Computing the Horribleness of Soft Condensed Matter

Eric Jakobsson
University of Illinois at Urbana-Champaign
National Center for Supercomputing Applications
National Center for the Design of Biomimetic Nanoconductors

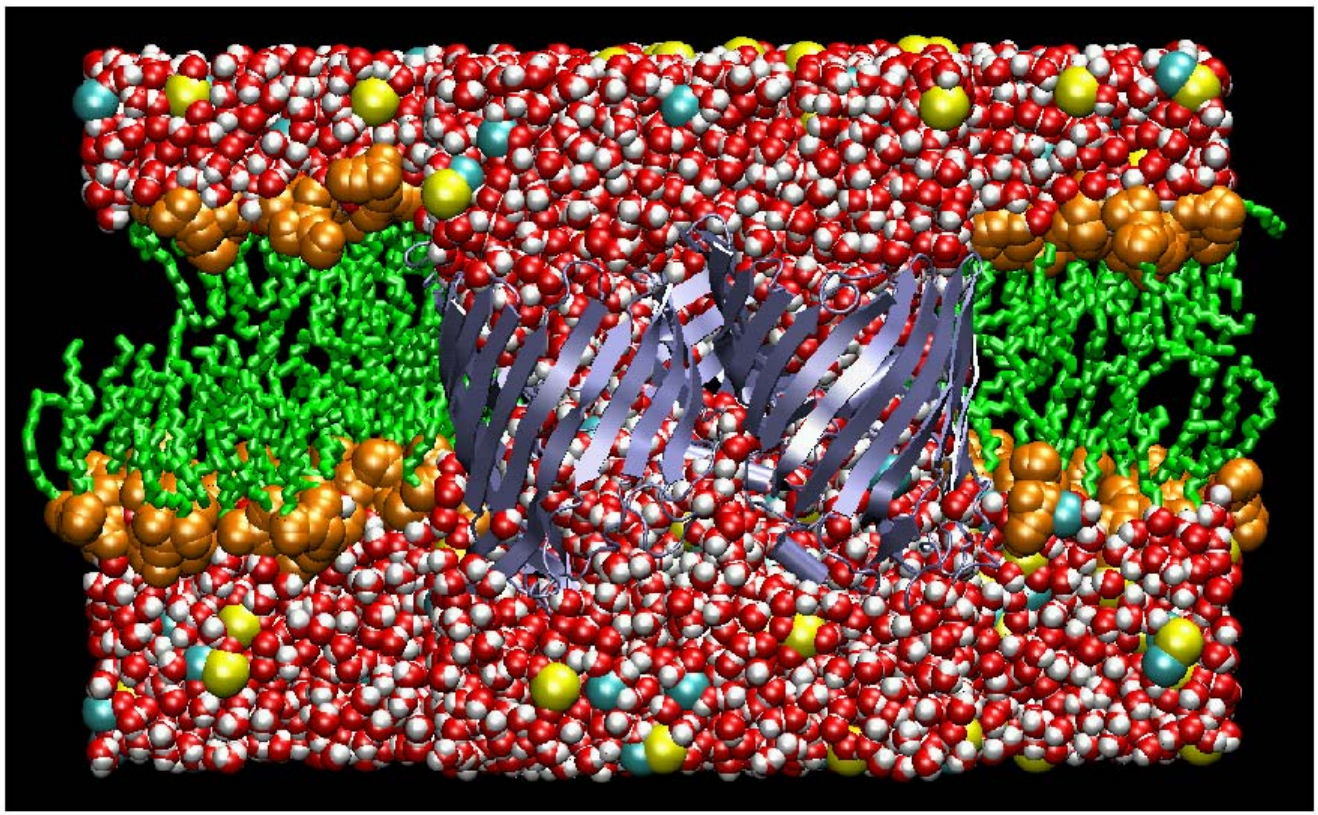
For the Purdue CRI Seminar

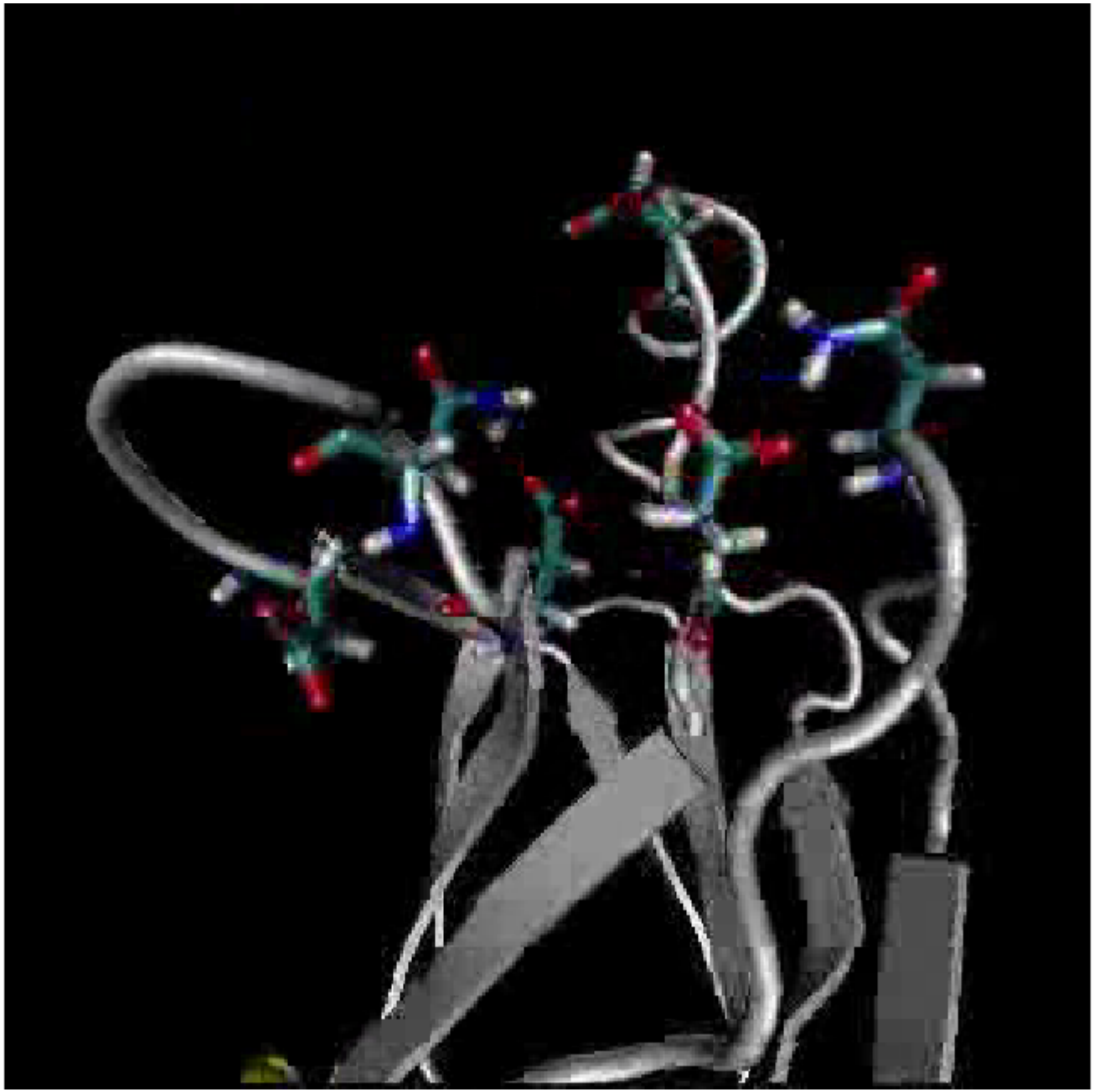
October 3, 2007

The three big areas of computational (mathematical?) biology

- 1. Physics-based computational biology---an extension to biological systems of physical chemistry. In this view, biological systems are messy soft condensed matter. In this view of biology, the complexity of biological behavior arises from the complexity of soft condensed matter. The mathematical horribleness (intractability) of this description is that all systems described in this way are connected to the rest of the world by an effectively infinite number of variables. However, this is a very satisfactory mode of describing biology for those who just love doing enormous computations.
- 2. Dynamical systems based computational biology---The next stage up the hierarchy of organization of biological systems---mathematical language is differential equations or random walk (agent based simulations), with the variables or the agents being lumped entities.
- 3. Central Dogma based computational biology---Biology is based on information flow from the code of the DNA through the RNA to the gene product (proteins).

An ion channel protein embedded in a biological membrane—a typical bit of biological soft condensed matter

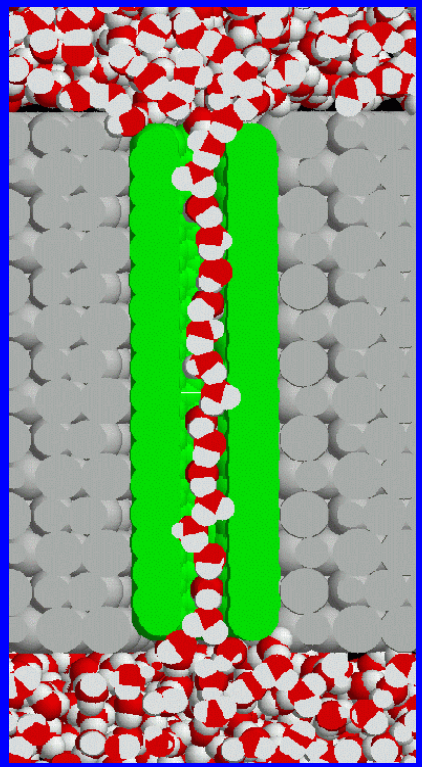




Weirdness at the Nanoscale---why nano and bio are intertwined and both strange

- Biological systems are heterogenous at the nanoscale.
- The nanoscale is where two distance scales clash—the atomic scale (angstroms) and the continuum scale (down to microns)

System Setup



- Carbon nanotube (fixed)

  8 sizes of nanotubes ranging 5.4 - 16.3 Å dia., armchair (5,5) - (12,12). Length ~40 Å.

- Bilayer mimetic (hcp CH_2's, fixed)
- SPC/E water (T = 300 K)

  $2q_H = -q_O = 0.8476\ e$

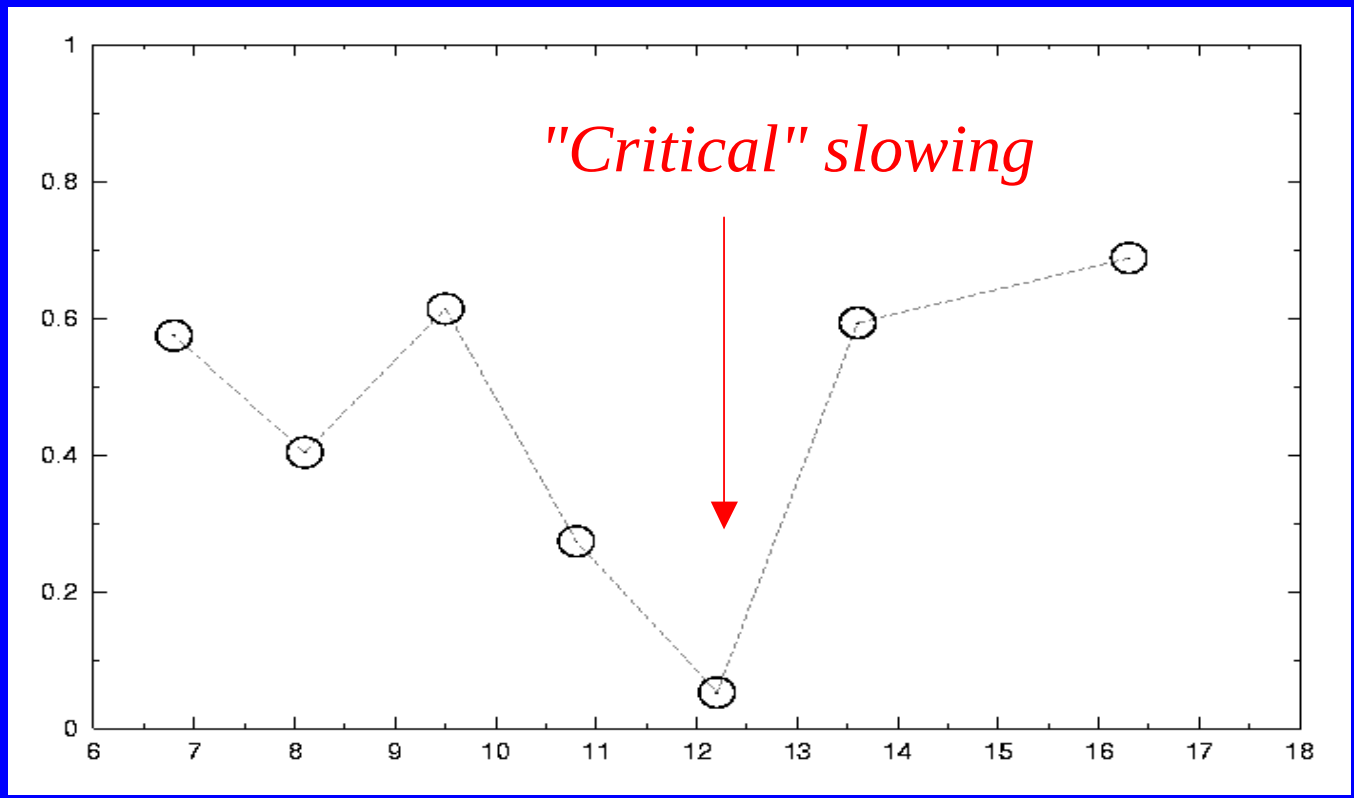
  - Electrostatics: PME
  - Nose-Hoover coupling
  - Pressure piston (P_z = 1 bar)
  - Runs of ~2 ns each using GROMACS

