To see a World in a Grain of Cells: Statistical Auguries of Nonequilibrium

a thesis by
Gabriel L. Salmon
in partial fulfillment of the requirement for the
Degree of Doctor of Philosophy
in Bioengineering

California Institute of Technology
Pasadena, CA, USA

2026
(Defended July 14, 2025)

©2025
all rights reserved
Gabriel L. Salmon
ORCID: 0000-0003-2163-8399

Acknowledgements

The best parts of my life are not written down.

This thesis is no exception.

This page has the worst gap of all between what is, and what is said.
So, *all* of the larger words of this thesis I have written are a searching way to say: **Thank you,** to all of you who have helped make these years a series of astonishments about the natural, mathematical, and joyful world.

To **Rob**, where can I begin? Thank you for the opportunity to think with you. To learn and grow with you. Most ecstatically, to practice a sharpening of wonder at the broader natural world with you. I consider myself a perpetual and thrilled student of Nature 1, and hold extraordinary gratitude for all of the opportunities to reach for dreams you advocate for us. I have much more I can and will say, but for now let me just express from the bottom of my heart and mind, I truly cherish the lessons I continue to learn from you and the am honored by the fellowship and friendship you build so generously. One of your most earnest motto's—"How can I help?"—says it all about your formidable and wonderful earnestness. I am excited to think together (and hone in both mathematical and literal whales together) deep into the future!

I am especially grateful to **Erik Winfree**, **Dianne Newman**, and **Lea Goentoro** as members of my committee for your willingness to articulate, share, and model extremely different—sometimes contradictory!—*ways of knowing*. One of the gifts of learning new mathematics and biology both is that the *form* of an informative or satisfactory answer to a question gains new substance. These new forms of manipulating knowledge are huge psychological catalysts for discovery and inquiry and are just fun to phase shift through. I am not sure whether Dianne is better at reading the mind of *Pseudomonas* or how to make people feel welcome as humans and scientists, even to ask very silly questions. Erik, thank you for constantly and generously advocating (and beautifully demonstrating) the power of thinking with ways of knowing from theoretical computer science and beyond! I am grateful for your introspection about what it conditions for being creative and independent. Lea, I think a lot about both your Systems Biology class those years ago (whose discussions ring in my ears with your characteristic infectious enthusiasm!) and

the times you have taken to sit down with me and go through with a pen in hand a report. Thank you all for your mentorship.

Life and science are most fun in coconspiracy. To **Sara**, thank you for the wide adventures (fesenjoon and all), and for inspiring both such rare fun and depth (capers which only continue). To **Ana**, thank you for your wide fellowship, spontaneous chocolate, dancing and LA explorations, and more. To **Hannah**, thank you for jamming richly together on energy-limited metabolism and your inclusiveness. To **John** and **Fang**, I have learned incredible science and math from you, from dumplings across LA to math on the beach in San Diego to beyond: thank you for making Koalag night and day science bloom. To **Dongyang** for great fun in fertile (and continuing) proofreading jams and friendship.

To **Avi**—thank you for always sharpening our thinking, slinging your killer persimmon babka, delivering zips of much needed gentle satire, and so generously sharing your considerable physiological and further wisdom. To **Hernan**—thank you for generously engaging with science at any notice, giving granular and hugely incisive manuscript feedback, and showing what bold science means.

To **Christian**, and **Charlotte**, for being extraordinary friends and flatmates; what an honor. To **Isabelle**, **Perry**, **Kellan**, for Sichuan-graced excursions and movie nights and profound discussions. To **Jan** for new music and deep philosophy, **Elin** for hikes and food and debate, **Jonathan** and **Jerry** for always energizing discussions and shared meals. To the whole Casa Bike Team, queridos amigos (particularly **Eli** and **Nico**), for showing what building creative and incisive lives means and bringing fun.

To many other friends and labmates of graduate school, including the whole continuum of the inimitable Phillips lab crew (**Griffin**, **Vahe**, **Soichi**, **Manuel**, **Suzy**, **Heun Jin**, **Amin**, **Rosalind**, **Tom**, **Kian**, **Minakshi**, **Doran**, **Rebecca**, **Rachel**, **Scott**, **Muir**, **Niko**) and **many more**, thank you for your guidance, wisdom, and examples.

To so many formative mentors and friends, for giving me wide chances across human years to try this thing called science, showing what creativity means, and teaching so generously. You particularly include: **Madhav Mani**, **Simon DeDeo**, **Sterl Phinney**, **Michael Fink**, **Mohit Verma**, **George Whitesides**, **Marcia Levitus**, **Alexandra Ros**, **Paul Jones**, **Rebecca Whelan**, **Wendy Beth Hyman**, **Dan Styer**, **Steven Fitzgerald**, **Yumi Ijiri**, **Jason Stalnaker**, **Rob Owen**, **Manish Mehta**, **Bradley Hann**, **Kevin Bennett**, **Ed Moore**, and **Brett Loucks**.

Last, to **my family**. Thank you for all of it. Every cactus you have taken me to visit, every thorn you helped pluck from my exploring feet, every amused return to that same cactus. Every story about immune cells going "Awoogah!," healing those thorns of desert and of life. **Naomi**, everyday I learn more from how you invent joy, notice generously, think inclusively, rebel to advance justice, value story. **Mother**, thank you for your fierceness and advocacy, care packages defying postal physics, deep and wise love; you teach me across all scales. **Dad**, in your love and example, among so much else, I learn from you what both true craft and care mean: human spirit built from 4 AM wakeups to care for strangers, noticing each heron and lichen, dry wit and deep support.

Family-mine, both given and chosen, I hope to be and act like you amazing souls someday. I will always be working on it.

Abstract

Precisely what new mathematical and biological behaviors are unlocked as cells—and their collectives—operate out of equilibrium? In this thesis, we take simple steps towards accounting for the characteristic scales of biological energy expenditures and tracing their functional destinies.

After embarking on an initial *Census of lively dissipation* (Chapter 1) to set the stage, Part I explores cells on the bioenergetic brink of survival and death. In Chapter 2 we explore the sweep of the smallest measured cellular metabolic rates; reflect on their empirical and philosophical underpinnings; and endeavor an initial order of magnitude accounting of the biophysical processes that could most plausibly dominate the energy budgets of starving cells. Followed by this tour, Chapter 3 hones in on fresh, provocative experiments probing anaerobic maintenance metabolism of *Pseudomonas*, proposing simple math to predict how microbes die in balance with their energetic resources and interpreting measured reactions to salt.

Part II travels to intracellular gene regulation. Among other regulatory settings, in Chapter 4 we investigate the most common regulated architecture of gene regulation in prokaryotes and ask how its behaviors categorically change under different investments in biochemical drive. Chapter 5 narrates how graph theory gives powerful tools for humanly thinking about even large regulatory state graphs.

In Part III, we work closely with inspiring new experimental collaborations and data on patterns of microtubules made by molecular motors. Chapter 6 reports on measuring biochemical energy profiles, and their rates of change, over time and space. Using this phenomenology, in Chapter 7 we develop theory to ask about the fundamental costs required to build or maintain biochemical gradients.

Last, in Chapter 8 (Part IV) we ask how a common mathematics might unify a large class of biological dynamics with disperse initial conditions but coherent final conditions, a setting we refer to as *exploratory dynamics*.

Collectively, we hope these case studies give quantitative glimpses of precisely how energy so exquisitely animates biology.

Published Content and Contributions

- S. Mahdavi*, **G. Salmon***, P. Daghlian, H. G. Garcia, and R. Phillips. "Flexibility and sensitivity in gene regulation out of equilibrium," *PNAS*, **October 2024**. https://doi.org/10.1101/2023.04.11.536490. (*equal contribution)
  (GS co-initiated and co-conducted research and analysis; GS wrote the final paper.)
- J. Kondev, M. Kirschner, H. G. Garcia, **G. Salmon**, & R. Phillips. "Biological Processes as Exploratory Dynamics," in submission to *Biophysical Journal*, **June 2025**. http://arxiv.org/abs/2506.04104
  (GS co-conducted research and cowrote the paper.)
- **G. Salmon**, H. Jeckel, A. Flamholz, D. Newman, …, R. Phillips. "Accounting for the plasticity and major drivers of minimal cellular metabolism," *in preparation*.
  (GS conducted research, analysis, and wrote the manuscript with coauthors.)
- A. Duarte, **G. Salmon**, HJ Lee, B. Najma, R. Banks, S. Hirokawa, M. Ashok, …, and R. Phillips. "Resolving the spatiotemporal power demand to sustain ordered motor-microtubule assemblies," *in preparation*.
  (GS co-conducted research, performed data analysis, and wrote the manuscript with coauthors.)

Contents

Acknowledgements iii

Abstract v

Published Content and Contributions vi

Table of Contents vii

List of Illustrations x

1 Life is widely—and plastically—dissipative 1
1.1 A lively census of dissipation . . . 1
1.2 Mathematical notions of (non)equilibrium . . . 6
1.2.1 Implications of this dissipation for theory . . . 11

I Survival out of equilibrium 19

2 Accounting for the plasticity and major drivers of minimal cellular metabolism 20
2.1 Introduction . . . 21
2.1.1 Overview . . . 21
2.2 Experimental measures and philosophical notions of basal cellular metabolism 23
2.3 Plausible Energy Sinks in the Cell . . . 35
2.3.1 The Cost of Protein Turnover . . . 35
2.3.2 Cost of post-translational protein modifications . . . 45
2.3.3 Cost of sustaining vital membrane potential and osmobalance . . . 55
2.3.4 Experimental inference of ion pumping costs . . . 59
2.3.5 Cost of cellular motility . . . 69
2.3.6 Cost of membrane synthesis . . . 72
2.4 Minimal maintenance power from empirical data in ocean sediment . . . 73
2.5 Outlook . . . 78
2.6 Acknowledgments . . . 79

3 Predictive theory for experiments on microbes on the energetic brink 84
3.1 Phenomenology of death differs under terminal e^- acceptor or donor deprivation . . . 86
3.1.1 Generalizing energetic resource balance with cellular energy consumption under broader dynamics . . . 89
3.2 Explicit kinetic theories for maintenance experiments . . . 103
3.2.1 Model C: Zeroth-order cells fed by first-order electrodes . . . 107
3.2.2 Model D: First-order cells fed by first-order electrodes . . . 109

3.2.3 Models A and B: Zeroth-order electrodes feed cells . . . 112
3.3 Quantitative glimpses of metabolic adaptation to ionic stress in *Pseudomonas* 115

II Genes and networks out of equilibrium 126

4 Flexibility and sensitivity in gene regulation out of equilibrium 127
4.1 Introduction . . . 128
4.2 Results . . . 132
4.2.1 Modeling pervasive gene regulatory motifs . . . 132
4.2.2 Nonequilibrium steady-state output responses . . . 134
4.2.3 Equilibrium output curves are constrained . . . 137
4.2.4 New regulatory shape phenotypes unlocked by nonequilibrium . . . 138
4.2.5 Leakiness, saturation, and EC50 are tunable at equilibrium . . . 140
4.2.6 Nonequilibrium control of sensitivity obeys shape-dependent global bounds . . . 141
4.2.7 Breaking detailed balance along each edge . . . 144
4.2.8 Breaking detailed balance two edges at a time . . . 147
4.2.9 Generic rate conditions forbid access to nonmonotonic responses . 149
4.2.10 More complex regulatory architectures obey similar constraints . . 151
4.3 Discussion . . . 154
4.3.1 Nonequilibrium steady-state probabilities via the Matrix Tree Theorem 159
4.3.2 Emergent shape parameters & shape phenotypes . . . 160
4.3.3 Single edge and edge pair perturbations . . . 161
4.4 Data & Availability . . . 161

5 Understanding larger graphs 167
5.1 Summary of nonequilibrium generalization of important transcriptional motifs . . . 168
5.2 Census of known regulatory interactions in *E. coli* . . . 170
5.2.1 Natural statistics of regulatory architectures in *E. coli* . . . 177
5.3 Incidence of known transcriptional architectures in Drosophila from Redfly 183
5.4 Understanding the hypercubic graph: regulation by N transcription factors 190
5.4.1 Illustration with the 3−cube case . . . 194
5.4.2 Common graphs formed by subsets of the hypercube . . . 209
5.4.3 All nontrivial, transcriptionally-potent subgraphs of the $n = 4$ hypercube . . . 214

III Patterns out of equilibrium 220

6 Accounting for the fundamental dissipative costs of gradient formations across space and time in motor-microtubule assemblies 221
6.1 Overview . . . 222
6.2 Introduction . . . 223
6.3 ATP Concentrations in Space and Time . . . 226

7 Power as a localized field in space and time 235
7.0.1 Power to Maintain a Motor Gradient . . . 235
7.0.2 Power Stored in a Three-Dimensional Gradient . . . 240

7.0.3 How much power does it take to *build* a gradient compared to that to *maintain* it? . . . 246

IV Exploration out of equilibrium **277**

8 Biological Processes as Exploratory Dynamics **278**
8.1 Theories of Inanimate and Animate Dynamics . . . 279
8.2 Exploratory Dynamics by Variation and Selection . . . 285
8.3 Zero-Dimensional Exploratory Dynamics . . . 289
8.4 Chromosome Search Via Microtubule Reset Dynamics . . . 294
8.4.1 One-Dimensional Chromosome Capture by Exploratory Dynamics 294
8.4.2 Three-Dimensional Chromosome Capture by Exploratory Dynamics 300
8.4.3 Molecular Control of Exploratory Dynamics . . . 306
8.4.4 Exploratory Power . . . 311
8.5 Transcription Factor Search for Binding Sites as Exploratory Dynamics . . . 318
8.6 Discussion . . . 331

List of Figures

1.1 Comparison of the power densities of inanimate matter and living matter. 3



1.2 Power densities across contexts, cell states, and different organisms. 4



1.3 States of a system represented as a network (a finite state, continuous time Markov chain). 7



1.4 Steady-states in and out of equilibrium. 9



1.5 The Boltzmann law for computing the equilibrium probabilities of different microstates. 11



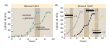
1.6 Gene expression for a great diversity of different situations all fall onto one master curve. 12



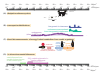
1.7 The superficially artificial nature of the distinction between equilibrium and nonequilibrium. 15



2.1 Choice of carbon source in the lives of bacterial cells. 22



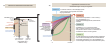
2.2 Extraordinary metabolic plasticity in microbes, as reported by different experimental means. 24



2.3 Continuous culture chemostats give correlations between bulk carbon consumption rates and fast growth rates. 26



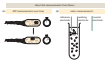
2.4 Conceptual subtleties of inferring basal metabolism using extrapolation from fast growth measurements in a chemostat. 27



2.5 Inferring maintenance rate from death rates under carbon limitation. 30



2.6 Inferring cellular energy consumption by (A) direct measurement of ATP consumption or (B) direct measurement of electron transport by electrode exchanges with extracellular electron carriers such as phenazines (including phenazine-1-carboxamide (PCN)). 31



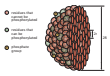
2.7 Processes that make up microbial energy budgets. 36



2.8 Degradation and active resynthesis of proteins enacts a steady-state with characteristic turnover rates and accompanying costs. 38



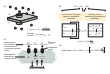
2.9 Phosphorylation of the amino acids on the surface of a protein. 46



2.10 Nonequilibrium steady state for phosphorylation-dephosphorylation cycle. 49



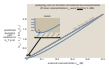
2.11 Maintaining gradients of ions in a cell requires active energy expenditure lower-bounded by a simple calculation. 56



2.12 Rate at which *E. coli* populations die (aerobically under carbon limitation), as a function of the concentration of sodium chloride in the external medium. 60



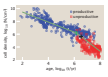
2.13 Predicted cost of pumping ions as function of the external medium concentration c_{out} up to 500 mM, assuming that cells fix cytosolic concentrations to c_{in} at a value of 5 mM, according to Eq. 2.70. 68



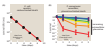
2.14 Fraction of organic carbon in sediment versus carbon age, as reported by deep marine sediment samples. 76



2.15 Cellular abundance in deep marine sediment versus local carbon age. 76



3.1 Empirical death curves for cells subjected to different deprivations. 86



3.2 Distinct macroscopic physiology of death under deprivation of electron donor (e.g. carbon) versus deprivation of terminal electron acceptor (e.g. oxygen). 88



3.3 Diverse environmental and laboratory scenarios where energetic resources $E(t)$ are replenished at a supply rate s adopting the same linear mathematical form, $s(E(t)) = k(F - E(t))$, for some appropriate constants k, F. 91



3.4 Model proposed to relate the rate of change dN/dt of the abundance of cells to the state variables of the current energy $E(t)$ and the current abundance of cells $N(t)$. 96



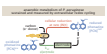
3.5 Remaining abundance of cells at late times (at $t = 5$ days), as a function of the extracellular phenazine concentration, extracted from the data depicted earlier in Fig. 3.2(B). 101



3.6 Measured survival data compared to the model of Eq. 3.19. 102



3.7 Schematic for how cellular metabolism at rate $\beta N(t)$, driving reduction of extracellular phenazine, is counteracted by oxidation by the electrode at rate k_{el}. 104



3.8 Presiding differential equations for the functional energy source. 106



3.9 Mathematical properties of Model (C) (as portended by Fig. 3.8). 107

3.10 Mathematical properties of Model (D) (as portended by Fig. 3.8). 111

3.11 Model (A) (as portended by Fig. 3.8). 113

3.12 Model (B) (as portended by Fig. 3.8). 114

3.13 Impact of salt and external phenazine concentration on cellular abundance and current over time. 117

3.14 Impact of salt and external phenazine concentration on per-cell metabolic rate over time. 119

3.15 The relationships between metabolic rate per cell to phenazine and sodium chloride changes over time. 120

3.16 Metabolic rate per cell versus external phenazine concentrations, at increasing sodium chloride levels (indicated by purple, blue, teal, and orange lines, respectively), over time. 121

3.17 Metabolic rate per cell versus external sodium chloride concentrations, at different phenazine levels (indicated by blue, pink, and purple lines), over time. 122

3.18 Representative subset of data, with early to intermediate timepoints ($t \leq 48$ hr) displayed, for relatively mild salt concentrations ([NaCl] ≤ 400 mM) for a modest, physiologically-well-tolerated phenazine concentration. 123

4.1 Structure of a fundamental gene-regulatory motif. 130

4.2 Incidence of regulatory architectures for transcription, where a square cycle's response is common. 131

4.3 Nonequilibrium response of the four-state fundamental transcriptional motif. 135

4.4 Global bounds, in or out of equilibrium, restrict maximal (normalized) response sensitivity (with respect to input concentrations $[X]$ on a log scale). 142

4.5 Systematically breaking detailed balance edge-by-edge. 145

4.6 Breaking detailed balance along two edges unlocks higher sensitivity and multiply-inflected outputs with smaller drive than required for breaking detailed balance along single edges. 153

5.1 Fresh ("hairball") visualization of the bipartite regulatory network of transcription factors and promoters in *E. coli*. 171

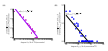
5.2 Hairball visualization of the bipartite regulatory network of transcription factors and promoters in *E. coli.*, partitioned by regulatory type. 173



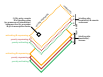
5.3 *E. coli*'s regulatory network visualized as a biadjacency matrix. 176



5.4 The two bipartite degree distributions in *E. coli* are acceptably well described by power laws. 177



5.5 Scheme for counting regulatory interactions. 178



5.6 Schematic for counting convention over architectures. 179



5.7 Summary of regulatory architectures in *E. coli*. 181



5.8 Separate experimental survey of regulatory architectures in *E. coli*. 182



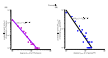
5.9 Hairball network visualization of annotated regulation in the fly. 184



5.10 The fruit fly's regulatory network visualized as a biadjacency matrix. 186



5.11 Power laws in the regulatory network of the fly. 187



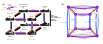
5.12 Summary of incidence of regulatory architectures in the fly. 188



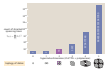
5.13 Summary of regulatory incidence in the fly. 189



5.14 Emergence of hypercubic state spaces from multiple transcription factors, in the case $n = (2 + 1)$. 191



5.15 Superexponential explosion of the number in rooted spanning trees with respect to the hypercube dimension. 193



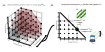
5.16 All directed spanning trees that are rooted in state S, computed by symbolic determinants, on the $n = 3$ cube graph of states. 195



5.17 Examples of spanning trees rooted in the cubic graph, organized by effector powers. 197



5.18 Powers attained by the transcription factors x_1 and x_2, in addition to the polymerase p, over all spanning trees in the cubic $n = 3$ graph. 200



5.19 Visualization of a hyperplane bounding permitted effector powers in trees setting structure of responses. 202



5.20 Adding a third transcription factor X_3 forms a hypercubic state space in $n = 4$ dimensions, known as a tesseract, represented in a concentric layout in panel (A). 204



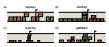
5.21 Allowed powers in transcriptional response of a tesseract graph. 205



5.22 Allowed powers in transcriptional response of an arbitrary hypercubic graph. 208



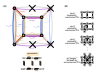
5.23 Four examples showing how *E. coli* promoters regulated by multiple transcription factors often present extensive overlaps between binding sites. 209



5.24 Regulatory motif with two transcription factors, whose binding is mutually exclusive. 210



5.25 Relation between the cubic graph and the "book graph" and a generalisation of this "book graph" in m dimensions corresponding to the binding of m exclusive transcription factors. 211



5.26 Structures of all subset graphs of the tesseract cube formed by all possible exclusions. 216



5.27 Allowed powers in responses of all subset graphs of the tesseract cube formed by all possible exclusions. 217



6.1 Schematic of the experimental system used to measure spatiotemporal evolution of ATP. 225



6.2 Experimental measurements and Finite Element simulations of ATP concentration in space and time. 227



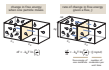
6.3 The power of aster formation. 228



6.4 Mechanistic processes taking place during aster formation and their estimated power. 231



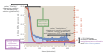
7.1 The change in free energy when a particle is transported in a gradient. 236



7.2 Dissipation in a spherically symmetric concentration gradient. 241



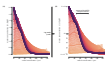
7.3 Experimental data motivating the question of how much power it takes to build versus maintain an aster's concentration gradients. 246



7.4 The dynamics of building a gradient. 254



7.5 Example motor distributions over space, angularly averaged to show variation of motor concentration over radial coordinate r. 255



7.6 Example of the dissipative behaviors of a steepening (number-preserving) exponential gradient in time, where the characteristic length scale $\lambda(t)$ decreases as a linear ramp in time, $\lambda(t) = \lambda_0 - \gamma t$. 259



7.7 Contrasting the locations in space where distinct dissipative mechanisms are expected to be maximal. 263



7.8 Visualization of the distinct spatial localizations of maintenance power (blue, lower left); building power (red, top); and concentration (green, lower right). 267



7.9 Temporal changes in apparent length scale of motor gradients of the data in Figure 7.5. 274



8.1 Gallery of dynamical equations. 280



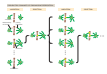
8.2 Comparison of human-scale dynamics and cellular-scale dynamics approaches to building a railroad between New York City and Boston. 284



8.3 Gallery of exploratory dynamics. 286



8.4 Microtubule search for kinetochores as exploratory dynamics characterized by stochastic variation in the dynamical trajectories of the system coupled with selection and stabilization of those trajectories that culminate in attachment to kinetochores. 288



8.5 Schematic of the process of microtubules searching for chromosomes. 290



8.6 Trajectories and weights for one-dimensional model of chromosome search and capture. 297



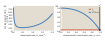
8.7 Search time as a function of the catastrophe rate for the one-dimensional model of chromosome search and capture. 300



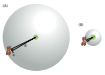
8.8 Trajectories and weights for three-dimensional model of chromosome search and capture. 302



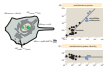
8.9 Optimal search time as a function of the catastrophe rate. 305



8.10 The geometry of chromosome search and cell size. 307



8.11 The power of biological processes. 312



8.12 Summary of the power required for various cellular processes per volume. 317



8.13 Searching for a binding site on the chromosome. 319



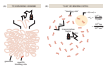
8.14 Simple model of combined 1D and 3D transcription factor search for a specific binding site. 320



8.15 Exploratory dynamics of transcription factor search as an example of variation and selection. 321



8.16 Estimating time scales of encounters between a transcription factor and the genome. 328



8.17 Time for a transcription factor to find its binding site in a model of combined 1D and 3D search. 329



Chapter 1

Life is widely—and plastically—dissipative

When alive, cells spend biochemical energy. How much? What are its mathematical consequences? What are its biological consequences? In this salvo chapter, we briskly pose these questions in turn.

1.1 A lively census of dissipation

Across contexts, cell states, and organisms, the rates that cells and organisms consume energy vary dramatically. The useful free energy processed and dissipated by cells adopts many forms—as ATP gradients, nutrient transformations, ion gradients, fundamental losses as heat, and beyond. This diversity gives way to useful insight when cellular energy expenditures are compared in common thermodynamic units of energetic equivalents expended per time. To think about these scales, consider the power spent per volume across living matter, a spectrum glimpsed in Figure 1.1. From energy-limited microbes to the flight muscle cells of the honey bee, the power densities of cells vary more than an

incredible seven orders of magnitude, as we compute and show in Figure 1.1(B-C). How do such lively metabolisms compare to their inorganic counterparts? Figure 1.1(A) gives a comparative visual answer to that question by familiar examples from the world of everyday macroscopic experience. Some of the examples shown there are drawn from our astronomical neighbors, including the Sun, joined by technologies developed by humans.[1] The most powerful states of living matter, including exponentially growing microbes like *E. coli*, are so dissipative as to be within a factor of a few hundred of the power density inside nuclear reactor cores. On the slow end, environmental microbes in ocean sediments consume power over a million times slower by volume, representing an extraordinary dissipative plasticity.

In the spirit of the Bionumbers database [1], we aim to take stock of the facts and data that have been amassed to describe the energies and powers associated with living matter. We give an even fuller census in the plot of Figure 1.2, which also illustrates how widely certain cell types tend to change their own dissipative status by exertion.[2] There, we also highlight a new experimental measurement, a basal metabolic rate of the opportunistic pathogen *P. aeuriginosa* reported by Ciemniecki *et al.* [2], whose value and attendant questions so intrigue us that we engage with it across many chapters of this thesis.

Understanding why the fastest-growing cells spend the amount of energy they do rewards us with useful rules of thumb that organizes our appreciation of the whole ladder of dissipations visible in both Figs. 1.1 and 1.2. Consider exponentially growing *E. coli*, with a typical doubling time of $\tau \sim 20 - 30$ min (or few $\times 10^3$ s). If each cell to be duplicated has order $N \sim$ few $\times 10^6$ proteins; each protein is composed of $\ell \sim$ few $\times 10^2$ amino acids; and the cost of each such amino acid incorporation is $\eta \sim 4$ ATP equivalents,

[1] Note that although for millennia humans have had control of fire, we are of the mind that the capacity to make things cold is even more impressive; both heating and cooling technologies are shown in Fig. 1.1(A).

[2] Note that in Fig. 1.2, we omitted the slowest metabolizing ocean microbes earlier depicted in Fig. 1.1.

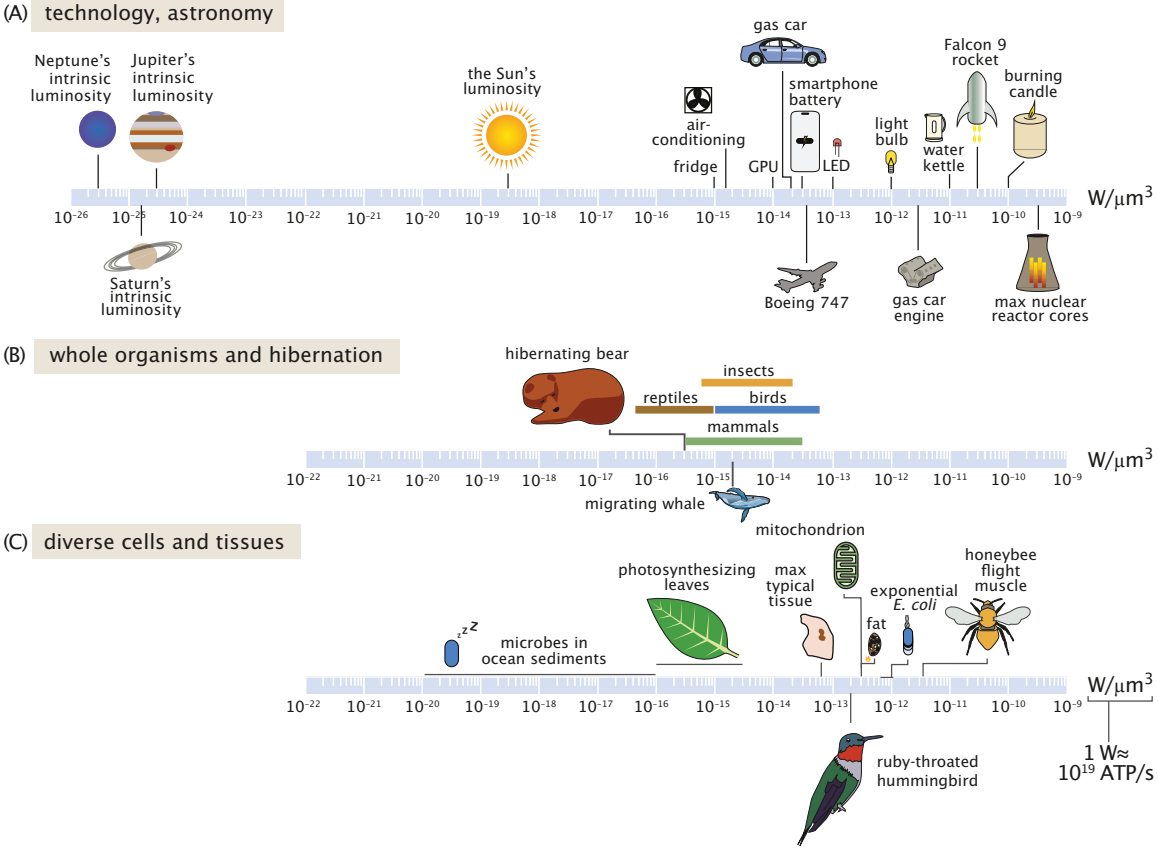


Figure 1.1: Comparison of the power densities of inanimate matter and living matter. The power density of organisms, and their plasticity, is impressive even on astronomical terms. (A) Astronomical, technological, and abiotic sources with familiar dissipations. We converted stellar data from NASA to power densities; other quantities were computed from first principles or manufacturer information. (B) Power densities at the organismal scale, including the fascinating process of hibernation. (C) Power densities at the cellular and tissue scales.

then just the cost of duplicating the proteome into a new daughter cell is

$$\text{power to duplicate proteome: } \frac{N\ell\eta}{\tau} \tag{1.1}$$

$$\approx \frac{(\text{few} \times 10^6 \text{ proteins}) \left(\frac{\text{few}\times 10^2 \text{ aa}}{\text{protein}}\right) \left(\frac{\text{few ATP}}{\text{aa}}\right)}{\text{few} \times 10^3 \ s} \tag{1.2}$$

$$\sim 10^6 \text{ ATP/s/cell.} \tag{1.3}$$

This simple calculation also invites us to recall the extremely useful scale that one ATP

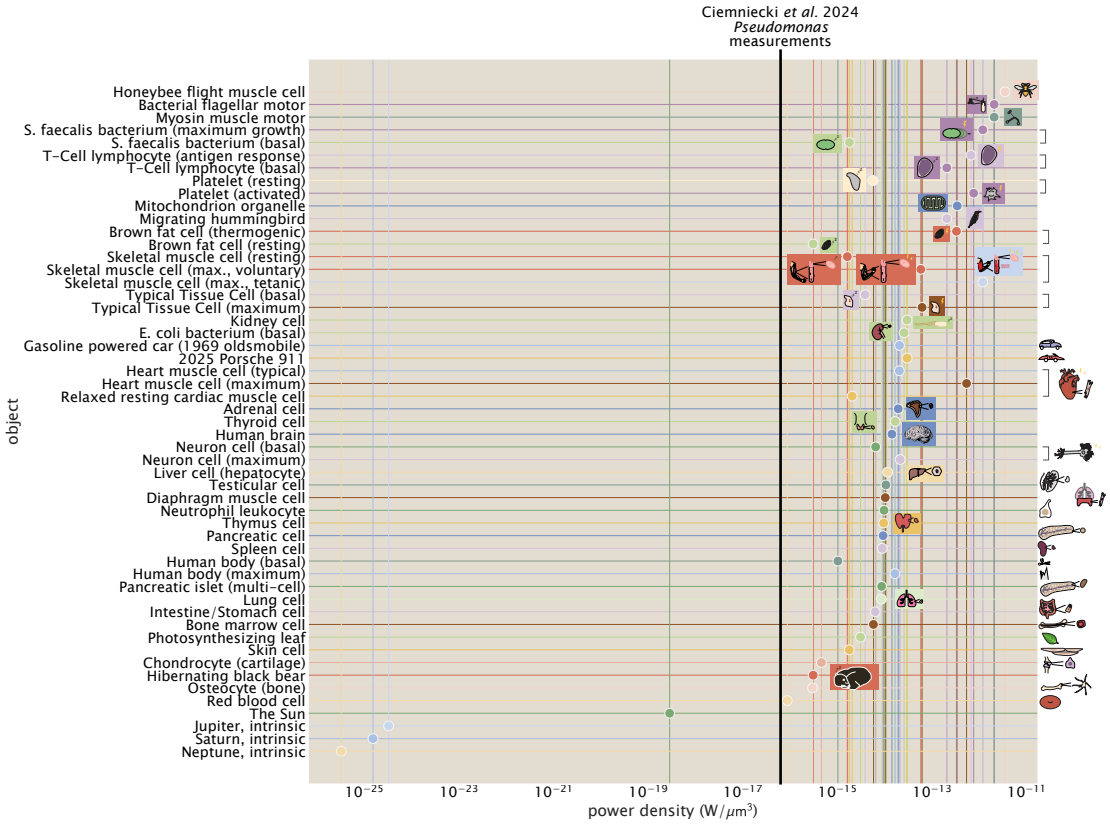


Figure 1.2: Power densities across contexts, cell states, and different organisms. This representation given in units of power/volume (on a log scale) shows that the power densities of organisms and cells vary over many orders of magnitude.

hydrolysis's worth of energy typically liberates approximately 20 k_BT worth of energy at temperature T, where k_B is Boltzmann's constant (Bionumbers ID (BNID) 103385 [1]); at typical temperatures $T \sim 300\ K$ found comfortable by an array of life, this implies that 1 Watt $\approx 10^{19}$ ATP/s. Accordingly, that power *E. coli* pays to duplicate its proteome is also expressible as $\sim 10^{-13}\ W/\mu m^3$ in its one cubic micron volume. Consulting Fig. 1.1(C), we recall that the total metabolic rate of exponential *E. coli* is measured as $\sim$ few $\times\ 10^6$ ATP/s/μm^3 to 10^7 ATP/s/μm^3 (or, equivalently, few $\times\ 10^{-13}$ W/μm^3 to 10^{-12} W/μm^3). In other words, duplicating proteins is so expensive that it is within a factor of few to ten of the total power budget of rapidly growing *E. coli* cells. Protein duplication appears to explain much, though not all, of the scale for why fast *E. coli* cells

consume the energy that they do [3].

What about the remaining, non-protein-duplicating, dissipation measured in fast growing cells? Does the remainder of the power budget have obvious (and precisely accounted for) biophysical destinies? Intriguingly, carefully tracking known expenditures (towards other accumulation of biomass and known biochemistry) does not always fully account for the measured total metabolism. Such outstanding energy expenditures were suggested back in the 1980s [4], but more fastidious modern accounting still ratifies "missing" fractions of the energy budget in common examples of aerobic metabolism. Writing in their 2023 study deploying detailed flux balance analysis modeling, and fully invoking sophisticated modern insights about physiology and constraints, Mori, Hwa, and coworkers write, "Surprisingly, the ATP generated during the biosynthesis of building blocks from glucose almost balances the demand from protein synthesis, the largest energy expenditure known for growing cells. This leaves the bulk of the energy generated by fermentation and respiration unaccounted for" [5]. One practical way these unexplained energy expenditures plague our field's collective quantitative understanding is the need for flux balance analysis to invoke *ad hoc* multiplicative inefficiency or mechanistically-unmotivated "maintenance" factors in order to match total metabolic rates [6, 7].

Not only do fast-growing cells spend large (sometimes only partly explained) fluxes of ATP, they also sustain large standing pools of ATP. Interestingly, across many cell types from bacteria to human muscle, brain, and retina, the steady-state concentration of ATP is consistently on the scale of a few millimolar in concentration, varying by only a factor of a few in typical homeostatic conditions [8]. This is a big standing pool, namely a few million molecules per cubic micron. While it is of course ultimately the instantaneous ratio of ATP to ADP and phosphate levels (not merely the total ATP concentration) that sets the energetic value of an ATP hydrolysis event, this interesting abundance further hints at a

broad standing capacity for nonequilibrium operations in many cells.

Regarded collectively, these numbers loudly announce some irresistible and important mystery stories. Towards what rich purposes could these plausible, hidden biological dissipations be directed? Which of those purposes can *only* be accomplished out of equilibrium, and which are reproducible without spending energy? What empirically accessible signatures would help us distinguish between equilibrium and nonequilibrium operations in biological networks? If, mathematically and physiologically, a function *could* be performed without spending as much energy as biology might do, what abstractions better replace a brittle analogy of "efficiency," and more expansively accommodate the high-dimensional priorities that really shape an organism? What are the limits below and above which functions underlying life is impossible? These are questions ripe for human wonder.

While we built initial intuition and met questions by appealing to the high dissipation end of the biological dissipative ladders, we can—and rather, must—learn from the fullness of the dissipative spectrum. The high end of the spectrum invites marveling at the regimes and mechanisms of extraordinary biomass accumulation, high-flux component turnovers, and the buzz of honeybee wings. The low end of the spectrum, in the great "struggle for existence," promises to teach us about the true boundaries between living and nonliving matter. In between lies principles of physiological adaptation that beckon discovery.

1.2 Mathematical notions of (non)equilibrium

Having confronted just how out of equilibrium biology often is, we now recall what it means mathematically for a system to be in or out of equilibrium. To set the stage for this discussion, imagine a system of interest like a cell, or a small subset of molecules inside the cell, that can be found in one of a finite number of states, as shown in Fig. 1.3(A). If the

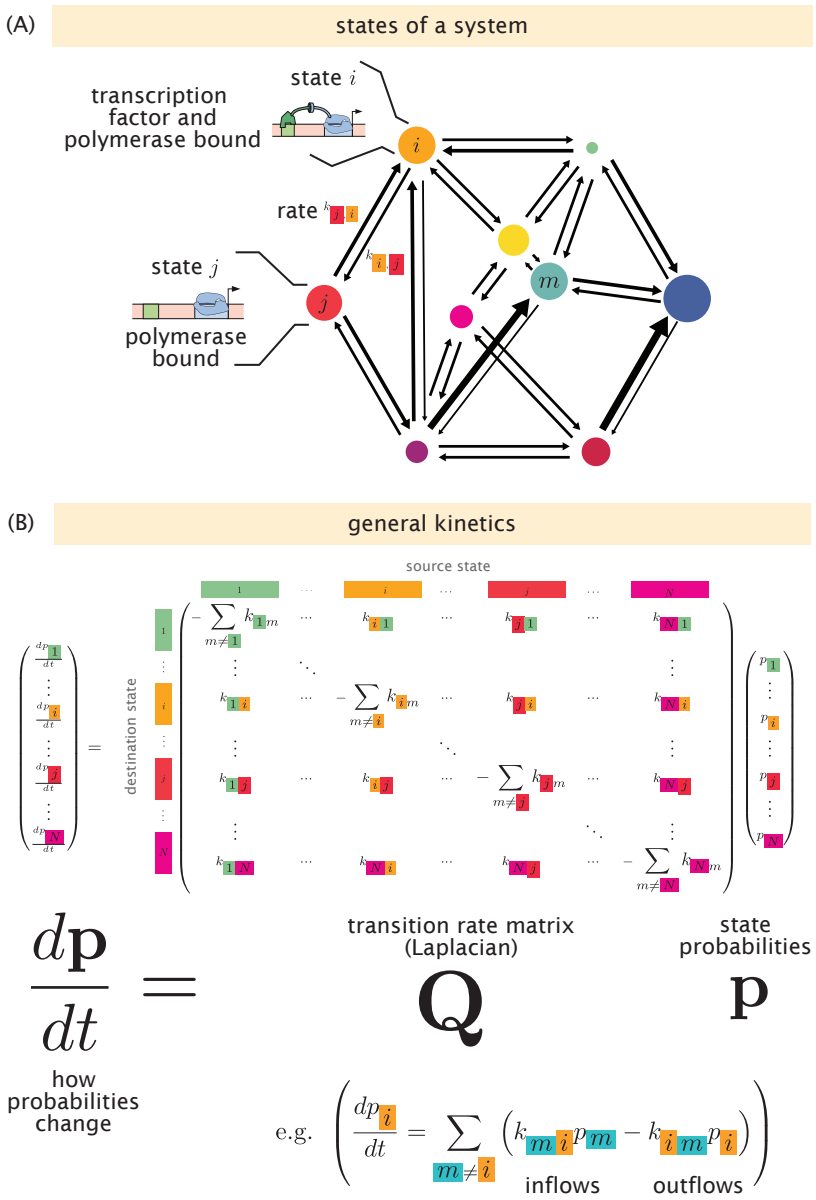


Figure 1.3: States of a system represented as a network (a finite state, continuous time Markov chain). (A) States of a cell, genome, or set of molecules schematized as a network or graph. Transitions between any pair of eligible connected states i and j occur with characteristic rate k_{ij}. (B) The probabilities $\mathbf{p}(t)$ of being found in any of the states change depending on the balance of inflows and outflows to each state, which can be mathematically summarized by the matrix equation $\frac{d\mathbf{p}}{dt} = \mathbf{Q}\mathbf{p}$, where the transition rates k_{ij} populate the entries of the transition rate (Laplacian) matrix $\mathbf{Q}$.

system transitions from state i to j at a characteristic rate k_{ij}, then the probability $p_i(t)$ of being in state i at time t changes depending on the balance of inflows of probability into

state i from other states $m \neq i$, and vice versa. That is,

$$\frac{dp_i}{dt} = \sum_{m \neq i} \left(k_{mi} p_m - k_{im} p_i \right), \tag{1.4}$$

as shown in Fig. 1.3(B). Equivalently, these rules for how the probabilities of all the states change can be tidily written as a matrix equation by squeezing the probabilities of every state into a vector $\mathbf{p} \equiv [p_1, \ldots p_N]^\top$ and the transition rates into a matrix $\mathbf{Q}$ whose i, jth entry represents the transition rate from state j to state i (and the diagonal entries of $\mathbf{Q}$ are set to ensure that probability is conserved to sum to one).[3] Or, all together (as shown in Fig. 1.3(B)),

$$\frac{d\mathbf{p}}{dt} = \mathbf{Q}\mathbf{p}. \tag{1.5}$$

This expression is also known as the Master Equation for the stochastic trajectory of the system. If the system's set of states is sufficiently well behaved (including, for instance, the requirement that every state i is reachable by some path from any other state j), then then there is a probability vector $\mathbf{p}_{\text{ss}}$ over states for which the the probability will not change, $\left.\frac{d\mathbf{p}}{dt}\right|_{\mathbf{p}_{\text{ss}}} = \mathbf{0}$. In general, all that the steady-state condition enforces is that, for all states $i \in [1, \ldots, N]$, the net inflow and net outflow of probability mass for every state balances each other (as schematized by Fig. 1.4(A)). This condition imposes N algebraic constraints on the behavior of the system.

[3] This convention, where the indices of k_{ij} are transposed to be stored in Q_{ji}, may seem artificial, but this is just cosmetic. When considering the transition from state i to j, the columns of $\mathbf{Q}$ (containing i) are regarded as the source states, and the rows of $\mathbf{Q}$ (containing j) are the destination states.

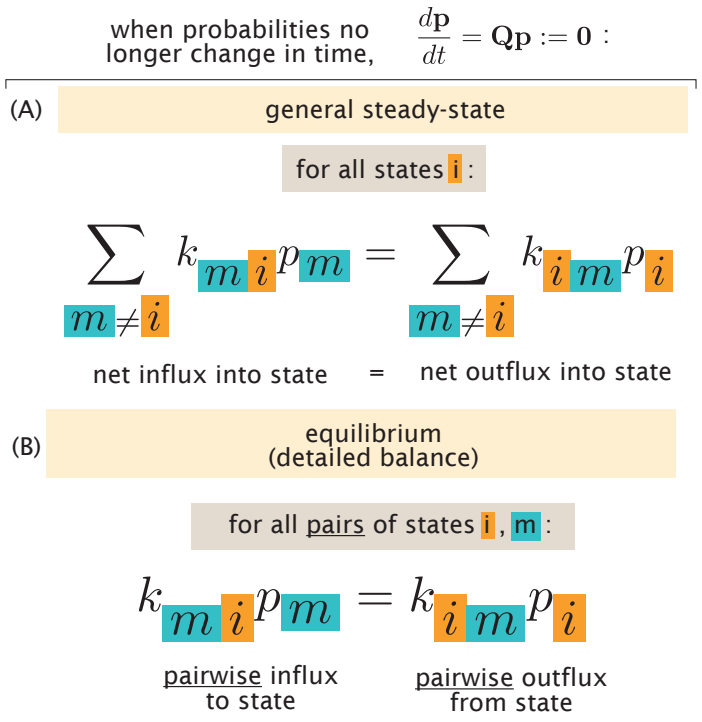


Figure 1.4: Steady-states in and out of equilibrium. (A) Conditions on the elements of the transition matrix and probabilities for a system that has reached a nonequilibrium steady state. (B) The special steady state known as equilibrium demands specific conditions on the transition rates and probabilities.

However, one very special way of achieving this steady-state condition is by a more granular constraint, considered on every pair of states. Specifically, instead of merely requiring that the *net* inflow and net outflow balance each other for each state, as in Fig. 1.4(A), consider instead the scenario where each and every *pair* of states enjoys a mutual balance of probability flow. That is, if,

$$\text{detailed balance (equilibrium): } \text{flux}_{i\to j} = k_{ij}p_i = k_{ji}p_j = \text{flux}_{j\to i}, \tag{1.6}$$

then the system is referred to as being in detailed balance. Shown in Fig. 1.4(B), this condition is exactly the condition of thermodynamic equilibrium (when the system is otherwise closed to material inputs and Markovian). Crucially, for many state network topologies, steady-state distributions can be reached without satisfying this much more restrictive constraint of local detailed balance enforcing equilibrium. Notice that there are a greater number, $N(N-1)/2$, of constraints imposed on the system compared to merely the steady-state condition. This mathematical framing concretely grounds our more empirical and philosophical allusions to dissipation in the preceding discussion to a specific, underlying gulf between defined mathematical universes.

An incredibly powerful consequence of the constraint of detailed balance is given by Boltzmann's equation, which assures that at equilibrium, the probability of a state i with an energy parameter E_i is proportional to $\exp[-E_i/k_BT]$. This relationship, whose form facilitates a huge fraction of successful statistical mechanical calculations, is so formidable and useful in calculations because it gives a *local* rule for predicting the relative probability of a system being in a state. This idea is depicted in Fig. 1.5. However, Boltzmann's equation only applies to systems in equilibrium. One theme we value and pursue aggressively with this thesis is to explore candidate "cognitive technologies" that could adopt the guiding navigational role that Boltzmann's equation gives, but for

nonequilibrium domains.

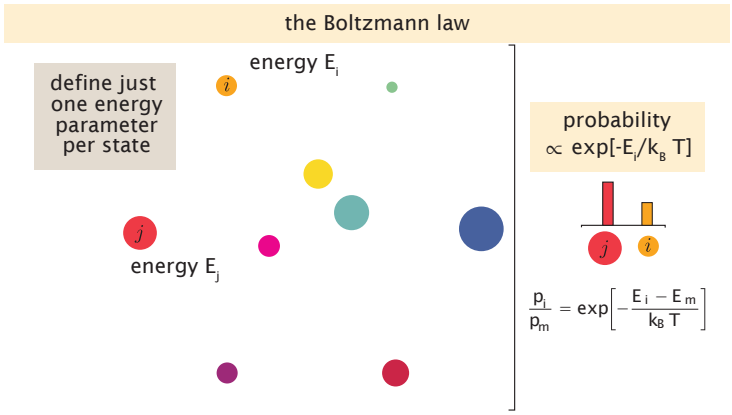


Figure 1.5: The Boltzmann law for computing the equilibrium probabilities of different microstates.

1.2.1 Implications of this dissipation for theory

Although the primordial mathematical differences between equilibrium and nonequilibrium behaviors are quickly stated (viz., the statements of Fig. 1.4), their actual consequences for the attainable complexity of biological functions of interest is a much richer, more interesting, and more subtle question. This is due to at least two surprising facts.

First, purely equilibrium systems are themselves capable of surprisingly complex and interesting behaviors, especially if they are large enough (and given enough time). This capability essentially underlies the success of sufficiently large Boltzmann machines to reproduce arbitrary probability distributions (probabilistic models by definition instantiated to be at equilibrium). Further, recently a body of stimulating work illustrates just how capable equilibrium self-assembly can be, in a formal computational sense [9].

Second, nonequilibrium systems are regularly capable of looking nearly equilibrium. This is a theme we now explore in two evocative vignettes.

A thermodynamic irony: the curious success of equilibrium theory.

In a famous article in the early 1960s, Eugene Wigner talked about the "unreasonable effectiveness of mathematics in the natural sciences." That justly famous essay highlights the repeated successes, both big and small, of recasting our understanding of the natural world in mathematical form. There are similarly curious "effectivenesses" to be celebrated in the natural sciences. One of the most remarkable aspects of many quantitative dissections of biological phenomena is the power of equilibrium thinking, one of the most unreasonably effective theories in all of science — with "equilibrium" even serving as the primary tool used to understand the degree of ionization of atoms in stars —- a manifestly out of equilibrium situation. A compelling biological example of this effectiveness is given in Figure 1.6 in the context of gene expression, as we now describe.

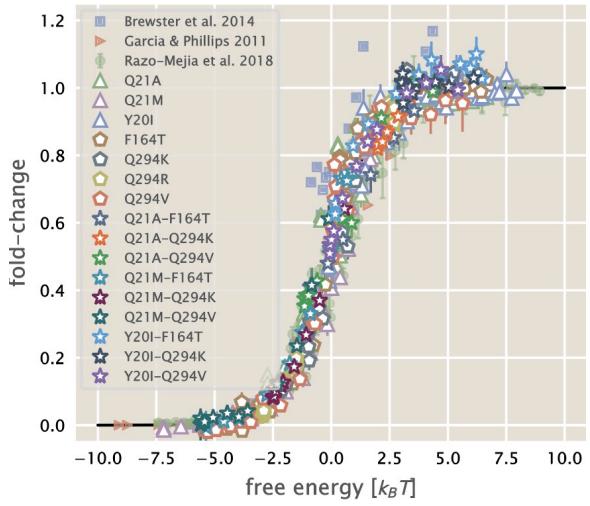


Figure 1.6: Gene expression for a great diversity of different situations all fall onto one master curve. The data collapse shown here highlights two key points: i) the effectiveness of equilibrium models of the process of transcription and ii) the emergence of "natural variables" that reveal what the cells "cares about" as opposed to the tuning variables used in the laboratory such as how tightly transcription factors bind to DNA, their copy number and the concentration of some inducer. (This figure is reproduced from Chure *et al.* [10].)

There are many ways to tune the level of gene expression of a bacterium. For example, we can artificially alter the sequence of the binding site for some transcription factor, thus

changing its K_d, and hence, how tightly the transcription factor binds the DNA. Alternatively, by tuning the ribosomal binding site for the mRNA that produces those transcription factors, we can control the copy number of the transcription factor. Or, we might tune the amount of some inducer, which essentially changes the number of *active* transcription factors. How do we account for all of these different ways of perturbing the level of gene expression?

As already seen in Figure 1.6, one route to computing the different states of occupancy of a promoter, and thereby, it's level of gene expression is the use of the Boltzmann law. In the context of transcription, the Boltzmann protocol asks us to enumerate the microscopic states of occupancy of some promoter (i.e. bound by polymerase, empty, bound by repressor, etc.) and then to compute their respective probabilities by invoking the Boltzmann distribution $p_i \propto \exp\left[-E_i/k_BT\right]$. For all of the different experiments reflected in Figure 1.6, the Boltzmann distribution was used to compute a single quantity that we will call p_{bound}, the probability that the promoter is occupied by RNA polymerase and thereby, producing mRNA. Without entering into any of the important technical details on the experimental or theoretical fronts, the reason for showing Figure 1.6 is to highlight how subtle the question of equilibrium and nonequilibrium really is in the biological setting. Let's now turn to a larger scale example in the context of embryonic development that illustrates the same subtlety.

Subtlety in pattern formation

The topic of how inorganic matter is turned into living organisms is one whose importance is felt in subjects other than the study of the growth of microbes. Indeed, its borders reach far beyond the traditional confines of the subject we call biology. In the 19th century, the subject of thermodynamics arose in large measure because of attempts to understand the workings of heat engines and how energy could be converted between different currencies.

Even discovering that there are different energetic currencies, especially the nature of the "kind of motion we call heat" was a great achievement [11, 12]. Unfortunately, the subject of thermodynamics felt short of its original name since almost always, the classical use of that term refers to thermostatics, the study of thermal systems in their terminal privileged state of equilibrium in which they are no longer changing in time. One of the deepest questions in modern science is how to generalize the ideas of traditional equilibrium systems to handle the full complexity of nonequilibrium steady states of living systems. Living organisms are *thermodynamic* systems in the true sense of the word in that they are both thermodynamic and involve the interconversion of matter and energy at every single moment that a chunk of matter is "alive."

The extraordinary numbers we are staring at about the scales of dissipation might give the impression that the physical consequences of nonequilibrium are always biologically obvious or conspicuous. To remember to disabuse ourselves of falling into this impression, here we offer an evocative thought experiment that clarifies that the accomplishments of energy expenditure are often subtle or barely perceptible at the scale of mean free paths of molecules—but of course emergently and importantly functional at larger scales or on statistical average.

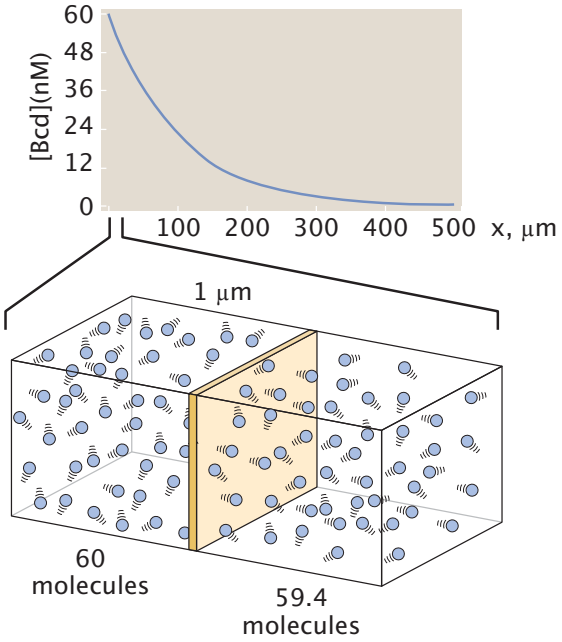


Figure 1.7: The superficially artificial nature of the distinction between equilibrium and nonequilibrium. The long axis of the 500 μm long fly embryo is discretized the into a collection of 1 μm boxes, the difference in the number of Bicoid molecules from one box to the next is less than a single molecule.

As shown in Figure 1.7, the subject is far more subtle than to simply make pronouncements that living organisms are out of equilibrium. In particular, there we see that given the measured Bicoid gradient in the developing fly embryo, if we discretize the 500 μm long embryo into a collection of 1 μm boxes, the difference in the number of Bicoid molecules from one box to the next is less than a single molecule. Clearly these individual molecules "know" nothing about a gradient: the distinction between equilibrium and nonequilibrium requires a macroscopic concept of fluxes.

In our view, these vignettes showing that equilibrium and nonequilibrium systems can surprisingly resemble each other makes even more interesting the effort to precisely define the boundaries of reachable behavior of each class of system, especially in biology. This question finds echoes in all of the chapters of this thesis.

In our view, the studies considered here help point the way to finding more general thermodynamic principles that describe active matter in all of its generality.

Bibliography

[1] Ron Milo, Paul Jorgensen, Uri Moran, Griffin Weber, and Michael Springer. Bionumbers—the database of key numbers in molecular and cell biology. *Nucleic acids research*, 38(suppl_1): D750–D753, 2010.

[2] John A Ciemniecki, Chia-Lun Ho, Richard D Horak, Akihiro Okamoto, and Dianne K Newman. Mechanistic study of a low-power bacterial maintenance state using high-throughput electrochemistry. *Cell*, 187(24):6882–6895, 2024.

[3] Ron Milo and Rob Phillips. *Cell biology by the numbers*. Garland Science, 2015.

[4] Ian S Farmer and Colin W Jones. The energetics of escherichia coli during aerobic growth in continuous culture. *European journal of biochemistry*, 67(1):115–122, 1976.

[5] Matteo Mori, Chuankai Cheng, Brian R Taylor, Hiroyuki Okano, and Terence Hwa. Functional decomposition of metabolism allows a system-level quantification of fluxes and protein allocation towards specific metabolic functions. *Nature Communications*, 14(1):4161, 2023.

[6] Amit Varma and Bernhard O Palsson. Metabolic capabilities of escherichia coli: I. synthesis of biosynthetic precursors and cofactors. *Journal of theoretical biology*, 165(4):477–502, 1993.

[7] Hong Zeng and Aidong Yang. Modelling overflow metabolism in escherichia coli with flux balance analysis incorporating differential proteomic efficiencies of energy pathways. *BMC systems biology*, 13:1–18, 2019.

[8] Jack V Greiner and Thomas Glonek. Intracellular atp concentration and implication for cellular evolution. *Biology*, 10(11):1166, 2021.

[9] Cameron Chalk, Salvador Buse, Krishna Shrinivas, Arvind Murugan, and Erik Winfree. Learning and Inference in a Lattice Model of Multicomponent Condensates. In Shinnosuke Seki and Jaimie Marie Stewart, editors, *30th International Conference on DNA Computing and Molecular Programming (DNA 30)*, volume 314 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 5:1–5:24, Dagstuhl, Germany, 2024. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-344-7. doi: 10.4230/LIPIcs.DNA.30.5. URL https://drops.dagstuhl.de/entities/document/10.4230/LIPIcs.DNA.30.5.

[10] Griffin Chure, Manuel Razo-Mejia, Nathan M Belliveau, Tal Einav, Zofii A Kaczmarek, Stephanie L Barnes, Mitchell Lewis, and Rob Phillips. The energetics of molecular adaptation in transcriptional regulation. *bioRxiv*, page 638270, 2019.

[11] A. Einstein and L. Infeld. *The Evolution of Physics*. Cambridge University Press, Cambridge, England, 1938.

[12] SG Brush. *The Kind of Motion We Call Heat*. North Holland, Amsterdam, 1986.

Part I

Survival out of equilibrium

Chapter 2

Accounting for the plasticity and major drivers of minimal cellular metabolism

When model microbes like *E. coli* and *S. cerevisiae* grow fast, they coordinate their growth, metabolic rate, macromolecular makeup, and even morphology with remarkable precision. Such "growth laws" have inspired calibrated theoretical models of physiological adaptation, and also have enabled empirical predictions of cellular attributes. Yet, most of these patterns have been observed during relatively rapid growth or along stereotypical environmental perturbations, such as carbon limitation or sublethal antibiotic dosage. Does comparable coordination exist for cells that are not, or only slowly, growing? What processes set the scale of minimal cellular metabolism, and how plastic are such power requirements across cells and conditions? Towards answering these questions, we formulate simple biophysical estimates that bound the most relevant processes—including protein turnover, ion leakage across membranes (osmoregulation), and signal transduction—that likely dominate the energy budgets of non-growing cells. We give context to these estimates by building a small census of empirical parameters and power densities of cells and organisms in diverse states of activity, unifying diverse data sources.

These data include recent direct laboratory measurements of the power output of non-growing cells. Our analysis stresses the plasticity of basal power requirements across conditions; highlights persistent empirical mismatches between maintenance extrapolated at high growth and measured at low growth; and recommends specific experiments to resolve molecular processes enabled by cellular nonequilibrium. These findings help clarify the minimal and empirical costs of maintaining living matter.

This chapter proceeds in three parts. First, we review both new and classic characteristic scales of minimal microbial metabolism in context of both history and new measurements, contributing to this comparison by inferring new numbers from unusual sources. Second, we discuss where these numbers actually come from and the extremely different conceptual assumptions underlying them. Last, we sweep through an exploratory order of magnitude accounting of what processes struggling cells mainly pay for to keep alive.

2.1 Introduction

2.1.1 Overview

This chapter is motivated by one of the most classic types of experiments in the history of microbiology, namely, those in which the metabolism and growth of bacterial cells is measured in real time in the controlled setting of the lab. Specifically, the cells are provided with some "environment" and the subsequent dynamics is measured. Indeed, Jacques Monod, as one of the greatest modern practitioners of this art, had a beautiful thesis on this topic culminating in the discovery of gene regulation as well as the allostery concept. Monod waxed eloquently on the importance of these types of experiments: "The study of the growth of bacterial cultures does not constitute a specialized subject or branch of research: it is the basic method of Microbiology....In the present review we shall

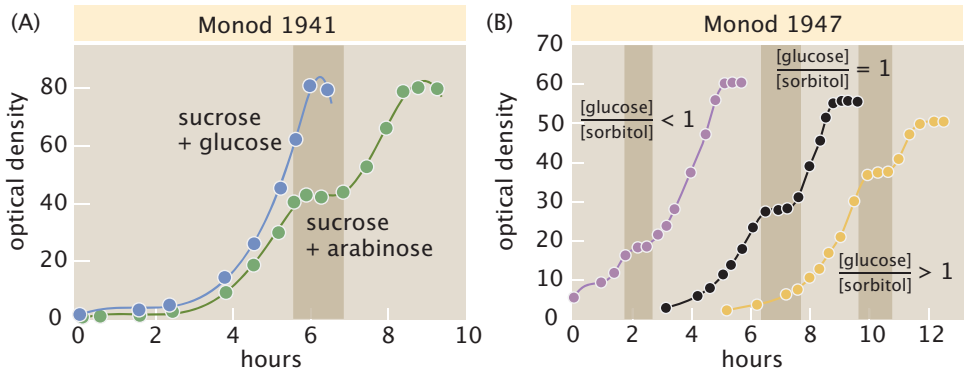


Figure 2.1: Choice of carbon source in the lives of bacterial cells. (A) Growth curves using optical density as a readout of number of cells reveal "enzymatic adaptation," where cells shift from one carbon source to another as indicated by the vertical brown band. (Specifically, cells were provided fixed initial composition of a mixture of sugars, and the brown bands emphasize when cells switch their physiology to metabolize the second sugar once the first is meaningfully depleted.) (B) Illustration of the shift from glucose to sorbitol as the carbon source for three different experiments involving different ratios of glucose to sorbitol. The shift from one carbon source to the other is indicated by the vertical brown bands. In fact, the lag in growth corresponding to those brown bands reflect the turning on of the genes for the enzymes needed to metabolize the secondary carbon sources. Adapted from Monod (1941) [2] and Monod (1947) [3].

consider bacterial growth as a method for the study of bacterial physiology and biochemistry" [1]. Some of the Monod's experimental results are shown in Figure 2.1 [2, 3]. In particular, here we see his measurements on bacterial growth where several different carbon sources are offered to the growing bacteria and their growth (as measured by optical density) is monitored over time. These measurements reveal that bacteria have preferences of some carbon sources over others. The shoulders in those curves correspond to the time period in which the cell is turning on the genes associated with the utilization of the less preferred carbon sources.

The work in this chapter was inspired by ongoing experimental efforts in the Newman Laboratory that call for a careful reckoning of bacterial physiology in circumstances that are widely divergent from those used in the famed experiments of Monod described above. In particular, new technologies made it possible to carefully measure the fluxes in electron transfer reactions in *Pseudomonas aeruginosa* in environmental conditions in which the cells

are deprived of oxygen (terminal electron acceptors), resulting in a form of "maintenance metabolism" and we were curious to quantitatively understand what changes to physiology are required to survive in these circumstances.

One of the founders of the modern field of bacterial physiology, Fred Neidhardt, wrote a beautiful article entitled, "Obsession with dN/dt" that highlights his own lifetime love affair with bacterial growth curves and the allied mathematics [4]. Neidhardt was cognizant that there is more to bacterial physiology than was revealed by the focus on bacterial growth curves: "Did Monod's views divert attention from the stationary phase and delay serious study of the developmental processes preparing cells for survival during nongrowth? I doubt it. I suggest the opposite is true – that the attention given by Monod to the usefulness of growth analysis in bacterial physiology and biochemistry hastened the day when nongrowth could be meaningfully studied. To recognize the right questions and pursue them required an understanding and an information base gained from growth studies. *The physiology of nongrowth is not the absence of the physiology of growth. Understanding both the phenotypic and genotypic processes involved in stationary-phase physiology involves building on what we understand about growth physiology*" (emphasis added). How ready are we to seize Neidhardt's gauntlet to build understanding in the nongrowing setting? We direct the remainder of this chapter precisely to begin asking how ideas may need to be generalized or superseded in the case of "maintenance metabolism"—life on the energetic brink.

