Net2: Bio-algorithms (Last week)

- Biology to aid Computing … to aid biology
- Molecular nano-computing (DNA)
- Self-assembly (nano-I/O)
- Intra cellular network computing (oscillators)
- Genetic algorithms (multi-align)
- Neural nets (exons)

Net3: Global integration

- Multi-cellular models -- e.g. sensory integration
- Systems biology, simulation & integration
- Organ systems
- Multi-organism - Ecological modeling
  - predator/prey - host/parasite - HIV
- Global & socioeconomic considerations
- Education
  - Model evaluation & sharing

Faster than exponential

The human neural net

“...The retina's 10 million detections per second [0.02 g] ... a risky extrapolation ... 100 trillion instructions per second to emulate the 1,500 gram human brain. ... Computer power for a given price roughly doubled each year in the 1990s, ... thirty more years at the present pace would close the millionfold gap.” (Moravec99)

Basic olfactory tasks

1) Odor memory and recognition.
2) Background elimination (one known + unknown thoroughly mixed)
3) Component separation, (few known odors mixed)
4) Odor separation (turbulent unknowns)

\[
\text{cov}_i = \text{cov}_{\text{min}} \left( \frac{c_i/c_{\text{thresh}}}{1 + f_b} \right), \quad \text{and} \quad \text{cov}_{\text{min}} = \text{cov}_{\text{min}}(c_i/c_{\text{thresh}}) (1 + f_b)
\]

\[
\text{ci} = \left( \frac{c_{\text{thresh}}}{f_b} \right) \text{cov}_{\text{min}}
\]

\(i=\) target, \(b=\) background, \(u=\) unknown, receptors \(i=1\) to \(1000\).

minimum covrage for concentration threshold.

\(f_b = \) fraction bound random 1 to \(10^4\) (log uniform pdf)

Hopfield 1999: PNAS 96:12506-11
Odor space and olfactory processing: Collective algorithms & neural implementation

80 adapting neurons, two sniffs: 100-500 msec has a mixed odor 50% + 100% y. At 500 msec 75% + 1100% y. The sniff at 100 milliseconds strongly activates more than half the neurons, after which they adapt. The changed sniff at 500 milliseconds is almost invisible. b) as in a), but the y-axis firing rate at the time of each action potential. The second sniff is now clearly visible, and most spikes appear to belong to one of three patterns. A 20% spread in D was included to produce parameter-spread noise.

Hopfield 1999; PNAS 96:12506-11 & MATLAB

Olfaction code

% figshare.m is a matlab script for one odor
% its two graphics panels are like Fig 1b and 1e in the paper
% by adjusting strength (below) other parts of Fig 1 can be recovered
% span of binding constants 10^6
% N odor receptor types
clear
N=2000
strength=.1 % for saturation, set strength=1
% strength=.003 is 3 times threshold
clf;
logtarget = 6*rand(N,1)-6; % the target at saturating concentration
target = exp(2.3*logtarget); % the signal due to the target
% when present at saturating strength
logbackground = 6*rand(N,1)-6; % the logbackground at saturating conc.
background = exp(2.3*logbackground); % actual signal from saturating bkgnd.
y=log10(target*strength)+0.1*randn(2000,1); % log10 of signal due to target at (strength) + plus noise

