

Lecture 1: Biomolecular Modeling



Richard M. Murray Caltech CDS/BE

Goals:

- Motivate the role of dynamics and feedback in the the cell
- Provide mathematical tools for analyzing and predicting the dynamics of transcriptional regulation in the cell
- Work through case study for the lac operon (model control system)

References

- Alberts et al, *Essential Cell Biology*, 3rd edition, 2009.
- R. Phillips, J. Kondev and J. Theriot, *Physical Biology of the Cell*, 2008.
- N. Yildirim, M. Santillan, D. Horiki and M. C. Mackey, "Dynamics and bistability in a reduced model of the lac operon". Chaos, 14(2):279-292, 2004.
- D. Del Veccho and R. M. Murray, *Biomolecular Feedback Systems*, http:// www.cds.caltech.edu/~murray/amwiki/BFS (preprint)

2009 Asian Control Conference (ASCC), 26 August 2009



Prokaryote (bacterium)

- Simplest type of cell: cell wall, cytoplasm, flagella/pili
- DNA is circular, single copy
- Simple cell division (mitosis): DNA duplicated in each daughter cell
- Examples: E. coli, Bacillus subtilis
 - *E. coli:* 1 µm long, 20 min cycle

Eukaryote: single and multi-cell

- Much more complicated structure
- DNA resides in nucleus; proteins are formed in the cytoplasm
 - DNA organized in chromosomes
- Organelles implement cell functions
 - Nucleus: contains DNA
 - Mitochondria: energy production

Central Dogma: DNA to Proteins



Transcription: DNA to mRNA

- Double stranded DNA contains nucleotide sequence (A, C, T, G) on a sugar (deoxyribose) backbone
- Watson-Crick pairing: A:T, C:G
- RNA polymerase *transcribes* DNA sequence to RNA sequence (A, C, U, G sequence on a ribose backbond)

Translation: mRNA to protein

- (Eukaryotes) pre-mRNA is spliced and processed to form mature mRNA
- mRNA is *translated* by ribosomes into a chain of amino acids using the genetic code (3 bp code for 1 aa)
- Amino acid chain (polypeptide chain) folds into a protein

Regulation: control gene expression

• Proteins bind to DNA, RNA and proteins to modulate expression

Richard M. Murray, Caltech CDS

Model Control System: The Lac Operon



Key idea (Jacob & Monod, 1950s): produce proteins when you need them

- Bacterial growth dependent on nutrient environment
- Lactose only consumed if glucose is not present
- Q: How does E. coli decide when to produce proteins?
- A: Lac "operon" (≈ control system)
- KPT07: "Lac operon is the hydrogen atom of regulation"



How does it work?

- Proteins for digesting lactose are controlled by binding of repressor and activator
- CAP (activator): recruits RNA polymerase when bound to cAMP, whose concentration is controlled by absence of glucose (+glucose \Rightarrow no cAMP)
- Repressor: blocks transcription by causing DNA looping unless it is bound to allolactose, a byproduct of lactose metabolism
- Feedback: if lactose present, then create proteins required to turn on lac operon, which creates the proteins required to matabolise lactose



DNA facts

- Typical gene is about 1000 bp (300 aa) long
- E. coli genome is 4.6 Mb, with ~4000 genes
- Human genome is about 4 Gb, with ~25,000 genes

Lac control circuitry

- Main circuitry is about 6000 bp (genes plus regulatory regions)
- Can see structure of the promotor in DNA sequence

Operon Layout and Molecular Census



Repressor and DNA looping



Repressor acts through DNA looping

 Looking can occur between binary binding site (O1) and secondary sites on DNA (O2, O3)





ession =
$$\frac{p_{\text{bound}}(R=0)}{p_{\text{bound}}(R)}$$

- Modeling via statistical mechanics: compute *p*_{bound} of RNA binding to DNA for different situations
- Experiments: relocate secondary binding sites and see what happens
- Still get some repression even if there is only one operator site present (no looping?)