2.2 Experimental measures and philosophical notions of basal cellular metabolism

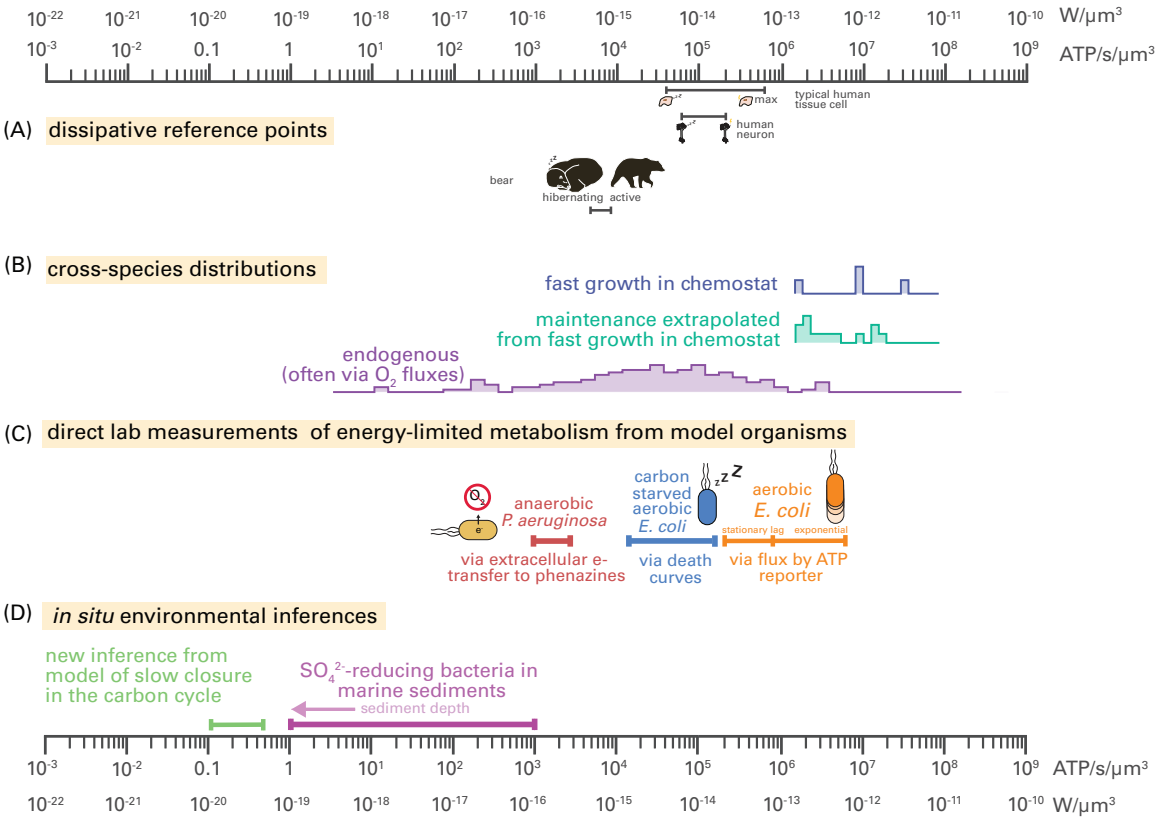


Figure 2.2: Extraordinary metabolic plasticity in microbes, as reported by different experimental means. (A) Metabolic plasticity of broader cell types and famously-hibernating organisms. (We calculated the metabolism of hibernating bears from experimental data in [5].) (B) Traditional laboratory measures of microbial metabolism, including chemostat measurements of species at high growth; extrapolated maintenance values from chemostat measurements; and various microbes persisting using just endogenous energy stores (often measured as oxygen fluxes). Note that maintenance values extrapolated from chemostats are large. (Compiled metabolic data across species were retrieved from [6, 7].) (C) Case studies of direct laboratory measurements of model organisms under nongrowth conditions. (*Pseudomonas* metabolic measurements, in red, were reported by [6]. Metabolic rates of aerobic carbon starved *E. coli*, in blue, were converted from death curve measurements by [8]. [9] reported values of *E. coli* metabolism at different growth stages, measured by ATP reporters, given in orange.) (D) Inferred maintenance values *in situ*. (Metabolic rates of sulfate reducing bacteria, in purple, were converted from values summarized by [10]. In green we give our fresh deduction of metabolism as consequence of the slow closure model in [11] and described in §2.4.)

Just as they amaze us by how rapidly they can devour energy, cells must also astonish us by how slowly they can consume it when conditions demand. While the fastest-growing microbial cells can metabolize almost 100 million ATP equivalents per second per cubic micron, cells deep in ocean sediments have been reported to metabolize at nearly 100 million times slower, as shown by Fig. 2.2(B-D). Cells of the same species (say *E. coli*) regularly slow metabolisms by three orders of magnitude when warranted. This immense metabolic plasticity is extraordinary on broader biological terms. For instance, both bears accomplishing their famous hibernation, or many typical human tissue cells, only reduce their metabolic rates by a factor few (to few times ten) between sustained states of relaxation and maximal exertion (as intimated in Fig. 2.2(A)).

To assemble the numbers reported in Fig. 2.2, we combined both existing compilations of metabolic rates [6, 10] (given as the histogram distributions in Fig. 2.2(B)) and also inferred fresh numbers from distinct sources, next converting them to common units to assess their communion with extant values of different origins.

Where do these highly provocative "basal metabolism" numbers, of the sort visualized in Fig. 2.2, come from? In other words, by what measurements do we arrive at what we purport to know about slow metabolism? Do these measurements cohere, or do they expose numerical and philosophical fractures? We now discuss some dominant means of inferring slow metabolic rates to frame these abiding questions.

Measuring metabolic rates has a distinguished and ingenious history in microbiology. Just thirteen years after he invented the nuclear chain reaction ushering in our nuclear age (and five years after he passionately petitioned the U.S. President and the Manhattan Project to avoid dropping a bomb), physicist Leo Szilard together with biologist Aaron Novick introduced a culturing device known as the chemostat [12, 13, 14], now a foundational technique across biology. As illustrated in Fig. 2.3, the chemostat smoothly controls the abundance of microbes consuming a limiting resource in culture medium.

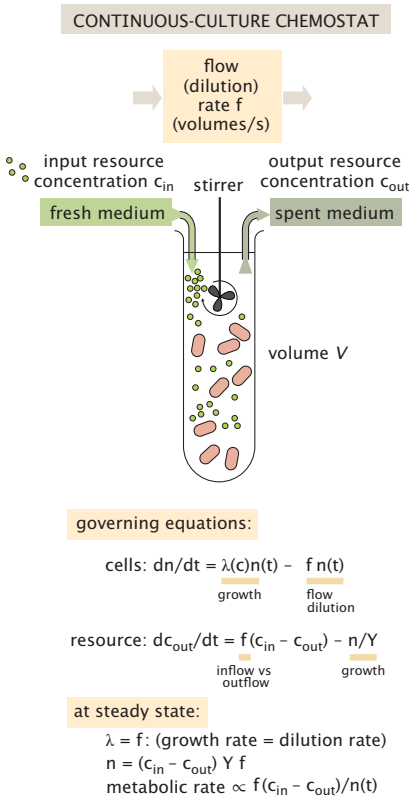


Figure 2.3: Continuous culture chemostats give correlations between bulk carbon consumption rates and fast growth rates. Cells at number density $n(t)$, consuming limiting resource at average concentration $c(t)$ outflowing at the same concentration $c_{\text{out}}(t) = c(t)$ at rate f, with a resource yield of Y, manifest a balance between their rate of growth λ and the flow rate. This allows the growth or metabolic rate to be inferred directly from teh flow rate and the difference between the resource concentrations of media flowing out and that of the fresh medium flowing in (c_{in}).

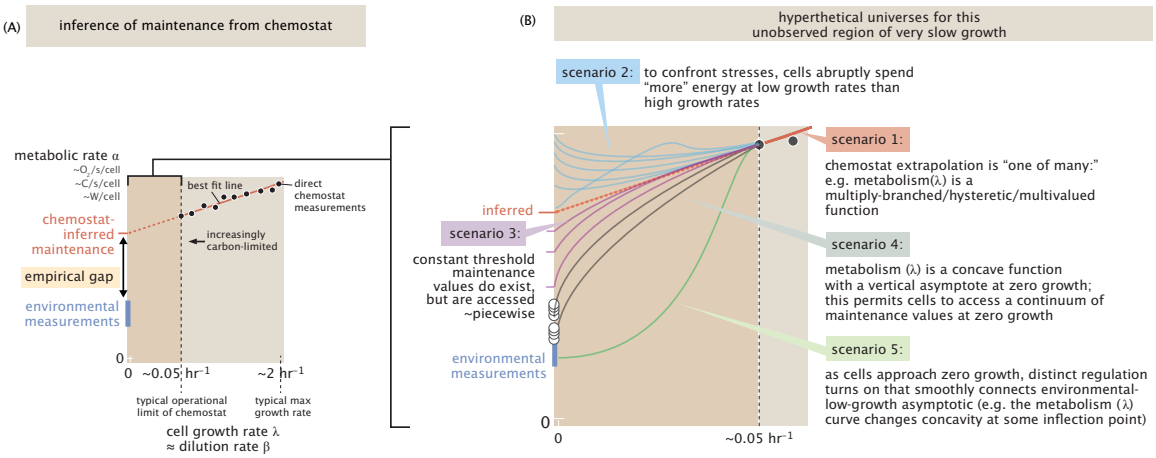


Figure 2.4: Conceptual subtleties of inferring basal metabolism using extrapolation from fast growth measurements in a chemostat.

The operating principle is simple and inspired. With inflow of fresh medium and outflow of cells and spent medium in a fixed volume, cells can impose a steady-state on the abundance of a resource (canonically, the carbon source) required for metabolism and growth. The flow rate of fresh medium sets the steady-state abundance of cells (as hinted at by the equations in Fig. 2.3), and crucially, the difference in resource level between the fresh inflowing medium and the spent outflowing medium reports on the resource consumption rate (and thus, biochemical energy consumption) of the cells.

The ability of the chemostat to rapidly balance and control the context of cells has unlocked a whole industry of deep insights about physiology, especially probing the regimes of rapid, exponential, steady-state bacterial growth and the extraordinary regularities manifested in that biochemical regime. Growth measurements across species soon indicated that when the dilution rate (which entrains the growth rate of the population) of fresh medium is decreased away from the fastest exponential growth attainable by cells, the metabolic rate on a per-cell basis also usually decreases. As illustrated schematically in Fig. 2.4(A), remarkably—for many laboratory strains and medium conditions—this decrease is often linear from fast growth rates to modest growth rates [15]. This line between metabolic rate and growth rate usually has nonzero y-intercept, and foundational studies (including by Pirt and colleagues

[15]) proposed that this extrapolated energy consumption rate at zero growth rate could be interpreted as the basal metabolic rate cells need just to stay alive, not to produce biomass.

This extrapolated intercept from growing cells in chemostats, often denoted β, is perhaps the dominant operationalization of what "basal maintenance metabolism" means when values are reported in the literature [15, 7, 6]. It is invoked in wonderful, deservedly influential, theoretical studies such as Lynch & Marinov's "The Bioenergetic Costs of a Gene" [16] and beyond. Fixed parameter values inferred from chemostat and culturing studies also enter applied climate models [17]. However, comparing the maintenance values extrapolated from chemostats across species and contexts (the data shown as teal histogram in Fig. 2.2(B)) with the basal metabolisms separately inferred from natural environments (Fig. 2.2(D)) or different laboratory methods (Fig. 2.2(C)) emphasizes that the chemostat estimates of basal metabolism are frequently many orders of magnitude greater than environmental inferences or direct laboratory measures.

This conspicuous empirical gap between chemostat measurements and other routes to basal metabolism values prompts us as a field to introspect about the underlying conceit of these extrapolated measurements. We remark that chemostats have limits on the slowest flow (hence cellular growth) rates they are able to stably operate at, without burgeoning technical tricks and complexity. A characteristic lower limit to the comfortable (common) operation of chemostats is a flow (growth) rate λ of the order of $\lambda_{\text{min}} \gtrsim \text{few} \times 10^{-2}\ \text{hr}^{-1}$, corresponding to a generational doubling time of faster than a few tens of hours (as schematized by the dark brown band of a Pirt style plot in Fig. 2.4(A)). By the nature of extrapolation, slower growth and therefore metabolic rates—precisely the regimes of say geobiological interest, anticipating doubling times slower than months—remain directly unobserved.

Accordingly, what nonlinear behaviors linking metabolic rate and growth rate could truly lurk in the limit as growth rate vanishes? Fig. 2.4(B) schematically imagines some

mathematical and physiological behaviors that connect a linear trend of metabolic rate with high growth rate to a spectrum of true underlying basal metabolic rates as bacteria stop growing. Specifically, we speculate microbes could very plausibly change the concavity (in either direction) of their metabolism's relationship with growth as they become more energy constrained or stressed; this could manifest as the emergence of a burst of energy expenditure or conservation as microbes drop their growth rates relative to the linear extrapolation expectation.

More generally, we offer this cartoon *not* to stray into a pedantically conspiratorial stance of asking, "What if biology is more complicated than a linear extrapolation?"[1] Instead we register these possibilities to ask whether the very act of quantitatively expecting growth rate λ to stay a sufficient, closed "equation of state" for the physiology of a challenged microbial cell, in the bonafide physics sense, may be overly optimistic. Of course, exactly this optimism has been richly rewarded in the case of *fast* growth. As Neidhardt incisively remarked decades ago, "Certain major phenotypic characteristics are coordinated with the absolute growth rate and are completely independent of the chemical nature of the medium. . . No description of the average bacterial cell of a given species is possible unless one specifies the growth rate" [4]. The near-sufficiency of growth rate alone to specify many aspects of the state of a cell has been more recently ratified in model organisms (growing exponentially in simple environments) to inspirational quantitative precision. This includes the establishment of "growth laws" by the groups of Terence Hwa and many other investigators [18]. We harvest from these examples big assurance that explanatory state variables (like growth rate) do emerge in biology (whose recognition yields thrilling understanding), but *not* that it will be specifically growth rate that is still the "right" state coordinate that best organizes the complexity of microbial metabolism in the limit of energetic challenge. In other words, these ideas require our community be bold

[1] Biology always exceeds the complexity of the line; what matters is the conceptual leverage a line does or does not give us.

enough to ask for the new abstractions that replace the growth rate as governing (but measurable) x-coordinate(s) in plots describing microbial cells in the energy-limited setting. Is this a membrane potential? An NAD^+/NADH ratio? Some new organizing variable entirely?

Progress towards new, effective abstractions demands being able to test them out ruthlessly in experiments. Our field has never enjoyed a time like ours when more of the right technological and biological (e.g. genetic) tools are at hand to do this. We now briefly remark on three, newly-direct laboratory measurements contributing values of total basal metabolism. We anticipate these measurements are especially the sort that may help probe physiology in particularly challenging but revealing conditions, each with a corresponding datum shown in Fig. 2.2(C). These are exploiting the quantitative phenomenology of death curves of bacteria to infer cellular energy demands (as shown in Fig. 2.5); direct measurements of ATP flux by reporters over time (as shown in Fig. 2.6(A)); and measurements performed using electrodes on cells reducing terminal oxidants (as shown in Fig. 2.6(B)).

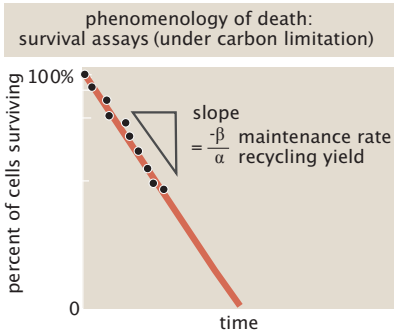


Figure 2.5: Inferring maintenance rate from death rates under carbon limitation. Under suitable conditions, internal population recycling of the energetic content of dying cells implies that the slope of survival curves on a semilog axis reports on the maintenance rate divided by the recycling yield.

How can death report on life? An ascendant tradition has emerged from the work of Schink, Gerland, Basan, and coworkers [19, 8] whose beautiful experiments show how—under stereotypical carbon limitation—*E. coli* cells often die at constant-per-capita rates, as

schematized by Fig. 2.5. This cursorily simple behavior is actually quite a deep and surprising regularity. Why should eight cells become four, four cells become two; two cells become one; and so on, all at a constant kinetic rate? Pulling on this conceptual thread can be a route to measuring valuable metabolic parameters. In the case of carbon limitation, it transpires that a straight line in the decline of (log) cell population emerges from dying cells being recycled by their still-viable neighbors. In the next chapter, we explore the existing math of this scenario much more granularly (when asking how it might be generalized to the richer and entirely physiologically different case of oxygen deprivation). However, for our present discussion, we remark that since (as is intuitive) struggling cells die faster when they harvest energy faster under a limited supply, the rate at which cells die reports quantitatively on the basal energy consumption rate β per cell in these conditions. To enable comparison with the other repertoire of reported values at hand, we translate some of the values of β identified in recent experiment's units [8] to common comparative metabolic units in Fig. 2.2(C) as the blue bar.

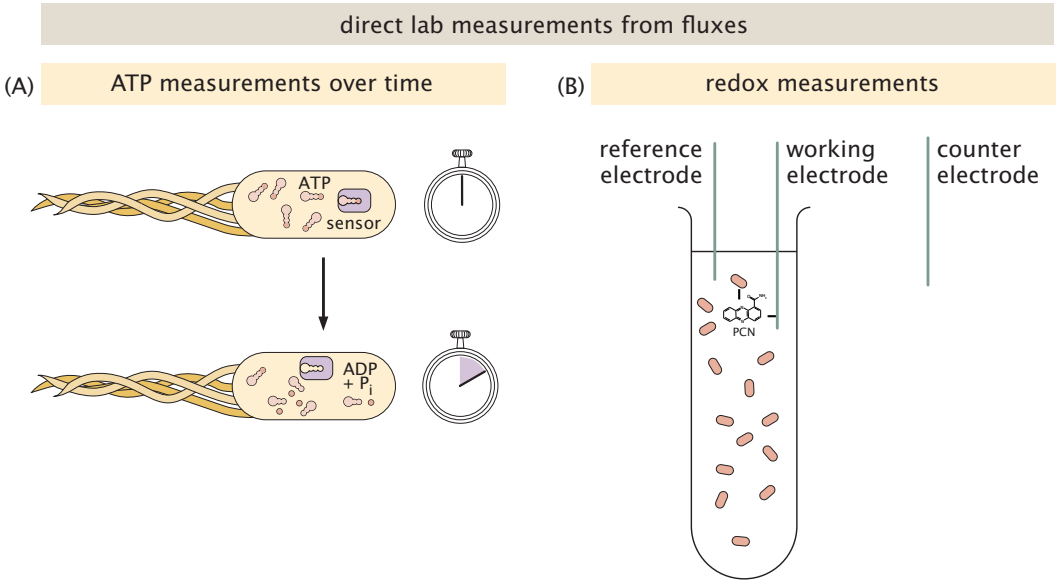


Figure 2.6: Inferring cellular energy consumption by (A) direct measurement of ATP consumption or (B) direct measurement of electron transport by electrode exchanges with extracellular electron carriers such as phenazines (including phenazine-1-carboxamide (PCN)).

Another quite recent, and exciting, glimpse of the real-time metabolism of microbial cells is by directly measuring ATP abundance by fluorescent reporters and watching resulting resolved fluxes [9]. This gives, for instance, the ability to both measure current endogenous ATP levels, and also finely temporally resolve the metabolic status of cells (from changes in this ATP pool) as they exit exponential phase and enter stationary or lag phase. We summarize Deng and coworker's measurements [9] on aerobically-grown *E. coli* in various stages of growth in Fig. 2.2(C) as the orange bars. These measurements show that aerobic *E. coli* slows its metabolic rate from $\sim 6.4 \times 10^6$ ATP/s/cell in exponential phase, to $\sim 8 \times 10^5$ ATP/s/cell in lag phase, to $\sim 2 \times 10^5$ ATP/s/cell in stationary phase. Complementing these changes in metabolic rates, Deng & coworkers also report how the total quasi-steady state abundance of ATP changes across these growth stages; *E. coli* reduces its typical cytosolic ATP pool from ~ 3 mM standing ATP in exponential phase to order ~ 0.6 mM standing ATP in stationary phase [9]. Incidentally, the purview and immediacy of this class of direct measurements of ATP is much wider than whole-cell energetic metabolism. We advance work with members of our laboratory in this direction, using ATP fluorescent reporters to ask how much building microtubule structures costs across space and time, in a later chapter in this thesis.

Most of these methods of experimentally measuring basal metabolisms depend on operating conditions that compel cells to be in particular regimes of physiology. That is, these direct measurements often access the regime of nongrowth by imposing stress along stereotypical axes—overwhelmingly commonly, carbon deprivation (or antibiotics). However, the reasons that cells survive, but do not grow, in natural environments can be much richer and more varied than just lacking carbon sources. Across the biosphere, cells instead often lack terminal electron acceptors like oxygen crucially required to allow them to perform metabolism. Unfortunately, many of the empirical means to assess metabolism we have just discussed do not directly translate effectively to oxygen deprivation. For

instance, fluorescent ATP reporters (like many assays depending on fluorescent proteins) often have completely different and abrogated structure or kinetics in anoxic settings. Similarly, the lovely quantitative picture that so enabled rates of death to report on basal metabolism in the case of carbon limitation is completely upended with terminal electron acceptor deprivation. (How this happens, and whether there may be some quantitative analog to allow death alone to report on metabolic rates, is a question we broach seriously in the next chapter.)

However, compelling work by the laboratory of Dianne Newman and colleagues instead turns the challenges presented by measuring anaerobic metabolism into an advantage. Specifically, some cells who manage to cope with anaerobic environments produce their own extracellular electron acceptors. This is manifested by the opportunistic pathogen *P. aeruginosa*, which produce phenazines, a class of redox active compounds that can mediate extracellular electron transfer. Experimentally, these molecules give a rare handle on measuring metabolic activity, since electrodes can be tuned to the midpoint potential of phenazine and thus selectively resolve the rate at which cells metabolize at high temporal resolution. We depict this scheme schematically in Fig. 2.6(B), and explore fun mathematics animated by these scenarios throughout the next chapter. We indicate the intriguing characteristic scale of metabolic rates measured in these systems in Fig. 2.2(C) as a red bar. Specifically, Ciemniecki *et al.* announce that anaerobic *P. aeruginosa* can metabolize at the striking rate of just $\sim 10^3 - 10^4$ ATP/s/cell.

This small number is tantalizing, in part, because it approaches the amazingly small values of microbial metabolism inferred from natural environments *in situ*, as visualized by Fig. 2.2(D). The tradition of laboratory measurements of microbial metabolism is rich and smart; in this great tradition, however, typical values at the bench are often separated from inferred maintenance rates of microbes in say marine sediments [10] by several orders of magnitude (as seen in the gulf between the lab values of Fig. 2.2(A-C) with the

values of metabolism *in situ* in Fig. 2.2(D)).

To deduce how quiescent microbes in these natural environments metabolize poses daunting technical challenges. Microbes in the ocean floor resist many forms of scrutiny. Mathematics can be one of the only microscopes available to understand these settings. This tradition has a big history. We now offer a very small entry in the spirit of this tradition. Observables like empirical scaling laws in the age of organic carbon in the ocean implicitly report on how slowly microbes metabolize. Recently, Rothman [11] proposed a model of the global carbon cycle that makes a slew of empirically-successful predictions about even the precise values of aged carbon and cellular abundance in the ocean. This model posits that past some age of carbon, microbial metabolism ultimately transitions from accumulating biomass—albeit very slowly—to a constant fundamental basal metabolism independent of growth. We were inspired by Rothman's elegant model to ask what it predicts for the basal metabolic rate of these bacteria *in situ*. In the appendix of this chapter (§2.4) we derive a simple prediction for this basal metabolic rate and evaluate it numerically using exponents of empirical scaling laws. We arrive at a phenomenally slow metabolic rate of plausibly less than a single ATP consumed per second per microbe, as indicated by the green bar in Fig. 2.2(D). This quantitatively buttresses a view of natural microbes apparently subsisting on extraordinarily slow rates of energy consumption.

By calculating and comparing empirical or inferred numbers such as those in Fig. 2.2 and throughout this chapter, we do not mean to canonize them, or pretend that these quantities are truly well "known" to acceptable precision and rigor. To the contrary. Numbers that matter, as basal metabolisms do both practically (in e.g. climate models) and conceptually (as inputs into nonequilibrium statistical models of living matter), must be measured many times in multiple different ways. Differences between reported basal metabolism numbers are already visible from the subset of settings and separate measurement sources we explored in this chapter. These differences must be reconciled or

attributed to particular features of these biological systems. We view these numbers as urgent provocations for us to account for both true underlying physiological plasticity in metabolism, and also differences in philosophy, measurements, and environmental context.

2.3 Plausible Energy Sinks in the Cell

Having seen just how plastic total basal metabolism appears to be, we now ask, what activities fundamentally set the scale of minimal metabolic rates? In the sections that follow, we strive for a series of increasingly sophisticated attempts to quantitatively reckon with the various dissipative processes within cells. This strategy is illustrated schematically in Figure 2.7 where we show some of the processes that make up the energy budget of microbial cells.

2.3.1 The Cost of Protein Turnover

As seen in Figure 2.7, one of the primary costs incurred by bacterial cells whether in a state of maintenance metabolism or rapid growth is associated with the synthesis of proteins. To that end, in this section, we attempt to make some rough estimates of the power budget associated with the synthesis or resynthesis of proteins.

Cells constantly remake proteins to compensate for their active or spontaneous degradation. The energetic cost to resynthesize proteins can be formidable, as a simple estimate reveals. Specifically—assuming that the cellular abundance of proteins is maintained at a steady-state—the stable copy number N, overall resynthesis rate q, degradation rate k, and typical lifetime τ can be related by Little's Theorem in queuing

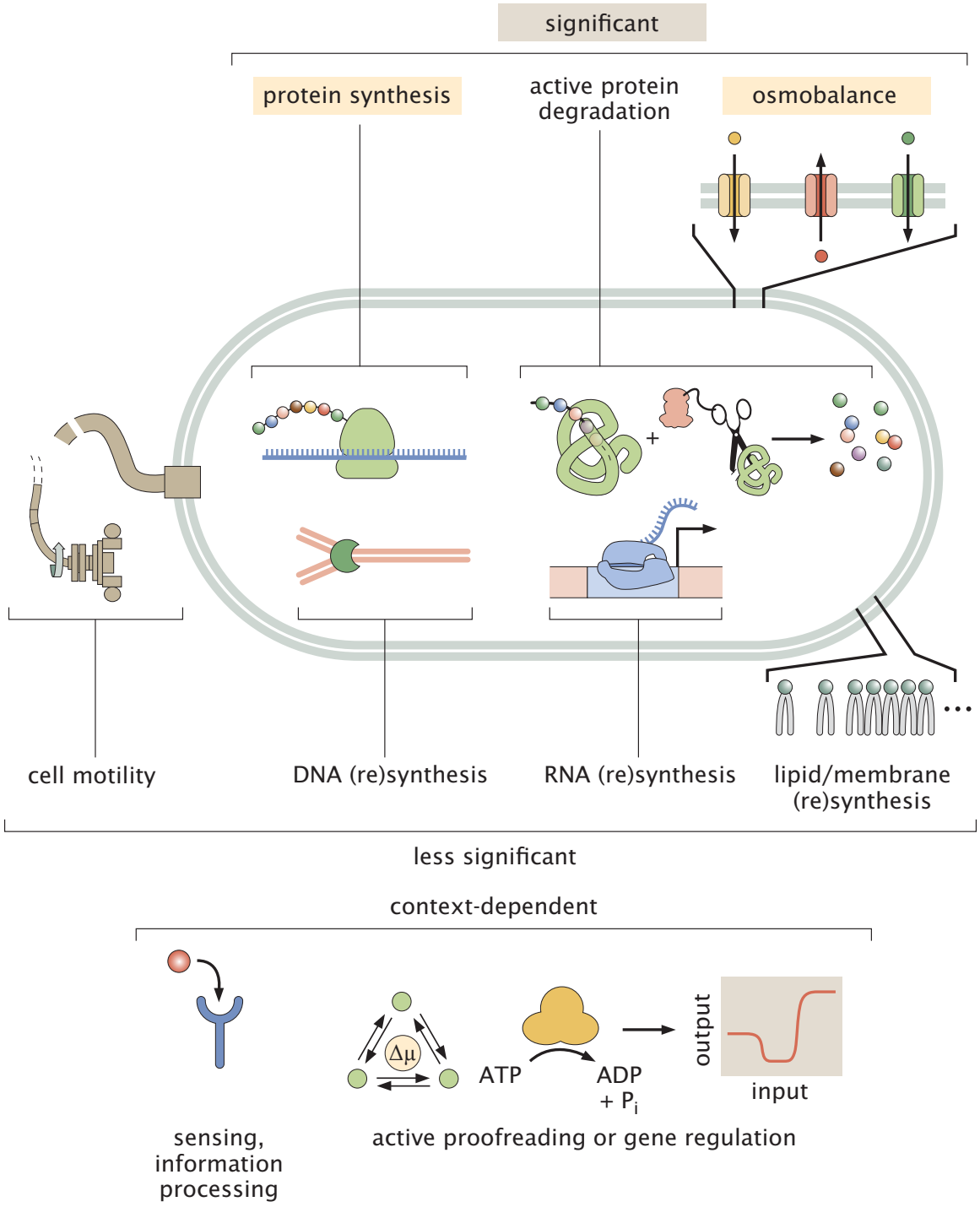


Figure 2.7: Processes that make up microbial energy budgets. Each of the processes shown in the figure can be subjected to quantitative estimates and compared; their balance surely adjusts across conditions of severe energy limitation to active growth. We attempt to categorize these processes by how large a contribution they make to the cell's energy budget, seeking bounds to focusing on the energy-limited case.

theory [20],

$$\frac{N}{\tau} = q = k. \tag{2.1}$$

Typical protein turnover rates in bacteria [21] and eukaryotes [22] are both of order $k \approx f \times 10^{-5}$ s^{-1}, consistent with measured typical half lives of a few hours (and where remark we adopt the useful convention that $f =$ few [23]).

We now ask, given these turnover times, how much does it cost to constantly replenish these degraded proteins? Figure 2.8 provides a schematic of our strategy for estimating the power associated with protein replenishment due to turnover. In particular, we imagine that under maintenance conditions, only a fraction (e.g. 10%) of the proteome is replenished. Given the rate of degradation $k \approx f \times 10^{-5}$ s^{-1}, we can compute the number of new proteins synthesized every second as

$$\text{protein synthesis rate} \approx \frac{1}{10} \times (f \times 10^{-5}\ \text{s}^{-1}) \times (f \times 10^{6}\ \text{proteins}) \approx 10\ \frac{\text{proteins}}{\text{s}}, \tag{2.2}$$

where the $1/10$ on the right hand side is the fraction of the proteome we imagine is replenished and $f \times 10^6$ proteins is the typical number of proteins per μm^3 (or per bacterial cell). Given that a typical protein is composed of $f \times 10^2$ amino acids and that the cost of each such amino acid incorporation is 4 ATP equivalents, then the cost of maintaining the proteome is

$$\text{power of protein synthesis} \approx \left(10\ \frac{\text{proteins}}{\text{s}}\right) \times (f \times 10^{2}\ \text{aa}) \times \left(4\ \frac{\text{ATP}}{\text{aa}}\right) \approx 10^{4}\ \frac{\text{ATP}}{\text{s}}. \tag{2.3}$$

We can formalize the calculations above as follows. Each protein is $\ell \approx$ a few hundred amino acids long and each peptide bond requires $\eta \approx 4$ ATP equivalents to polymerize [25]. Since mammalian and bacterial cells alike typically have concentrations $c \approx$ a few

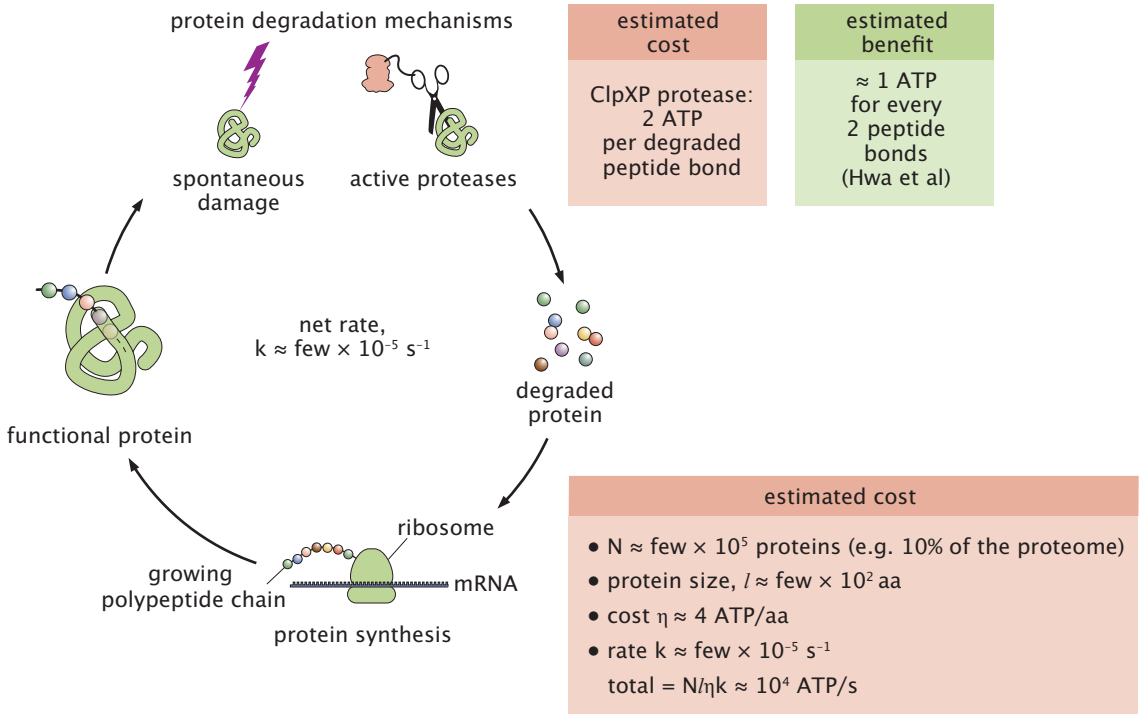


Figure 2.8: Degradation and active resynthesis of proteins enacts a steady-state with characteristic turnover rates and accompanying costs. The cycle on the left shows the life history from newly synthesized protein to degradation. The bottom right recapitulates the key features of the estimates made in the body of the paper. (We note that one enormous part of the picture for which reliable quantitative data are currently in severe deficit is the question of whether protein degradation pathways exist that could, in the net balance, *produce* energy; compelling initial impressions of this possibility are being investigated by Hwa and colleagues [24].)

million proteins per cubic micron [26], protein resynthesis to maintain the standing proteome pool should demand of order $\gamma c k \ell \eta \approx \gamma 10^5$ ATP/(μm^3 s), where we introduce the parameter γ to represent the fraction of the proteome that is resynthesized during maintenance conditions. It is amusing to speculate on the parameter γ by recalling efforts to create minimal genomes, some of which are ongoing even here at Caltech in the laboratory of Kaihang Wang. Given that the total number of genes in *E. coli* is $\approx$ 4500 and the minimal genome size is of order 500, one might imagine that under maintenance conditions, a variety of proteins do not undergo replenishment. That said, like with much of the guesswork in this chapter, there is clearly so much more experimental work to be done to pin down the mechanisms of microbial maintenance metabolism.

Cost of protein turnover inferred from cellular data

In a paper now three decades old, Russell and Cook [27] had already thought deeply about these matters, claiming that the empirical rate at which proteins turnover in *E. coli* is not sufficient to account for inferred maintenance rates in this organism. We find it instructive to quote their observations in their entirety. "Radiolabeling experiments indicate that the rate of protein turnover in exponentially growing *E. coli* cells ranged from 0.5 to 2.5 %/h and that RNA turnover paralleled the breakdown of protein. During stationary phase, protein turnover was higher (5% for *E. coli* and 8% for *Bacillus cereus*, but even these rates cannot account for all of the maintenance in *E. coli*. On the basis of a cell composition of 0.5 g of protein per g of cells, 4 ATP equivalents per amino acid polymerized, an average molecular mass for an amino acid of 100 Da, and a maximum ATP production of 24 ATP per glucose, the glucose consumption rate needed to sustain a protein turnover rate of 5%/h would be 0.04 mmol of glucose per g of cells per h. The maintenance rate of *E. coli* is 0.31 mmol of glucose per g of cells per h [as extrapolated from chemostat measurements]."

Let us try to come to terms with what Russell and Cook have said. First, they note that

the cost of protein turnover seems to fall far short of the total power cost of maintenance. However, a protein turnover rate of a few %/hr in stationary phase, assuming a total protein copy number of $N \sim \text{few} \times 10^6$ proteins/cell, corresponds to,

$$\frac{\text{few turned over}}{100 \text{ total}} \times (\text{few} \times 10^6 \text{ total}) \sim 10^5 \frac{\text{proteins turned over}}{\text{hr cell}} \tag{2.4}$$

$$\sim \boxed{\text{few} \times 10 \frac{\text{proteins turned over}}{\text{s cell}}}. \tag{2.5}$$

Since a protein is typically $\ell \sim 300\ aa$ long and polymerizing each amino acid into a growing peptide chain costs about 4 ATP hydrolyses worth of energy, it takes order 10^3 ATP/protein. This anticipates that protein turnover appears to demand $\boxed{\text{few} \times 10^4 \text{ ATP/s/cell}}$. This numerical estimate is intriguing in the context of recent measurements in *Pseudomonas* which uses phenazine to cycle a median value of a few $\times 10^3$ e^-/s/cell [6]. Since each electron transported along relevant metabolic pathways provides about the equivalent of one ATP hydrolysis's worth of energy (up to a factor few), the measurement corresponds to about ten times smaller an ATP consumption rate per *Pseudomonas* cell than demanded by protein turnover in *E. coli* in stationary phase. We consider it plausible that *E. coli* on glucose could easily vary physiologically within a factor ten of the governing properties of *P. aeruginosa* using phenazine. This calculation gives grounds for inferring that a large part of the observed metabolic rate of *Pseudomonas* might indeed be explained by just protein turnover, but explicit turnover data in *Pseudomonas* are obviously warranted.

Specifically, to inform a more useful and representative estimate of the proportion of an anaerobic cell's energy budget destined for protein resynthesis, absolute (not just relative) copy numbers of proteins *and* their turnover rates should be measured. While these values would seem very basic and essential quantities, they appear to be only rarely measured simultaneously in well-defined conditions where other metabolic features are also

measured. We view this deficiency as a direct call to action for our experimental efforts in the community.

The above quick calculation assumes that the total copy number of proteins in stationary growth is still a few million per cell, an abundance informed by growing cells; if the protein count is different in stationary phase, then a few percent turned over per hour corresponds to a different quantitative number turned over per hour than estimated above.

Cost of protein turnover from first principles and chemical rates

One hypothesis advanced for the minimal metabolic energy flux cells need to survive is that slow spontaneous degradation of the proteome must be counteracted by some nonzero rate of protein synthesis, even without growth. Here, we make this claim quantitatively precise using a simple stochastic kinetic model. Next we coarsely estimate whether known mechanisms and kinetics of protein degradation or spontaneous racemetization could account for measured maintenance requirements.

Consider a proteome consisting of N proteins, each of which has ℓ amino acids. Each amino acid can fail at rate k (per amino acid per time). Say that m amino-acid failures per protein are required before the whole protein fails (requiring replacement by active synthesis). What is the expected rate at which proteins fail (hence approximately the total rate of new protein synthesis required to maintain a steady pool)?

Our first approximation is to treat the failures of amino acids as statistically independent and following the same Poisson process with failure rate k.[2] Thus, inside a protein, the net rate at which amino acids fail is the total number of amino acids times this failure rate,

[2]This treatment of independently-failing-entitites is analogous to how the net rate at which rhodopsin photoreceptors show spurious activation with temperature is rare per receptor, but meaningful in aggregate. For instance, one paper writes, "As a toad rod contains about 2×10^9 rhodopsin molecules, and as all these combined rhodopsin molecules give a mean interval between thermal breakdowns of 50 sec, then the mean thermal lifetime per rhodopsin molecule is the product of these, about 10^{11} sec or roughly 3000 years. So, despite the occurrence of the discrete electrical events, the rhodopsin molecule is incredibly stable" [28].

namely

$$\ell k \text{ amino acids failing/second/protein.}$$

The time τ that is needed to wait for m exponentially-distributed events each with failure rate λ is described by a Gamma distribution, $\tau \sim \text{Gamma}(m, \lambda)$, which has mean $\langle \tau \rangle = \frac{m}{\lambda}$. Thus the effective rate at which one protein fails (by reaching ℓ amino acid failures) is $\frac{\ell k}{m}$. Since there are in fact N proteins, each of which can be regarded as failing statistically independently, this predicts that there are

$$k^{\text{tot}} = \boxed{\frac{N \ell k}{m} \text{ proteins failing per second}}. \tag{2.6}$$

Note that this expression is intuitive: the greater the required number m of amino acids per protein failure, the smaller the overall protein failure rate; but the larger the number ℓ of amino acids that could fail per protein, the per-amino-acid failure rate k, or the total number N of proteins, the higher the total protein failure rate.

Now, we hazard a tentative numerical estimate of this total spontaneous protein degradation rate, based on the risk of amino acid "failure" attributed to spontaneous deamidation, isomerization, and racemization of protein residues. Geiger and Clark [29], focusing on asparaginyl/aspartyl residues whose associated chemical transformation rates are larger than for other resides, used hexapeptides to measure that L-Asn undergoes spontaneous deamidization with a half-life of $\tau_1 \sim 1.4$ days, followed by further hydrolysis (a reaction with half-life $\tau_2 \sim 2.3$ hr), and racemization (a reaction with half-life $\tau_3 \sim 19.5$ hr). Viewing these transformations as sequential kinetic processes, we

assimilate these values to estimate the overall rate to be

$$k \sim \frac{\ln 2}{\tau_1 + \tau_2 + \tau_3} \tag{2.7}$$

$$\sim \frac{\ln 2}{34\ \text{hr} + 2.3\ \text{hr} + 19.5\ \text{hr}} \tag{2.8}$$

$$\sim 3.5 \times 10^{-6}\ \text{amino acids failing/second/protein.} \tag{2.9}$$

Next, recalling the sensitivity of protein function to even single nucleotide polymorphisms in underlying sequence, we imagine that a protein fails when one out of a few amino acids has failed, namely $m/\ell \sim 1/\text{few} = \text{few} \times 10^{-1}$. Last, we take the typical value of a proteome having $N \sim \text{few} \times 10^6$ proteins. These values predict an effective overall rate,

$$k^{\text{tot}} \sim \left(\text{few} \times 10^6\ \text{proteins}\right) \left(\frac{\text{few total aa}}{\text{1 aa failure req'd}} \times \text{failed protein}\right) \left(3.5 \times 10^{-6}\ \frac{\text{aa failing}}{\text{s protein}}\right) \tag{2.10}$$

$$\sim \boxed{\text{few} \times 10\ \text{proteins ``failing''/second/cell}}. \tag{2.11}$$

Note that we must regard this estimate as a (plausibly loose) approximate upper-bound of the spontaneous protein failure rate, given that the rates reported by Greiger and Clark [29] were measured for extremely short polypeptides (reducing steric effects expected to retard these chemical transformations). Further, these rates also describe especially chemically-labile Asp and Asn; other amino acids have reduced analogous spontaneous transformation rates. Last, just reaching a racemized state may not yet render an amino acid functionally deficicient, without subsequent chemical and biological transformations not acknowledged here; so the rate of actual protein failure may be slowed by the times these other processes need to complete. Nonetheless, this estimate provides an interesting scale of how quickly spontaneous injurious transformations occur in a cell's proteome. However approximate, we note with surprise and curiosity the remarkable correspondence of the crudely estimated

spontaneous protein "failure" rate, Eq. 2.11, and the empirically-reported rate of protein turnover, Eq. 2.5.

In closing this broad discussion of the cost of protein turnover, we note that we could just as well have asked for the cost of the other processes in the central dogma such as replication and transcription during "maintenance metabolism." That is, costs of RNA and DNA repair and turnover are also incurred by the cell. Here we simply note that as was already examined in detail by Lynch and Marinov [16]: the cost of transcription is far lower than that of translation, and for now, we hypothesize that that same hierarchy will be true in the context of maintenance metabolism as well. Hence, we view the focus on protein turnover as the key process in the context of the central dogma.

2.3.2 Cost of post-translational protein modifications

Proteins of all types are continuously subjected to modifications by the addition and removal of chemical groups. In a very interesting article ("Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding"), Jeremy Gunawardena asks us to think about post-translational modifications as a way of expanding the natural repertoire of the 20-letter amino acid alphabet. Similarly, Prof. Christopher Walsh wrote a whole book entitled, "Posttranslational Modifications of Proteins: Expanding Nature's Inventory," again making the point that by adding chemical groups to proteins we can expand nature's repertoire, significantly changing their properties. However, we cannot forget, this widespread escape from genetic imprisonment has an energetic cost. In this section, we explore simple estimates of both the ubiquity and the cost of such post-translational modifications, once again trying to establish a baseline of what processes within cells are energetically "costly" and what processes are "cheap."

To get a feeling for the problem, let's consider the amendment of proteins by the addition of phosphate groups. The point is that proteins are all made up of the same 20 amino acids. But once those proteins are synthesized, some residues such as threonine, serine and tyrosine, can be modified through the addition of phosphate groups (there are a host of other post-translational modifications such as acetylation, palmitoylation and methylation, to name a few, but for concreteness we focus on phosphorylation). If we consider amino acid usage, we could say that threonine, serine and tyrosine make up roughly 10% of the residues and these are the ones that are available for phosphorylation. Our aim here is to make a simple, but essentially first-principles, estimate of the number of distinct states of phosphorylation there are for a typical protein and then to examine the cost of running the kinases and phosphatases that maintain a steady flux of post-translational modifications.

Figure 2.9 provides a crude representation of proteins that can help us estimate the fraction of surface residues available for post-translational modification. In our toy model we assume

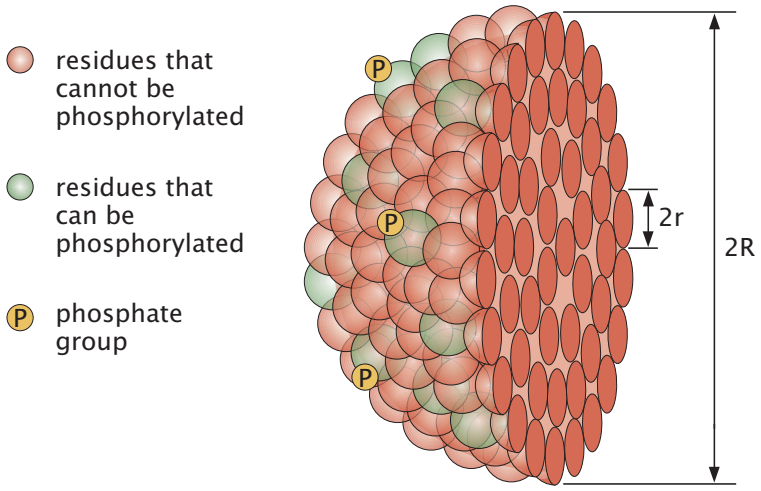


Figure 2.9: Phosphorylation of the amino acids on the surface of a protein. The simple estimate uses the known size of proteins and amino acids to estimate what fraction of the $f \times 100$ amino acids are on the surface of the protein, and then out of those surface exposed amino acids, what fraction are available for phosphorylation, for example.

that each of the N residues making up the protein is itself a sphere. The crude scaling of the radius of our spherical protein will go as $N^{1/3}$. However, we can do a little better. If the amino acids which we imagine as tiny spheres are maximally close packed, then the volume of the protein will be roughly $0.75N \times V_{aa}$. The factor of 0.75 is approximately equal to the volume fraction occupied by close packed spheres. Therefore, the radius of our protein will be given by $R = (0.75N)^{1/3} R_{aa}$. Given that a typical protein has $f \times 10^2$ amino acids and a radius of f nm, the radius of a single amino acid can be estimated as $R_{aa} \approx 0.4$ nm. Therefore, the scaling of the protein radius with the number of amino acids will go approximately as

$$R(N) \approx 0.4 \times (0.75N)^{1/3} \text{ nm}. \tag{2.12}$$

As a very crude estimate for the total number of residues on the surface of the protein, we can take the total surface area and divide it by the 2D projection area of a single residue, that

is

$$S_{\text{array}} = 4\pi R^2, \tag{2.13}$$

$$S_{aa}^{\text{proj}} = \pi R_{aa}^2, \tag{2.14}$$

$$N_{\text{surface}} = \frac{S_{\text{protein}}}{S_{aa}^{\text{proj}}} = 4 \times (0.75N)^{2/3} \approx f \times N^{2/3}. \tag{2.15}$$

Taking $N = f \times 10^2$ for a typical protein, we find that the number of surface residues is approximately

$$N_{\text{surface}} \approx 130. \tag{2.16}$$

A few (f) of 20 amino acids are commonly phosphorylated in bacteria. Making the crude assumption that all amino acids are equally represented on the surface, only 15% of the residues, or $130 \times 0.15 \approx 20$, will on average be phosphorylatable. This means that a total of $2^{20} \approx 10^6$ phosphorylation states are available to a given protein. Given that there are several million proteins in a given *E. coli* cell, this means that there are

$$\text{number of possible proteins in one bacterium} \approx (10^6)^{f \times 10^6}, \tag{2.17}$$

distinct possible proteins in one bacterium! We really do not have any good way to think about these kinds of huge numbers.

To estimate the cost of a phosphorylation-dephosphorylation cycle, we need to now get explicit about the reactions themselves. Here, we follow the development of Hong Qian's excellent 2007 piece entitled, "Phosphorylation energy hypothesis: open chemical systems and their biological functions," which shows how to compute the fluxes in these cycles, and from those fluxes, their cost. We borrow some of these calculations verbatim from *The Restless Cell* by Hueschen and Phillips. Unfortunately, none of these earlier treatments attempted to put in any numbers, which is a shortcoming we try to remedy here. To that

end, Figure 2.10 shows a hypothetical enzyme E that can also exist in a phosphorylated state E^*. We can think of this nonequilibrium steady state as arising from two coupled reactions given by

$$E + ATP \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} E^* + ADP \tag{2.18}$$

and

$$E^* \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} E + P_i. \tag{2.19}$$

The rate equations describing that process are given by

$$\frac{d[E]}{dt} = -\frac{d\,[E^*]}{dt} = -k_1[E][ATP] + k_{-1}\,[E^*]\,[ADP] + k_2\,[E^*] - k_{-2}[P_i][E], \tag{2.20}$$

where $[E]$ is the concentration of enzyme that is not phosphorylated, $[E^*]$ is the concentration of phosphorylated enzyme, $[ATP]$ is the concentration of ATP, $[ADP]$ is the concentration of ADP and $[P_i]$ is the concentration of phosphate. We have read this equation directly off the two kinetic equations describing the dynamics of interconversion between E and E^*.

We can use eqn. 2.20 to find the steady-state values of $[E]$ and $[E^*]$ by setting the left side equal to zero and using the constraint that $[E] + [E^*] = [E_{tot}]$. Following this strategy, we find

$$[E]^{ss} = \frac{k_{-1}[ADP] + k_2}{k_1[ATP] + k_{-2}[P_i] + k_{-1}[ADP] + k_2}[E_{tot}] \tag{2.21}$$

and

$$[E^*]^{ss} = \frac{k_1[ATP] + k_{-2}[P_i]}{k_1[ATP] + k_{-2}[P_i] + k_{-1}[ADP] + k_2}[E_{tot}], \tag{2.22}$$

where we have introduced the superscript $[E^*]^{ss}$ notation to indicate steady state. Intuitively, the qualitative behavior of these steady-state concentrations make sense since we see that as the excess concentration of ATP is increased, this tilts the balance towards higher levels of

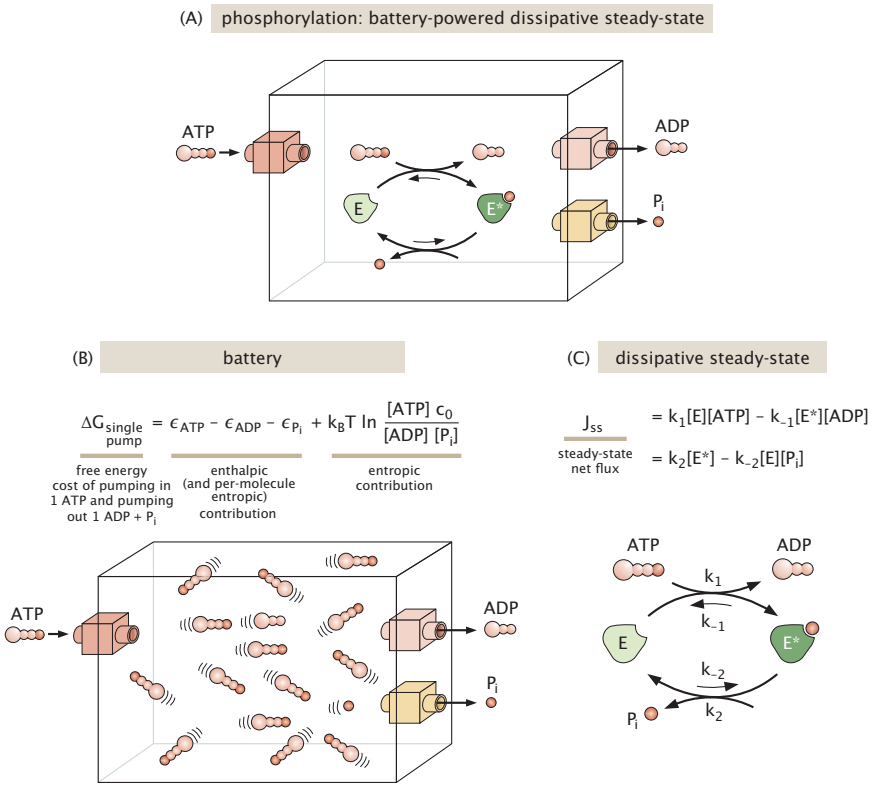


Figure 2.10: Nonequilibrium steady state for phosphorylation-dephosphorylation cycle. (A) Schematic of the chemical reactor system in which the phosphorylation and dephosphorylation of enzyme E takes place. The pumps at the system boundary represent the work that needs to be done to maintain the out of equilibrium concentrations of ATP, ADP and P_i. (B) The free energy cost of active maintenance. For every phosphorylation-dephosphorylation cycle, one ATP needs to be added to the system and one ADP and P_i needs to be removed. (C) Evaluation of the flux in the phosphorylation-dephosphorylation cycle. Adapted from CL Hueschen and R Phillips, The Restless Cell, Princeton University Press, 2024.

E*. We can find the steady-state net flux passing through the reaction given in eqn. 2.18 as

$$J = k_1[ATP][E] - k_{-1}[ADP]\,[E^*]\,. \tag{2.23}$$

Using the steady-state values of $[E]$ and $[E^*]$ we find the steady-state flux as

$$J^{ss} = \frac{k_1 k_2[ATP] - k_{-1}k_{-2}[ADP][P_i]}{k_1[ATP] + k_{-2}[P_i] + k_{-1}[ADP] + k_2}. \tag{2.24}$$

The units of this quantity are $1/s$ and the interpretation is that J^{ss} tells us the net number of reactions per volume per unit time. The flux in reaction given by eqn. 2.19 is the same since the whole reaction network is in steady state (and it follows from the math too). Note that if the reaction were allowed to equilibrate, then $J^{ss} = 0$ implying the allied condition

$$\frac{[ADP]_{eq}[P_i]_{eq}}{[ATP]_{eq}} = \frac{k_1 k_2}{k_{-1}k_{-2}}. \tag{2.25}$$

Given the steady rhythm of phosphorylation of the proteins of the proteome, what can we say about its typical energetic costs? Figure 2.10 gives us one hint as to how to think about the cost of these events. It is convenient to think of the processes of the cell as being powered by a suite of different biological batteries. For photosynthetic cells, that battery is provided by the steady stream of photons impinging on the Earth, each of which is worth roughly $100\ k_BT$ of energy. Photosynthetic reactions produce organic matter such as sugars which harbor a huge amount of stored energy ($\approx 1200\ k_BT$ per sugar) which can in turn be converted into other kinds of biological batteries such as membrane potentials and ATP. As a reminder, we recall the rule of thumb that 1 ATP $\approx$ 20 k_BT.

To provide more everyday intuition for thinking about these biological batteries, it is helpful to think of one of the most underrated of human technologies, namely, refrigerators. First, remember that a refrigerator doesn't work unless we plug it into an

electric outlet in the wall. In the same way that the dynamics of photosynthesis and carbon fixation does not work if we are not "plugged into" the external energy source of the sun, the refrigerator doesn't work unless we are plugged into the power grid. In both cases, the energy source is "over there." That is, we don't need to concern ourselves with how that energy source works or how much it costs to maintain it, it is just there. For the refrigerator, there is a heat pump that exploits the external energy source to reduce the temperature. The second law of thermodynamics and the corresponding insights about the Carnot cycle tell us how much energy we need to extract from the external power source to achieve that cooling.

Figure 2.10 attempts to make the refrigerator analogy more concrete by considering a system coupled to ATP, ADP and P_i pumps. In the same way that a refrigerator pumps heat, we can think of the reactions within a cell that keep a fixed ratio of ATP to ADP and P_i as mediated by pumps as shown in the figure. Every time one full phosphorylation-dephosphorylation cycle is completed, the pumps will pump in a single ATP and pump out an ADP and P_i with the result that the ratio of these molecules remains constant. Keeping those reservoirs at fixed values is the equivalent of keeping our refrigerator plugged into an outlet. But doing that pumping action comes with a free energy cost of its own. Our goal now is to work out how much energy is consumed for each such pumping event. Once we have that energy per cycle, we can think invoke our usual concept that $\partial F/\partial t = J\ \Delta\mu$, where $\Delta\mu$ is the cost per cycle and J is the number of such reactions per unit time.

In one of the most inspired and subtle acts of coarse-graining in the history of science, Josiah Willard Gibbs showed how we could capture the nearly infinite degrees of freedom with some reservoir in the simple idea of the chemical potential. The chemical potential is defined as the difference in free energy of the reservoir when it has N and $N-1$ molecules

of the species of interest. In the present context, the chemical potential of ATP is defined as

$$\mu_{ATP} = \Delta U_{ATP} - T\Delta S_{ATP} \tag{2.26}$$

where we have

$$\Delta U_{ATP} = N\varepsilon_{ATP} - (N-1)\varepsilon_{ATP} \tag{2.27}$$

and

$$\Delta S_{ATP} = k_B \ln \frac{W(N)}{W(N-1)}, \tag{2.28}$$

where $W(N)$ is the number of states of the reservoir when there are N ATP molecules present; ε_{ATP} is a phenomenological parameter that captures the energy associated with having a single ATP molecule in solution; k_B is Boltzmann's constant. The ΔS term accounts for the entropy change because during a hydrolysis event, the number of ATPs was reduced by one while the number of ADP and P_i was increased by one. Putting all of this together yields

$$\mu_{ATP} = \varepsilon_{ATP} + k_B T \ln \frac{[ATP]}{c_0}, \tag{2.29}$$

where c_0 is the standard state concentration, usually taken as 1 M. We have similar expressions for ADP and P_i, namely,

$$\mu_{ADP} = \varepsilon_{ADP} + k_B T \ln \frac{[ADP]}{c_0} \tag{2.30}$$

and

$$\mu_P = \varepsilon_P + k_B T \ln \frac{[P_i]}{c_0}. \tag{2.31}$$

These equations are entirely analogous to those we used in the previous section in describing the ions and sugars that were being transported into and out of the cell.

When reactions 1 and 2 both take place, the biochemical system has taken one complete

cycle. The free energy cost associated with one such reaction can therefore be reckoned by appealing to the chemical potentials of the ATP, ADP and P_i in the form

$$\Delta F_{\text{one cycle}} = \mu_{ATP} - \mu_{ADP} - \mu_P. \tag{2.32}$$

The reason for the different signs is that we have to add an ATP and remove and ADP and P_i, and the definition of the chemical potential therefore mandates these signs. Given the functional form of the chemical potentials derived above, we can rewrite this expression as

$$\Delta F_{\text{one cycle}} = \Delta\varepsilon + k_B T \ln \frac{[ATP]c_0}{[ADP][P_i]}. \tag{2.33}$$

When the system is in equilibrium, as we saw above, the net flux $J^{ss} = 0$ which corresponds to the condition

$$k_1 k_2 [ATP]_{eq} = k_{-1} k_{-2} [ADP]_{eq} [P_i]_{eq} \tag{2.34}$$

which can be rewritten as

$$\frac{[ADP]_{eq}[P_i]_{eq}}{[ATP]_{eq}} = \frac{k_1 k_2}{k_{-1} k_{-2}}. \tag{2.35}$$

Further, at equilibrium, we know that $\Delta F = 0$, permitting us to relate the equilibrium concentrations to $\Delta\varepsilon$. Specifically, in that case we have

$$\Delta\varepsilon = -k_B T \ln \frac{[ADP]_{eq}[P_i]_{eq}}{[ATP]_{eq}} = -k_B T \ln \frac{k_1 k_2}{k_{-1} k_{-2}}, \tag{2.36}$$

permitting us to rewrite the total free energy change in one full cycle as

$$\Delta F_{\text{one cycle}} = k_B T \ln \frac{k_1 k_2 [ATP]}{k_{-1} k_{-2} [ADP][P_i]} = k_B T \ln \frac{J_1 J_2}{J_{-1} J_{-2}}. \tag{2.37}$$

The result worked out here is a special case of a much more general result that can be

written in the form

$$\frac{\partial F}{\partial t} = k_B T \sum_i (J_i - J_{-i}) \ln \frac{J_i}{J_{-i}}, \tag{2.38}$$

which instructs us to sum over all the separate reactions. For our case of interest here, we have two reactions to sum over with the result that

$$\frac{\partial F}{\partial t} = k_B T \, (J_1 - J_{-1}) \ln \frac{J_1}{J_{-1}} + k_B T \, (J_2 - J_{-2}) \ln \frac{J_2}{J_{-2}}. \tag{2.39}$$

We can work out the separate contributions by appealing to our analysis above where we worked out the various fluxes, yielding

$$\ln \frac{J_1}{J_{-1}} = \ln \frac{k_1 [ATP][E]}{k_{-1}[ADP][E^*]} \tag{2.40}$$

and

$$\ln \frac{J_2}{J_{-2}} = \ln \ \frac{k_2 [E^*]}{k_{-2}[E][P_i]}. \tag{2.41}$$

We note that $J_1 - J_{-1} = J_2 - J_{-2}$, so that we can make the simplification

$$(J_1 - J_{-1}) \ln \frac{J_1}{J_{-1}} + (J_2 - J_{-2}) \ln \frac{J_2}{J_{-2}} = (J_1 - J_{-1}) \ln \frac{J_1 J_2}{J_{-1} J_{-2}}. \tag{2.42}$$

We see that

$$k_B T \ln \frac{J_1 J_2}{J_{-1} J_{-2}} = k_B T \ln \frac{k_1 k_2 [ATP]}{k_{-1} k_{-2} [ADP][P_i]}, \tag{2.43}$$

which is the free energy change in one cycle, while the prefactor $(J_1 - J_{-1})$ gives us the net number of cycles per second. We have found it exceedingly difficult to find measurements of the parameters k_1, k_{-1}, k_2 and k_{-2} and as a result, we attempt instead to use

$$\frac{\partial F}{\partial t} = k_B T \, (J_1 - J_{-1}) \ln \frac{J_1 J_2}{J_{-1} J_{-2}}. \tag{2.44}$$

If we take $J_1 = J_2 = 100\ \mathrm{s}^{-1}$ and make the conservative estimate that $J_{-1} = J_{-2} = 1\ \mathrm{s}^{-1}$, this tells us that

$$\frac{\partial F}{\partial t} = k_B T \times \frac{100}{s} \times \ln 10^4 \approx f \times 10\ \frac{\mathrm{ATP}}{\mathrm{s}}. \tag{2.45}$$

Note that this estimate is for a single protein. Clearly if every protein in the proteome were being phosphorylated at this rate, this would be a huge cost. Measurements and modeling performed in the 1980's on cyclic (de)phosphorylation cascades [30] report that typical operating conditions for such cascades demand at most of order 1 mM ATP/min, corresponding to at most a power per volume of few $\times\ 10^5$ ATP/μm^3 s. Clearly, there is far more to learn about the power associated with escaping genetic imprisonment through post-translational modifications.

