Integrate 1: Minimal & exponential Systems (last week)

Life & computers: Self-assembly
Math: be aware of assumptions & approximations
Catalysis & Replication
Differential equations: $\frac{dy}{dt} = ky(1-y)$
Mutation & the single molecule: Noise is overcome
Directed graphs & pedigrees
Bell curve statistics: Binomial, Poisson, Normal
Selection & optimality

Integrate 1: Minimal & exponential Systems (last week)

Big questions: How can we define (& design) “General biology” (self-replicating systems)?
How soon might the exponential of IT and/or Biotech overtake human intelligence?
DNA single molecule stochastics is a given.
How does life reduce the noise by dozens of logs?

Integrate 2: Optimal BioSystems

• Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection

From atoms to (bio)molecules

<table>
<thead>
<tr>
<th>Elements</th>
<th>Gas</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂O</td>
<td>H₂, O₂</td>
</tr>
<tr>
<td>CH₄</td>
<td>C₆₀</td>
</tr>
<tr>
<td>NH₃</td>
<td>N₂</td>
</tr>
<tr>
<td>H₂S</td>
<td>Sⁿ</td>
</tr>
<tr>
<td>PH₃</td>
<td>K⁺PO₄⁻Na⁺</td>
</tr>
<tr>
<td></td>
<td>Elemental</td>
</tr>
<tr>
<td></td>
<td>Salt</td>
</tr>
</tbody>
</table>

Purify

Elements, molecules, assemblies, organelles, cells, organisms
chromatography

Purified history

Pre 1970s: Column/gel purification revolution
Mid-1970s: Recombinant DNA brings clonal (single-step) purity.
1984-2002: Sequencing genomes & automation aids return to whole systems.
Integrate 2: Optimal BioSystems

Elements & Purification
• Systems Biology & Applications of Models
• Life Components & Interconnections
• Continuity of Life & Central Dogma
• Qualitative Models & Evidence
• Functional Genomics & Quantitative models
• Mutations & Selection

Why Genomes & Systems?

#0. Why sequence the genome(s)? To allow #1,2,3 below.
#1. Why map variation?
#2. Why obtain a complete set of human RNAs, proteins & regulatory elements?
#3. Why understand comparative genomics and how genomes evolved? To allow #4 below.
#4. Why quantitative biosystem models of molecular interactions with multiple levels (atoms to cells to organisms & populations)?
To share information. Construction is a test of understanding & to make useful products.

Number of component types (guesses)

<table>
<thead>
<tr>
<th></th>
<th>M.gen</th>
<th>Worm</th>
<th>Human</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bases</td>
<td>.58M</td>
<td>&gt;97M</td>
<td>3000M</td>
</tr>
<tr>
<td>DNAs</td>
<td>1</td>
<td>7</td>
<td>25</td>
</tr>
<tr>
<td>Genes</td>
<td>.48k</td>
<td>19k</td>
<td>21k</td>
</tr>
<tr>
<td>RNAs</td>
<td>.4k</td>
<td>&gt;30k</td>
<td>.2-3M</td>
</tr>
<tr>
<td>Proteins</td>
<td>.6k</td>
<td>&gt;50k</td>
<td>.3-10M</td>
</tr>
<tr>
<td>Cells</td>
<td>1</td>
<td>959</td>
<td>1014</td>
</tr>
</tbody>
</table>

http://www.nature.com/cgi-taf/DynaPage.taf?file=/nature/journal/v409/n6822/full/409860a0_fs.html

Integrate 2: Optimal BioSystems

Elements & Purification
• Systems Biology & Applications of Models
• Life Components & Interconnections
• Continuity of Life & Central Dogma
• Qualitative Models & Evidence
• Functional Genomics & Quantitative models
• Mutations & Selection

Grand (& useful) Challenges

A) From atoms to evolving minigenome-cells.
• Improve in vitro macromolecular synthesis.
• Conceptually link atomic (mutational) changes to population evolution (via molecular & systems modeling).

B) From cells to tissues.
• Model combinations of external signals & genome-programming on expression.
• Manipulate stem-cell fate & stability.
• Engineer reduction of mutation & cancerous proliferation.
• Programmed cells to replace or augment (low toxicity) drugs.

C) From tissues to physio- & eco- systems
• Programming of cell and tissue morphology.
• Quantitate robustness & evolvability.
• Engineer sensor-effector feedback networks where macro-morphologies determine the functions; past (Darwinian) or future (prosthetic).

http://www.wolframscience.com/reference/notes/1003g
From monomers to polymers

Complementary surfaces
Watson-Crick base pair
(Nature April 25, 1953)

Nucleotides
dATP
rATP

The simplest amino acid component of proteins

Glycine
Gly
G

config[glycine,]
substituent(aminoacid_L_backbone),
substituent(hyd),
linkage(from(aminoacid_L_backbone,car(1)),
to(hyd,hyd(1)),
nl,single)].