Phillips, Kondev, Theriot (2008)

ASCC, Aug 09

Rates of Transcriptional Regulation



Primary timescales:

- mRNA production: 10-30 bp/sec
- Protein production: 10-30 aa/sec
- Protein folding: (depends)

DNA, protein from PKT08, mRNA production from Vogel & Jensen

Other important rates

- mRNA half life : ~100 sec
- Protein half life: ~5 x 10⁴ sec
- Protein diffusion (along DNA): up to 10⁴ bp/sec
- Typically assume that activators and repressors reach equilibrium state much more quickly than other time scales

Half life times from Yildirim and Mackey, 2003; Protein diffusion from Blainey et al, PNAS 2006.

Richard M. Murray, Caltech CDS

From Numbers to Equations: Dynamical Modeling



Modeling philosophy: Ask questions first, build model later

- Many different models possible for the same system; no such thing as "the model"
- The model you use depends on the questions you want to answer
- Never build a model without first posing the questions

Analysis and design based on models

- A model provides a prediction of how the system will behave
- Feedback can give counter-intuitive behavior; models help sort out what is going on
- Models don't have to be exact to be useful; they just need to help explain (and predict)

Dynamic Modeling Approaches



Choice of model depends on the questions you want to answer

- Modeling takes practice, experience and iteration to decide what aspects of the system are important to model at different temporal and spatial scales
- Good analysis models make testable predictions and produce "surprising" results
- Good design models capture enough of the important behavior to give allow good design decisions

Statistical Mechanics Viewpoint



• Based on energy arguments; allows study of complex situations

ASCC, Aug 09

The Master Equation: Detailed Events and Rates



Key idea: transition rates between microstates

- Enumerate micro-states corresponding to the system of interest
- Define the system "state" in terms of the individual probabilities of each microstate at each instant in time
- Dynamics are given by rate of change of the probability of each individual state

Transition rates depend on detailed model of molecular dynamics

• Example: lattice structure captured by Botzmann model

$$k_{\rm des,i} = \nu \exp\left(-\frac{E_{\rm des,0} + i\Delta E}{k_b T}\right)$$

($v = vibrational frequency, 10^{13}$)

 For DNA, RNA, protein interactions, can look at probability of collision + binding energy to reason about rates; also possible to identify via experiments

Simulating the Master Equation for Chemical Reactions

Stochastic Simulation Algorithm (SSA)

- N states $(X_1, X_2, ..., X_N)$ where X_i is the number of copies of species S_i in the system
- *M* reaction channels R_i that define changes in the state; $v_{ii}(X)$ = change in X_i for R_i
- Propensity function: $a_i(X) dt$ = probability that a reaction R_i will occur in time interval dt
- SSA ("Gillespie algorithm"): determine how many reactions occur in a given time step and execute; then update propensity functions and repeat
- Choose time steps to be small enough that propensity functions are roughly constant

Example

$X_{\infty} + Y_1 \longrightarrow 2 Y_1$	with propensity a_1
$\mathbf{Y}_1 + \mathbf{Y}_2 \longrightarrow 2\mathbf{Y}_2$	with propensity a_2
$\mathbf{Y}_2 \longrightarrow \mathbf{Z}$	with propensity a_3

• Propensity function: $a1(X) = c_1 X_{\infty} Y_1$ to account for multiple copies of each species

Tools

- SBML modeling language + many tools
- StochKit (Linda Petzold) C++ libraries
- SimBiology (MATLAB) includes SSA



Chemical Kinetics: The Law of Mass Action

Alternative approach: keep track of concentrations

- If number of molecules is large, we can keep track of concentration of each species
- No longer track individual events; assume an average rate of events and use ODEs