2.3.3 Cost of sustaining vital membrane potential and osmobalance

Microbial cells build and sustain very different concentrations of ions than their environmental milieux—for instance intracellularly accumulating potassium and reducing sodium significantly relative to typical ambient concentrations [31]. As shown in Figure 2.11, if no active import and export processes were performed by the cells to sustain these concentration differences, ions would leak along their entropic incentive and these gradients would decay to equilibrium uniformity. While a rich array of fascinating enzymatic machinery orchestrates this pumping activity, and the battery of distinct molecular actors and feedbacks are extremely intricate and merit their own individual dissection, here we abstract some of this considerable complexity and formulate some basic fundamental lower bounds and phenomenological estimates on the dissipative costs needed to perform these functions. The concept in this first round of calculations is simply to get a feeling for the relevant orders of magnitude.

As shown in Figure 2.11, despite a very small permeability to ions, there is a leak current

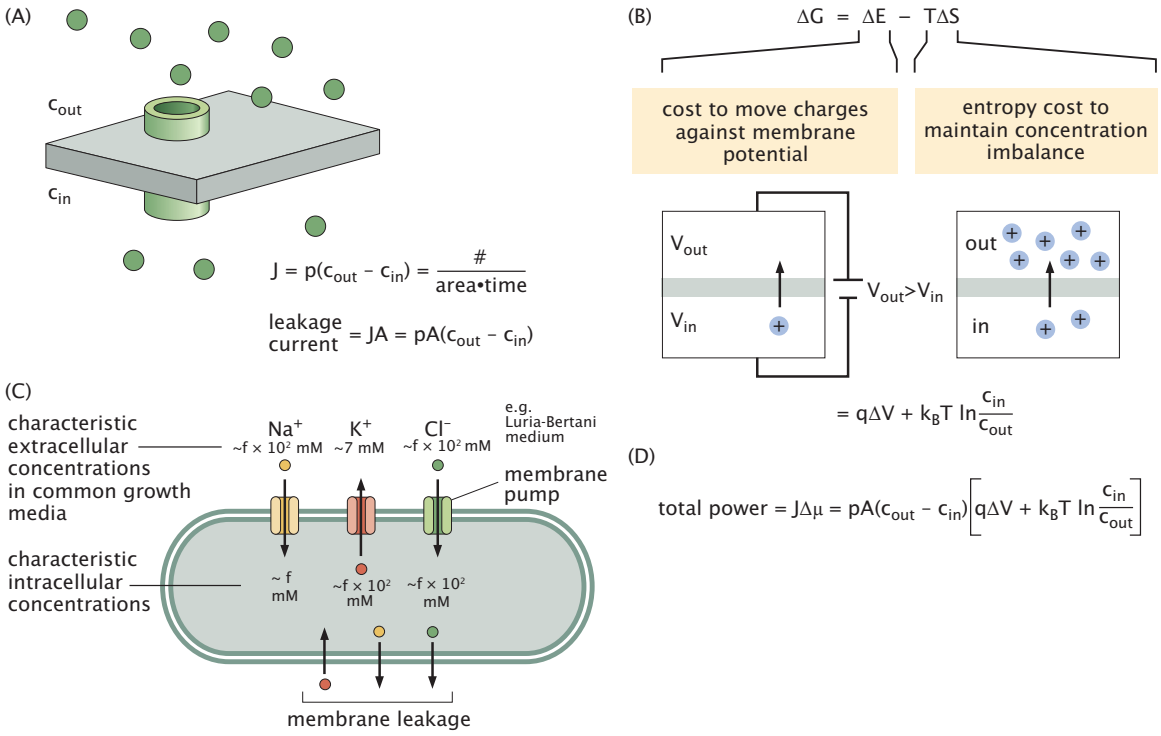


Figure 2.11: Maintaining gradients of ions in a cell requires active energy expenditure lower-bounded by a simple calculation.

of ions across lipid bilayer membranes. In particular, if we assume a typical permeability of $p = 10^{-4}$ nm/s $= 10^{-7}$ μm/s [32], and a concentration difference across the membrane of order 100 mM, then the flux is given by

$$J_{ion} = p\Delta c \approx 10 \frac{\text{ion}}{\mu\text{m}^2\,\text{s}}. \tag{2.46}$$

For a microbial cell with a surface area of f μm^2, this corresponds to a leak current of about $f \times 10$ ions/s. As a first sense of scale, the cost to transport one such ion up its gradient is roughly $\Delta G_{ion} = q\Delta V + k_B T \ln(c_{in}/c_{out}) \approx 1$ ATP, indicating that the power of this process is about $f \times 10$ ATP/s, a tiny fraction of the overall power of the cell in either maintenance or growth stages.

More precisely, as discussed by Kedem and Katchalsky [33] and others [32], for dilute

solutions neglecting the contribution of water fluxes[3], the change in chemical potential that accompanies a single solute molecule moving across a membrane from a concentration c_{out} to a concentration c_{in} is

$$\Delta\mu \equiv \mu_s^{\text{out}} - \mu_s^{\text{in}} \approx \frac{\bar{v}}{N_A}\Delta p + k_B T \ln \frac{c_{\text{out}}}{c_{\text{in}}} + q\Delta V, \tag{2.47}$$

where $\bar{v}$ is the partial molar volume; $\Delta p = p_{\text{out}} - p_{\text{in}}$ is the change in pressure; q is the solute's charge; ΔV is the cell membrane potential; and N_A and k_B are Avogadro's constant and Boltzmann's constant, respectively. Under nice conditions, a bacterial cell has a standard membrane potential of $\Delta V \in [70, 200]\ mV$ (BNID 103386; [35]). It is reasonable to take $\Delta V = 0.15\ V$ as a characteristic value in this range.

At this point, our plan is to examine each of the contributions to $\Delta\mu$ separately, but from the outset, treating the first term involving the pressure as negligible. Thus, we begin by examining the entropic contribution to the chemical potential. This term can be written as

$$\Delta\mu \approx k_B T \ln \frac{c_{\text{out}}}{c_{\text{in}}}. \tag{2.48}$$

The flux j of solute molecules expected to pass a membrane with permeability p (with apologies that the permeability has the same symbol as the pressure) against a concentration difference Δc is well approximated by a linear transport law of the form

$$j = -p\Delta c, \tag{2.49}$$

[3] In an ideal solution with composition c varying in space as $c(r)$, we can more generally write $\mu(r) = \mu_0 + k_B T \ln c(r)$. See e.g. Atkins & de Paula, *Physical Chemistry*, 9th Ed., pg. 164, eq 5.19 [34] for review of the basic ideal logarithmic variation.

so the total current dN/dt of solvents experienced across a membrane of area J is

$$\frac{dN}{dt} = Aj = -Ap\Delta c. \tag{2.50}$$

Thus, to resist this spontaneous leakage current across the membrane, at least an accompanying free energy dissipation rate dS/dt must be expended by the cell, given by,

$$\frac{dS}{dt} = \frac{dN}{dt}\Delta\mu \tag{2.51}$$

$$= -\boxed{Ap\,(c_{\text{out}} - c_{\text{in}}) \ln \frac{c_{\text{out}}}{c_{\text{in}}}\, k_B T}. \tag{2.52}$$

To illustrate the scale of this required dissipation rate to maintain a membrane potential, consider an ion with the following representative features:

- permeability, $p \sim 10^{-7}\ \mu m/s$: taken to near be that of potassium, see Reference [32].
- concentration difference: $c_{\text{out}} \sim$ few mM (in Luria-Bertain medium); $c_{\text{in}} \gtrsim (1-2) \times 10^2$ mM (potassium) (see BNID 115760) Thus, $\Delta c \sim 100\ mM = 10^8\ \mu m^{-3}$.
- cell membrane area, $A \sim$ few μm^2.

These numbers suggest a characteristic dissipation scale of about,

$$\left|\frac{dS}{dt}\right| \sim \frac{dN}{dt}\Delta\mu = \underbrace{\left(\text{few}\ \mu m^2\right)\,(10^{-7}\ \mu m/s)(10^8\ \mu m^{-3})}_{\text{few}\times 10\ \text{ions}/s} \underbrace{\ln \frac{150\ mM}{5\ mM}}_{\sim\text{few}}\ k_B T \tag{2.53}$$

$$\sim 100\ k_B T/s \tag{2.54}$$

$$\sim \boxed{5\ \text{ATP}/(\text{ion}\cdot s)}. \tag{2.55}$$

With this result for the entropic contribution to the free energy change associated with moving one ion across the membrane in hand, we now turn to the contribution in Eqn. 2.47 due to the membrane potential.

Quantifying the $\frac{dN}{dt} q\Delta V$ term

For an ion of charge $|q|= +1$ (electron/proton equivalents), such as for potassium, $q = 1.6 \times 10^{-19}$ Coulomb. This implies

$$q\Delta V \approx (0.15\ V)(1.6 \times 10^{-19}\ C) \tag{2.56}$$

$$\approx 2.4 \times 10^{-20}\ J \tag{2.57}$$

$$\approx 6\ k_B T/\text{ion}. \tag{2.58}$$

This is of the same order as the former $\Delta\mu = k_B T \ln(c_{\text{out}}/c_{\text{in}})$ term. To see that, we note that for a typical ratio $c_{out}/c_{in} \approx 100$ and recalling that $\ln 10 \approx 2.3$, we see that in this case $\Delta\mu = k_B T \ln\ 100 \approx\ 5\ k_B T$, quite in line with the voltage contribution to the chemical potential.

2.3.4 Experimental inference of ion pumping costs

The estimate we just reached, giving an approximate lower bound for the power cost to pump ions and maintain a concentration gradient in a cell, is wholly agnostic to how cells instantiate their balance of these ions using their real and nonideal biochemistry and ion channels. How does the actual power paid by cells to balance ions under environmental challenge relate to this lower bound?

How cells respond to osmotic provocations can be experimentally probed by adjusting the concentration of a salt like sodium chloride in the medium and watching cells metabolize or die differently. Here we turn to a revealing set of such experiments from Schink, Basan & coworkers on *E. coli*.

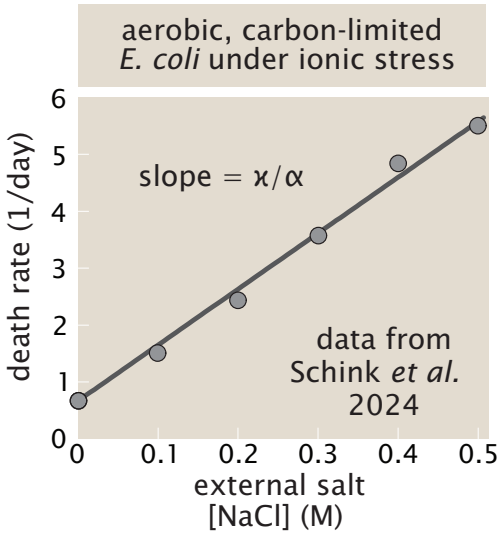


Figure 2.12: Rate at which *E. coli* populations die (aerobically under carbon limitation), as a function of the concentration of sodium chloride in the external medium. (Adapted from Schink *et al.* [8].)

Schink *et al.* report that starving populations of *E. coli* die faster when their medium has more sodium chloride (relative to a typical reference condition). Specifically, as shown by their measurements reproduced in Figure 2.12, the per-capita death rate $\gamma \equiv \frac{1}{N}\frac{dN}{dt}$ of these cells[4] increases remarkably linearly with the ambient concentration of sodium chloride. Under these (carbon-limited, aerobic) culture conditions, Schink and coworkers also argue that the basal metabolic rate β at which cells consume energy is proportional to this death rate, $\beta = \alpha\gamma$, where α is a fixed proportionality constant specifying the harvestable energetic content released by a dead cell. We alluded to this proportionality in §2.2 and will return to this idea even more forcefully in the next chapter. Accordingly, the linear variation of death rate with external salt level announced by measurements like those in Fig. 2.12 reports that the basal metabolic rate β also varies linearly with salt, namely,

$$\beta = \beta_0 + \kappa[\mathrm{NaCl}], \tag{2.59}$$

where β_0 is the basal metabolic rate at some (negligible) reference level of salt, and κ represents the additional power cost incurred by the cell per increment in sodium chloride concentration past this reference level.

Before theoretically understanding and contrasting this relationship Eq. 2.59's behavior to our earlier theoretical estimate of Eq. 2.52, we take it as a phenomenological fact and immediately probe its numerical implications. Schink *et al.*'s analysis stopped short of calculating or reporting κ, the increase in metabolic activity demanded per additional increment of salt concentration, but the quantities they otherwise report suffice to yield it. Specifically, the slope of the line between death rate and salt in Figure 2.12 gives the value of κ/α, so invoking a physiological value of α can deliver a value of κ. To estimate α, we use a condition where both β and γ under the same conditions were reported or can be

[4] As described further in their methods sections, Schink *et al.* performed these measurements in N-C-minimal medium [8].

estimated, since $\gamma = \beta/\alpha$. Specifically, Schink *et al.* report that cells consume energy at a rate $\beta \sim 0.6$ fmol glucose/cell/day (see [19], Extended Data Figure 10(B)), which we may express as 10^5 ATP/cell/s[5], in a particular and characteristic medium condition. Under these ambient conditions (43 mM sodium chloride), the trend of Figure 2.12 reports a death rate $\gamma \sim 1$ day$^{-1} = 1.2 \times 10^{-5}\ s^{-1}$. This implies that,

$$\alpha = \frac{\beta}{\gamma} \approx \frac{10^5 \text{ ATP/cell/s}}{1.2 \times 10^{-5}\ s^{-1}} \sim \boxed{8 \times 10^9 \text{ ATP/dead cell}}. \tag{2.62}$$

We now briefly pause to subject this estimated number for α to a smell test. Does this value make sense, if it is to represent the energy effectively harvestable from a dead cell? To ask whether this is plausible, we recall that the energy required to *make* a fresh *E. coli* cell is of order few$\times 10^{10}$ ATP. This building cost can be called to mind quickly, by remembering that while growing exponentially at a fast doubling time of order $\tau_2 \sim 20-30$ min $=$ few$\times 10^3\ s$, a very good rule of thumb is that cells are spending of order $\rho \sim 10^7$ ATP/s/μm^3 [38, 39], immediately giving a total energy expenditure of $\rho\tau_2 = (1-\text{few}) \times 10^{10}$ ATP to build a cell. Accordingly, the fact that our estimated number suggests a dead cell releases a few times less energy than a fresh one requires to be made is eminently reasonable. Indeed, in other studies, Schink and coworkers measured the relative recycling yield delivered by dead cells compared to the abundance when alive, and also found by distinct and explicit experiments that about a few to ten dead cells give the energy required to build a new cell [19]. This scale—that (under aerobic carbon limitation) dead cells yield a few times less energetic content in recyclable

[5] To understand the inferred metabolic expenditure of *E. coli* in units of ATP/cell/s (permitting comparison to other meaningful dissipative numbers at hand), we recall that the ideal, complete oxidation of 1 molecule of glucose generates the energetic equivalent of 26 ATP molecules (BNID 114969; [36, 37]). This gives,

$$\beta \sim 0.6\ \frac{\text{fmol glu}}{\text{CFU day}} \times \frac{10^{-15}\text{ mol}}{\text{fmol}} \frac{26\text{ mol ATP}}{1\text{ mol glu}} \frac{6 \times 10^{23}\text{ ATP}}{\text{mol}} \frac{1\text{ day}}{86400\ s} \tag{2.60}$$

$$\sim \boxed{10^5 \text{ ATP/s/cell}}. \tag{2.61}$$

necromass than a new cell needs—is an interesting and evocative bioenergetic rule of thumb in its own right.

Armed with this plausible value for the parameter α, we now at last estimate our desired value of κ setting the metabolic burden per an increment in external sodium chloride concentration. Consulting Figure 2.12 reveals that total death rate changes from less than one day^{-1} to close to six day^{-1} over a range of 500 mM external sodium chloride, or a slope of

$$\kappa/\alpha \approx \frac{5 \text{ day}^{-1}}{500 \text{ mM NaCl}} = 10^{-2}\text{day}^{-1}/(\text{mM NaCl}) \tag{2.63}$$

$$= 1.2 \times 10^{-7}\ s^{-1}/(\text{mM NaCl}). \tag{2.64}$$

Combining our estimates gives

$$\kappa = \frac{\kappa}{\alpha} \times \alpha \tag{2.65}$$

$$\approx \left(8 \times 10^{9} \text{ ATP}\right)\left(1.2 \times 10^{-7}\ s^{-1}/(\text{mM NaCl})\right) \tag{2.66}$$

$$\sim \boxed{10^{3} \text{ ATP/s/cell/(mM NaCl)}}. \tag{2.67}$$

This characteristic scale for κ announces, with huge quantitative drama, the metabolic plasticity demanded of microbes like *E. coli* in their variable environments. Consider the scale of fluctuations in ambient salt that cells easily experience. For instance, if one of the human tears of joy shed while writing this thesis fell into one of the 96-well plates hosting cells that inspired us to write it, the unlucky well could experience a $\Delta c = 10$ mM boost in the concentration of sodium ions[6]; our estimated value of κ anticipates *E. coli* would increase its metabolic rate by a considerable $\kappa\Delta c \sim 10^{4}$ ATP/s/cell. For context, such an

[6]Photometry measurements report human tears typically have sodium ion concentrations of $c \sim 150$ mEq/L = 150 mM [40], and since teardrops have volumes of a few microliters, eg $V \sim 7\mu L$ [41], implying that a typical teardop offers about 1 micromol of sodium; if added to a well with a typical working volume of $\sim 100\ \mu L$, this anticipates a boost of 10 mM sodium, if e.g. there were negligible sodium to begin with.

increase is about a tenth of the starting reference metabolic rate at 43 mM sodium chloride (see Eq. 2.61), but of the order of the *total* metabolic rate of *P. aeruginosa* directly measured by Ciemniecki and colleagues in the Newman lab [42] using extracellular phenazine cycling. Clearly much more dramatic ionic fluctuations could shape a life of *E. coli* in soils or human guts. Our inferred value for κ predicts that these environmental fluctuations may induce substantial, coupled energetic demands on struggling bacteria as well.

However, intriguingly, the metabolic costs incurred under different conditions or by other species may be substantially different. For instance, just to foreshadow a more rigorous discussion, our analysis of new measurements performed by collaborators in the Newman laboratory at Caltech appear to resolve a much milder variation of *Pseudomonas aeruginosa*'s metabolic rate with ambient sodium chloride than here with *E. coli*. We discuss this evidence in further detail in the succeeding chapter.

Where does the mathematical form and numerical sensitivity of a cell's metabolic response to salt come from? Should Eq. 2.59's linear relationship between total metabolic rate β and external salt concentration be regarded mostly phenomenologically, e.g. as a Taylor expansion to linear order of some more complicated physiological dependence $\beta \equiv \beta([\mathrm{NaCl}])$? Even under this generic interpretation, κ serves as a useful governing effective parameter that concisely summarizes how sensitive metabolism is to the insult of external salt. Indeed, Schink and colleagues motivate this relationship literally by positing the total metabolic rate β is dominated by active pumping against a diffusive flux $P\Delta c$ for fixed permeability P (which subsumes the cell membrane's area over which the leaks can happen) with a concentration difference Δc, where the ion cost is κ. However, we search for more explicit biophysical insight into this question by returning to contrast the empirically linear variation of death (thus metabolic) rate with external salt, Eq. 2.59, with

our earlier coarse biophysical estimate, Eq. 2.52, which predicts that,

$$\text{power cost to pump: } \frac{dS}{dt} = \Delta\mu(pA\Delta c) \tag{2.68}$$

$$= \left(k_B T \ln \frac{c_{\text{out}}}{c_{\text{in}}}\right) pA\,(c_{\text{out}} - c_{\text{in}}) \tag{2.69}$$

$$= k_B T A p\,(c_{\text{out}} - c_{\text{in}}) \ln \frac{c_{\text{out}}}{c_{\text{in}}}. \tag{2.70}$$

On its surface, this prediction appears to contradict the simple linear scaling of power with external ion concentration of Eq. 2.59. That is, unlike the literal invocations by Schink and co, the cost κ to pump an ion out of the cytoplasm should manifestly vary with the present balance of inner and outer concentrations, specified by the chemical potential. However, examining two limiting and intentionally extreme cases of ionic regulation gives some clues about how to mathematically reconcile these pictures.

First, imagine that cells maintain their cytoplasmic concentration of ions c_{in} at a fixed level despite any variation in the external levels c_{out}. This evokes a scenario where cytoplasmic functionality is extremely sensitive to the homeostasis of particular ions, compelling cells to exert stringent control to precisely clamp an ionic concentration inside the cell. [7] For evocative illustration, consider the finding that *E. coli* tightly can maintain cytoplasmic sodium ion concentration to a level within $\sim 5 - 14$ mM Na^+ while subject to external sodium concentrations between 1-85 mM sodium in specific culture conditions [43], illustrating a partial homeostasis. In this simple case, the predicted power cost to pump as the external concentration varies is often only mildly nonlinear over experimentally relevant regimes of external salt concentration; for instance, Figure 2.13 illustrates the relative pumping cost predicted by Eq. 2.70 when the cell fixes its cytoplasmic concentration to 5 mM despite an external salt concentration variation up to

[7] Imagining that cytoplasmic ion concentration c_{in} is constant with respect to even a ten-fold variation in $c_{\text{out}} \gtrsim c_{\text{in}}$ will *not* necessarily be reasonable for all bacterial strains.

500 mM. Note that if the ambient concentration c_{out} is not so different than the cytoplasmic concentration c_{in}, then a Taylor expansion anticipates that the cost varies quadratically in c_{out}. This behavior can be seen by writing the logarithm of the ratio in the pumping cost as

$$\ln \frac{c_{\text{out}}}{c_{\text{in}}} = \ln \frac{c_{\text{in}} + (c_{\text{out}} - c_{\text{in}})}{c_{\text{in}}} = \ln \left(1 + \frac{\Delta c}{c_{\text{in}}}\right) \tag{2.71}$$

$$\approx \frac{\Delta c}{c_{\text{in}}}, \tag{2.72}$$

so $\frac{dS}{dt} \sim (\Delta c)^2 / c_{\text{in}}$. Even more usefully, Figure 2.13 shows that over large typical concentration ranges the true nonlinearity of the pumping cost is mild, and for example a line of average slope over a tested domain,

$$\langle \text{slope} \rangle = k_B T \, p \, A \, \ln\left(\frac{c_{\text{out}}^{\text{max}}}{c_{\text{in}}}\right). \tag{2.73}$$

that reproduces the average slope from $c_{\text{out}} = 0$ to a maximum concentration tested $c_{\text{out}} = c_{\text{out}}^{\text{max}}$) often adequately describes the variation of the predicted pumping cost with external salt concentration. In summary, if the inner concentration of an ion is clamped to some fixed value and the external concentration of ion is varied over a wide range relative to this internal value (while the cell manifests a constant permeability and other properties), the corrections to a linear behavior (with average slope scaling with the permeability of the cell and the area of the cell) should be modest.

Next, we consider a second separate physiological extreme, where the inner concentration of ion is not fixed to a literal constant, but pinned to some multiplicative factor of the external concentration, namely $\frac{c_{\text{in}}}{c_{\text{out}}} := f$ for some fixed multiplicative factor f. (This idea abstractly stylizes some classic observations made on *P. aueriginosa* where sodium concentrations inside cells were reported to be $\sim$ 90% of the external medium up

to hundreds of millimolar in external salt [44].) In this case, the predicted cost Eq. 2.70 can be written as,

$$\frac{dS}{dt} = k_B T A p \, (c_{\text{out}} - c_{\text{in}}) \ln \frac{c_{\text{out}}}{c_{\text{in}}} \tag{2.74}$$

$$= k_B T A p \; c_{\text{out}} (1 - f) \ln f \tag{2.75}$$

$$= \underbrace{(k_B T A p (1 - f) \ln f)}_{\equiv k} \; c_{\text{out}}. \tag{2.76}$$

Here we see that the metabolic cost is exactly linear in the external salt—without any further approximation required—with the interpretation of the slope k as something proportional to $k_B T A p$ and depending (gently) on the value of the maintained target fraction f.

Thus in either physiological extreme of these two examples, we conclude that the characteristic slope relating pumping cost and external salt concentration should enjoy the interpretation of being nearly linearly proportional to $k_B T A p$. Of course, the real dynamics of ionic regulation in cells is much more complex than acknowledged by these toy scenarios. Specifically, for instance, the inner concentration c_{in} is probably set by a much more intricate web of feedbacks than mere constancy or proportionality to ambient salt.

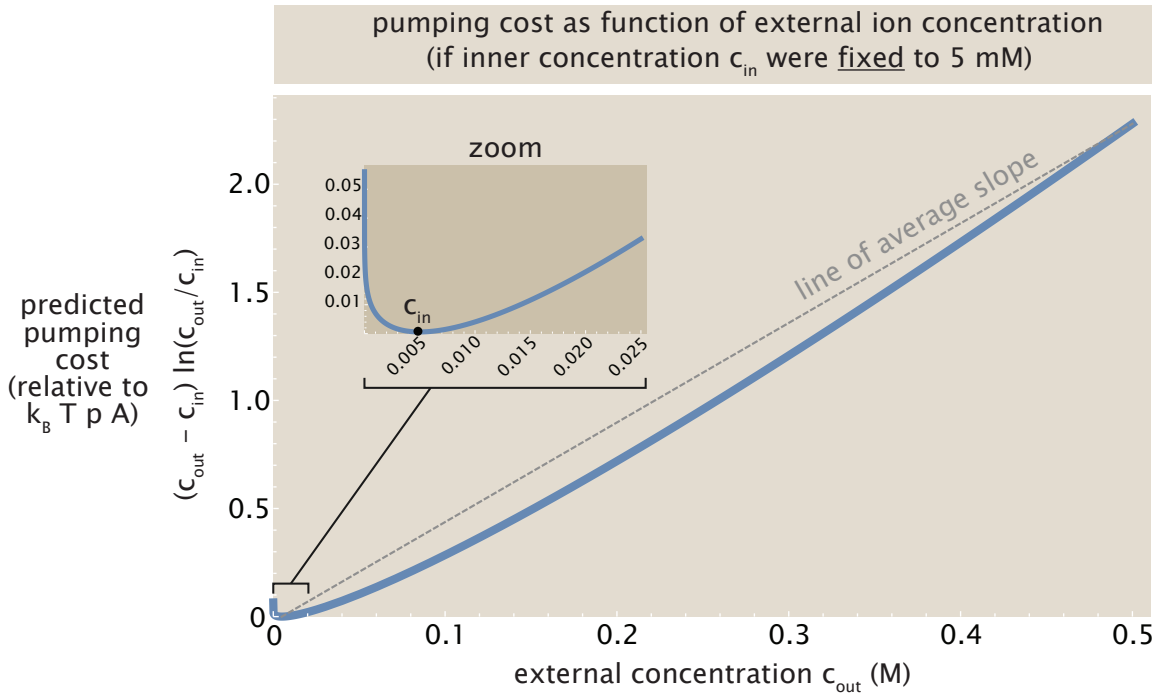


Figure 2.13: Predicted cost of pumping ions as function of the external medium concentration c_{out} up to 500 mM, assuming that cells fix cytosolic concentrations to c_{in} at a value of 5 mM, according to Eq. 2.70. (This power cost is reported in units of $k_B T p A$ on the y-axis.) Notice that the curvature of this function over this range is not dramatic, as judged by only small bowing relative to the average slope over the domain that varies external concentration to $10^3\times$ the fixed inner concentration. Inset: the cost to pump has the intuitive minimum of zero cost at $c_{\text{in}} = c_{\text{out}}$, about which the cost function is approximately quadratic.

2.3.5 Cost of cellular motility

Another process highlighted in Figure 2.7 that is shared by many microbes is that of motility. For the purposes of the present estimates, we consider the case of swimming motility powered by the motion of exterior flagella, though there are a number of other flagellum based motilities. In this section, we explore two different ways of assessing the energetic cost and power of cellular motility. First, we consider the fact that swimming organisms must overcome fluid drag. Second, we consider the power consumption of the molecular machines (e.g. the bacterial flagellar motor) that drive motion through viscous fluids.

Before embarking on these estimates, we remark that whether cells move (and so spend accompanying mechanical energy to power their flagella) depends sensitively on their environment. Specifically, many microbes on the severe energetic brink, including in senescent biofilms, do not appear to activate their flagella in a number of natural conditions for a large fraction of their lifetimes. This contextual dependence implies that there is a number of maintenance lifestyles where the mechanical costs of motility are not especially imposed on cells.

The Cost of Overcoming Drag

Whether giant whales migrating across oceans, long distance flying migrations such as that of the bar tailed godwit which travels nonstop between Alaska and New Zealand or a swimming bacterium, in all of these cases, the living organism must overcome the forces of viscous drag. But there the similarity ends. There is roughly a trillion-fold difference in the Reynolds number in these distinct situations, with the whale in a regime of extremely high Reynolds number and the bacterium in the opposite very low Reynolds number regime. As

a result, the drag laws that characterize these two cases are extremely different. To get a feeling for the forces involved in the low Reynolds number setting, we imagine a spherical bacterium of radius a which is met with a drag force known as Stokes drag and described by

$$F_{drag} = 6\pi\eta a v, \tag{2.77}$$

where η is the viscosity and v is the velocity of the swimming bacterium. (We note that in general rod-shaped bacteria would experience lesser drags, but such differences in geometry usually matter only within factors of few for typical cellular aspect ratios.) We can make a naive estimate of the power associated with overcoming this mechanical drag by multiplying the drag force and a typical velocity for *E. coli*, namely $v \sim 35\ \mu\text{m}/s$ [45], resulting in

$$\text{power of swimming} = 6\pi\eta a v^2 \tag{2.78}$$

$$\approx 20 \times (10^{-3}\ \text{Pa}\cdot\text{s}) \times (10^{-6}\ \text{m}) \times (f \times 10^{-5}\ \text{m/s})^2 \tag{2.79}$$

$$\approx f \times 10^2\ ATP/s. \tag{2.80}$$

This is a very interesting number that shows that the cost of overcoming fluid drag is small in comparison with many of the other much larger powers showcased in Figure 2.7. However, there is more to consider.

The Cost of Running the Flagellar Motors

The swimming motion of a bacterium such as *E. coli* ultimately derives from the rotation of flagella. And each such rotation has a cost, regardless of how much fluid drag the bacterium needs to overcome. In this section, we attempt to estimate this cost for the purposes of comparing it to the cost of overcoming the fluid drag and the many other cellular processes

highlighted in Figure 2.7.

Experiments on swimming in *E. coli* tell us that for every rotation of the flagellar motors that power swimming, the cost is roughly 1200 protons per revolution, and the motors are rotating at a speed of 10 Hz [46]. Further, we will assume that there are multiple such motors. As usual, we turn to our strategy of computing the power as

$$\text{power of flagellar motors} = J_{motors} \times \Delta\mu_{H^+}, \tag{2.81}$$

where J_{motors} is really a statement of how many rotations per second total occur and $\Delta\mu_{H+}$ is the free energy cost associated with the $H+$ ions moving across the membrane to power the motor. To maintain the hydrogen ion biological battery, those $H+$ ions needed to be pumped back up their concentration gradient. We recall that the proton motive force is given by

$$\Delta\mu_{H+} = q\Delta V + k_B T \ln \frac{c_{out}}{c_{in}}, \tag{2.82}$$

where the first term represents the cost of transporting charge against the membrane potential and the second term is associated with the entropy change associated with transporting molecules against their concentration gradient. For characteristic scales in these problems, we estimate $\Delta\mu_{H+} \approx 10\ k_B T$. Hence, for one motor rotating at 10 Hz, this corresponds to a free energy of

$$\begin{aligned}
\text{flagellar motor power} &= J_{motors} \times \Delta\mu_{H^+} && (2.83)\\
&\approx 10\ \frac{\text{rotations}}{s^{-1}} \times 10^3\ \frac{H+}{\text{rotation}} \times \frac{10\ k_B T}{H+} && (2.84)\\
&\approx 10^5\ \frac{k_B T}{s} && (2.85)\\
&\approx 5 \times 10^3\ \frac{ATP}{s}. && (2.86)
\end{aligned}$$

Note that this is the power for only a single flagellar motor. In fact, there are often multiple

flagella and hence we can imagine that this power is another five-fold higher than this value, in the range of $f \times 10^4$ ATP/s. As often appears to be the case, the cost of running the machines of the cell often comes out much higher than the dissipative processes that they must act against. In this case, there is a 100-fold difference between the motor power and the mechanical power of swimming.

2.3.6 Cost of membrane synthesis

For one cell to become two, not only must the contents of the cellular interior be doubled, but so too must all of the surface components such as the membrane and the cell wall. If we imagine the bacterial cell membrane as having an area of something like $5\ \mu m^2$ and recognizing that there is both an inner and outer membrane, this gives a total area of $20\ \mu m^2$. However, given that roughly half of that area is taken up by membrane proteins, we estimate a total lipid area of $10\ \mu m^2$. If we now assume that the typical cross sectional area of a lipid molecule is in the range $1/4 - 1/2\ nm^2$, this implies that the number of lipids per bacterial cell is something like

$$\text{number of lipids} \approx \frac{10 \times 10^6\ \text{nm}^2}{1/4 - 1/2\ \text{nm}^2/\text{lipid}} \approx 2 - 4 \times 10^7\ \text{lipids}. \tag{2.87}$$

Let's now try to estimate the cost of synthesizing all of these lipids in the case of fast growing cells. Roughly, the cost is $f \times 10$ ATP equivalents per phospholipid. Obviously, there is a great diversity of different lipid types and the biochemical pathways for each such lipid are different. But the cost we assume here is a reasonable and conservative guess. Now, given that there are $f \times 10^7$ lipids, we can work out a power

$$\text{power of lipid synthesis} \approx \frac{(f \times 10^7\ \text{lipids}) \times (f \times 10\ \text{ATP/lipid})}{f \times 10^3\ \text{s}} \approx f \times 10^5\ \text{ATP/s}, \tag{2.88}$$

a significant power investment in comparison with protein synthesis, which is an order of magnitude higher. Note that our calculation, as evidenced by the denominator of $f \times 10^3$ s, reflects the fast growth regime with a division time of less than an hour. To pivot to the case of "maintenance metabolism" would imply knowing something about the rate of turnover of lipids for cells that are not growing at all. Here we are starved of the relevant facts and can only make educated guesses, though we hope that someone might point us to solid experimental facts. If we assume that instead of being replenished on the hour time scale that instead the turnover time is of order a day, this means that the power will be reduced by a factor of between $f \times 10$ and 100, to the much more modest power of between $f \times 10^3$ and 10^4 ATP/s.

A key point of this overall exercise is to take stock of what is already known experimentally and what new experiments should be done in order to refine our understanding of the unusual molecular census of cells in a state of "maintenance metabolism." As concerns the subject of lipids, we have the strong sense that there is so much more to be known since it is clear that cells "care" about their lipid composition since when that composition is perturbed, cells invoke steps to restore the previous lipid composition, which is a very complex mix. Even the phospholipids have a strange collection of tail lengths whose properties and functional origins remain, for now, unclear: Why are there so many diverse phospholipids? Why their particular lengths? These questions await substantive further inquiry.

2.4 Minimal maintenance power from empirical data in ocean sediment

Next, to complement our more mechanistic and indirect first-principles reasoning about dissipation rates above, we endeavor to estimate how environmental carbon data, interpreted

from the lens of a recent and successful kinetic theory, report on the basal power requirement cells need to survive in real environments.

In a recent tour de force, Prof. Dan Rothman built a beautiful model aiming to account for how quickly microbes degrade carbon in marine and sediment environments [11]. The model is built from assuming that increasingly rare fluctuations in energy available to microbes constrain the degradation of organic carbon. In brief, Rothman's model predicts power laws for the total abundance of carbon $g(t)$; the cell density per volume $N(t)$; and hence the cell-specific carbon degradation rate $b(t)$, all with precise, theoretically-ordained power law exponents in carbon age t. These predictions accord impressively with available data.

Crucially for our interest in quantifying maintenance metabolism, Rothman predicts that the cell-specific carbon oxidation rate $b(t)$ decreases with age until it reaches a minimum viable basal rate $b(t_*)$ at a critical carbon age t_*, after which this rate is constant, $b(t > t_*) \equiv b(t_*)$. Though Rothman's compelling model specifies all the ingredients needed to back-calculate the basal metabolic rate $b(t_*)$ from other numerical features, Rothman's paper stopped short of algebraically or numerically calculating the value demanded of the model given by internal consistency. Here we quickly solve for this basal metabolic rate. Next we evaluate it numerically using Rothman's identified scalings, and so ask what this model predicts as a type of consensus value for the transition to basal metabolism across microbes in the ocean in their natural environments.

As the rate dg/dt at which carbon abundance changes with age is attributable to microbes, specifically

$$\frac{dg}{dt} = -N(t)b(t) \tag{2.89}$$

by definition, this assertion amounts to an unambiguous quantitative prediction for the

value of basal power requirement $b(t_*)$. Rothman's model predicts that carbon ages as

$$g(t) = g_0 t^{-\alpha} \tag{2.90}$$

for a numerical exponent α (a value fixed by theory, and dependent on whether $t < t_*$ or $t > t_*$) and some prefactor g_0[8]. This implies that $dg/dt = -\alpha g_0 t^{-\alpha-1}$, or by the definitional equation Eq. 2.89,

$$g(t) = -\frac{t}{\alpha}\frac{dg}{dt} \tag{2.91}$$

$$= \frac{t}{\alpha} N(t) b(t), \tag{2.92}$$

so the basal power requirement rate is given by

$$b(t_*) = \boxed{\frac{\alpha}{t_*}\frac{g(t_*)}{N(t_*)}}. \tag{2.93}$$

[8]More precisely, to enforce continuity at the crossover time t_*, the condition insists that $\lim_{t\to t_*^+} g(t) = \lim_{t\to t_*^-} g(t)$, so we have $g(t \le t_*) = g_1 t^{-\alpha_1}$ and $g(t \ge t_*) = (g_0 t_*^{\alpha_2-\alpha_1}) t^{-\alpha_2}$. Rothman's model, based on fixed point equations follwoing from extreme value theory, predicts that $\alpha_1 = \exp[-\gamma]/2$ and $\alpha_2 = \exp[-\gamma]$, where $\gamma \approx 0.5772\ldots$ is the Euler-Mascheroni constant.

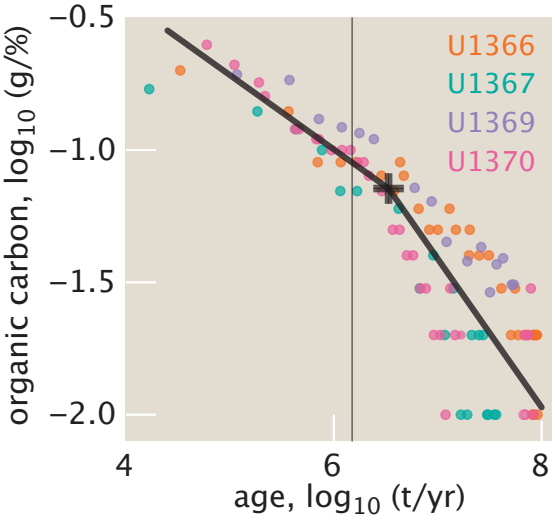


Figure 2.14: Fraction of organic carbon in sediment versus carbon age, as reported by deep marine sediment samples. The gray vertical line indicates the cross-over time $t_* \sim 1.5 \times 10^6$ yr taken by Rothman. Our further simple piecewise fit (indicated by the black line) to these carbon-abundance data $g(t)$ finds the cross-over time $t_*^g = (3.4 \pm 1.0) \times 10^6$ yr, emphasized by the black cross in the plot, within a factor few of Rothman's estimate.

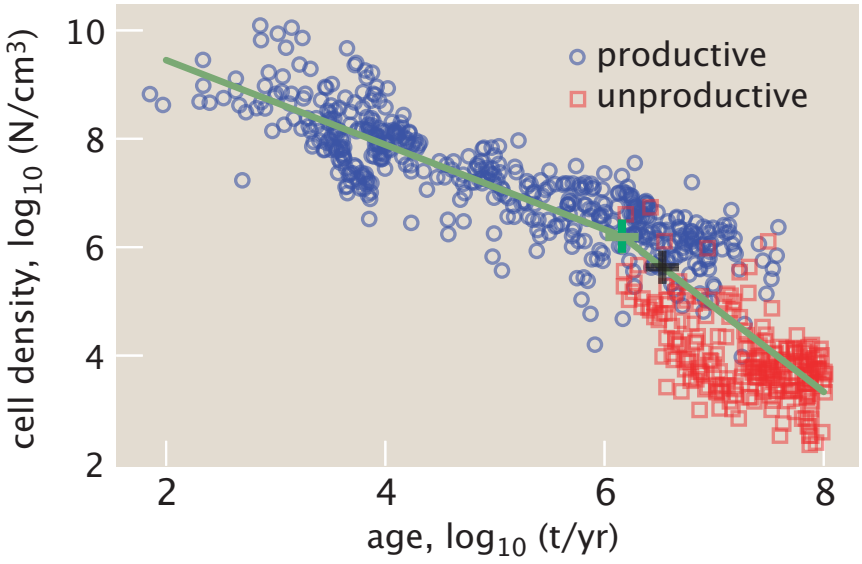


Figure 2.15: Cellular abundance in deep marine sediment versus local carbon age. Measurements from metabolically "productive" and "unproductive" marine sites are colored in blue and red, respectively A piecewise linear fit to these cell data $n(t)$ alone, indicated by the solid green line, predicts a cross-over time $t_*^n = (1.4 \pm 0.2) \times 10^6$ yr, emphasized by the green cross. For reference, the cross-over time t_*^g inferred earlier from the separate fit to the carbon abundance data $g(t)$ is also displayed, as a black cross.

From organic carbon $g(t)$ and cell abundance $N(t)$ versus age, Rothman approximately takes the cross-over time to be $t_* \sim 1.5 \times 10^6$ yr. We can read off the approximate values of the carbon abundance $g(t_*)$ and the cellular density $N(t_*)$ at the crossover time t_* from plots of data considered by Rothman; Figures 2.14 and 2.15 replot these data, along with our own simple piecewise linear fits to these profiles. With Rothman's original cross-over time $t_* \sim 1.5 \times 10^6$ yr, we observe $g(t_*) \sim 10^{-1}\%$ organic carbon (e.g. a fraction 1/1000th) and $N(t_*) \sim 10^6$ cells/cm^3. Finally, we must convert the carbon abundance fraction $g(t_*)$ reported in a mass-percent unit to an actual mass abundance. To do so, we take the density of sediment to be $\rho_s \sim 0.34$ g/cm^3, as reported by the original sources of data compiled by Rothman.

Thus,

$$g(t_*) \sim 10^{-3} \frac{\text{g organic carbon}}{\text{g sediment}} \times 0.34 \frac{\text{g sediment}}{\text{cm}^3} \tag{2.94}$$

$$\sim 3.4 \times 10^{-4} \frac{\text{g organic carbon}}{\text{cm}^3}. \tag{2.95}$$

Collectively these values predict,

$$b(t_*) = \frac{\exp[-\gamma]}{1.6 \times 10^6 \text{ yr} \times (\pi \times 10^7 \text{ s/yr})} \frac{10^{-4} \frac{\text{g organic carbon}}{\text{cm}^3}}{10^6 \text{ cells/cm}^3} \tag{2.96}$$

$$\sim 10^{-24} \text{ g organic carbon/cell/s} \tag{2.97}$$

$$\sim 3 \times 10^{-3} \text{ organic carbon molecules/cell/s}. \tag{2.98}$$

Clearly this is a very small uptake rate—it says that a cell waits an average of over five minutes between each uptake of an organic molecule.

Complete oxidation of one gram of glucose yields 1.6×10^4 J of energy; since $k_B T \sim$

$4.1 \times 10^{-21} J$ and 1 ATP $\sim 20 k_B T \sim 8 \times 10^{-20}\ J$,

$$b(t_*) \sim (10^{-24}\text{g organic carbon/cell/s}) \left(1.6 \times 10^4 J/(g\ C)\right) \left(\frac{1\ ATP}{8 \times 10^{-20}\ J}\right) \tag{2.99}$$

$$\sim \boxed{0.2\ \text{ATP/cell/s}}. \tag{2.100}$$

This is $\boxed{10^4}$-fold smaller than the measurement reported by Ciemniecki *et al.* on *Pseudomonas* [6].

We remark that the cross-over age taken above is close to slightly-more-fussily-inferred values. Specifically, we performed separate piecewise linear fits of carbon abundance and cell density to age, where the two allowed slopes are fixed to the theoretically predicted values and the only free parameter is the cross-over age under enforced continuity. On the basis of the carbon abundance data, we found a cross-over time $t_*^g = (3.4 \pm 1.0) \times 10^6$ yr, within a factor few of Rothman's cross-over time. On the basis of cell density data, we found $t_*^n = (1.4 \pm 0.2) \times 10^6$ yr, essentially the same as Rothman's cross-over time. Thus the cross-over behaviors are indeed likely close to those taken above, within a factor few.

2.5 Outlook

A flurry of foundational discoveries over the past decades have revealed a surprising regularity to how cells act while *growing fast*. Microbial physiology textbooks can now celebrate a set of impressively-general growth laws for the fastest growth rates [47], specific macromolecular compositions [42, 18], and even regulation of size and shape [48] of cells, at least for particular environments and species. This apparent universality in growth invites an optimism for discovering complementary *non*-growth laws that quantitatively account for how cells manage their metabolisms at the precipice of vitality. This setting of cells just surviving in challenging environments has urgent relevance across scales from

single-cell motility [10] or antibiotic persistence, to models of the global atmosphere and carbon cycle [11].

On the other hand, almost from the instant that fixed "maintenance rates" were conceptually and mathematically invoked, microbial physiologists have quarreled about their usefulness or biological interpretability as fixed numbers or mechanistic invariants; see the perspective posed by Tempest & Neijssel [49] forty years ago. In this chapter, to ground such discussion in concrete numbers and learn how they relate to underlying physiological processes, we endeavored an initial tour of how various total basal metabolism values quantitatively and philosophically differ. In many cases, our estimates of underlying physiological processes were separated by orders of magnitude from the purported total metabolic rates of the slowest cells in published data both old and new. These reckonings are approximate and subject to their necessary assumptions, required by a poverty of direct unambiguous measurements on features like membrane permeability, lipid turnover, protein abundance, and beyond specifically in the energy-challenged regimes of interest. Our analysis stress the urgency of more rigorous and cross comparative physiology in the energy-limited regime.

2.6 Acknowledgments

We gratefully acknowledge thoughtful and philosophically generative discussions with Steven Wilbert; Dongyang Li; John Marken; Fang Xiao; and Madhav Mani. We thank members of the Newman and Phillips laboratories for input. We thank the Center for Environmental Microbial Interactions (including a Pilot Grant to GLS) for funding that helped support and catalyze this collaborative work across labs.

Bibliography

[1] J. Monod. The growth of bacterial cultures. *Annu Rev Microbiol*, 3:371–394, 1949.

[2] J. Monod. Sur un phénomène nouveau de croissance complexe dans les culures bactériennes. *Comptes Rendus Des Séances de L'Académie Des Sciences*, 212:934–939, 1941.

[3] J. Monod. The Phenomenon of Enzymatic Adaptation - and Its Bearings on Problems of Genetics and Cellular Differentiation. *Growth*, 11(4):223–289, 1947.

[4] Frederick C Neidhardt. Bacterial growth: constant obsession with dn/dt. *Journal of Bacteriology*, 181(24):7405–7408, 1999.

[5] Øivind Tøien, John Blake, Dale M Edgar, Dennis A Grahn, H Craig Heller, and Brian M Barnes. Hibernation in black bears: independence of metabolic suppression from body temperature. *Science*, 331(6019):906–909, 2011.

[6] John A Ciemniecki, Chia-Lun Ho, Richard D Horak, Akihiro Okamoto, and Dianne K Newman. Mechanistic study of a low-power bacterial maintenance state using high-throughput electrochemistry. *Cell*, 187(24):6882–6895, 2024.

[7] Tori M Hoehler, Dylan J Mankel, Peter R Girguis, Thomas M McCollom, Nancy Y Kiang, and Bo Barker Jørgensen. The metabolic rate of the biosphere and its components. *Proceedings of the National Academy of Sciences*, 120(25):e2303764120, 2023.

[8] Severin Schink, Mark Polk, Edward Athaide, Avik Mukherjee, Constantin Ammar, Xili Liu, Seungeun Oh, Yu-Fang Chang, and Markus Basan. Survival dynamics of starving bacteria are determined by ion homeostasis that maintains plasmolysis. *Nature Physics*, 20(8):1332–1338, 2024.

[9] Yijie Deng, Douglas Raymond Beahm, Steven Ionov, and Rahul Sarpeshkar. Measuring and modeling energy and power consumption in living microbial cells with a synthetic atp reporter. *BMC biology*, 19(1):101, 2021.

[10] Tori M Hoehler and Bo Barker Jørgensen. Microbial life under extreme energy limitation. *Nature Reviews Microbiology*, 11(2):83–94, 2013.

[11] Daniel H Rothman. Slow closure of earth's carbon cycle. *Proceedings of the National Academy of Sciences*, 121(4):e2310998121, 2024.

[12] Aaron Novick and Leo Szilard. Description of the chemostat. *Science*, 112(2920):715–716, 1950.

[13] Aaron Novick and Leo Szilard. Experiments with the chemostat on spontaneous mutations of bacteria. *Proceedings of the National Academy of Sciences*, 36(12):708–719, 1950.

[14] Aaron Novick and Leo Szilard. Experiments on spontaneous and chemically induced mutations of bacteria growing in the chemostat. In *Cold Spring Harbor Symposia on Quantitative Biology*, volume 16, pages 337–343. Cold Spring Harbor Laboratory Press, 1951.

[15] S. J. Pirt. The maintenance energy of bacteria in growing cultures. *Proc R Soc Lond B Biol Sci*, 163(991):224–31, 1965.

[16] Michael Lynch and Georgi K Marinov. The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences*, 112(51):15690–15695, 2015.

[17] Gangsheng Wang, Melanie A Mayes, Lianhong Gu, and Christopher W Schadt. Representation of dormant and active microbial dynamics for ecosystem modeling. *PloS one*, 9(2):e89252, 2014.

[18] Matthew Scott, Carl W Gunderson, Eduard M Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330 (6007):1099–1102, 2010.

[19] Severin J Schink, Elena Biselli, Constantin Ammar, and Ulrich Gerland. Death rate of *E. coli* during starvation is set by maintenance cost and biomass recycling. *Cell systems*, 9(1):64–73, 2019.

[20] John DC Little et al. A proof for the queuing formula $l = \lambda w$. *Operations research*, 9(3): 383–387, 1961.

[21] Karthik P Jayapal, Siguang Sui, Robin J Philp, Yee-Jiun Kok, Miranda GS Yap, Timothy J Griffin, and Wei-Shou Hu. Multitagging proteomic strategy to estimate protein turnover rates in dynamic systems. *Journal of proteome research*, 9(5):2087–2097, 2010.

[22] Eran Eden, Naama Geva-Zatorsky, Irina Issaeva, Ariel Cohen, Erez Dekel, Tamar Danon, Lydia Cohen, Avi Mayo, and Uri Alon. Proteome half-life dynamics in living human cells. *Science*, 331(6018):764–768, 2011.

[23] Sanjoy Mahajan. *Street-Fighting Mathematics: The Art of Educated Guessing and Opportunistic Problem Solving*. MIT Press, Cambridge, Mass., 2010. ISBN 9780262514293 (pbk. alk. paper) 026251429X (pbk. alk. paper).

[24] Rohan Balakrishnan Miriam Abele Zhongge Zhang Chenhao Wu Lance Freiman Eray Enutson Kevin Corbett Christina Ludwig Matteo Mori Harish Kannan, Gabriel Mullin-Manzanarez and Terence Hwa. Bacteria leverage protein degradation to increase energy supply during starvation. APS March Meeting 2025, 2025. URL https://summit.aps.org/events/MAR-S71/8.

[25] George Stephanopoulos, Aristos A Aristidou, and Jens Nielsen. Metabolic engineering: Principles and methodologies. 1998.

[26] Ron Milo. What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays*, 35(12):1050–1055, 2013.

[27] James B Russell and Gregory M Cook. Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiological reviews*, 59(1):48–62, 1995.

[28] TD Lamb. Sources of noise in photoreceptor transduction. *Journal of the Optical Society of America A*, 4(12):2295–2300, 1987.

[29] Terrence Geiger and S Clarke. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. succinimide-linked reactions that contribute to protein degradation. *Journal of Biological Chemistry*, 262(2):785–794, 1987.

[30] E Shacter, P Boon Chock, and ER Stadtman. Energy consumption in a cyclic phosphorylation/dephosphorylation cascade. *Journal of Biological Chemistry*, 259(19):12260–12264, 1984.

[31] Ravish Sharma, Tomohiro Shimada, Vinod K Mishra, Suchitra Upreti, and Abhijit A Sardesai. Growth inhibition by external potassium of escherichia coli lacking ptsn (eiiantr) is caused by potassium limitation mediated by ycgo. *Journal of bacteriology*, 198(13):1868–1882, 2016.

[32] Rob Phillips. Membranes by the numbers. *Physics of biological membranes*, pages 73–105, 2018.

[33] Ora Kedem and Aharon Katchalsky. Thermodynamic analysis of the permeability of biological membranes to non-electrolytes. *Biochimica et biophysica Acta*, 27:229–246, 1958.

[34] P. Atkins and J. de Paula. *Atkins' Physical Chemistry*. OUP Oxford, 9th edition, 2010. ISBN 9780199543373. URL https://books.google.com/books?id=BV6cAQAAQBAJ.

[35] Quang Hon Tran and Gottfried Unden. Changes in the proton potential and the cellular energetics of escherichia coli during growth by aerobic and anaerobic respiration or by fermentation. *European journal of biochemistry*, 251(1-2):538–543, 1998.

[36] Ron Milo, Paul Jorgensen, Uri Moran, Griffin Weber, and Michael Springer. Bionumbers—the database of key numbers in molecular and cell biology. *Nucleic acids research*, 38(suppl_1): D750–D753, 2010.

[37] Christoph Kaleta, Sascha Schäuble, Ursula Rinas, and Stefan Schuster. Metabolic costs of amino acid and protein production in escherichia coli. *Biotechnology journal*, 8(9):1105–1114, 2013.

[38] Ian S Farmer and Colin W Jones. The energetics of*Escherichia coli*during aerobic growth in continuous culture. *European journal of biochemistry*, 67(1):115–122, 1976.

[39] R. Milo and R. Phillips. *Cell Biology by the Numbers*. Garland Press, New York, NY, 2016.

[40] Jørn Hess Thaysen and Niels A Thorn. Excretion of urea, sodium, potassium and chloride in human tears. *American Journal of Physiology-Legacy Content*, 178(1):160–164, 1954.

[41] S Mishima, A Gasset, SD Klyce, and JL Baum. Determination of tear volume and tear flow. *Investigative ophthalmology & visual science*, 5(3):264–276, 1966.

[42] Griffin Chure and Jonas Cremer. An optimal regulation of fluxes dictates microbial growth in and out of steady state. *Elife*, 12:e84878, 2023.

[43] Chien-Jung Lo, Mark C Leake, and Richard M Berry. Fluorescence measurement of intracellular sodium concentration in single *Escherichia coli* cells. *Biophysical journal*, 90(1):357–365, 2006.

[44] Tibor I Matula and Robert A Macleod. Penetration of *Pseudomonas aeruginosa* by sodium chloride and its relation to the mechanism of optical effects. *Journal of Bacteriology*, 100(1): 411–416, 1969.

[45] Graeme Lowe, Markus Meister, and Howard C Berg. Rapid rotation of flagellar bundles in swimming bacteria. *Nature*, 325(6105):637–640, 1987.

[46] Howard C Berg. The rotary motor of bacterial flagella. *Annual review of biochemistry*, 72(1): 19–54, 2003.

[47] Markus Basan, Sheng Hui, Hiroyuki Okano, Zhongge Zhang, Yang Shen, James R Williamson, and Terence Hwa. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature*, 528(7580):99–104, 2015.

[48] Leigh K Harris and Julie A Theriot. Relative rates of surface and volume synthesis set bacterial cell size. *Cell*, 165(6):1479–1492, 2016.

[49] David W Tempest and Oense M Neijssel. The status of yatp and maintenance energy as biologically interpretable phenomena. *Annual Review of Microbiology*, 38(1):459–513, 1984.

Chapter 3

Predictive theory for experiments on microbes on the energetic brink

Many cells—including microbes that power our planet's carbon cycle, nurture our agriculture, or devastate our human lungs—orchestrate their physiology under severe bioenergetic limitations. Such metabolically-constrained cells are not just slowed-down versions of the fast-growing cells well studied by most laboratory experiments. Instead, these cells on the metabolic brink are completely remodeled organisms and must follow distinct quantitative principles, tradeoffs, and strategies to survive [1, 2]. These extraordinary physiological states at energetic extremes give unique opportunities to identify critical consequences of the nonequilibrium energy expenditures that distinguish life from nonliving matter.

Careful experimental studies probing the physiology of energy-limited microbes are rare but reveal quantitative regularities for canonical perturbations like carbon deprivation [3]. On the other hand, there is far more to environmental challenges than carbon deprivation. In this chapter, we strive to make forceful links between experimentally tractable observables and the often quite different "theorist variables" actually underlying

energy-limited cells as dynamical systems. Our aim is to move slightly away from philosophizing grandly with cartoons and verbal statements. Here we seek to ask, what are the precise consequences of different mathematical and physiological universes on a spectrum of real measurable experimental quantities? Further, how can we learn from the inverse problem of how macroscopic phenomenology of abundance constrains the metabolic states and operations maintained by cells? We are inspired by unprecedented measurements on total metabolism, and rich accompanying trajectories of cell counts, by investigators in the laboratory of Dianne Newman and colleagues.

This chapter takes its journey in three parts. First, in §3.1, we ask about the phenomenology of death. Why do populations of dying cells drop so differently under oxygen starvation compared to carbon starvation? We give a simple quantitative origin story for these big macroscopic differences. Second, in §3.2, we pose humble kinetic pictures for the particulars of real experiments culturing cells with electrodes, and analyze these models' consequences. We hope these different experimentally-testable consequences might help rigorously distinguish between models that otherwise seem competitively credible. Third, in §3.3, we analyze pilot experiments on *Pseudomonas* in varied salt conditions. Amid some rich phenomenology, this allows us to preliminarily estimate a valuable characteristic parameter specifying the changes in metabolic rate demanded of cells in greater environmental salt conditions. We quantitatively contrast this parameter in *Pseudomonas* with an anlog we infer from very separate data in *E. coli*, uncovering striking differences.

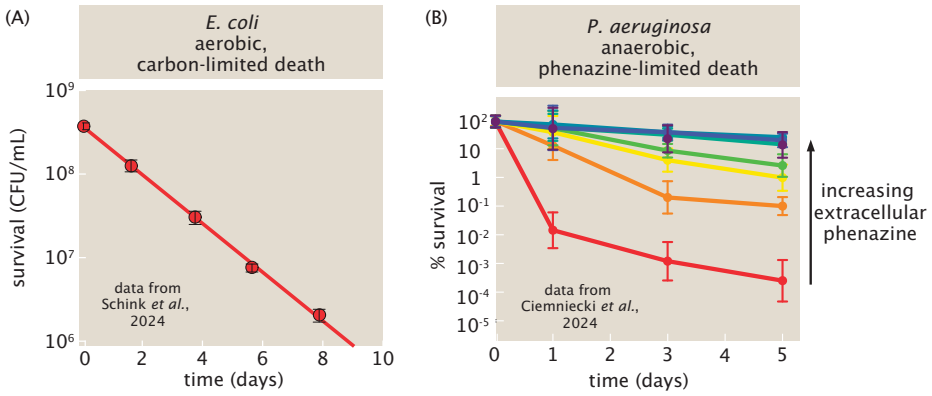


Figure 3.1: Empirical death curves for cells subjected to different deprivations. (A) *E. coli* cells deprived of carbon show a stereotyped death profile. (B) *P. aeruginosa* deprived of a terminal electron acceptor shows a markedly different death profile. Unlike the electron donor- (carbon-) limited growth case at left, the terminal electron acceptor case shows conspicuous changes to the per-capita death rate over time. Distinct color traces refer to different ambient concentrations of extracellular phenazine as a terminal electron acceptor. The carbon-limited data are from Schink *et al. Nat. Phys.* 2024 [4]; the phenazine-limited data are from Ciemniecki *et al.* 2024 [5].

3.1 Phenomenology of death differs under terminal e^- acceptor or donor deprivation

When starved of terminal electron acceptors, such as oxygen or phenazines, for example, cells die differently than when they are starved of electron donors such as carbon. This distinction manifests itself at a population level. As shown in Figure 3.1(A), when starved of carbon, viable cells often decline at constant exponential rates, showing linear declines in abundance measured on a log scale over time. In contrast, cells die at markedly time-varying rates when anaerobically starved of phenazine, making log abundance drop at slowing rates over time, as shown in Figure 3.1(B).

What physiology and dynamics of resource exchange with the environment account for these differences? Can these features further report on the underlying bioenergetic status of stressed cells, for example to eventually infer maintenance metabolic rates from population abundances alone (once calibrated by precise redox current measurements)? Towards these questions, here we formulate candidate models capturing the distinct

balances between cellular energy consumption and bioenergetic resource supply from the environment (or from other dying cells) underlying terminal electron acceptor or electron donor deprivation.

Our central contention for why these macroscopic phenomenologies of death differ when the electron donor is limiting, versus when the electron acceptor is limiting, is simply stated. When carbon is limiting, dying and lysing cells, which are made up of the very thing that is limiting the energetic state of the cells, can help feed and sustain the cells that remain alive. This dynamic is schematized in Figure 3.2(A), and is exactly the beautiful quantitative picture developed and quantitatively stress-tested by the experimental work of Schink, Gerland, Basan and coworkers for *E. coli* under carbon limitation [6, 7, 4]. In this picture, the energetic demand of an instantaneous abundance of cells is exactly balanced by a supply proportional to the current death rate, yielding the quantitative regularity of the per-capita death rate visible in survival curves like that in Figure 3.1(A). Specifically, if the limiting energy pool (putatively, set by carbon abundance) is called $E(t)$, then cells each consuming energy at a rate of β energy units per cell per time collectively deplete $E(t)$ at rate $-\beta N(t)$; but this is balanced by a resupply $-\alpha \frac{dN}{dt}$ from currently dying cells, where α is the energetic yield per dead cell (and we recall that death makes $dN/dt < 0$). Thus, we have the dynamical description of the energy supply of the form

$$\text{carbon limitation:} \quad \frac{dE}{dt} = \underbrace{-\beta N(t)}_{\substack{\text{viable cells consume} \\ \text{maintenance energy}}} \underbrace{-\alpha \frac{dN}{dt}}_{\substack{\text{dying cells} \\ \text{donate energy}}} . \tag{3.1}$$

Assuming that the energy reaches a steady-state (quickly relative to the longer timescales over which cells meaningfully change abundance), namely $dE/dt = 0$, implies that

$$\text{per-capita death rate} \equiv \frac{1}{N(t)} \frac{dN}{dt} = -\frac{\beta}{\alpha}. \tag{3.2}$$

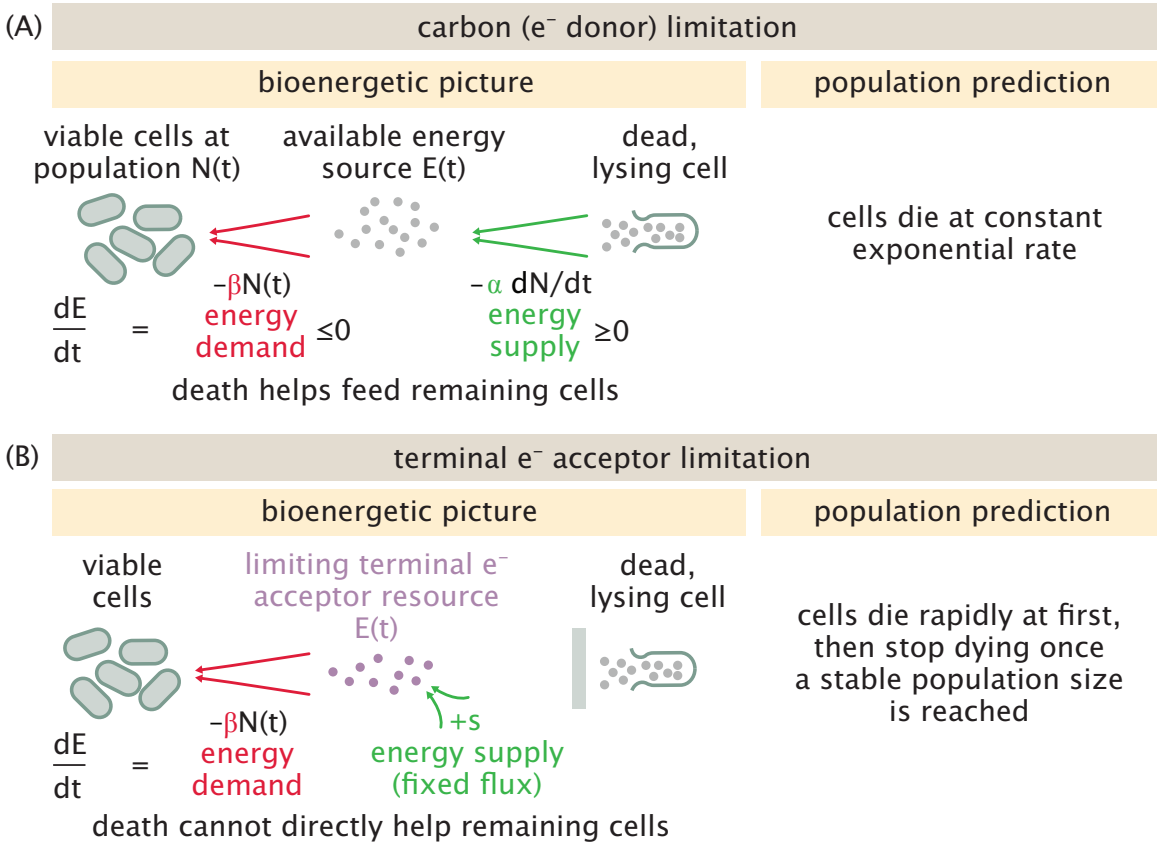


Figure 3.2: Distinct macroscopic physiology of death under deprivation of electron donor (e.g. carbon) versus deprivation of terminal electron acceptor (e.g. oxygen). (A) When cells are energy-limited by carbon, of which dying cells are made and can donate, viable cells can recycle nutrients from the "necromass" of their dying brethren and replenish their energy supply. The relevant energy source $E(t)$ changes in time due to consumption by cells at a rate $-\beta N(t)$ where β is the energy demand per time per cell, but replenished by dying cells at rate $\alpha\ dN/dt$ where α is the energetic content of a dead cell. (B) In contrast, when cells are limited by terminal electron acceptors (such as oxygen), dying cells do not meaningfully replenish the communal supply of the energetically-limiting resource. Instead a supply flux s fixed by environmental conditions participates in the energetic balance. Note that in these two cases, the putative nature of the limiting energetic resource has changed from the electron donor (grey dots in panel (A) to the electron acceptor (purple dots in panel (B)).)