Systems biology model sharing

<table>
<thead>
<tr>
<th>Simulators</th>
<th>Director</th>
<th>Inst.</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERATO</td>
<td>John Doyle</td>
<td>Caltech</td>
<td>SysBioWorkbench &amp; SBML</td>
</tr>
<tr>
<td>Gepasi</td>
<td>Pedro Mendes</td>
<td>Santa Fe</td>
<td>MCA, systems kinetics</td>
</tr>
<tr>
<td>JarnacScamp</td>
<td>Herbert Sauro</td>
<td>Caltech</td>
<td>MCA, Stochastic</td>
</tr>
<tr>
<td>StockSim</td>
<td>Dennis Bray</td>
<td>U.Camb.</td>
<td>Stochastic</td>
</tr>
<tr>
<td>BioSpice</td>
<td>Adam Arkin</td>
<td>LBL</td>
<td>Enzyme/receptor-ligand</td>
</tr>
<tr>
<td>DBSolve</td>
<td>Igor Gorynin</td>
<td>Glaxo</td>
<td></td>
</tr>
<tr>
<td>E-Cell</td>
<td>Masaru Tomita</td>
<td>Keio</td>
<td>Metab. Net ODE</td>
</tr>
<tr>
<td>Yeast</td>
<td>Jim Schaff</td>
<td>U.CT</td>
<td>Geometry</td>
</tr>
<tr>
<td>Net3</td>
<td>J. Bassingthwaighte</td>
<td>Seattle</td>
<td>Enzymes to body physiology</td>
</tr>
<tr>
<td>CellML</td>
<td>Peter Hunter</td>
<td>U.Auckland</td>
<td>Geometry, model sharing</td>
</tr>
<tr>
<td>GENESIS</td>
<td>James Bower</td>
<td>Caltech</td>
<td>Neural networks</td>
</tr>
<tr>
<td>STRePS</td>
<td>Luef Gatewood</td>
<td>U.MN</td>
<td>Stochastic micropopulation disease spread</td>
</tr>
</tbody>
</table>

P=java, w = windows, u=unix, x=XML, + = source/community input

Net3: Global integration

• Multi-cellular models -- e.g. sensory integration
• Systems biology, simulation & integration
• Organ systems
• Multi-organism - Ecological modeling
  – predator/prey - host/parasite - HIV
• Global & socioeconomic considerations
• Education
  – Model evaluation & sharing

Design and strategy for the Cardiome Project

The Physiome Project has the goal of providing the quantitative description of the integrated functions of the living organism ... a central scheme, a description of the spread of excitation and contraction through an anatomically detailed cardiac model with fiber directions. This will be augmented by the additions of regional blood flows, substrate uptake and metabolism, and energy production and utilization in serving contraction and ionic balances. Later stages will involve cellular regulation and responses to interventions. The organization of such projects is by the assembling of components whose linkages one to another are first minimized and then augmented to improve the approximation to reality.

Modeling SNP to 3D to cell morphology & pathology

**Red Blood Cell Function**

- Transport O$_2$ from lungs to tissues – using hemoglobin to carry the O$_2$
- Hemoglobin is maintained in its functional state (reduced) by the metabolic machinery
- Cell membrane separates the internal environment from the external environment
  - Subject to physicochemical constraints
    - Electro neutrality, Osmotic balance
  - Cause of the imbalance:
    - Impermeable polyions inside the cell
      - Hemoglobin, organic phosphates

**3D-structure to function to morphology to function**

- Normal RBC 6-8 µm; 4-5 x 10$^12$ per L
- Sickle cell; HbS
- Echinocytes (crenated); hyperosmotic medium
- Acanthocytes; abetalipoproteinemia
- Macrocytes 9-12 µm; megaloblastic anemia
- Spherocytes; enzyme deficiencies

**From SNPs to pathogen resistance mechanisms**

"Model...erythrocytes of [glutathione peroxidase] GPX1*2 heterozygotes should be more efficient in sheltering the cell membrane from irreversible oxidation and binding of hemoglobin caused by the oxidant stress exerted by Plasmodium falciparum... we observed a clear trend toward a dissociation between the HBB*A-*S and GPX1*2-*1 genotypes in the overall data."  

**From SNPs to toxicology mechanisms**

"Drug induced [e.g. primaquine] oxidative hemolysis... with certain enzymopathies, notably glucose 6 phosphate dehydrogenase deficiency... Others... disrupt mitochondrial function and... heme biosynthesis... including... alcohol & chloramphenicol"  

**A possible molecular mechanism governing human erythrocyte shape.**

Band3 = 10% RBC membrane

HCO$_3^-$, H$^+$, ATP
H$_2$O + Hb-CO$_2$


**Red cell distortion and conceptual basis of diffusing capacity estimates: finite element analysis**

Compute the uptake of CO across a two-dimensional geometric capillary model containing a variable number of equally spaced RBCs (circular or parachute shaped, with the same perimeter length). Total CO diffusing capacity (DLCO) and membrane diffusing capacity (DMCO) were calculated by a finite element method. DLCO calculated at two levels of alveolar PO$_2$ were used to estimate DMCO by the Roughton-Forster (RF) technique or by the morphometric random linear intercept method. Results: shape distortion of RBCs reduces diffusive gas uptake, & exaggerates errors in the RF technique at a high capillary hematocrit. Shape distortion error in morphometric estimates of DMCO are exaggerated at a low hematocrit.