    (See www. gromacs.org)

• Simulations done on NCSA IA32 and IA64 Linux superclusters

Relative Diffusion coefficients

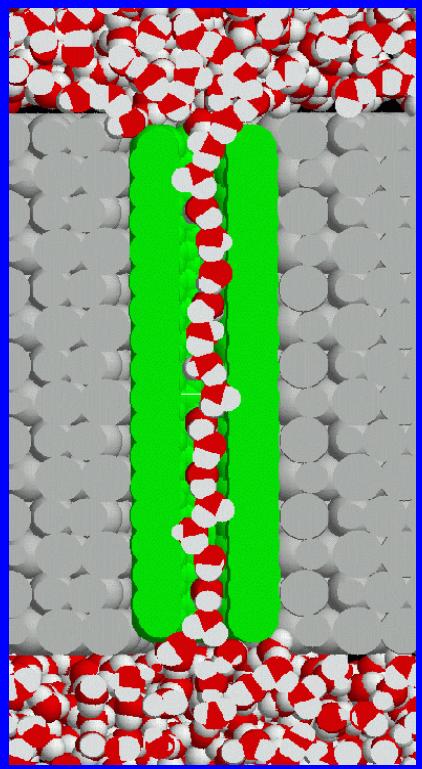
Water in Nanotube vs. bulk(=1)



Snapshots of Water Configurations

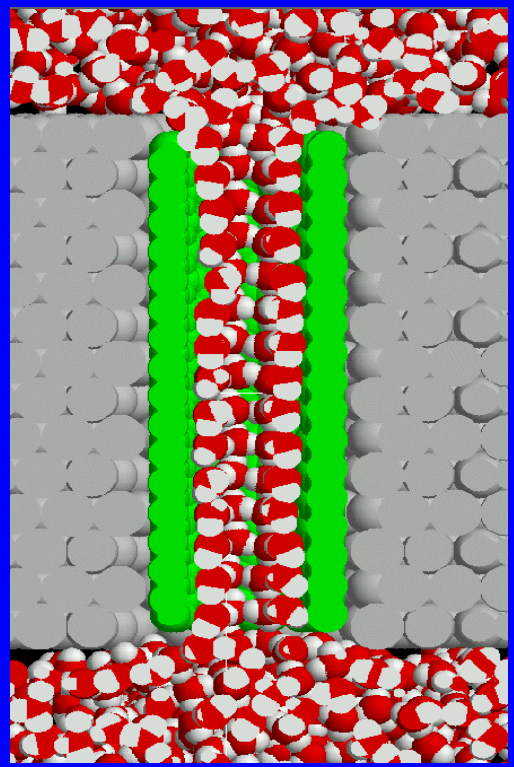
T = 300 K (water), fixed tube & slab

(6,6)



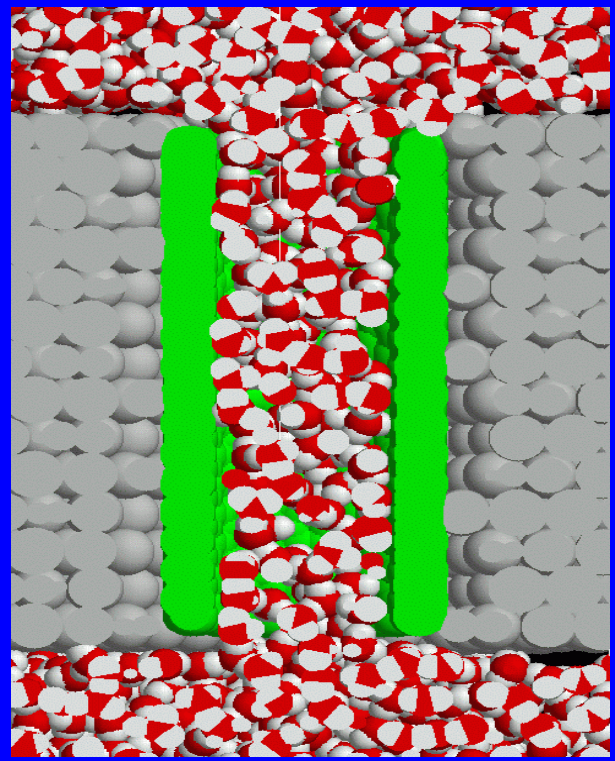
- Single file
- 1-D hydrogen bonding

(9,9)



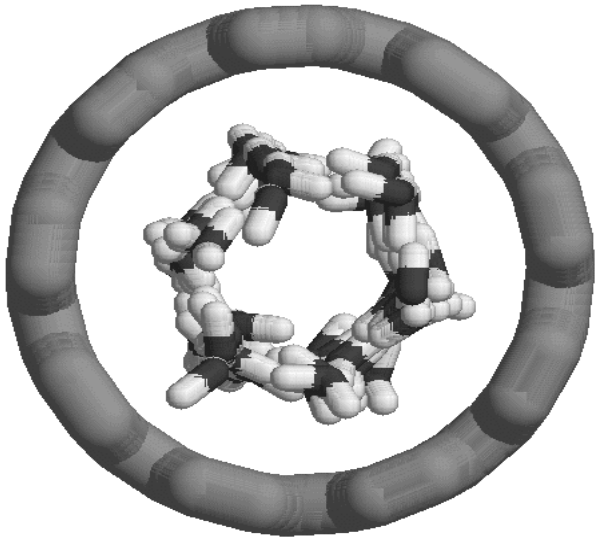
- Critical size for order
- 2-D hydrogen bonding

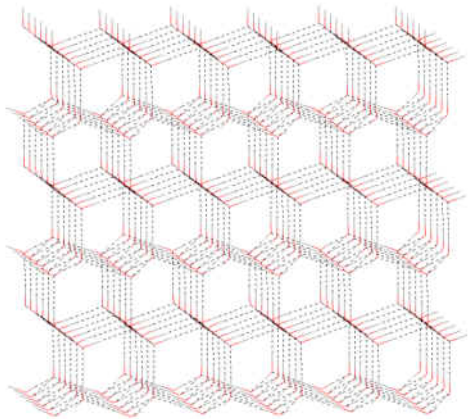
(12,12)

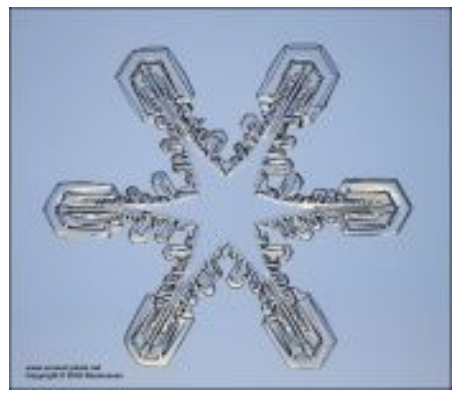


- Bulk water properties not yet achieved

Snapshot from end of "critical diameter" nanotube of water structure, showing configuration similar to hex ice at 300K, (basis of snowflake symmetry) (Snowflake pictures by Susan Rasmussen)