This result gives the prediction that per-capita death rates should be constant in time, exactly the behavior affirmed by a straight line of abundance (measured on a log scale) versus time as shown in Figure 3.1(A). Since the slope of this death trajectory is proportional to the maintenance rate β at which bacteria consume energy, these curves can report precious information about the energetic status of cells. This conceit of inferring energetic status from death is useful and scalable, because a single viability experiment is relatively straightforward by the gymnastic standards of more complex microbiological measurements.

3.1.1 Generalizing energetic resource balance with cellular energy consumption under broader dynamics

When cells are limited by terminal electron acceptors such as in the absence of oxygen or molecules like oxidized phenazine, instead of carbon, we propose that dying cells do not substantially benefit the energetic state of their viable neighbors. This is because lysing cells are not generally significant sources of newly available terminal electron acceptors.[1] Instead, the supply of the limiting energetic resource is dictated by environmental boundary conditions distinct from dying cells, such as the flux of oxygen at a plane in a biofilm, or in the case of many recent experiments in the Newman laboratory [5], set by the action of the electrode reoxidizing extracellular phenazine. We illustrate this contrast in Figure 3.2(B).

This simple, foundational difference should drive very different macroscopic phenomenologies of death. Assume that the energetic pool $E(t)$ now interpreted differently as a pool largely set by terminal electron acceptor abundance, not carbon,

[1] At least, this statement appears appropriate for the externally-regulated extracellular redox carrier case of many of the experiments of interest in the Newman laboratory. Other physiological and environmental circumstances could change this picture and e.g. restore some dependence of the energetic supply term on dying cells.

changes in time as,

$$\text{terminal } e^- \text{ acceptor limitation: } \quad \frac{dE}{dt} = \underbrace{-\beta N(t)}_{\substack{\text{viable cells consume} \\ \text{maintenance energy}}} + \underbrace{s}_{\substack{\text{environment} \\ \text{supplies energy}}}, \tag{3.3}$$

where s is the rate at which energy is supplied by the environmental milieu.

Previously, in the carbon limitation case, dying cells replenished energetic resource at a rate $s = -\alpha \frac{dN}{dt}$. However, generically—and especially plausibly in the case of limited terminal electron acceptors—the supply rate will not just be proportional to the death rate. Instead, we observe that the supply rate s of the energetic resource resource can be some potentially physiologically or environmentally complex function of the instantaneous energetic resource $E(t)$; the instantaneous cell count $N(t)$; and/or the current rate of change of the cells dN/dt, namely,

$$\text{supply rate: } s \equiv s(N(t), E(t), dN/dt). \tag{3.4}$$

Immediate gifts from knowing energetic resource reaches steady-state

When the energy reaches a steady state $dE/dt = 0$[2], the coordinates $N(t), E(t)$, and dN/dt become subject to an implicit relation. That is, when the right side of Eq. 3.3 vanishes, the governing quantities s, N, E are constrained to depend on each other and jointly fall on some manifold specified by the shape of the supply rate function s, namely,

$$N(t) := \frac{s(N(t), E(t), dN/dt)}{\beta}. \tag{3.5}$$

[2]A steady-state in energy, $dE/dt \approx 0$, is often plausible over the timescales of the slower changes in the cell count $N(t)$. Specifically, for typical laboratory conditions of interest like those studied by Ciemniecki and coworkers [5] and Schink and coworkers [4], measurements typically report per-capita death rates on the order $\lesssim 1$/few per day; this timescale should be slow relative to the local equilibration of the molecular energy pool by uptake and metabolic reactions.

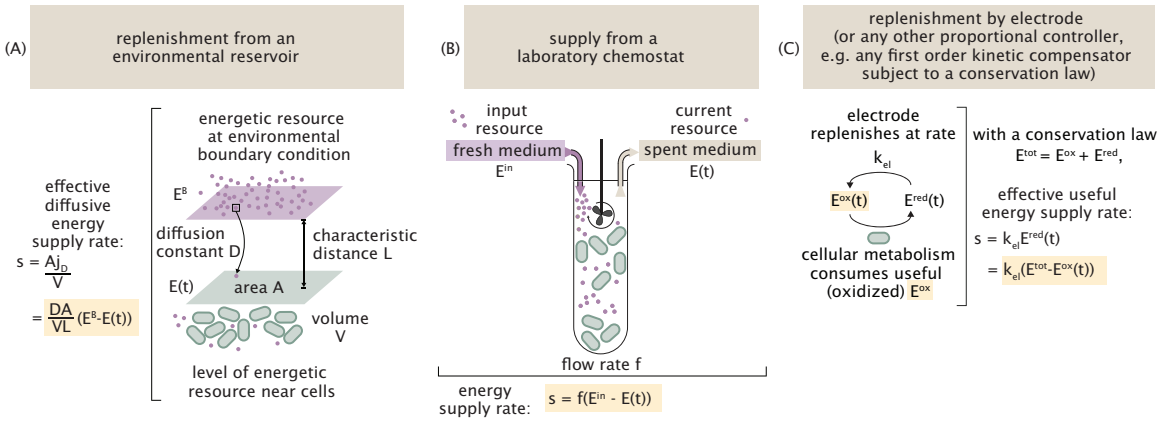


Figure 3.3: Diverse environmental and laboratory scenarios where energetic resources $E(t)$ are replenished at a supply rate s adopting the same linear mathematical form, $s(E(t)) = k(F - E(t))$, for some appropriate constants k, F. (A) If cells whose local abundance of energy is $E(t)$ are in contact with an environmental reservoir presenting a boundary condition of fresh energy resource at a higher concentration E^{B}, the typical diffusive current of resource to the body of cells can be modeled as $s = \frac{DA}{VL}(E^{\mathrm{B}} - E(t))$, where A and V are the characteristic area and volume of the cells; D is the diffusion coefficient of the energetic resource; and L is a characteristic distance between the cells and the environmental boundary condition. (B) A laboratory chemostat actively replenishing the abundance of energetic resources at current concentration $E(t)$ by flowing fresh medium at concentration E^{in} at flow rate f resupplies energy at a rate $s = f(E^{\mathrm{in}} - E(t))$. (C) When functional energetic resource is regenerated by a reaction mechanism with first-order kinetics in a depleted form of the resource, the supply rate of the functional form of the resource is proportional to the difference between the current abundance of functional resource and the total amount of resource. This example manifests in the setting of cells that need an energetic resource in its oxidized to accomplish metabolism; if an electrode reoxidizes reduced resource at rate k_{el}, then the supply term adopts the form $s = k_{\mathrm{el}}(E^{\mathrm{tot}} - E^{\mathrm{ox}}(t))$. (This scenario is discussed in much further mathematical and experimental detail shortly in §3.2.)

For suitably well-behaved supply rate functions s, this constraint can enable the determination of one of these quantities $\{E, N, dN/dt\}$ in terms of the other two—for instance, $N(t) = h(E, dN/dt)$ for some function h, and so on. This mathematical reasoning was exactly that which related N and dN/dt in the earlier simpler case of carbon limitation: there, the constraint simplified to $N = h(dN/dt)$ specified by the very simple proportionality between cell abundance and its time derivative, as highlighted in Eq. 3.2.

To illustrate how this constraint continues to deeply and quantitatively illuminate the population dynamics and energetic status of cells beyond the simple case of carbon limitation, we now specialize to the case where the energy supply term s depends linearly

on the present value of the energy $E(t)$, namely,

$$s \equiv s(E(t)) \equiv k(F - E(t)), \tag{3.6}$$

for some constants k and F. Here, k sets the timescale on which energy is resupplied, and F specifies the default, steady magnitude of the energy in the environment if no consumption by cells were occurring. As illustrated in Figure 3.3, this mathematical character emerges highly naturally and commonly across a variety of endogenous and laboratory environments for microbial cells. For instance, as depicted in Figure 3.3(A), if a cell population is in contact with an environmental reservoir delivering a boundary condition of fresh energetic resource (such as oxygen) at a larger concentration E^{B} than locally near the cells, the resulting concentration gradient can generate a diffusive current proportional to the difference between the concentration $E(t)$ near cells and the concentration found at the boundary. Specifically, Fick's law gives the diffusive flux as $j_D = -D\frac{\partial E}{\partial z}$ for an energetic resource with diffusion constant D varying in a spatial coordinate z; taking the concentration gradient as $\frac{\partial E}{\partial z} = \frac{E_B - E(t)}{L}$ if the cells are a characteristic distance L away from the environmental boundary gives the average rate at which the concentration of energy is replenished near the cells as $s = \frac{DA}{VL}(E^{\mathrm{B}} - E(t))$, where A and V are the area and volume of the cell population.

We contend that this transport scenario, which generically emerges from coupling to big external reservoirs, applies widely to important terminal electron acceptors across natural environments. Specifically, the atmosphere presents a huge, well-mixed store of oxygen that can generically encourage an oxygen-rich boundary condition over layers of soil on land, and water columns and sediments in oceans and freshwater. Accordingly, in surface

layers, soluble oxygen at levels set by Henry's law in equilibrium with the atmosphere[3] can provide a steady flux of oxygen to cells depleting it. Oxygen is hardly the only critical bioenergetic resource participating in nearly adiabatic exchange with large external reservoirs. Carbon dioxide (despite its low redox potential) is widely harnessed for metabolism and reduced by methanogenic archaea [10] and acetogenic bacteria [11, 12]. Dissolved sulfate at large ambient concentrations of few tens of millimolar [13] is also widely present across oceans and used by sulfate reducing bacteria. These examples—which span the dominant terminal electron acceptors enabling respiration across the globe—illustrate how atmospheric or hydraulic coupling to large external pools can drive replenishment of energetically-critical resources. In contrast, the most thermodynamically useful electron donor forms of carbon are less frequently backed by such well-mixed reservoirs, and can often instead arrive as finite boluses, solids and adsorbed species. Of course, this tendency for electron acceptors to be resupplied by large reservoirs and electron donors to arrive more finitely or episodically is hardly absolute. For instance, the nitrate crusts of deserts [14] and solid-phase iron minerals, built up over geological time as solids, can be as relatively non-renewable as fresh organic carbon. And, conversely, steady release of carbon from abiotic reservoirs or ecological crossfeeding can contribute steady streams of electron donors instead of finite pools. We highlight these archetypes simply to affirm that kinetic circumstances where metabolically-limiting resources arrive as steady supply fluxes (however slow) are extremely common across natural environments, and the physicochemical nature and abundance of electron acceptors make them particularly apt to exhibit these supply kinetics.

[3] Typical values of soluble oxygen are set by Henry's law; oxygen's Henry's law volatility constant is $H^{pc}_{V,O_2} \equiv \frac{p_{O_2}}{c_{O_2,aq}} \approx 770 \underbrace{\frac{L\,atm}{mol}}_{atm/molar}$ [8]. Since atmospheric oxygen is ≈20% of the atmosphere's composition, at 1 atmosphere oxygen's partial pressure is about $p_{O_2} \approx 0.2\ atm$. Thus we expect $c_{O_2,aq} \approx \frac{p_{O_2}}{H^{pc}_{V,O_2}} \approx \frac{0.2\ atm}{770/atm/molar} \approx 2.6\times10^{-4}\ M = 260\ \mu M$. This well agrees with values of oxygen measured at $c_\infty = 210\ \mu M$ at e.g. the surface of human mucus [9].

Joining these natural environments, common laboratory settings also deliver energy to cells at rates linear in the current energy supply. As shown in Figure 3.3(B), continuous flow chemostats flowing in fresh medium with resource at concentration E^{in} at volumetric flow rate f replenish energy at a net rate of $s = f(E^{\text{in}} - E(t))$. Alternatively, as shown in Figure 3.3(C), if the functional level E^{ox} of energetic resource is replenished by redox reactions accomplished by the environment or a reoxidizing laboratory electrode, first order kinetics in the reduced form of the resource and a conservation law implies that useful energy is resupplied at a rate $s = k_{\text{el}}E^{\text{red}}(t) = k_{\text{el}}(E^{\text{tot}} - E^{\text{ox}}(t))$, where k_{el} is the first-order oxidizing rate of the electrode and E^{tot} is the total extracellular concentration of the resource. We will return in much greater detail to experiments exploiting this latter technology to probe energy-limited physiology of *P. aeruginosa* in §3.2. More generally, we remark that the form of a linear dependence of s on the present resource level is precisely the form of a proportional integral control scheme; can develop from any coupled first order kinetics that share E as a substrate but generate it in the opposite direction as the cells consume it; or else can also emerge as any dynamical linear approximation to a departure from equilibrium in $E(t)$.

Now that a supply term is specified, we find the consequences of a steady-state in the overall functional energy dE/dt.[4] The rate of change of energy in Eq. 3.3 becomes

$$\frac{dE}{dt} = -\beta N(t) + k(F - E(t)) := 0 \tag{3.8}$$

$$\rightarrow E_{\text{ss}} = \boxed{F - \frac{\beta N(t)}{k}}. \tag{3.9}$$

Armed with this expression for the energy pool E_{ss}, we now ask for the consequences for

[4] Alternatively, if cells deplete energy $E(t)$ at a rate $\beta N(t)E(t)$—that is, first order in $E(t)$ instead of zeroth order—the steady state value of the energy is instead,

$$E_{\text{ss}}^{(1)} = F\frac{k}{k + \beta N(t)}. \tag{3.7}$$

how populations of cells change in time, dN/dt.

However, even by adopting a specific form for the supply term $s(E(t))$ and enforcing a steady-state in E does not deliver closed expression for dN/dt automatically. That is, in general we no longer automatically enjoy a closure of the system of equations that specifies how cell counts change in time given just how the energy supply changes in time. To make progress, we need to provide an extra theoretical ingredient: an *ansatz* for the form of dN/dt's dependence on E and N.

Closing the system of equations to predict dN/dt and $N(t)$, how viability changes over time

We will now hazard the audacity of such an *ansatz* (namely, a brazen quantitative guess). Often, the biochemical, environmental, and genetic context of cells under study will provide principled suggestions or clues for plausible functional forms for the population changes in time, dN/dt. However, absent more precise physiological information, one phenomenological way to proceed is to anticipate that dN/dt is a smooth function f of just the variables $N(t)$ and $E(t)$ (and time derivatives thereof), namely,

$$\frac{dN}{dt} = f\left(N(t), E(t), \frac{dE}{dt}, \dots\right). \tag{3.10}$$

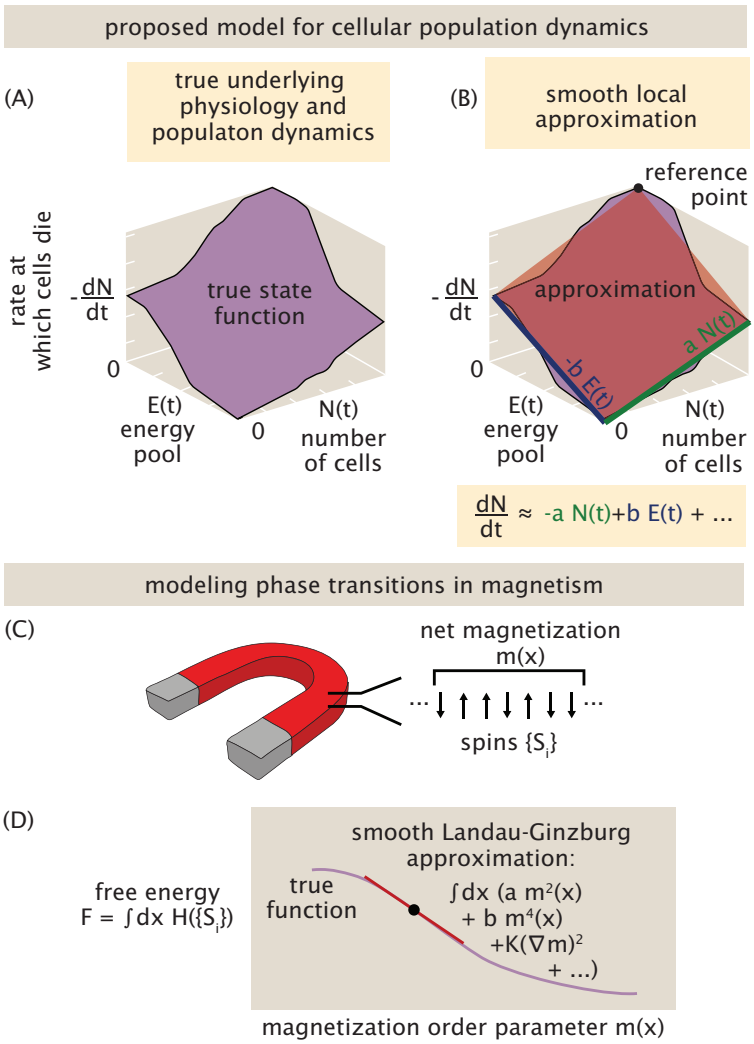


Figure 3.4: Model proposed to relate the rate of change dN/dt of the abundance of cells to the state variables of the current energy $E(t)$ and the current abundance of cells $N(t)$. (A) In principle, the rate that cells die can depend in complex ways on microscopic biochemical conditions of cells. We assume these dependencies can be subsumed into smoothly-varying dependencies on macroscopic state variables—an overall energy pool $E(t)$ (including possibly its time derivatives), and the current abundance of cells $N(t)$ (depicted schematically as the surface in purple). (B) When dN/dt varies smoothly in $N(t)$ and $E(t)$, it can admit a smooth local approximation as an expansion in each coordinate. The simplest expansion has the form $dN/dt = -a\ N(t) + b\ E(t)$, for constants $a, b \geq 0$ on physiological grounds (namely, more cells should cause death rate to increase; more energy should cause death rate to decrease). (C) This approach—approximating a macroscopic function of many microscopic variables (death rate) by low order expansions in mesoscopic state variables (current cell count and communal energy level)—is ratified by analogy to the success of the Landau-Ginzburg approach to phase transitions in (for instance) magnetic materials. In the Landau-Ginzburg approach, microscopic variables (magnetic spins $\{S_i\}$) are summarized by order parameters (including the magnetization $m(\mathbf{x})$), and the local Hamiltonian of the whole system is expressed as a Taylor expansion of the free energy density with respect to the order parameters and its derivatives, but only retaining terms permitted by physical symmetries in the problem. Keeping only the lowest non-trivial terms gives simple field equations consistent with the field variables and known constraints.

As illustrated in Figure 3.4, this formulation immediately invites a Taylor expansion, about some reference state $N(t) = N_*, E(t) = E_*$. In general, this gives,

$$f \approx f_0 + \left.\frac{\partial f}{\partial N}\right|_{N^*,E^*} (N - N_*) + \left.\frac{\partial f}{\partial E}\right|_{N^*,E^*} (E - E_*) + \frac{1}{2} \left.\frac{\partial^2 f}{\partial E \partial N}\right|_{N_*,E_*} (E - E_*)(N - N_*) + \cdots \tag{3.11}$$

$$= \underbrace{\left(f_0 - \frac{\partial f}{\partial N} N_* - \frac{\partial f}{\partial E} E_* + \cdots\right)}_{\equiv c} + \underbrace{\frac{\partial f}{\partial N}}_{\equiv -a} N + \underbrace{\frac{\partial f}{\partial E}}_{\equiv b} E + \frac{1}{2} \frac{\partial^2 f}{\partial E \partial N} N E + \cdots, \tag{3.12}$$

where we have defined constants a, b, and c as indicated for notational convenience, foreshadowing their expected physical signs.

If this function does not vary aggressively with its variables $N(t)$, $E(t)$, then it is reasonably approximated by truncating this series to the first few terms. Retaining various orders of terms in this expression amounts to different complexities acknowledged of physiology. This tradition is simple but validated by the expressiveness of both the generalized Lotka Volterra equations and consumer resource models of ecology[5] and a broader tradition of parsimonious expansions of macroscopic field variables throughout statistical physics (as illustrated by the analogy in Figure 3.4(C-D)).

For purposes of our immediate discussion here, however, we will consider the simplest linear theory of physiology, given by the first order terms in this expansion, namely

$$\frac{dN}{dt} \approx c - aN(t) + bE(t). \tag{3.13}$$

Now, we argue that on physiological grounds, we anticipate that $c = 0$ and $a, b \geq 0$—that

[5]For instance, taking $b = 0$ and $\frac{1}{2} \left.\frac{\partial^2 f}{\partial E \partial N}\right|_{N_*,E_*}$ nonzero gives a linear variant of the famous MacArthur Consumer Resource Model for one consumer and one resource (see Eq. A2 in Cui, Marsland, and Mehta [15]).

is, we expect that no change in the cell population should occur without any cells or energy ($N(t) = E(t) = 0$), and further that more cells should cause at most more death but more energy should cause at most less death. This gives,

$$\boxed{\frac{dN}{dt} \approx -aN(t) + bE(t)}. \tag{3.14}$$

The parameters a and b capture overall couplings of cellular death rate to current population size and energy states as shown by the approximate red plane in Figure 3.4(B) and carry natural physical meanings. The constant a represents the per-capita death rate (e.g. exponential death rate) if no excess available resource E were available, $E = 0$; it is reasonable to expect that a monotonically increases with the basal metabolic rate of cells β. Conversely, the energy coefficient b represents the decrease in the number of cells dying per time (eg the unit change in dN/dt) expected per additional unit of energy; it should be related to the thermodynamic yield/biochemical quality of the energetic resource in question. Just as for the coefficients entering Landau-Ginzburg field theory of critical phenomena, these parameters are phenomenological but, in principle, directly experimentally measurable, and their qualitative features can be integrated with thermodynamic understanding. For instance, different electron donors and acceptors fall on a hierarchy based on the reducing or oxidizing potential of their half reactions, in addition to the biochemical efficiency with which given organisms accomplish these reactions for given resources. Accordingly, different energetic resources should impose a corresponding hierarchy on the numerical values of the parameters b.

Finally, armed with both the consequences of an energetic steady-state in Eq. 3.9 and the proposed closure of the systems of equations governing cell count and energy Eq. 3.14, we solve how cell count changes over time. Substituting the expression for energy into the

proposed dN/dt gives,

$$\frac{dN}{dt} = -aN(t) + b\underbrace{E(t)}_{\equiv E_{ss}} \tag{3.15}$$

$$= -aN(t) + b\left(F - \frac{\beta N(t)}{k_{\text{el}}}\right) \tag{3.16}$$

$$= -\left(a + \frac{\beta}{k_{\text{el}}}\right)N(t) + bF. \tag{3.17}$$

This differential equation for the abundance of cells Eq. 3.17 has a simple, easily expressed analytical solution, predicting that cell count will change in time according to,

$$N(t) = \boxed{\underbrace{\frac{bF}{a + \frac{\beta}{k}}}_{\equiv N_\infty} + \left(N_0 - \frac{bF}{a + \frac{\beta}{k}}\right)\exp\left[-\left(a + \frac{\beta}{k}\right)t\right]} \tag{3.18}$$

$$= N_\infty + (N_0 - N_\infty)\exp\left[-\left(a + \frac{\beta}{k}\right)t\right], \tag{3.19}$$

where N_0 is the initial population of cells at time $t = 0$. These population dynamics in Eq. 3.19 reflect an exponential decay to a constant, nonzero, steady-state population size,

$$N_\infty \equiv \frac{bF}{a + \frac{\beta}{k}}. \tag{3.20}$$

This steady-state population scales linearly with the parameter b setting the reduction in death rate per unit of excess energy and the default environmental value F of the energy (if no cells were consuming energy). It also increases (hyperbolically) with the rate k at which energy is resupplied. In contrast, this steady-state population decreases with the default per-capita death rate a (if no energy were present) and the maintenance rate β at which cells consume energy. Notice that the dynamics of this solution Eq. 3.19 interpolates between two slopes for log survival: at early times, if $N(t)$ is far above the steady-state abundance,

the population declines approximately exponentially with a slope $a + \frac{\beta}{k}$ in time. At late times, however, the approach to a the constant steady-state population flattens this slope, asymptotically converging to zero.

To illustrate that this model, however extraordinarily simple, nonetheless makes precise and falsifiable predictions that commune with real measurements, we return to the motivating abundance curves of Fig. 3.2(B), measured by Ciemniecki and coworkers [5]. In this (carbon-plush, anaerobic) setting, oxidized phenazine is the limiting energetic resource, which invites us to mathematically identify it as setting the variable $E(t)$ in our preceding equations, $E(t) \sim [\mathrm{PCN}]^{\mathrm{ox}}(t)$. Under the conditions of the indicated measurements, explained much more thoroughly in the following section, we anticipate that the action of an electrode accomplishing the measurement replenishes the oxidized phenazine at a first order rate in the current amount of reduced phenazine. This identifies the default amount F of resource (phenazine) if the cells were absent as the total amount of phenazine, $F = [\mathrm{PCN}]^{\mathrm{tot}} = [\mathrm{PCN}]^{\mathrm{ox}}(t) + [\mathrm{PCN}]^{\mathrm{red}}(t)$. Armed with these interpretations, our model of Eq. 3.20 makes the strong and unambiguous prediction that the late-time (approximating steady-state) population size of cells should scale linearly with the total phenazine concentration,

$$\text{prediction: } N_{ss} \propto [\mathrm{PCN}]^{\mathrm{tot}}. \tag{3.21}$$

Is this prediction ratified? Fig. 3.5 largely affirms this prediction, registering an impressively linear variation of total population size at late times with external phenazine concentration.

Moreover, importantly, the model also adequately describes the fuller dynamics of

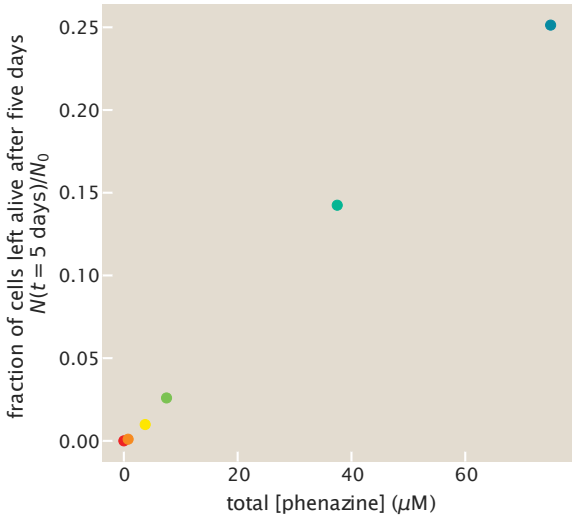


Figure 3.5: Remaining abundance of cells at late times (at $t = 5$ days), as a function of the extracellular phenazine concentration, extracted from the data depicted earlier in Fig. 3.2(B) from Ciemniecki *et al.* [5]. Late cell abundance clearly varies largely linearly with external phenazine concentration. Note the small departure from linearity visible in the condition at 75 μ phenazine at the upper right, which might be interpretable as the onset of physiological saturation or toxicity. Another mild departure is visible for the condition without phenazine entirely (red dot at lower left), since our model does not comment on the nature of energy sources orthogonal to phenazine.

survival curves, as supported by Fig. 3.6.[6]

The adequate agreement of even the simplest model we discussed and both quantitative and qualitative features of these population abundance curves over time invites both optimism for broader applicability and extensions. These models connect, in a minimally viable way, the unique phenomenology and underlying bioenergetic status of cells

[6]To resolve the two remaining fit parameters for this model's prediction for $N(t)$, we can rewrite the prediction assuming that the only parameter varying was F (set by the total phenazine concentration) and not the physiological or electrode parameters a, b, k, β. Dividing Eq. 3.18 by N_0 gives,

$$\text{fractional viability}: \frac{N(t)}{N_0} = \frac{bF}{N_0\left(a+\frac{\beta}{k}\right)} + \left(1 - \frac{bF}{N_0\left(a+\frac{\beta}{k}\right)}\right)\exp\left[-\left(a+\frac{\beta}{k}\right)t\right]. \tag{3.22}$$

Assuming fixed b, N_0, a, β, k but varying F, we can define

$$\frac{N(t)}{N_0} = \sigma F + (1-\sigma F)\exp\left[-\frac{\xi}{\sigma}t\right], \tag{3.23}$$

where we appealed to just two parameters $\sigma \equiv \frac{b}{N_0(a+\beta/k)}$ and $\xi \equiv b/N_0$. Note that to model death requires that $(1 > \sigma F)$.

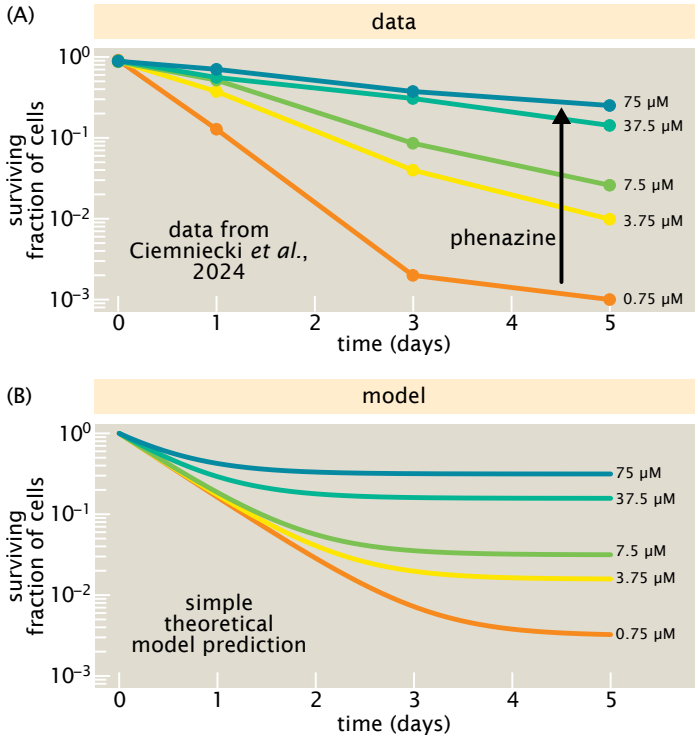


Figure 3.6: Measured survival data compared to the model of Eq. 3.19. (A) A physiologically tolerated subset of data. (B) Our simplest model prediction from Eq. 3.17, enforcing the energy source term F proportional to the phenazine concentration (otherwise leaving two fit parameters).

challenged by limited terminal electron acceptors. Once paired with appropriate calibrating experiments and measurements, such models could infer and monitor the metabolic rates of cells scalably across environmental conditions and perturbations, similarly to how such inferences have been developed and productively applied to the case of carbon limitation.

3.2 Explicit kinetic theories for maintenance experiments

Having seen that both qualitative and quantitative features of how cells survive when limited by terminal electrode acceptors can be predicted from a very simple model with interpretable physiological parameters, we now strive to make close contact with particular culturing experiments. Specifically, we aim to build candidate "theories of the experiment" for how the raw measured quantities (metabolic current and cell abundance) interact in experiments on *Pseudomonas* sustained anaerobically [5] in the Newman laboratory. Our aim is practical and granular: we want to offer distinct predictions for observables that help distinguish between different possible underlying kinetic pictures that could describe how cells and experimental electrodes interact. We feel that adjudicating among these underlying kinetic pictures is important to interpret exactly how trends in observables report on underlying physiology. However, while we now specialize to the language of *Pseudomonas*, phenazines, and electrodes in what follows, we underscore that essentially identical mathematics would make quantitative predictions of a plethora of broader environmental scenarios, such as those intimated in Fig. 3.3, *mutatis mutandis*.

At a high conceptual level, eliding many fascinating and complex biochemical details, cells produce energy by directing electrons from an electron donor (e.g. a carbon source like a sugar) and enzymatically delivering these electrons to an electron acceptor (such as oxygen in aerobic conditions), gleaning Gibbs free energy of the coupled half reactions

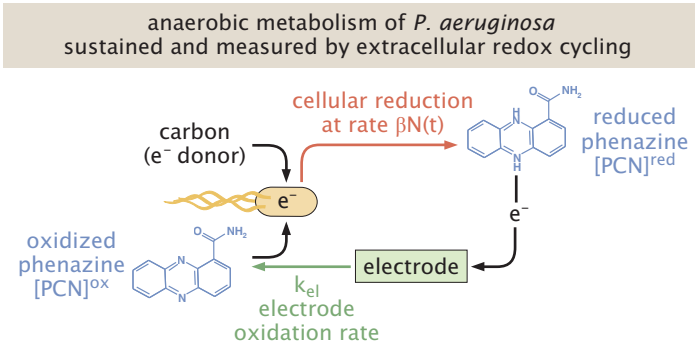


Figure 3.7: Schematic for how cellular metabolism at rate $\beta N(t)$, driving reduction of extracellular phenazine, is counteracted by oxidation by the electrode at rate k_{el}.

along the way. Oxygen is often a privileged electron acceptor for many aerobe species. Yet, when it is unavailable—as is exceedingly common in microbial biofilms in soil, mucus-filled lungs, marine sediments, and beyond—cells have developed a spectrum of ingenious alternative energy conservation strategies. One such tactic is implemented by *Pseudomonas*, whose strains can produce a class of extracellular redox-active molecules known as phenazines. Often, producing phenazines is key to how *Pseudomonas* can persist and exhibit pathogenicity, despite energy limitation or antibiotics, across natural environments.

To explore the beautiful metabolic dynamics attending anaerobic situations compensated for by phenazines, cells where the natural phenazine biosynthesis pathways have been knocked out ($\Delta phz1/2$ strains) can be provided titrated amounts of exogenous extracellular phenazine. As discussed in both recent [5] and now-classic [16, 17] studies, while these struggling cells perform basal metabolism and reduce extracellular phenazine as a terminal electron acceptor, an electrode can reoxidize phenazine to replenish its supply and simultaneously offer a readout of cellular metabolism. We illustrate this procedure in Fig. 3.7. Further, under some relevant physiological conditions, calibrations and biochemistry establish that 1 electron per second per cell measured by electrode can correspond to the energetic equivalent of a cell spending $\sim$1 ATP hydrolysis's worth of

energy [5]. Thus, measured current gives an unprecedentedly direct readout of cellular metabolism in a variety of challenging conditions of enduring environmental relevance—namely, anaerobic energy limitation.

Now, we strive to minimally mathematize such cellular metabolism balanced by electrode kinetics. Specifically, we model cells as reducing oxidized phenazine in either a zeroth order process (that is, cells reduce oxidized phenazine at a rate independent of the current level of oxidized phenazine), or as a first order process (that is, the rate at which oxidized phenazine is reduced scales with the current concentration of oxidized phenazine). Complementarily, we model the electrode as accomplishing reduction that is either zeroth order or first order in the current level of reduced phenazine in the culture medium.

These different kinetic orders are each plausibly well justified in different physiological and physical scenarios. Cells would reduce oxidized phenazine at a constant rate (zeroth order in the external phenazine) if their internal metabolic repertoire, not uptake per se, bottlenecks cellular reduction, and/or if their emission of reduced phenazine itself is saturated. Alternatively, however, if uptake or transport were limiting cellular metabolism, reduction kinetics would more plausibly be dependent on (plausibly first order in) the level of oxidized phenazine. Analogously, if transport to the electrode were truly limiting the rate at which it could reoxidize phenazine, the electrode would plausibly depend on the current level of reduced phenazine (credibly to first order, for example in a diffusion-limited sense). Otherwise, if the electrode's internal kinetics were a bottleneck, its action would better be described as zeroth order.

It is not fundamentally obvious, *a priori*, that the operating regimes of microbial culture (or environmental settings) necessarily ordain that any of these kinetic scenarios is never relevant. (This said, to interpret the electrode as measuring something about the ambient metabolic environment, its current is probably not zeroth order (which would register a constant current no matter the action of cells).) That is, there may be physical scenarios

where any of these kinetic combinations manifest. In particular, a combination of these scenarios might develop simultaneously in natural biofilms or planktonic cultures.

Accordingly, here, we specify, solve, and explore all four sets of kinetic equations setting the balance between cellular reduction of phenazine and the reoxidation by an electrode up to first order reaction kinetics in each process. These joint models are anticipated in Figure 3.8.

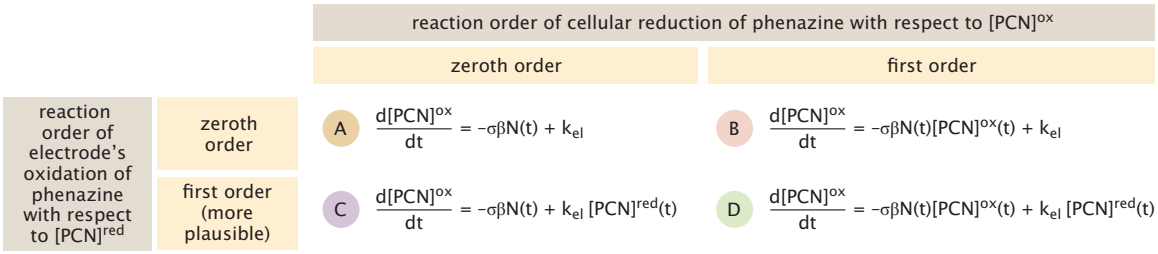


Figure 3.8: Presiding differential equations for the functional energy source. All four kinetic order combinations possible when cells and electrodes each engage in zeroth or first order kinetics in their respective substrate forms of phenazine. (A) Zeroth order kinetics by cells, zeroth order kinetics by the electrode. (B) First order kinetics by cells, zeroth order kinetics by the electrode. (C) Zeroth order kinetics by cells, first order kinetics by the electrode. (D) First order kinetics by cells, first order kinetics by the electrode. (Note that across these models, the units of the presiding constants change according to dimensional consistency: e.g. σ carries units of phenazine reduced per second per joule metabolized by cells if cells are zeroth order, but alternatively has units of per second per joule metabolized by cell if cells are first order in oxidized phenazine. Analogously, k_{el} has units of oxidized phenazine produced per time if the electrode is zeroth order, but otherwise has units of oxidized phenazine molecules per reduced phenazine per time if the electrode is first order.)

We discuss models (C) and (D) previewed in Fig. 3.8, corresponding to first order kinetics by the electrode, first and in greatest detail. These are the especially relevant and interesting cases where the electrode's activity actually varies with the present concentration of oxidized phenazine. However, for completeness, we include and remark in passing on aspects of the solutions of models (A) and (B) (which reflect zeroth order kinetics by the electrode).

model (C)	
cells: zeroth order in $[PCN]^{ox}$ electrode: first order in $[PCN]^{red}$	
differential equation:	$\frac{d[PCN]^{ox}}{dt} = -\sigma\beta N(t) + k_{el}[PCN]^{red}(t) = -\sigma\beta N(t) + k_{el}([PCN]^{tot} - [PCN]^{ox}(t)) = -\sigma\beta N(t) - k_{el}[PCN]^{ox}(t) + k_{el}([PCN]^{tot}$ (of the form dx/dt = − a N(t) − b x(t) + c)
resulting solution:	$[PCN]^{ox}(t) = \exp[-k_{el}(t)]\left([PCN]_0^{ox} + \int_0^t d\tau \exp[k_{el}\tau](k_{el}[PCN]^{tot} - \sigma\beta N(\tau))\right)$
simplification if N(t) is ≈ constant N_0:	$[PCN]^{ox}(t) = \exp[-k_{el}(t)]\left([PCN]_0^{ox} - \left([PCN]^{tot} - \frac{N_0\sigma\beta}{k_{el}}\right)\right) + \left([PCN]^{tot} - \frac{N_0\sigma\beta}{k_{el}}\right)$
steady-state PCN:	$[PCN]_\infty^{ox} = [PCN]^{tot} - \frac{\sigma\beta N(t)}{k_{el}}$
steady-state current:	$I_{ss} \propto k_{el}\left([PCN]^{tot} - \left([PCN]^{tot} - \frac{\sigma\beta N(\tau)}{k_{el}}\right)\right) = \sigma\beta N(t)$
per-cell steady-state current:	$\frac{I_{ss}}{N(t)} \propto \sigma\beta$

Figure 3.9: Mathematical properties of Model (C) (as portended by Fig. 3.8).

3.2.1 Model C: Zeroth-order cells fed by first-order electrodes

Back in Eq. 3.8, we visited a form of this model, evaluated when the energetic resource (e.g. oxidized phenazine) reached a steady-state. Here we return to set up the scenario and probe some of its more dynamical properties in the explicit setting of cells sustained by extracellular phenazine cycling.

As Fig. 3.7 recalls, phenazine transitions between two chemical forms, oxidized ($[PCN]^{ox}$) or reduced ($[PCN]^{red}$), with total phenazine concentration giving the total $[PCN]^{tot} = [PCN]^{ox} + [PCN]^{red}$. Assume cells reduce phenazine at a constant rate $\sigma\beta N(t)$; namely, proportional to their total maintenance energy demand $\beta N(t)$, where β is the basal metabolic power per cell (in e.g. ATP/s/cell or Watt/cell), $N(t)$ is their current abundance, and σ describes the number of phenazine reductions required per unit of energy. If electrodes replenish the oxidized form of phenazine at a first-order rate in the current level of reduced phenazine, $k_{el}[PCN]^{red}(t)$, then the competition of these reactions sets how the current level of oxidized phenazine changes in time, namely,

$$\frac{d[\mathrm{PCN}]^{\mathrm{ox}}}{dt} = \underbrace{-\sigma\beta N(t)}_{\text{cells reduce oxidized phenazine}} + \underbrace{k_{\mathrm{el}}[\mathrm{PCN}]^{\mathrm{red}}}_{\text{electrode reoxidizes phenazine}} = -\frac{d[\mathrm{PCN}]^{\mathrm{red}}}{dt} \tag{3.24}$$

$$= -\sigma\beta N(t) + k_{\mathrm{el}}([\mathrm{PCN}]^{\mathrm{tot}} - [\mathrm{PCN}]^{\mathrm{ox}}(t)). \tag{3.25}$$

At steady-state in the phenazine level, $\frac{d[\mathrm{PCN}]^{\mathrm{ox}}}{dt} = 0$, this differential equation predicts a stable level of oxidized phenazine given by,

$$[\mathrm{PCN}]^{\mathrm{ox}}_{\infty} = [\mathrm{PCN}]^{\mathrm{tot}} - \frac{\sigma\beta N(t)}{k_{\mathrm{el}}}. \tag{3.26}$$

The electrode reports a current, $I(t)$, proportional to the rate at which it is taking electrons off of reduced phenazine. We specified that rate to be $k_{\mathrm{el}}[\mathrm{PCN}]^{\mathrm{red}}(t)$ (by proscribing first order kinetics to the electrode). So, at steady-state in the oxidized phenazine concentration, the current will be,

$$I_{\mathrm{ss}} \propto k_{\mathrm{el}} \left([\mathrm{PCN}]^{\mathrm{tot}} - \left([\mathrm{PCN}]^{\mathrm{tot}} - \frac{\sigma\beta N(t)}{k_{\mathrm{el}}} \right) \right) \tag{3.27}$$

$$= \sigma\beta N(t). \tag{3.28}$$

Accordingly, this model makes the firm prediction that the measured per-cell current, $I_{ss}/N(t) \propto \sigma\beta$, should be proportional only to the maintenance rate that cells consume energy (thereby reducing phenazine); this steady-state current should *not* vary with the absolute population size $N(t)$, the total phenazine concentration $[\mathrm{PCN}]^{\mathrm{tot}}$, or the characteristic rate at which the electrode performs oxidation, k_{el}. As we will now see, these invariances are not assured for other kinetic scenarios.

More dynamically, if we knew $N(t)$ as a trajectory in time (despite its fundamental dependence on $E(t)$ itself), we could write an (implicit) closed form solution to this

differential equation Eq. 3.25, via an integrating factor. The result is the not-very-illuminating expression,

$$[\text{PCN}]^{\text{ox}}(t) = \exp[-k_{\text{el}}t]\left([\text{PCN}]_0^{\text{ox}} + \int_0^t d\tau \exp[k_{\text{el}}\tau]\left(k_{\text{el}}[\text{PCN}]^{\text{tot}} - \sigma\beta N(\tau)\right)\right). \quad (3.29)$$

However, this equation becomes more comprehensible and informative when the cell count $N(t)$ is essentially constant, $N(t) \approx N_0$, over the timescale over which we are asking dynamical questions of $[\text{PCN}]^{\text{ox}}(t)$. Then, phenazine relaxes in time as,

$$[\text{PCN}]^{\text{ox}}(t) = \exp\left[-k_{\text{el}}t\right]\left([\text{PCN}]_0^{\text{ox}} - \left([\text{PCN}]^{\text{tot}} - \frac{N_0\sigma\beta}{k_{\text{el}}}\right)\right) + \left([\text{PCN}]^{\text{tot}} - \frac{N_0\sigma\beta}{k_{\text{el}}}\right). \quad (3.30)$$

These properties of this model, which must be ruthlessly compared with behaviors registered by experiments, are summarized in Fig. 3.9.

3.2.2 Model D: First-order cells fed by first-order electrodes

When microbes also perform redox chemistry with phenazines with first order kinetics, the cellular term that was $-\sigma\beta N(t)$ in Eq. 3.25 becomes $-\sigma\beta N(t)E(t)$[7]. This yields the differential equation,

$$\frac{d[\text{PCN}]^{\text{ox}}}{dt} = \underbrace{-\sigma\beta N(t)[\text{PCN}]^{\text{ox}}}_{\text{cells reduce oxidized phenazine}} + \underbrace{k_{\text{el}}[\text{PCN}]^{\text{red}}}_{\text{electrode reoxidizes phenazine}} = -\frac{d[\text{PCN}]^{\text{red}}}{dt} \quad (3.31)$$

$$= -\sigma\beta N(t)[\text{PCN}]^{\text{ox}} + k_{\text{el}}([\text{PCN}]^{\text{tot}} - [\text{PCN}]^{\text{ox}}) \quad (3.32)$$

$$= k_{\text{el}}[\text{PCN}]^{\text{tot}} - (\sigma\beta N(t) + k_{\text{el}})[\text{PCN}]^{\text{ox}}. \quad (3.33)$$

[7]Notice here that the units of σ have changed between these two models, as required by dimensional consistency.

At steady-state in the balance of reduced contra oxidized phenazine, we see that

$$[\mathrm{PCN}]_{\infty}^{\mathrm{ox}} = [\mathrm{PCN}]^{\mathrm{tot}} \frac{k_{\mathrm{el}}}{k_{\mathrm{el}} + \sigma\beta N(t)}. \tag{3.34}$$

This implies that the electrode will report a steady-state current,

$$I_{\mathrm{el}} \propto k_{\mathrm{el}}[\mathrm{PCN}]^{\mathrm{red}} \tag{3.35}$$

$$= k_{\mathrm{el}}([\mathrm{PCN}]^{\mathrm{tot}} - [\mathrm{PCN}]^{\mathrm{ox}}) \tag{3.36}$$

$$= k_{\mathrm{el}} \left([\mathrm{PCN}]^{\mathrm{tot}} - [\mathrm{PCN}]^{\mathrm{tot}} \frac{k_{\mathrm{el}}}{k_{\mathrm{el}} + \sigma\beta N(t)} \right) \tag{3.37}$$

$$= k_{\mathrm{el}}[\mathrm{PCN}]^{\mathrm{tot}} \left(\frac{\sigma\beta N(t)}{k_{\mathrm{el}} + \sigma\beta N(t)} \right), \tag{3.38}$$

Here, doubling the cell count has a multiplicative effect smaller than a doubling of the current. Only when $k_{\mathrm{el}} \gg \sigma\beta N(t)$ is the measured current approximately proportional to the cell count $N(t)$, $I_{\mathrm{el}} \propto \approx k_{\mathrm{el}}[\mathrm{PCN}]^{\mathrm{tot}} \frac{\sigma\beta N(t)}{k_{\mathrm{el}}} = [\mathrm{PCN}]^{\mathrm{tot}} \sigma\beta N(t)$. That is, the value of the current normalized by the total cell count, $\frac{I_{\mathrm{el}}}{N(t)}$, only enjoys the interpretation of being literally proportional to the per-cell metabolic rate β we are interested in when the electrode's reoxidation rate k_{el} is much faster than the cellular rate $\sigma\beta N(t)$.[8]

[8] A casual empirical remark included in Ciemniecki *et al.* [5] might be consistent with this picture: "Accordingly, on day 5 of the survival assay, we pooled ΔPCN1/2 cells from multiple technical replicate wells in the 375 mM PCN condition, pelleted them, resuspended them in a small volume, and added them to other wells containing ΔPCN1/2 cells and 375 mM PCN such that the total cell concentration was approximately doubled. We observed a 1.5x increase in current generation following the increase in cells that was sustained for at least 10 h (Figure S4B), demonstrating that the electrode was not limiting. Accordingly, the observed PCN oxidation rate at the electrode (Figure 5D) can be interpreted as equivalent to the cellular PCN reduction rate."

model (D)		
cells:	first order	in $[PCN]^{ox}$
electrode:	first order	in $[PCN]^{red}$

differential equation:	$\frac{d[PCN]^{ox}}{dt} = -\sigma\beta N(t)[PCN]^{ox}(t) + k_{el}\,[PCN]^{red}(t)$ $= -\sigma\beta N(t)[PCN]^{ox} + k_{el}\,([PCN]^{tot} - [PCN]^{ox}(t))$ (of the form $dx/dt = a - f(t)\,x(t)$) $= k_{el}[PCN]^{tot} - (\sigma\beta N(t) + k_{el})[PCN]^{ox}(t)$
resulting solution:	$[PCN]^{ox}(t) = \exp\,[-k_{el}(t)]\,\exp\left[-\sigma\beta\int_0^t d\tau\, N(\tau)\right]\left([PCN]_0^{ox} + k_{el}[PCN]^{tot}\int_0^t d\tau\,\exp\,[k_{el}\tau]\,\exp\left[\int_0^t ds\, N(s)\right]\right)$
simplification if N(t) is ≈ constant N_0:	$[PCN]^{ox}(t) = [PCN]^{tot}\,\frac{k_{el}}{k_{el} + \sigma\beta N_0} + \left([PCN]_0^{ox} - [PCN]^{tot}\,\frac{k_{el}}{k_{el} + \sigma\beta N_0}\right)\exp\,[-\,k_{el} + \sigma\beta N_0)t]$
steady-state PCN:	$[PCN]_\infty^{ox} = [PCN]^{tot} - \frac{k_{el}}{k_{el} + \sigma\beta N(t)}$
steady-state current:	$I_{ss} \propto k_{el}\,[PCN]^{tot}\left(\frac{\sigma\beta N(t)}{k_{el} + \sigma\beta N(t)}\right)$
per-cell steady-state current:	$\frac{I_{ss}}{N(t)} \propto k_{el}\,[PCN]^{tot}\left(\frac{\sigma\beta}{k_{el} + \sigma\beta N(t)}\right)$

Figure 3.10: Mathematical properties of Model (D) (as portended by Fig. 3.8).

This has the solution, assuming that $N(t)$ does not rapidly change over timescale of interest, namely $N(t) \approx N_0$,

$$[\mathrm{PCN}]^{\mathrm{ox}}(t) = \underbrace{[\mathrm{PCN}]^{\mathrm{tot}} \frac{k_{\mathrm{el}}}{k_{\mathrm{el}} + \sigma\beta N_0}}_{[\mathrm{PCN}]^{\mathrm{ox}}_{\infty}} + \left([\mathrm{PCN}]^{\mathrm{ox}}_0 - [\mathrm{PCN}]^{\mathrm{tot}} \frac{k_{\mathrm{el}}}{k_{\mathrm{el}} + \sigma\beta N_0}\right) \exp[-(k_{\mathrm{el}} + \sigma\beta N_0)t] \tag{3.39}$$

$$= [\mathrm{PCN}]^{\mathrm{ox}}_{\infty} + ([\mathrm{PCN}]^{\mathrm{ox}}_0 - [\mathrm{PCN}]^{\mathrm{ox}}_{\infty}) \exp[-(k_{\mathrm{el}} + \sigma\beta N_0)t]. \tag{3.40}$$

Notice that if there is a rapid steady-state in the phenazine dynamics (e.g. phenazine rapidly adjusts to a steady-state upon slower changes to cell counts), then

$$[\mathrm{PCN}]^{\mathrm{ox}} \longrightarrow [\mathrm{PCN}]^{\mathrm{ox}}_{\infty} = [\mathrm{PCN}]^{\mathrm{tot}} \frac{k_{\mathrm{el}}}{k_{\mathrm{el}} + \sigma\beta N(t)};$$

this holds instantaneously, even if $N(t)$ is not literally a constant N_0. This expression for the steady-state concentration of oxidized phenazine is illuminating from the perspective of our bioenergetic questions, since it anticipates a linear scaling of the oxidized phenazine concentration with total external phenazine, and also a hyperbolic (saturating) dependence on the current cell abundance $N(t)$. Further mathematical behaviors are summarized in Fig. 3.10.

3.2.3 Models A and B: Zeroth-order electrodes feed cells

Mathematical features of Models A and B are summarized in Figs. 3.11 and 3.12.

model (A)		
cells:	zeroth order	in $[PCN]^{ox}$
electrode:	zeroth order	in $[PCN]^{red}$

differential equation:	$\frac{d[PCN]^{ox}}{dt} = -\sigma\beta N(t) + k_{el}$ (of the form dx/dt = − a N(t) + c)
resulting solution:	$[PCN](t) = k_{el}t - \sigma\beta \int_0^t dt\, N(t)$
steady-state implication:	$N(t) = \frac{k_{el}}{\sigma\beta}$ (constant population size)
implied current:	constant $I_{ss} \propto k_{el}$ (independent of cell number N(t) or PCN)

Figure 3.11: Model (A) (as portended by Fig. 3.8).

model (B)	
	cells: first order in $[PCN]^{ox}$ electrode: zeroth order in $[PCN]^{red}$
differential equation:	$\frac{d[PCN]^{ox}}{dt} = -\sigma\beta N(t)[PCN]^{ox}(t) + k_{el}$ (of the form dx/dt = − a x(t) N(t) + c)
resulting solution:	$[PCN]^{ox}(t) = \exp\left[-\sigma\beta\int_0^t dt\, N(t)\right]\left([PCN]_0^{ox} + k_{el}\int_0^t d\tau \exp\left[\sigma\beta\int_0^\tau ds\, N(s)\right]\right)$
simplification if N(t) is ≈ constant N_0:	$[PCN]^{ox}(t) = \frac{k_{el}}{\sigma\beta N_0} + \left([PCN]_0^{ox} - \frac{k_{el}}{\sigma\beta N_0}\right)\exp\left[-\sigma\beta N_0 t\right]$
steady-state PCN:	$[PCN]_\infty^{ox}(t) = \frac{k_{el}}{\sigma\beta N(t)}$
implied current:	constant $I_{ss} \propto k_{el}$ (independent of cell number N(t) or PCN)

Figure 3.12: Model (B) (as portended by Fig. 3.8).

In summary, we expect that Model C or Model D may be especially effective models that quantitatively capture the phenazine dynamics and culturing experiments performed by the Newman laboratory. Which model is more appropriate demands targeted measurements, probably involving the independent measurement of the electrode's kinetic rate k_{el} itself, and asking at high resolution how current varies systematically with modulated properties like total cell count and total phenazine content in a spectrum of conditions. It is also very feasible to generalize these dynamics to a smooth interpolation of these models. For instance, cells implementing Michalis Menten kinetics at a rate $N(t)\beta_{\max}\frac{E(t)}{E(t)+K_M}$ (for resource $E(t)$, characteristic saturation constant K_M setting the typical onset of reductive activity, and maximum per-capita reduction rate $\beta_{\max}$) interpolates between zeroth and first order kinetics in energetic resource (namely phenazine).

3.3 Quantitative glimpses of metabolic adaptation to ionic stress in *Pseudomonas*

In the preceding chapter, we discussed how to newly extract from published death data [4] a coefficient—referred to as κ—specifying how much additional metabolic power *E. coli* cells spend per unit of environmental salt (sodium chloride). Under these aerobic, carbon-limited conditions, we inferred $\kappa_{E.\,coli} \sim 10^3$ ATP/s/cell/(mM NaCl). This large value betrays that *E. coli* can feel a considerable metabolic burden when buffeted by environmentally-feasible fluctuations in ion concentrations. Do other species, such as energy-limited *Pseudomonas*, show a comparably big sensitivity to its osmotic environment? How does their metabolic responses to environmental salt conditions change with other variables of environmental context?

Towards these questions, we worked with members of the Newman laboratory

(specifically Hannah Jeckel, Tracy (Chia Lun) Ho, Kemal Demirer) to design experiments, which they intrepidly performed, to measure the impacts of both extracellular salt and phenazine concentrations on the viability and metabolic rates of *P. aeruginosa* narrowly surviving in anaerobic conditions. These experiments harness the powerful experimental conceit of measuring extracellular phenazine reduction by *Psuedomonas* as a measureable reporter of metabolic rate, as discussed and illustrated in the preceding section and in Fig. 3.7. Specifically, we could directly and precisely measure the metabolic rate of anaerobic cells in a big set of environmental conditions using a parallel set of electrodes compatible with 96 well culture plates, as discussed and deployed recently by Ciemniecki and coworkers [5]. As theorists, we are deeply galvanized by these experimental settings and measurements. Enlightenment about how cells work, if it is to come, will be grounded in probing the high-dimensional landscape of cellular physiology by testing many different environmental modulations and their effects in calibrated, real units. Refuting, ratifying, or ultimately building candidates for emergent quantitative simplicities in microbial life can be greatly catalyzed by strains under genetic control; environments under biochemical and atmospheric control; and high-resolution measurements resolving their interplay, capabilities the work by the Newman lab advances.

In a set of pilot experiments, *P. aueriginosa* cells were provided plenty of carbon; no oxygen; three concentrations of extracellular phenazine to act as the terminal electron acceptor; and four different levels of sodium chloride, all in physiologically relevant concentration ranges. These cells were nurtured in standard MOPS minimal medium, as described previously [5, 18, 17].

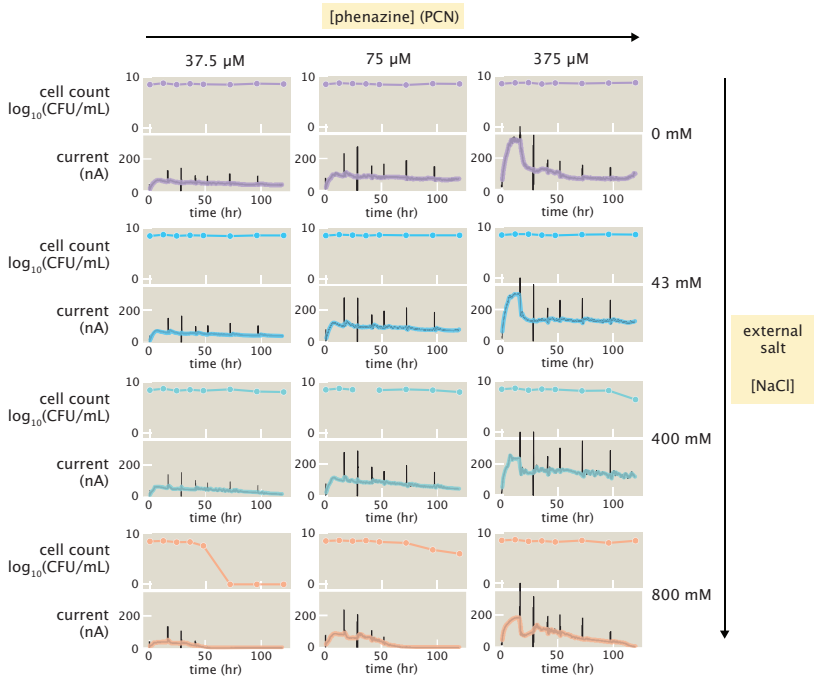


Figure 3.13: Impact of salt and external phenazine concentration on cellular abundance and current over time. The (i, j)th subplot in this grid shows the viable cell count (top axis) and total current (bottom axis) measured for the ith external sodium chloride concentration and the jth extracellular phenazine concentration condition. (In these current plots, the thin black trajectories show the raw current signal over time, including the presence of large transient spikes when very small volumes were withdrawn for purposes of counting viable cell abundance. To filter out these spikes as artifacts of measurement, we computed the median of current over one hour increments, shown as thick colored lines.)

As shown in the top axes of each plot in Figure 3.13, the bulk survival of such energetically and ionically challenged *Pseudomonas* over time varies richly with both external salt and phenazine. Specifically, more than eighty percent of cells largely survived over 120 hours and longer if they were supplied with any of the tested 37.5 μM, 75 μM, and 375 μM concentrations of phenazine and less than 400 mM of sodium chloride. However, at the substantial concentration of 800 mM sodium chloride (shown as orange traces in the bottom row of Fig. 3.13), cells with only 37.5 μM phenazine largely died between 48 hours and 72 hours; more phenazine at 75 μM enabled longer survival, and 375 μM phenazine mostly protected the cells from much death for up to 120 hours.

Accompanying these trajectories in cell abundance, the total currents sustained by these populations of cells also vary systematically as shown in the lower axes of each plot in Fig. 3.13. These trajectories in metabolic rate are also visualized by regarding these data on a per-cell basis, with per-capita metabolic rates shown in Fig. 3.14. Some general trends are manifest. Especially during the initial transient increases of measured current at early times (e.g. up to ∼twelve hours) under some conditions, larger levels of phenazine tend to manifest larger currents (as reflected by the rightmost column of Fig. 3.13). Clearly, more salt slows how quickly cells rev up their metabolic activity in the first twelve hours under the high (375 μM) phenazine condition, hinting at an evidently but curiously antagonistic relationship between phenazine and ionic context for cells as they initially strive to acclimatize (see the rightmost column of Fig. 3.14). However, the general impact of different salt concentrations is much more context- and even time- dependent, a theme we now discuss.

As shown by the surface plots in Fig. 3.15 and the full complement of two-dimensional slices in Fig. 3.16 and Fig. 3.17, the geometry of the relationship the metabolic rate per-capita to external salt and phenazine concentration changes in slope, directionality, and monotonicity over time. For instance, between zero and twelve hours (Fig. 3.15(A)

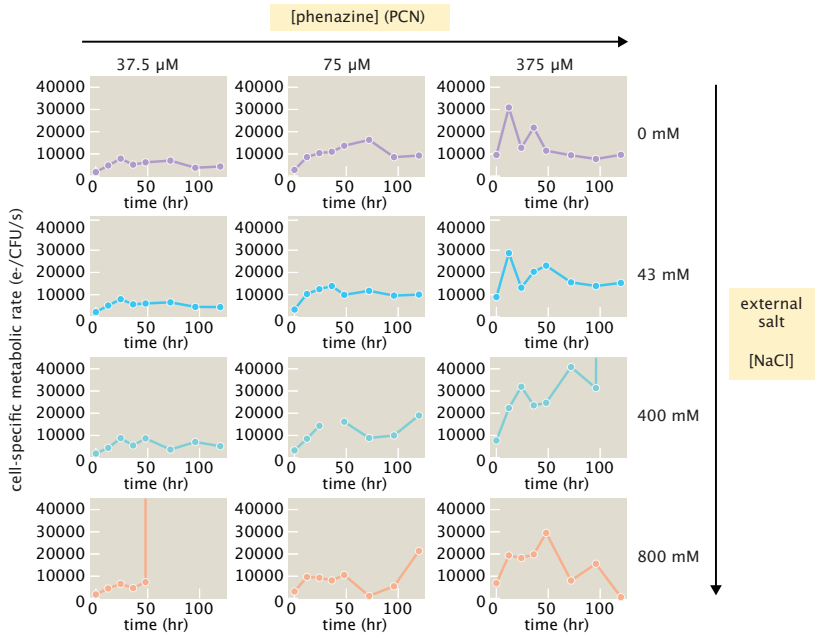


Figure 3.14: Impact of salt and external phenazine concentration on per-cell metabolic rate over time. These are are the same data as shown earlier in Figure 3.13, but visualized as a ratio to give a current per-cell. As discussed earlier in Reference [5], a current of one electron per second per cell corresponds to approximately the energy equivalent of 1 ATP hydrolysis's worth of energy per second per cell. (Here, in calculating the scale of the y-axis of metabolic rate per cell, we took the working volume of wells that cells were grown in to be 150 μL, as previously described [5]. Note that very rapid increases in per-cell metabolic rate are probably attributable to crashes in population and should be regarded as subject to the accompanying uncertainty.)