A possible molecular mechanism governing human erythrocyte shape.  

Action Potential Model for Canine Ventricular Cell


Whole body recirculation model (NSR) with four chambered heart, seven organs, four arteries and four veins

Multi-Organ System Failure

Is a, life-threatening complication of otherwise mild sickle cell disease. (as well as of injury in non-sickle cell individuals). The system model is largely unknown. Will genomic data help prediction and/or prevention?

Net3: Global integration

- Multi-cellular models -- e.g. sensory integration
- Systems biology, simulation & integration
- Organ systems
- Multi-organism - Ecological modeling
  - predator/prey - host/parasite - HIV
- Global & socioeconomic considerations
- Education
  - Model evaluation & sharing

The Sims:

Sim City’87
Sim Earth’88
Sim Ant’91
Sim Life’92
Sim Farm’93
Sim Isle’95
...

Integration of multiple inputs: Think globally; act locally:
Think globally; act locally

Lithosphere (0.2% C, 75% SiO₂) 110°C at 4 km
Diameter = 1.3e6 m = 5e22 g    (5000 species / g soil)
Biosphere: 3e15 g (dry wt. marine); 2e18 g (land)
Microbial hydrosphere 1.4e21 ml = 1e27 cells = 4e33 bp
Anthroposphere (23% C) = 6e23 cells = 4e32 bp.

Biodiversity for nanoengineering

Protein Eng 2000 Feb;13(2):121-8 Protein engineering of cytochrome b450(cam) (CYP101) for the oxidation of polycyclic aromatic hydrocarbons.


Modeling bio-effects on global warming

The equatorial Pacific, sub arctic pacific and Southern Ocean are high-nutrient low-chlorophyll (HNLC) areas which may support higher plant biomass if micro-nutrients such as Fe were added... No ocean fertilization study has been long lived enough to follow the effects of iron fertilization through the food web, and hence determine the potential for long term carbon sequestration.


Crossing the Hopf bifurcation in a live predator-prey system.

\[ \frac{dN}{dt} = \delta (N_1 - N) - F_C(N)C \]
\[ \frac{dC}{dt} = F_C(N)C - F_B(C)B(\epsilon - \delta) - RC \]
\[ \frac{dB}{dt} = F_B(C)R - (\delta + m + \lambda)B \]

\( C = \) concentration of Chlorella
\( R = \) concentration of reproducing Brachionus
\( B = \) concentration of total Brachionus
\( m = \) demographic mortality of Brachionus
\( \lambda = \) decay of fecundity of Brachionus
\( \beta = \) maximum birth rates
\( K = \) half-saturation constants

Fussmann et al. Science 2000; 290:1358-60. (Pub)

Predator -- prey

Human microbial ecology: Sorting out cause, effect & prescription (vaccine paradox)

- Helicobacter pylori
- Cytomegalovirus
- Porphyromonas gingivalis
- Chlamydia pneumoniae
- Human papilloma virus
- Hepatitis B, C virus
- EpsteinBarr Virus

Stomach cancer
Stomach ulcers
Coronary artery disease
Alzheimer’s
Cervical cancer
Liver cancer
Breast cancer
Nasopharyngeal carcinoma


www.ncbi
New nucleotide sequences processed in GenBank per month (above)

Today's total is:

HIV-1 resequencing

HIV/AIDS models

D. Kirschner & G. Webb

Resistance, Remission, & Qualitative Differences in HIV Chemotherapy

Therapy T(0)/Vs(0) in mm-3

306, 21 (5.8 years), 217, 31 (7.7 years), 100, 69 (8.4 years), 43, 156 (8.6 years).

