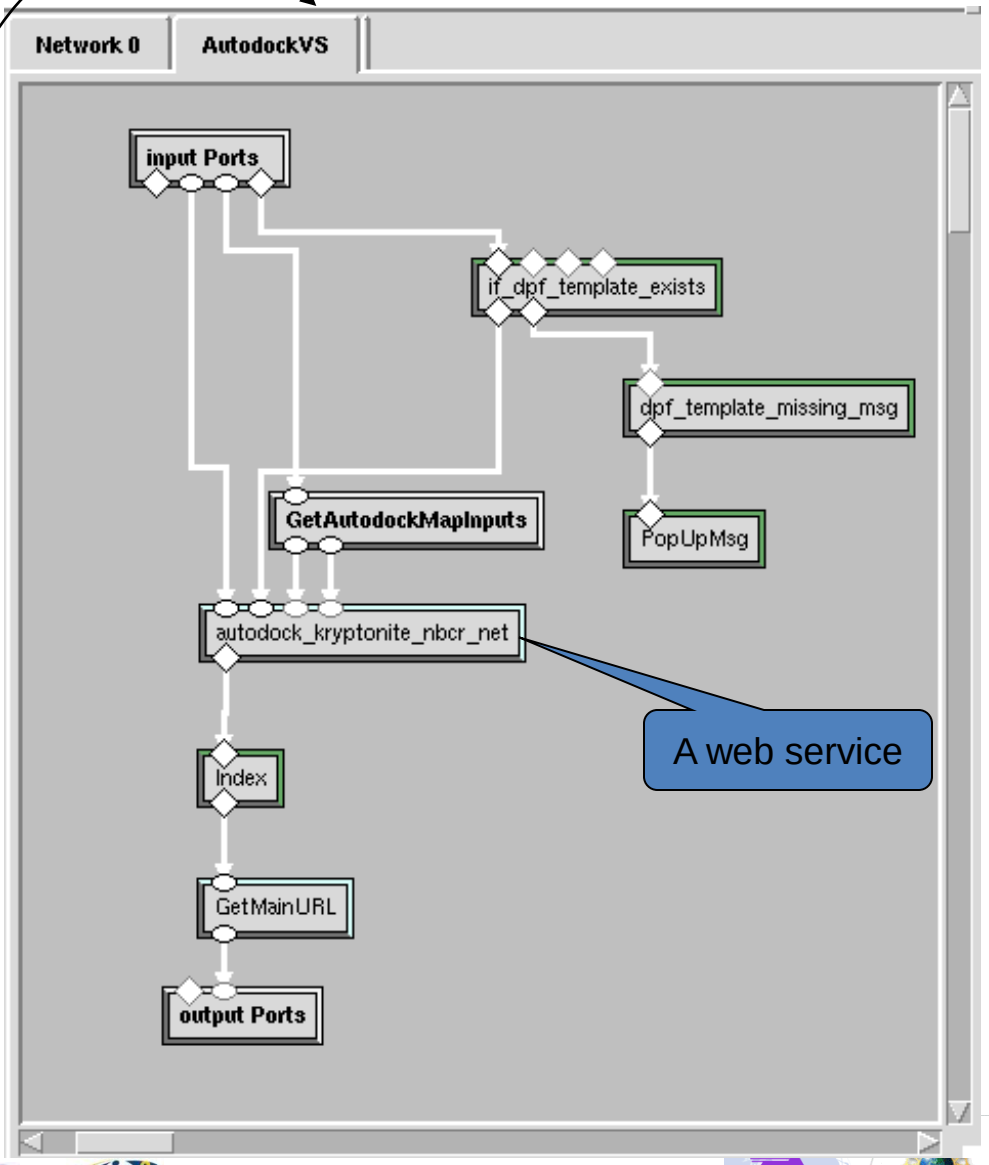
AutoDock Workflow



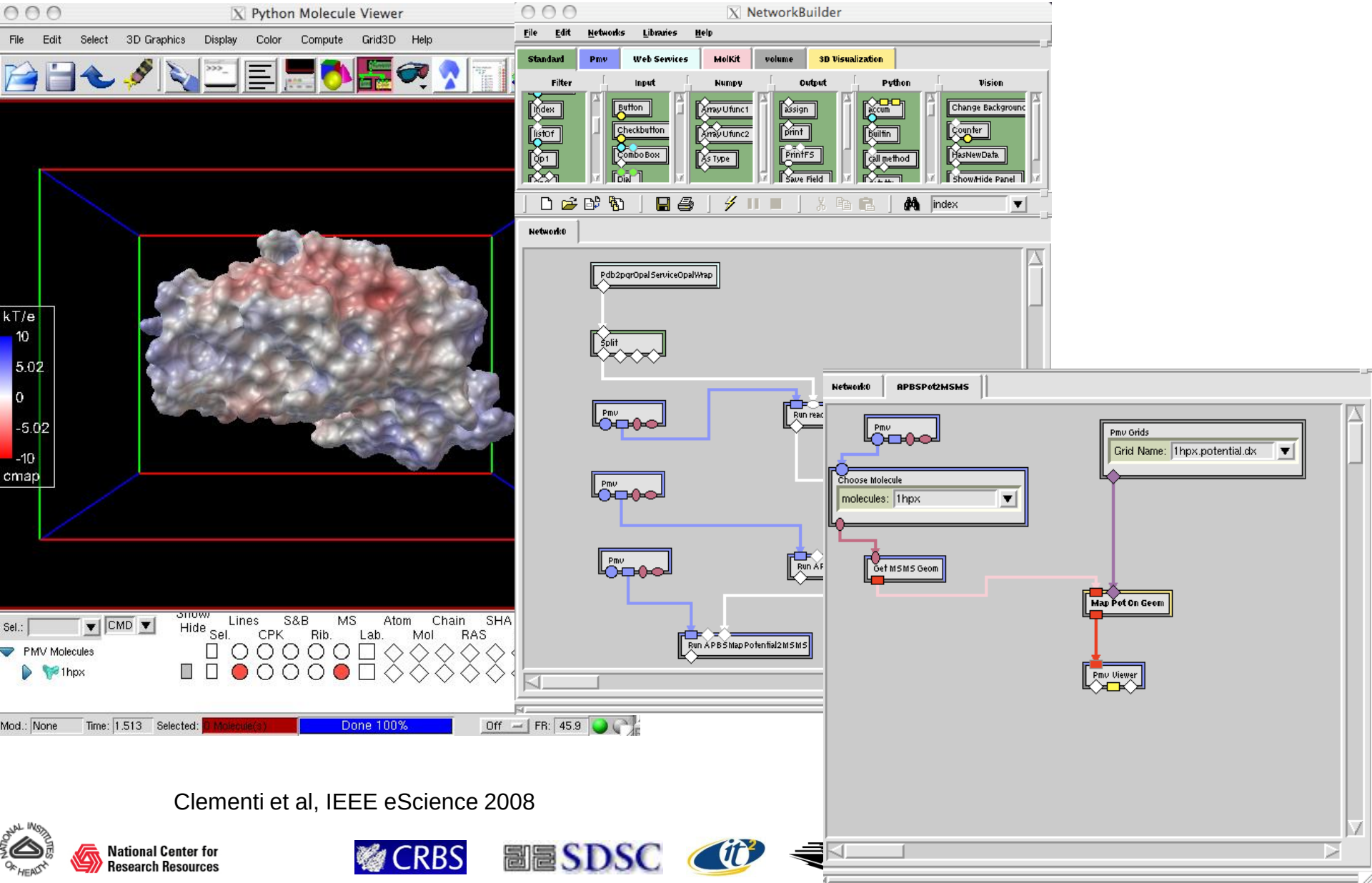
A Virtual Screening Vision Workflow



Ren et al, NAR, 2010



