Topics

• Optimal gene circuits
  – Optimal protein expression levels
  – Selection for regulation in variable environments
  – Selection for regulatory modules
Selection for Regulation

- Everything about the structure and function of biomolecular mechanisms are subject to selection
  - Anything observed beyond a random occurrence must confer a selection advantage
    - Network motifs in gene expression and signal transduction networks
- The cells also operate in a constrained environment
  - Resource constraints
    - Nutrients
    - Oxygen
    - Metabolites
    - ...
  - Time constraints
- Realizing and maintaining regulatory constructs requires investment with non-negligible effort expenditures
- The question to be answered is this:
  When does it become beneficial to spend the effort to regulate?
Optimality of Gene Circuits

• Evolution operates by random changes to the genes
  – The changes that offer extra benefits are likely to achieve fixation
    • Wide-spread expression across the population after a few generations
  – The non-beneficial changes are much less likely to achieve fixation
    • Depletive changes are quickly selected against

• The rule of thumb:
  The mechanisms observed in cells must have accumulated through time specifically because they improved the likelihood of the organisms of surviving and leaving descendants

• The likelihood of leaving descendants is expressed as a fitness function
  – Therefore, changes are more likely to be kept if they improve the fitness function
Fitness Optimization

• The evolutionary fitness of an organism is difficult to define explicitly
  – Does not necessarily correspond to the physical strength or health

• The general approach is to take the average number of descendants as the fitness
  – Works reasonably well for bacterial cells
  – Awkward and severely inadequate for complex organisms
    • Problems with defining evolutionary fitness based on the number of descendants in a human society
Fitness Optimization in Simple Organisms

• The changes that allow leaving more descendants are favored
  – NB: Everything is, still and always, stochastic

• Making optimal use of the environmental resources is selected for with a high probability
  – Because it improves the chances of leaving descendants
  – Making optimal trade-offs also qualifies

• The stochastic nature of the selection process suggests a softer optimization than seeking “the” optimal configuration
  – The configurations with “sufficiently good” performance survive
  – Multiplicity of choices also provide alternative avenues of responding to environmental changes
    • Never-ending cycle of variations
    • Law of diminishing returns
Law of Diminishing Returns

- Cells do display a tendency for greater organization, but greater organization itself is ultimately not the goal
  - Higher organization is costly
    - Synthesis and maintenance of molecules involved in the regulatory mechanism
  - Organization is worth the effort up until a point of maximal benefit for the effort spent
  - After that level, more organization does not enjoy a selective advantage
    - Same reasoning follows for the cost/benefit of intelligence in animals
→ The law of diminishing returns in cellular organization

- Engine power vs. maximum speed (automotive industry)
- Detection rate vs. false alarm rate (pattern recognition)
- …
Optimal Protein Expression

• Expression levels of proteins are regulated by the biomolecular mechanism
  – Gene transcription networks operating based on the input from the signal transduction networks

• The regulatory mechanism adjusts the expression of a protein at a specific level

• The level at which the expression is maintained ought to optimize the fitness of the organism
  – The expression of a protein incurs a cost and provides a benefit
    • To achieve an increase in the fitness
  – The level itself is adjusted so that the benefit is worth the overall cost and the additional benefit is maximized

• Example: The Lac system
Optimal Expression of the LacZ Protein

• The LacZ gene is a component in the (E. coli) lactose metabolism
  – The protein product of LacZ breaks down lactose for use as energy and carbon source
• The level of LacZ expression in the cell balances two outcomes:
  – The benefit
  – The cost
• Both the benefit and the cost are measured in terms of the increase and the decrease observed in a fitness function: the growth rate
The Benefit Function of LacZ

- The benefit of LacZ is the relative increase in the growth rate per additional amount of LacZ
  - Hence, the benefit is proportional to the rate at which LacZ breaks down lactose

\[ b(|Z|, L) = \frac{\delta |Z|L}{K + L} \]

where

- \(|Z|\) denotes the number of LacZ proteins
- \(\delta\) denotes maximal growth rate advantage per LacZ protein
- \(K\) is the Michaelis constant
- \(L\) is the amount of lactose present
The Benefit Function of LacZ

- Experimental setup:
  - The cells are induced for maximal expression of LacZ
    - $|Z_{WT}|$ is maximal
  - The growth rate measured for varying levels of external lactose
  - The maximum benefit in saturating amounts of lactose was observed to be around 17%
- The values for the parameters $\delta$ and $K$ were determined as
  - $\delta = 0.17 |Z_{WT}|^{-1}$
  - $K = 0.4$ mM
The Cost Function of LacZ

• The cost of LacZ was also measured experimentally
  – The cell was induced to express LacZ at specific levels in complete absence of lactose
  – The cost was measured as the relative reduction in the growth rate
• The cost of LacZ was determined to be a monotonically increasing function of LacZ expression
  – Synthesis of LacZ reduces resources that otherwise would be available for other beneficial mechanisms
• A reasonably accurate model for the cost that incorporates this nonlinear behavior is
  \[ c(|Z|) = \eta |Z|/(1+|Z|/M) \]
  where \( M \) denotes the level at which \(|Z|\) production overwhelms the cell, and \( \eta \) is a scalar constant
  – When \(|Z| \approx M\), the cell begins to halt all other cellular operations and recruits all resources toward LacZ synthesis
The Cost Function of LacZ