Kloxo

Smiles String: [CH2][([NH3+])][C]=-[O]]-[O-]

Integrate 2: Optimal BioSystems

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
•Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection

Continuity of Life & Central Dogma
Self-assembly, Catalysis, Replication, Mutation, Selection
Regulatory & Metabolic Networks

Polymers: Initiate, Elongate, Terminate, Fold, Modify, Localize, Degrade
"The" Genetic Code

Adjacent mRNA codons

‘Silent’ codon changes

Translation t-, m-, r-RNA

Large macromolecular complexes:
Ribosome: 3 RNAs (over 3 kbp plus over 50 different proteins)
The ribosome is a ribozyme.

Perl Dogma (EditPlus)

Continuity & Diversity of life

http://www.colorado.edu/psych/psych305/continuity_2.html

Evidence

Radiochemical dating:
Initial atoms remaining f = 1 - exp(-kt) (exponentially)
Molecular: endpt codon bias b.
Conserved silent codons f = b + (1 - b)exp(-kt).

How many living species?

5000 bacterial species per gram of soil (<70% DNA bp identity)
Millions of non-microbial species (& dropping)
Whole genomes: 100 done since 1995, 700 in the pipeline! (ref)
Sequence any: 16234 (in 1995) to 79961 species (in 2000) NCBI

& Why study more than one species?
Comparisons allow discrimination of subtle functional constraints.
Genetic codes (ncbi)

1. "Standard Code"
Base1 = TTTTTTTTTTTTTTTTCCCCCCCCCCCCCCCCAAAAAAAAAAAAAAAAGGGGGGGGGGGGGGGG
Base2 = TTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGGTTTTCCCCAAAAGGGG
Base3 = TCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG
AAs = FFLLSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ---M---------------M---------------M----------------------------

2. The Vertebrate Mitochondrial Code
AAs = FFLLSSSSYY**CCWWLLLLPPPPHHQQRRRRIIMMTTTTNNKKSS**VVVVAAAADDEEGGGG
Starts = --------------------------------MMMM---------------M------------

3. The Yeast Mitochondrial Code
AAs = FFLLSSSSYY**CCWWTTTTPPPPHHQQRRRRIIIMMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ----------------------------------MM----------------------------

11. The Bacterial "Code"
AAs = FFLLSSSSYY**CC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = ---M---------------M------------MMMM---------------M------------

14. The Flatworm Mitochondrial Code
AAs = FFLLSSSSYYY*CCWWLLLLPPPPHHQQRRRRIIIMTTTTNNNKSSSSVVVVAAAADDEEGGGG
Starts = -----------------------------------M----------------------------

22. Scenedesmus obliquus mitochondrial Code
AAs = FFLLSS*SYY*LCC*WLLLLPPPPHHQQRRRRIIIMTTTTNNKKSSRRVVVVAAAADDEEGGGG
Starts = -----------------------------------M----------------------------

Translational reprogramming


Herbst KL, et al. 1994
PNAS 91:12525-9
A mutation in ribosomal protein L9 affects ribosomal hopping during translation of gene 60 from bacteriophage T4.
"Ribosomes hop over a 50-nucleotide gap during translation..."

Integrate 2: Optimal BioSystems

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
• Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection

Jacob & Monod
http://www.nobel.se/medicine/laureates/1965/jacob-lecture.html

Gene Ontology (nature of being)

The objective of GO is to provide controlled vocabularies for the description of the molecular function, biological process and cellular component of gene products.

... Many aspects of biology are not included (domain structure, 3D structure, evolution, expression, etc.)... small molecules (Klotho or LIGAND)
**Gene Ontology (GO)**

- **Molecular function**
  What a gene product can do without specifying where or when. (e.g. broad "enzyme"; narrower "adenylate cyclase")

- **Biological process**
  >1 distinct steps, time, transformation (broad: "signal transduction." narrower: "cAMP biosynthesis.")

- **Cellular component**
  part of some larger object, (e.g. ribosome).