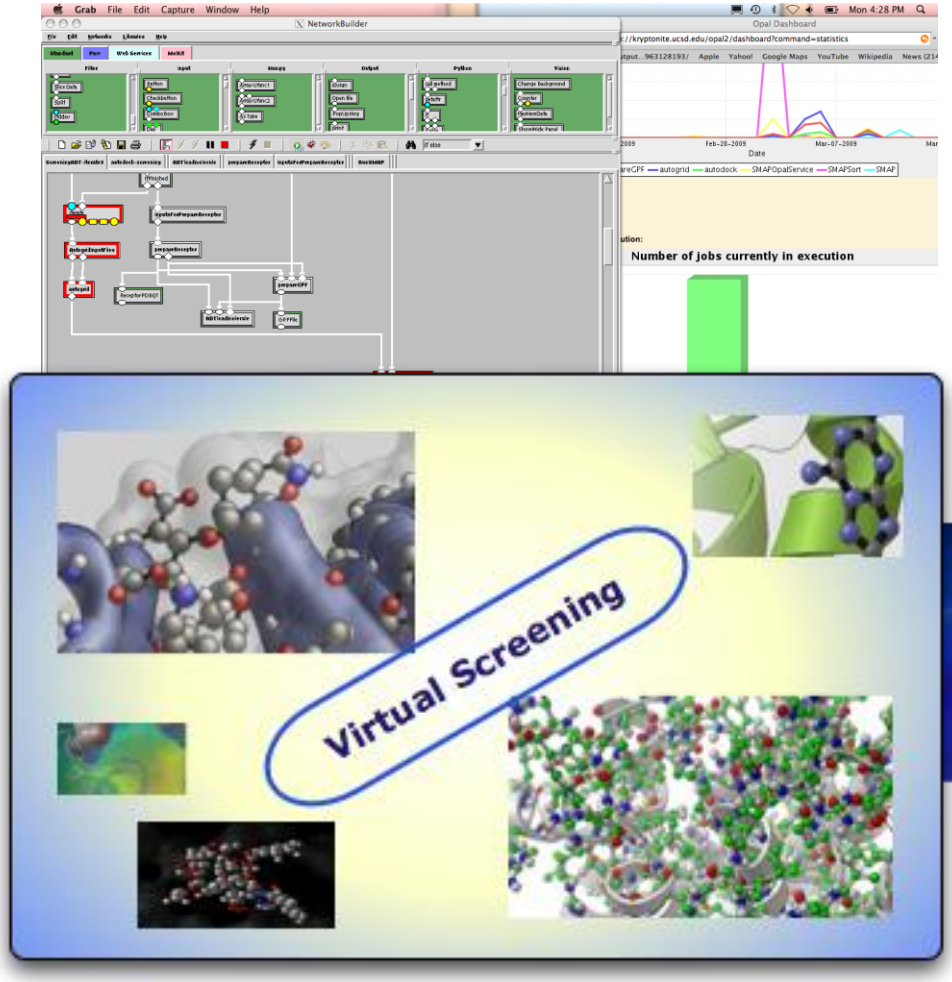
Vision Based Grid Workflow Environment

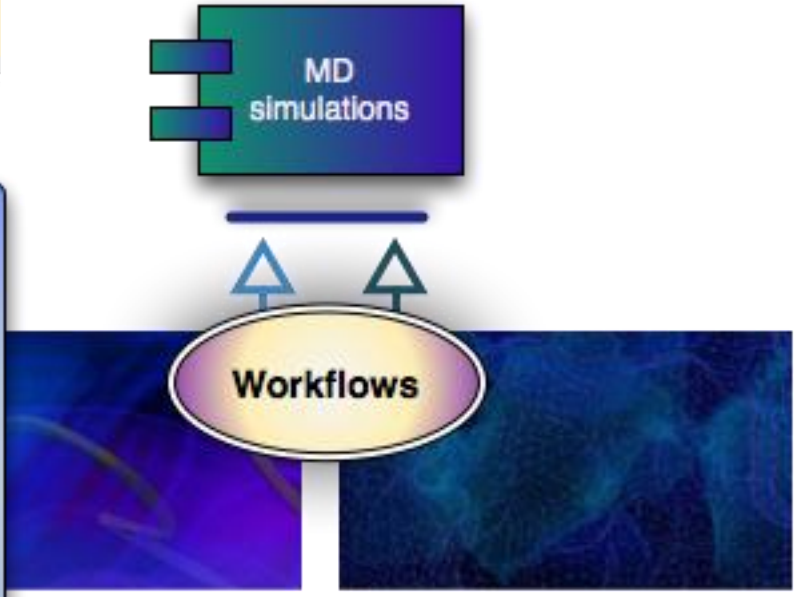


Clementi et al, IEEE eScience 2008