 $X_{\infty} + Y_1 \longrightarrow 2 Y_1$ with rate k_1 $Y_1 + Y_2 \longrightarrow 2 Y_2$ with rate k_2 $Y_2 \longrightarrow Z$ with rate k_3

Keep track of change due to each reaction

$$d[Y_1]/dt = k_1[X_{\infty}][Y_1] - k_2[Y_1][Y_2]$$

$$d[Y_2]/dt = k_2[Y_1][Y_2] - k_3[Y_2]$$

More complicated reactions:

$$2 \mathbf{A} + \mathbf{B} \xrightarrow[k_{-1}]{k_{-1}} \mathbf{A}_2 \mathbf{B}$$

$$d[A]/dt = 2k_{-1}[A_2B] - 2k_1[A]^2 \cdot [B]$$

$$d[B]/dt = k_{-1}[A_2B] - k_1[A]^2 \cdot [B]$$

$$d[A_2B]/dt = k_1[A]^2 \cdot [B] - k_{-1}[A_2B]$$



Lotka Two Ways



- x = concentration vector
- *S* = stochiometry matrx
- v(x) = Rate vector

ASCC, Aug 09

Michaelis-Menten Kinetics

Enzymatically controlled reactions

• Enzyme E acts as a catalyst for reaction

$$\mathbf{E} + \mathbf{S} \xrightarrow[k_{-}]{k_{-}} \mathbf{ES} \xrightarrow{r} \mathbf{E} + \mathbf{P}$$

• Dynamics follow from mass action kinetics

$$\frac{d[E]}{dt} = k_{-}[ES] - k_{+}[E] \cdot [S] + r[ES]$$
$$\frac{d[S]}{dt} = k_{-}[ES] - k_{+}[E] \cdot [S]$$
$$\frac{d[ES]}{dt} = k_{+}[E] \cdot [S] - k_{-}[ES] - r[ES]$$
$$\frac{d[P]}{dt} = r[ES]$$

• If we assume enzyme dynamics are fast compared to other species, we can simplify the resulting dynamics:

 $d[\mathbf{P}]$

dt

$$[\text{ES}] = [\text{E}] \cdot [\text{S}] / K_m$$
$$K_m = \frac{k_- + r}{k_+}$$

ASCC, Aug 09



 $= r[\mathbf{E}]_{\mathrm{tot}}$

ES

Phosphorylation

Post-translational modification of proteins

- Addition of a phosphase(PO₄) group to a protein
- Modifies shape of protein; affects binding to other proteins (Xp "active")

Ε ATP ADP bindina kinase release ATP E:ADP E:ATP kinase binding kinase B:ADP:Xp E:ATP:X ADP + catalvsis substrate х substrate F:XP substrate release substrate F

Kinase

- Binds ATP and catalyzes a transfer of phosphate group to protein
- Enzymatic reaction (kinase is available again after reaction)

Phosphatase

- Removes phosphate group
- Non-specific enzyme

Phosphorylation $E + ATP \Longrightarrow E:ATP$ $E:ATP + X \Longrightarrow E:ATP:X$ $E:ATP:X \longrightarrow E:ADP:Xp$ $E:ADP:Xp \Longrightarrow E:ADP + X$ $E:ADP \longrightarrow E + ADP$

Dephosphorylation $Xp + F \Longrightarrow Xp:F$ $Xp:F \longrightarrow X + F$

> Consider adding hydrolysis more explicitly in dephosphorylation

ADP

 X_{P}

Transcriptional Regulation



Yildirim-Mackey Model for the Lac Operon



Questions:

- In the absence of glucose, what concentration of lactose is required for the *lac* operon to become "active"?
- Focuses on "bistability": *lac* operon has two stable equibrium points:
 - low lactose: machinery off
 - high lactose: machinery on

Model

- Ordinary differential equation for rates of transcription, translation and degradation of β-galactosidase (β-gal) and allolactose
- Assume levels of lactose outside the cell is constant and level of permease (from *lacY*) is constant to simplify the model
- Takes into account time delays in producing proteins (RBS transcription + protein translation)