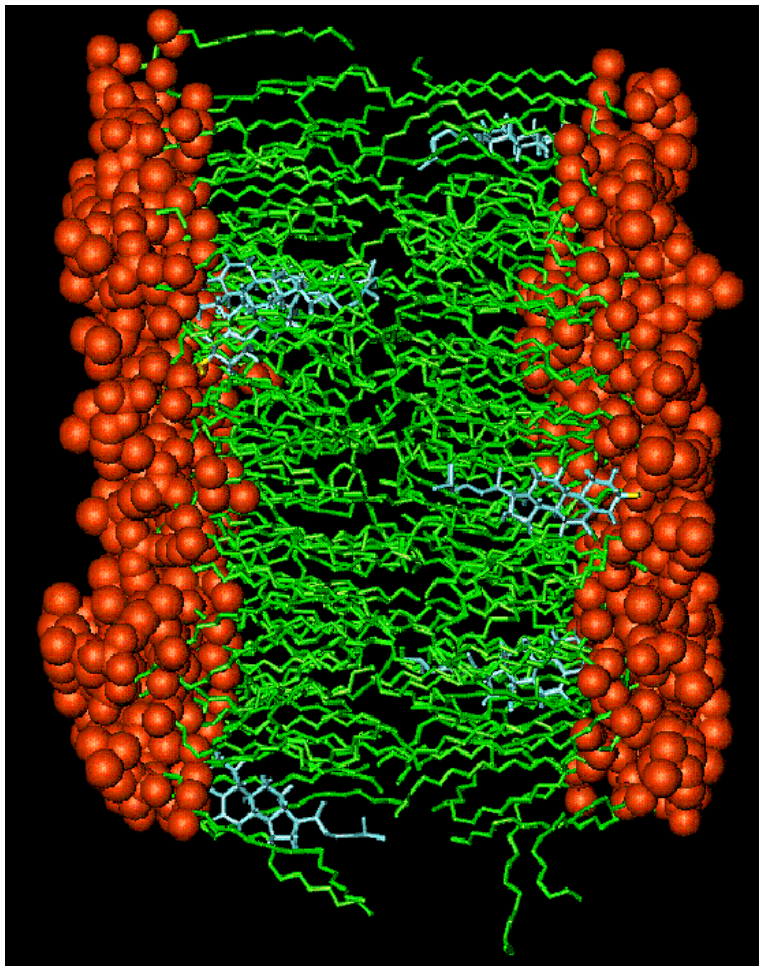


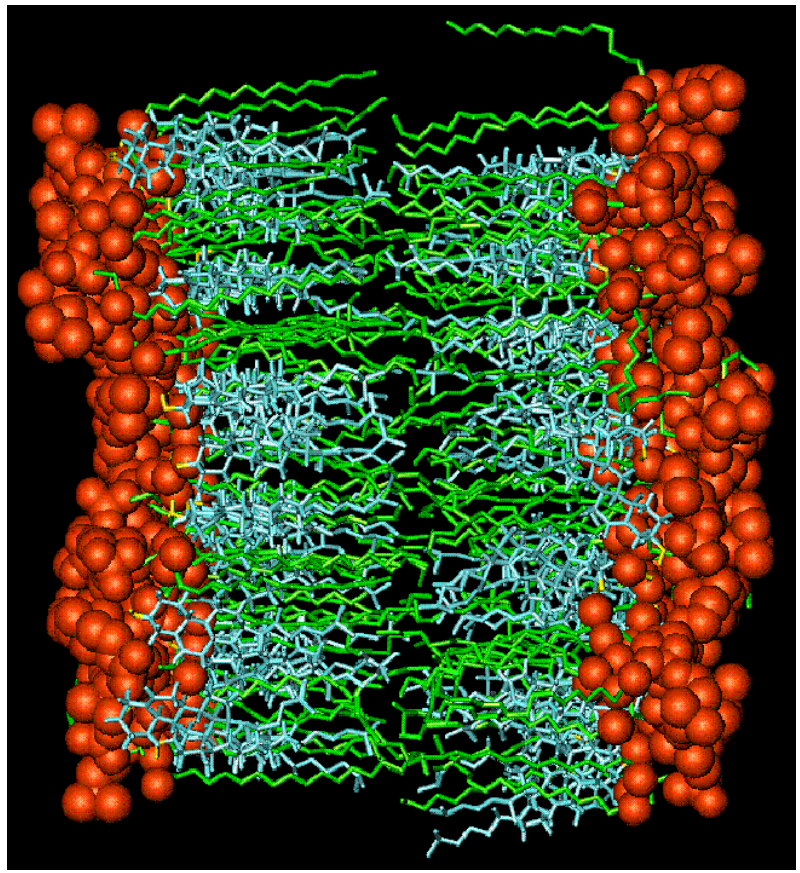


Cholesterol in Membranes

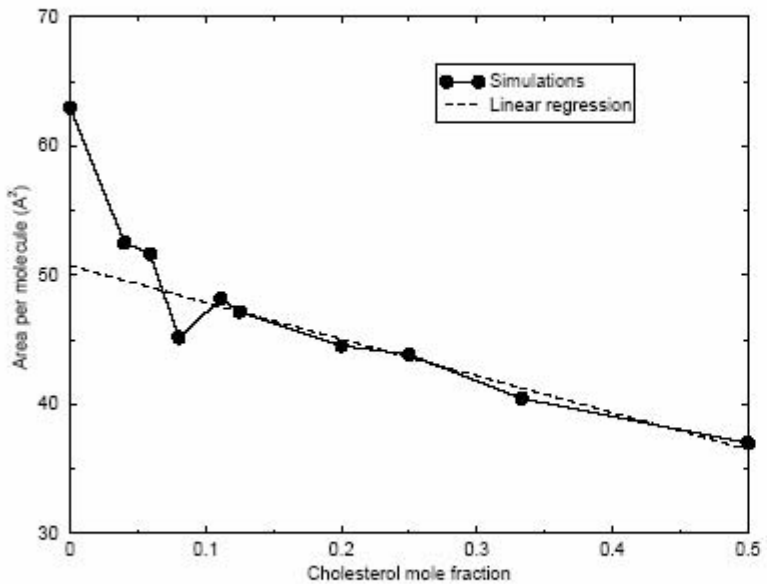
- Background: Multi-gene functional networks in human and other animal membranes are organized into cholesterol-rich regions called “rafts”. Driving force for raft formation is action of cholesterol to “condense” lipid membranes.
- Problem: How can we understand the molecular physics of the process by which cholesterol causes condensation of lipid membranes?
- Solution: Massive computer simulations for cholesterol-lipid membranes at different composition, coupled with analysis of overall phase and detailed molecular conformations—computational physical biochemistry and computational spectroscopy.

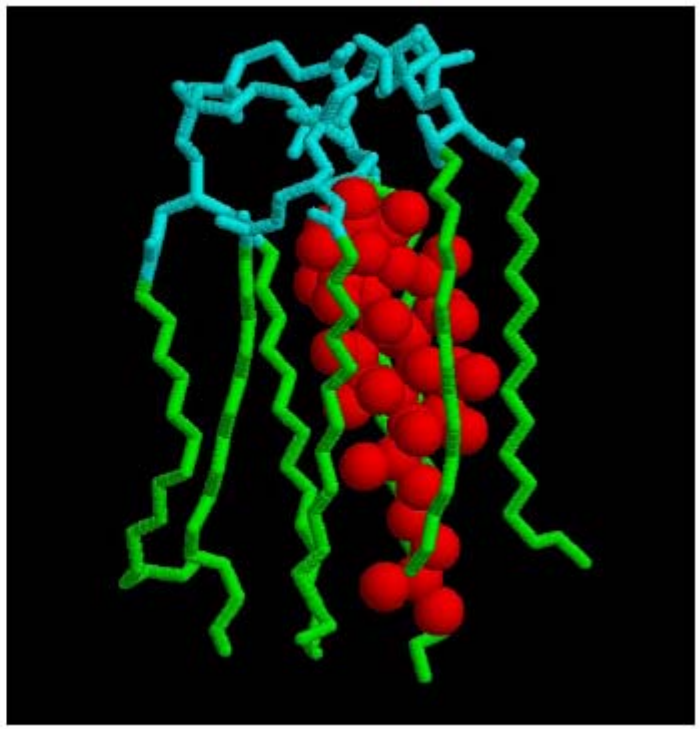
Representative results of the simulations. On left hand side is 16/1 lipid/cholesterol ratio. The lipid chains (green) are highly disordered, indicating fluid state. Right-hand side is 1/1 lipid/cholesterol. Lipid chains are much straighter, indicating gel-like state.



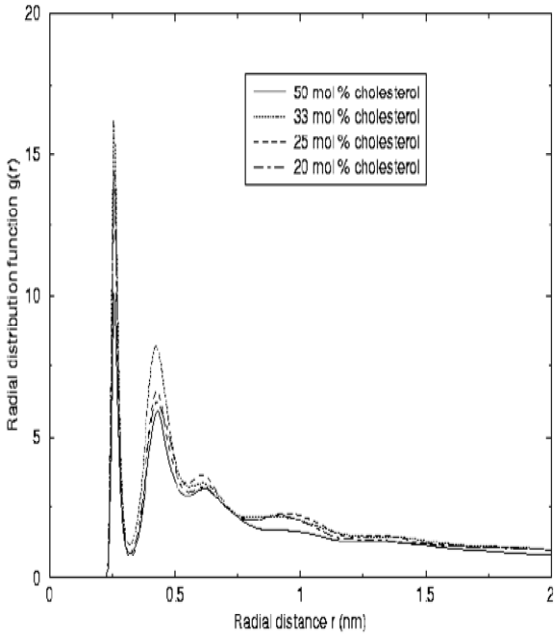


Condensing effect of cholesterol in membranes as simulated (matches experiments).





Computational Spectroscopy: The figure below left analyzes the distribution of distances between selected atoms in the cholesterol and lipid molecules for simulations at different cholesterol concentrations. The fact that the peaks are in the same place for all cholesterol concentrations is a "signature" of a characteristic structural building block for the cholesterol-lipid condensed complex.

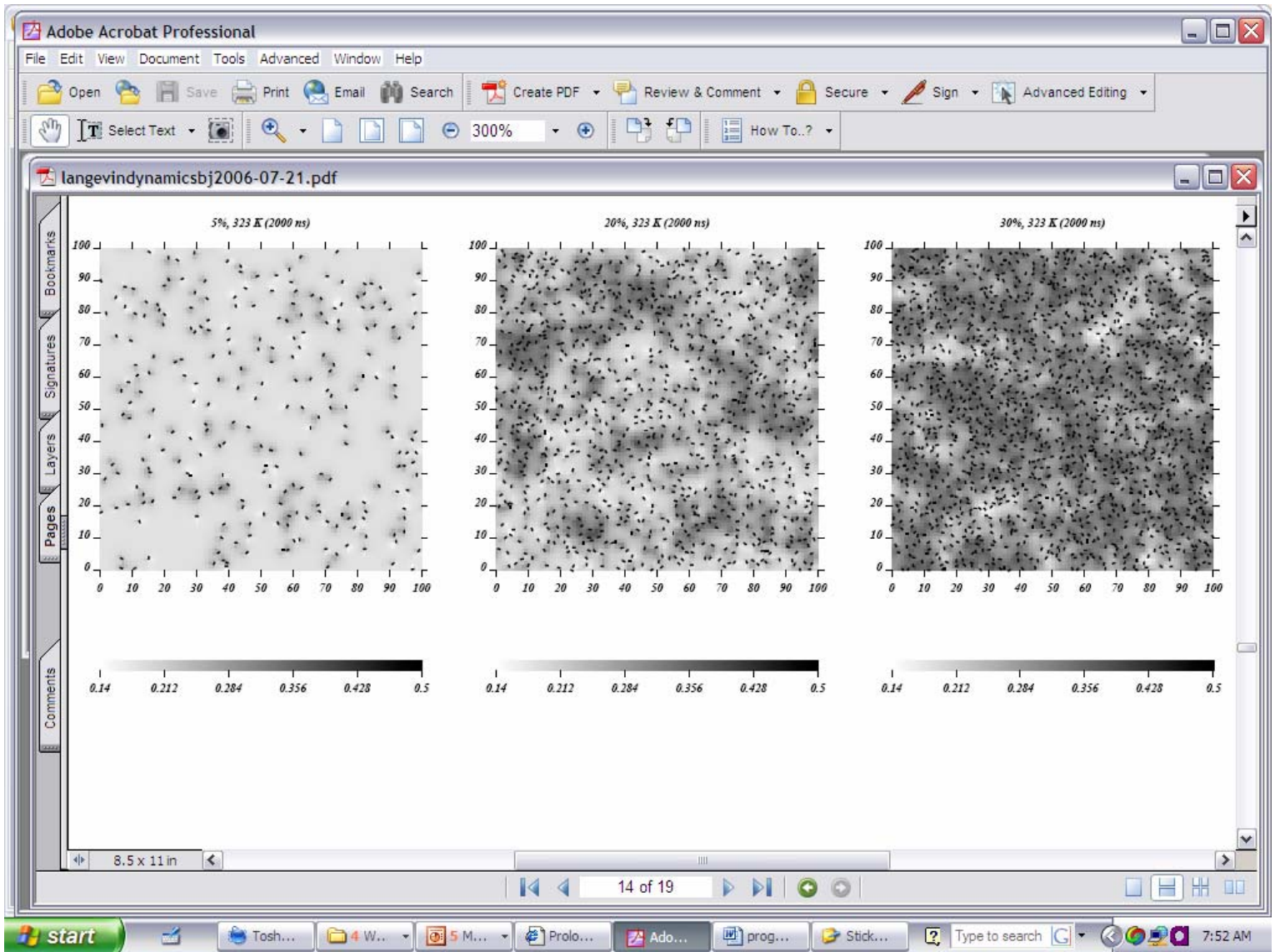


- **Statistical Mechanical Significance:** By Boltzmann statistics, it is possible to convert the radial distribution function into a pairwise potential of mean force, and do 2-D Brownian dynamics for how cholesterol and PC move across the membrane surface relative to each other.