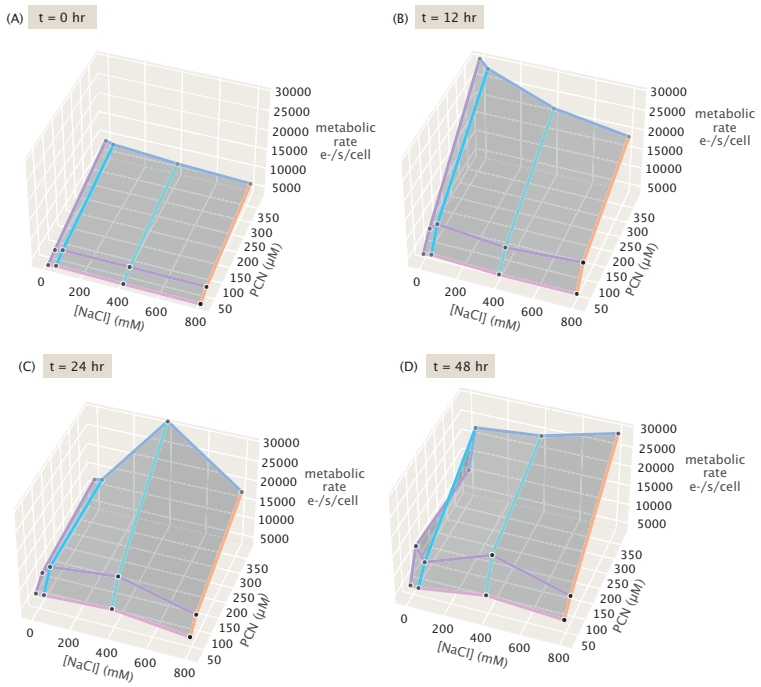


Figure 3.15: The relationships between metabolic rate per cell to phenazine and sodium chloride changes over time. (A)-(D) Metabolic rate measurements as a function of the three phenazine and four sodium chloride concentrations at timepoints $t = 0, 12, 24$, and 48 hr.

versus Fig. 3.15(B)), metabolic rate increases its steepness of variation with respect to both salt and phenazine. At intermediate times (e.g. at 24 hours), these relationships even develop peaks—for instance, metabolic rate gains a local maximum at the intermediate value of 400 mM sodium chloride, reproducible across the three phenazine concentrations (see the locally-maximal measurements connected by the light blue line in Fig. 3.15(C)). We speculate that since ions can affect or challenge virtually all aspects of cellular biochemistry, and 800 mM of external salt is a considerable ionic insult, the fact that metabolic rates can often be slower or saturating at this salt concentration may be connected to retarding a number of separate metabolically-demanding activities than those specifically associated with ion pumping. However, adjudicating such speculation demands much deeper measurements and experiments.

Many of the physiological surfaces of Figs. 3.15-3.17 insinuate cross-talk between

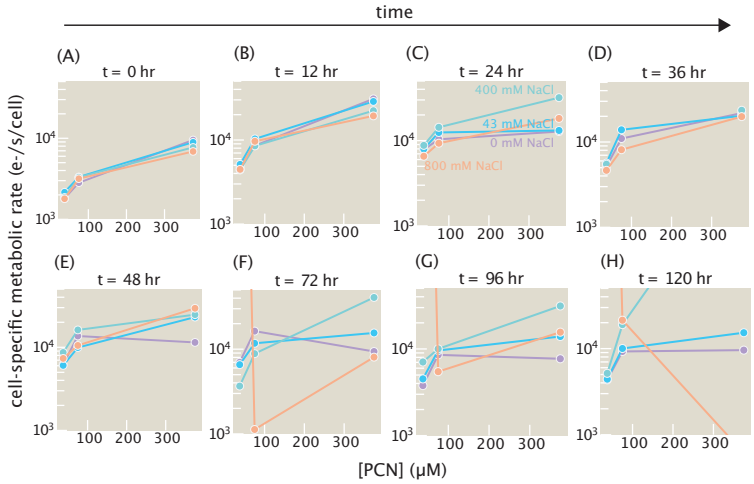


Figure 3.16: Metabolic rate per cell versus external phenazine concentrations, at increasing sodium chloride levels (indicated by purple, blue, teal, and orange lines, respectively), over time. (Note that at late times (72 hr and later), some populations have died to very small abundances, giving some frantically-varying curves to be regarded with appropriate uncertainty.)

cellular responses to external salt and external phenazine. For instance, whereas metabolic rate looks to mildly increase then gently saturate with salt concentration for low (37.5 μM and 75 μM) concentrations of phenazine, the 375 μM phenazine condition shows a continuous *decrease* in metabolic rate with external salt, as is conspicuous in the zero-hour and twelve-hour timepoints shown in Fig. 3.15(A-B) or Fig. 3.17(A-B). We hazard the tentative speculation that such interrelationships could be tied to the fact that some efflux pumps and transporters involved in exchanging toxins between the cell and the environment are coupled to sodium ions as well [19]; thus, if phenazine must be exported by the cell when its ambient concentration becomes harmful, sodium ion homeostasis may operate differently. We propose that these speculations are precisely and readily testable.

At last, we return to the quantitative question posed at the beginning of our discussion and ask quantitatively how dramatically external salt changes the metabolic demand of *Pseudomonas*. Unlike the tidy linear increase of metabolic rate with salt inferred for *E. coli* by Schink [4] (discussed in the preceding chapter), the repertoire of behaviors measured of *Pseudomonas* in Fig. 3.17 are not manifestly linear in salt, and are subject to the context

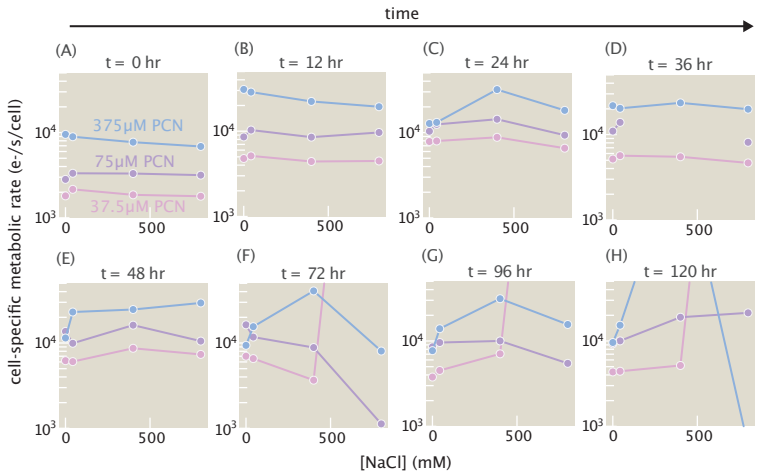


Figure 3.17: Metabolic rate per cell versus external sodium chloride concentrations, at different phenazine levels (indicated by blue, pink, and purple lines), over time. Note that at late times (72 hr and later), some populations have died to very small abundances, giving some frantically-varying curves to be regarded with appropriate uncertainty.

dependencies we just discussed. However, across these times and conditions, we see that the largest changes in metabolic rates per cell with increments of sodium chloride are much more gentle than implicated for *E. coli*. Specifically, even the instantaneous slopes of the most dramatic of these relationships are at most,

$$\kappa_{P.\ aeruginosa} \lesssim \text{few} \times 10 \text{ ATP/s/cell/(mM NaCl)}. \tag{3.41}$$

For instance, Fig. 3.18 illustrates that (for the modest perturbations of 400 mM external sodium chloride and below) and a modest level of (physiologically-well tolerated) phenazine, metabolic rate changes only of order $\kappa \lesssim 7$ ATP/s/cell/(mM NaCl). This is hugely different than *E. coli*'s number we referred to earlier ($\kappa_{E.\ coli} \sim 10^3$ ATP/s/cell/(mM NaCl)).

Two factors may contribute to the much smaller sensitivity we observed in anaerobic *Pseudomonas*'s metabolic response to salt compared to that of aerboic *E. coli*. First, these strains may themselves exhibit intrinsically different aversions to sodium ions (and

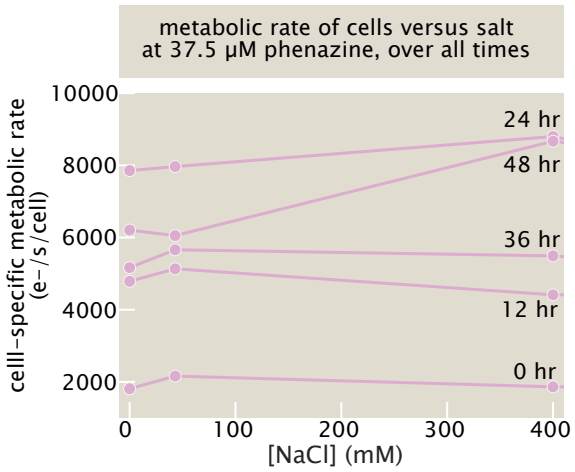


Figure 3.18: Representative subset of data, with early to intermediate timepoints ($t \leq 48$ hr) displayed, for relatively mild salt concentrations ([NaCl] $\leq$ 400 mM) for a modest, physiologically-well-tolerated phenazine concentration. Compared to the sensitivity inferred for *E. coli*, the per-cell metabolic rate for these *P. aeruginosa* clearly changes more gently over these salt ranges.

therefore pay very different pumping and other costs). Studies report that some *Pseudomonas* can allow the accumulation of cytosolic sodium to levels of order 80% of the extracellular medium in ranges of several tens of mM external sodium [20]; this apparent tolerance is in contrast to some reports on *E. coli*'s tight regulation of internal sodium concentrations [21]. If *E. coli* is intriniscally more halophobic than *Pseudomonas*, it may pay more to maintain ion homeostasis when the environment demands it. Second, the presence or absence of oxygen may fundamentally change the insults delivered by salts to many aspects of the functioning cell; in particular, damage to cellular components may occur via mechanisms dependent on salt and oxygen jointly. Absent oxygen, then, the repair costs attending these processes may be required at lower rates, overall causing less metabolic burden in response to salt.

More generally, the rich physiological manifolds hinted at in this preliminary analysis beckon to be reconciled with much more extensive measurements, perturbations, and attendent theory-building.

Bibliography

[1] Tori M Hoehler and Bo Barker Jørgensen. Microbial life under extreme energy limitation. *Nature Reviews Microbiology*, 11(2):83–94, 2013.

[2] Rachel M Walker, Valeria C Sanabria, and Hyun Youk. Microbial life in slow and stopped lanes. *Trends in microbiology*, 32(7):650–662, 2024.

[3] William R Shoemaker, Stuart E Jones, Mario E Muscarella, Megan G Behringer, Brent K Lehmkuhl, and Jay T Lennon. Microbial population dynamics and evolutionary outcomes under extreme energy limitation. *Proceedings of the National Academy of Sciences*, 118(33): e2101691118, 2021.

[4] Severin Schink, Mark Polk, Edward Athaide, Avik Mukherjee, Constantin Ammar, Xili Liu, Seungeun Oh, Yu-Fang Chang, and Markus Basan. Survival dynamics of starving bacteria are determined by ion homeostasis that maintains plasmolysis. *Nature Physics*, 20(8):1332–1338, 2024.

[5] John A Ciemniecki, Chia-Lun Ho, Richard D Horak, Akihiro Okamoto, and Dianne K Newman. Mechanistic study of a low-power bacterial maintenance state using high-throughput electrochemistry. *Cell*, 187(24):6882–6895, 2024.

[6] Severin J Schink, Elena Biselli, Constantin Ammar, and Ulrich Gerland. Death rate of *E. coli* during starvation is set by maintenance cost and biomass recycling. *Cell systems*, 9(1):64–73, 2019.

[7] Elena Biselli, Severin Josef Schink, and Ulrich Gerland. Slower growth of *Escherichia coli* leads to longer survival in carbon starvation due to a decrease in the maintenance rate. *Molecular systems biology*, 16(6):e9478, 2020.

[8] Rolf Sander. Compilation of Henry's law constants (version 5.0. 0) for water as solvent. *Atmospheric Chemistry & Physics*, 23(19), 2023.

[9] Elise S Cowley, Sebastian H Kopf, Alejandro LaRiviere, Wiebke Ziebis, and Dianne K Newman. Pediatric cystic fibrosis sputum can be chemically dynamic, anoxic, and extremely reduced due to hydrogen sulfide formation. *MBio*, 6(4):e00767–15, 2015.

[10] Rudolf K Thauer, Anne-Kristin Kaster, Henning Seedorf, Wolfgang Buckel, and Reiner Hedderich. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nature Reviews Microbiology*, 6(8):579–591, 2008.

[11] Kai Schuchmann and Volker Müller. Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nature Reviews Microbiology*, 12(12):809–821, 2014.

[12] Georg Fuchs. CO_2 fixation in acetogenic bacteria: variations on a theme. *FEMS Microbiology Reviews*, 2(3):181–213, 1986.

[13] Alexander J Krause, Graham A Shields, Robert J Newton, and Benjamin JW Mills. Modelling sulfate concentrations in the global ocean through phanerozoic time. *Journal of the Geological Society*, 181(4):jgs2023–184, 2024.

[14] JK Böhlke, GE Ericksen, and K Revesz. Stable isotope evidence for an atmospheric origin of desert nitrate deposits in northern Chile and southern California, USA. *Chemical Geology*, 136 (1-2):135–152, 1997.

[15] Wenping Cui, Robert Marsland III, and Pankaj Mehta. Diverse communities behave like typical random ecosystems. *Physical Review E*, 104(3):034416, 2021.

[16] Yun Wang, Suzanne E Kern, and Dianne K Newman. Endogenous phenazine antibiotics promote anaerobic survival of *Pseudomonas aeruginosa* via extracellular electron transfer. *Journal of bacteriology*, 192(1):365–369, 2010.

[17] Nathaniel R Glasser, Suzanne E Kern, and Dianne K Newman. Phenazine redox cycling enhances anaerobic survival in *Pseudomonas aeruginosa* by facilitating generation of atp and a proton-motive force. *Molecular microbiology*, 92(2):399–412, 2014.

[18] Alexa Price-Whelan, Lars EP Dietrich, and Dianne K Newman. Pyocyanin alters redox homeostasis and carbon flux through central metabolic pathways in *Pseudomonas aeruginosa* pa14. *Journal of bacteriology*, 189(17):6372–6381, 2007.

[19] Wakano Ogawa, Yusuke Minato, Hayata Dodan, Motoyasu Onishi, Tomofusa Tsuchiya, and Teruo Kuroda. Characterization of mate-type multidrug efflux pumps from *Klebsiella pneumoniae* mgh78578. *PloS one*, 10(3):e0121619, 2015.

[20] Tibor I Matula and Robert A Macleod. Penetration of *Pseudomonas aeruginosa* by sodium chloride and its relation to the mechanism of optical effects. *Journal of Bacteriology*, 100(1): 411–416, 1969.

[21] Chien-Jung Lo, Mark C Leake, and Richard M Berry. Fluorescence measurement of intracellular sodium concentration in single *Escherichia coli* cells. *Biophysical journal*, 90(1):357–365, 2006.

Part II

Genes and networks out of equilibrium

Chapter 4

Flexibility and sensitivity in gene regulation out of equilibrium

Cells adapt to environments and tune gene expression by controlling the concentrations of proteins and their kinetics in regulatory networks. In both eukaryotes and prokaryotes, experiments and theory increasingly attest that these networks can and do consume biochemical energy. How does this dissipation enable cellular behaviors forbidden in equilibrium? This open question demands quantitative models that transcend thermodynamic equilibrium. Here we study the control of simple, ubiquitous gene regulatory networks to explore the consequences of departing equilibrium in transcription. Employing graph theory to model a set of especially common regulatory motifs, we find that dissipation unlocks nonmonotonicity and enhanced sensitivity of gene expression with respect to a transcription factor's concentration. These features allow a single transcription factor to act as both a repressor and activator at different concentrations or achieve outputs with multiple concentration regimes of locally-enhanced sensitivity. We systematically dissect how energetically-driving individual transitions within regulatory networks, or pairs of transitions, generates a wide

range of more adjustable and sensitive phenotypic responses than in equilibrium. These results generalize to more complex regulatory scenarios, including combinatorial control by multiple transcription factors, which we relate and often find collapse to simple mathematical behaviors. Our findings quantify necessary conditions and detectable consequences of energy expenditure. These richer mathematical behaviors—feasibly accessed using biological energy budgets and rates—may empower cells to accomplish sophisticated regulation with simpler architectures than those required at equilibrium.

4.1 Introduction

Gene regulation—to which biology owes much of its exquisite sophistication [1]—is achieved by many network architectures that allow (and credibly depend on) energy expenditure [2, 3, 4, 5]. To adapt to varying environments, cells often dynamically tune concentrations of transcription factors [6] or inducers as their available control variables. This biochemical control adjusts the probabilities of cellular states by regulating rate constants that depend on the transcription factor or effector. Organisms use the specific shapes of these input (transcription factor concentration) to output (transcription) relationships to execute extraordinary signal processing [7]. Given their centrality, these induction curves also promise to help clarify how spending biochemical energy empowers the very dynamism and fidelity of the living. While unexplained energy expenditures in aerobically-growing bacteria have been documented since at least the 1980s [8], new measurements and models [9, 10, 11, 12] confirm that cells often operate with surprising surpluses in energetic capacity relative to known expenditures. These observations place fresh urgency on deciphering how dissipation modifies vital functions such as gene regulation [13].

How can nonequilibrium relieve fundamental constraints on physiological adaptation,

or enhance the flexibility of cellular behavior? To confront this question, here we investigate the nonequilibrium output behaviors of networks formed by regulatory binding proteins. These systems can represent the dynamic behaviors of genetic transcription executed by RNA polymerase (RNAP) and regulated by transcription factors as control variables (Fig. 4.1). While our analysis generalizes to accommodate more complex regulation, including combinatorial action by many transcription factors and DNA looping (Fig. 4.2 and Supporting Information), we focus on the widely-relevant case of a single transcription factor governing a cycle of four states. This architecture is putatively the most common regulatory architecture in well-characterized bacteria (Fig. 4.2). This motif is also among the simplest closed systems capable of breaking equilibrium using basic reactions pervasive in biology. Accordingly, quantifying the behavior and limitations of the square graph is also a crucial step towards understanding more complex transcriptional networks.

Given their simplicity, equivalents of the system in Fig. 4.1 have enjoyed earlier study in guises such as enzymatic control [14]; remodeling of nucleosomes [5]; and other settings in transcription [15, 16]. In this work, we use tools from graph theory [17, 18] to explore the full space of transcriptional steady-state outputs available for this system under different energetic drives, compared to equilibrium control. We find that all equilibrium responses must be monotonic (with one inflection point) as a function of control variables, such as the concentration of transcription factor, measured in a conventional logarithmic scale. In contrast, we discover that nonequilibrium models can exhibit three types of output: an "equilibrium-like," monotonic response with one inflection point, potentially displaced from equilibrium; a new—but still-monotonic—shape with three inflection points; and a new, surprising non-monotonic shape with two inflection points, where, for instance, increasing a control variable can change its effect from repression to activation. Combining analytical and numerical analysis, we globally bound the maximal sensitivities

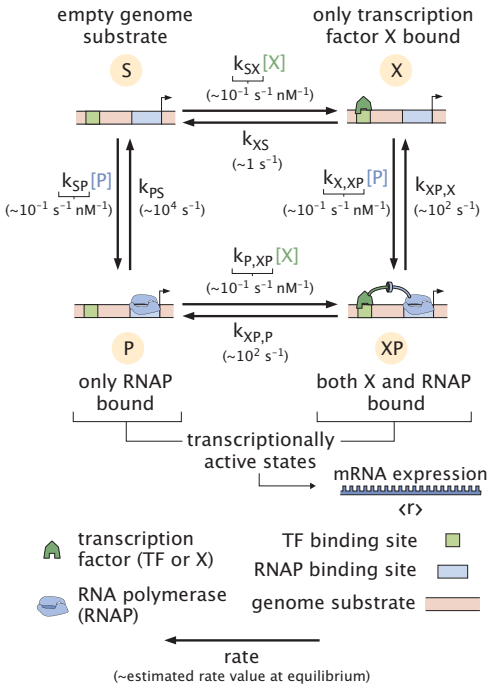


Figure 4.1: Structure of a fundamental gene-regulatory motif. A square cycle of four-states emerges when up to two molecules (such as a transcription factor X and polymerase P) can bind to a common substrate (say a genome). Output observables $\langle r \rangle$ are linear combinations of the state probabilities; for instance, mRNA production scales with the probabilities of transcriptionally active states where polymerase is bound to the genome (states P and XP). These outputs vary with the control parameter $[X]$, here schematized as the concentration of a transcription factor. (Numerical values of rate constants underneath edges are biologically plausible values, as inferred in the SI §2B.)

of transcriptional responses. Demonstrating that these mathematical behaviors are feasible to access within biological energy expenditures around typical rates, we systematically analyze the impact of breaking detailed balance along each transition rate. This analysis establishes design principles for optimizing sensitivity and unlocking behaviors that are especially prone to implicate nonequilibrium in measurements.

These broader, multiply-inflected transcriptional responses unlocked by nonequilibrium could be harnessed to achieve useful physiological functions. Our findings illustrate surprising regularity revealed by graph theoretic tools, and dissect how even primordial biological networks operating out of equilibrium can rival the regulatory sophistication of (plausibly) larger, slower networks at equilibrium.

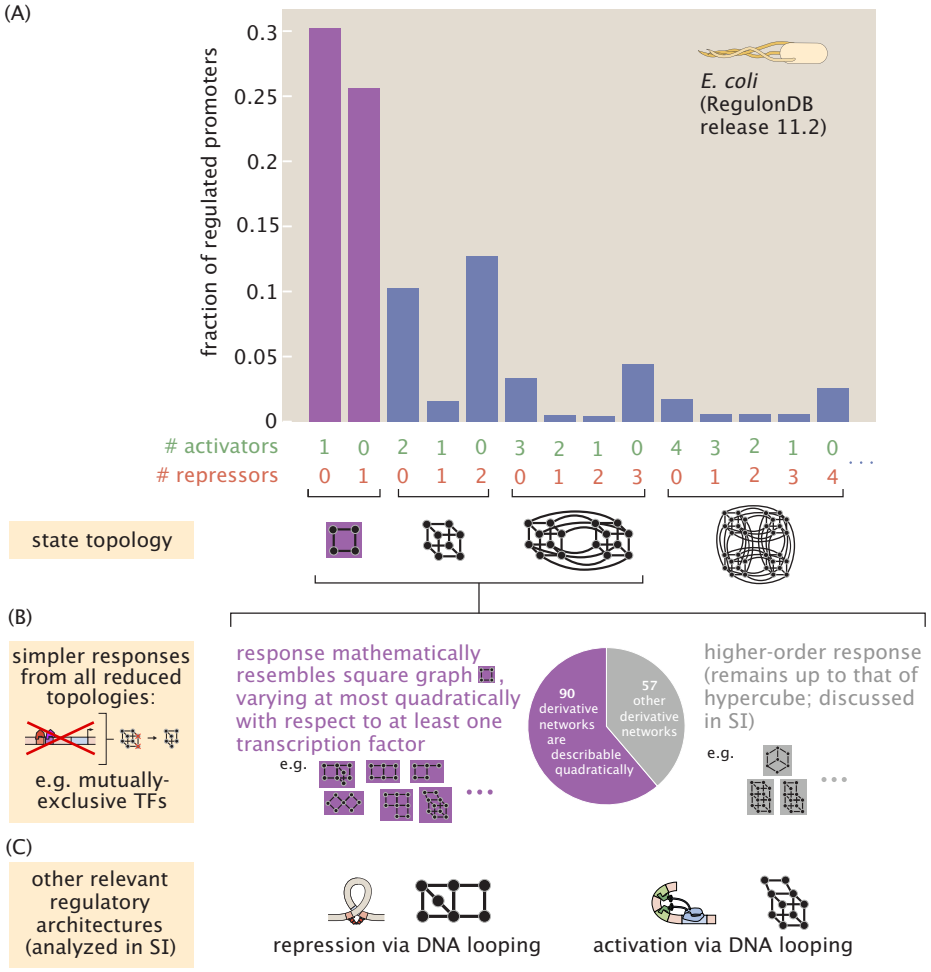


Figure 4.2: Incidence of regulatory architectures for transcription, where a square cycle's response is common. (A) Regulatory architectures reported in *E. coli*, organized by the numbers of activating and repressing transcription factors associated with each regulated promoter, according to data in the RegulonDB database [19] release 11.2. Over half of these regulated promoters are reported to be regulated by the most common architecture of a single transcription factor (purple bars), and thus can be described by a square graph of states (Fig. 4.1). Less common network architectures with multiple transcription factors form hypercubic state spaces whose complete responses are analysed in the Supporting Information (SI). (B) While larger binding networks can generically show more complex regulatory responses than the square cycle, simpler topologies and responses result from restrictions like overlapping binding site (see SI §3D). Over half of all nontrivial transcriptionally-potent derivative networks formed by up to three transcription factors show a response that resembles the square graph with respect to at least one transcription factor. (C) Transcriptional regulation by DNA looping accomplishes distinct responses, whose properties are analyzed algebraically and numerically in the SI.

4.2 Results

4.2.1 Modeling pervasive gene regulatory motifs

At steady-state, a system is in *equilibrium* (or, equivalently, at *detailed balance*) if, for all pairs of states (i, j), the probability flux $k_{ij}p_i$ into state j from state i equals the flux $k_{ji}p_j$ into state i from state j, where p_i is the probability of state i and k_{ij} is the transition rate from state i to j. Otherwise, the system is out of equilibrium and requires energetic dissipation to sustain the system's steady-state. Nonequilibrium steady-states can only be achieved with systems that contain at least one cycle; linear or branched architectures at steady-state must be at equilibrium (for systems closed to external material inputs; see Supporting Information (SI), §1B and [20, 21]). A single cycle is thus the simplest closed setting where the intriguing new consequences of nonequilibrium become possible.

A cycle of four states emerges naturally from up to two molecules binding or unbinding to a substrate. When the substrate S is a promoter site on the genome, one molecule is RNA polymerase P, and the second molecule is a transcription factor protein X that can enhance or impede polymerase binding to the genome, the resulting cycle captures transcriptional regulation. Specifically, the four states represent the empty site of the genome substrate ("S"); the genome substrate bound to the transcription factor only ("X"); to the polymerase only ("P"); or to both ("XP"). Figure 4.1 illustrates this central, motivating setting. (Note that the transcription factor and polymerase concentrations $[X]$ and $[P]$ do not affect whether the system is in or out of equilibrium, and can be tuned while maintaining any extent of disequilibrium; see SI §1C.)

This square cycle of states pervades gene regulation. We analyzed the incidence of regulatory architectures reported for *E. coli*, according to the most recent release of the canonical *RegulonDB* database of known interactions [19]. As summarized in Figure 4.2A, over half of all regulated promoters in *E. coli* are reported to accommodate a single copy of

a transcription factor. The four-state cycle also finds widespread examples or structural-equivalents in models of eukaryotic gene regulation [15, 5, 22, 23]. Accordingly, these promoters and genes can be modeled by the four-state cycle in Figure 4.1. These cyclic architectures also contrast the more commonly-studied, non-cyclic, motif of simple repression that cannot break detailed balance (see SI §1B) [1, 24, 25, 6, 26]. We also provisionally quantified an empirically-common role for the square graph motif in the eukaryote Drosophila, according to interactions reported by the *Redfly* [27] database (see SI §2A.2). However, this apparent commonality of a single transcription factor binding site per gene is probably more an artifact of our field's substantial regulatory ignorance for eukaryotes than a well-tested empirical finding, since only $\sim 1.8\%$ of genes are canonically annotated with known regulatory interactions in Drosophila.

Figure 4.2 also documents other more complex regulation, including by multiple transcription factors and via DNA looping, used by prokaryotes and eukaryotes. Tellingly, eukaryotic gene expression is a setting where explicit ATP-consumption is especially plausible [4, 3] yet poorly understood [2, 13, 15]. These larger networks often form hypercubic state spaces or variants thereof. We analyze extensively these more complex networks, and their reachable behaviors, in the Supporting Information, S. Often, however, we find that biophysical constraints (such as overlapped binding sites) compel even these more complex networks to resemble the mathematical behavior of the square graph with respect to at least one transcription factor as a control variable (Figure 4.2B and SI §3D). In sum, the prevalence of architectures that are exactly square cycles across organisms—and those that mathematically resemble the square cycle—establish the importance of understanding the regulatory capabilities of the square graph. Further, understanding how the square graph operates is a key prerequisite for understanding more complex regulatory networks. However, for completeness, we derive how a wide variety of other common regulatory architectures perform using similar analyses in the SI (§3).

Kinetic measurements often justify the assumption that transcription factors bind and unbind with genomes quickly relative to transcription by polymerase. This separation of timescales makes macroscopic gene expression proportional to the steady-state probability of finding the system in transcriptionally-active microstates. (We precisely validate this assumption for our setting using plausible transcriptional rates in the SI §2C.)

We note that the average gene production rate $\langle r \rangle_{\text{mRNA}}$, proportional to gene expression, is a typical and crucial output of interest. This response grows with the net probability that the polymerase is bound, $\langle r \rangle_{\text{mRNA}} = r(p_P + p_{XP})$, where r is the transcription rate once the polymerase is bound, p_p is the probability of the state P where just the polymerase is bound, and p_{XP} is the probability of the state XP where both polymerase and transcription factor are bound.

However, other outputs (that depend on other states) may also be biologically or experimentally significant. For instance, the localization of the transcription factors themselves to the genome (to recruit other co-factors or epigenetic modifications) can shape biological function independent of the polymerase, e.g. invoking the probability p_X. We accommodate the breadth of these possible outputs by studying how any (nonnegative) linear combination $\langle r \rangle = \sum_{\text{states } i} r_i p_i$ of state probabilities varies with the transcription factor concentration X as a control variable, where r_i gives the potency of the ith state. These different outputs and problem settings are captured by adopting particular $\{r_i\}$, but as we will now see, all are subject to universal behavior.

4.2.2 Nonequilibrium steady-state output responses

To explore how these input-output functions operate away from equilibrium, we cannot apply equilibrium statistical mechanical models that assign thermodynamic energies to each state to calculate their probabilities [1, 28, 6, 29, 30]. Instead, we now require a fully kinetic description (also known as a chemical master equation or continuous-time Markov

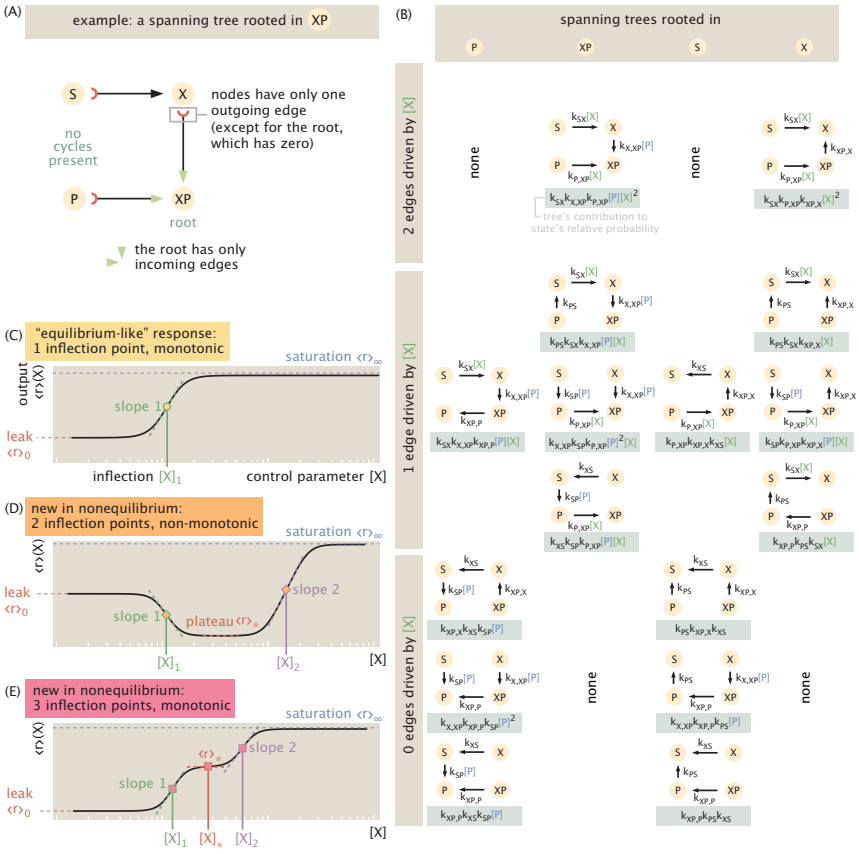


Figure 4.3: Nonequilibrium response of the four-state fundamental transcriptional motif. (A) An example of a spanning tree (rooted in state XP) like those that define steady-state probabilities via the Matrix Tree Theorem. (B) All 16 directed, rooted spanning trees of the four-state cycle in 1(A): trees are grouped by the root state (in columns) and by how many participating edges depend on the control parameter X (in rows). As guaranteed by the Matrix Tree Theorem, the steady-state probability of any state—in or out of equilibrium—is given by the sum of the weights of these spanning trees, introducing up to a quadratic dependence in X in any output, as represented by Eq. 4.1. (C-E) Three universal output behaviors (*regulatory shape phenotypes*) can result from this architecture. A monotonic "equilibrium-like" output (C) manifests a Hill-like or MWC-like response, behavior familiar from equilibrium thermodynamic models. However, exclusively out of equilibrium, new multiply-inflected regulatory shape phenotypes become possible. Under drive, outputs can (D) vary non-monotonically and reach two inflection points with the control parameter; or show three inflection points and vary monotonically (E). These richer phenotypes show a wider set of properties that characterize each curve: these include the "leak" value of the observable when the control variable is absent ($\langle r \rangle_0 = \langle r \rangle([X] = 0)$, in orange; the saturation asymptotic limit as the control variable is maximally present ($\langle r \rangle_\infty = \lim_{[X]\to\infty} \langle r \rangle$; in light blue); the observable's values at intermediate plateau regions ($\langle r \rangle_*$; in red); and slopes 1 and 2 at inflection points $[X]_1$ and $[X]_2$ when they are defined (in green and purple, respectively).

chain) based on transitions between states. A large increase in complexity and the number of parameters typically accompanies this generalization. Fortunately, these dynamics admit a beautiful and powerful correspondence to graph theory that helps tame this complexity [17, 18, 20]. Our guide is the Matrix Tree Theorem, which gives a simple diagrammatic procedure on a network's structure to find stationary probabilities (see Methods and SI §2D).[1] In brief, the Matrix Tree Theorem asserts that at steady-state, the probability of any state is proportional to the sum of products of rate constants over all spanning trees rooted in that state. A rooted spanning tree of a graph G is a subset of (directed) edges that collectively visits every state exactly once, has no cycles, and designates a specific root vertex u. As illustrated by the example in Figure 4.3A, each state in a spanning tree has exactly one outgoing edge (except the root state, which has only incoming edges); thus, spanning trees contain a unique directed path from any state to the root state.

Counting all sixteen rooted spanning trees of the four-state transcriptional system (Figure 4.3B) and deploying the Tree Theorem explains how probabilities must vary with the transcription factor control parameter $[X]$. Depending on the root (separated by column in Figure 4.3B), each spanning tree carries two edges that depend on $[X]$ (top row of Fig. 4.3B); one edge (middle row, Fig. 4.3B); or no $[X]$−dependent edges (bottom row, Fig. 4.3B). This structure yields relative probabilities with up to quadratic scaling with $[X]$. Hence we find that, for this class of architectures, the form of any output function $\langle r \rangle$, in or out of equilibrium, is a ratio of quadratic polynomials in $[X]$,

$$\langle r \rangle = \frac{A + B[X] + C[X]^2}{D + E[X] + F[X]^2}, \tag{4.1}$$

[1] Though the square graph could also be analyzed directly by (more arduous) linear algebra, this setting already shows how the Matrix Tree Theorem dramatically simplifies practical calculations and organizes terms in powers of the control variable (as shown in Fig. 4.3B). It also shows power, calculationally and conceptually, in more complex graphs where linear-algebraic operations become forbidding (see SI, §3).

where the coefficients A, B, C, D, E and F are sums of subsets of (weighted) directed spanning trees carrying various $[X]$-dependencies (see SI §2D). The denominator, the sum of all rooted spanning trees and hence also a quadratic polynomial, serves as a normalizing factor that converts statistical weights to probabilities and represents a nonequilibrium partition function.

Note that while we derived the output form Eq. 4.1 using the particular choice of $[X]$-dependent arrows appropriate for this transcriptional setting, the same formalism can treat many other control parameters that appear quite (structurally or biologically) distinct from these details, such as a concentration of another internal molecule (for instance polymerase, $[P]$) or an external molecule (for instance explicit drive by $[ATP]$). Specifically, any graph need only display up to two powers of the control variable $[X]$ among its spanning trees—a modest structural condition—for its response to exactly reproduce that of the square graph. The SI §2H gives further examples of different placements of controlled edges that still produce a network output with the functional form of Eq. 4.1, and therefore remain precisely addressable by the analysis of this paper. Other outputs, like those of the more complicated networks of multiple transcription factors analyzed in the SI, invite a fresh application of the Matrix Tree Theorem but benefit from the same framework; see the Supplementary Information for an ensemble of such analyses relevant to transcription (SI §3).

4.2.3 Equilibrium output curves are constrained

Eq. 4.1 describes all input-output curves, in or out of equilibrium, produced by this four-state transcriptional system. When detailed balance does hold, this equation becomes equivalent to thermodynamic statistical-mechanical models (as it must). We explain algebraic correspondences to thermodynamic models, like those communing with earlier transcriptional experiments [6, 30], in the SI §2G.3. Importantly, we find that the

equilibrium condition demotes any observable output to the simpler form of a ratio of *linear* polynomials in $[X]$, namely

$$\langle r \rangle^{\text{eq}} = \frac{A' + B'[X]}{C' + D'[X]}, \tag{4.2}$$

for constants $\{A', B', C', D'\}$ set wholly by thermodynamic parameters (see the SI §2G.1). Not coincidentally, this functional form formally reproduces or evokes the Hill induction, Michaelis-Menten, Langmuir-binding, Monod-Wyman-Changeux, or two-state Fermi function forms from the equilibrium statistical mechanics of binding commonly used to model and fit induction curves in natural [6, 31] or synthetic [32] settings. This equilibrium curve is paradigmatic of our biochemical intuition—smoothly saturating, with one point of inflection, with respect to transcription factor concentration $[X]$ in a conventional logarithmic scale (see Fig. 4.3A and SI §2E).[2]

While earlier insightful work identified raw algebraic expressions similar to Eqs. 4.1 and 4.2 in different contexts (including References [16, 34]), their functional flexibility has not been deeply explored, and such studies largely asked distinct questions (e.g. focusing on sensitivity; see SI §2G.6). Next, we derive new biological consequences of these behaviors that represent precise contrasts between nonequilibrium and equilibrium regulation.

4.2.4 New regulatory shape phenotypes unlocked by nonequilibrium

How much more complex is the regulation realizable by nonequilibrium outputs $\langle r \rangle$ (Eq. 4.1), compared to that of their equilibrium special case, $\langle r \rangle^{\text{eq}}$ (Eq. 4.2)? To reach the qualitative essence of this question, we first investigate the possible *shapes* of the output

[2]Throughout our analysis and discussion in this paper, we monitor the shape, number of inflection points, and sensitivity of transcriptional outputs with respect to the control parameter of the concentration of transcription factor, on a *logarithmic* scale. We use this logarithmic convention in alignment with common practice in biochemical and transcriptional studies [6, 32, 33].

curve. Specifically, we monitor the output's changes in concavity with respect to the control parameter. We postpone comment on the characteristic positions and scales of output curves—any shifts in their horizontal position (*viz.* characteristic concentration scales) or vertical expanses (*e.g.* maximally-induced responses)—until shortly.

Neglecting scales and shifts allows us to collapse the general, six-parameter output curve of Eq. 4.1 to a normalized function of just two emergent shape parameters,

$$\frac{\langle r\rangle - \langle r\rangle_0}{\langle r\rangle_\infty - \langle r\rangle_0} = \frac{ax + x^2}{1 + bx + x^2}, \tag{4.3}$$

Here, the emergent shape parameters a and b are complicated functions of the coefficients in Eq. 4.1 (and hence of underlying rate constants), and x is the governing concentration $[X]$ measured in terms of a characteristic concentration scale (all defined in the SI §2F).The values $\langle r\rangle_0 \equiv \langle r\rangle([X] = 0)$ and $\langle r\rangle_\infty \equiv \lim_{[X]\to\infty} \langle r\rangle$ are the *leakiness* (uninduced) and *saturation* (maximally-induced) responses; we return to these values in the following subsections. This representation preserves the concavity of the response function, allowing us to explore shapes and quantitative features in a two-dimensional space more efficiently and comprehensively than possible in the space of the eight rates.[3]

Harnessing this collapsed representation, we discover that all output curves assume just three different universal shapes (see Methods & SI §2I). First, the output can be "equilibrium-like:" monotonic and saturating with a single inflection point with respect to the control parameter (on a log scale), recalling the shape of the equilibrium response (Fig. 4.3C). Uniquely out of equilibrium, however, two additional multiply-inflected response shapes become possible. Under energy expenditure, outputs can become nonmonotonic and show two inflection points (Fig. 4.3D), or remain monotonic with three inflection

[3]The two-parameter simplicity of Eq. 4.3 is one possible nonequilibrium sophistication of the (usually one-parameter) data collapses used to unify simpler, equilibrium, two-state physiological responses [31] and regulation [6] in bacteria.

points (Fig. 4.3E), with respect to the log of the control parameter. Responses with three inflections are always shaped as depicted in Fig. 4.3E: maximally steep at the first and third inflection points, but minimally steep at the second inflection point. Clearly, these nonequilibrium curves are marked departures from simple equilibrium-like curves, but show a remarkable parsimony and regularity, given that they describe all departures from equilibrium for any rate parameter values. These three regulatory behaviors can pose different physiological implications for an organism; admit distinct quantitative constraints on sensitivity (as we will soon see); and require different conditions on underlying rate constants to be reached. In view of their categorical differences, we refer to these possible shapes as *regulatory (shape) phenotypes*.[4]

4.2.5 Leakiness, saturation, and EC50 are tunable at equilibrium

Does spending energy enable finer control over these quantitative traits, beyond growing their number? In fact, only some quantitative traits are given extra adjustability by spending energy. Without the transcription factor, the system cannot be found in any microstate that involves it, collapsing four states into just the two $\{S, P\}$ states. This pair of states forms an acyclic graph, so these steady-state probabilities must show detailed balance (i.e. are set purely thermodynamically). Thus, leakiness $\langle r \rangle_0$, determined exclusively by S and P states, can be adjusted freely while maintaining detailed balance. Analogously, when the transcription factor concentration is saturating ($[X] \to \infty$), the system is never found in the two microstates without the transcription factor, again admitting an orthogonal description of a balance between two states, now $\{X, XP\}$. Hence, saturation $\langle r \rangle_\infty$ is also freely adjustable at equilibrium. These leakiness and saturation values are independently adjustable by two separate energy parameters—the binding energies of the polymerase to

[4]We use the phrase "regulatory (shape) phenotype," referring to the overall shape of a response curve, to distinguish our meaning from the usage of Reference [2], who instead referred to specific *quantitative traits* within curves of a single mathematical shape (such as sensitivity or noise) as "regulatory phenotypes."

the genome when the transcription factor is absent or present, respectively. At equilibrium, once the leakiness and saturation are fixed by energy parameters, the response's maximal sensitivity (slope at the inflection point) is predetermined and no longer tunable, as revealed by its algebraic dependencies (see SI §2G.2). In contrast, while the location of the governing inflection point depends on these two energy parameters, it can also be tuned—remaining at equilibrium—using another energy parameter (the binding energy between the transcription factor and genome). (See SI §2G.2 for details.)

4.2.6 Nonequilibrium control of sensitivity obeys shape-dependent global bounds

Out of equilibrium, the sensitivity of responses enjoys greater adjustability. Specifically, the diversity of input-output curves accessible under drive motivates us to assess sensitivity by a suitably normalized slope $s([X])$, defined by

$$s([X]) \equiv \left| \frac{d\langle r \rangle}{d \ln \left({}^{[X]}/_{[X]_0} \right)} \frac{1}{\langle r \rangle_{\max} - \langle r \rangle_{\min}} \right|, \tag{4.4}$$

where $\langle r \rangle_{\min} \equiv \min_{[X]} \langle r \rangle$ and $\langle r \rangle_{\max} \equiv \max_{[X]} \langle r \rangle$ are the extremal values of the observable over all $[X]$, and $[X]_0$ is a characteristic concentration scale ensuring dimensional consistency. For monotonic curves, the maximum $\langle r \rangle_{\max}$ and minimum $\langle r \rangle_{\min}$ responses are necessarily the uninduced leakiness $\langle r \rangle_0$ and the maximally-induced saturation $\langle r \rangle_\infty$ (or vice-versa), whereas for nonmonotonic responses with two inflections, the maximal and minimal responses can occur at intermediate finite values of $[X]$. This normalized sensitivity $s([X])$ is directly related to familiar measures such as the logarithmic sensitivity and the effective Hill coefficient, but more naturally and tightly describes sensitivities of nonmonotonic phenotypes (see SI §2J).This measure parameterizes sensitivity with greater indifference to absolute scales of the function.

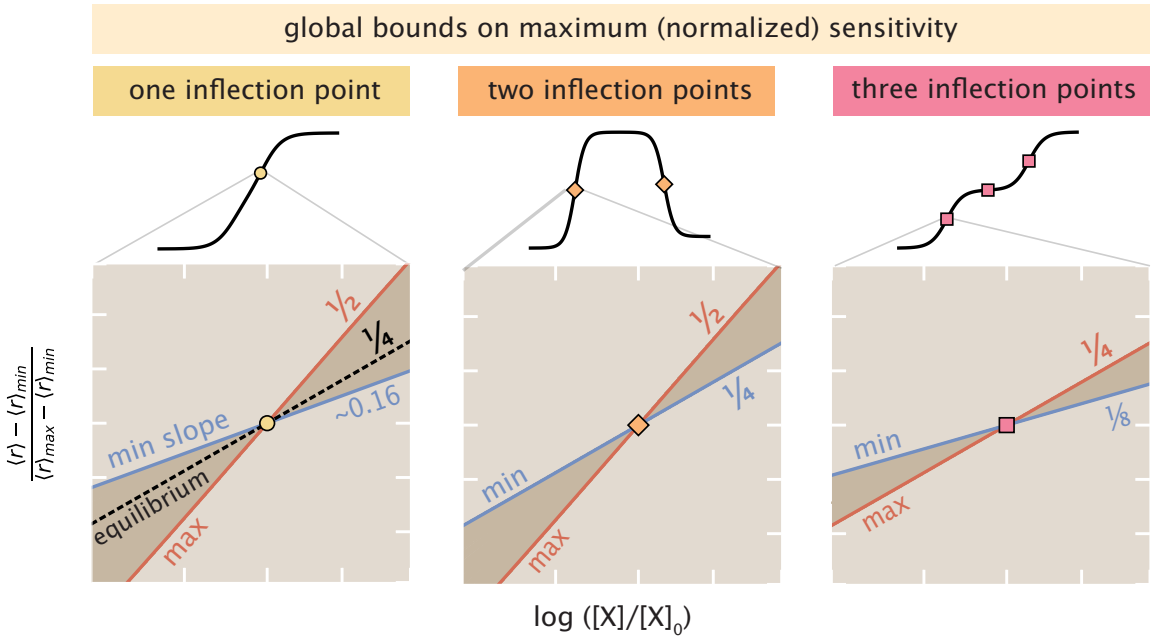


Figure 4.4: Global bounds, in or out of equilibrium, restrict maximal (normalized) response sensitivity (with respect to input concentrations $[X]$ on a log scale). The normalized responses $\frac{\langle r\rangle - \langle r\rangle_{\min}}{\langle r\rangle_{\max} - \langle r\rangle_{\min}}$ are plotted near points of inflection that maximize slope, separated by shape phenotype. When the output has one inflection point (left), the maximal sensitivity is bounded between a minimum of 0.16 (blue line) and a maximum of $1/2$ (red line) for any set of rate values or any dissipation; this subsumes the equilibrium case, whose normalized sensitivity is fixed at $1/4$ (black dotted line). When the output has two inflections (middle), the maximal sensitivity is bounded between $1/4$ and $1/2$. When the output has three inflections (right), the maximal sensitivity is bounded between $1/8$ and $1/4$.

By combining wide numerical sampling, symbolic inequality solving, and analytical arguments (see SI §2J), we investigated the maximal normalized sensitivity $s([X])$ any response curve can exhibit for the four-state system across its three possible shape phenotypes. We discovered that sensitivity is tightly bounded above and below by precise finite limits; these limits vary by phenotype. Figure 4.4 summarizes these bounds, visualized by how normalized and centered response curves $\frac{\langle r\rangle - \langle r\rangle_{\min}}{\langle r\rangle_{\max} - \langle r\rangle_{\min}}$ behave around inflection points of maximal slope. Equilibrium response curves always show a normalized sensitivity of exactly one-fourth. Out of equilibrium, singly-inflected response curves can increase this maximal sensitivity up to one-half, or *decrease* maximal sensitivity below the equilibrium value to a numerical value of about 0.16. (We lack a coherent explanation for this curious numerical lower bound, but verified it by precise symbolic inequality solving; see SI §2J). Nonquilibrium with two inflection points all have maximal sensitivity of *at least* the equilibrium level of one-fourth, but up to one-half. Nonequilibrium curves with three inflection points all show maximal sensitivity of *at most* the equilibrium level of one-fourth, and at least a sensitivity of one-eighth.

Cast in terms of the *raw* maximal sharpness $d\langle r\rangle/d\ln([X]/[X]_0)$ of each response curve, these bounds report that raw maximal sharpness is always between one eighth and one half of the distance between the maximum and minimum responses per $e \approx 2.7$-fold increase in the concentration $[X]$. We stress that these bounds on sensitivity, in terms of the observed $\langle r\rangle_{\min}$ and $\langle r\rangle_{\max}$, are tighter quantitative constraints than bounds merely in terms of the maximal or minimal transcription rates (or other potency values) $\max_i\{r_i\}$ or $\min_i\{r_i\}$ that any microstate of the system can show, as can be connected to recent, related upper bounds [33]. This follows since the extrema of the *average* observable response curve over all $[X]$ are usually more restricted than the most extreme potencies over microstates (namely, $\max_i\{r_i\} \geq \langle r\rangle_{\max}$ and $\min_i\{r_i\} \leq \langle r\rangle_{\min}$). (See SI §2J.4).

These findings emphasize that network architecture and dissipation are not the only

hard global constraints that bound sensitivity. The global shape of the response curve further categorically constrains the possible sensitivity. This relationship is potentially biologically relevant: for instance, it is impossible for an organism regulated by the square-graph transcriptional motif to achieve both a triply-inflected output curve and a normalized sensitivity greater than that at equilibrium. This represents a tradeoff between the shape complexity of a response and its maximal sensitivity.

4.2.7 Breaking detailed balance along each edge

Our foregoing analysis has been mathematically general. That is, the constrained shapes and bounds on sensitivity hold for any response following Eq. 4.1, over all rate constant values and energetic dissipations. These constraints also apply even—as previously noted—if the response is produced by a different underlying graph architecture than the particular transcriptional motif shown in Fig. 4.1, as long as the graph still yields spanning trees that depend up to quadratically on the control variable. Just because multiply-inflected or adjustable response curves are mathematically possible, however, does not establish that they are biologically plausible. To assess whether these behaviors can be accessed using physiologically-plausible amounts of energy expenditure or typical biological rates, we now specialize to the plausible particulars of transcription as in Fig. 4.1. In the remainder of this paper, we quantify the extent of dissipation sustaining a nonequilibrium steady-state by focusing on the free energy $\Delta\mu$ coupled to the system, with units of k_BT or Joule; we refer to this quantity as the *nonequilibrium driving force* or simply as the *(net) drive* (see SI §1D). In addition, we now adopt the transcriptional potencies $r_P = r_{XP} = 1$ and $r_S = r_X = 0$. This choice makes our response observable $\langle r \rangle_{\text{mRNA}}$ the probability that polymerase is bound to the genome. (More general responses can be easily analyzed analogously; see SI.)

Typical empirical binding energies, diffusion-limited rates, and single-molecule kinetic

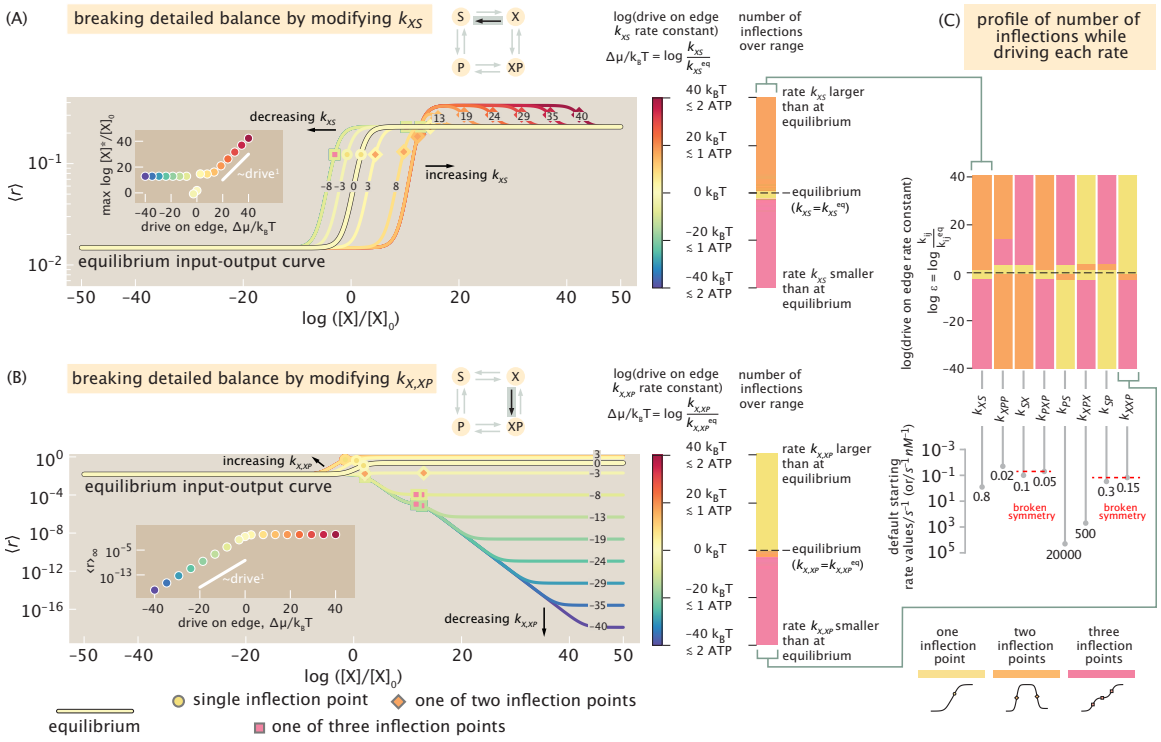


Figure 4.5: Systematically breaking detailed balance edge-by-edge. (A) Example of how spending energy to modify a single rate (here, k_{XS})—while the seven other rates remain fixed—changes the response curve away from default equilibrium behavior (pale yellow curve labeled "0" net drive and outlined in black). Responses from rate values larger than (or smaller than) at equilibrium are shown in increasingly red (or blue) colors, respectively; curves are also labeled with the numerical values of the net drive that generated them in k_BT units (positive for an increase; negative for a decrease). Each curve's resulting inflection points are marked by yellow, orange, or pink markers, denoting one to three inflection points (respectively), and summarized in the associated one-dimensional (shape phenotypic) phase-diagram with the same colors on the right. Inset: the position of the final inflection point $\max \ln [X]^*/[X]_0$ versus net drive (power law exponent is ~ 1). (B) Another representative behavior is displayed when $k_{X,XP}$ is instead the rate varied. Inset: the saturation $\langle r \rangle_\infty$ versus net drive (power law exponent is ~ 1). (C) Summary of how all eight rates respond to energy expenditure to realize different regulatory shape phenotypes. Below, stem plots give precise values of each default rate constant at equilibrium. (These rates satisfy initial "broken symmetries" that violate the conditions in Eq. 4.5 by default, facilitating more ready access to nonmonotonicity. The SI §2K documents the impact of departing from different default starting rates that instead satisfy Eq.4.5, in addition to the impacts of driving all other edges.) (Here, the reference concentration scale setting the horizontal offset of the concentration axis is $[X]_0 \equiv 1$ nM.)

measurements yield order-of-magnitude estimates for the eight rates governing transcription at equilibrium (see SI §2B and Fig. 4.1). First, we choose a set of default rates consistent with these orders-of-magnitude (given in the lower right stem plot of Fig 4.5C).

Next, we investigate how breaking detailed balance by spending energy to increase or decrease a single rate constant at a time—while keeping the seven other rates fixed at biological default values—modulates the transcriptional response curve. Hydrolyzing an ATP molecule makes available $\approx 20\ k_BT$ of energy (BNID 101701, [35]; [36]) that can be used as a chemical potential gradient to drive transitions (for instance, by powering an enzymatically-assisted pathway [37]). This amount of free energy is also the scale observed to power active processes like biomolecular motors [38]. Accordingly, to conservatively emulate a biological energy budget, we allot a maximum of just two ATP hydrolyses' worth of free energy, $|\Delta\mu| \leq\ 40\ k_BT$, to break detailed balance. This budget for drive allows a given individual rate to be scaled by up to a factor $\exp[\Delta\mu/k_BT] = \exp[\pm 40]$.

Applied edge-by-edge, this procedure reveals that biologically-feasible energy expenditures dramatically modify the response curve and easily attain all three regulatory shape phenotypes. Illustrating this regulatory plasticity, Fig. 4.5A shows how breaking detailed balance by scaling a rate up (increasingly red curves) or down (increasingly green-blue curves) can shift response curves to the left or right on the horizontal $\log[X]$ axis (effectively tuning what EC50 formerly represented at equilibrium), and also smoothly change the number of inflection points (see SI §2K for many further example response curves). Yet even for the same net nonequilibrium driving force, the consequences of breaking detailed balance depend significantly on the edge it is broken along. Fig. 4.5B shows another representative behavior by modifying a different edge, where the major effect of departing equilibrium is to modulate the leakiness, saturation, or intermediate scales of the response. Despite the diversity of this regulation, quantitatively-regular control behavior emerges as well: inset plots emphasize that phenotypic properties such as the position, $\max\{\log[X]^*\}$, of the final inflection point and the saturation, $\langle r \rangle_\infty$, scale as power laws with the net drive over some regimes.

This broad regulatory flexibility is sustained over all eight rate constants, whose

comprehensive response behaviors under drive are analyzed in the SI §2K. Fig. 4.5C summarizes how driving each rate attains different shape phenotypes (number of inflections). Notably, any rate can be driven to access any of the three response shape phenotypes at some small, biologically-feasible dissipation. Yet the minimum nonequilibrium driving force values needed to unlock a given phenotype—and the fraction of rate space manifesting this phenotype—varies markedly across the rates. For instance, the two-inflection-point nonequilibrium response shape (orange) is only reached for a fairly narrow, fine-tuned region of drive for the rates k_{PS}, $k_{XP,X}$, k_{SP}, and $k_{X,XP}$, but is the most common shape phenotype over finite net drives for the rates k_{XS}, $k_{XP,P}$, k_{SX}, and $k_{P,XP}$. Such variable consequences of injecting energy along different rate transitions reflect the privileged roles that states XP and P play in the graph, given that their probability is the transcriptionally-potent response we monitor. The contrasting impacts of modifying each edge are also sensitive to the default rates that define the system's biological equilibrium starting point, a revealing dependence that we will return to in the penultimate Results section.

4.2.8 Breaking detailed balance two edges at a time

Adjusting one edge at a time, as we have just investigated, is one of many ways a network could invest energy to control its input-output function. For instance, the classical scheme of kinetic proofreading recognized that many steps could each be driven independently [39], as has later been repeatedly observed in the multistep ways that T-cell or MAPK activation implement kinetic proofreading [40, 41, 42, 43] or in mechanochemical operation of myosin motors [44]. How do such distributed investments of energy afford expanded control of response functions? To understand this question, we now appraise how breaking detailed balance along up to two edges at a time expands how different response behaviors may be accessed. With two independent drives (one for each edge's

departure from its default biological value), the formerly-one-dimensional phase diagrams of Fig. 4.5 become slices of two-dimensional phase diagrams that map where response shapes are reached (see Fig. 4.6A-B; and also the census of how all twenty-eight rate pairs behave found in the SI §2K).

Geometrically more complex than their one-edge equivalents in Fig. 4.5, these two-edge phase diagrams expose new ways to transition between the shape phenotypes. One measure of this new facility is the energetic cost needed to reach nonmonotonic (two inflection-point) response curves. Starting from biological equilibrium, what is the minimum net drive $\Delta\mu_0$ required for the response to become nonmonotonic, when energy can be injected along just one edge at a time (Fig. 4.5) or up to two edges at a time (Fig. 4.6A & 4.6B)? Regarding this question, we find that the $\binom{8}{2} = 28$ possible pairs of edges can be divided into two types. A few—like the edge pair (k_{XS}, k_{SX}) illustrated in Fig. 4.6B—require the same finite total dissipation to reach nonmonotonicity as needed if only driving either individual edge. However, the majority of rate pairs—such as the edge pair (k_{SX}, k_{PS})—offer a dissipative bargain: by controlling both rates it is possible to find a point in rate space where only an infinitesimal departure from detailed balance activates nonmonotonicity (as circled in 4.6A). These infinitesimal minimal drives contrast the finite drives always required while modifying single edges (Fig. 4.5C). This new economy is enjoyed by the 22 rate pairs that include at least one of the four special rates $k_{X,XP}, k_{SP}, k_{XP,X}$, or k_{PS}; their membership is a clue for identifying critical conditions on nonmotonicity as we deduce in the next Results section.

The richer behaviors achievable by breaking detailed balance along two rates (instead of just one) become even more pronounced from the lens of sensitivity. The heatmap of Fig. 4.6C depicts the maximal unnormalized sharpness $d\langle r\rangle/d\ln[X]$ reached by modifying the rate pair (k_{SX}, k_{PS}) (the same rates mapped phenotypically in the phase space of Fig. 4.6A). If only one rate constant at a time were allowed to be driven, only the slices of

sharpness along the white dotted $x = 0$ and $y = 0$ vertical and horizontal lines would be accessible, at most realizing a maximal unnormalized sharpness of $\lesssim 0.15$ with respect to the concentration $[X]$ on a log scale. However, once both edges can be modified, it becomes possible to access the maximal slope region on the lower right, yielding a greater maximum sensitivity of about 0.35. Repeating this procedure for all 28 rate pairs, as shown in Fig. 4.6D, we find that the points in rate space that maximize slope all require *both* rate constants in each pair to be modified from their default equilibrium values (lying away from the $x = 0$ and $y = 0$ vertical and horizontal lines). To maximize sensitivity, all rate pairs show one (but usually not both) rate constant that has been driven to the maximal extent allowed by the nonequilibrium driving force budget (localizing optimal points to the borders—but not necessarily corners—in Fig. 4.6D). The net drive $\Delta\mu$ ensuing from both rate's departure from their equilibrium values is often distinct from those independent departures. Fig. 4.6E recasts the same slope-maximizing points in Fig. 4.6D in terms of these two separate properties (the net drive $\Delta\mu$, and the average geometric distance, D, each edge moved from its biological starting point.) Different rate pairs show dramatically different optimal maximum sensitivities at varying cost: choosing to break detailed balance along the (k_{SX}, k_{PS}) can achieve a maximal slope of about 0.35 (probability units per e-fold change in [X]) at a net drive of only $\Delta\mu \approx 10\ k_BT$ (dark grey marker), but choosing less wisely the rate pair (k_{SX}, k_{PXP}) at best attains a slope of about 0.054 (probability units per e-fold change in [X]), even while spending a net energy $\Delta\mu \gtrsim 35\ k_BT$ almost four times as large. Collectively, these findings highlight how distributing dissipation over the transitions in a network can achieve more precise and dramatic responses.