National Center for Research Resources

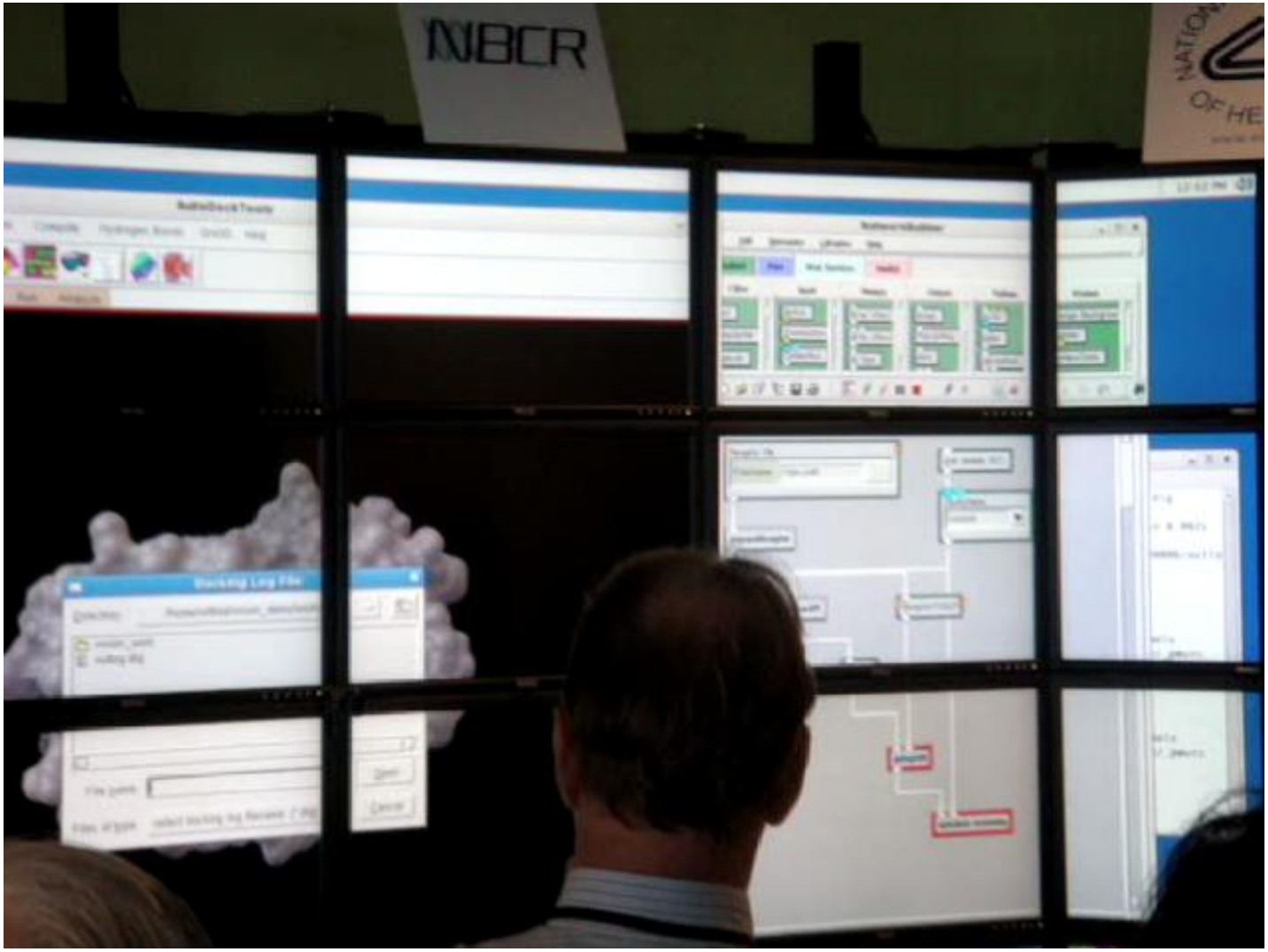
Service as Software: CADD Workflows based upon distributed services