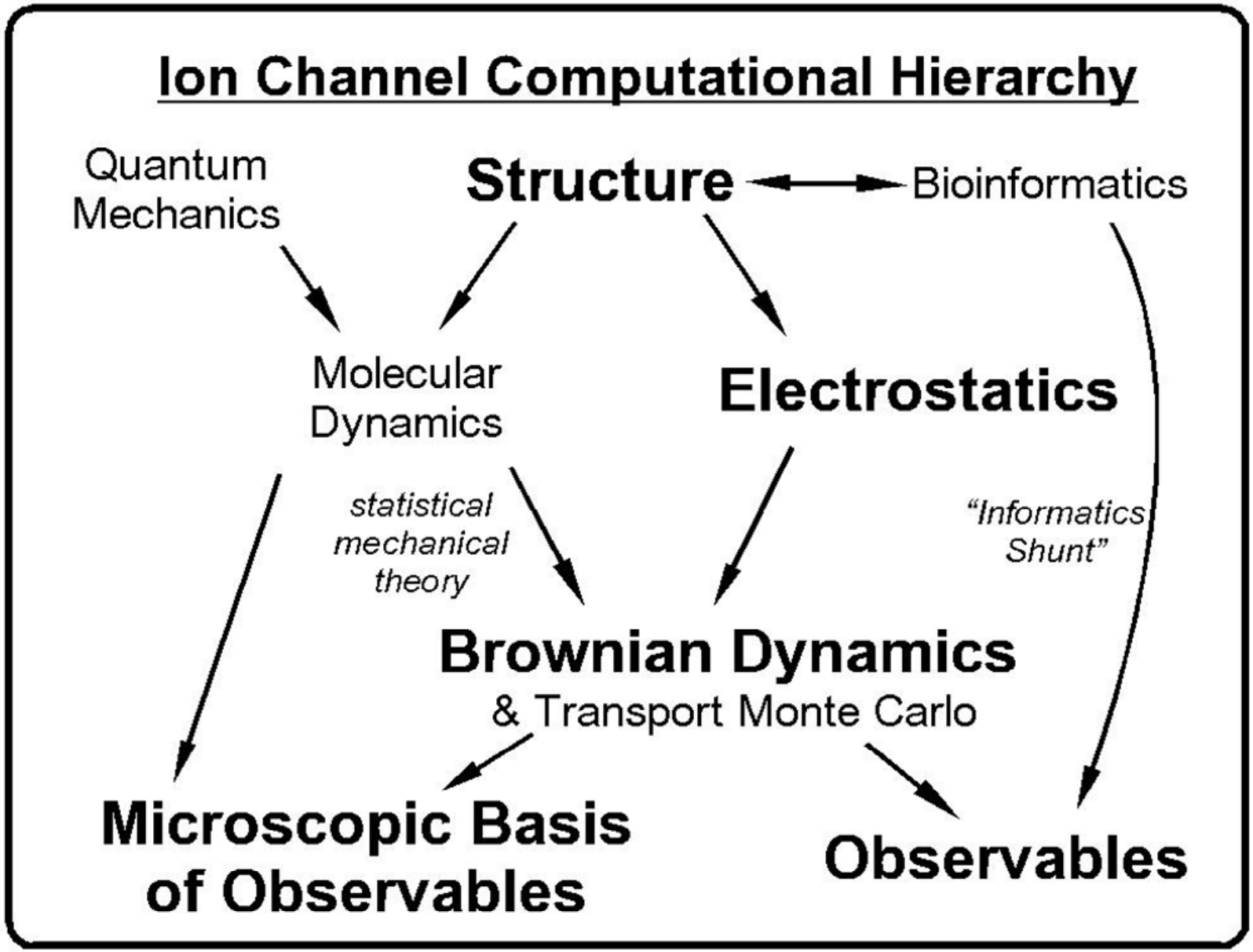
From the Scott/Jakobsson/Grama Labs--bridging time and length scales in understanding membrane dynamics and organization

- Problem: No supercomputer that we can foresee will be able to simulate, from atomistic molecular dynamics, domain formation and phase relationships in heterogeneous membranes
- Our labs' (Scott/Grama/Jakobsson) approach is to use atomistic molecular dynamics simulations to parameterize Mean Field Langevin Dynamics simulations that can span large distances and long time scales. In initial implementation, we do molecular dynamics on DPPC-cholesterol
- In our specific implementation, cholesterol molecules are discrete particles, while other membrane lipids are represented by continuous fields of concentration and order parameter. Field evolution dynamics are derived directly from analysis of molecular dynamics output of corresponding membrane, specifically from correlation analysis of neighbor interactions.
- Method has been validated by successfully reproducing heat capacity through phase transitions, and phase boundary tie lines, for dppc-cholesterol mixtures.

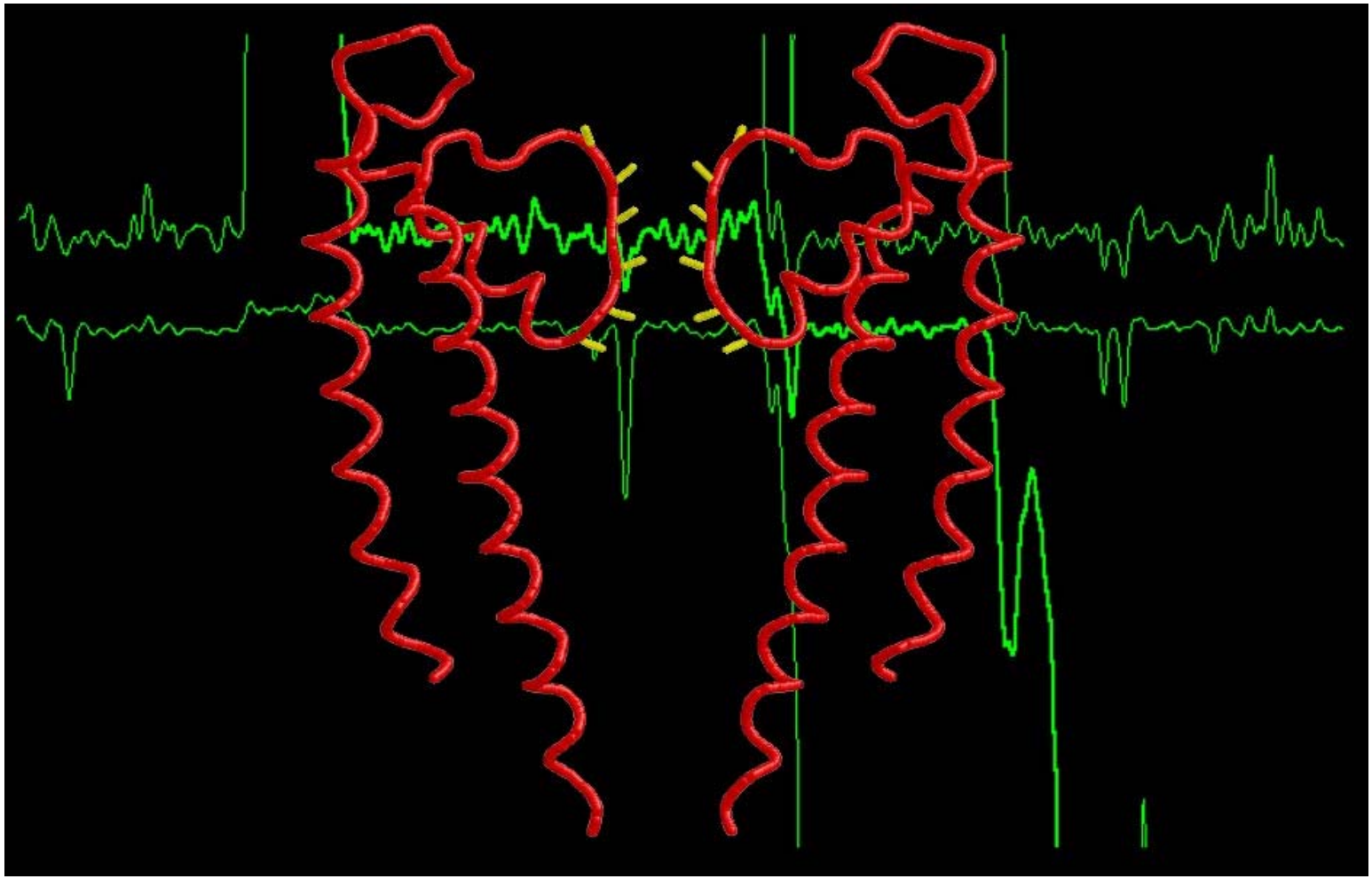
Results of large scale, long time Langevin Dynamics simulations, parameterized by the md.



Social analogue to membrane domain formation---swimmers in a public pool in New York City


Ion Channel Computational Hierarchy
Quantum Mechanics
Structure
Bioinformatics
Molecular Dynamics
Electrostatics
statistical mechanical theory
"Informatics Shunt"
Brownian Dynamics
& Transport Monte Carlo
Microscopic Basis of Observables
Observables

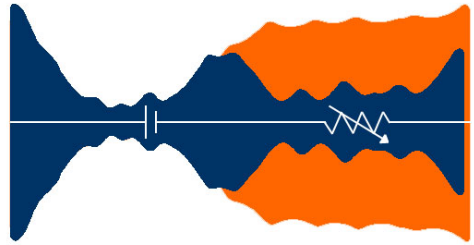
Brownian dynamics simulation of potassium ions moving through a protein channel—parameters derived from molecular dynamics and electrostatics.



Next big category of computational biology

- Dynamical systems based computational biology---The next stage up the hierarchy of organization of biological systems---mathematical language is differential equations or random walk (agent based simulations), with the variables or the agents being lumped entities.

Ncdbn… *The National Center for the Design of Biomimetic Nanoconductors*



Eric Jakobsson 1, Narayan Aluru 1, Hagan Bayley 2, Jeff Brinker 3, Scott Feller 4, Mark Humayun 5, David A. LaVan 6, Gerhard Klimeck 7, Kevin Leung 8, Michael McLennan 7, Steve Plimpton 8, Umberto Ravaioli 1, Susan Rempe 8, Benoit Roux 9, Marco Saraniti 10, H. Larry Scott 10, Xinguang Zhu 1

1 University of Illinois at Urbana-Champaign

2 Oxford University

3 University of New Mexico

4 Wabash College

5 University of Southern California

7 Yale University

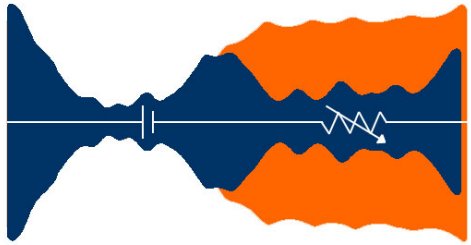
8 Sandia National Laboratories

9 Cornell University

10 Illinois Institute of Technology

Eric J

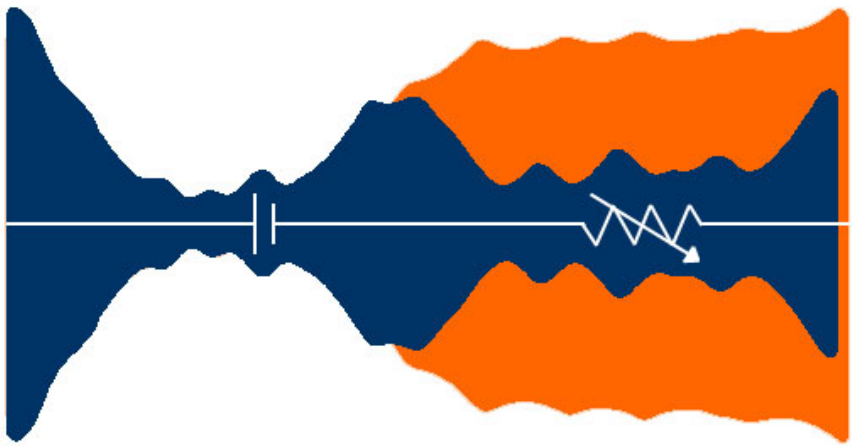
Ncdbn… *The National Center for the Design of Biomimetic Nanoconductors*