4.2.9 Generic rate conditions forbid access to nonmonotonic responses

Why, as we have seen, are nonmonotonic responses accessed with different ease while driving some rates—or still more economically, rate pairs—rather than others? How do

the default equilibrium rates from which biology departs affect the tunability of responses? Confronting these questions leads us to glean general kinetic conditions that enable or forbid nonmonotonicity. We reformulate the criterion for nonmonotonicity to explicitly invoke net drive and rate constants (see SI §2L). Using these analytical arguments, we determine that nonmonotonicity is forbidden for any net drive when transition rates satisfy the following, surprisingly loose, conditions:

$$\begin{matrix}\langle r \rangle \\ \text{is always} \\ \text{monotonic} \\ \text{in } [X]\end{matrix} \equiv \begin{cases} k_{X,XP} \geq k_{SP} \text{ and } k_{XP,X} \leq k_{PS}, \text{ or} \\ \\ k_{X,XP} \leq k_{SP} \text{ and } k_{XP,X} \geq k_{PS}. \end{cases} \tag{4.5}$$

That is, if the presence of the transcription factor on the genome increases or decreases the polymerase's binding rate in a sense opposite to its effect on the unbinding rate (or leaves either unchanged), the response must depend on the transcription factor monotonically. Only when the transcription factor plays a functionally "ambiguous," dualistic role—coherently changing both the polymerase's binding and unbinding rates (that themselves have opposite effects on the response)—may the response become nonmonotonic under a sufficient net drive. Since access to nonmonotonicity is governed by kinetic conditions in (4.5)—but thermodynamic parameters instead set whether a response is globally activating or repressing (SI §2L)—the qualitative origin of nonmonotonicity stems from when kinetic and thermodynamic aspects in the system oppose each other.

This condition of Eq. 4.5 helps explain why some rates and rate pairs reach regulatory shape phenotypes so differently under drive, and how default starting rate constants matter. A comprehensive census of responses while driving one edge at a time when default rates *satisfy* Eq. 4.5 is provided in the SI §2K. Instructively, Eq. 4.5 demands that when the transcription factor does *not* change the polymerase's (un)binding rates—namely, either $k_{X,XP} = k_{SP}$ or $k_{XP,X} = k_{PS}$—the response must be monotonic. By

default, under the often reasonable classical assumption that the binding rate of polymerase is purely diffusion-limited [1], the transcription factor indeed may not affect the polymerase's binding rate, thus forcing the response to be monotonic.[5] This type of biophysical constraint may contribute to why transcriptional responses are most canonically pictured as monotonic. However, while plausible, this biophysical scenario is hardly inescapable or universal. In fact, even for architectures as "simple" as *lac* repression, there is gathering empirical evidence that proteins associate with DNA binding sites under more intricate regulation than merely diffusion [45]. Transcription factors that mediate steric access to the genome (dissipatively or not), such as via DNA looping [46], may also be especially prone to contravene this condition.

4.2.10 More complex regulatory architectures obey similar constraints

As documented earlier in Figure 4.2, more complex regulatory networks than the square cycle also govern transcription, albeit less commonly. In the Supporting Information (§3), we use the Matrix Tree Theorem to derive how combinatorial regulation by multiple transcription factors—forming hypercubic graphs—generalize the algebraic behavior of the square graph. Further, we also investigate two distinct architectures for DNA looping, thus generalizing quantitative input-output functions characterized in previous experiments (e.g. [47]) but previously studied with models confined by equilibrium assumptions [30, 47]. These larger networks also produce outputs that are ratios of particular (multivariate) polynomials in transcription factor concentrations as control variables. Like in the square graph, these responses attain restricted subsets of monomials in the control variables, whose constrained algebraic forms we calculate in the SI (§3). At

[5] By contrast, by the assumption that the transcription factor has the typical biophysical effect of changing the affinity between the polymerase and genome, the polymerase's off-rate from the genome *is* affected by the transcripton factor's presence, and $k_{XP,X} \neq k_{PS}$. So usually it is not an equality between polymerase's off-rates that prevents a response from being nonmonotonic.

equilibrium, these responses analogously collapse further to much simpler expressions with constrained shapes. To learn which nonequilibrium behaviors are not just mathematically possible but also biologically plausible to reach in such systems, we further systematically broke detailed balance along each edge in the architecture of DNA looping by repression. This operation illustrates how the edge-by-edge analysis in Fig. 4.5 can extend to more complex graphs. These algebraic and numerical analyses promise to identify constraints on the ultimate signal processing and combinatorial logic that cells can perform even with larger networks, in or out of equilibrium. For instance, we also derive separate kinetic criteria needed for DNA looping architectures to attain nonmonotonic response shapes. These conditions reveal how even in these more complex networks, narrow and biophysically-challenging parameter regimes are required to reach specific response features (see SI§3).

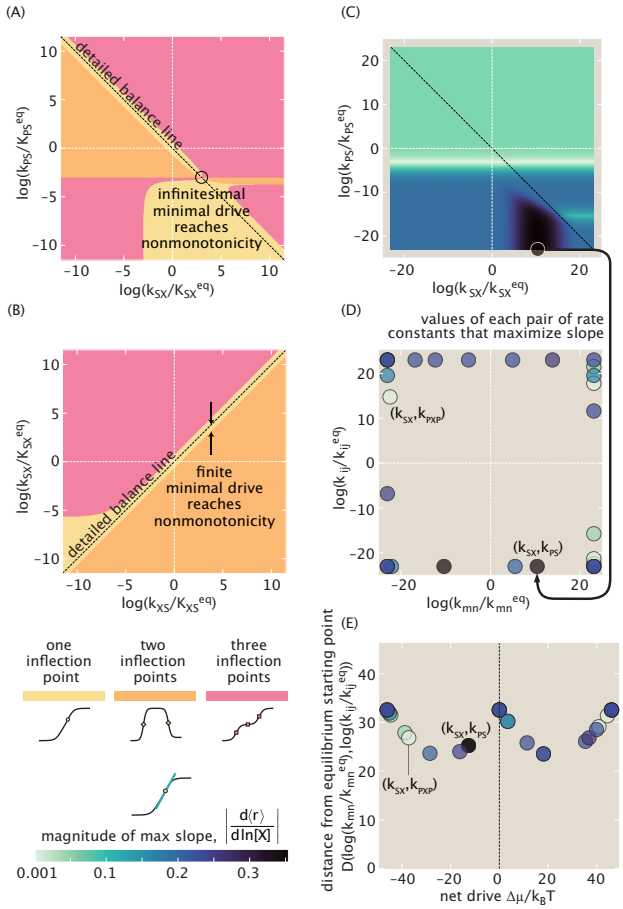


Figure 4.6: Breaking detailed balance along two edges unlocks higher sensitivity and multiply-inflected outputs with smaller drive than required for breaking detailed balance along single edges. (A) Adjusting the rate pair (k_{SX}, k_{PS})—while fixing the other six rates at their default biological values at equilibrium (of Figure 4.3A and Figure 4.5C's stem plot)—varies the number of inflection points (light yellow: one inflection, orange: two inflections, pink: three inflections), in a 2D analog of Figure 3. Specifically, this rate pair illustrates a case where nonmonotonic two-inflection curves can be reached with only an infinitesimal net drive. (B) In contrast, when tuning (k_{XS}, k_{SX}), a finite minimum drive is needed to access nonmonotonicity; numerical sampling reveals that this total drive is the same as required while only tuning one edge at a time. (C) Maxima of raw slope $d\langle r\rangle/d\ln[X]/[X]_0$ over the same modulations (axes) of the rate pair (k_{SX}, k_{PS}) shown in (A), with slope-maximizing rates within the permissible rate space indicated with a circle. $[X]_0 \equiv 1$ nM is a reference concentration. (D) Overlaying the same positions of maximal slope for all twenty-eight rate pairs emphasizes that optimal slopes are found at the boundary of the permissible rate space. Marker colors reflect the maximal slope achieved for each rate pair. Panel (E) summarizes the behavior of panel (D) by representing each optimal rate pair value with two important natural parameters: the net drive $\Delta\mu/k_BT$ (either the log ratio or log product of each rate's difference from their equilibrium starting values, depending on the relative (counter)clockwise orientation of the rates in a pair); and the net total distance the optimal values are found from their starting values in rate space, $D\left(\ln\frac{k_{mn}}{k_{mn}{}^{eq}}, \ln\frac{k_{ij}}{k_{ij}{}^{eq}}\right) \equiv \sqrt{\left(\ln\frac{k_{mn}}{k_{mn}{}^{eq}}\right)^2 + \left(\ln\frac{k_{ij}}{k_{ij}{}^{eq}}\right)^2}\, dk_{mn}\, dk_{ij}$.

4.3 Discussion

In this work, we dissected how spending energy transforms the control of gene expression in minimal and common transcriptional architectures. Harnessing a kinetic description and diagrammatic procedure from graph theory, we found that any transcriptional outputs follow a universal form with respect to a control parameter like a transcription factor's concentration. Focusing on a particularly simple and pervasive regulatory motif, we discovered these responses may only adopt three shapes, including an equilibrium-like (monotonic, singly inflected, saturating) response. Uniquely out of equilibrium, however, two unexpected and noncanonical output behaviors become possible: a doubly-inflected , nonmonotonic response, and a triply-inflected, monotonic response. Underneath wide parametric complexity, we established tight global bounds on the maximal sensitivity of transcriptional responses and learned these can vary and tradeoff with response shape. Next, we systematically mapped how biologically-feasible amounts of energy along single rates or rate pairs control responses. These findings established that the noncanonical responses are easily accessed around rates plausible for transcription, especially when dissipation can be distributed more widely over a network. Last, we uncovered global, transparent kinetic conditions that forbid (or enable) novel nonmonotonic responses. An important aspect of the present work is that beyond focusing on point quantities such as the sensitivity, we characterize the entire input-output function over the full range of different detailed-balance-breaking behaviors.

The flexible regulation unlocked by nonequilibrium could be widely biological salient. Responses that can show three inflection points—instead of just one at equilibrium—could accomplish the role of two classical (singly-inflected) input-output functions. Since an inflection can mark a local region of enhanced output sensitivity, and effectively implement a threshold, this functionality could allow cells to achieve distinct cellular fates,

such as in Wolpert's classical French Flag model [48]. By contrast to our small architecture, canonical pictures of multiple thresholded responses usually require multiple genes—often at least one specific gene per threshold [49]. One imporant example is the celebrated Dorsal protein in *Drosophila*, where two critical thresholds have been proposed to accomplish *twist* gene activation and *decapentaplegic* gene repression to help establish distinct parts of dorsal patterns in embryonic development [50, Fig. 2.26, p. 64]. We propose that triply-inflected responses from a single gene could accomplish some of this same functionality with a smaller architecture.

Nonmonotonic response functions with two inflection points could empower cells to accomplish more sophisticated signal processing, such as band-pass or band-gap filtering of chemical inputs, and/or generate temporal pulses of chemical outputs. Similar implications have been been explored by Alon & coworkers, *inter alios*, who established how nonmonotonic outputs can be produced by chaining together incoherent feed-forward loops [51, 52, 53, 54]. To achieve more complex outputs, these networks use transcriptional interactions among multiple genes at equilibrium—e.g. from two to six (or more) genes in such examples. Hence these networks operate with comparatively larger sizes and timescales than mere binding-unbinding reactions on a single gene's regulatory network like the square graph we study in this report. We suggest these comparisons contribute new material to a maturing discourse about when and how biology uses thermodynamic or kinetic control mechanisms [37, 44].

Even responses that remain "equilibrium-like" with a single inflection benefit from energy expenditure, since our bounds establish they may be up to two times more sensitive than at equilibrium, and enjoy new kinetic (instead of merely thermodynamic) ways of controlling the location of the governing inflection point (EC50). Such adaptation evokes earlier influential models of ultrasensitivity by energy expenditure [55].

While only mild net drives unlock useful regulatory shapes and traits, our analysis

emphasizes other mechanistic factors that govern how easily these behaviors can be accessed, or measured as signatures of nonequilibrium in natural or synthetic settings. First, the biological network's architecture determines whether these new macroscopic behaviors can be attained at all. Although prokaryotic gene regulation has regularly shown a compelling coherence between quantitative measurements and equilibrium statistical mechanical models (including demanding studies from our own laboratories over the past two decades [6, 56, 57, 28, 25] and beyond [46]), many of the most fiercely interrogated systems (e.g. the *lac* repressor) are indeed exactly those with acyclic network topologies that make nonequilibrium steady-states impossible (without open fluxes) and guarantee detailed balance. This reflects a possible overrepresentation of biological settings where detailed balance may be expected *a priori* to apply on mere structural grounds. On the other hand, the means to spend energy biochemically clearly exist, even in bacteria through two-component regulatory systems [58] and other active settings like nucleosome remodeling in eukaryotes [5]. Accordingly, our findings invite and vigorous reappraisal of whether signatures of nonequilibrium are lurking in architectures that are more prone to accommodate it, such as the four-state "simple activation" motif we discussed here. Moreover, the measurements (or synthetic biological perturbations) needed to map the nonequilibrium landscape of transcriptional responses must differ from the convenient binding site modifications (e.g. parallel promoter libraries [59, 25]) previously used to test equilibrium models, since manipulating binding energies inherently preserves detailed balance. Developing fresh experimental approaches to augment or attenuate a single transition between microstates (or set of transitions) *in situ* to break detailed balance is a crucial direction of future empirical work, whose value is advocated for by our results. To manipulate and probe tractable models of transcription, these methods might include optogenetic control [60, 61], or adjustments of governing enzyme concentrations or activities.

Second, where energy is invested crucially dictates which regulatory behaviors are available. We found that investing energy along more than one rate at once was capable of achieving more dramatic response curves more economically. This finding may help explain the many observations in biological systems where energy is independently injected along multiple steps [39, 40, 41, 42, 43, 44]. However, since each independently-regulated injection of energy may also be accompanied by architectural costs, not all examples of biological regulation may contain the distributed dissipation machinery required to make novel nonequilibrium response signatures conspicuous.

Third, the structures of responses while breaking detailed balance edge-by-edge, and our general kinetic criteria that forbid nonmonotonicity, highlight that certain critical imbalances between rate constants are needed to produce some out of equilibrium signatures. On basic biophysical grounds, some natural systems may (or may not) exhibit the required rate imbalances to make novel responses as easy to activate (see SI §2L.2). Indeed, the rate imbalances required to produce nonmonotonicity we found are non-obvious. These kinetic criteria have significant implications for organizing parameter explorations. For instance, we show in the SI (§2M) that two exciting, completely separate studies recently published [15, 2] exploring the informational and functional consequences of nonequilibrium impose simplifying assumptions on rate constants that in fact preclude the possibility of non-monotonic responses, according to equivalents of our monotonicity criterion. We expect that this approach and our kinetic criteria will help future works include and capture the regulatory consequences of these rich behaviors. We anticipate this flexibility may be especially germane for environments that present nonuniform input statistics.

The contrast between the nonequilibrium steady-states possible to support using this "simple activation" architecture, and the difficulty of sustaining nonequilibrium steady-states in a simple repression architecture that lacks a cycle, also possibly provides a

new design principle to understand the timeless question of why both activators and repressors are employed as distinct architectures when they can produce the same mean gene expression. Intriguing rationalizations based on ecological demand have been offered for why these architectures are used differently in *E. coli*, such as the classical proposal by Savageau [62, 63, 64]. We speculate that another, quite distinct, feature—the very possibility of using nonequilibrium to steer input-output response curves so flexibly—may also contribute to why organisms might use a simple-activation (or other cycle-containing) architecture over acyclic architectures, all other features being equal. Whether this nonequilibrium controllability significantly shapes the natural incidence of regulatory architectures can only be assessed using quantitative measurements of input-output behaviors from a much broader set of architectures than the relatively narrow (e.g. Lac repressor, Bicoid, Pho5, CI in bacteriophage-λ switch) subjects of existing analyses.

Our work urges the value of measuring the global *shapes* of input-output functions. This approach complements but contrasts the more common practice of attempting to fit and compare single scalar parameters like sensitivity or pointwise proofreading-ratios [13, 55]. Unexpectedly flexible global curve shapes might be even more obvious and robust signatures of nonequilibrium than pointwise estimators, since scalar measures like derivatives are often highly sensitive to measurement noise [65]. Yet, just as for pointwise measures, inferring nonequilibrium from global shape requires sufficient knowledge of a system's states. As our wider analysis of transcriptional architectures (SI §3) stresses, larger graphs can reach complex shapes at equilibrium only accessible by small graphs out of equilibrium. Whether a measured complex response stems from a simple system operating out of equilibrium, or a more architecturally-complex system operating near equilibrium, may depend on separate lines of empirical evidence, including knowledge of binding site topology. Nevertheless, such details are often known—or knowable—for promoters of interest.

Our calculations also provide explicit maps of parameter spaces that can guide the naturalist looking for whether this expanded regulation occurs naturally in specific examples of transcription. This information is also a guide to the synthetic biologist who endeavors to engineer such responses in genetic circuits and exploit the advantages of producing complex regulation using a small driven network, instead of a comparatively larger, more slowly tuned network of multiple genes and/or proteins at equilibrium.

Beyond advocating for new experiments, our findings invite many theoretical extensions. How dissipation affects the intricate tradeoffs between sensitivity, specificity, speed, and stochasticity in (steady-state or transient) gene regulation is a large, open, physiologically-relevant question amenable to further graph-theoretic dissection. In addition, we hope for deeper analytical rationalization of our bounds on sensitivity; our upper bounds surely share similar foundations with looser, more architecturally general, bounds recently and insightfully established by Owen & Horowitz [33], though our additional lower bounds and different mathematical quantities suggest separate theoretical ingredients.

Overall, we foresee that graph-theoretic treatments like we have deployed here—and as have been first so powerfully established and refined by foundational investigators [18, 20, 66, 33]—will help understand still more sophisticated networks. Just as Feynman diagrams and other diagrammatic reasoning catalyze field theory and particle physics [67, 68], drawing graphs promises to help build structural principles about how energy enlivens biology.

4.3.1 Nonequilibrium steady-state probabilities via the Matrix Tree Theorem

Consider a continuous-time Markov chain with N states, whose transition rates k_{ij} between states i and j are stored in the j, ith element of the transition matrix $\mathbf{L}$, and so the

probabilities $\mathbf{p}(t) = [p_1, \ldots, p_N]^\top$ of finding the system in these states evolve according to

$$\frac{d\mathbf{p}}{dt} = \mathbf{L}\mathbf{p}.$$

(With this convention of $\mathbf{p}$ as a column vector, the columns of the matrix $\mathbf{L}$ sum to zero and the diagonal entries are accordingly $L_{ii} = -\sum_{j \neq i} L_{ji} = -\sum_{j \neq i} k_{ij}$.) Identifying our Markov system as a weighted graph, a *spanning tree* over the states is a set of $N - 1$ edges that visits every state exactly once. A spanning tree $\spadesuit_i$ *rooted* in a state i contains no outgoing edges from state i (and exactly one outgoing edge for every other state $j \neq i$). These notions are summarized in the example of Fig. 4.3A. The **Matrix Tree Theorem** (MTT), also known as the Markov Chain Tree Theorem, states that at steady state $\left(\frac{d\mathbf{p}}{dt} = \mathbf{L}\mathbf{p} = \mathbf{0}\right)$, the statistical weight of the ith microscopic state is the sum of products of rate constants over spanning trees rooted in node i

$$\rho_i = \sum_{\text{span. } \spadesuit_i}^{N_{T_i}} \left(\prod_{k_{rs} \in \spadesuit_i}^{N-1} k_{rs} \right), \tag{4.6}$$

where N_{Ti} is the number of spanning trees rooted in i [18, 24]. This weight ρ_i is the relative odds of finding the system in state i as a fraction of all the statistical weights $\rho_{tot} = \sum_j \rho_j$, namely $p^i = \rho_i / \rho_{tot}$. Applying the MTT to the regulatory motif of Fig. 4.3A indicates that any steady-state probabilistic observable depends on the transcription factor control parameter $[X]$ according to Eq. 4.1 (see SI §2D).

4.3.2 Emergent shape parameters & shape phenotypes

The collapsed shape representation of Eq. 4.3 allows us to solve for the number of positive solutions to $d\langle r \rangle / d \ln([X]/[X]_0)$, yields the numbers of possible inflection points (via, for instance, Descartes' rule of signs or explicit inequality solving) and hence shapes (see SI §2I). Numerical and symbolic analysis of the space formed by these two emergent shape

parameters (a, b) (Eq. 4.3 and SI §2I.2) helps establish our global bounds on sensitivity. Ultimately, this collapsed representation is also a crucial theoretical stepladder to find the generic conditions forbidding nonmonotonicity given in Eqs. 4.5 (see SI §2I).

4.3.3 Single edge and edge pair perturbations

We estimated default biological rates for transcription at equilibrium by synthesizing reported binding affinities, association rates, and diffusion constants (see SI §2B). We solved the condition for an inflection point symbolically and numerically (see SI §2I-K).

Acknowledgements. We gratefully acknowledge feedback from and formative discussions with Forte Shinko; Nicholas Lammers; Avi Flamholz; Jordan Horowitz; Jeremy Owen; and Jané Kondev. G.L.S. thanks the US NSF GRFP under Grant DGE-1745301 and Caltech's Center for Environmental Microbial Interactions for support. R.P. is supported by the Rosen Center at Caltech and the NIH 1R35 GM118043 (MIRA). H.G.G. is supported by NIH R01 Award (R01GM139913), NIH R21 Award (R21HD107436), and the Koret-UC Berkeley-Tel Aviv University Initiative in Computational Biology and Bioinformatics. H.G.G. is also a Chan Zuckerberg Biohub–San Francisco Investigator.

4.4 Data & Availability

Symbolic and numerical code used for this study's analyses and figures is available open-source. See https://github.com/RPGroup-PBoC/graphnoneq.

Bibliography

[1] Muir Morrison, Manuel Razo-Mejia, and Rob Phillips. Reconciling kinetic and thermodynamic models of bacterial transcription. *PLoS Comput. Biol.*, 17(1):e1008572, 2021.

[2] Rok Grah, Benjamin Zoller, and Gašper Tkačik. Nonequilibrium models of optimal enhancer function. *Proc. Natl. Acad. Sci. U.S.A.*, 117(50):31614–31622, 2020.

[3] Felix Wong and Jeremy Gunawardena. Gene regulation in and out of equilibrium. *Annu. Rev. Biophys.*, 49:199–226, 2020.

[4] Antoine Coulon, Carson C Chow, Robert H Singer, and Daniel R Larson. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat. Rev. Genet.*, 14(8): 572–584, 2013.

[5] Robert Shelansky and Hinrich Boeger. Nucleosomal proofreading of activator–promoter interactions. *Proc. Natl. Acad. Sci. U.S.A.*, 117(5):2456–2461, 2020.

[6] Manuel Razo-Mejia, Stephanie L Barnes, Nathan M Belliveau, Griffin Chure, Tal Einav, Mitchell Lewis, and Rob Phillips. Tuning transcriptional regulation through signaling: a predictive theory of allosteric induction. *Cell Systems*, 6(4):456–469, 2018.

[7] François Jacob and Jacques Monod. On the regulation of gene activity. In *Cold Spring Harb. Symp. Quant. Biol.*, volume 26, pages 193–211. Cold Spring Harbor Laboratory Press, 1961.

[8] Ian S Farmer and Colin W Jones. The energetics of escherichia coli during aerobic growth in continuous culture. *European journal of biochemistry*, 67(1):115–122, 1976.

[9] Matteo Mori, Chuankai Cheng, Brian R Taylor, Hiroyuki Okano, and Terence Hwa. Functional decomposition of metabolism allows a system-level quantification of fluxes and protein allocation towards specific metabolic functions. *Nature Communications*, 14(1):4161, 2023.

[10] Jonathan Rodenfels, Karla M Neugebauer, and Jonathon Howard. Heat oscillations driven by the embryonic cell cycle reveal the energetic costs of signaling. *Developmental cell*, 48(5):646–658, 2019.

[11] Qiwei Yu, Dongliang Zhang, and Yuhai Tu. Inverse power law scaling of energy dissipation rate in nonequilibrium reaction networks. *Physical review letters*, 126(8):080601, 2021.

[12] Xingbo Yang, Matthias Heinemann, Jonathon Howard, Greg Huber, Srividya Iyer-Biswas, Guillaume Le Treut, Michael Lynch, Kristi L Montooth, Daniel J Needleman, Simone Pigolotti, et al. Physical bioenergetics: Energy fluxes, budgets, and constraints in cells. *Proceedings of the National Academy of Sciences*, 118(26):e2026786118, 2021.

[13] Benjamin Zoller, Thomas Gregor, and Gašper Tkačik. Eukaryotic gene regulation at equilibrium, or non? *Curr. Opin. Syst. Biol.*, page 100435, 2022.

[14] Hao Ge, Min Qian, and Hong Qian. Stochastic theory of nonequilibrium steady states. part II: Applications in chemical biophysics. *Phys. Rep.*, 510(3):87–118, 2012.

[15] Nicholas C Lammers, Avi I Flamholz, and Hernan G Garcia. Competing constraints shape the nonequilibrium limits of cellular decision-making. *Proc. Natl. Acad. Sci. U.S.A.*, 120(10): e2211203120, 2023.

[16] Kee-Myoung Nam, Rosa Martinez-Corral, and Jeremy Gunawardena. The linear framework: using graph theory to reveal the algebra and thermodynamics of biomolecular systems. *Interface Focus*, 12(4):20220013, 2022.

[17] Jeremy Gunawardena. A linear framework for time-scale separation in nonlinear biochemical systems. *PLoS One*, 7(5):e36321, 2012.

[18] Inomzhon Mirzaev and Jeremy Gunawardena. Laplacian dynamics on general graphs. *Bull. Math. Biol.*, 75(11):2118–2149, 2013.

[19] Víctor H Tierrafría, Claire Rioualen, Heladia Salgado, Paloma Lara, Socorro Gama-Castro, Patrick Lally, Laura Gómez-Romero, Pablo Peña-Loredo, Andrés G López-Almazo, Gabriel Alarcón-Carranza, et al. Regulondb 11.0: Comprehensive high-throughput datasets on transcriptional regulation in *Escherichia coli* K-12. *Microb. Genom.*, 8(5), 2022.

[20] Terrell L Hill. *Free energy transduction and biochemical cycle kinetics*. Courier, 2013.

[21] Hong Qian. Open-system nonequilibrium steady state: statistical thermodynamics, fluctuations, and chemical oscillations. *J. Phys. Chem. B*, 110(31):15063–15074, 2006.

[22] Robert Shelansky, Sara Abrahamsson, Michael Doody, Christopher R Brown, Heta P Patel, Tineke L Lenstra, Daniel R Larson, and Hinrich Boeger. A telltale sign of irreversibility in transcriptional regulation. *bioRxiv*, pages 2022–06, 2022.

[23] Leonid A Mirny. Nucleosome-mediated cooperativity between transcription factors. *Proc. Natl. Acad. Sci. U.S.A.*, 107(52):22534–22539, 2010.

[24] Rob Phillips and Hernan G. Garcia. *Physical Genomics: From E. coli to Elephants*. Princeton University Press, 2024.

[25] William T Ireland, Suzannah M Beeler, Emanuel Flores-Bautista, Nicholas S McCarty, Tom Röschinger, Nathan M Belliveau, Michael J Sweredoski, Annie Moradian, Justin B Kinney, and Rob Phillips. Deciphering the regulatory genome of *Escherichia coli*, one hundred promoters at a time. *eLife*, 9:e55308, 2020.

[26] Mattias Rydenfelt, Hernan G Garcia, Robert Sidney Cox III, and Rob Phillips. The influence of promoter architectures and regulatory motifs on gene expression in *Escherichia coli*. *PLoS One*, 9(12):e114347, 2014.

[27] Soile VE Keränen, Angel Villahoz-Baleta, Andrew E Bruno, and Marc S Halfon. REDfly: an integrated knowledgebase for insect regulatory genomics. *Insects*, 13(7):618, 2022.

[28] Hernan G Garcia and Rob Phillips. Quantitative dissection of the simple repression input–output function. *Proceedings of the National Academy of Sciences*, 108(29):12173–12178, 2011.

[29] Hernan G Garcia, Jane Kondev, Nigel Orme, Julie A Theriot, and Rob Phillips. Thermodynamics of biological processes. In *Methods in enzymology*, volume 492, pages 27–59. Elsevier, 2011.

[30] Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Terence Hwa, Jané Kondev, and Rob Phillips. Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.*, 15(2):116–124, 2005.

[31] Lee R Swem, Danielle L Swem, Ned S Wingreen, and Bonnie L Bassler. Deducing receptor signaling parameters from *in vivo* analysis: LuxN/AI-1 quorum sensing in *Vibrio harveyi*. *Cell*, 134(3):461–473, 2008.

[32] Adam J Meyer, Thomas H Segall-Shapiro, Emerson Glassey, Jing Zhang, and Christopher A Voigt. *Escherichia coli* "marionette" strains with 12 highly optimized small-molecule sensors. *Nat. Chem. Biol.*, 15(2):196–204, 2019.

[33] Jeremy A Owen and Jordan M Horowitz. Size limits the sensitivity of kinetic schemes. *Nat. Commun.*, 14(1):1280, 2023.

[34] David Hartich, Andre C Barato, and Udo Seifert. Nonequilibrium sensing and its analogy to kinetic proofreading. *New Journal of Physics*, 17(5):055026, 2015.

[35] Ron Milo, Paul Jorgensen, Uri Moran, Griffin Weber, and Michael Springer. Bionumbers—the database of key numbers in molecular and cell biology. *Nucleic acids research*, 38(suppl_1): D750–D753, 2010.

[36] Quang Hon Tran and Gottfried Unden. Changes in the proton potential and the cellular energetics of escherichia coli during growth by aerobic and anaerobic respiration or by fermentation. *European journal of biochemistry*, 251(1-2):538–543, 1998.

[37] Jordan M Horowitz, Kevin Zhou, and Jeremy L England. Minimum energetic cost to maintain a target nonequilibrium state. *Phys. Rev. E.*, 95(4):042102, 2017.

[38] Milo M Lin. General efficiency relation for dissipative molecular machines. *Proceedings of the Royal Society A*, 480(2282):20230388, 2024.

[39] John J Hopfield. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. U.S.A.*, 71(10):4135–4139, 1974.

[40] Derek M Britain, Jason P Town, and Orion David Weiner. Progressive enhancement of kinetic proofreading in T cell antigen discrimination from receptor activation to dag generation. *eLife*, 11:e75263, 2022.

[41] Peter S Swain and Eric D Siggia. The role of proofreading in signal transduction specificity. *Biophys. J.*, 82(6):2928–2933, 2002.

[42] Timothy W McKeithan. Kinetic proofreading in T-cell receptor signal transduction. *Proc. Natl. Acad. Sci. U.S.A.*, 92(11):5042–5046, 1995.

[43] Wenping Cui and Pankaj Mehta. Identifying feasible operating regimes for early T-cell recognition: The speed, energy, accuracy trade-off in kinetic proofreading and adaptive sorting. *PLoS One*, 13(8):e0202331, 2018.

[44] Joel D Mallory, Anatoly B Kolomeisky, and Oleg A Igoshin. Kinetic control of stationary flux ratios for a wide range of biochemical processes. *Proc. Natl. Acad. Sci. U.S.A.*, 117(16):8884–8889, 2020.

[45] Emil Marklund, Guanzhong Mao, Jinwen Yuan, Spartak Zikrin, Eldar Abdurakhmanov, Sebastian Deindl, and Johan Elf. Sequence specificity in DNA binding is mainly governed by association. *Science*, 375(6579):442–445, 2022.

[46] Thomas Kuhlman, Zhongge Zhang, Milton H Saier Jr, and Terence Hwa. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, 104(14):6043–6048, 2007.

[47] James Q Boedicker, Hernan G Garcia, and Rob Phillips. Theoretical and experimental dissection of DNA loop-mediated repression. *Phys. Rev. Lett.*, 110(1):018101, 2013.

[48] Lewis Wolpert. Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.*, 25(1):1–47, 1969.

[49] Spyros Papageorgiou and Yannis Almirantis. Gradient model describes the spatial-temporal expression pattern of HOXA genes in the developing vertebrate limb. *Dev. Dyn.*, 207(4):461–469, 1996.

[50] Lewis Wolpert, Cheryll Tickle, and Alfonso Martinez Arias. *Principles of development*. Oxford University Press, USA, 6 edition, 2015.

[51] Shai Kaplan, Anat Bren, Erez Dekel, and Uri Alon. The incoherent feed-forward loop can generate non-monotonic input functions for genes. *Mol. Syst. Biol.*, 4(1):203, 2008.

[52] Shuji Ishihara, Koichi Fujimoto, and Tatsuo Shibata. Cross talking of network motifs in gene regulation that generates temporal pulses and spatial stripes. *Genes Cells*, 10(11):1025–1038, 2005.

[53] Subhayu Basu, Yoram Gerchman, Cynthia H Collins, Frances H Arnold, and Ron Weiss. A synthetic multicellular system for programmed pattern formation. *Nature*, 434(7037):1130–1134, 2005.

[54] Robert Entus, Brian Aufderheide, and Herbert M Sauro. Design and implementation of three incoherent feed-forward motif based biological concentration sensors. *Syst. Synth. Biol.*, 1:119–128, 2007.

[55] Yuhai Tu. The nonequilibrium mechanism for ultrasensitivity in a biological switch: Sensing by Maxwell's demons. *Proc. Natl. Acad. Sci. U.S.A.*, 105(33):11737–11741, 2008.

[56] Daniel L Jones, Robert C Brewster, and Rob Phillips. Promoter architecture dictates cell-to-cell variability in gene expression. *Science*, 346(6216):1533–1536, 2014.

[57] Robert C Brewster, Franz M Weinert, Hernan G Garcia, Dan Song, Mattias Rydenfelt, and Rob Phillips. The transcription factor titration effect dictates level of gene expression. *Cell*, 156(6): 1312–1323, 2014.

[58] Alexander Y Mitrophanov and Eduardo A Groisman. Signal integration in bacterial two-component regulatory systems. *Genes. Dev.*, 22(19):2601–2611, 2008.

[59] Timothy C Yu, Winnie L Liu, Marcia S Brinck, Jessica E Davis, Jeremy Shek, Grace Bower, Tal Einav, Kimberly D Insigne, Rob Phillips, Sriram Kosuri, et al. Multiplexed characterization of rationally designed promoter architectures deconstructs combinatorial logic for IPTG-inducible systems. *Nat. Commun.*, 12(1):325, 2021.

[60] Wei Zhang, Alexander W Lohman, Yevgeniya Zhuravlova, Xiaocen Lu, Matthew D Wiens, Hiofan Hoi, Sine Yaganoglu, Manuel A Mohr, Elena N Kitova, John S Klassen, et al. Optogenetic control with a photocleavable protein, PhoCl. *Nat. Methods*, 14(4):391–394, 2017.

[61] Hongtao Liu, Gustavo Gomez, Sophia Lin, Shuo Lin, and Chentao Lin. Optogenetic control of transcription in zebrafish. *PLoS One*, 7(11):e50738, 2012.

[62] Michael A Savageau. Design of molecular control mechanisms and the demand for gene expression. *Proc. Natl. Acad. Sci. U.S.A.*, 74(12):5647–5651, 1977.

[63] Michael A Savageau. Genetic regulatory mechanisms and the ecological niche of *Escherichia coli*. *Proc. Natl. Acad. Sci. U.S.A.*, 71(6):2453–2455, 1974.

[64] Uri Alon. *An introduction to systems biology: design principles of biological circuits*. CRC press, 2nd edition, 2019. See especially §7.6 "Demand rules for gene regulation can minimize errors," p. 129.

[65] Håkan Lanshammar. On precision limits for derivatives numerically calculated from noisy data. *Journal of biomechanics*, 15(6):459–470, 1982.

[66] Jürgen Schnakenberg. Network theory of microscopic and macroscopic behavior of master equation systems. *Rev. Mod. Phys.*, 48(4):571, 1976.

[67] David Kaiser. *Drawing Theories Apart: The dispersion of Feynman diagrams in postwar physics*. University of Chicago Press, 2005.

[68] Martinus Veltman. *Diagrammatica: the path to Feynman diagrams*. Number 4. Cambridge University Press, 1994.

Chapter 5

Understanding larger graphs

This chapter presents especially illuminating findings and approaches from a broader set that GS and others derive in the Supplement of Reference [1].

In the preceding chapter, we illustrated how to use diagrammatic reasoning to learn about the nonequilibrium response of a small, extremely pervasive network motif. In this expository chapter, we use the Matrix Tree Theorem to quantify the behaviors of a panoply of important regulatory scenarios beyond the simple square graph, in or out of equilibrium. The same approaches deployed on the square graph yield insights into the output behaviors of much larger graphs and more complex networks. We organize this discussion as follows. First, as a motivating preview, we curtly summarize algebraic results developed for networks later in the chapter. Second, we recall the need for understanding such networks by parsing and visualizing the distribution of currently known regulatory interactions in *E. coli* and the fly, extracting revealing statistical summaries of how promiscuously transcription factors and promotors appear to interact. Last, we give a conceptual exposition on how the tree theorem is a tool for thinking even about such complex scenarios and derive the regulatory response functions for arbitrary hypercubic

graphs by understanding their structural constraints using an arboreal lens.

5.1 Summary of nonequilibrium generalization of important transcriptional motifs

We summarize key new network calculations in the Table 5.1 that follows.

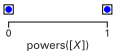
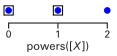
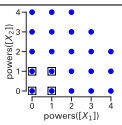
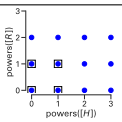
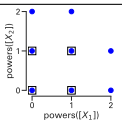
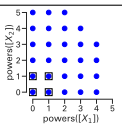
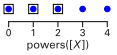
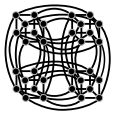
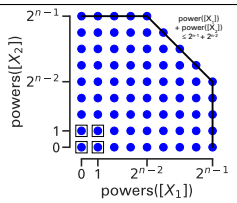
regulatory architecture	control parameters	network topology	equilibrium behavior, $\langle r \rangle^{\text{eq}}$	general nonequilibrium behavior, $\langle r \rangle$	attained powers in $\langle r \rangle$
simple (strict) repression [*cf.* Bintu [2] case 1]	one: $[X]$		$\frac{A+B[X]}{C+D[X]}$	$\frac{A+B[X]}{C+D[X]}$	
simple "activation" (this study's emphasis) [*cf.* Bintu [2] case 2]	one: $[X]$		$\frac{A+B[X]}{C+D[X]}$	$\frac{A+B[X]+C[X]^2}{D+E[X]+F[X]^2}$	
activator recruited by a helper (symmetric) [*cf.* Bintu [2] cases 3 and 7]	two: $[X_1]$, $[X_2]$		$\frac{A+B[X_1]+C[X_2]+D[X_1][X_2]}{E+F[X_1]+G[X_2]+H[X_1][X_2]}$	$\frac{\sum_{i\le 4, j\le 4, i+j\le 6} A_{ij}[X_1]^i[X_2]^j}{\sum_{i\le 4, j\le 4, i+j\le 6} B_{ij}[X_1]^i[X_2]^j}$	
strict repressor recruited by a helper [*cf.* Bintu [2] case 4]	two: $[R]$, $[H]$		$\frac{A+B[H]+C[R]+D[H][R]}{E+F[H]+G[R]+J[H][R]}$	$\frac{\sum_{i\le 3, j\le 2} A_{ij}[H]^i[R]^j}{\sum_{i\le 3, j\le 2} B_{ij}[H]^i[R]^j}$	
dual strict repressors [*cf.* Bintu [2] cases 5 and 6]	two: $[X_1]$, $[X_2]$		$\frac{A+B[X_1]+C[X_2]+D[X_1][X_2]}{E+F[X_1]+G[X_2]+H[X_1][X_2]}$	$\frac{\sum_{i\le 2, j\le 2, i+j\le 3} A_{ij}[X_1]^i[X_2]^j}{\sum_{i\le 2, j\le 2, i+j\le 3} B_{ij}[X_1]^i[X_2]^j}$	
dual activators cooperating via looping[*cf.* Bintu [2] case 8]	two: $[X_1]$, $[X_2]$		$\frac{A+B[X_1]+C[X_2]+D[X_1][X_2]}{E+F[X_1]+G[X_2]+H[X_1][X_2]}$	$\frac{\sum_{i\le 4, j\le 5, i+j\le 7} A_{ij}[X_1]^i[X_2]^j}{\sum_{i\le 4, j\le 5, i+j\le 7} B_{ij}[X_1]^i[X_2]^j}$	
repressor with two DNA binding units and DNA looping [*cf.* Bintu [2] case 9]	one: $[X]$		$\frac{A+B[X]+C[X]^2}{D+E[X]+F[X]^2}$	$\frac{A+B[X]+C[X]^2+D[X]^3+E[X]^4}{F+G[X]+H[X]^2+I[X]^3+J[X]^4}$	
m independent transcription factors [*cf.* Spirit of Bintu [2] case 10]	m, $\{[X_i]\}$	 (full hypercube)	$\frac{\sum_{z_1,z_2,z_3,\ldots z_N \in \{0,1\}} \alpha_{(z_i)} \prod_{i=1}^{m} [X_i]^{z_i}}{\sum_{z_1,z_2,z_3,\ldots z_N \in \{0,1\}} \beta_{(z_i)} \prod_{i=1}^{m} [X_i]^{z_i}}$	$\frac{\sum_{\substack{\max(i_j)=i_{(n)}\le 2^{n-1} \\ i_{(n-1)}\le 2^{n-2} \\ \vdots \\ i_{(1)}=\min(i_j)\le 2^0}} A_{\{i_j\}} \prod_{i=1}^{m} [X_i]^{i_j}}{\sum_{\substack{\max(i_j)=i_{(n)}\le 2^{n-1} \\ i_{(n-1)}\le 2^{n-2} \\ \vdots \\ i_{(1)}=\min(i_j)\le 2^0}} B_{\{i_j\}} \prod_{i=1}^{m} [X_i]^{i_j}}$	

Table 5.1: Summary of response functions of a variety of important transcriptional regulatory scenarios, generalized here to apply in or out of equilibrium. These examples include generalizing those formerly analyzed by equilibrium statistical mechanics in Reference [2]. The rightmost column visualizes the attained powers of the response function (namely, the domains of the sums in the middle two columns); square boxes denote the constrained equilibrium response. Note that the denominator and the numerator of the response $\langle r \rangle$ have the same algebraic form; they differ only on the coefficients of each monomial term. By construction, the coefficients of the polynomial in the denominator (the sum of all rooted and directed spanning trees, behaving like a partition function) are all positive. However, some of the coefficients of the numerator can be zero for a particular observable.

Specifically, later in this chapter we will focus in detail on our derivations of the regulatory response functions for the hypercubic graphs and their structural derivatives. (We leave details about our additional derivations of DNA looping interactions to be found in the Supplementary Information of Reference [1].)

Next, we review the state of knowledge about regulation in *E. coli*, and *D. melanogaster*, by extracting fresh up-to-date empirical histograms of regulatory architectures and their incidence. The data of interest are how many distinct transcription factors (whether distinct in biological identity or not) can simultaneously occupy each promoter or gene; this gives a putative upper bound on the dimensionality of the state space of the underlying regulatory graph.

5.2 Census of known regulatory interactions in *E. coli*

To understand what state space topologies are most common in *E. coli*, we reanalyzed the bipartite[1] network of all known regulatory interactions between transcription factors and promoters, as reported in the most recent release 11.2 of the canonical database *RegulonDB* [3]. These data report 216 unique transcription factors and 1152 unique regulated promoters. During data processing, we parsed all interactions (conservatively retaining all interactions, reported as strong or weak evidence; this choice does not significantly the incidences observed).

[1] A network is called *bipartite* if its nodes can be divided into two classes, and all edges in the network connect nodes of one class to the other class. This is naturally the case for how promoters and transcription factors interact.

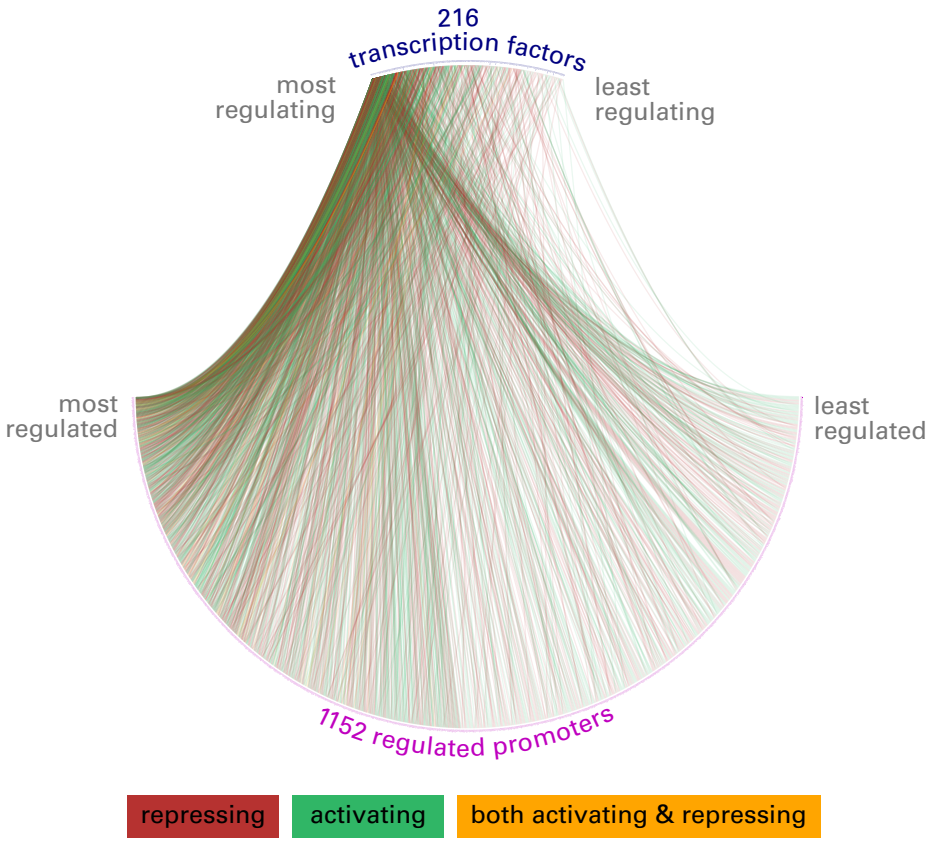


Figure 5.1: The bipartite network of all known regulatory interactions between transcription factors (top, navy nodes) and genomic promoters (bottom, magenta nodes) in the bacterium *E. coli*, constructed from raw data scraped from the RegulonDB database release 11.2 as of October 2023 [3]. Each arc visualizes an interaction between a transcription factor node on top and a promoter node on the bottom; repressing interactions are red, activating interactions green; and interactions that can be activating or repressing are orange. Nodes are arranged in decreasing degree from left to right on their respective arcs, such that the most highly regulated promoters (starting with csgDp, flhDp, gadAp, hdeAp, ...) appear on the left on the bottom arc, and the most promiscuous/highly regulating transcription factors (starting with CRP, FNR, Fis, H-NS, IHF, ArcA, ...) appear on the left on the top arc.

While on aggregate, it is true that the most common regulated architecture across these interactions is putatively the square graph (the subject of the preceding chapter), gazing at the whole network as a gestalt makes clear that plenty of interesting biological promoters are associated with more than one transcription factor.

To gain some emotional and quantitative understandings of this fact, we visualize all the network interactions of *E. coli* in a three separate ways. First, we visualize the network as a bipartite graph, where one set of nodes are transcription factors and the other set of nodes are promoters, and we draw a subtle arc edge between a transcription factor and a promoter when the database reports an interaction. To make this overwhelming network slightly more informative, we also graphically sort nodes in decreasing order of their number of regulated or regulating interactions. The result, shown in Fig. 5.2, gives the indelible impression that at least some promiscuous interactions are central to the regulatory logic of this organism.

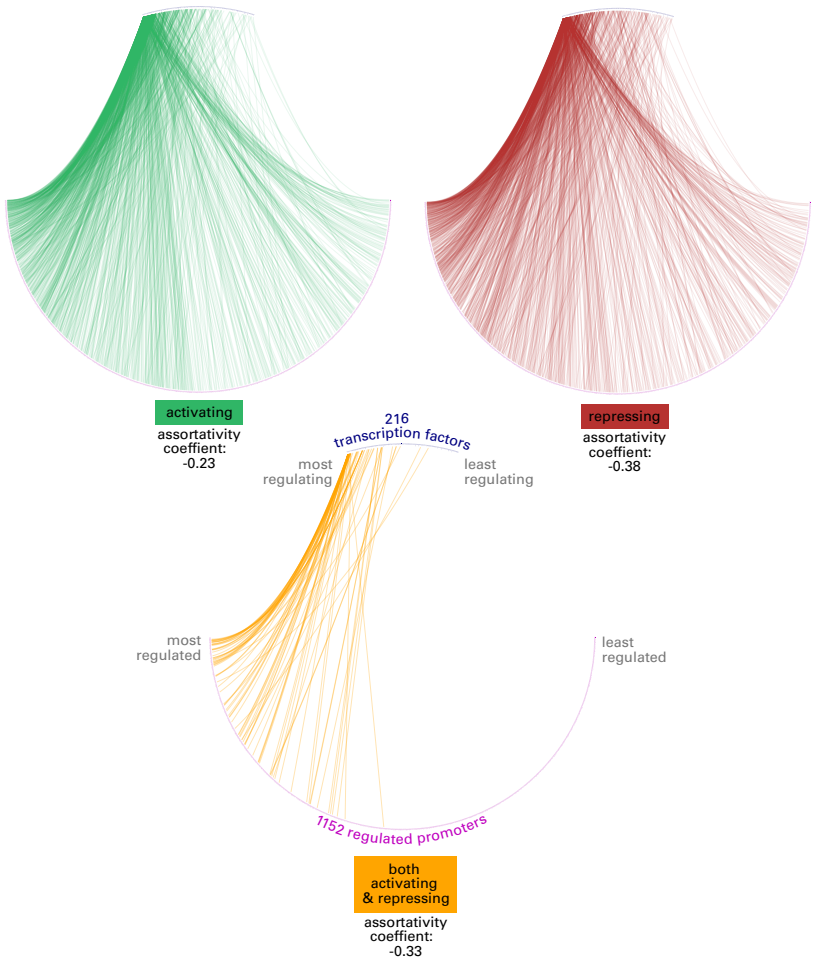


Figure 5.2: *E. coli*'s regulatory network as in Fig. 5.2, explicitly partitioned by the type of regulatory interaction (whether a transcription factor represses a promoter, activates it, or both).

Notice that while highly-regulating transcription factors are densely connected to highly-regulated promoters—as expected by chance alone given that highly-connected nodes are highly represented among edges—there is also a significant density of connections between highly regulating transcription factors and sparsely-regulated promoters, and vice versa (streaks from upper left to lower right, and from lower left to upper right). In fact, this tendency to connect highly-connected nodes to sparsely-connected nodes is quantitatively significant beyond what is expected by chance. This is reflected by appreciably negative "assortativity coefficients" that measure the average correlation between a node's degree and the degrees of nodes it is connected to. Assortativity coefficient values (here, consistently of order -0.3) are normalized to vary from 1 (completely assortative; highly-connected nodes completely prefer connecting to other highly-connected nodes) to -1 (completely disassortative; highly-connected nodes completely prefer connecting to sparsely-connected nodes) [4]. Interestingly, activating interactions (assortativity coefficient -0.23) are detectably less dissortative than repressing interactions (assortativity coefficient -0.38). That is, while both activators and repressors tend to regulate promoters whose connectivity is different than their own, repressing transcription factors that regulate many promoters are even more likely to regulate promoters that are not already very degenerately repressed, compared to activating transcription factors. Also, note the curious fact that all of the (small number of) transcription factors that can act either as repressors or activators (shown in the orange network at bottom) appear to regulate only the top half of highly-regulated promoters.

Such hairball visualizations gives one gestalt of this network, but taken alone give an incomplete impression of this regulation. This is true because the promiscuity of some central interactions (obviously on display as dense node connections) nontheless coexists with the fact that most promoters are regulated only by one transcription factor. To better appreciate how these facts coexist, second, we visualize the sparsity of this interaction

matrix using an explicit heatmap, as shown by Fig. 5.3. We order the rows (promoters) and columns (transcription factors) of this biadjacency matrix's heatmap in descending order of their promiscuity, and also annotate their empirical cumulative distribution functions as marginal histograms. These visualizations (as in the vast swaths of white (no-interaction) space in the heatmap) reflect the sparsity of this regulatory network, ultimately reflecting the sharply decaying probability of interactions reflected by these distribution functions. See https://www.its.caltech.edu/~gsalmon/gs069_Ecoli_biadjacency_RegulonDB11pt2.html for an interactive version of a heatmap where individual interactions can be gleaned by hovering with a mouse.

To understand the rapid decays in these probabilities of interactions with their promiscuity, we also consider them on a log-log axis, as shown by Fig. 5.4, and find they are acceptably described by power law behaviors. These plots affirm that, for instance, observing promoters regulated by k transcription factors is strongly (superlinearly) unlikely in k. This behavior, along with the large total number of possible interactions, reconciles the fact that the most common regulatory motif remains one transcription factor per promoter, but that with long tails in these distributions, there are some rare promoters regulated by many transcription factors.

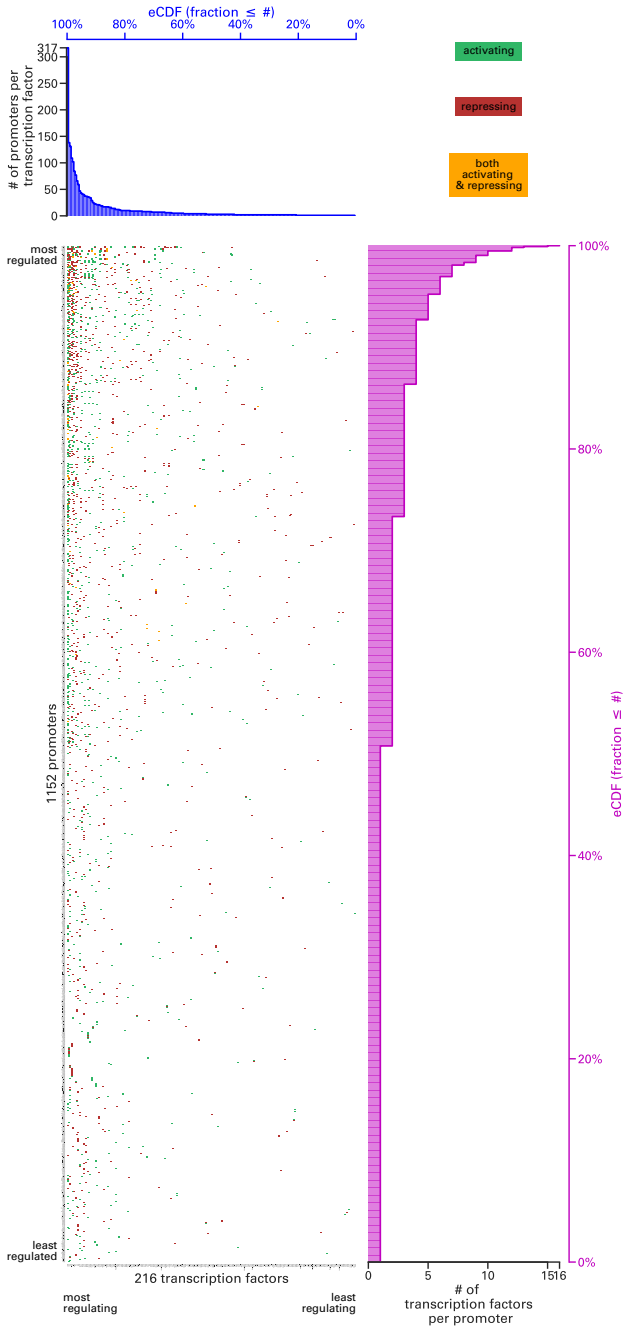


Figure 5.3: *E. coli*'s regulatory network as in Fig. 5.2, now visualized as a biadjacency matrix where a pixel (i, j) is white if no interaction is reported between promoter i and transcription factor j, and colored otherwise if this regulation is reported. Rows and columns are arranged in decreasing degree from top to bottom and left to right; this means the total number of transcription factors associated with a given promoter, or the total number of promoters associated with a given transcription factor, trace out the empirical cumulative distributions of the respective degrees, visualized as marginal plots. Note the striking sparsity of the regulatory network; only 2,355 out of the 248,832 possible (transcription factor, promoter) interactions in this table are present (0.95%), and if all the other putatively "unregulated" promoters (not visualized in this table) are included this fraction is only 0.27% (2,355 out of 868,968).

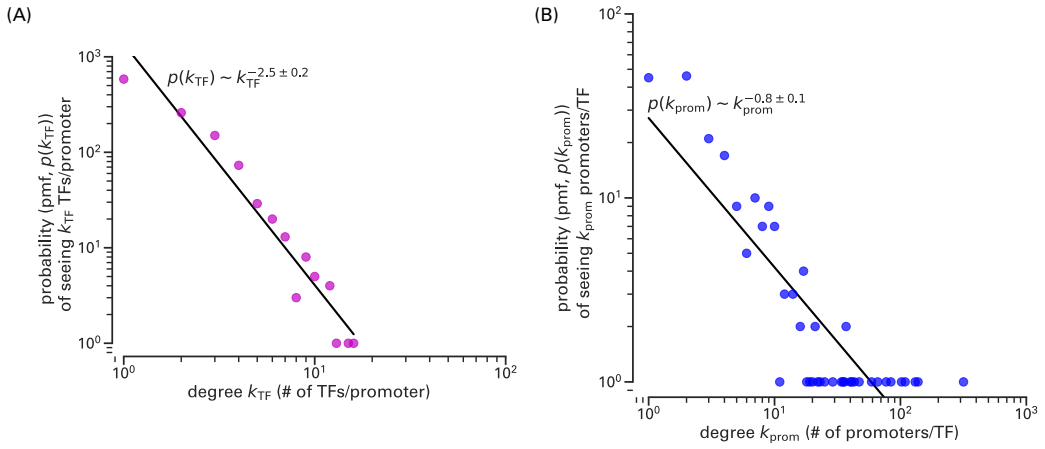


Figure 5.4: The two bipartite degree distributions in *E. coli* are acceptably well described by power laws. (A) The probability $p(k_{\text{TF}})$ that k_{TF} transcription factors regulate a promoter scales as $\sim k_{\text{TF}}^{-2.5}$ (black line; given by a regression minimizing relative error). This strongly-superlinearly decaying relationship says, for instance, that seeing a promoter with twice a given number of associated transcription factors is $2^{-2.5} \approx$ over 5.6 times less likely. (B) The probability of seeing a given number of promoters associated with a transcription factor is weakly sublinear in this number, $p(k_{\text{prom}}) \sim k_{\text{prom}}^{-0.8}$.

5.2.1 Natural statistics of regulatory architectures in *E. coli*

For a more coarse-grained understanding of the statistics of this regulation, we summarized the raw interaction data just discussed in a yet separate way. We counted each distinct binding site according to whether the transcription factor is reported to be repressing, activating, or more rarely both at each distinctly-positioned binding site on the genome. Fig. 5.5 schematizes this count; the indicated layers stratify both the type of interactions (indicated by layer color) and whether the interactions are annotated with the required granularity of the number of binding sites between the interacting transcription factor and promoter (indicated by dotted or solid layer outline).

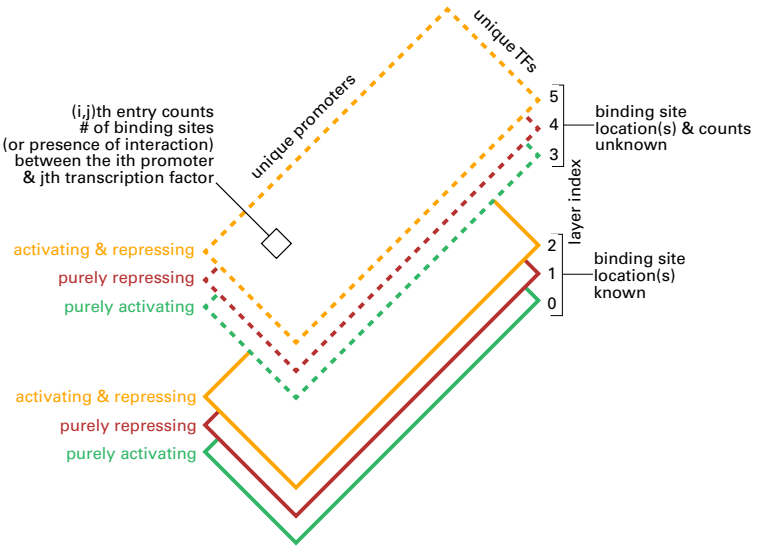


Figure 5.5: Categorical classification of counts of binding sites between promoters and transcription factors, separately acknowledging regulatory interactions where the locations or counts of binding site are not known/reported (dashed outlined layers). Since interactions without any localized binding sites cannot report how many individual transcription factors plausibly reside on the given promoter, the entries without specific binding site locations reported (layers 3-5) were omitted from future counts.

Next, we chose and followed a convention for counting the regulatory architectures achievable by the rare but present cases of binding sites identified as both repressing and activating.

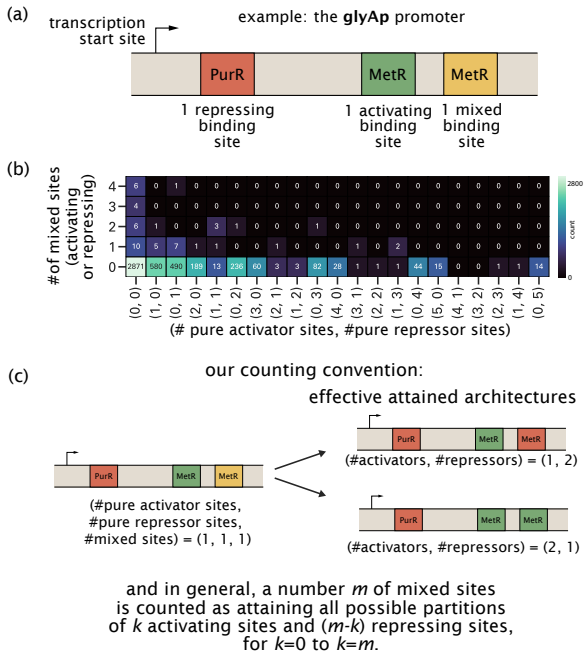


Figure 5.6: Accommodating cases in the regulatory network where individual binding sites (on a promoter for a particular transcription factor) are reported to be both activating and repressing. (A) One example of this type of mixed binding site is found in the glyAp promoter, where repressing, activating and mixed binding sites are respectively represented in red, green and yellow. (B) Heatmap showing counts of these mixed binding sites as rows, against columns depicting different counts of binding sites annotated as pure activating or repressing. Mixed sites are uncommon, but present. (C) Since such mixed sites can act as either repressors or activators, we adopt the convention of counting all possible effective architectures where the mixed binding sites assume either repressing roles (e.g. top) and/or activating role (e.g. bottom)s. Mathematically, this means that when there are m mixed binding sites, k sites can act as extra activating sites and $(m-k)$ can act as extra repressing sites, for $0 \leq k \leq m$. This convention allows us to collapse the architectures to represent examples of some number of activators and some number of repressors on the promoter.

As illustrated in Fig. 5.6, when we see a promoter with r purely repressive binding sites; a purely activating binding sites; and m sites that can be either repressing or activating, we count this scenario as attaining multiple architectures, namely the set $\{(a+k, r+(m-k))\}_{0 \leq k \leq m}$. Note that this convention only (very mildly) shapes the observed empirical distributions of architectures, since appreciably less than a percent of

all binding sites are annotated as both repressing and activating. Further, by construction, this convention does not affect the total copy number of transcription factors that are counted per promoter, the number of interest in defining what state space topologies promoters have. The resulting raw counts of regulatory architectures are summarized in Fig. 5.7. We also note that this observed regulatory architectures from an aggregating database is also highly compatible with the results of individual concerted experimental surveys.

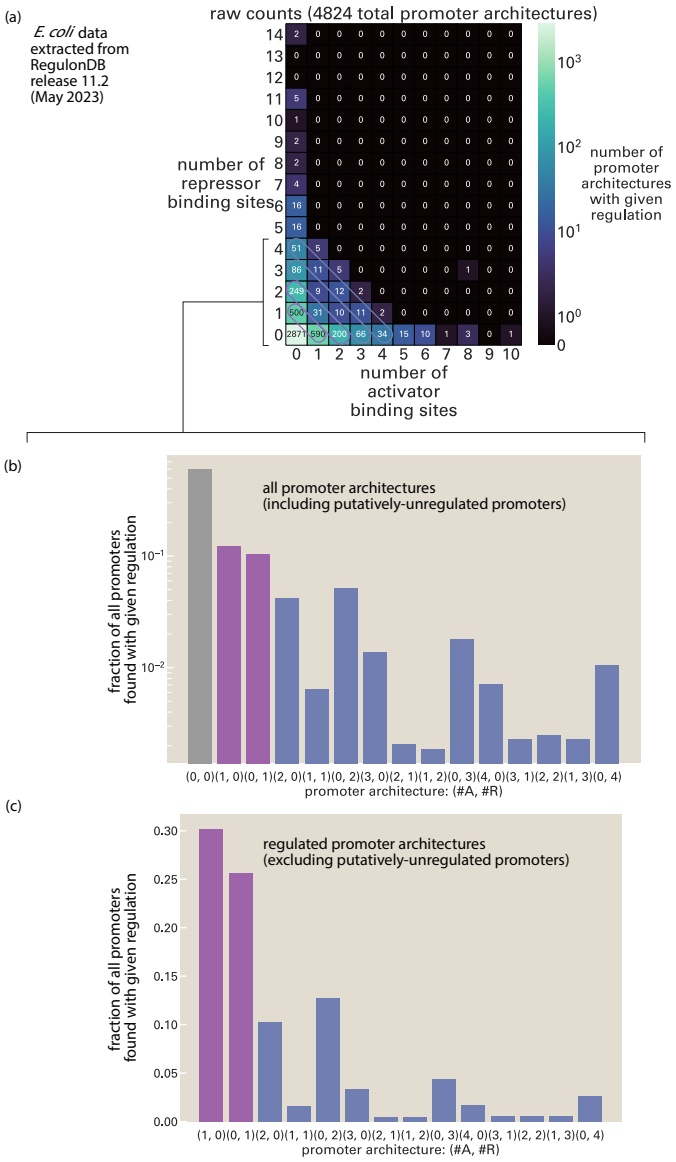


Figure 5.7: Incidence of regulatory architectures reported in *E. coli*, according to data in the RegulonDB database [3] release 11.2 (May 2023), organized by the joint numbers of activating and repressing transcription factors associated with each promoter. Architectures are regarded as the copy number of possibly-non-distinct proteins that can be bound to sites on each promoter. (A) The joint empirical distribution of the number of repressors and activators is largely symmetric (that is, $p(\#A, \#R) \approx p(\#R, \#A)$). (B) The joint distribution projected on a one-dimensional histogram on a logarithmic y-scale, including putatively-unregulated promoters with no known regulatory interactions (in grey) that comprise the majority ($\sim$ 71%) of promoters. Purple bars highlight promoters reported to have one associated transcription factor. (B) The incidence of regulatory architectures among those promoters with at least one known regulatory interaction. Over half of these regulated promoters are reported to be be regulated by the most common architecture of a single transcription factor (purple bars), and thus are plausibly describable by the square graph of states analyzed most thoroughly in the preceding chapter's main text. (This figure forms the basis of Fig. 2A in the preceding chapter's main text.)