ASCC, Aug 09

Richard M. Murray, Caltech CDS

Model Derivation: **B-Gal Production**

$$\frac{dM}{dt} = \alpha_M \frac{1 + K_1 (e^{-\mu \tau_M} A(t - \tau_m))^n}{K + K_1 (e^{-\mu \tau_M} A(t - \tau_m))^n} - \bar{\gamma}_M M$$

$$\frac{dB}{dt} = \alpha_B e^{-\mu\tau_B} M(t - \tau_B) - \tilde{\gamma}_B B$$

M = lacZ mRNA concentration

B = β -gal concentration

A = allolactose concentration



mRNA production

 Production rate related p_{bound} via a modified Hill function, ala

$$\frac{dM}{dt} = \alpha \left(\left[1 - p_{\text{bound}}(A) \right] + \epsilon \right) - \gamma M$$

- RNA degradation via exponential decay
- Account for time delay in translation of RBS via τ_M :
 - Use allolactose concentration, *A*, at time *t* τ_M
 - Exponential factor to account for dilution due to cell division

Protein production

- Assume rate of production is proportional to amount of mRNA
- Include protein degradation via exponential decay
- Add time delay to account for time to produce functional protein, τ_B

Model Derivation: Allolactase Dynamics



- A = allolactose concentration
- L = internal lactose concentration
- P = permease concentration
- B = β -gal concentration

Allolactose

- 1. Converted from lactose with Michaelis Menten-like kinetics (Huber et al)
- 2. Converted back to glucose and galactose
- 3. Degradation

Lactose (internal)

- 1. Transported to interior of cell by permease
- 2. Loss back to external environment
- 3. Converted to allolactose by β -gal
- 4. Degradation

Permease

- 1. Produced by *lacY* gene (after delay)
- 2. Degradation

20

Richard M. Murray, Caltech CDS

Determining the Constants

Parameter	Value
$\mu_{ m max}$	$3.47 \times 10^{-2} \text{ min}^{-1}$
$\overline{\mu}$	3.03×10 ⁻² min ⁻¹
am	997 nM-min ⁻¹
α_{B}	1.66×10 ⁻² min ⁻¹
a	1.76×10 ⁴ min ⁻¹
YM	0.411 min ⁻¹
YB	8.33×10 ⁻⁴ min ⁻¹
YA	1.35×10 ⁻² min ⁻¹
n	2
K	7200
K ₁	$2.52 \times 10^{-2} (\mu \text{ M})^{-2}$
KL	0.97 mM
KA	1.95 mM
BA	2.15×104 min-1
τ_M	0.10 min
τ_B	2.00 min

Yildirim, Santillan, Horike and Mackey:

- μ dilution rate, based on 20 minute cell division time
- α_x production rate, based on steady state values
- γ_x decay rate, based on half life experiments
- τ_M time delay to produce RBS, based on RNA elongation rates
- τ_B time delay to translate protein, based on protein length and translation speed
- *n* Hill coefficient (no justification!)
- *K* based on basal rate of production (Yagil & Yagil)
- *K*₁ based on dissociation constant (Yagil & Yagil)
- K_x measured by Wong, Gladney and Keasling (97)
- β_A loss of allolactase, through conversion to glucose and galactose. Measured by Hubert et al (75)

Note: repressor binding model is pretty ad hoc...

$$\frac{dM}{dt} = \alpha_M \frac{1 + K_1 (e^{-\mu \tau_M} A(t - \tau_m))^n}{K + K_1 (e^{-\mu \tau_M} A(t - \tau_m))^n} - \bar{\gamma}_M M$$

$$= \alpha_A B \frac{L}{K_L + L} - \beta_A B \frac{a}{K_A + A} - \tilde{\gamma}_A A$$