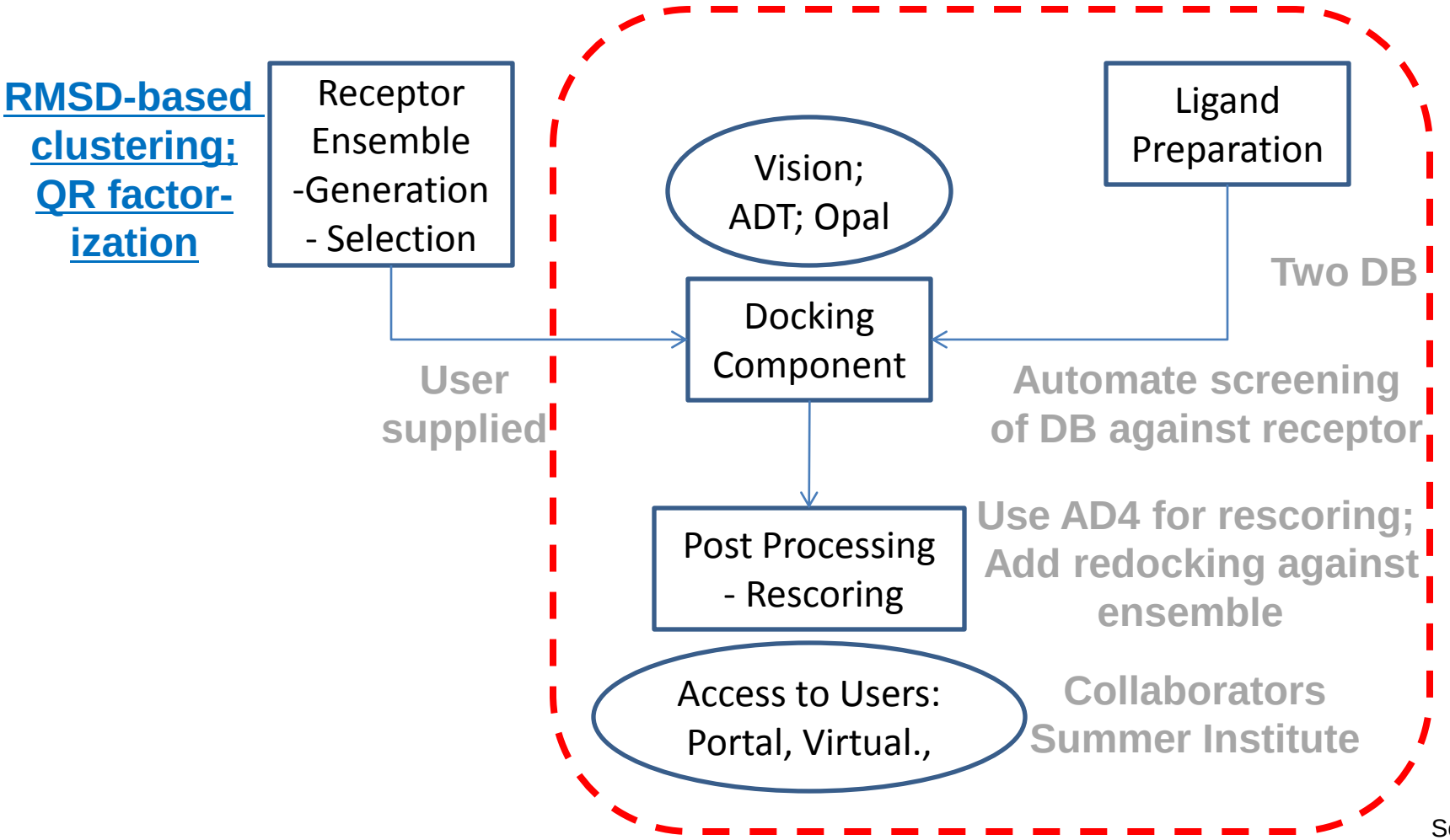


http://cadd.nbcr.net

CADD Pipeline

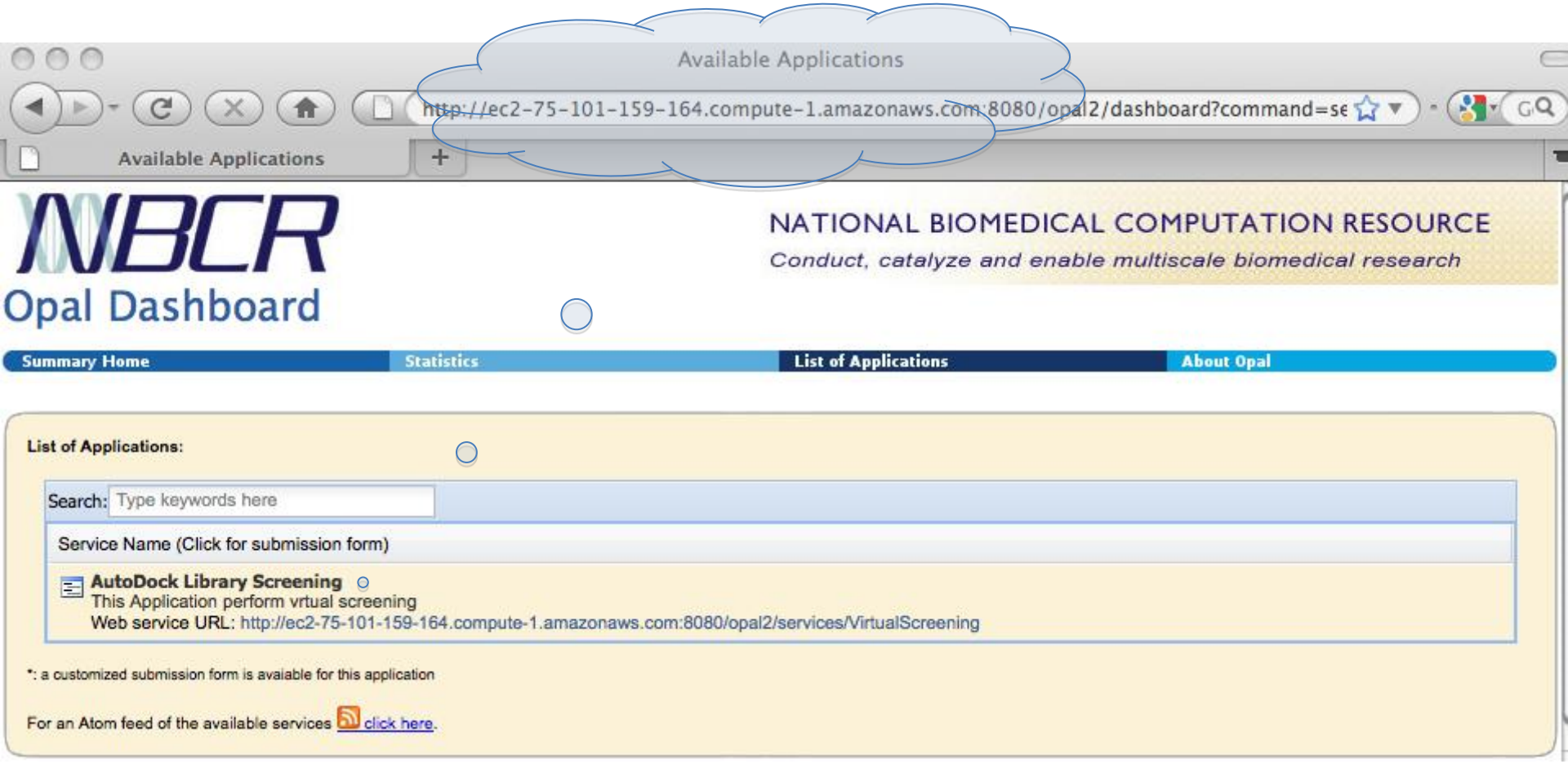


Phase 3: CADD Pipeline
End of Year 2 (April 2011)