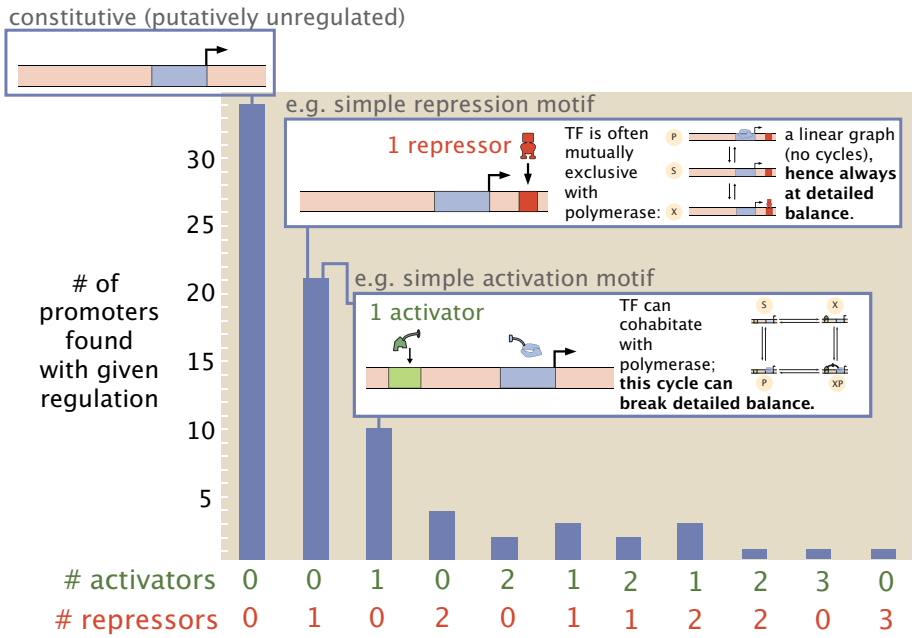


Figure 5.8: An experimental histogram of empirically-observed gene regulatory motifs in *E. coli* [5] reveals that many promoter sites are regulated by a single repressor or activator. A single repressor can often implement the simple repression motif, where the repressor excludes the polymerase from binding, allowing just three states in a linear graph. (Raw histogram data are courtesy of Reference [5].) The reason that the "simple activation motif" is schematized as linked to both the (0 activator, 1 repressor) and (1 repressor, 0 activator) histogram bar is that while steric exclusion commonly occurs for repressors, often making single repressors well described by a linear graph of three states, some repressors do not completely exclude the polymerase, permitting a cycle motif too.

Reference [5] is among the widest experimental censuses discovering regulatory interactions in *E. coli* in the recent literature. This study also found that transcriptional architectures with one activator or repressor are the most commonly observed regulated transcriptional architectures.

Thanks to common steric overlaps between the repressor binding site and polymerase binding site [6], a repressor is often—though not necessarily—mutually exclusive with the polymerase (see Fig. 5.8). In this case, the repressor implements a simple repression motif, a graph which lacks a cycle [7]. However, when the repressor does not sterically exclude the polymerase, a cycle of four states emerges. The same cycle of four states emerges with activators, whose binding sites rarely directly overlap with the polymerase binding site [6]; this produces a "simple activation motif." These observations affirm that a single cycle of four states is a common motif in prokaryotic gene regulation, but other motifs are also widely observed.

5.3 Incidence of known transcriptional architectures in Drosophila from Redfly

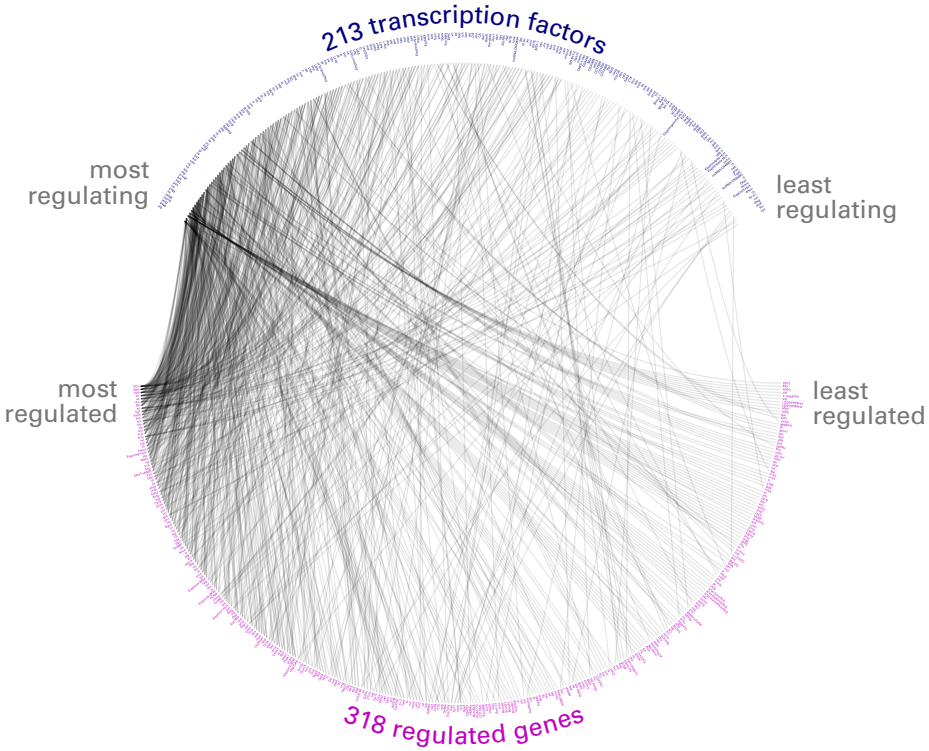


Figure 5.9: The bipartite network of all known regulatory interactions between transcription factors (top, navy nodes) and genes (bottom, magenta nodes) in the fruit fly eukaryote *D. melanogaster*, constructed from raw data scraped from the Redfly database v9.6.2 as of October 2023 [8] (`http://redfly.ccr.buffalo.edu/about_redfly.php`). Each black arc visualizes an interaction between a transcription factor node on top and a gene node on the bottom. Nodes are arranged in decreasing degree from left to right on their respective arcs, such that the most highly regulated genes (starting with Ubx, dpp, eve, en...) appear on the left on the bottom arc, and the most promiscuous/highly regulating transcription factors (starting with Dif, Mad, Derf, Su(h), ...) appear on the left on the top arc. As hinted at by the surprisingly dense connections from upper left to lower right or lower right to upper left, Drosophila's network is also detectably disassortative, with an apparent assortative coefficient of -0.18; this is moderately less disassortative than *E. coli*'s regulatory network.

Having just analyzed how common regulatory interactions of different promiscuity (and so, different dimensionality of the underlying graph of state spaces) appear to be in a prokaryote, it is interesting to contrast this with that in a eukaryotic context. Fig. 5.9 visualizes the regulatory network in the fruit fly we parsed from the *Redfly* database; we visualized these interactions according to a similar scheme as in our hairball network visualization of *E. coli* in Fig. 5.2.

Similarly, Fig. 5.10 reports the biadjacency matrix of this regulatory network; Fig. 5.11 depicts power law behavior to regulatory degrees, as in *E. coli*; and Figs. 5.12 and 5.12 summarize these networks in terms of the incidence of certain types of regulatory architecture.

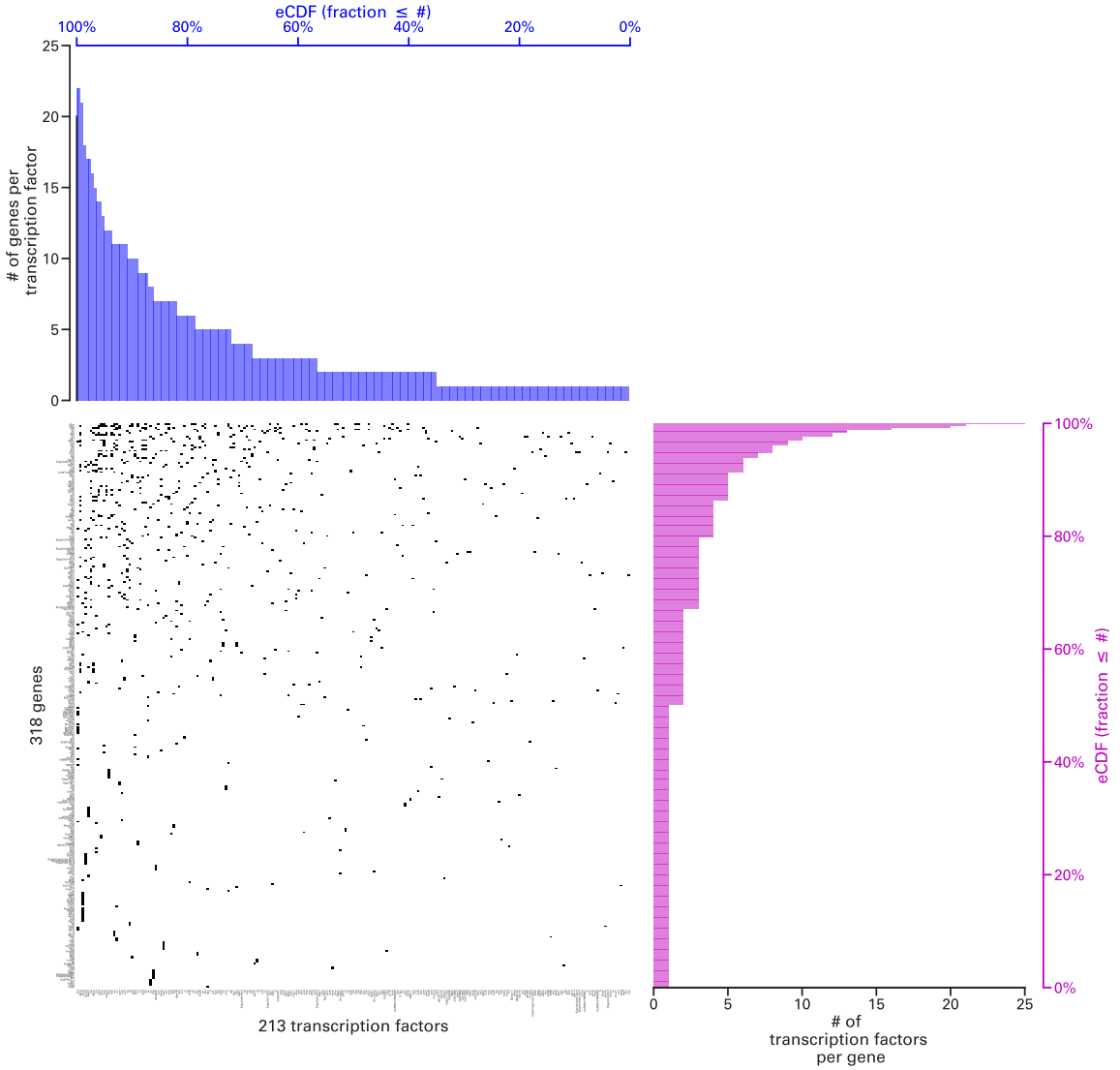


Figure 5.10: Like *E. coli*'s regulatory network, Drosophila's regulatory network is very sparse: only 831 of 67,734 (regulated TF, gene) possible pairs are genuinely reported to be regulatory ($\sim 1.2\%$). If we also count the 17,410 additional putatively-unregulated genes, the proprotion drops to 831 of 3,776,064 (or just 0.02%). See https://www.its.caltech.edu/~gsalmon/gs069_Dmelanogaster_biadjacency_Redfly202310.html for an interactive version of the heatmap where individual interactions can be gleaned by hovering with a mouse.

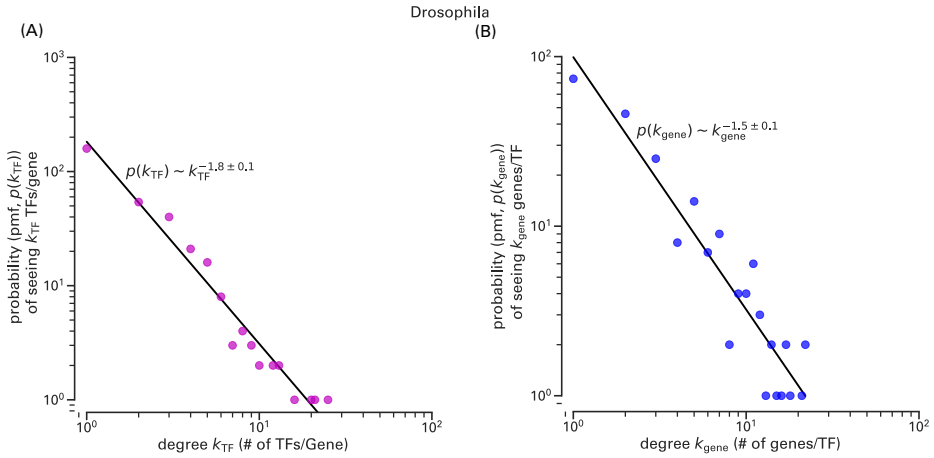


Figure 5.11: The two bipartite degree distributions in Drosophila are acceptably well described by power laws. (A) The probability $p(k_{\mathrm{TF}})$ that k_{TF} transcription factors regulate a gene scales as $\sim k_{\mathrm{TF}}^{-1.8}$ (black line; given by a regression minimizing relative error). This superlinearly decaying relationship says that seeing a gene with twice a given number of associated transcription factors is $2^{-1.8} \approx 3.5$ times as unlikely. (B) The probability of seeing a given number of promoters associated with a transcription factor also superlinearly decays with degree, $p(k_{\mathrm{prom}}) \sim k_{\mathrm{prom}}^{-1.5}$.

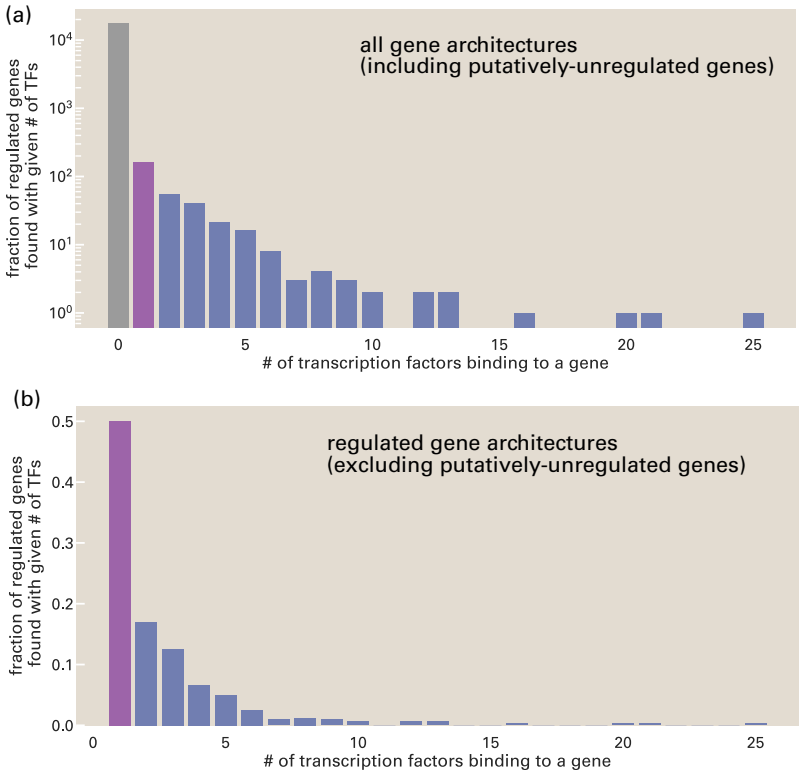


Figure 5.12: Histograms of number of distinct transcription factors interacting with each gene in Drosophila, according to data reported the Redfly database [8] v9.6.2 as of October 2023. (A) Histogram of the number of distinct transcription factors reported to bind a gene, including the putatively-unregulated genes in Drosophila (gray bar). (B) Histogram of the number of transcription factors associated with genes with at least one reported regulatory transcription factor; the purple bar represents the large proportion (about half) of regulated genes reported to be regulated by a single transcription factor. These data (up to five transcription factors) form the inset of Fig. 2A in the preceding chapter.

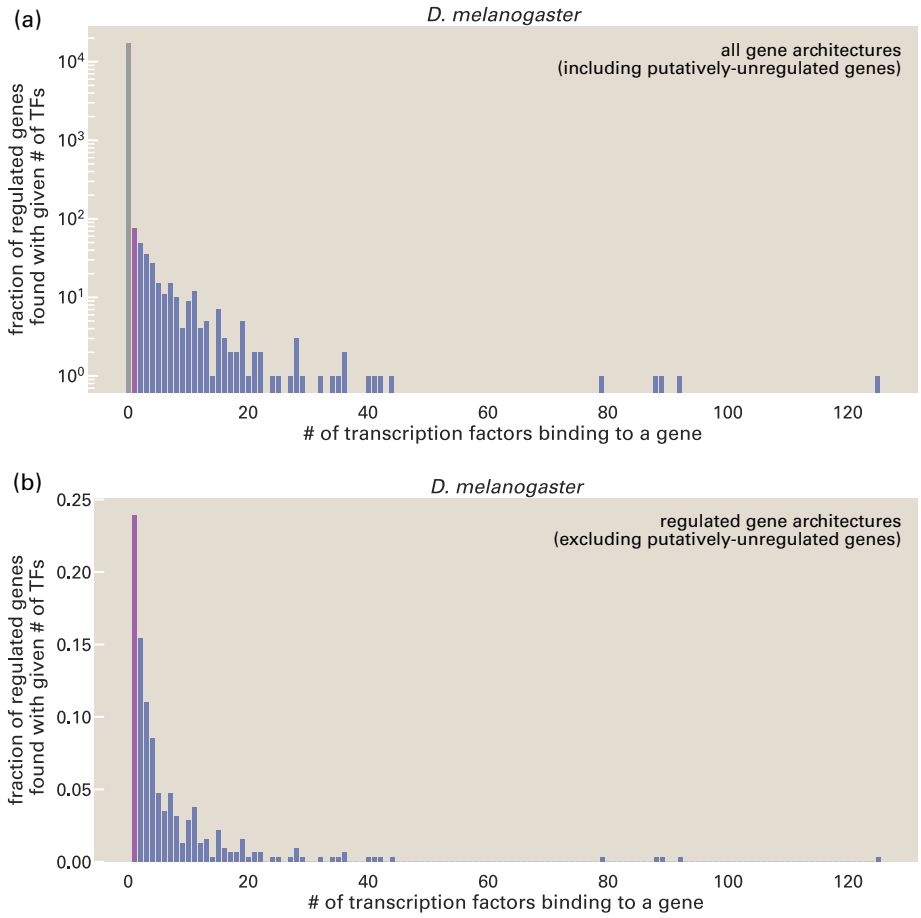


Figure 5.13: Histograms of the copy numbers of transcription factors capable of interacting with each gene in Drosophila, equivalent to the numbers of transcription factor reported binding sites per gene, according to data reported the Redfly database [8] v9.6.2 as of October 2023. (A) Histogram of the number of distinct transcription factors reported to bind a gene, including the putatively-unregulated genes in Drosophila (gray bar). (B) Histogram of the number of transcription factors associated with genes with at least one reported regulatory transcription factor; the purple bar represents the large proportion (about half) of regulated genes reported to be regulated by a single transcription factor. These data (up to five transcription factors) form the inset of Fig. 2A in the preceding chapter. (Incidence of regulatory architectures regarded by number of binding sites.)

5.4 Understanding the hypercubic graph: regulation by N transcription factors

The empirical distributions of known regulatory architectures for *E. coli* and *D. melanogaster* summarized in §5.2, §5.3, and the preceding chapter's Fig. 2A show that fully $\sim$ a quarter to a half of promoters or genes are reported to be regulated by the most common case of just one transcription factor. These scenarios are often plausibly described by a square graph of states and subject to the restrictions and behaviors we explore thoroughly in the preceding chapter. Yet clearly, promoters and genes can also be under the control of multiple transcription factors. The presence or absence of these transcription factors on the genome form hypercubic state spaces; some simple examples appear as annotations in Fig. 2A of the preceding chapter.

In this section, we derive the mathematical behavior of any response observable defined by these hypercubic networks, in or out of equilibrium. This derivation follows by understanding the strong structural constraints on the possible spanning trees in a hypercube. Following an illustration with small hypercubes and general arguments, Equations 5.6 and 5.8 report these generic nonequilibrium and equilibrium behaviors, respectively, thus fully generalizing Equations 1 and 3 of the preceding chapter to arbitrary combinatorial binding topologies. Obviously, such more complex network topologies can show more complex and higher-dimensional mathematical behaviors than simple regulatory cycles. However, their flexibility remains constrained, and later in §5.4.2 we also discuss how biophysical constraints—such as the common scenario of overlapping binding sites—can significantly simplify responses (see Figures 5.26 and 5.27).

As new experimental measurements become available, these equations can be used for quantitative fits to gene expression input-output curves, accommodating promoters known to be under the control of many transcription factors. As the contrast between nonequilibrium and equilibrium behaviors grows in higher dimension, these results give

tools to discern whether more complicated networks show signatures of operating out of equilibrium. These mathematical behaviors also prelude more specific investigations of attainable sensitivity, newly possible curve shapes, and specific access to interesting regulatory phenotypes, like those found for the square graph in later sections of the preceding chapter.

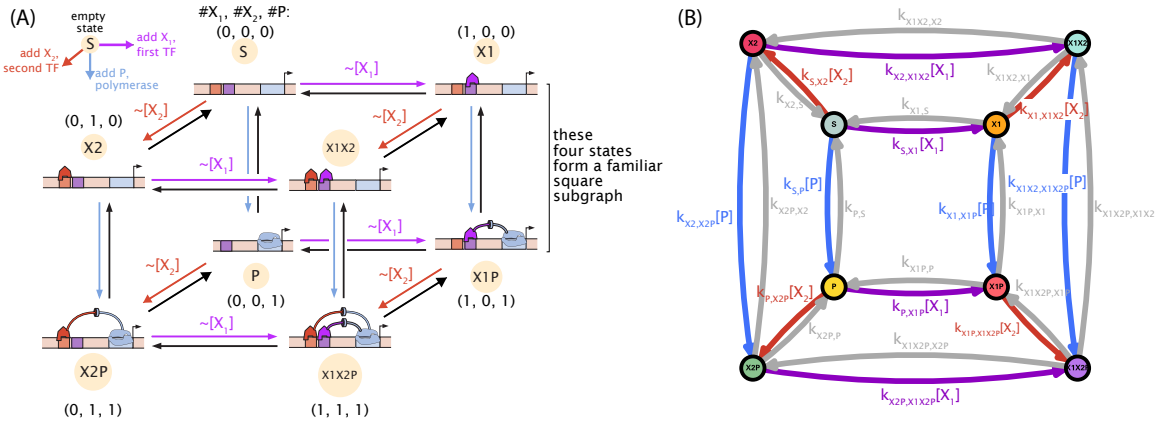


Figure 5.14: Regulation with two transcription factor binding sites: a cubic graph. (A): Allowing another transcription factor extends the state space from a square to a cube. In general, for m transcription factors and one polymerase, the state space will become a $n = (m+1)$ dimensional hypercube, where the $2^n = 2^{m+1}$ vertices encode the presence or absence of each of the transcription factors or polymerase. This description assumes that every transcription factor is mutually compatible with the presence of other transcription factors; these transcription factors need not be distinct, if some proteins each have multiple binding sites, in which case some concentrations control multiple sets of orthogonal edges. Here we visualize the $m = 2$ transcription factors case, giving a $n = (2+1)$-dimensional cube. Purple rightwards arrows $\searrow$ illustrate binding with transcription factor X_1; red leftwards arrows $\swarrow$ denote binding with transcription factor X_2; and downwards vertical $\downarrow$ arrows denote binding with polymerase P. Each state is annotated with the appropriate binary code (a vector of length $n = (m+1)$ whose ith entry is 1 if transcription factor i is present in that state, and 0 otherwise, with the convention that the final $(m+1)$th entry refers to the polymerase). We refer this code to help derive how algebraic responses collapse at detailed balance. (B): The same $n = 2+1$ hypercube-graph , visualized as a planar Tutte embedding, with rates associated labeled.

As a hint to the impending complexity, while we only had sixteen directed spanning trees on the square ($n = 2$-cube), many more trees confront us for hypercube graphs in higher dimensions. In general, the number of undirected spanning trees on the n-cube is,

$$T_u(n) = \frac{1}{2^n} \prod_{i=1}^{n} (2i)^{\binom{n}{i}}. \tag{5.1}$$

This enumeration result follows from nontrivial combinatorial arguments [9] or from finding the eigenvalues of the Laplacian by analyzing the vertices of the hypercube as an Abelian group via representation theory [10, § 5.5, p. 62]. (These counts of the number of undirected trees on the n-cube are reported in the Online Encyclopedia of Integer Sequences as terms of the sequence OEIS A006237.) Specifically, there are $T_u = 384$ undirected spanning trees on the 3-cube. Since any undirected spanning tree can be rooted at each vertex and generate a unique directed spanning tree, there are $N_{\text{vertices}} = 2^n$ rooted spanning trees per undirected spanning tree, giving

$$T_d = N_{\text{vertices}} T_u(n) = 2^n T_u(n) = \prod_{i=1}^{n} (2i)^{\binom{n}{i}}$$

directed spanning trees. Or, for $n = 3$, there are $T_d = 2^n T_u = 8 \times 384 = 3{,}072$ directed spanning trees on the 3-cube. This complexity explodes rapidly: the next largest cube, the tesseract ($n = 4$), has a whopping 42,467,328 undirected spanning trees (so 679,477,248 directed spanning trees), and so on. We plot this explosion of rooted spanning trees in Fig. 5.15.

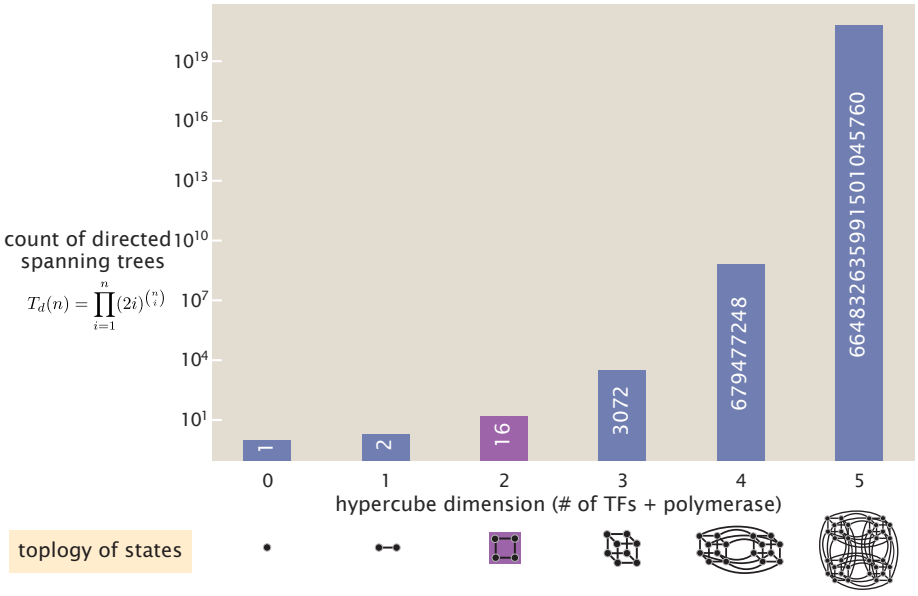


Figure 5.15: Superexponential explosion of the number in rooted spanning trees with respect to the hypercube dimension.

Notice that by the time the network is governed by only four transcription factors (forming a hypercube in five dimensions), the number of rooted spanning trees has exploded to over 664 quintillion. This number is over a "millimole" of spanning trees, a supercosmological number: if we managed to speedily sketch 1 spanning tree per second, it would take us $\boxed{\sim 1500\times}$ the current age of the universe to find all these trees. Clearly, brute force must be replaced by the foresight of structural reasoning about these trees.

5.4.1 Illustration with the $3-$cube case

We first illustrate the consequences of multiple transcription factors with the tractable case of the $n = 3$ dimensional cube, capturing the impact of two transcription factors (call them X_1 and X_2) which can be present at up to one copy each, in addition to the polymerase. This architecture, where appropriate transitions scale with each of these control variables, is depicted in Figures 5.14.

To enumerate the spanning trees, we exploit the Matrix Tree Theorem: removing row and column i in the Laplacian matrix and taking the determinant gives, up to a sign, the sum of all spanning trees rooted in state i, a procedure amenable to symbolic computation. Figure 5.16 illustrates the 384 trees rooted in just the state S, where the node positions are the same as those in Fig. 5.14.

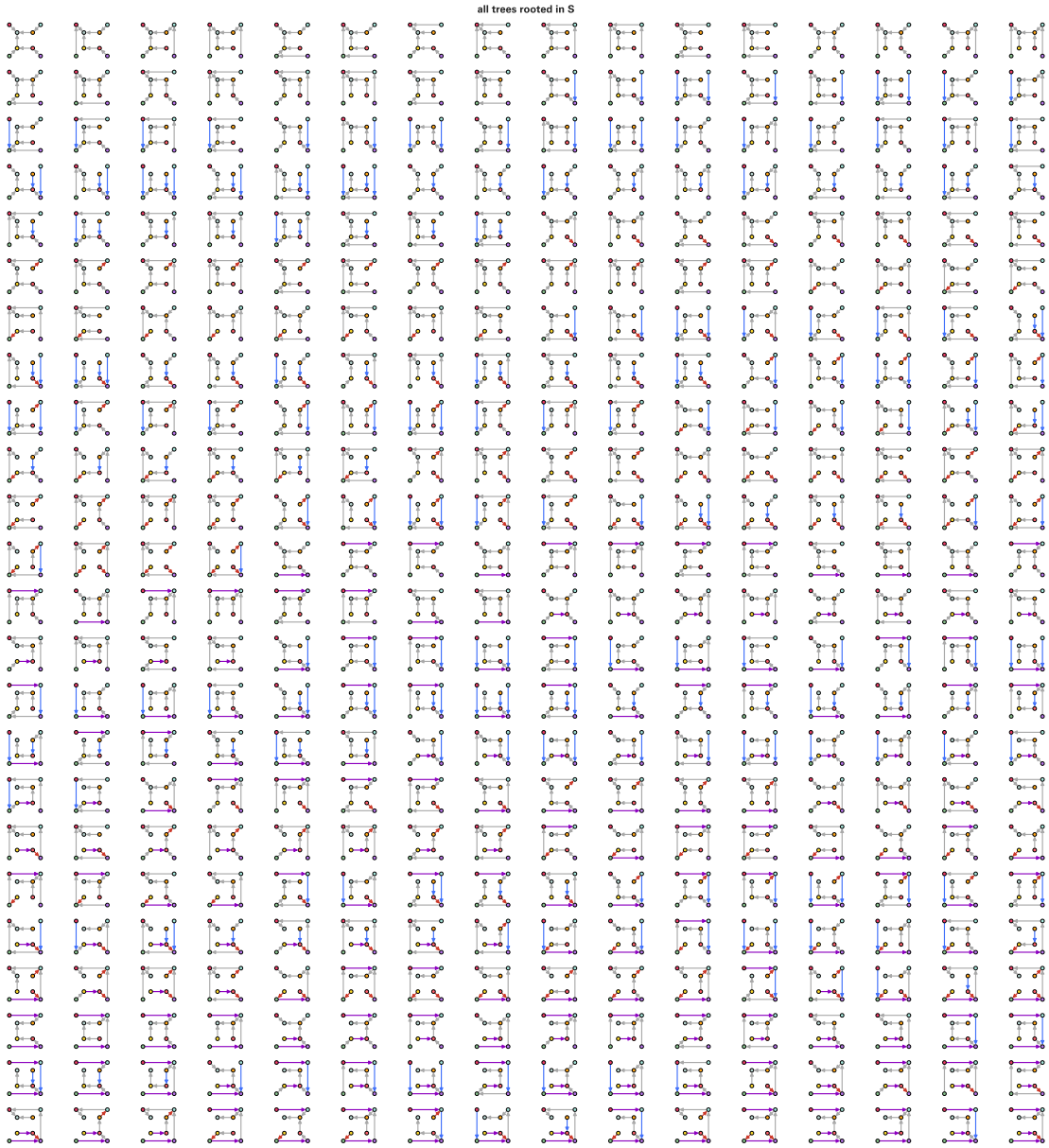


Figure 5.16: All directed spanning trees that are rooted in state S, computed by symbolic determinants, on the $n = 3$ cube graph of states. The powers of transcription factors $[X_1]$ and $[X_2]$ found in each spanning tree increases from the top left to the bottom right in this grid.

A subset of these trees can be organized into a two-variable analog of our main text's tabular figure 3B, showing how different root states adopt varied dependencies on powers of the control variable(s). We map examples of all attained mixed powers of the transcription factor concentrations for just one representative state, for illustration, in Figure 5.16. (These trees, being just those rooted in state S, are one eighth of all the trees; those rooted in the seven other states are shown in full in e.g. our Reference [1], SI.) Accumulated over states, these dependencies give ρ_{tot}, and therefore any response $\langle r \rangle$, up to a quartic dependence on each control variable, since it is possible to form rooted trees that use all four $[X_i]$-controlled edges present in the graph. However, the graph's structure enforces that not all combinations of powers of the transcription factors are accessible, as we now discuss.

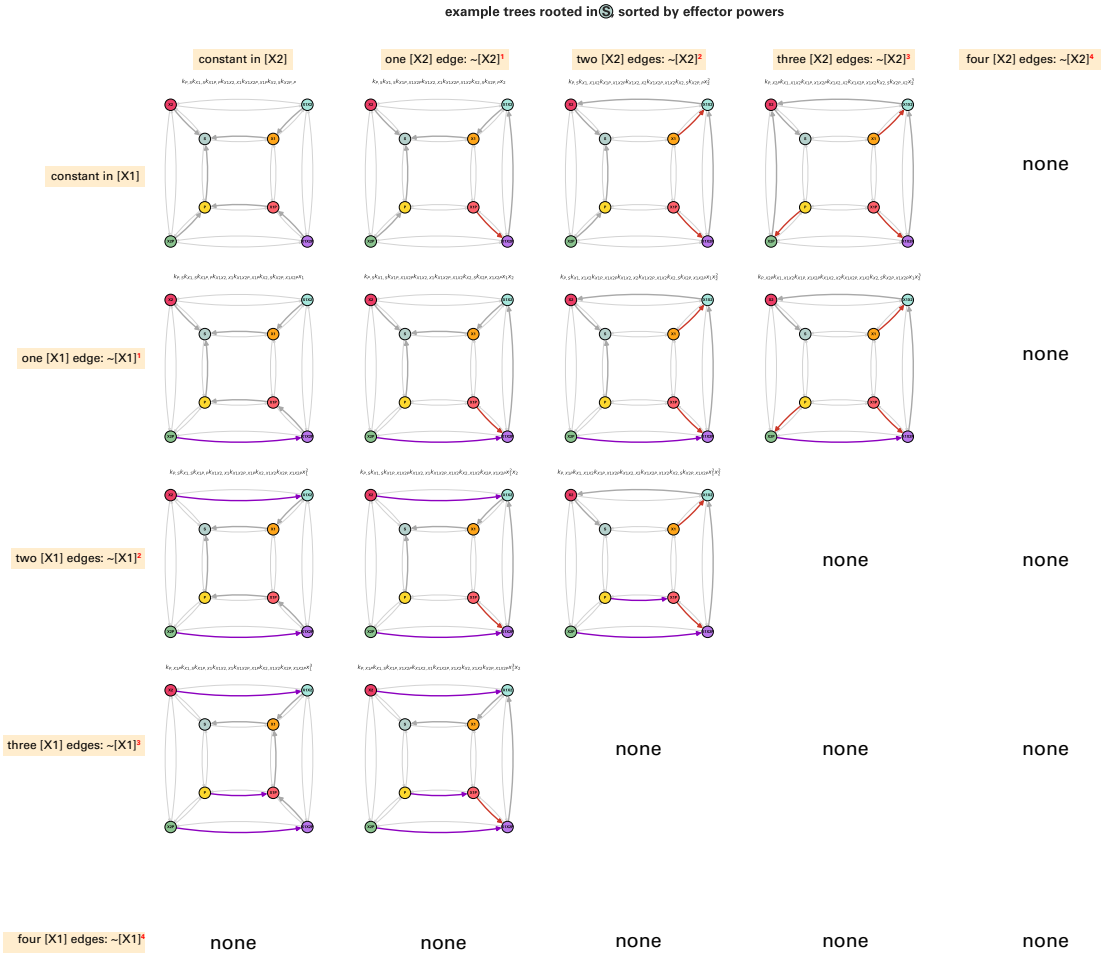


Figure 5.17: Examples of spanning trees rooted in the (empty) state S that contribute different combined powers of each transcription factor. Specifically, the entry at row m and column l of the grid exemplifies a spanning tree (when it exists) that has m edges dependent on transcription factor $[X_1]$ and l edges dependent on transcription factor $[X_2]$, contributing a monomial term $\sim [X_1]^m[X_2]^l$ to any response function. This establishes that, for some constants A_{ij} entirely determined by rate constants,

$$\rho_S = A_{00} + \sum_{i=1}^{3} A_{i0}x_1^i + x_2\sum_{i=0}^{3} A_{i1}x_1^i + x_2^2\sum_{i=0}^{2} A_{i2}x_1^i + x_2^3\sum_{i=0}^{2} A_{i3}x_1^i. \tag{5.3}$$

(State P attains the same possible powers of $[X_1]$ and $[X_2]$ shown here.)

In summary, these allowed spanning tree structures over all roots give any response (in or out of detailed balance) for the $m = 2$ transcription factors case (two control variables $[X_1], [X_2]$) of the form,

$$\langle r \rangle = \frac{A_{00} + \sum_{i=1}^{4} A_{i0}[X_1]^i + [X_2]\sum_{i=1}^{4} A_{i1}[X_1]^i + [X_2]^2\sum_{i=1}^{4} A_{i2}[X_1]^i + [X_2]^3\sum_{i=1}^{3} A_{i3}[X_1]^i + [X_2]^4\sum_{i=0}^{2} A_{i4}[X_1]^i}{B_{00} + \sum_{i=1}^{4} B_{i0}[X_1]^i + [X_2]\sum_{i=1}^{4} B_{i1}[X_1]^i + [X_2]^2\sum_{i=1}^{4} B_{i2}[X_1]^i + [X_2]^3\sum_{i=1}^{3} B_{i3}[X_1]^i + [X_2]^4\sum_{i=0}^{2} B_{i4}[X_1]^i}. \tag{5.4}$$

(Note that while these terms are organized into groups by constant powers of $[X_2]$ for convenience, this ordering is arbitrary: this response is indeed symmetric under interchange of $[X_1]$ and $[X_2]$. The "democratic" structure of the hypercube graph enforces that no transcription factor is special.)

Slicing the response function when one transcription factor is fixed

Notice that if say the second transcription factor were held fixed at some concentration $[X_2] := [X_2]_0$, the response function is a ratio of quartic polynomials in the remaining transcription factor control parameter. Clearly this response, even with respect to just one control parameter, has acquired new flexibility. Whether a complex experimentally-observed response is attributable to a simple system operating out of equilibrium, or a more architecturally-complex system operating near equilibrium, will depend on specific biological details and separate experimental and biological context—namely, whether many transcription factors' binding sites are indeed found in a promoter and whether these proteins are actually expressed. Nevertheless, such details are often known—or knowable—for genes and promoters of interest.

Full generalization to the hypercube in n dimensions: cubes within cubes

What explains which transcription factor powers are attained (or not) in this cube graph, and how does this structure generalize to an arbitrary number of transcription factors? To answer this question, it is instructive to consider how a spanning tree could maximize both the numbers of edges controlled by one transcription factor x_1, and those controlled by another x_2, simultaneously.

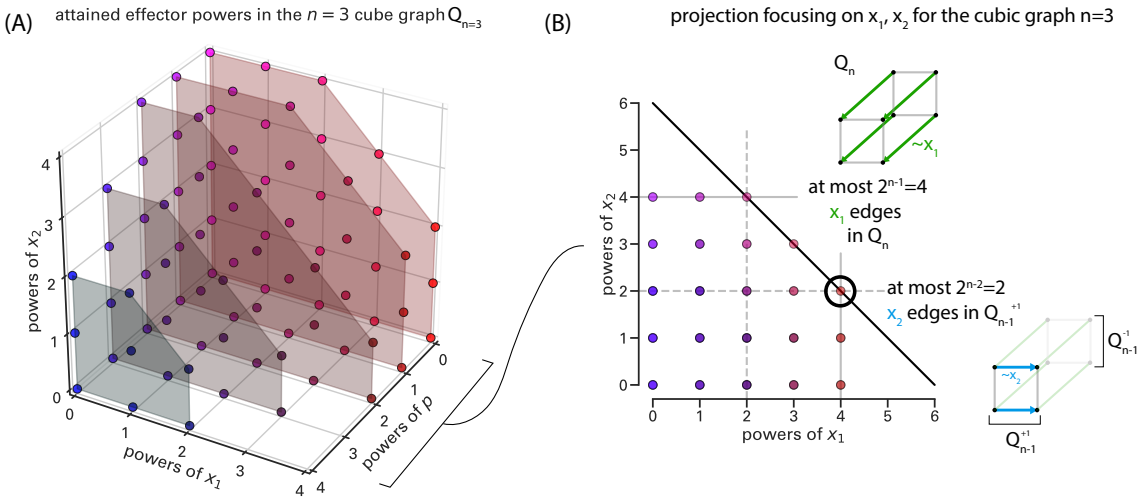


Figure 5.18: Powers attained by the transcription factors x_1 and x_2, in addition to the polymerase p, over all spanning trees in the cubic $n = 3$ graph. (A) Powers of $x_1^i x_2^j p^k$ involved in the numerator and denominator of the output function (5.4). If there is a dot at position (i, j, k) in space, then there is a term proportional to $x_1^i x_2^j p^k$ in both numerator and denominator of (5.4). The color of a dot corresponds to a power of p. Pink dots are term with a contribution of order 0 in p, the blue ones of order 5 in p. The plotted planes with the corresponding colors are the planes are constant power of p. (B) Powers of $x_1^i x_2^j$ involved in the numerator and denominator of the output function (5.4), not taking into account the polymerase P. (Indeed, this concentration of polymerase is often taken to be constant in regulatory models.) Q_n represents the full hyper-cubic graph of dimension n; in this example, $n = 3$. The edges having an x_1 dependency are represented in green. In a graph representing the binding of two transcription factors, there are at most 2^{n-1} edges depending on the concentration of one of the transcription factors. (In dimension n the hypercube represents the binding of two transcription factors and of the polymerase to the DNA, 2^{n-1} states don't have x_1 bound to the DNA and could acquire it, creating 2^{n-1} edges with a x_1 dependency). In dimension 3 the number of this type of edges is 4. The circled dot represents a spanning tree in which, these 4 edges are participating. As there are 6 nodes in a cube, two edges are still needed to create a spanning tree. The two edges presented in blue, depending on x_2, complete the four green edges, depending on x_1, in a spanning tree. In fact, to complement the 4 green edges into a spanning tree, the two edges have to be in Q_{n-1}^{+1}. Q_{n-1}^{+1} and Q_{n-1}^{-1} are the $n-1$ dimensional cubes formed by states in which X_1 is respectively bound or unbound to the DNA and the edges connecting them. Indeed, the 2 edges have to depart from a node in which X_1 is bound to the DNA (included in Q_{n-1}^{+1} by definition), because the participation edges in Q_{n-1}^{-1} would create nodes with two outgoing edges. In Q_{n-1}^{+1}, there are up to 2^{n-2} edges depending on x_2. Therefore, complementing the 4 green edges, with a x_1 linear dependency, by the two blue edges, with a x_2 linear dependency, creates a spanning tree with a quartic dependency in x_1 and quadratic in x_2, corresponding to the circled dot.

Figure 5.18 summarizes the attainment of powers of x_1, x_2, and polymerase p, where the powers attained by x_1 and x_2 are emphasized by a projection over the attained polymerase powers at right. A tree on the full (hyper)cube can easily contain all of the possible edges driven by a particular transcription factor, say x_i: there are 2^{n-1} such x_i-driven edges in the whole graph (since half of the 2^n states in a full hypercube lack this particular transcription factor, and from each of these x_i-deprived states we count one x_i-driven edge), explaining how trees can attain powers of $2^{n-1} = 4$ in either x_1 or x_2 individually. (This boundary is visualized by the unbroken grey line in Fig. 5.18.) However, these trees that contain all of the possible x_i-driven edges cannot simultaneously contain all x_j-driven edges in the graph. This is because, by construction, all of the x_i-deprived states already have exactly one outgoing edge (precisely to assure each of the x_i-driven edges belong to the tree), so the remaining edges that build the spanning tree must be found wholly among the half of the graph's states where x_i is *present*. This restriction halves the number of possible x_j edges that can now participate in the same tree.

In other words, if Q_n represents the full n-dimensional hypercube, then it can always be partitioned into two sets of states, one set deprived of transcription factor x_i and the other enriched by x_i. (In these sets of states, we also consider the interconnecting edges among them.) These sets of states and edges each form an $(n-1)$-dimensional hypercube; call Q_{n-1}^{+i} the subcube where X_i is present, and Q_{n-1}^{-i} the subcube where X_i is absent. The full n-dimensional hypercube is formed by joining these smaller hypercubes together by the perfect matching formed by the set of edges describing the binding of transcription factor X_i. Inside each smaller hypercube Q_{n-1}, there are only 2^{n-2} possible edges driven by the other transcription factor X_j; and demanding that all of the X_i edges appear in a tree restricts the remaining edges to be taken from the Q_{n-1}^{+i} cube of states (where x_i is present). This scenario is illustrated by the example of the power combination circled in black in Fig. 5.18.

If we relax our demand that *all* 2^{n-1} edges of some transcription factor X_i participate in

a tree by allowing one fewer x_i-driven edge to appear, we can find a tree that now includes up to one extra power of another transcription factor X_j; this extra power will come from one of the edges in the Q_{n-1} formed by the X_i-deprived states (formerly forbidden from contributing to the powers of x_j that could participate). Continuing to remove X_i edges continues to add new possible x_j edges in valid spanning trees, up to the 2^{n-1} allowed (by symmetric reasoning). This explains the diagonal boundary of attained powers marked by a black line with slope -1 in the right of Fig. 5.18; in general, this maximal boundary is a hyperplane assuring that the total powers of transcription factors (and polymerase) is at most

$$\#\text{ of total regulated edges in any tree (hence powers in response)} \leq \sum_{k=0}^{n} 2^{n-k} = 2^n - 1. \tag{5.5}$$

For instance, Figure 5.19 visualizes this planar constraint that powers of $[X_1]$, $[X_2]$, and P must ultimately sum to less than or equal to $2^3 - 1$ = seven regulated edges in the $n = 3$-cube.

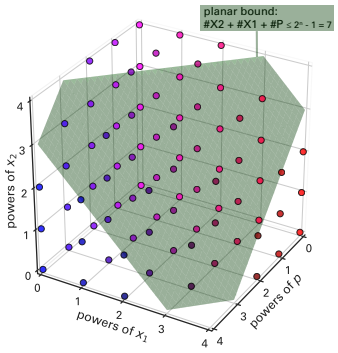


Figure 5.19: Visualization of how the plane $\#X_2 + \#X_1 + \#P \leq 2^n - 1 = 7$ (plotted in green) bound the attainable powers in any tree (hence response), in the $n = 3$-cube case. $\#X_2$, $\#X_1$ and $\#P$ respectively correspond to exponent of x_2, x_1 and p. Powers of $x_1^i x_2^j p^k$ involved in the numerator and denominator of the output function (5.4). If there is a dot in the (i, j, k) space, then there is a term proportional to $x_1^i x_2^j p^k$ in both numerator and denominator of (5.4). The color of a dot corresponds to a power of p. Pink dots are term with a contribution of order 0 in p, the blue ones of order 5 in p.

Furthermore, this logic applies recursively to each *subsequent* transcription factor $x_k \notin \{x_i, x_j\}$—for instance, this structure gives trees of maximal power $x_i^{2^{n-1}} x_j^{2^{n-2}} x_k^{2^{n-3}} \cdots$. This reasoning is agnostic to the dimensionality of the state space (total number of transcription factors). Since the maximal power over all transcription factors is at most 2^{n-1}, the second highest power over transcription factors is 2^{n-2}, and so on, we use the notation of order statistics to write the general form. Recall that the ith *order statistic* $y_{(i)}$ of a set of values $\{Y_1, Y_2, \ldots, Y_n\}$ refers to the value that occupies the ith position when the set of values is arranged in ascending order. In other words, it represents the ith smallest value in the set, so the ith largest value of a set is $y_{(n-i)}$. Using this notation, the generic response of any observable (in or out of detailed balance) formed by the hypercubic graph from n transcription factors can be written as,

$$\boxed{\langle r \rangle = \frac{\displaystyle\sum_{\substack{\max\{i_l\}=i_{(n)}\leq 2^{n-1},\ i_{(n-1)}\leq 2^{n-2},\\ \ldots\ i_{(1)}=\min\{i_l\}\leq 2^0}} A_{\{i_l\}} \prod_{i=1}^{m} [X_i]^{i_l}}{\displaystyle\sum_{\substack{\max\{i_l\}=i_{(n)}\leq 2^{n-1},\ i_{(n-1)}\leq 2^{n-2},\\ \ldots\ i_{(1)}=\min\{i_l\}\leq 2^0}} B_{\{i_l\}} \prod_{i=1}^{m} [X_i]^{i_l}}}, \tag{5.6}$$

where we have temporarily adopted the convention that one of the "transcription factor"s (say the last one $[X_n]$) represents the polymerase. (Note that this response form obeys the interchange symmetry required of all the transcription factors.)

Further illustration with $n = 4$ tesseract graph

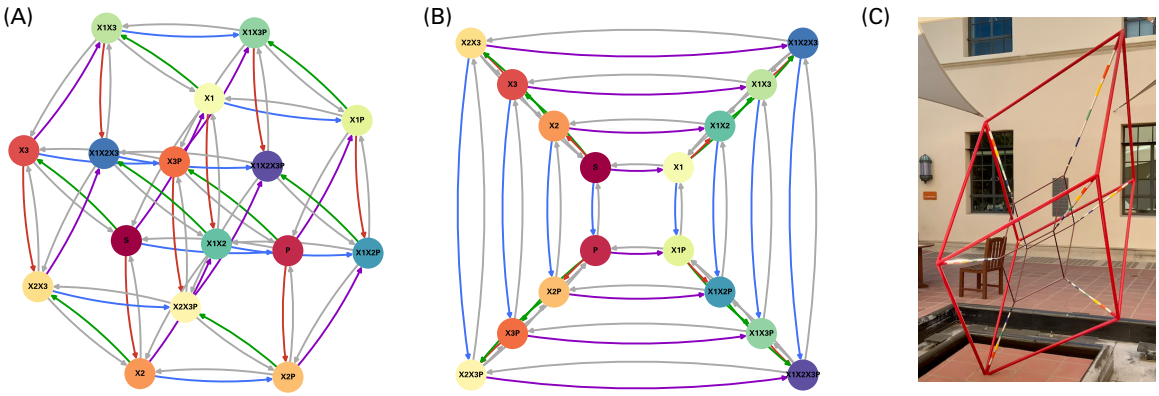


Figure 5.20: Adding a third transcription factor X_3 forms a hypercubic state space in $n = 4$ dimensions, known as a tesseract, represented in a concentric layout in panel (A). Panel (B) is a sculpture showing another popular arrangement of nodes in the tesseract, from artist Ronald Linde, pictured at Caltech.

Explicit enumeration of all 679,477,248 *directed* spanning trees for the 4-cube (tesseract) graph (giving the state space of three transcription factors acting in concert with the polymerase) is not very computationally tractable while preserving all symbolic dependencies on the 32 rate constants associated with the graph. However, we restrict our ambition to discerning just the algebraic form of the x_i dependencies, by replacing all of the rate constant prefactors by 1 but preserving the x_i-dependencies on the rates. This simplifies the symbolic computation of the determinant substantially, enough that *Mathematica* (but still no longer *Python* via `sympy` on consumer computing hardware) can now perform the computation, and reveal the powers attained for all transcription factors x_i.

We visualize the powers attained in this graph in Figure 5.21. These terms are consistent with the structure of the generic response form we indicate in Eq. 5.6.

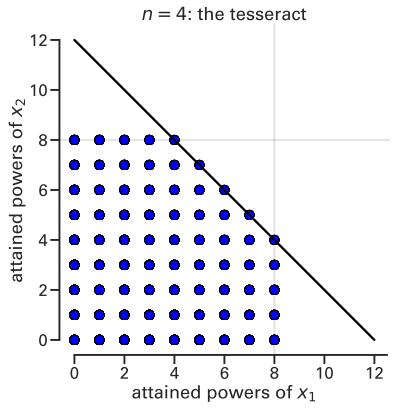


Figure 5.21: Powers of $x_1^i x_2^j$ involved in the numerator and denominator of the output function of a tesseract graph (hypercube for $n = 4$), representing a regulation unit with three transcription factor binding sites and depicted in Figure 5.20. x_2 and x_1 are the respective concentrations of two transcription factor.If there is a dot in the (i, j) plane, then there is a term proportional to $x_1^i x_2^j$ in both numerator and denominator of the output function, at constant x_3 and p (concentration of the third transcription factor and the polymerase). In this graph there are $\sim 6.8 \times 10^8$ spanning trees dependent on x_1, x_2, x_3, p.

Collapse of the output function at detailed balance

Earlier, in the square graph ($n = 2$-cube), we originally arrived at the response's demotion to a ratio of linear polynomials at detailed balance by explicitly using the cycle condition (namely, that the product of rate constants in one direction around the cycle equals that around the opposite direction at equilibrium). However, now for the $n = 3$-cube, we have no fewer than twelve coupled cycles, complicating a direct translation of our earlier tact to such more complicated graphs. Instead of explicit appeals to cycle conditions, we use the alternative structural reasoning we discus in Reference [1], SI.

By emphasizing the powers of control parameter accumulated along paths between states that are disconnected, this reasoning helps us understand how the $n = 3$-cube demotes its response at detailed balance. As before, we are essentially interested in expressing a given state j's statistical weight ρ_j relative to the statistical weight of a reference state (call it P again without loss of generality). The state P is directly connected to states $\{S, X_1P, X_2P\}$, so these terms contribute at most a pure linear power $[X_1]$ or $[X_2]$ to any responses. This leaves the remaining, disconnected states: $\{X_2, X_1, X_1X_2P\}$ are two edges away from P in the graph, and state X_1X_2 is a full graph diameter of three edges away from P. While it *is* possible to create paths between P and these other states that involve two edges dependent on an individual control parameter, actually this does *not* imply that the response is demoted to only a quadratic polynomial in that parameter, since all of these paths also must contain edges scaling as the inverse of that control parameter, canceling at least one power of this dependence.

This cancellation immediately becomes more clear when we return to the binary encoding of states we mentioned in Fig. 5.14. Every state in the hypercube graph, no matter how many edges apart, differs from any other state by at most a negation in each of its binary digits. When comparing two states in the graph, a negation in a particular binary coordinate i is exactly a difference in whether or not transcription factor i is present or

absent between the two states. Since every edge at most changes one binary coordinate at a time, this just says that the net number of binding versus unbinding events of transcription factor i along any path connecting two states is at most the maximum binary distance along a binary coordinate possible allowed by the graph, namely one. Summarizing in other words, even if two states are connected by a long path of possibly $[X_i]-$dependent arrows, the simple state space requires that every such path ultimately contribute at most a factor $[X_i]^{\pm 1}$ to the ratio of the statistical weights of the two states.

To wit, the response of a hypercube graph formed by the interaction of N transcription factors $\{X_i\}_{i=1:N}$ must ultimately collapse at equilibrium to the dramatically simplified multilinear form,

$$\langle r \rangle^{\text{eq.}} = \boxed{\frac{\sum_{z_1,z_2,z_3,\ldots z_N \in \{0,1\}} \alpha_{z_1,z_2,z_3,\ldots z_N} [X_1]^{z_1}[X_2]^{z_2}[X_3]^{z_3}\cdots[X_n]^{z_n}}{\sum_{z_1,z_2,z_3,\ldots z_N \in \{0,1\}} \beta_{z_1,z_2,z_3,\ldots z_N} [X_1]^{z_1}[X_2]^{z_2}[X_3]^{z_3}\cdots[X_n]^{z_n}}} \tag{5.7}$$

$$= \frac{\sum_{z_1,z_2,z_3,\ldots z_N \in \{0,1\}} \alpha_{z_1,z_2,z_3,\ldots z_N} \prod_{i=1}^{n} [X_i]^{z_i}}{\sum_{z_1,z_2,z_3,\ldots z_N \in \{0,1\}} \beta_{z_1,z_2,z_3,\ldots z_N} \prod_{i=1}^{n} [X_i]^{z_i}}. \tag{5.8}$$

For our $n = 3$ (two-transcription factor) example, the detailed balance response has the form,

$$\langle r \rangle^{\text{eq.},\, m=2} = \frac{\alpha_{111}[X_1][X_2]p + \alpha_{110}[X_1][X_2] + \alpha_{100}[X_1] + \alpha_{010}[X_2] + \alpha_{001}p + \alpha_{000}}{\beta_{111}[X_1][X_2]p + \beta_{110}[X_1][X_2] + \beta_{100}[X_1] + \beta_{010}[X_2] + \beta_{001}p + \beta_{000}}. \tag{5.9}$$

In this example, if e.g. $[X_2]$ is held at some constant, then this form collapses dramatically all the way back down to a ratio of linear polynomials in $[X_1]$. This collapse is still true in n dimensions: if all but a specific transcription factor's concentration $[X_i]$ were held fixed,

then the functional form with just this non-fixed control variable is still just a ratio of linear polynomials in $[X_i]$.

In summary, we have shown how the response function attains a restricted subset of possible combinatorial powers of control variables, in or out of equilibrium, on the n-cubic graph. We graphically summarize the algebraic constraints on these responses in Figure 5.22, which shows the attainable powers of any two control variables among the full n-space.

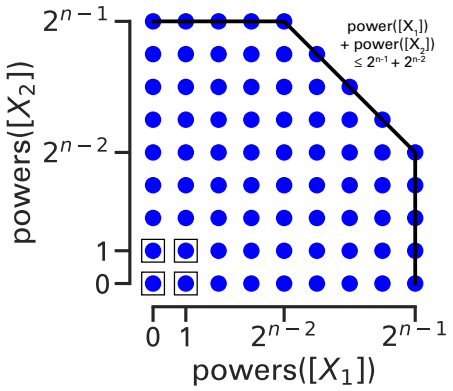


Figure 5.22: Graphical summary of how any two control variables $[X_1]$ and $[X_2]$ participate algebraically in the nonequilibrium or equilibrium response on an n-cubic graph. If a blue dot appears at position (i, j), then a monomial of the form $[X_1]^i[X_2]^j$ participates in the numerator and denominator of the algebraic form of the response $\langle r \rangle$. The dots outlined by the black boxes are the only powers attainable when the response is at equilibrium. The boundary of the permitted jointly-attained powers is a result of the structural reasoning explained in this section.

Accommodating multiple binding sites for a transcription factor

The general response function for the hypercubic graph, derived in Eq. 5.6 and Eq. 5.8, accommodates equally well the cases where all transcription factors bound to the promoter are molecularly distinct, and the cases where there are multiple copies of a given transcription factor bound to the promoter. For instance, if one transcription factor X_j has two binding sites (and thus can be found at a copy number of zero, one, or two copies on

the promoter), against a backdrop of m other transcription factors with single binding sites, then the response function is given by Eq. 5.6 and Eq. 5.8 for a $(m+2)$-dimensional hypercube, except that two concentrations appearing in these response formulas have been set to the same value $[X_j]$.

5.4.2 Common graphs formed by subsets of the hypercube

N strictly exclusive transcription factors: the book/rolodex graph

Often, the binding sites for transcription factors overlap. For instance, four promoters with appreciably-overlapping binding sites in *E. coli* are visualized in Fig. 5.23. These common overlaps suggest that the presence of transcription factors on a promoter might exclude others from also being bound.

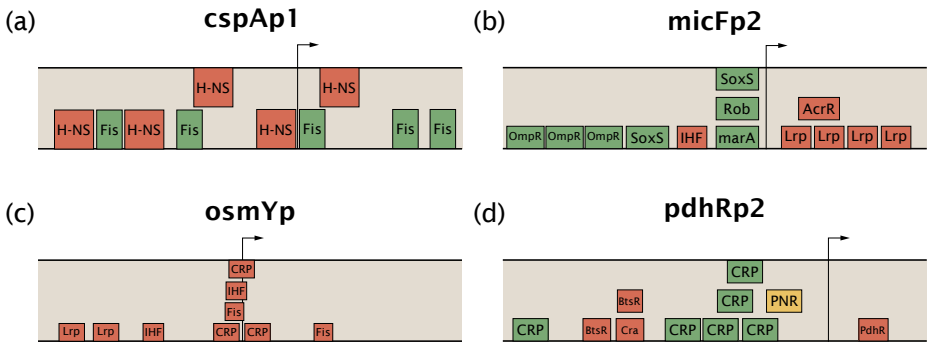


Figure 5.23: Four examples showing how *E. coli* promoters regulated by multiple transcription factors often present extensive overlaps between binding sites. (A) Regulation of the promoter cspAp1. (B) Regulation of the promoter micFp2. (C) Regulation of the promoter osmYp. (D) Regulation of the promoter pdhRp2. Binding sites are annotated by their name and colored in green, red, or yellow if the sites are understood to be activating, repressing, or both, respectively. The transcription starting site is indicated by the black arrow. (Note that these schematics reflect the topology/overlaps and relative positions of binding sites, but the size of each binding site's box is not to scale in genomic coordinates.)

If some transcription factors are mutually exclusive with each other and/or the polymerase, the hypercubic graph is pruned to a smaller state space, significantly changing the observed response functional form of the response with respect to the control variables.

Consider the highly biologically plausible case where two transcription factors X_1 and X_2 cannot coexist with each other, but each can coexist with the polymerase. The trees that span the resulting graph have particularly simple structure and give correspondingly simple forms for the ultimate response function of the system.

fig 4a from Lammers, Flamholz, Garcia 2023
https://doi.org/10.1073/pnas.2211203120

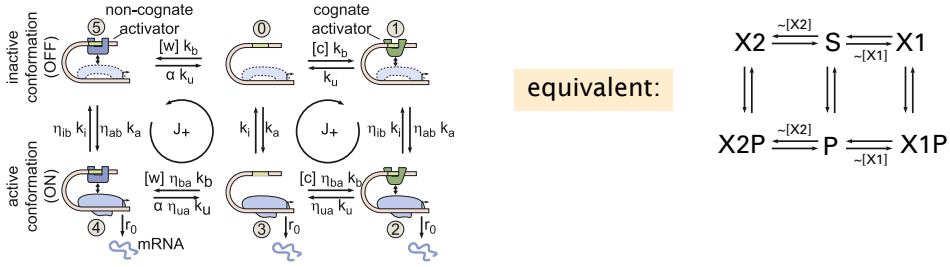


Figure 5.24: Regulatory motif with two transcription factors, whose binding is mutually exclusive, from Reference [11]. When only one transcription factor can bind at a time, but both can bind in the presence of the polymerase, a smaller state space graph is formed. One biological scenario yielding this simple exclusion structure is the setting studied by Reference [11], where both cognate and noncocagnate enhancers modulate the action of polymerase in cellular decision making.

For instance, consider the biological example invoked by Fig 4A of Lammers, Flamholz and Garcia [11], which they use to model interference by noncognate factors in cellular decision making. The state space collapses from a three-dimensional cubic graph (with eight conjoined cycles) to just two conjoined square cycles, as shown in Fig. 5.24.