ASCC, Aug 09

Richard M. Murray, Caltech CDS

 $\frac{dA}{dt}$



Bistable behavior (saddle node bifurcations)

- Can have single or multiple equilibrium points depending on parameters
- Bifurcation plot: change in stability versus params
 - Note: ossible hysteresis from saddle node
- Parametric stability plot: stability regions
- Simulations: nearby initial conditions can lead to different steady state solutions
- Use to predict behavior (for future experiments)



ASCC, Aug 09

Richard M. Murray, Caltech CDS

Comparison to Experiment



β -gal activity for L_e = 8 x 10⁻² mM

- Experimental data from Knorre (1968) for *E. coli* ML30 (◊) and Pestka et al. (1984) for *E. coli* 294 (●)
- Model simulation using constants from Table 1 (slide 16) with μ = 2.26 x 10⁻² min⁻¹ and γ_x (= ??) fit to data



Oscillation in β -gal w/ phosphate feeding

- Periodic phosphate feeding from Goodwin (1969)
- Simulation used $\mu = 2.26 \times 10^{-2} \text{ min}^{-1}$ and γ_x (= ??). Other parameters unchanged.
- Q: how should we assess these data?

Modeling Noise in Chemical Reactions

Fundamental assumption (Gillespie, 1992)

• Treat species concentrations as a random variable X(t)

 $a_j(x) dt$ = the probability, given X(t) = x, that one R_j reaction will occur in the next infinitesimal time interval [t, t + dt]

Chemical Master equation (exact)

• Describes the evolution of the probability distribution for the microstate X(t)

$$\frac{\partial P(\mathbf{x}, t \mid \mathbf{x}_0, t_0)}{\partial t} = \sum_{j=1}^{M} [a_j(\mathbf{x} - \mathbf{v}_j) P(\mathbf{x} - \mathbf{v}_j, t \mid \mathbf{x}_0, t_0) - a_j(\mathbf{x}) P(\mathbf{x}, t \mid \mathbf{x}_0, t_0)]$$

Chemical Langevin equation (CLE)

- Assume that time steps are small enough but not too small
 - (i) Time increment *dt* small enough that propensity functions are approx constant
 - (ii) Time increment dt large enough that number of events is much larger than 1
- Under these assumptions, can derive stochastic ODE (Gillespie, 2000):

$$X_{i}(t+dt) = X_{i}(t) + \sum_{j=1}^{M} v_{ji}a_{j}(X(t))dt + \sum_{j=1}^{M} v_{ji}a_{j}^{1/2}(X(t))N_{j}(t)(dt)^{1/2}$$
White noise w/

Remarks

Mass action kinetics are deterministic version of mean of the CLE

ASCC, Aug 09

unit variance

Cell Noise (Elowitz et al, 2002)

Noise in cells

- Experiments by Elowitz, Levine, Siggia, Swain. *Science* 2002
- Put RFP and GFP under identical promoters; *should* get yellow
- Results: get range of colors

Extrinsic Noise:

 global to a single cell, but varies from one cell to the next (e.g. cell volume, plasmid copy number)

Intrinsic Noise:

• inherent stochasticity in gene expression (e.g. what order reactions occur in)

$$\dot{x}_i = E(t) \cdot f_i(x_i, I_i(t))$$

Coefficient of variation: $CV(X) = \frac{std(X)}{mean(X)}$

• Normalized measure of noise, commonly used in biology



Modeling Summary

Core processes in biological circuits

- Transcription, translation, binding
- Activation and repression
- Enzymatic reactions, (phosphorylation)

Quantitative modeling approaches

- Statistical thermodynamics
- Chemical master equation, SSA
- Chemical Langevin equation
- Mass action, Michaelis-Menten, Hill functions and ODEs

Lac operon

- Model system for transcriptional regulation
- Activation via CAP:cAMP
- Repression via Lac Repressor (DNA folding)

Next

- Control analysis techniques
- More complex biological circuits