---

**Evidence for facts (GO)**

- IMP inferred from mutant phenotype
- IGI genetic interaction
- IPI physical interaction
- ISS sequence similarity
- IDA direct assay
- IEP expression pattern
- IEA electronic annotation
- TAS traceable author statement
- NAS non-traceable author statement

---

**Sources of Data for Qualitative Models:**

- **Direct observation**
  Capillary electrophoresis (DNA Sequencing): 0.4Mb/day
  Chromatography-Mass Spectrometry (e.g. peptide LC-ESI-MS): 20Mb/day
  Microarray scanners (e.g. RNA): 300 Mb/day
  Other microscopy (e.g. subcell, cell, tissue networks)

---

**Signaling Pathway Database (SPAD)**

- >500 bio-databases
- How are the data & models entered?

---

**Dynamic simulation of the human red blood cell metabolic network.**

- Dominant alleles affecting variety of RBC proteins, malaria, drug-hemolysis, etc.
- Rare individually, common as a group.

---


Bioinformatics 17: 286-287.
Enzyme Kinetic Expressions

\[
\frac{v_{\text{meq}}}{N_{\text{meq}}} = \frac{F6P/K_{\text{meq}}} {1 + F6P/K_{\text{meq}}} \left( \frac{\text{Mg} \cdot \text{ATP}/K_{\text{meq}}}{1 + \text{Mg} \cdot \text{ATP}/K_{\text{meq}}} \right) \\
N_{\text{meq}} = 1 + \frac{K_{\text{meq}}}{K_{\text{meq}} + \text{AMP}} \\
\]

How do enzymes & substrates formally differ?

\[
A \xrightarrow{\text{E}} \text{EATP} \xrightarrow{\text{ADP}} \text{E2} + P \\
E \xrightarrow{\text{E2} + P} \text{ATP} \\
\]

Catalysts increase the rate (and specificity) without being consumed.

Continuity of Life & Central Dogma
Self-assembly, Catalysis, Replication, Mutation, Selection
Regulatory & Metabolic Networks

Integrate 2: Optimal BioSystems

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
• Functional Genomics & Quantitative models
  Mutations & Selection

Sources of Data for Quantitative Models:

- Capillary electrophoresis (DNA Sequencing):
  0.4 Mb/day

- Chromatography-Mass Spectrometry (eg. peptide LC-MS-MS):
  20 Mb/day

- Microarray scanners (eg. RNA):
  300 Mb/day

- Other microscopy (eg. subcell, cell, tissue networks)

Structural Genomics
(the challenge of distant homologs)

\[? \quad ?\]

Functional Genomics
(quantitative ligand interactions)

100% Sequence Identity:
1. Enolase Enzyme
2. Major Eye Lens Protein

100% Sequence Identity:
1. Thioredoxin Redox
2. DNA Polymerase Processivity

Polymers: Initiate, Elongate, Terminate, Fold, Modify, Localize, Degrade
mRNA expression data & protein binding & mutant growth …

What is functional genomics?

Function (1): Effects of a mutation on fitness (reproduction) summed over typical environments.
Function (2): Kinetic/structural mechanisms.
Function (3): Utility for engineering relative to a non-reproductive objective function.

Proof: Given the assumptions, the odds are that the hypothesis is wrong less than 5% of the time, keeping in mind (often hidden) multiple hypotheses.

Is the hypothesis suggested by one large dataset already answered in another dataset?

Genomics Attitude

Whole systems: Less individual gene- or hypothesis-driven experiments; Automation from cells to data to model as a proof of protocol.

Quality of data: DNA sequencing raw error: 0.01% to 10%. Consensus of 5 to 10 error: 0.01% (1e-4)

Completion: No holes, i.e. regions with data of quality less than a goal (typically set by cost or needs of subsequent projects).

Impossible: The cost is higher than reasonable for a given a time-frame and quality assuming no technology breakthroughs. Cost of computing vs. experimental “wet-computers”.

Integrate 2: Optimal BioSystems

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
• Mutations & Selection

Mutations and selection

Types of Mutants

Null: PKU
Dosage: Trisomy 21
Conditional (e.g. temperature or chemical)
Gain of function: HbS
Altered ligand specificity
In-frame mutants + wild-type Pool Select Multiplex PCR size-tag or chip readout

Multiplex competitive growth experiments:

80% of 34 random yeast insertions have s<-0.3% or s>0.3%
for t=160 generations, e=1 (rich media); ~50% for t=15, e=7.
Should allow comparisons with population allele models.

Multiplex competitive growth experiments:

Non-optimal evolves to optimal

Integrate 2: Optimal BioSystems

Elements & Purification
Systems Biology & Applications of Models
Life Components & Interconnections
Continuity of Life & Central Dogma
Qualitative Models & Evidence
Functional Genomics & Quantitative models
Mutations & Selection