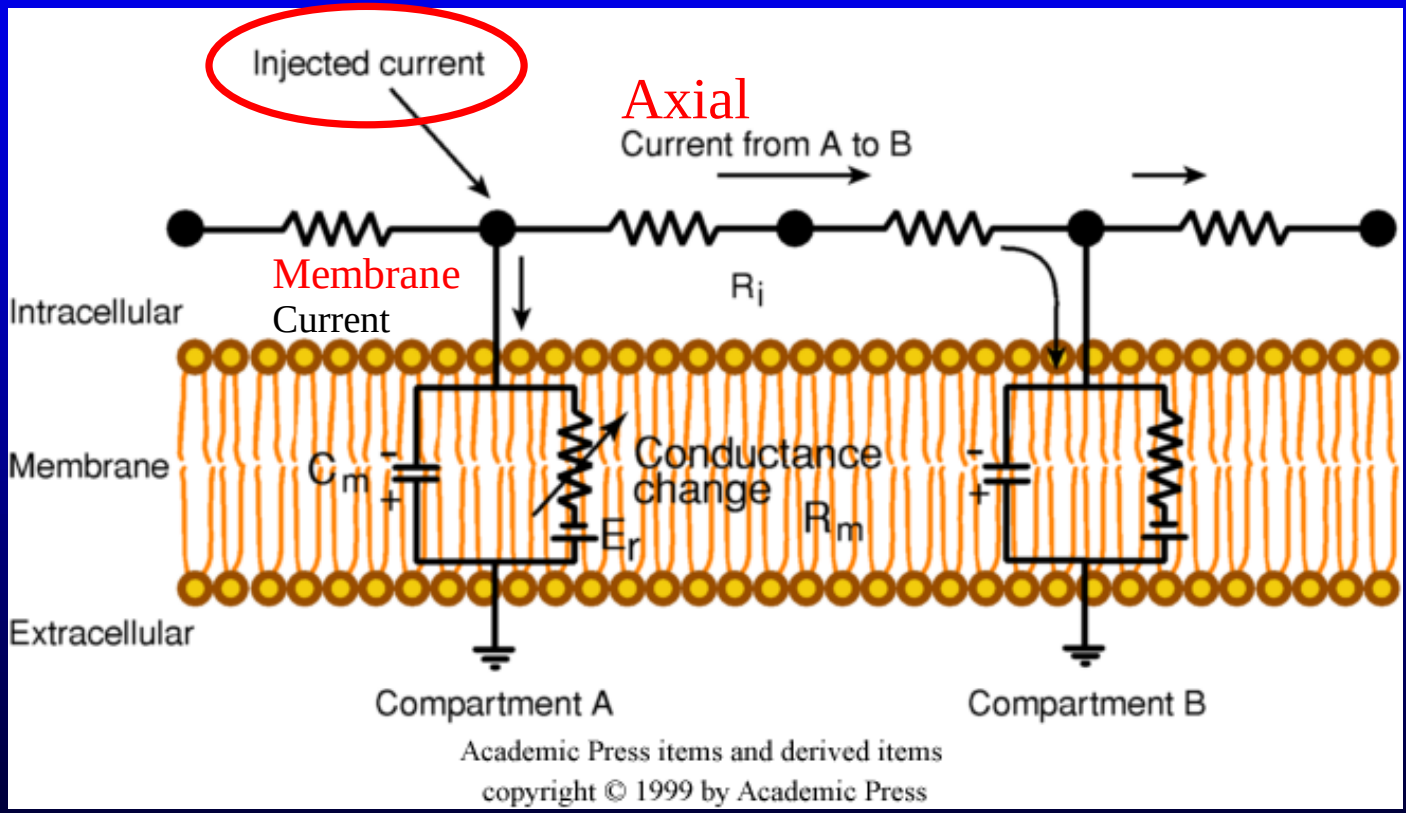
- *Integration of Theory, Computation, Experiment, and Design to develop the Scientific and Technological Foundation for Synthetic Cells and Organelles.*
- Core Technology---Membranes self-assembled on nanoporous solids.
- Overall approach---Concentrate intracellular contents in nanoporous solid by evaporation of surrounding aqueous solution; then self assemble “designer membrane” on the surface by evaporation of surrounding organic solvent--or possibly by self assembling membrane on nanoporous film support that separates the synthetic cell interiors and exteriors
- Expertise the team brings---Computational studies of membranes, membrane proteins, and other soft materials at multiple levels of detail, molecular design and engineering of ion channels and transporters, electrophysiology, cell modeling, nanofabrication of structures from nanoporous solids, supported membrane technology, engineering design at the nano- and micro-scales.
- Value added by the NDC---focusing the efforts of the multi-institutional, multidisciplinary team on a single goal---development of synthetic cells and organelles for signaling, power generation, absorption and delivery of selected substances, and transport.

A symbolic representation for biomimetic nanoconductors

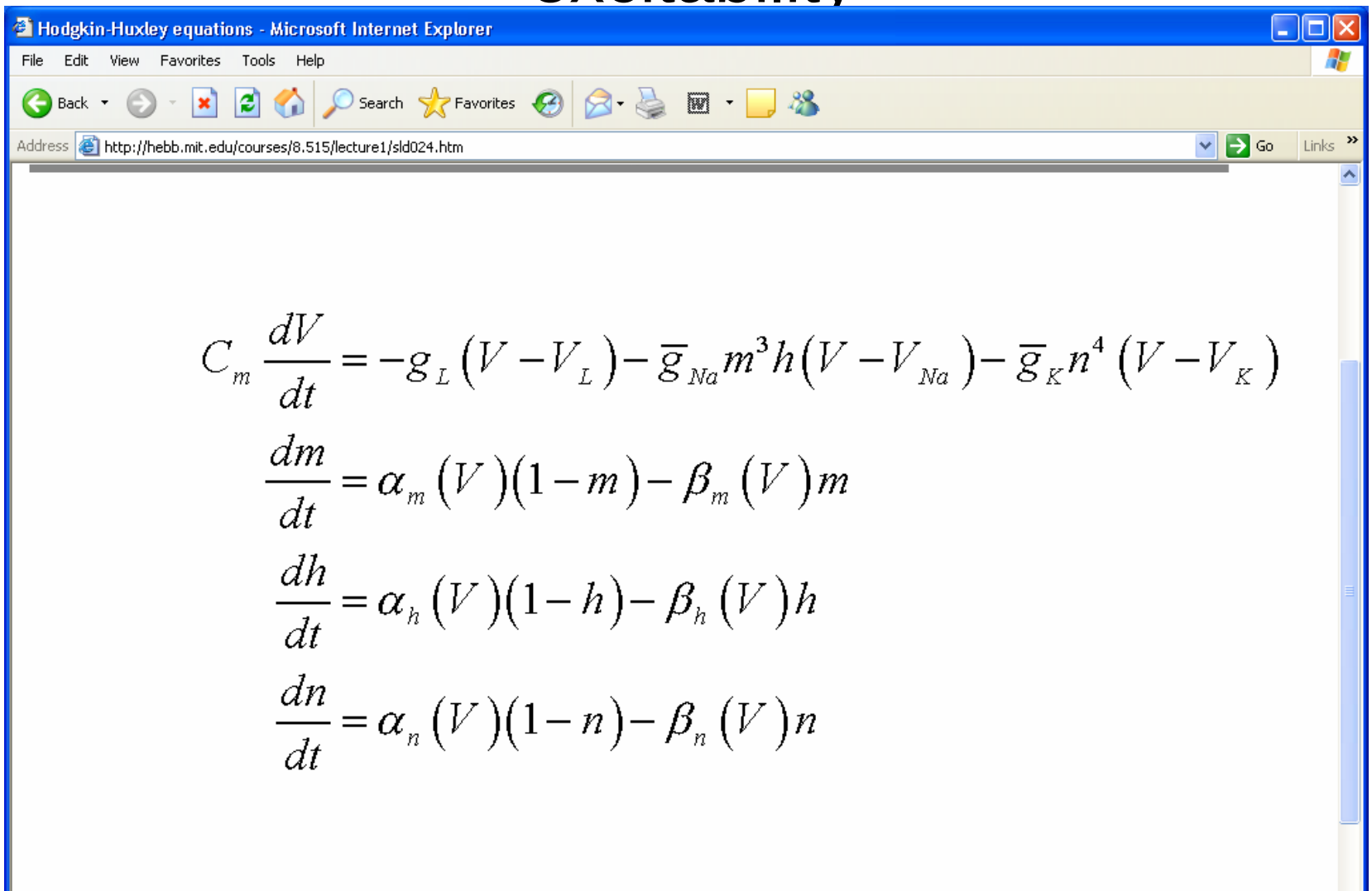
- Blue and orange represent closed and open states of K channels, based on crystal structures from MacKinnon lab, with lumen defined by SAXA program (Sameer Varma).
- White is equivalent circuit deduced by Hodgkin and Huxley before knowing anything about molecular basis of ion transport.
- We now know that emf source is in the selectivity filter
- We know that most of the resistance is in the wide part (to understand that apparent paradox, we did the Brownian dynamics)



Transient Signals - Two Segments



The Hodgkin-Huxley equations for the voltage-dependent variation of specific ionic conductances that underly electrical excitability



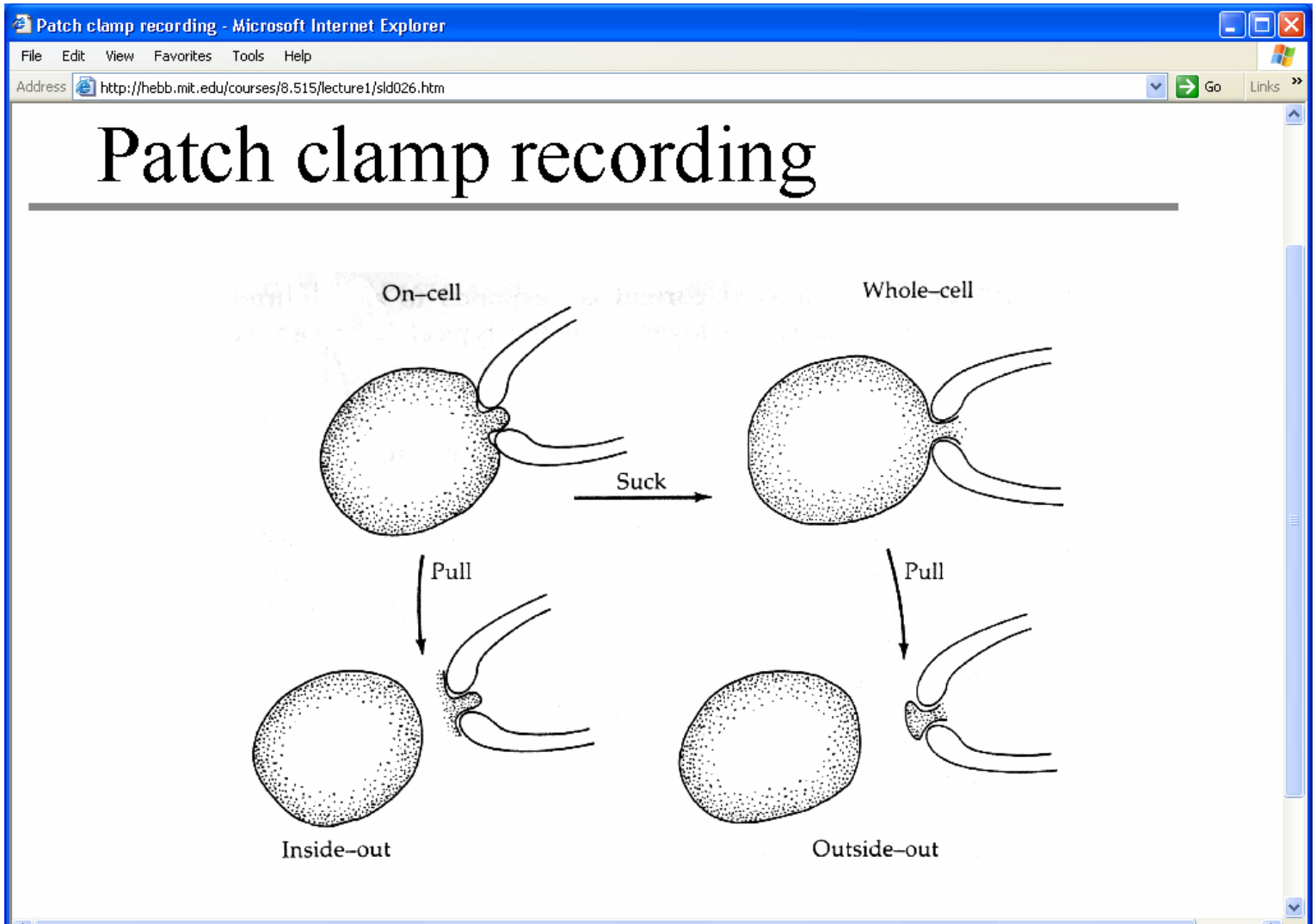
$$C_m \frac{dV}{dt} = -g_L (V - V_L) - \bar{g}_{Na} m^3 h (V - V_{Na}) - \bar{g}_K n^4 (V - V_K)$$

