Looking for a place for theory

Andrew Murray

NSF/Simons Center for the Mathematical and Statistical Analysis of Biology

Harvard
Outline

Some thoughts on theory

Brief MathBio Center intro

Simplifying hypotheses?

Do yeast hedge their bets?

How do stem cells commit to a fate?
What should biological theory look like?

How does reproductive fitness, $w$, change during evolution?

\[
\frac{dw}{dt} = f(N_e, U_b, U_d, \rho(s), \text{Epistasis, Sex})
\]

Simplify: no sex, deleterious mutation, epistasis, $N,s =$ constants

Result:

\[
\frac{dw}{dt} = s^2 \left[ \frac{2 \ln(Ns) - \ln\left(\frac{s}{U_b}\right)}{\ln^2\left(\frac{s}{U_b}\right)} \right]
\]

speed proportional to $s^2 \ln(N_e)$

speed increases with $\ln(U_b)$

Desai & Fisher, Genetics, 2007
What should biological theory look like?

Fit surprisingly good given violence of the assumptions

*But is it good and extensible enough?*

Desai et al, Current Biology, 2007
What should biological theory look like?

J. Z. Young

\[ I = C_m \frac{dV_m}{dt} + \bar{g}_K n^4 (V_m - V_K) + \bar{g}_{Na} m^3 h (V_m - V_{Na}) + \bar{g}_l (V_m - V_i), \]

\[ \frac{dn}{dt} = \alpha_n (V_m) (1 - n) - \beta_n (V_m) n \]

\[ \frac{dm}{dt} = \alpha_m (V_m) (1 - m) - \beta_m (V_m) m \]

\[ \frac{dh}{dt} = \alpha_h (V_m) (1 - h) - \beta_h (V_m) h \]

Alan Hodgkin

Andrew Huxley

Rod MacKinnon

K⁺ channel
What should biological theory look like?

Other examples

Circadian clocks
Bacterial chemotaxis

What do they have in common?

Single, well-defined phenomenon
Appropriate level of abstraction
Simplifying assumptions productive
Experiment-theory linkage

Used simple, existing mathematics
What do we want?

More examples for single phenomena?

Even more accurate models for our successes?

Incorporating additional levels of abstraction?

Models relevant to the real world (disease, ecological change)?

Going beyond single phenomena

Purely predictive: leave it to Google?

Curiosity satisfying: what should we study?
From genotype to phenotype to function

Genetic networks

Cell fate decisions

Cells make tissues make organs

Morphologies show adaptation
Operating principles

Understand emergence of functional phenotypes:
- Molecular circuitry
- Form
- Function

At multiple scales and in multiple systems
- Gene → cell → organism → environmental adaptation

Through a virtuous cycle
- Assumptions → mathematical model → experiment
  
  statistical inference
Layered understanding

Interlayer interactions: chemical, mechanical, electrical, optical
Leadership

Andrew Murray (MCB)
Cassandra Extavour (OEB/MCB)
L. Mahadevan (Applied Math)
Sharad Ramanathan (MCB/SCRB)
Johan Paulsson (HMS Sys Bio)
Decision Making

Paola Arlotta (SCRB)
Richard Losick (MCB)
Johan Paulsson (HMS Sys Bio)
Sharad Ramanathan (MCB/SCRB)
David Scadden (SCRB)
Morphogenesis

Victoria D'Souza (MCB)
Ethan Garner (MCB)
Doeke Hekstra (MCB/Appl. Math)
L. Mahadevan (Applied Math)
Vinothan Manoharan (SEAS)
Daniel Needleman (MCB/SEAS)
Adaptation
Our goals

Science goals:
- Connect layers, find new principles
- Predict
- Exploit extant & develop new math
- Collaboration between mathematicians, statisticians, biologists

People goals: train & disperse new Hodgkins and Huxleys
- Trainee-led community
- Infrastructure for learning across disciplines
- Strongly mentored independence
- Interact & influence through workshops and visitor program
Might life be simpler than we think?
Competition yields complexity
Simplifying two complex problems?

Maintaining cellular composition

Responding to variable environments
Maintaining cellular composition

Cells are complex systems

Many components

- Boeing 787: 2.3 million parts in 15 million liters
- Budding yeast cell: 50 million parts in 50 pl

Part density: yeast:787 = $6 \times 10^{18}:1$

Proportions matter

Sophisticated circuits control synthesis & destruction

Who regulates the regulators?
Thinking from first principles

6000 genes in budding yeast

10 parameters per protein

amount, activity, partners, modification, location, etc.

Dividing cells follow paths in ≈60000-D space

How similar are their paths?

How unimodal are distributions?

How do cells transit between very different states?
Plasmid segregation is a bad model

DNA is special: omne DNA ex DNA

You must never lose the last molecule

High mean copy numbers are no protection

If one cell’s copy number = 20, \( p(\text{loss}) = 2^{-20} \)

If population mean = 20, \( p(\text{loss}) >> 2^{-20} \)
Error 1: proteins aren’t like DNA
Growth without division is a bad model

\[
\frac{dX}{dt} = k_{\text{synthesis}} - k_{\text{destruction}}X
\]

Deterministically, steady state \( X = \frac{k_{\text{synthesis}}}{k_{\text{destruction}}} \)

But molecules are indivisible

Molecularly, synthesis & destruction are stochastic

Temporally, number of \( X \) is a 1-D random walk
Error 2: cells must divide

\[ \frac{dA}{dt} = k_{\text{synthesis}} \]
In reality, cells divide

\[
\frac{dA}{dt} = k_{\text{synthesis}} - k_{\text{dilution}}A
\]
Keeping error at bay

Howard Berg

Johan Paulsson
Dilution bounds variation

100 different protein species
Make $100 \pm 10$ copies of each, segregate binomially

Protein Abundances for Mean = 100 over 201 cycles

- Most abundant Pr
- Least abundant Pr
- Median abundance Pr
- 91st most abundant Pr
- 10th most abundant Pr
So why bother to regulate?