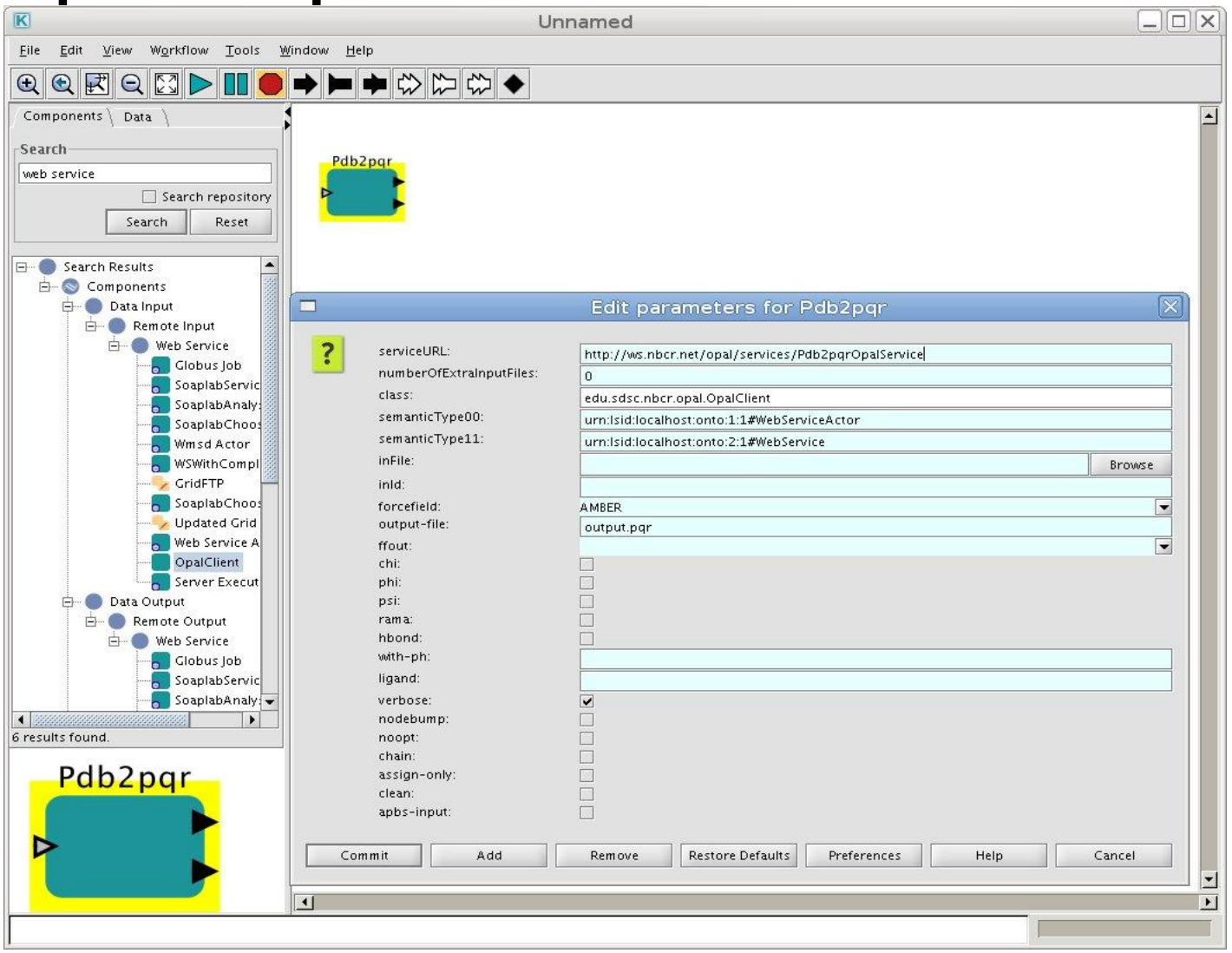


Source: Amaro

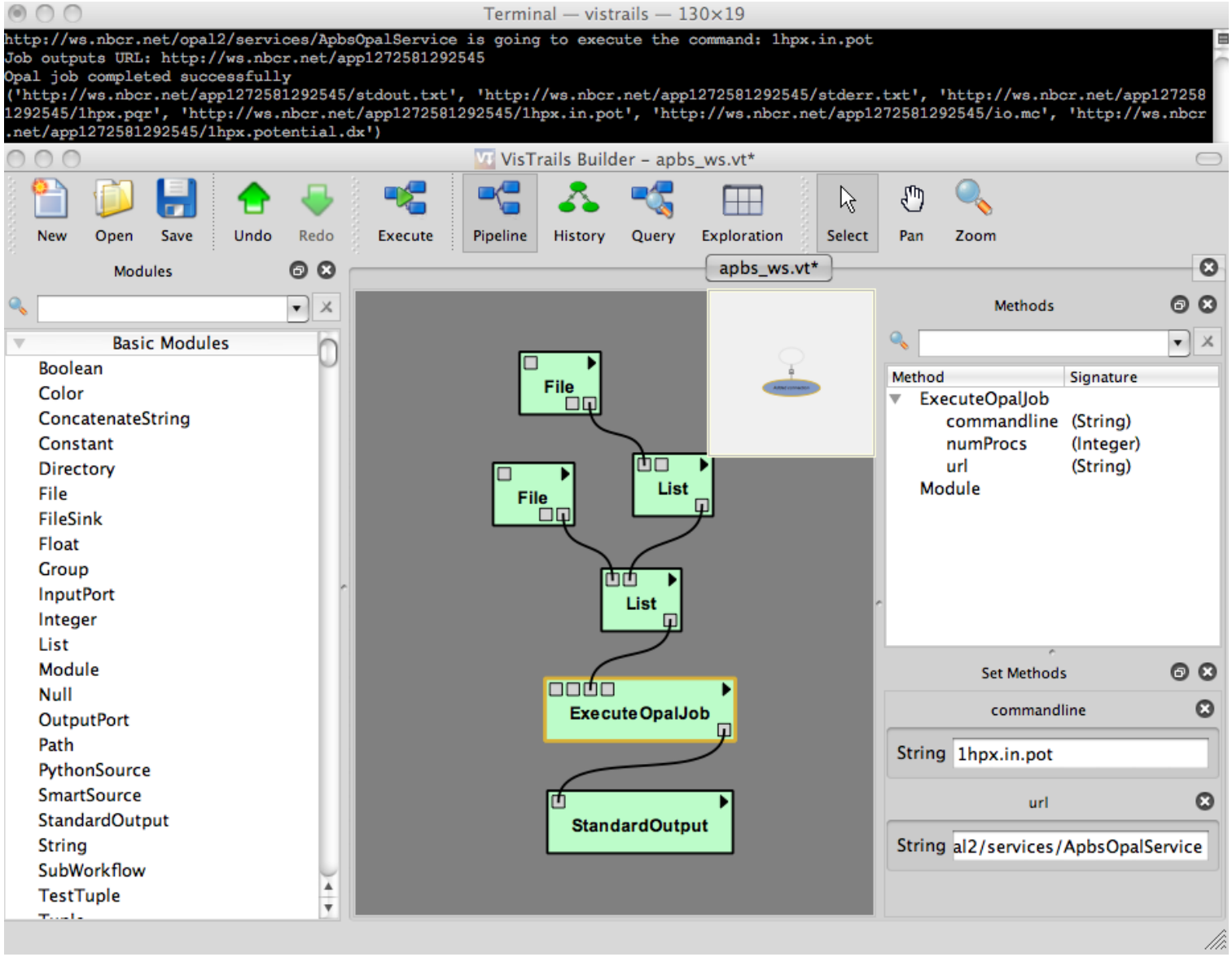
Cloud Computing with Amazon EC2



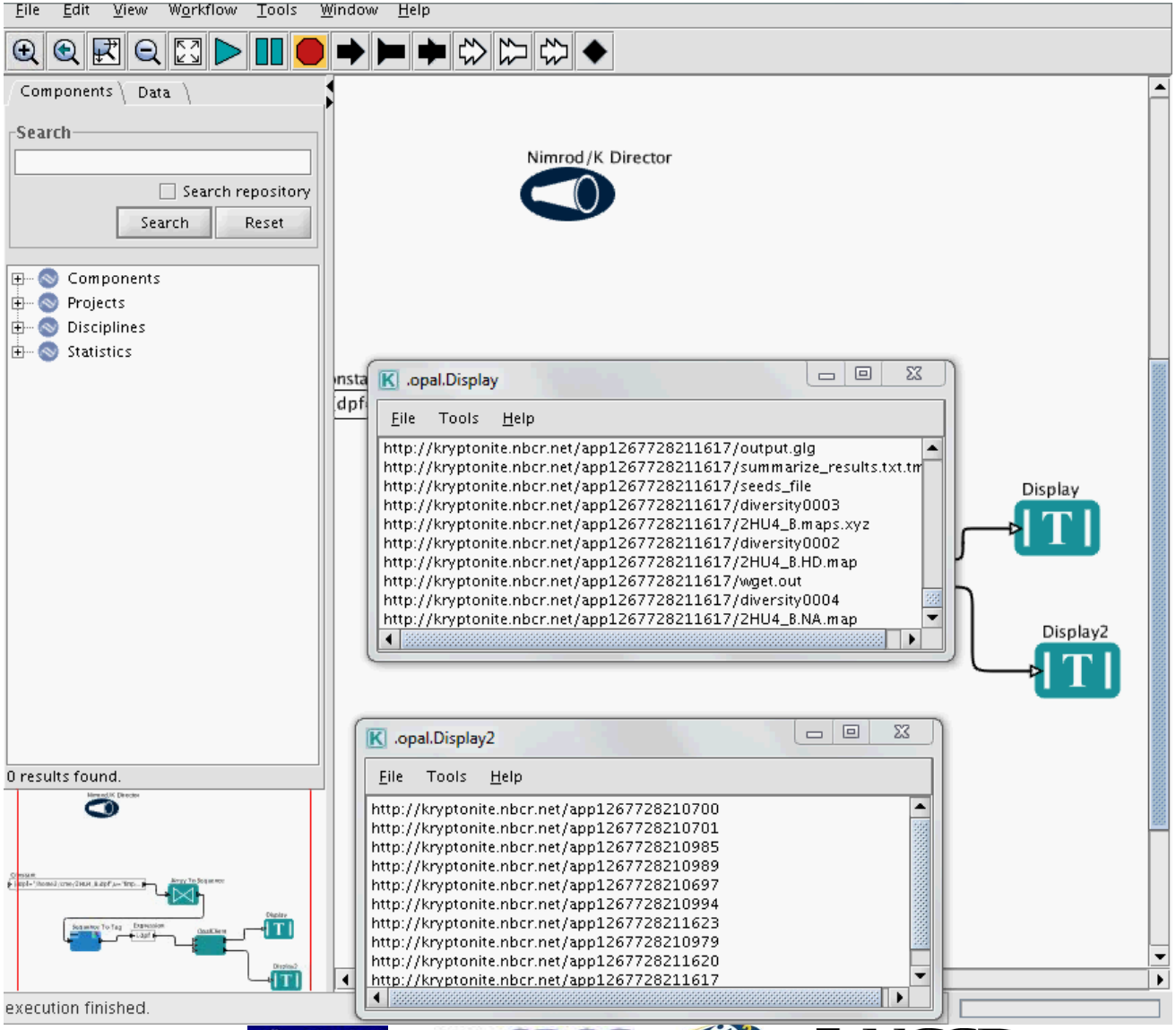
Kepler Opal Web Services Actor



Vistrails and Opal Service for APBS



Nimrod/K and Opal Service for AutoDock4



Enticott et al, PRAGMA 18

Social Networks and Collaborative Environment

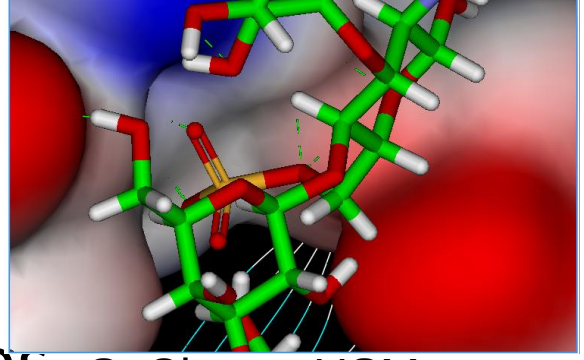
Social Network Site	Number of Users	Features	API Examples
Google	170 million (Gmail)	Google Integrated Suite of Tools	Google Apps Engine
LinkedIn	65 million	Professional	Huddle/Zoho Office Online
Twitter	100 million	Short MMS/SMS	TwitPic