The rates of exponential increase in Fig. 2a (.03, .02, .01, .005) are inversely correlated to starting CD4+ T-cell count; decay rates in Fig. 2b (all -.2) are not correlated to starting viral levels. The lack of correlation of viral decay rates is an indication of slower clearance of wild-type virus in the external lymphoid compartment. The time to the downward spike in Figure 2b is correlated to starting viral levels (1). The treatment parameters (c1, c2, c3, q) are the same in all four simulations.

Vaccines for the 21st Century

http://books.nap.edu/html/vacc21/

Level I: Most favorable: saves money & Quality-Adjusted Life Years (QALY)

Level II: <$10,000 < Level III: <$100K per QALY saved < Level IV

Level I candidate vaccines:

- Viral: CMV vaccine for 12 year olds, Flu vaccine for 20% of the US per year.
- Therapeutic vaccines: IDDM diabetes, MS, Rheumatoid arthritis
- Bacterial: Streptococcus B & pneumoniae vaccine for infants & 65 year olds.

“A quantitative model that could be used by decision makers to prioritize the development of vaccines against a number of disparate diseases” 1985 & 1999.

Role of genomics & computational biology in vaccine R&D?

DNA vaccines, Intracellular vaccines, RNAi, multiplexed...


Malaria & Mosquito genomes

Net3: Global integration

- Multi-cellular models -- e.g. sensory integration
- Systems biology, simulation & integration
- Organ systems
- Multi-organism - Ecological modeling
  - predator/prey - host/parasite - HIV
- Global & socioeconomic considerations
- Education
  - Model evaluation & sharing
Human Genome Project Ethical, Legal & Social Issues (ELSI)

- Fairness - Genetic non-discrimination
- Privacy
- Reproductive rights - cloning
- Psychological stigmatization
- Clinical quality-control
- Safety and environmental issues - GMO & biowarfare
- Uncertainties - testing minors
- Conceptual & philosophical implications - diversity?
- Commercialization of products - Who owns?

--Clinton, June 26, 2000--

The Genetic Nondiscrimination in Health Insurance and Employment Act of 1999, introduced by Senator Daschle and Congresswoman Slaughter, that will extend these employment protections to the private sector and finish the job of helping to extend protections to individuals purchasing health insurance, begun with the Health Insurance Portability and Accountability Act.

ELSI: Do Races Differ? Not Really, DNA Shows

Hb variants "evolved to help the ancestors of these groups resist malaria infection, but both prove lethal when inherited in a double dose. As with differences in skin pigmentation, the pressure of the environment to develop a group-wide trait was powerful, and the means to do so simple and straightforward, through the alteration of a single gene.

A founder effect explains the high incidence of Huntington's neurodegenerative disease in the Lake Maracaibo region of Venezuela, and of Tay-Sachs disease among Ashkenazi Jews.

But Dr. Naggert emphasized that medical geneticists had a much better chance of unearthing these founder effects by scrutinizing small, isolated and well-defined populations, like the northern Finns, the Basques of Spain, or the Amish of Pennsylvania, than they did by going after "races."

Dangers of model-free science

Lysenko: inheritance of acquired characteristics.
"His habit was to report only successes. His results were based on extremely small samples, inaccurate records, and the almost total absence of control groups. An early mistake in calculation, which caused comment among other specialists, made him extremely negative toward the use of mathematics in science."

Dangers of ethics-free science

The 1979 release of Anthrax-836 spores in Sverdlovsk.
"In 1953 a leak...In 1956, Sizov found that one of the rodents captured in the Kirov sewers had developed a new strain more virulent than the original. The army immediately ordered him to cultivate the new strain...to install in the SS-18s targeted on western cities."

Alibek & Handelman "Biohazard" 1999 (Davis)

How can we improve our genome engineering tools preferentially toward defense and away from terrorism?

Genetically modified organisms (GMO)
Genetically modified organisms

Developing world needs:
- Agri-vaccines, salt & drought tolerance

Terminators: Allergen dispersal vs reseeding

“Organic”: no inorganic fertilizers means high animal load.