$$\frac{dm}{dt} = \alpha_m (V)(1 - m) - \beta_m (V) m$$

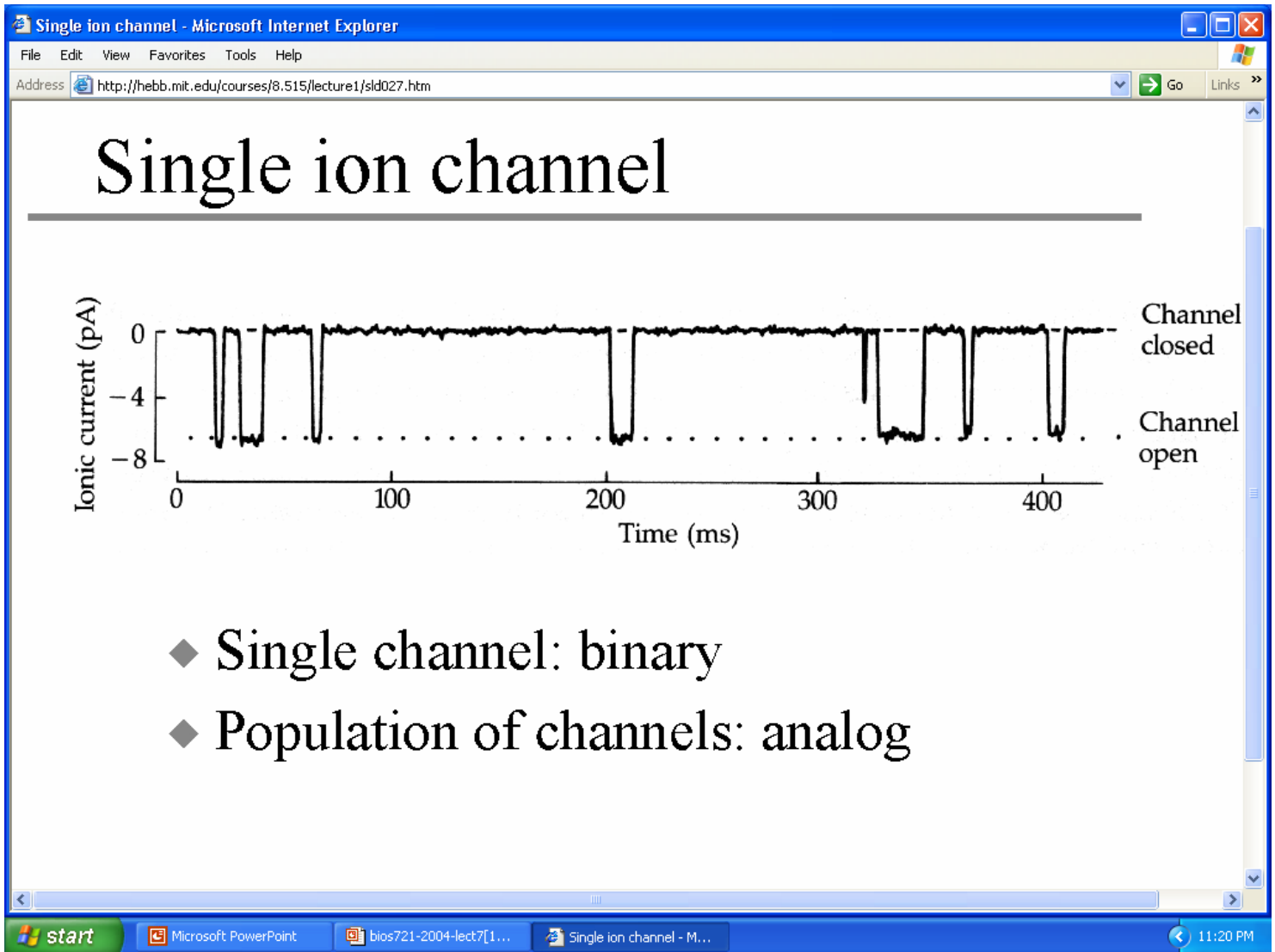
$$\frac{dh}{dt} = \alpha_h (V)(1 - h) - \beta_h (V) h$$

$$\frac{dn}{dt} = \alpha_n (V)(1 - n) - \beta_n (V) n$$

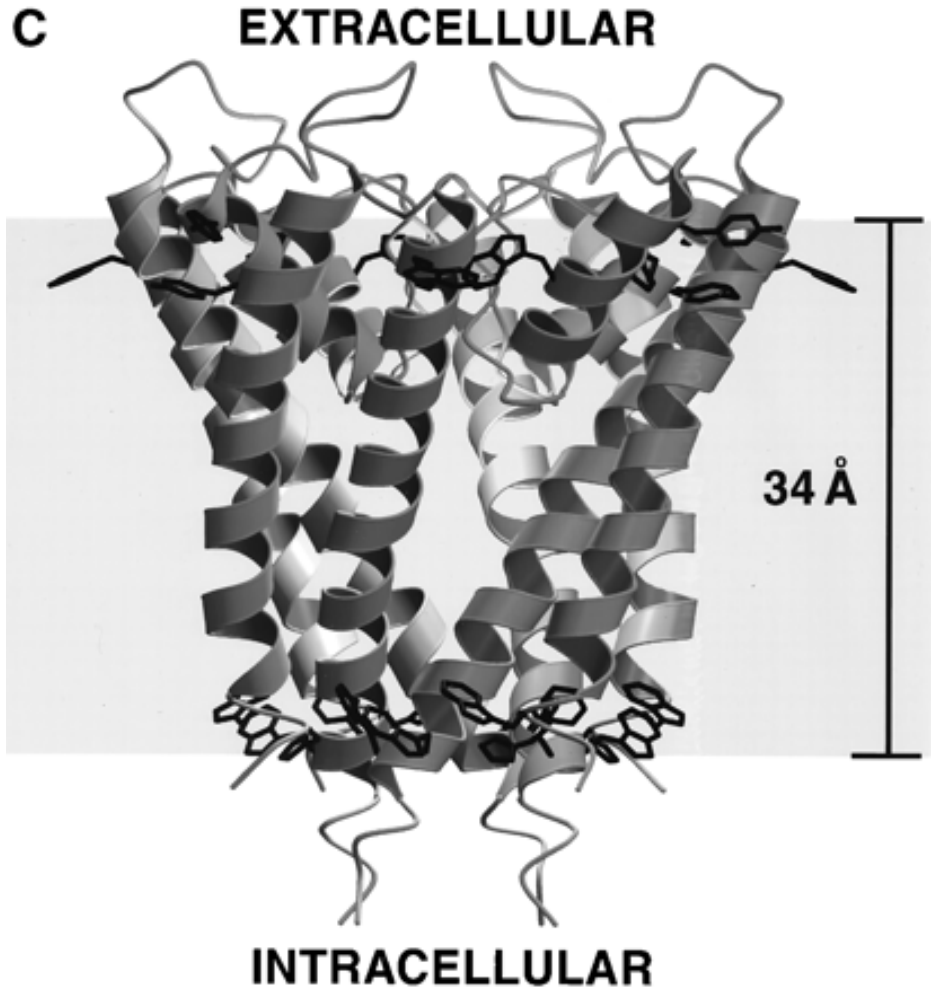
Fast forward to 1976—Neher and Sakmann invent the patch clamp



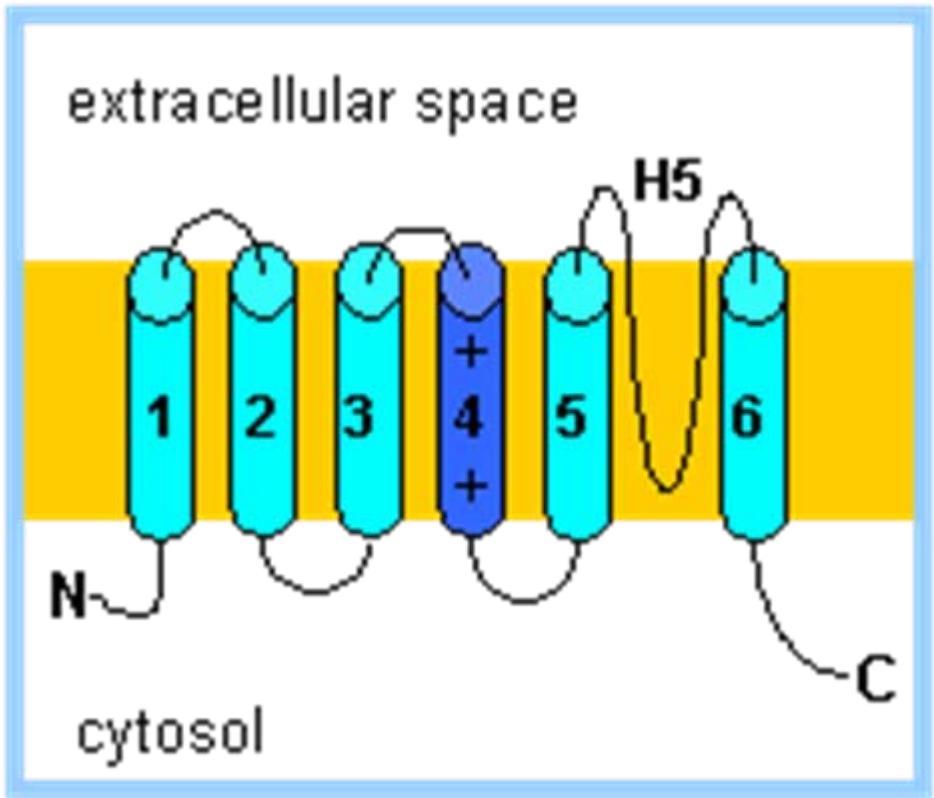
The patch clamp records permit us to see single channels opening and closing—functional properties of a single protein molecule.



Fast forward to 1998—The structure of the permeation pathway of the potassium channel.



We now know there is a common domain composition for the channel protein, comprised of 1) the selectivity filter (H5), the pore (S5 and S6), the voltage sensing domain (S1 through S4) and the inactivation gate (N terminus)---which is a segue into……



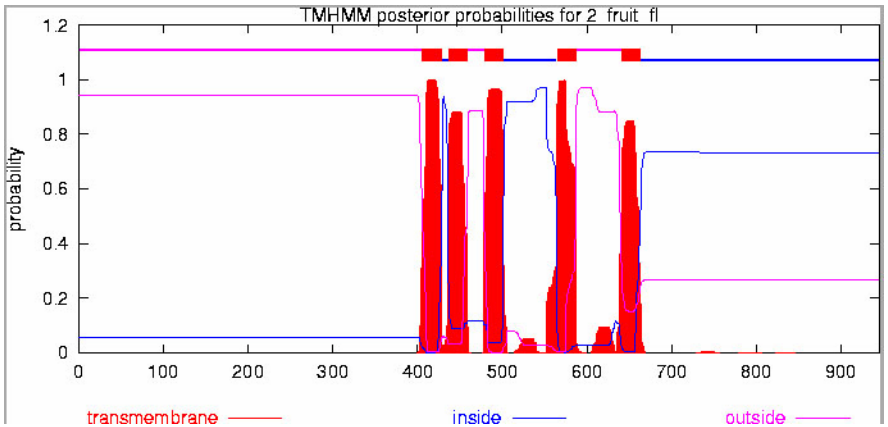
The third big area of computational biology

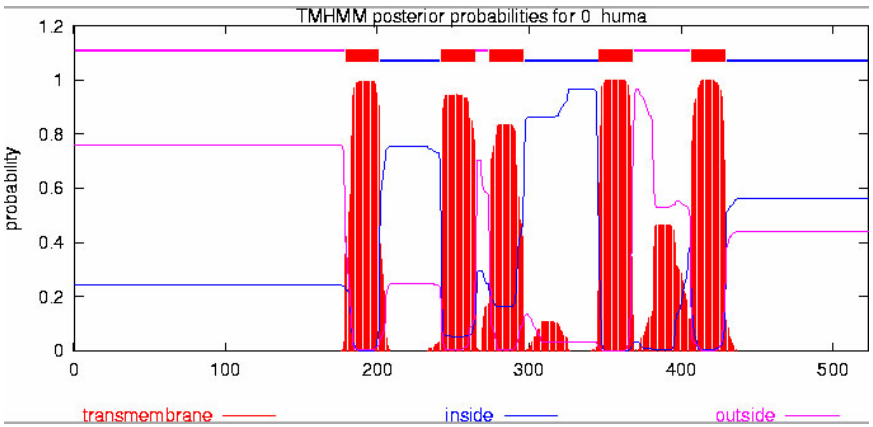
- 3. Central Dogma based computational biology---Biology is based on information flow from the code of the DNA through the RNA to the gene product (proteins). Mathematics and computation here is based on pattern matching, computing is measured in SOPS (string operations per second) rather than FLOPS (floating point operations per second).