~~Google Wave~~	100,000 X 7?	Upload any file	Google Wave Robot
Facebook	500 million+	Social network	Facebook Apps

Are these too big to fail?
Utility Computing finally?

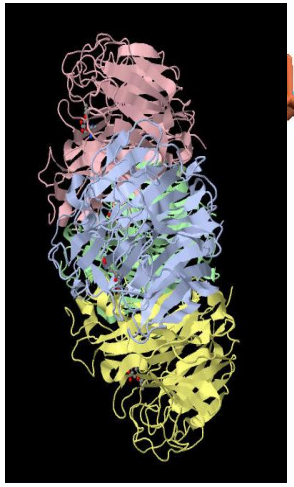
NBCR

PRIME 2010

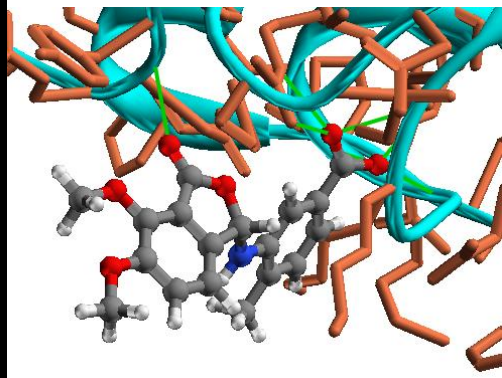
- 13 host sites
- Expanded to five sites
  - Doshisha University, Kyoto
  - National Institute of Information and Communication Technology (NiCT)
  - National Museum of Marine Biology and Aquarium (NMMBA), Kenting, Taiwan,
  - National Taiwan University (NTU),
  - University of Hyderabad
- Continued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