- By experimental evidence
  - $\eta = 0.02 |Z_{WT}|^{-1}$
  - $M = 1.80$
- The cost when $|Z| = |Z_{WT}|$ is about 4.5%
  - For small $|Z|$, the cost is approximately linear with a slope of $\eta$
  - As $|Z|$ grows, the monotonically increasing behavior becomes more apparent
- Note that the cost is measured by the relative reduction in the growth rate
The Cost-Benefit Analysis of LacZ

- For a given external lactose concentration of \( L \), the optimal \(|Z|\) level is determined by maximizing the fitness function
  - The fitness function is defined by the relative growth rate with the benefit and the cost of expressing \( Z \) at a given level
    \[
    f(|Z|) = b(|Z|, L) - c(|Z|)
    = \delta |Z| L/(K+L) - \eta |Z|/(1+|Z|/M)
    = |Z|(\delta L/(K+L) - \eta/(1+|Z|/M))
    \]
  - Note that this function does have a maximum at
    \[
    \frac{\partial f}{\partial |Z|} = \delta L/(K+L) - \eta/(1+|Z|/M) + |Z|\eta/(M(1+|Z|/M)^2) = 0
    \]
    - The benefit grows linearly
    - The cost grows exponentially
    - Thus, however small it starts, the cost eventually grows much larger than the benefit
  - The level of \(|Z|\) that maximizes this function provides the optimal level of LacZ protein
  - The maximal value of the fitness function itself is the relative growth rate achieved at the optimal \(|Z|\)
The Cost-Benefit Analysis of LacZ
Remarks

• The cost-benefit analysis of LacZ reveals several behavior types
  – When there is little or no LacZ, the cells produce no LacZ at all
    • The benefit of LacZ does not exceed its cost
    • If the cells are maintained for sufficiently many generations, they would tend to lose the “useless” lactose metabolism
  – When lactose is present in significant amounts, the cells determine an optimal expression level for LacZ with respect to the maximal wild-type rate
    • This is achieved in a few hundreds of generations
  – The model also predicts that if the cells are kept in environments with lactose concentrations consistently higher than 0.6, they would evolve a higher wild-type expression level
Optimal Regulation in Variable Environments

• Regulation endows the biomolecular mechanism with the ability to determine when a gene product is needed
• But not all genes are regulated
  – Some genes are always expressed
  – Other (hypothetical) genes are never expressed (and have thus been lost through generations)
• The answer to the question of whether a gene is regulated or expressed at all is again subject to cost-benefit analysis
  – If a gene is always beneficial or needed, it would always be expressed without a need for a regulatory mechanism
  – On the other hand, if a gene is never beneficial, it is not needed at all and would gradually be lost through generations
  – Regulation would be selected for if determining when it is beneficial would improve the fitness
Optimal Regulation in Variable Environments

• The cost-benefit analysis for maintaining a regulatory mechanism:
  – Regulation entails a cost
  • Synthesis and maintenance of all molecules taking part in a regulatory mechanism requires expenditure in
    – raw materials,
    – energy, and
    – processing time
  – The cost is compensated for if the regulated system provides sufficient benefits when the favorable conditions are present
Optimal Regulation in Variable Environments

• Consider a gene product Z
• Let $p$ be the fraction of time when the conditions in which Z is utilized are present
  – Hence, $p$ signifies the demand for the gene product Z
• The fitness with respect to Z is then given by
  \[ f_Z = pb - pc - r \]
  where $b$ and $c$ denote the benefit and the cost of utilizing Z during the favorable times, and $r$ denotes the cost of maintaining its regulatory mechanism
Optimal Regulation in Variable Environments

• The alternatives to regulation are
  – Constant expression without regulation:
    \[ f_Z = pb - c \]
  – No expression:
    \[ f_Z = 0 \]

• Regulation will be selected when
  \[ pb - pc - r > pb - c \text{ and } pb - pc - r > 0 \]

• Constant expression will be selected when
  \[ pb - c > pb - pc - r \text{ and } pb - c > 0 \]

• No expression will be selected when
  \[ 0 > pb - pc - r \text{ and } 0 > pb - c \]
Optimal Regulation in Variable Environments

\[ b = b_0, \ c = c_0 \ (b_0 > c_0) \]

- no expression
- constant expression
- regulation

\[ r \]

\[ p \]
Selection for Regulatory Modules

• A regulatory system is selected for as long as the net benefit in favorable times compensates for the cost of maintaining it

• But among different regulatory mechanisms, some seem to be selected for instead of others
  – These modules are recognizable as network motifs
  – Network motifs confer specific benefits to the network
    • Quicker response
    • Sign-sensitive delay
    • Temporal programming
    • …
  – Their benefit must be calculated with respect to those of alternative (simpler) regulatory mechanisms
Selection for C1-FFL