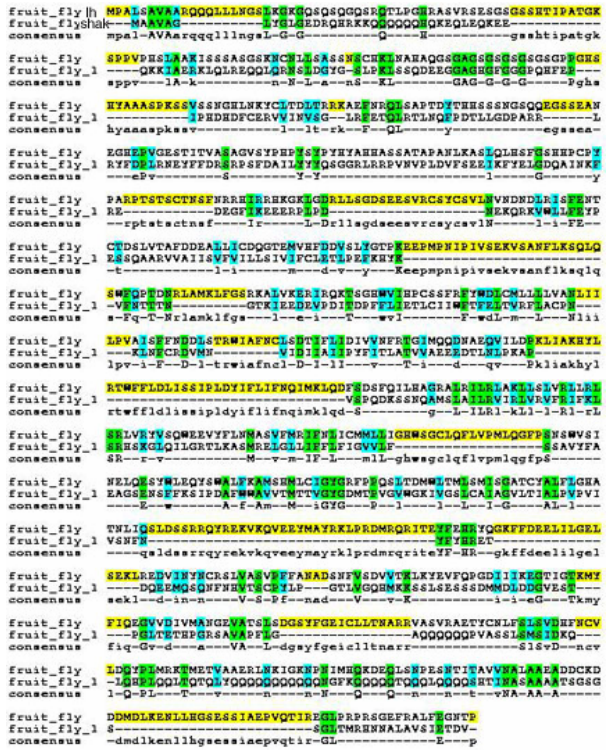
Conservation maps of K channels mapped on to KcsA structure for prokaryotic channels (left) and human channels (right). Dark areas are most highly conserved, light areas least.

Alignments of Shaker (left) and I_h pacemaker (right) K channels from widely varying organisms, illustrating hyperconservation within these subclasses.

Left hand side—Transmembrane Hidden Markov Model Analysis for Topology for Pacemaker (top) and Shaker (bottom) showing preservation of topology in both subfamilies. Right hand side—Alignment of Shaker vs. Pacemaker showing very low homology

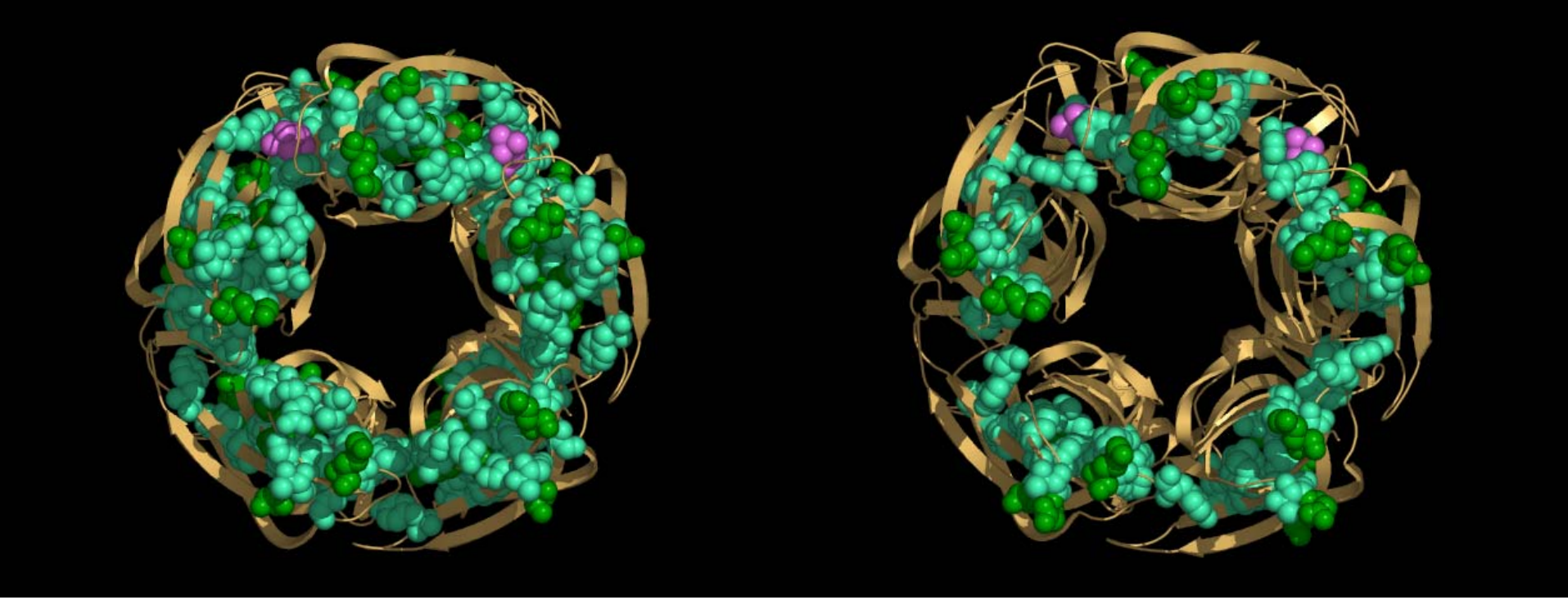






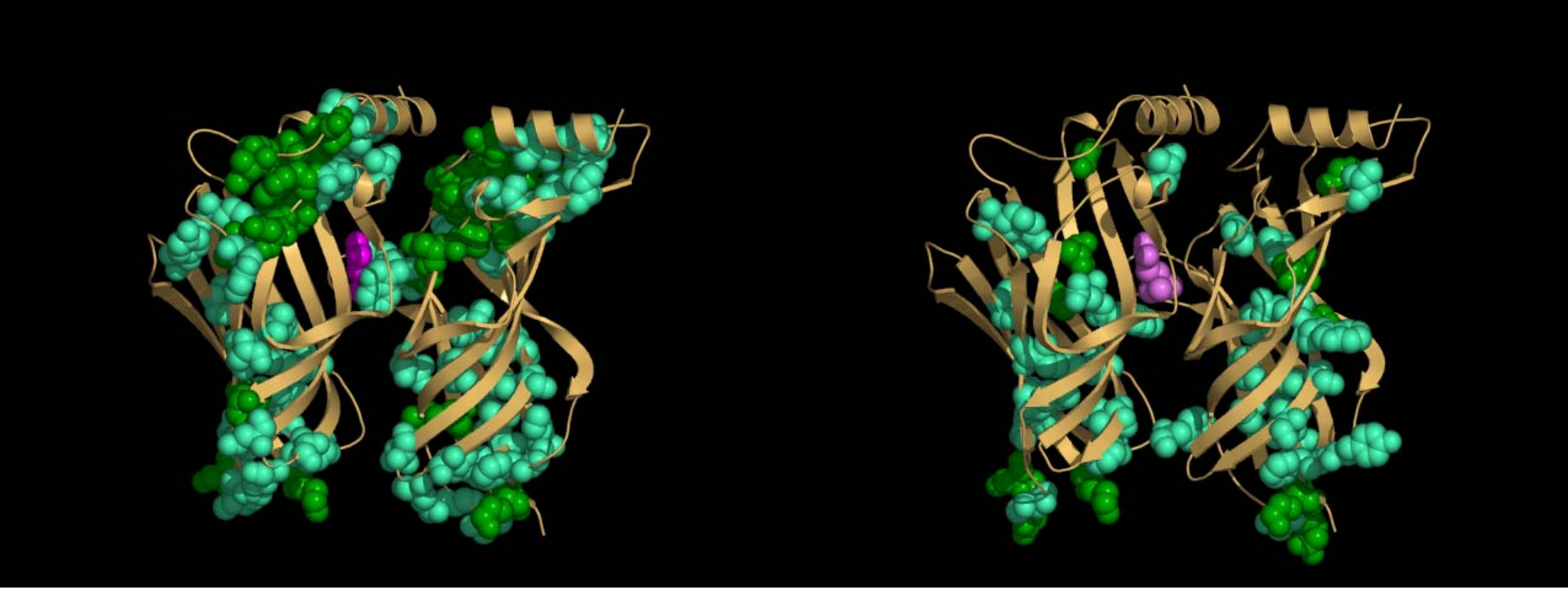
The discovery of ligand-gated channels in prokaryotes.

- This search was suggested by I. Aravind, whose group had earlier discovered prokaryotic homologues of eukaryotic genes that code for enzymes involved in the metabolism of messengers in the neuroendocrine system.
- Ordinary blast searches fail to find such genes.
- In this study, psi-blast using only the distinctive ligand-binding domain of an Ach-receptor channel as the original probe, was employed.
- 15 prokaryotic homologues were revealed. In all cases, in addition to the homologous ligand-binding domains, topology analysis reveals the four transmembrane helices characteristic of the eukaryotic members of the family.



Eukaryotes **Prokaryotes**

Bottom View
80% Consensus
Protein: 1UV6



Eukaryotes **Prokaryotes**

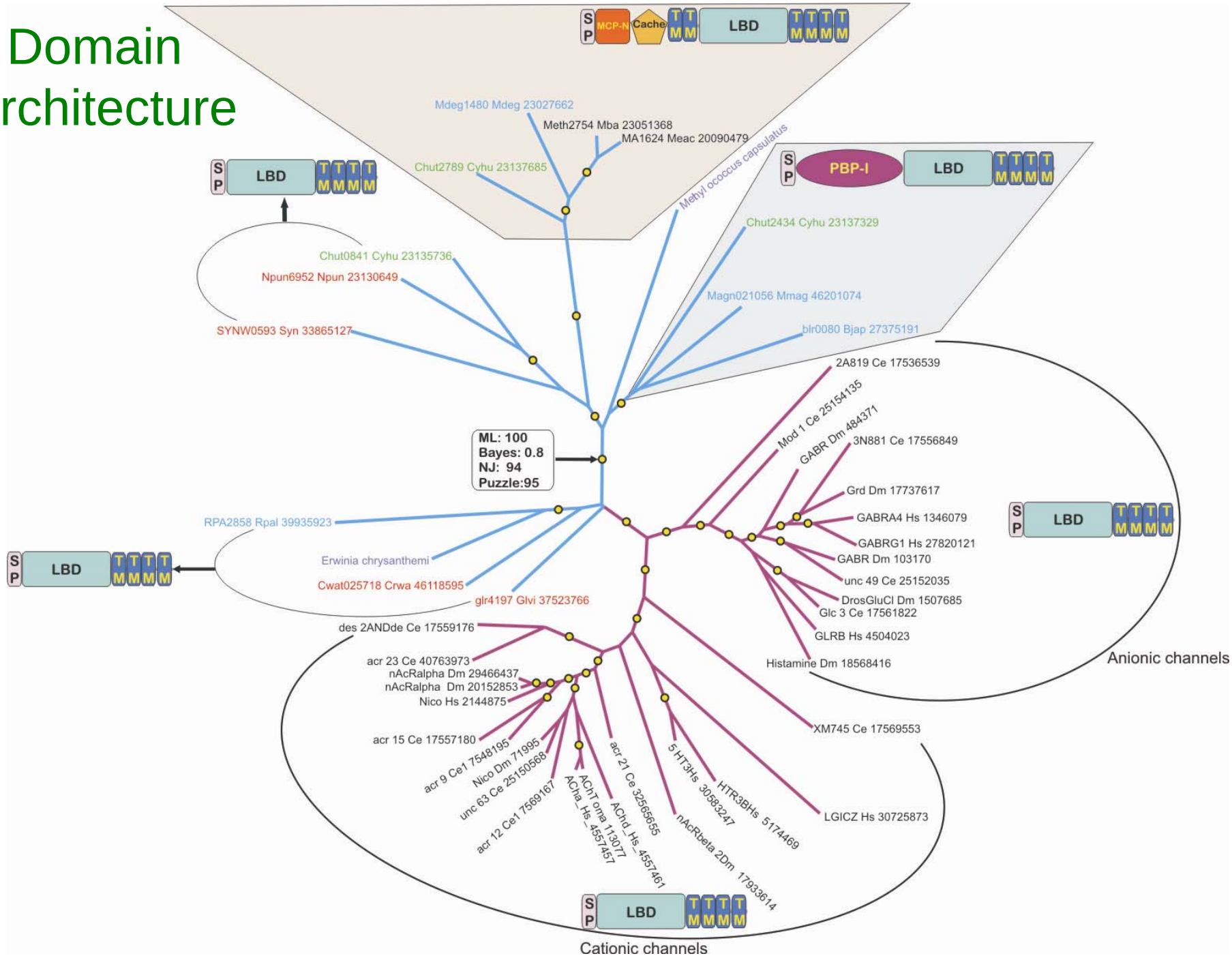
Side View [Chain C and D]
80% Consensus
Protein: 1UV6