S. Chang, USM

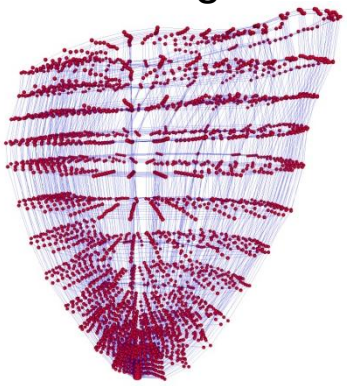


Jade Kwan, NiCT



C. Wong CNIC

M. Mui, U Hyderabad



S. Revelli, Monash

C Lau, Osaka

National Center for Research Resources

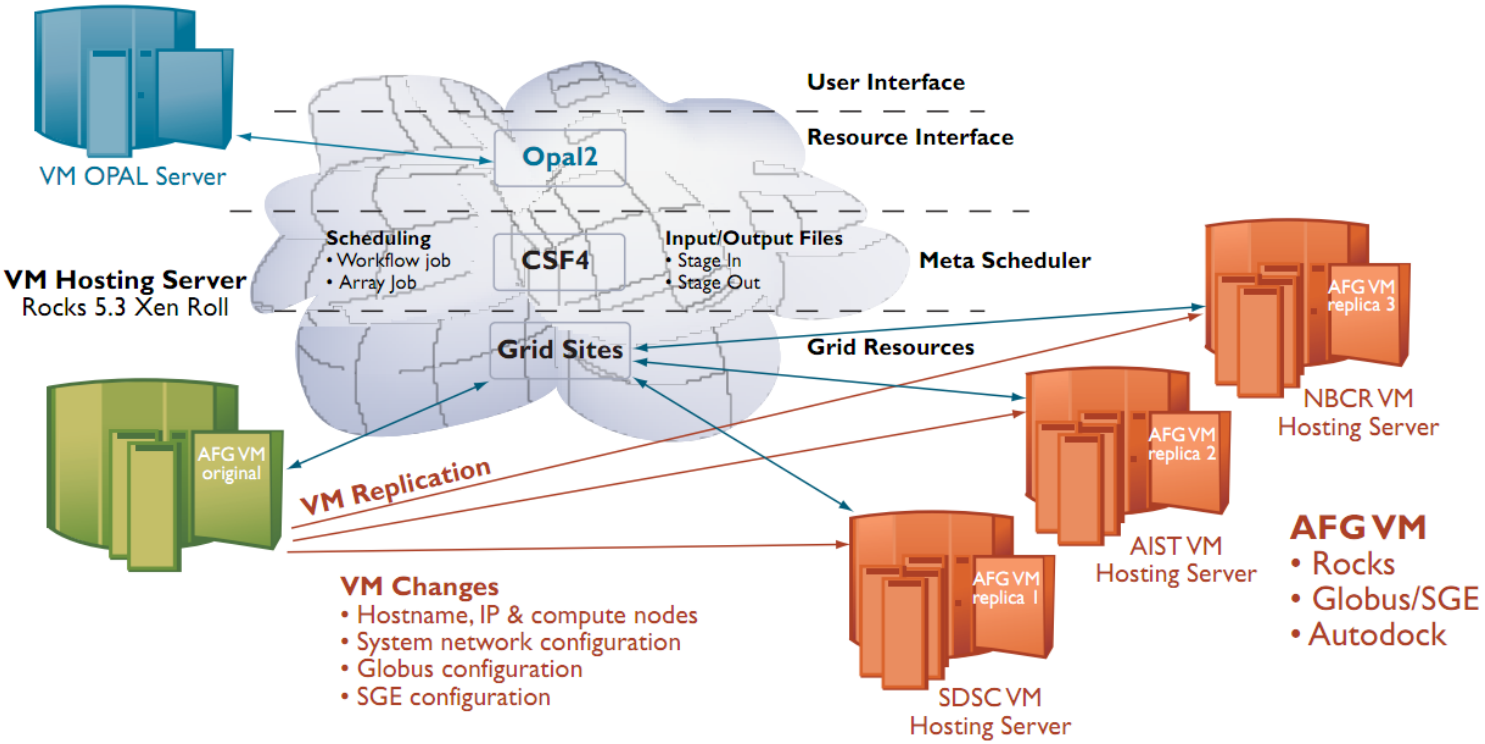
CRBS

SDSC

UCSD

PRAGMA: model for international collaboration in Technology and Science

Fig. 1 PRAGMA Rocks Virtualization Experiment



Broadening Impact of Technology Engaging Future Generations

PRIME Student 2009: Jessica Hsieh, USM

PRAGMA 18

Like the Grid

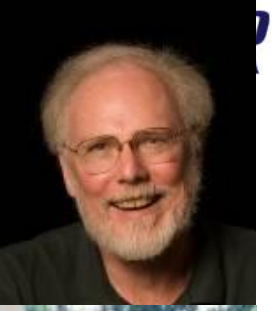
Like the Cloud

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

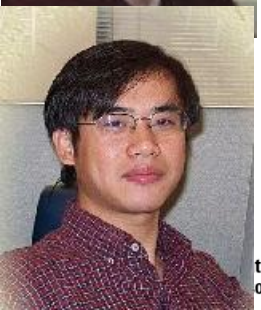
March 2 – PRAGMA Institute on Implementation & Welcome Reception

March 3-4 – PRAGMA Workshop

March 5 – Technology for Coral Reef Observatory

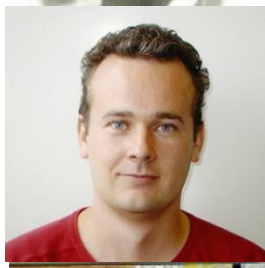
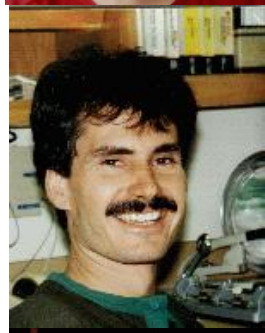








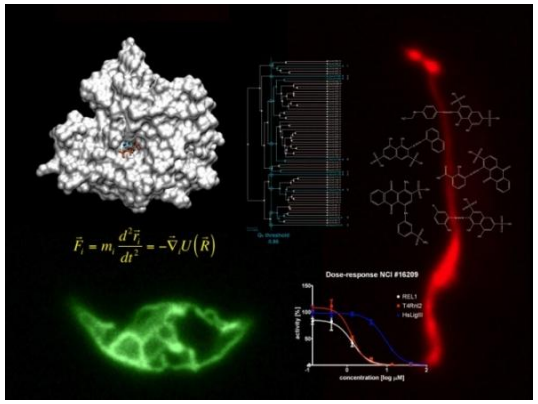


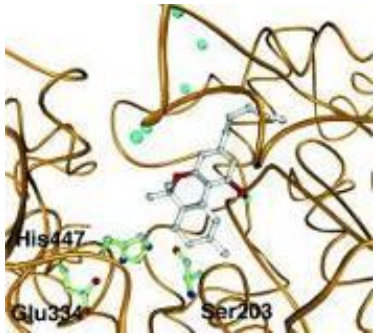

NBCR Researchers
Leading Interdisciplinary Team

- J. Andrew McCammon, Chemistry
- Rommie Amaro, Chemistry
- Michael Holst, Math
- Nathan Baker, Chemistry/Math
- Zeyun Yu, Comp Science
- Anushka Michailova, Biophysics
- Peter Arzberger, Math/Pop BIO
- Andrew McCulloch, Bioengineering
- Roy Kerckhoff, Bioengineering
- Michel Sanner, Comp Science
- Art Olson, Chemistry
- Mark Ellisman, Neuro,
- Philip Papadopoulos, Electrical Engineering
- Wilfred Li, Biochemistry