Many natural pesticides are carcinogens including estragole (basil), safrole (natural root beer), symphytine (comfrey tea), hydrazine (mushrooms) & allyl isothiocyanate (brown mustard); psoralen (celery) & aflatoxin (nuts & cereals).

Cloning & stem cells

Why do clones exhibit developmental defects?
- Study epigenetic reprogramming with expression profiles?

Can we increase fraction of stem cells without going through cloning?
- “Radial glial cells that lacked a functional form of a transcription factor called Pax6 could not generate neurons. But when Pax6 was introduced into glial precursor cells, these cells started to produce neurons.” [ref]

Net3: Global integration

- Multi-cellular models -- e.g. sensory integration
- Systems biology, simulation & integration
- Organ systems
- Multi-organism - Ecological modeling
  - predator/prey - host/parasite - HIV
- Global & socioeconomic considerations
- Education
  - Model evaluation & sharing

Models for education & decision-making

Improve our ability to deal with:
- Uncertainty
- Complexity
- Quantitation
- Exceptions (collect and cherish)
- Comparisons of diverse entities
- Translation & integration
- Continuity over time

Measures of quality of structural & functional genomic data

<table>
<thead>
<tr>
<th>Automate</th>
<th>Data quality</th>
<th>Model quality</th>
<th>Similarity search</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray diffraction</td>
<td>resolution &lt; 0.2nm</td>
<td>Σo-c/Σo &lt; 0.2</td>
<td>DALI</td>
</tr>
<tr>
<td>Sequence</td>
<td>bp &lt;0.01%</td>
<td>conserved proteins</td>
<td>BLAST</td>
</tr>
<tr>
<td>Function</td>
<td>cc, t-test</td>
<td>AlignACE</td>
<td>Correlation Map &amp; specificity</td>
</tr>
</tbody>
</table>

Biophilia & Consilience

Biophilia -- the connections that human beings subconsciously make with other living beings. (Cute animals, snake dreams, therapeutic greenery & natural sounds …)

Consilience - Long-separated fields come together and create new insights; e.g. chemistry & genetics created the powerful new science of molecular biology. Is all human endeavor, from religious feeling to financial markets to fine arts, ripe for explaining by hard science?

How might genomics & computational biology contribute?

Kellert & Wilson 1993 *The Biophilia Hypothesis*,
E. O. Wilson 1999 *Consilience: The Unity of Knowledge*
PET & MRI

Positron emission tomography

555MBq of $^{15}$O butanol, scan for 60s; effective image resolution of 9mm (FWHM).

Significant activations for the contrast religious-recite vs. rest in religious subjects, rendered onto canonical T1-weighted image of SPM97d ($P < 0.001$, uncorrected for multiple comparisons) for task comparisons, an ancova (analysis of covariance) model was fitted to the data for each voxel.


MRI & gene expression

Suitable for intact, opaque organisms in 3D at cellular resolution (10 µ)


Multi-thanks to

1999:
Jeremy Edwards, Tim Chen, Tao Wei,
Bob Freeman, Dan Janse

2000: Jason Comander, Adnan Derti, Bob Freeman, Yonatan Grad, Haley Hieronymus, Dan Janse, Peter Kharchenko, Douglas Selinger, Allegra Petti, Nikos Reppas

Teaching Fellows 2001

Douglas Selinger
Bob Freeman
Hui Ge
Haley Hieronymus
Suzanne Komili
Sharyl Wong
Matt Wright
Zhou Zhu

Teaching Fellows 2002

Suzanne Komili
Woodie Zhao
Joon Lee
Lan Zhang
Tom Patterson
Jon Radoff
Yonghong Xiao
Gary Gao
Laksman Iyer

Teaching Fellows 2003

Lan Zhang
Woodie Zhao
Griffin Weber
Tom Patterson
J Singh
Xiaoxia Lin
Mike Jones
ChihLong Liu
Faisal Reza
101 after '03

1) We will need Teaching Fellows if there is to be a course in Fall 2004. Please contact us.

2) A small number of projects based on need, merit, and interest may be selected for additional support, resources, and/or mentors.

Binary code

Over 5,300 yr BP, China's first emperor, Fu Xi, 64 hexagrams of the I Ching.
See Walter & Smith