To ensure correct ratios of components?

To match cellular composition to environment?

What do cells know about their environment?
More senses than you think!

We have >> 5 senses ($P_{O_2}$, $P_{osm}$, pH, etc.)

So do micro-organisms

Bacteria use two component systems
  Histidine kinase phosphorylates response regulator
  Number of pairs averages 40, can reach 200
How are global responses integrated?

Imagine 40 two component systems
Each has three levels of activity: 0, med, hi
There are $3^{40} = 10^{19}$ combinations
Books in Widener $\approx 5 \times 10^{12}$ characters

Dimension reduction badly needed!
Option 1: central processing unit

We use a nervous system to compute

Could microbes have a protein based CPU?
   How could it work (multisite phosphorylation?)?
   How many components would it need?
   What is the phenotype of mutants?
   Could we have missed it?
Option 2: a single processing rule

Every pathway has three output levels
- No corrective action required
- Local homeostatic response
- Stop growth & cell cycle

Signals from different pathways combined

Possible rules
- Respond only to strongest signal
- Respond to sum of signals
- Non-linear response

Accessible to experimental test
Do yeast hedge their bets?
Yeast have two strategies for energy production:

- **High glucose**
  - Glycerolysis
  - Fermentation
  - 2 ATP / molecule

- **Low/no glucose**
  - Glycerolysis
  - Respiration
  - Other sugars
  - Mitochondrion
  - ~32 ATP / molecule
Interconversion between metabolic states is slow
Following the transition between metabolic states
Anatomy of an experiment

Cells trapped in microfluidic chamber

High-glucose media flow
Glucose repression

Abrupt switch to glucose-free media flow

No-glucose media flow
Shift to glucose depression
Cell state: quantitative analysis of mitochondrial networks
Phenotypic heterogeneity after glucose deprivation

- + Glucose
- - Glucose

+ Brightfield

Mito-mNeonGreen

Scale: 10 µm
Mitochondrial heterogeneity after glucose deprivation
Predictive power of mitochondrial volume

10th percentile

90th percentile

-60'

30'

120'

240'

405'

10th percentile

90th percentile

Density

Sphericity

Density

Sphericity

Density

Sphericity

Density

Sphericity
Arrest or recovery is a heritable trait

Sisters

Random pairs

\( r^2 = 0.47 \)

\( r^2 = 0.00 \)
Metabolic regulators control physiology

Encourage fermentation

Encourage respiration
Metabolic regulators control cell state

+ Glucose - Glucose
+ Brightfield Mito-mNeonGreen

HXK2

hxk2Δ
Metabolic regulators control cell state

Wildtype

\( snf1\Delta \)

\( hxk2\Delta \)
Arrest/recovery states are reflected in pH dynamics
Recovery during starvation is memory-dependent

Respiration (acetate) → Glucose (0 to 24 hrs) → Glucose deprivation
Recovery during starvation is memory-dependent

\[ y = \frac{-0.809}{1 - e^{-0.835 \times (t - 4.502)}} + 1.033 \]
Half-life: 4.50 hr

Switching rate, recoverers > arresters:
0.23 hr\(^{-1}\)

Switching rate, arresters > recoverers:
\[
\frac{0.23}{3.46} = 0.07 \text{ hr}^{-1}
\]

Recovery during starvation is memory-dependent
Tradeoff: faster specialist vs slower generalist?

Environment:
High glucose  No glucose

Rate of glycolysis  

Relative fitness  

Rate of glycolysis  

Relative fitness  

Tradeoff: faster specialist vs slower generalist?
How do stem cells commit to a fate?

Jim Valcourt

Sharad Ramanathan
Signals recycled in development

5-6 signaling pathways \( \rightarrow \) 100s of cell types

Option 1: details (BMP2 vs BMP4) matter

Option 2: combinatorial & concentration code

Option 3: history matters
Stem cells can adopt different fates

Stem Cell → Mesendoderm

Stem Cell → Bipotent ectoderm

Bipotent ectoderm → Non-neural ectoderm
Response to signals change as cells differentiate

A stem cell’s signal response changes during differentiation
Can’t isolate cells at commitment point

Differentiation markers only seen in brown or purple state

Differentiation takes 24 hrs after commitment
Can we read a cell’s mind in real time?

Math on gene expression reveals binary decision factors

Mesoderm: Oct4 UP, Sox2 DOWN

Bipotent ectoderm: Oct4 DOWN, Sox2 UP

Jang et al, eLife, 2017
Differentiation in living stem cells

- Signal added (BMP + Activin A)
- Differentiating towards neural (Activin/Nodal inhibition)
- Differentiation (54 h)
- Signal (38 h)
Oct4:Sox2 predicts the signal response

- cells that end up as mesendoderm
- cells that end up as bipotent ectoderm

OCT4:SOX2 predicts eventual fate

time after BMP4 + Activin A addition (h)
Collect “omic” data from three points for analysis:

- Epiblast
- Mesendoderm
- Bipotent ectoderm
- Non-neural ectoderm

Mesendoderm (outgroup to identify lineage-specific changes)

Data types: RNA-seq, ATAC-seq, & ChIP-seq of key factors
Develop math model of commitment network

Test model’s predictions