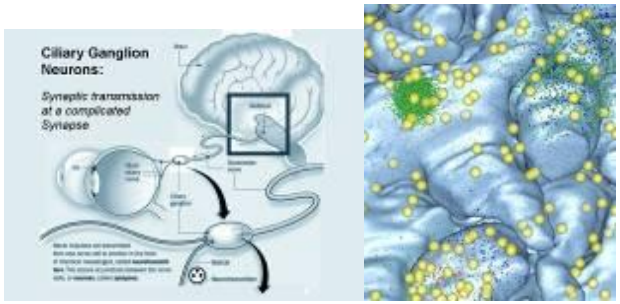
Biomedical Advances Powered by NBCR Tools



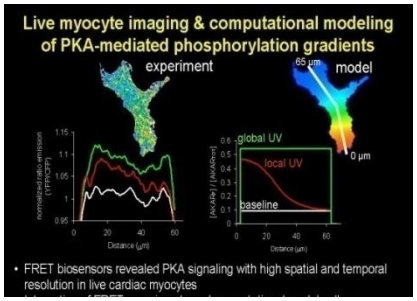
African Sleeping Sickness, Amaro et al, PNAS 2008



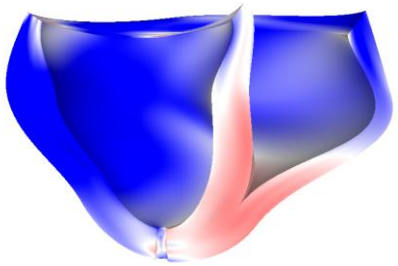
Alzheimer's Treatment Eubanks et al, Mol. Pharm 2006



Neurotransmission Insight: Imaging and Modeling; Coggan et al Science 2005



Imaging, Modeling and Cell-level Understanding Saucerman et al PNAS 2006



Biventricular pacing and scar size Kerckhoffs et al.. Med Image Analysis 2008

- Training Sessions
  - Cluster/Grid Computing & Workflow Management
  - Programming Scalable Scientific Applications
  - Virtual Screening & Computer Aided Drug Design
  - Computational Cardiac Electrophysiology and Mechanics
  - Molecular Electrostatics and Diffusion
  - Multiscale Imaging Analysis and Visualization

Builds Community

Leads to improved on-line training materials

http://si.nbcr.net

National Center for Research Resources | CRBS | SDSC | it2 | UCSD

Acknowledgment

Science:

- *Irene Newhouse*
- *Dong Xu**
- **Youngjin Choi**
- Hsing Pao
- **Jung-Hsin Lin**
- ***Habibah Wahab***
- Maqsudul Alam
- *Doman Kim*

Technology:

- Zhaohui Ding
- Guanyuan Liu
- **Osamu Tatebe**
- Yusuke Tanimura
- Jonghyun Lee
- **Kai Nan**
- **Xiaohui Wei**
- **Suntae Hwang**
- Aimee Li, MURPA

Collaborators:

Peter Arzberger,UCSD

David Abramson, Monash University, MUPRA, PRAGMA

Andy McCammon, UCSD

Rommie Amaro, UCI

Michel Sanner, TSRI

Sriram Krishnan, SDSC

Programming Staff:

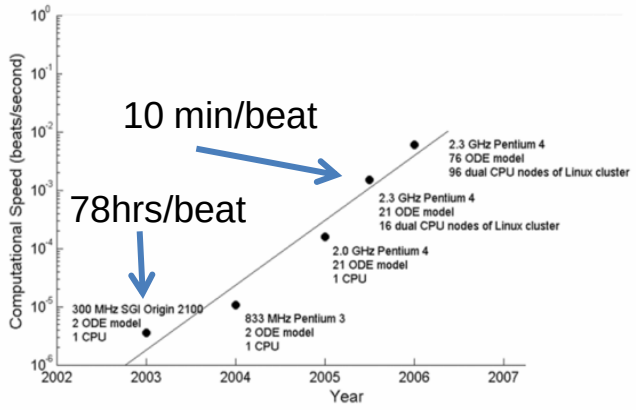
Nadya Williams,

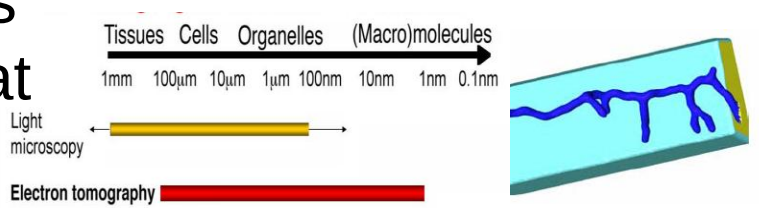
Jane Ren

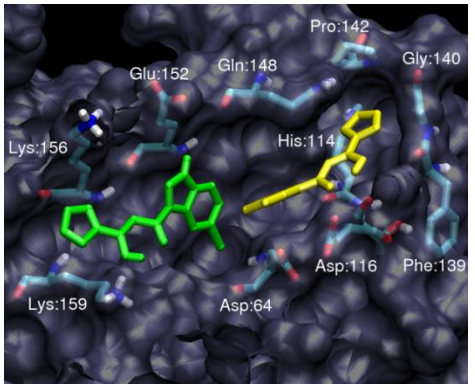
http://www.pragma-grid.net
http://www.nbcr.net

Computational Translational Medical Science: Potential in the next 5 years

- Improving treatment planning for heart disease through individual patient modeling
- Understanding role of realistic subcellular structure on functions of cells via multiscale modeling at the mesoscale
- Speeding drug discovery by creating, improving the efficiency of, and enhancing computer aided drug discovery pipeline







Tools, Simulation Packages and Pipeline, CI