- Consider two alternatives for expressing the gene Z only when the signals for both genes X and Y are present
  - Simple regulation
    - The signals trigger the expression of their respective genes
    - When both genes pass their respective thresholds, the gene Z starts to be expressed
  - Feed-forward loop
    - The signals for the gene X triggers its expression
    - Only after the gene product X is above a threshold does the expression of Y begins
    - The expression of the gene Z begins after a delay
Selection for C1-FFL

• Suppose that the expression of the gene product Z has a certain relative fitness $f(t)$
  – The dependence on time indicates the varying cost-benefit structure of Z as the level of Z rises in response to expression signals
• Let the triggering of the Z expression for $D$ units of time be represented by a pulse of duration $D$
  – From the time when $S_X$ and $S_Y$ are both turned on until $S_X$ is turned off
• In case of a pulse of duration $D$, the total acquired relative fitness $\phi(D)$ can be computed by
  $$\phi(D) = \int_0^D f(t)dt$$
• Considering that pulses can have random durations, the average relative fitness is
  $$\Phi = E\{\phi(D)\} = \int_0^\infty \phi(D)p(D)dD$$
  where $p(D)$ is the probability distribution associated with the pulse durations $D$
Selection for C1-FFL

• For simple regulation:
  – Supposing that the gene product Z becomes available after a time delay $T$ following the signal, the fitness function is given by
    $$f(t) = b \cdot 1(t - T > 0) - \beta \eta$$
  • $b$ is the benefit, $\beta$ is the production rate, and $\eta$ is the cost per produced protein product
  – The accumulated fitness from a pulse of duration $D$ is then
    $$\phi(D) = b(D - T) - \beta \eta D$$
    if $D \geq T$, and
    $$\phi(D) = -\beta \eta D$$
    otherwise
  – The corresponding average accumulated benefit is given by
    $$\Phi_{sr} = \int_{T}^{\infty} b(D - T) p(D)dD - \int_{0}^{\infty} \beta \eta D p(D)dD$$
Selection for C1-FFL

• For feed-forward loop regulation:
  – The expression of Z begins after a delay of $\tau$, and the gene product Z becomes available $T + \tau$ times after the pulse begins
  – The associated fitness function becomes
    $$ f(t) = b \mathbf{1}(t - T - \tau > 0) - \beta \eta \mathbf{1}(t - \tau > 0) $$
  – The accumulated fitness becomes
    $$ \phi(D) = b(D - T - \tau) - \beta \eta(D - \tau) $$
    if $D \geq T + \tau$,
    $$ \phi(D) = -\beta \eta(D - \tau) $$
    if $\tau < D < T + \tau$, and
    $$ \phi(D) = 0 $$
    if $D < \tau$
  – The corresponding average accumulated fitness is then given by
    $$ \Phi_{ffl} = \int_{T+\tau}^{\infty} b(D - T - \tau)p(D)dD - \int_{\tau}^{\infty} \beta \eta(D - \tau)p(D)dD $$
Selection for C1-FFL

• The selection between simple regulation and feed-forward loop regulation is determined based on the distribution $p(D)$ of pulse durations
  – If all pulses are very short, the simple regulation would incur costs with no benefits, while the feed-forward loop would prevent these costs
  – If all pulses last very long, then the simple regulation would be preferred since the feed-forward loop would always loose the benefits to be acquired during the extra delay $\tau$
  – If, on the other hand, the probability mass in $p(D)$ is accumulated on a combination of very short pulses and very long ones, then the feed-forward loop would be selected for
    • Due to its ability to filter out the costly and non-beneficial short pulses
Selection for C1-FFL

• Example:
  – Consider the probability distribution
    \[ p(D) = (1 - \lambda)p_1(D) + \lambda p_2(D) \]
    where the elementary distributions \( p_1(D) \) and \( p_2(D) \) are shown to the right
  – By varying the level of \( \lambda \), we can compute the two average accumulated fitness functions \( \Phi_{str} \) and \( \Phi_{ffl} \)
  – The values of \( \lambda \) for which one is greater than the other then determine when the former would be selected for
    • At the expense of the other
Selection for C1-FFL

\[ b = 1.25 \beta \eta \]
Summary

• Bacterial cells must make optimal use of the resources available to them and maximize their growth rates
  – Otherwise, they lose out to the others who do manage to maximize their growth rates in the same conditions and thus come to dominate the populations
• By this token, among all possible ways of carrying out a required operation, the one that is most suitable in the given environmental conditions are favored
  – Maintaining an optimal protein expression level
  – Regulating gene expression
  – Identifying the optimal regulatory mechanism
• The principle of optimality transcends organisms and species
  – All cells, in single-cellular and multi-cellular organisms are endowed with similar mechanisms to ensure effective use of limited resources
  – This suggests
    • Common lineage
    • Convergent evolution