Phylogenetic relationship of the bacterial and eukaryotic ART-LGICs

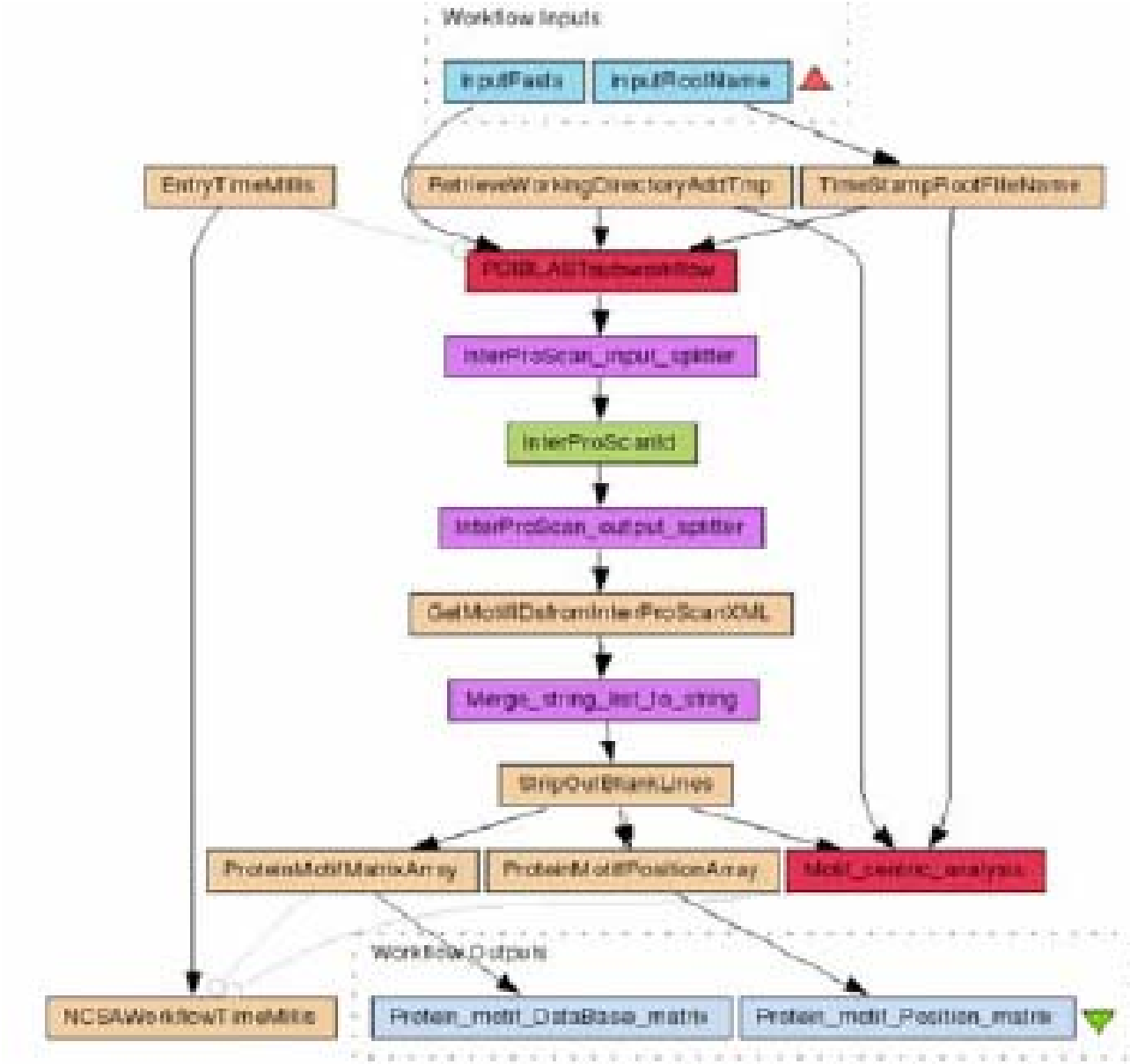
- The tree shows a strongly supported monophyletic animal branch that in turn split up into the two major families corresponding to the “Cationic” and “Anionic” type of receptors

- Some bacterial sequences are closer to the animal ART-LGICs than versions from other bacteria, and vice versa

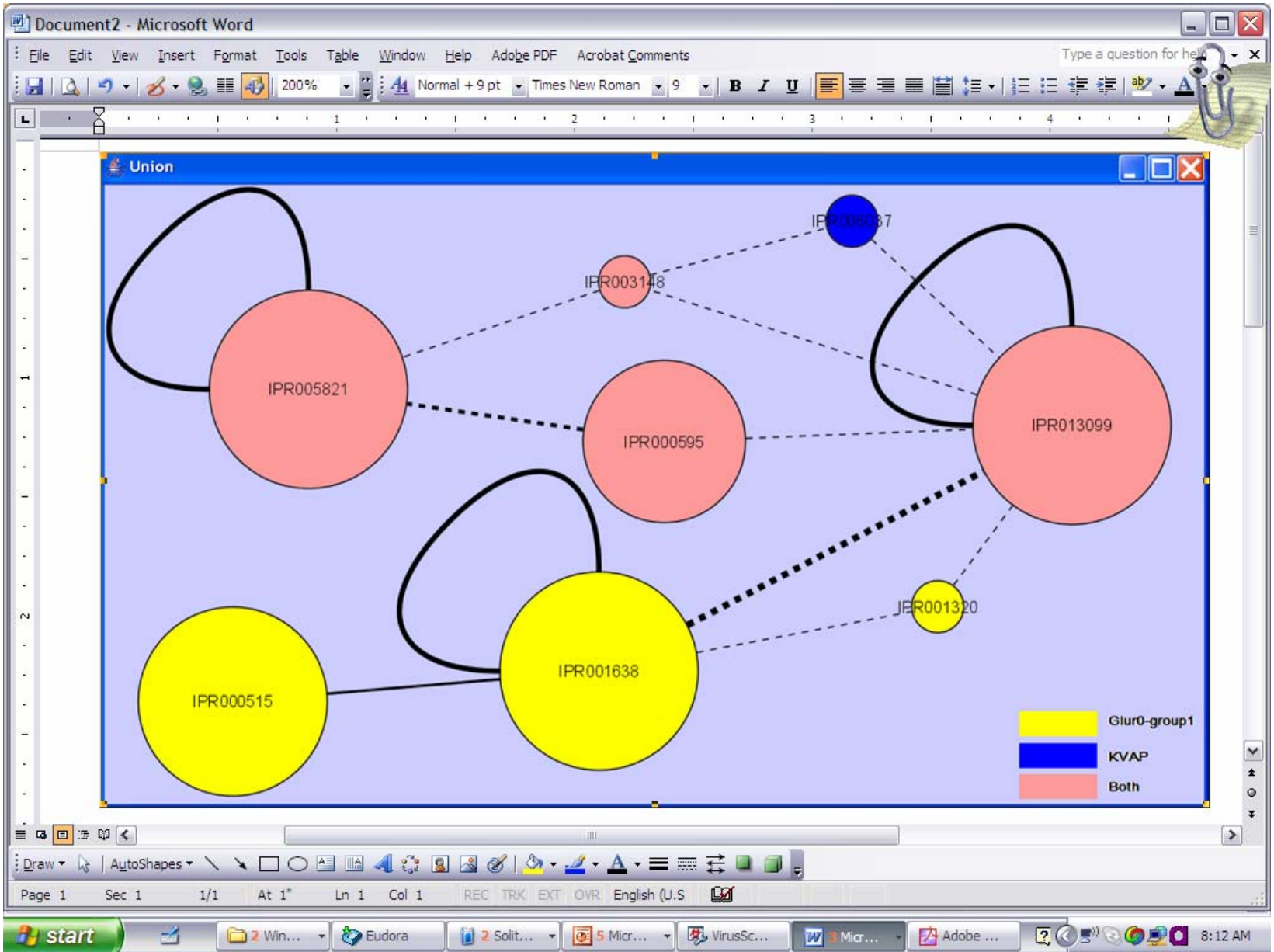
Domain Architecture



Moving from “boutique” to high-throughput domain analysis--- MotifNetwork from NCSA/RENCI



MotifNetwork output shared domains among prokaryotic potassium channels and glutamate receptor channels



Does God really play dice? Reflections on the ubiquity of the Fokker-Planck Equation

In general a Fokker-Planck equation contains a set of terms for deterministic forces and a set of terms for random forces, and describes how a system evolves as the sum of deterministic and random forces.

Undisputable fact: This is a useful representation for complex systems because we can “smear” lots of detail into the random forces.

Semi-scientific question: Does the usefulness of such representations arise because that is the way nature is, or just reflect our inability to know the details? (see letters between Neils Bohr and Albert Einstein)

Other applications of Fokker-Planck

- Stochastic extension of Hodgkin-Huxley equations to give proper threshold behavior of neurons
- Intelligent agent models to understand dynamics of complex systems (for example infectious disease epidemiology)

Properties of Solutions to Fokker-Planck equations

- Local roughness or irregularities. (as in thermal motion of individual particles)
- Larger scale smoothness (as in mass-action laws described by continuous differential equations.)

Roots of Fokker-Planck Equation—Correspondence between Diffusion and Random Walk (the stochastic term in Fokker-Planck)

$$\frac{\partial \phi}{\partial t} = D\nabla^2 \phi(\vec{r}, t)$$

$$D = \frac{\epsilon^2}{6\delta t}$$

Bottom line—all mass-action equations have a corresponding Fokker-Planck representation

Universality of Fokker-Planck: Schrodinger's equation is formally a diffusion equation and therefore a Fokker-Planck equation

$$\frac{\partial \phi}{\partial t} = D\nabla^2 \phi(\vec{r}, t)$$

$$H(t)\left|\psi\left(t\right)\right\rangle = \mathrm{i}\hbar\frac{\partial}{\partial t}\left|\psi\left(t\right)\right\rangle$$

Bottom line: It isn't turtles all the way down, it's Fokker-Planck equations all the way down.

Argument that God really throws dice

- Physicists believe: The most powerful universal forces operating in the Big Bang are reflected in the large scale structure of the universe.
- Cosmologists observe: The universe has irregular structure up through galactic superclusters and filaments—then on larger scale becomes smooth.
- Note: the large scale structure of the universe has the properties of the solution to a Fokker-Planck equation.
- Therefore—I allege---God (or something) threw dice (i.e., injected noise) into the Big Bang

Closing thoughts

For me, and for many scientists, pondering the largest questions produces a feeling something akin to a religious experience, a relationship to God, but…….

….The Devil is in the details (so be sure to do your maths right). Thanks for your attention

And thanks to Collaborators and Lab Members: Narayan Aluru, Iyar Aravind, See-Wing Chiu, Sony Joseph, Jay Mashl, Larry Scott, Ashok Paliniappan, Sagar Pandit, Asba Tasneem, Sameer Varma

Question 1

Question 2