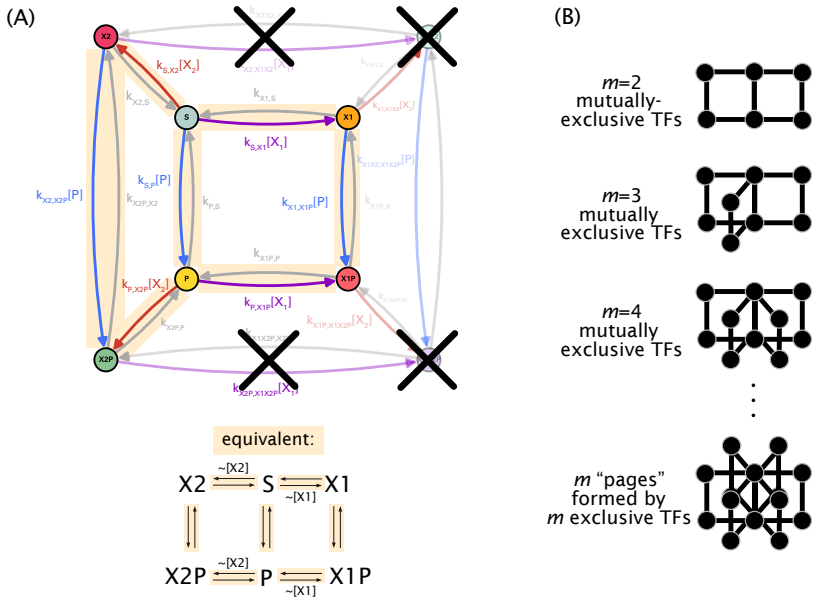


Figure 5.25: Relation between the cubic graph and the "book graph" and a generalisation of this "book graph" in m dimensions corresponding to the binding of m exclusive transcription factors. (A) Schematic highlighting the set of states (outlined in tan) of two mutually-exclusive transcription factors found inside a full hypercube graph (subject to no exclusions). Forbidden interactions and states are crossed out in black. (B) When multiple transcription factors govern a network and all are mutually exclusive (except with the polymerase), the resulting graph is always a set of m square cycle "pages," each associated with a given transcription factor, conjoined along one central "spine"/edge-pair between the empty S and polymerase-bound P states.

Figure 5.25 localizes the resulting states of this example as a subgraph of the hypercube graph we just investigated in §5.4. This graph also achieves analogous structures in higher dimension (with a larger number of transcription factors, still mutually exclusive except for the polymerase): extra "pages" are added around the "axis" of the $S \leftrightarrow P$ transition, creating a structure rather like a cylindrical rolodex. Each of the independent square pages is a cycle between the empty state S, the polymerase-bound state P, plus the states X_j and X_jP where a particular transcription factor X_j is bound alone or bound along with the polymerase, respectively; the exclusions among transcription factors enforce that these cycles are joined only along the central axis. It transpires that some literature refers to this general simple graph structure as a "book graph" (see e.g. Wolfram MathWorld).

Another important, biologically-distinct example where this regulatory architecture manifests is the setting of post-translational modifications, such as phosphorylation, made to a standing pool of transcription factors. For instance, if both unphosphorylated and phosphorylated forms of a transcription factor bind (perhaps with distinct affinity) to the same site on the genome, then these species are effectively mutually-exclusive distinct transcription factors, and will produce a book graph like those visualized in Figure 5.25B.

General response function of the book graph from multiple exclusive transcription factors

Having understood that multiple transcription factors form a book graph when they are mutually exclusive, we can deduce the form of any resulting response function. The crucial point is that the states S and P are the only states in the graph which have any outgoing edges dependent on transcription factor concentrations. We conclude that in any spanning tree, at most one outgoing edge can depend on a transcription factor (say X_i) outgoing from S, and at most one edge can depend on a transcription factor (say X_j, possibly with $i = j$) from P. This structural constraint immediately requires that any tree (hence any response function)

can only attain exactly the following dependencies on transcription factor concentrations: $\sim [X_i]^2, \sim [X_j]^2, \sim [X_i][X_j], \sim [X_i]$, or $\sim [X_j]$, for all pairs of distinct transcription factors (i, j). Accordingly, any response function of a book graph always exhibits the remarkably simple form[2],

$$\boxed{\langle r \rangle^{\text{book}} = \frac{a_0 + \sum_{i=1}^{m} (a_{i,1}[X_i] + a_{i,2}[X_i]^2) + \sum_{i \neq j \leq m} b_{ij}[X_i][X_j]}{c_0 + \sum_{i=1}^{m} (c_{i,1}[X_i] + c_{i,2}[X_i]^2) + \sum_{i \neq j \leq m} d_{ij}[X_i][X_j]}}. \tag{5.10}$$

Importantly, observe that if only one of the transcription factor concentrations $[X_i]$ is taken to vary (namely, regarding the function above from a hyperplanar slice with all but one transcription factors fixed), then this form collapses all the way down to a ratio of quadratic polynomials in this concentration, exactly the mathematical form studied extensively in preceding chapter.

At equilibrium, detailed balance enforces that every ratio of statistical weights has at most one power of a transcription factor's concentration. This collapse can be understood by the same means as the collapses found for the hypercube, or DNA looping graphs, analyzed earlier: focusing on the possible paths between any pair of states, we see that at most one

[2] In this work, we have usually absorbed the response function's dependence on polymerase concentrations into coefficients since we regard polymerase as a completely separate (and plausibly much more constrained) biological control variable. However, here it is structurally amusing to notice that the $[X_i]^2$-dependent terms of the response function present all possible powers of $[P]$ from $[P]^0$ (constant) to $[P]^m$. Why? Each tree delivering a quadratic dependence on some transcription factor $\sim [X_i]^2$ cannot accumulate any further dependencies on other transcription factor concentrations, nor can the $S \to P$ edge participate in the tree (since both S and P have "spent" their possible outgoing edges exactly to achieve the $\sim [X_i]^2$ dependence). However, every page of the book graph formed by the transcription factors can easily either include a $[P]$-dependent edge (going from $X_j \to X_j P$) or not (then involving the unbinding edge $X_j P \to P$), and still span the graph's states. In addition, if we relax the requirement that $[X_i]^2$ appear, a separate $[P]^{m+1}$ dependency can be attained by taking all $X_j \to X_j P$ edges with the $S \to P$ (all downward edges) and including up to one $[X_i]$-dependent edge, where the tree is completed by all other radial edges being oriented inwards as unbinding edges, $X_j P \to P$. We propose these structural considerations could be useful for other scenarios where an analog of the book graph appears and the variable P, with which all other transcription factor control variables are compatible, is perhaps instead understood as itself a separate transcription factor.

power of any given transcription factor's concentration $[X_i]$ is accumulated along any simple path between any two states in the graph. Thus any response of a book graph is a linear rational function of transcription factor concentrations at equilibrium, such that Eq. 5.10 collapses to,

$$\langle r \rangle^{\text{book, eq.}} = \frac{a_0 + \sum_{i=1}^{m} a_i [X_i]}{b_0 + \sum_{i=1}^{m} b_i [X_i]}. \tag{5.11}$$

5.4.3 All nontrivial, transcriptionally-potent subgraphs of the $n = 4$ hypercube

While we just analyzed the limiting case where all transcription factors are mutually exclusive, more relaxed and complex exclusion topologies (where only some transcription factors or polymerase are incompatible with each other on the genome) are also plausible. In this section, we systematically find the architectures (and associated response functions) of all possible networks with up to three transcription factors. We find these networks by beginning with the tesseract ($n = 3 + 1$ dimensional hypercube) graph and removing all states formed by possible individual, pairwise, triplet, or quadruplet mutual exclusions among the three transcription factors and polymerase. (For instance, the book graph is generated by forbidding states where any pair of transcription factors are present, here specifically excluding states X_1X_2, X_2X_3, and X_1X_3 and their accompanying transitions.) Further, we require that the resulting graphs remain transcriptionally potent (e.g. retain at least one state where polymerase can be bound to the genome) and discard redundant (e.g. literally isomorphic) networks generated during this enumeration. Figs. 5.26 and 5.27 summarize the results of this exercise.

This exercise reveals that a large fraction of networks constrained by exclusions show simple mathematical dependencies in at least one of their transcription factors.

Specifically, the majority (90 out of the total nontrivial 147 networks) show at most a quadratic dependence in at least one transcription factor; such networks are shaded in purple in Figs. 5.26 and 5.27. This quadratic dependence is the same as that analyzed thoroughly in the preceding chapter; when the simply-varying transcription factor is the only transcription factor that varies, the resulting response function is subject to all the precise and quantitative mathematical constraints born by the square graph. The result of this exercise is also summarized in Figure 2B of the preceding chapter.

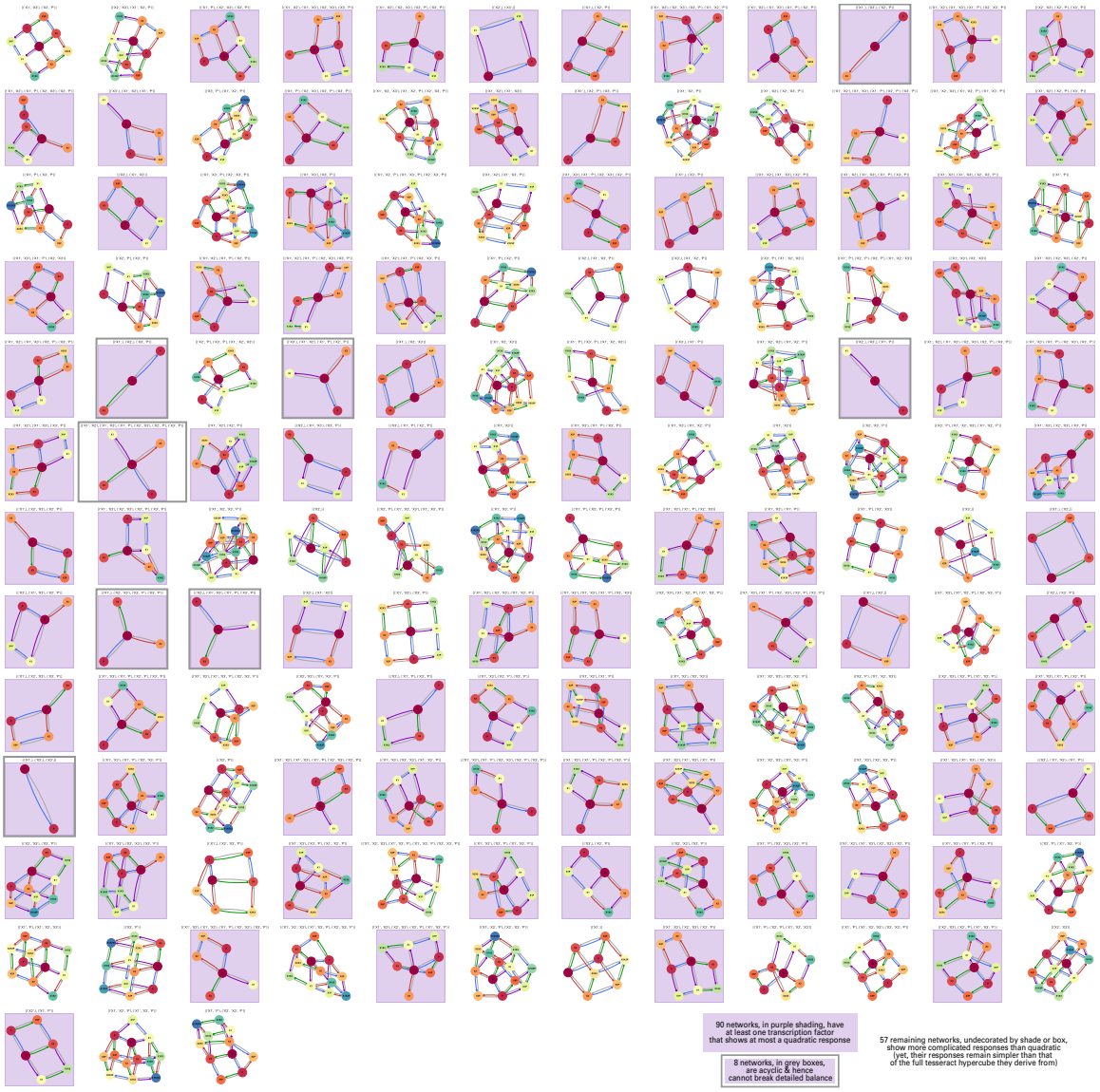


Figure 5.26: Comprehensive census of all subsets of the tesseract cube formed by all possible pairwise, triplet, or individual exclusions among the transcription factors and polymerase, restricted to the set where polymerase can still be present. Each panel depicts the state topology of a distinct possible set of mutual exclusions between transcription factors (among themselves) and/or polymerase, as long as the polymerase can be found bound to the genome in at least one state. (The ordering of panels follows an arbitrary enumeration of while maintaining that the polymerase can bind in at least one state.) Purple shading emphasizes those networks that display at least one transcription factor whose response varies as a quadratic response or simpler (namely, matching the mathematical response of a square graph).

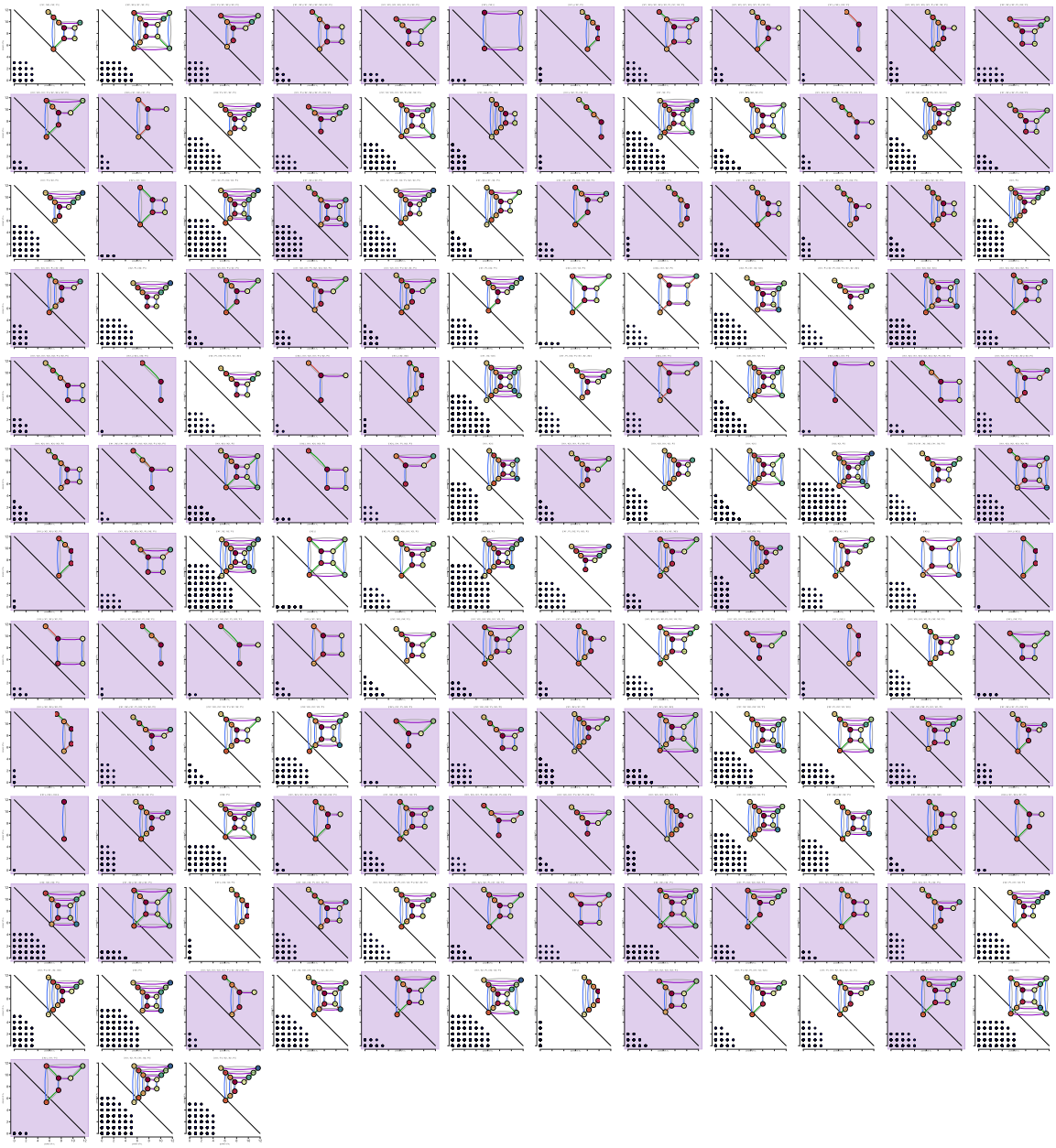


Figure 5.27: The same three-transcription-factor-networks identified in Fig. 5.26, now visualized alongside their attained powers in two of their transcription factor concentrations, $[X_1], [X_2]$. The networks (and their ordering) are the same as those in the earlier Fig. 5.26; the nodes are just arranged according to a different layout to emphasize their nature as a subgraph of the tesseract graph, whose full set of nodes is shown in Fig. 5.20B.

Bibliography

[1] Sara D. Mahdavi, Gabriel L. Salmon, Patill Daghlian, Hernan G. Garcia, and Rob Phillips. Flexibility and sensitivity in gene regulation out of equilibrium. *Proceedings of the National Academy of Sciences*, 121(46):e2411395121, 2024.

[2] Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Terence Hwa, Jané Kondev, and Rob Phillips. Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.*, 15(2):116–124, 2005.

[3] Víctor H Tierrafría, Claire Rioualen, Heladia Salgado, Paloma Lara, Socorro Gama-Castro, Patrick Lally, Laura Gómez-Romero, Pablo Peña-Loredo, Andrés G López-Almazo, Gabriel Alarcón-Carranza, et al. Regulondb 11.0: Comprehensive high-throughput datasets on transcriptional regulation in *Escherichia coli* K-12. *Microb. Genom.*, 8(5), 2022.

[4] Mark EJ Newman. Mixing patterns in networks. *Physical review E*, 67(2):026126, 2003.

[5] William T Ireland, Suzannah M Beeler, Emanuel Flores-Bautista, Nicholas S McCarty, Tom Röschinger, Nathan M Belliveau, Michael J Sweredoski, Annie Moradian, Justin B Kinney, and Rob Phillips. Deciphering the regulatory genome of *Escherichia coli*, one hundred promoters at a time. *eLife*, 9:e55308, 2020.

[6] Mattias Rydenfelt, Hernan G Garcia, Robert Sidney Cox III, and Rob Phillips. The influence of promoter architectures and regulatory motifs on gene expression in *Escherichia coli*. *PLoS One*, 9(12):e114347, 2014.

[7] Muir Morrison, Manuel Razo-Mejia, and Rob Phillips. Reconciling kinetic and thermodynamic models of bacterial transcription. *PLoS Comput. Biol.*, 17(1):e1008572, 2021.

[8] Soile VE Keränen, Angel Villahoz-Baleta, Andrew E Bruno, and Marc S Halfon. REDfly: an integrated knowledgebase for insect regulatory genomics. *Insects*, 13(7):618, 2022.

[9] Olivier Bernardi. On the spanning trees of the hypercube and other products of graphs. *arXiv preprint arXiv:1207.0896*, 2012.

[10] Richard Stanley. *Enumerative Combinatorics: Volume 2*. Cambridge University Press, 1999.

[11] Nicholas C Lammers, Avi I Flamholz, and Hernan G Garcia. Competing constraints shape the nonequilibrium limits of cellular decision-making. *Proc. Natl. Acad. Sci. U.S.A.*, 120(10): e2211203120, 2023.

Part III

Patterns out of equilibrium

Chapter 6

Accounting for the fundamental dissipative costs of gradient formations across space and time in motor-microtubule assemblies

This chapter describes joint work with Ana Duarte, Rob Phillips, and others in the Phillips lab. In this project in progress, GS has formulated theory, copiloted the data analysis pipelines, and cowritten the manuscript.

So far in our journey through this thesis, the values of measured energy expenditures we have discussed have not been explicitly space-dependent. Our empirical coordinates have been whole-cell basal metabolic rates, or the nonequilibrium driving force operating in a mean field gene-protein binding network understood as well mixed. Thankfully, however, biology is profoundly nonuniform. These patterns are a flavor of lifeiness that surely costs energy to durably sustain. Over the past years, colleagues in the Phillips laboratory (especially Ana Duarte, HJ Lee, and many others) have carefully built and ushered into use a fluoresent probe that can report on local concentrations of ATP in space in controllable

in vitro assays. Accordingly, these measurements afford the means to measure power density as a field variable that explicitly and meaningfully depends on space and time. Over the past years we have worked with these intrepid colleagues to make more robust the local inference of ATP from noisy biological data. Simultaneously we have enjoyed scaling up minimal but, to our knowledge, fresh theory that makes strong predictions for where and when power should be maximal, and from what specific physical origins power demands likely stem.

In this first of two chapters, we proceed through our draft findings of the quantitative phenomenology of gradients of ATP resolved while molecular motors conspire collectively to form patterned arrangements of microtubules. In the chapter that follows (Chapter 7), we spelunk through nascent theory that, one day, should converse with experiments.

6.1 Overview

Living matter produces a variety of beautiful spatiotemporal structures and patterns that are not present in their nonliving counterparts. Often, these ordered, non-equilibrium steady states are sustained through an influx of energy. Careful examination of when and where living systems direct energy consumption helps us understand the principles dictating such ordering and can motivate development of non-equilibrium theories. Here, we investigate the energetic cost of assembling an ordered aster from a disordered, uniform mixture of cytoskeletal microtubules and kinesin motors. Using a fluorescent ATP sensor, we perform a careful measurement of ATP consumption through space and time on an in vitro cytoskeletal network. Our experiments resolve the emergence of radial ATP gradients. We successfully predict how a given motor profile generates these ATP distributions using reaction-diffusion models in conjunction with finite element simulations. With our results, we compare the power per volume required by our

cytoskeletal networks with the known power per volume expenditure in cells. Evoking the direct quantification of energetic fluxes from our spatial measurements, unlocks future exploration of what steady states are accessible to cells and how the cytoskeleton drives spatial organization.

6.2 Introduction

One of the main drivers maintaining the rich patterns in living matter is a steady investment of energy. For example, the existence of morphogen gradients in developmental patterning is paid for through a steady flux of protein synthesis and degradation. Similarly, cytoskeletal-motor systems hydrolyze ATP and GTP to achieve processes ranging from intracellular transport, to cell motility, to chromosome segregation. Motivated by these processes, we were inspired to develop a physical understanding of the connection between energy fluxes and the emergence of biological order in space and time in the context of the particular example of microtubule-motor assemblies.

The energetic basis of the processes within living cells are based upon a few fundamental energy currencies, which can be thought of as a sort of biological batteries. This metaphor is useful because it reminds us that batteries are indifferent to the particulars of what they are wired up to - they can drive anything from the light in a flashlight to motorized toys. Biological processes are powered by several key biological batteries including membrane potentials, redox reactions and trinucleotide hydrolysis. Indeed, for the molecular motor driven reactions that power the structures of the cytoskeleton, ATP and GTP hydrolysis are central. Thus, we were curious about how ATP consumption in space and time drives the dynamics of structure formation.

Recently, it has become possible to measure the total energy consumption of both living

organisms and the molecular machinery that drive them [1, 2, 3, 4]. We aimed to complement these foundational studies by investigating the spatial distribution of energy consumption. In particular, here we report the visualization of spatial, ATP concentration gradients across cytoskeletal networks, giving insight into how structure, composition and morphology drive energy dissipation.

We use motor-microtubule assemblies as a highly controllable, tunable system due to their minimal components and self-organizing properties in vitro as well as their biological ubiquity. An abundance of work has established that connected dimeric motor domains can cross-link microtubules, creating ordered networks [5, 6, 7, 8]. To control the position, size and start of microtubule cross-linking, we optogenetically link motor proteins together as shown in Figure 6.1, as previously developed in our labs [8]. Motors harness energy to drag microtubules into ordered structures by hydrolyzing an ATP molecule for each step they take along the microtubule. We measure the energy consumed by the motors throughout space and time using a fluorescent, QUEEN based ATP probe [9] as shown in Figure 6.1.

In addition, to complement the measurements and to provide a framework for understanding them, we develop a reaction-diffusion equation that describes the emergent ATP spatiotemporal gradients and explore the implications of that model with finite element calculations. Combining our experimental and theoretical results, we determine that the measured network formation power is indeed many orders of magnitude greater than the theoretical power of equilibrium processes, and we demonstrate the novel value of spatial measurements by bounding energy dissipation with Jensen's inequality.

We note that the work presented in the remainder of the paper is written with the intention of providing the essential narrative structure in the paper itself. In parallel, the Supplementary Material is written with the aim of making every technical detail of how the experiments were done, how the data was analyzed, how conclusions were drawn and how

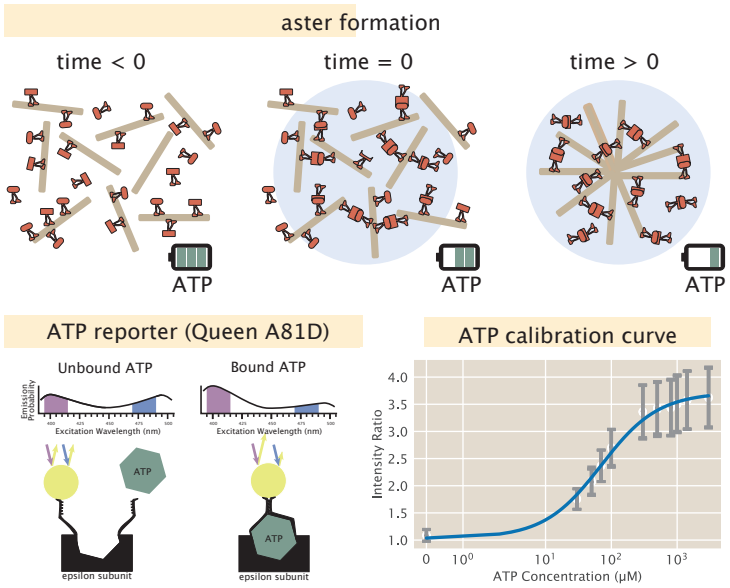


Figure 6.1: Schematic of the experimental system used to measure spatiotemporal evolution of ATP. The top panel illustrates the formation of an aster using light activated motor dimerization. Before light activation, motors independently walk on microtubules hydrolyzing ATP. At $t = 0$ a circular light pattern is projected onto the sample. Motor proteins inside the illuminated region dimerize, crosslinking microtubules. As time elapses, microtubules are dragged into an aster resulting in the depletion of ATP. The lower left panel illustrates the binding mechanism for ATP to the ATP probe. Binding ATP to the probe causes the number of emission counts due to an excitation of 405 nm light to increase, while emission counts from 480 nm light excitation decreases. By comparing the ratio of the emission counts at 405 nm and 480 nm light excitations, the concentration of ATP can be inferred. The lower right panel is a calibration curve mapping known ATP concentrations to fluorescent light intensity ratios. Each black circle represents the mean ratio value for a given image and gray error bars report the standard deviation of the image.

the data was interpreted using quantitative models completely and rigorously transparent.

6.3 ATP Concentrations in Space and Time

The key elements of our experimental design are shown in Figure 6.1. As noted above, using spatially and temporally controlled illumination, we can generate patterns such as the radially symmetric aster shown in the schematic. Our principal experimental goal is to measure the rate of consumption of ATP as a function of position and time, a goal that is realized by using the fluorescent, ratiometric ATP reporter [9] depicted in Figure 6.1(B). The probe mechanism creates a change in the protonation state [10] of the fluorophore when ATP binds [11], triggering a shift in the fluorophore's absorption spectrum [12]. By using known standards, as shown in Figure 6.1(C), we can construct a calibration curve that now permits us to measure the ATP concentration in a given spatial region. Given that the characteristic scale of ATP concentrations in our experiments are of order 100 mM, we see that our ATP probe can measure ATP consumption in a spatially resolved fashion since, as seen in the calibration curve, the ATP reporter is sensitive in precisely the concentration ranges over which the reporter is effective. Note that in the Supplemental Information, we give an extremely detailed description of how we handled uneven illumination and photobleaching, a prerequisite to a calibration curve like that shown in Figure 6.1(C).

With this measurement scheme in hand, we were able to characterize the quantity of ATP in space and time during aster formation, like that shown in Figure 6.2. Note that since we measure the ATP concentration from one time step to the next, this allows us to characterize ATP dynamics, namely ATP movements due to diffusion and depletion of the molecules as they are hydrolyzed by molecular motors. In particular, as shown in Figure 6.2(A), we monitor the amount of ATP as a function of distance from the aster center over a period of

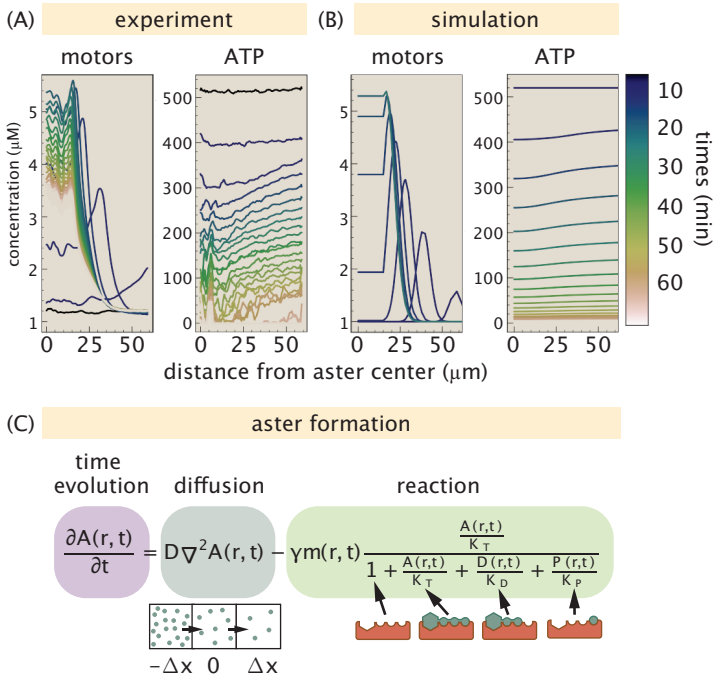


Figure 6.2: Experimental measurements and Finite Element simulations of ATP concentration in space and time. (A) The radial concentration profile of motors and ATP as measured experimentally. (B) The radial concentration profile of motors and ATP as computed using finite element calculations of a continuum reaction-diffusion equation. As time evolves, motor proteins become concentrated near the aster center while ATP depletes. Both the experiment and simulation find emergence of an ATP gradient with the greatest depletion in the aster center, where motors are most concentrated. (C) The reaction-diffusion equation used to simulate ATP concentration is written with illustrations of the diffusive term and the binding states of ATP (A), ADP (D) and phosphate (P) to the motor protein schematized in red.

more than an hour, revealing not only the development of a spatial gradient, but also a net overall consumption of ATP.

Another way of visualizing the results of our measurements is to take the ATP data from successive instants and convert it to a power. As shown in Figure 6.3, the power can be evaluated directly in units of ATP/s, making it possible to compare directly to other measurements, as well as estimates of what the energy from hydrolysis events is used to pay for. Next, we ask how one might develop quantitative intuition for the distribution of the ATP in the aster in both space and time.

The profiles revealed in Figure 6.2 characterize the radial and temporal dependence of

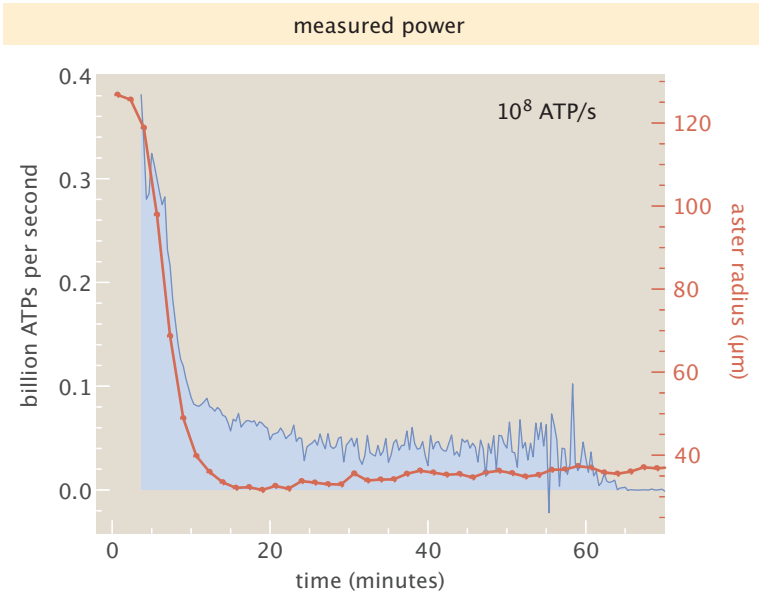


Figure 6.3: The power of aster formation. The number of ATPs consumed per second is plotted as a function of time in light blue. The red line on top reports the aster radius over time. The magnitude of power follows the size of the aster, consuming ATP faster while contracting and reaching a baseline power level once the aster is no longer changing size.

both the motors and the ATP. As seen in the dynamical equation in Figure 6.2(C), the change of concentration in some small material volume element can be attributed both to ATP molecules entering and leaving that small region and to the hydrolysis of those ATP molecules by molecular motors that are in the material volume element of interest. More precisely, as shown in the figure, we can write a reaction-diffusion equation that captures the rate of change of ATP in a material volume element. Note that there is another more pernicious dynamics taking place during our experiments, namely, photobleaching. (We have several sections in Supplemental Information appendices that describe how we measure and account for photobleaching, including finite element simulations of a diffusion-photobleaching equation.) As shown by the box underneath the equation in Figure 6.2(C), a material volume element of interest has molecules both leaving and entering as a result of diffusion, as described in more detail in the Supplement. The reaction term uses a Michaelis-Menten-like dynamics for the rate of ATP consumption by

motors. However, the denominator of that term also includes terms that reflect competitive inhibition of the reaction due to ADP and P_i. The local hydrolysis rate is proportional to the density of motors and to the concentration of ATP itself.

One convenient way to analyze the solutions to such equations in diverse geometries is by appealing to numerical methods. In our case, we used the finite element method to compute the space-time history of the ATP as shown in Figure 6.2(B). As seen in the figure, for reasonable parameter choices (the details of parameter selection are described in the Supplement), the time evolution of the ATP field is in reasonable accord with our measurements. One question that naturally arises is to ask what that ATP hydrolysis "pays for?" Of course, mechanistically, we know that the ATPs are being consumed by motors as they carry out their walk along microtubules. But here we mean it differently. That energy is dissipated through elementary processes such as ordering and contraction of the microtubule network. How much energy do these processes cost?

Microscopically, a huge variety of dissipative processes are taking place during the microtubule-motor rearrangements attending aster formation. As shown in Figure 6.3 the power varies considerably at different stages of the aster formation process. To that end, we explore the power associated with a variety of processes that we imagine are taking place concurrently and would cost different amounts at different stages of the aster formation process. Indeed, as seen in Figure 6.4 there are a number of possible estimates that one can make in units of ATP/s that reflect what ATP hydrolysis might be used to "pay for." Each of these estimates (and several others) is spelled out in detail in Supplemental Information appendices; here we simply describe the concept of each calculation and how much power consumption it implies.

Given the measured power, we were intrigued to compare that power to the powers associated with several elementary dissipative processes shown in Figure 6.4. For example, as is evident from the radius as a function of time in Figure 6.3, the volume of the

microtubule aster is decreasing over time. As shown in Figure 6.4(A) and described in detail in elsewhere in appendices, we can perform a simple estimate of the power associated with this contraction as the pV work done divided by the elapsed time. We find that the pV power is five orders of magnitude smaller than the measured power. As shown in Figure 6.4(B), a second dissipative process is the frictional sliding of the various microtubules during the contraction process. A naive estimate is obtained by replacing a given microtubule by a corresponding sphere of the same dimensions and to work out the Stokes drag. As in the case of the pV power, this results in a power that is five orders of magnitude smaller than the measured power of Figure 6.3. A better estimate can be made in which the microtubule is treated as a rod rather than a sphere and in this case the computed power is even smaller. Another approach to estimating the power is offered by field theories of nematic ordering in which the state of the system is characterized by the spatially varying tensor $Q_{ij}(\mathbf{f}, t)$. This estimate is trickier to make since the parameters in such a field theory of motor-microtubule systems are not well known. Nevertheless, as seen in Figure 6.4(C) (and described in more detail in the Supplement), the power we estimate associated with such ordering is many orders of magnitude smaller than the measured power. The final dissipative process highlighted in Figure 6.4(D) is that of maintaining a nonequilibrium gradient of motors radially outward from the center of the aster. One way to think about such a gradient is that if there were not some active transport carrying motors towards the aster center, then diffusion would smooth out that gradient. As we show elsewhere in appendices, there is a well defined prescription for estimating the power to maintain such a gradient and we find that it is of the same order of magnitude as the measured powers. This interesting result suggests the hypothesis that a significant fraction of the ATP hydrolysis consumed by the motors is "spent" to maintain this gradient.

It is practically a cliche to note that living organisms are "out of equilibrium." But we also find that this compact, binary statement, while true, is largely unhelpful. Equilibrium ideas

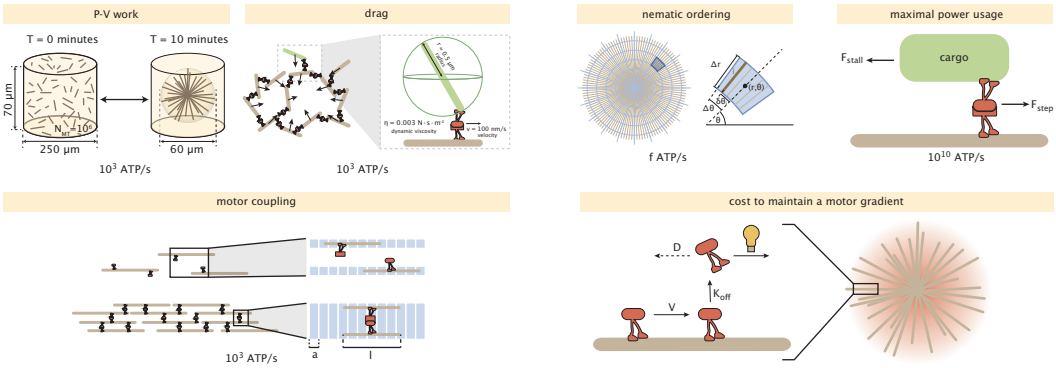


Figure 6.4: Mechanistic processes taking place during aster formation and their estimated power. Each schematic considers a different dissipative processes that occurs during the formation of an aster. (A) Schematic of the power of compressing an "ideal gas" of microtubules. (B) The power of dragging microtubules through a viscous medium. (C) Power estimate for inducing nematic ordering in a random array of microtubules. (D) The power to maintain a concentration gradient of motors in the aster.

are even used to describe the conditions in a star, and it behooves us to have quantitative measures of what we mean by the word "nonequilibrium" in different contexts. For example, although a bacterium can be thought of as a furnace burning fuel at a rate of 10^5 W/m^3, for single cells and even embryos, this power consumption leads to a temperature change relative to the external world of less than 10^{-7} K. Though each ATP hydrolysis event leads to a local temperature change of ≈ 10 K in a 1 nm^3 volume, that temperature change relaxes away at time scales far less than a microsecond. Similarly, the concentration gradients of morphogens are very shallow implying that the "out of equilibrium" concentration gradient is barely discernible at the molecular scale. To see this, we imagine discretizing the anterior-posterior axis of the fly embryo into little 1 μm^3 boxes, the measured gradients are so gentle that two adjacent 1 μm^3 boxes near the anterior part of the embryo would have an average of 60 and 59 molecules, respectively. The point here is that the question of the extent to which biological systems depart from equilibrium is both important and subtle.

In recent years, renewed efforts have been made both experimentally and theoretically to explore the deep question of when and where energy is being consumed within cells and what the energy consumption is used for. A seminal study in the 1970s by Stouthammer

already recognized the huge energetic burden of protein synthesis in fast growing cells. [13] This process sets a baseline of roughly 10^6ATP/μm^2 s because to produce the $\approx$ few $\times 10^2$ peptide bonds per protein each of which costs 4 ATP equivalents for the roughly few $\times 10^6$ proteins per cell results in a total cost of $\approx$ few $\times 10^9$ ATPs over the 1000 seconds of the cell cycle. Recent experiments in bacteria illustrated how in starved cells after roughly 24 hours they die due to a loss of membrane potential. Here too, we can make simple estimates of the power required to maintain the transmembrane potential against membrane leakage. Stated simply, the flux of ions out of the cell can be estimated as $dn/dt = JA$, where $J = p\Delta c$. Using the very low permeability to ions of roughly $p \approx 10^{-9}$ nm/s, we find that the net loss of ions occurs at a rate of 1 - 10 ions/s. Using the relation $\Delta Q = C\Delta V$, we can estimate that in under 24 hours the membrane potential will be nearly completely lost. The power to maintain the membrane potential against such leakage is a tiny value of 10s of ATPs per second.

We argue that the management of energy expenditure in space and time is of critical importance for understanding everything from the development of membrane potentials to the fidelity of biological polymerization to the assembly of structures such as the spindle. Recently, both calorimetry and fluorescence approaches have made it possible to dissect the dissipation in living cells. Such measurements generally tell a similar story, namely, that the scale of measured dissipation is often orders of magnitude larger than the dissipation one might expect by considering elementary processes such as flow. In this paper, we aimed to rigorously measure such energy consumption in space and time in the specific context of a microtubule-motor system.

Bibliography

[1] Peter J Foster, Jinhye Bae, Bezia Lemma, Juanjuan Zheng, William Ireland, Pooja Chandrakar, Rémi Boros, Zvonimir Dogic, Daniel J Needleman, and Joost J Vlassak. Dissipation and energy propagation across scales in an active cytoskeletal material. *Proceedings of the National Academy of Sciences*, 120(14):e2207662120, 2023.

[2] Jonathan Rodenfels, Karla M Neugebauer, and Jonathon Howard. Heat oscillations driven by the embryonic cell cycle reveal the energetic costs of signaling. *Developmental cell*, 48(5):646–658, 2019.

[3] Qiwei Yu, Dongliang Zhang, and Yuhai Tu. Inverse power law scaling of energy dissipation rate in nonequilibrium reaction networks. *Physical review letters*, 126(8):080601, 2021.

[4] Yuansheng Cao, Hongli Wang, Qi Ouyang, and Yuhai Tu. The free-energy cost of accurate biochemical oscillations. *Nature physics*, 11(9):772–778, 2015.

[5] FJ Nédélec, Thomas Surrey, Anthony C Maggs, and Stanislas Leibler. Self-organization of microtubules and motors. *Nature*, 389(6648):305–308, 1997.

[6] Thomas Surrey, François Nédélec, Stanislas Leibler, and Eric Karsenti. Physical properties determining self-organization of motors and microtubules. *Science*, 292(5519):1167–1171, 2001.

[7] Johanna Roostalu, Jamie Rickman, Claire Thomas, François Nédélec, and Thomas Surrey. Determinants of polar versus nematic organization in networks of dynamic microtubules and mitotic motors. *Cell*, 175(3):796–808, 2018.

[8] Tyler D Ross, Heun Jin Lee, Zijie Qu, Rachel A Banks, Rob Phillips, and Matt Thomson. Controlling organization and forces in active matter through optically defined boundaries. *Nature*, 572(7768):224–229, 2019.

[9] Hideyuki Yaginuma and Yasushi Okada. Live cell imaging of metabolic heterogeneity by quantitative fluorescent ATP indicator protein, QUEEN-37c. *BioRxiv*, pages 2021–10, 2021.

[10] Takeharu Nagai, Asako Sawano, Eun Sun Park, and Atsushi Miyawaki. Circularly permuted green fluorescent proteins engineered to sense ca2+. *Proceedings of the National Academy of Sciences*, 98(6):3197–3202, 2001.

[11] Hideyuki Yaginuma, Shinnosuke Kawai, Kazuhito V Tabata, Keisuke Tomiyama, Akira Kakizuka, Tamiki Komatsuzaki, Hiroyuki Noji, and Hiromi Imamura. Diversity in atp concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. *Scientific reports*, 4(1):6522, 2014.

[12] Roger Y Tsien. The green fluorescent protein. *Annual review of biochemistry*, 67(1):509–544, 1998.

[13] A. H. Stouthamer. Theoretical Study on Amount of Atp Required for Synthesis of Microbial Cell Material. *Antonie Van Leeuwenhoek Journal of Microbiology*, 39(3):545–565, 1973.

Chapter 7

Power as a localized field in space and time

Armed with the phenomena we measured in the preceding Chapter 6, we now develop simple field theory to anticipate how and where concentration profiles changing in time will cause local depletions of energy to vary in time and space.

7.0.1 Power to Maintain a Motor Gradient

As noted in the main body of the paper, the largest power estimate out of the suite of processes we considered is associated with the formation of the gradient of motors. Because of the gradient in motor density across the aster, it is of interest to estimate the free energy required to maintain that gradient. We begin by examining the free energy cost of a one-dimensional gradient to set notation and to explain the concept and follow that discussion by the case of interest involving spherical symmetry. The concept of the estimate is to compute the free energy change when we take one particle from a region with one concentration and put that particle in a nearby region with a slightly different concentration.

The left panel of Figure 7.1 illustrates the free energy change we need to evaluate. In

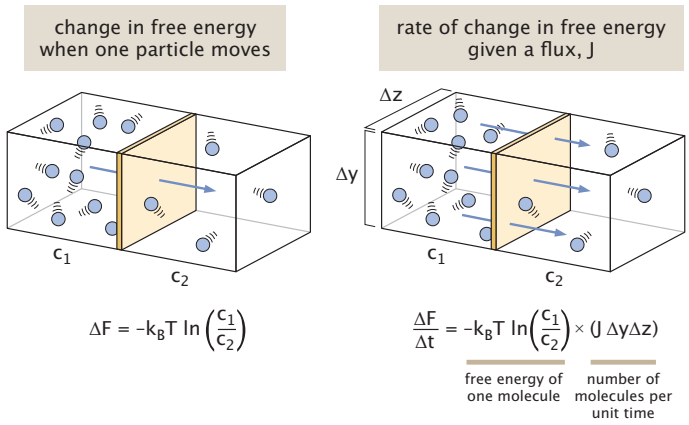


Figure 7.1: The change in free energy when a particle is transported in a gradient. The left panel shows the free energy change upon moving a particle from the left to the right.

particular, we compute the free energy of the system before and after allowing a single molecule to cross from one side of the partition to the other. The free energy is defined as

$$F = U - TS, \tag{7.1}$$

where U is either the energy or the enthalpy and S is the entropy. We ignore the subtleties associated with Helmholtz vs Gibbs free energies. We interest ourselves in the change in free energy

$$\Delta F = F_{final} - F_{initial} = (U_{final} - TS_{final}) - (U_{initial} - TS_{initial}). \tag{7.2}$$

Further, we simplify by ignoring any contribution due to energy (or enthalpy) changes (i.e. $\Delta U = 0$).

To compute the entropy change, we need to compute the entropy of the solutions on both sides of the partition, both before and after we have plucked a molecule from the left side and installed it on the right side. The total entropy is given by

$$S_{tot} = S_1 + S_2 \tag{7.3}$$

where the subscripts refer to the two compartments. This expression implies that the free energy in this state is given by

$$F = -TS_{tot}. \tag{7.4}$$

We can compute the entropy by counting the number of microscopic states both before and after the particle transfer. In particular, we have

$$S_{tot}^{(final)} = k_B \ln W_1^{(final)} + k_B \ln W_2^{(final)}, \tag{7.5}$$

with a similar expression for the entropy in the initial state. This can be simplified to the form

$$S_{tot}^{(final)} = k_B \ln (W_1^{(final)} W_2^{(final)}), \tag{7.6}$$

which makes sense given that the total number of microscopic states is equal to the product of the number of states in the first box and the number of states in the second box.

In a lattice model, we make the abstraction that space is subdivided into tiny lattice sites with a characteristic dimension of 1 nm^3 (i.e. molecular sizes). To compute the number of microstates W, we count the number of ways of arranging our L_i motors among the Ω lattice sites as

$$W_i(L_i) = \frac{\Omega!}{L_i!\,(\Omega - L_i)!},. \tag{7.7}$$

When $\Omega \gg L_i$ (i.e. the dilute limit), we can make the much simpler approximation

$$W_i(L) = \frac{\Omega^{L_i}}{L_i!}, \tag{7.8}$$

which amounts to the idea that every motor can sit on any of the Ω lattice sites.

We can now write the change in free energy which is strictly entropic as

$$\Delta F = -k_B T \left(\ln \frac{\Omega^{L_1+1}}{(L_1+1)!} \frac{\Omega^{L_2-1}}{(L_2-1)!} - \ln \frac{\Omega^{L_1}}{L_1!} \frac{\Omega^{L_2}}{L_2!} \right). \tag{7.9}$$

This can be simplified to

$$\Delta F = -k_B T \ln \frac{L_1!}{(L_1+1)!} \frac{L_2!}{(L_2-1)!} \approx -k_B T \ln \frac{L_2}{L_1}. \tag{7.10}$$

We can rewrite this in a more familiar form using the language of concentrations. If we multiply numerator and denominator within the logarithm by Ωv, where v is the volume of a single lattice site in our lattice model, then $\Omega v = V_{tot}$ and hence $c_1 = L_1/V_{tot}$ and $c_2 = L_2/V_{tot}$, permitting us to write

$$\Delta F = k_B T \ln \frac{c_2}{c_1}. \tag{7.11}$$

Now, as seen in the right panel of Figure 7.1 if we want to find the total rate of free energy change, we need to multiply the free energy per particle by the total number of particles transported between the two adjacent reservoirs per unit time using the flux resulting in

$$\left(\frac{\Delta F}{\Delta t} \right) = -\Delta \mu (J \Delta y \Delta z). \tag{7.12}$$

The factor $J \Delta y \Delta z$ counts the number of molecules carried down the gradient per unit time and the minus sign guarantees that if the left reservoir has more molecules than the left, then $\Delta F/\Delta t < 0$. We now interest ourselves in the case where the concentration is slowly varying, permitting us to write

$$\left(\frac{\Delta F}{\Delta t} \right) = k_B T \ln \frac{c(x+\Delta x)}{c(x)} \times (J \Delta y \Delta z), \tag{7.13}$$

where we have introduced the notation $c_1 = c(x)$ and $c_2 = c(x + \Delta x)$. Note now we switched the sign because we inverted the ratio in the logarithm with the concentration on the right now appearing in the numerator. Note that this expression is valid regardless of whether $c_1 > c_2$ or $c_2 > c_1$ since in those two cases the flux is in opposite directions and our expression reflects that. By invoking Fick's law we now have

$$\left(\frac{\Delta F}{\Delta t}\right) = k_B T \ln \frac{c(x + \Delta x)}{c(x)} \times \left(-D\frac{dc}{dx}\right) \Delta y \Delta z. \tag{7.14}$$

By carrying out the Taylor expansion $c(x + \Delta x) = c(x) + (\partial c/\partial x)\Delta x$, we can rewrite the logarithmic term as

$$\begin{aligned} \ln \frac{c(x + \Delta x)}{c(x)} &= \ln \left[\frac{c(x) + \frac{\partial c}{\partial x}\Delta x}{c(x)}\right] \\ &= \ln \left[1 + \frac{1}{c}\frac{\partial c}{\partial x}\Delta x\right]. \end{aligned} \tag{7.15}$$

Finally, we invoke a second Taylor series in the form $\ln(1 + \epsilon) \approx \epsilon$, resulting in the simple and powerful expression

$$\left(\frac{\Delta F}{\Delta t}\right) = -D k_B T \frac{1}{c(x)} \left(\frac{\partial c(x)}{\partial x}\right)^2 \Delta x \Delta y \Delta z. \tag{7.16}$$

This result tells us the free energy loss for two adjacent planes if the gradient is allowed to dissipate. We can now put this all together to explore the free energy dissipated over a continuous concentration field. In this case, we take the expression seen in eqn. 7.16 and add up the contribution from each set of planes in our discrete representation of the concentration field. We can work out the minimum power to maintain such a gradient where we imagine every time a molecule goes down its gradient, energy is consumed to push it back where it came from. Given this approach, we can now tackle the question of estimating the power associated with maintaining the radial motor gradient in our asters. However, to do so, we first need to reinterpret our one-dimensional analysis done here to

the case of a three-dimensional, but spherically symmetric concentration gradient.

$$\frac{\partial F}{\partial t} = D k_B T \int_0^\infty 4\pi r^2 dr \frac{1}{c(r)} \left(\frac{dc}{dr}\right)^2 \tag{7.17}$$

$$= D k_S T \left(\frac{c_0}{\lambda}\right)^2 4\pi \int_0^\infty \frac{r^2}{c_b e^{-r/\lambda}} e^{-2r/\lambda} dr \tag{7.18}$$

$$= 8\pi D k_B T_{c_D} \lambda \tag{7.19}$$

$$\frac{d}{dr}\left(\left.\frac{\partial F}{\partial t}\right|_r\right) = \left(2re^{-r/\lambda} - \frac{r^2}{\lambda} e^{-r/\lambda}\right) \text{ const.} \tag{7.20}$$

This result provides a very interesting and actionable insight. There is a competition between the fact that higher concentrations involve higher ATP consumption rates and the geometric effect that shells at larger radii have larger volumes. Together these two effects conspire to give rise to a particular radius at which the ATP consumption rate is highest.

7.0.2 Power Stored in a Three-Dimensional Gradient

Building on the analysis of the previous section which was performed in one-dimension, we now turn our attention to a three-dimensional structure with a spherically symmetric concentration gradient of motors of the form $c(r)$, where r is the radial distance from the aster center. Figure 7.2(A) shows the amendment that needs to be made to our one-dimensional analysis, where now in the three dimensional case, the "boxes" are spherical shells. Interestingly, the expression we derived earlier goes through essentially unchanged except that now we consider the radial concentration, and instead of integrating over planes along the x-direction, we now integrate over spherical shells in the

r direction. Given these adjustments, the power to sustain a gradient is now given as

$$P = k_B T D \int_0^{2\pi} \int_0^{\pi} \int_0^{\infty} \frac{1}{c(r)} \left(\frac{\partial c(r)}{\partial r} \right)^2 r^2 \sin\theta \, dr \, d\theta \, d\phi, \tag{7.21}$$

where D is the diffusion constant and $c(r)$ is the concentration profile.

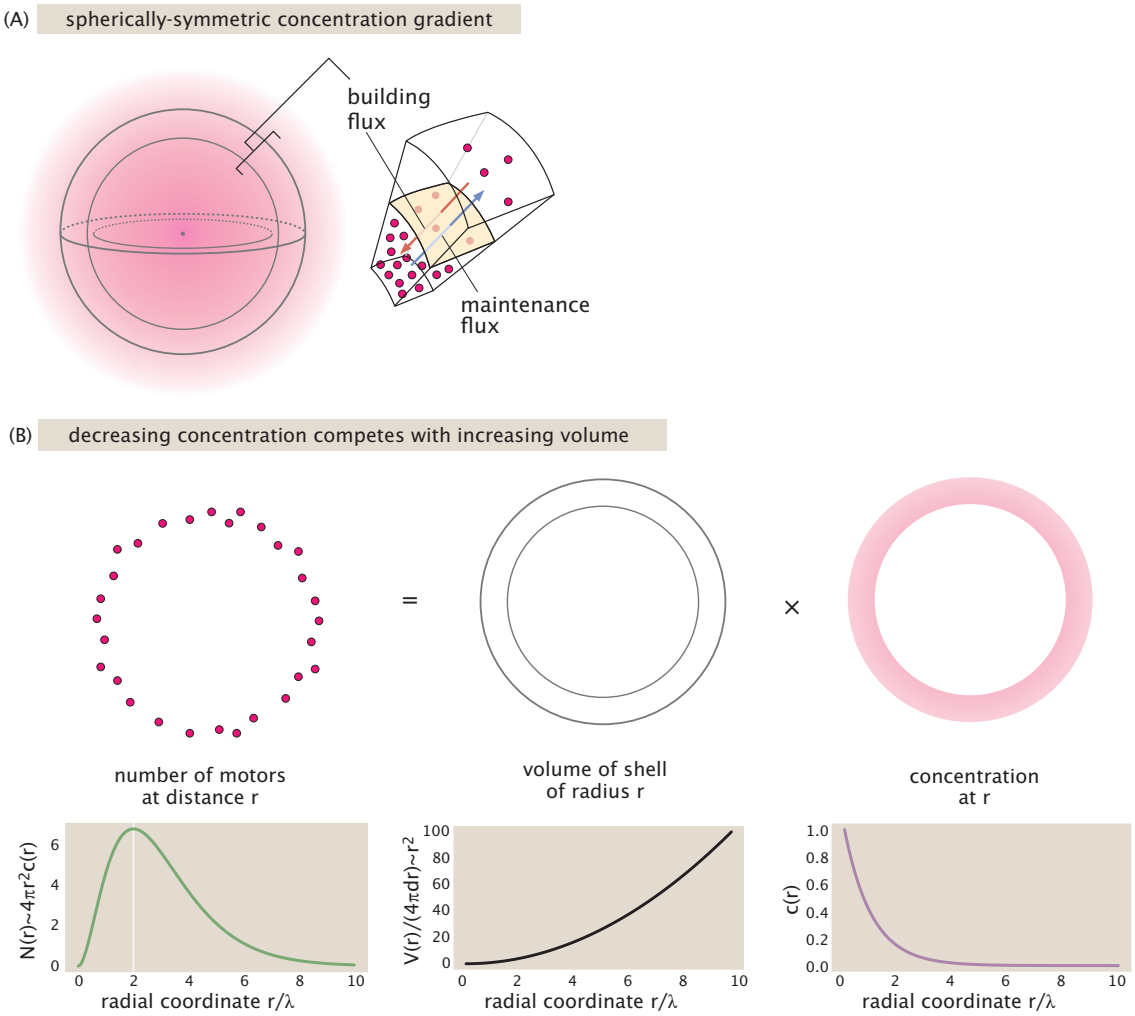


Figure 7.2: Dissipation in a spherically symmetric concentration gradient. (A) Maintaining a spherically symmetric concentration gradient results from a competition between the outward diffusive flux and the inward active flux provided by motors moving on microtubules. (B) There is a peak in the distribution of motors as a function of radial distance resulting from the competition between the monotonically decreasing motor concentration and the geometric effect that the spherical shells get larger with increasing r.

To get a qualitative feeling for the power scales, we begin by considering a fixed, radial

distribution of motors described as a decaying exponential of the form

$$c(r) = c_0 e^{-r/\lambda}, \tag{7.22}$$

where c_0 is the motor concentration at $r = 0$ and λ is the decay length. Plugging this profile into eqn. 7.21, we can immediately compute the power that must be expended to prevent the diffusive relaxation of this gradient as

$$\begin{aligned} P &= 4\pi k_B T D \int_0^\infty \frac{e^{r/\lambda}}{c_0} \left(-\frac{c_0}{\lambda} e^{-r/\lambda}\right)^2 r^2 \, dr \\ &= \frac{4\pi k_B T D c_0}{\lambda^2} \int_0^\infty r^2 \, e^{-r/\lambda} dr. \end{aligned} \tag{7.23}$$

Integrating by parts, we find that the integral evaluates to $\lambda^2 \int_0^\infty r^2 \, e^{-r/\lambda} dr = 2\lambda^3$ yielding,

$$P = 8\pi k_B T D c_0 \lambda. \tag{7.24}$$

Estimate of the magnitude of power to prevent diffusive spreading

As this paper has made evident repeatedly, there is much more to these analyses than obtaining formulae! Rather, we need to put in the numbers and try to understand the magnitude of the various effects that are being paid for by the ATP budget. To that end, we now compute the power required to hold motor proteins inside an aster region of radius $30\ ^-\mathrm{m}$.

To get a feeling for the numbers, we appeal to the measurements described in the paper for which the asters end up described by $c(r) = c_0 e^{-r/\lambda}$, with parameters $c_0 \approx 1\ \mu M$ and $\lambda \approx 30\ \mu m$. Given these values, we first ask roughly how many motors are there in the

gradient given by

$$\#\text{ of motors in aster} = \int_{aster} dV c_0 e^{-r/\lambda} = 8\pi c_0 \lambda^3. \tag{7.25}$$

This result can be seen by using integration by parts on the radial integral. Given that $c_0 = 1\mu M \approx 10^3/\mu m^3$ and $\lambda \approx 30\mu m$, we find that the number of motors in the aster is

$$\#\text{ of motors in aster} = 8\pi c_0 \lambda^3 \approx 8\pi \times \frac{10^3}{\mu m^3} \times (f \times 10\ \mu m)^3 \approx 10^9 \text{ motors.} \tag{7.26}$$

Now let's see how much power is required to maintain those motors in the gradient described above. Using eqn. 7.24, we find

$$P = 8\pi k_B T D c_0 \lambda \tag{7.27}$$

$$\approx (f \times 10) \times k_B T \times (10\ \frac{\mu m^2}{s}) \times \left(\frac{10^3}{\mu m^3}\right) \times (f \times 10\ \mu m) \tag{7.28}$$

$$\approx 5 \times 10^5\ \frac{\text{ATP}}{\text{s}}. \tag{7.29}$$

The numbers we took here for the size of the aster were for " late times" and thus we need to make a more careful analysis for the change in aster size over time to provide a full picture of the dynamics of the power in space and time. This calculation really served as a first cut to get a feeling for the numbers.

The Spatial Distribution of Power

The three dimensional dissipation calculated above suggests interesting spatial effects in the measured power. As seen in Figure 7.2, the radially decaying concentration gradient implies that although the concentration may be higher at smaller radii, the number of motors in a given radial shell is not maximum at the origin. This means that there is the possibility

of nonmonotonic power density. The key idea is illustrated in Figure 7.2(B), where we see that the number of motors as a function of distance from the origin is peaked. This is the result of a competition between the volume of the shells with increasing r (which increases) and the decrease in motor concentration as a function of distance as shown in the lower right panel.

To get a feeling for this effect, we once again turn to our model static, spherically symmetric gradient and show that there is a maximum in the power as a function of radius. Our first task is to compute the power in a spherical shell at radius r. One way to think of the power associated with the spherical shell at radius r is to evaluate eqn. 7.21 only between r and $r + \Delta r$, resulting in

$$P = 4\pi r^2 k_B T D \frac{1}{c(r)} \left(\frac{\partial c(r)}{\partial r} \right)^2 \Delta r. \tag{7.30}$$

If we now exploit the known concentration profile $c(r) = c_0 \exp(-r/\lambda)$ and substitute it into eqn. 7.30, we find the power in the shell at radius r is given by

$$\frac{\partial F}{\partial t}(r) = -(\text{const.}) \times r^2 e^{-r/\lambda}, \tag{7.31}$$

where we have suppressed all constants such as D, c_0, λ, π, etc. This very simple expression now permits us to show that for our simplified model of a static aster, there is a maximum in the dissipation rate as a function of r given by

$$\frac{d}{dr}\left(\frac{\partial F}{\partial t}(r) \right) = -(\text{const.}) \times \left[2re^{-r/\lambda} - \frac{r^2}{\lambda} e^{-r/\lambda} \right] = 0 \tag{7.32}$$

resulting in a maximum in the power dissipation at radius $r = 2\lambda$

To be thorough, we were also curious about the implications of asymmetries in the aster since three-dimensional imaging demonstrates that such asters are not completely spherical.

Here we consider the opposite extreme in which the aster is a cylinder instead.

$$\frac{\partial F}{\partial t}(r) = 2\pi r \Delta z \left(-D\frac{\partial c}{\partial r}\right)\frac{1}{c}\frac{\partial c}{\partial r}\Delta r k_B T. \tag{7.33}$$

This expression has units of power (i.e. J/s) and tells us the power dissipated in the annulus at radius r. That then raises the question of how does the power dissipated depend upon the radius, and specifically, is there a maximum. If we once again exploit the fact that the concentration has cylindrical symmetry and is characterized by $c(r) = c_0 \exp(-r/\lambda)$, then we see that we have

$$\frac{\partial F}{\partial t}(r) = -(\text{const.}) \times r e^{-r/\lambda}, \tag{7.34}$$

where we now suppress the constants such as D, Δr, etc. This result implies in turn that

$$\frac{d}{dr}\left(\frac{\partial F}{\partial t}(r)\right) = -(\text{const.}) \times \left[e^{-r/\lambda} - \frac{r}{\lambda}e^{-r/\lambda}\right] = 0 \tag{7.35}$$

which implies that the maximum dissipation occurs at $r = \lambda$.

The calculations of this section are both interesting and suggestive for the measured ATP consumption in our aster experiments. However, the discussion here was based upon a variety of simplifying assumptions, the most important of which is that the aster is fixed in shape over time. However, we know that asters evolve in time from the moment that light is used to cross link the motors leading to a time-dependent aster size and motor concentration. In the next sections, we begin by asking about the difference between the power to construct and maintain an aster. Of course, to construct an aster and the motor concentration gradient requires more than diffusion. There are active fluxes and we now examine the power of these fluxes.

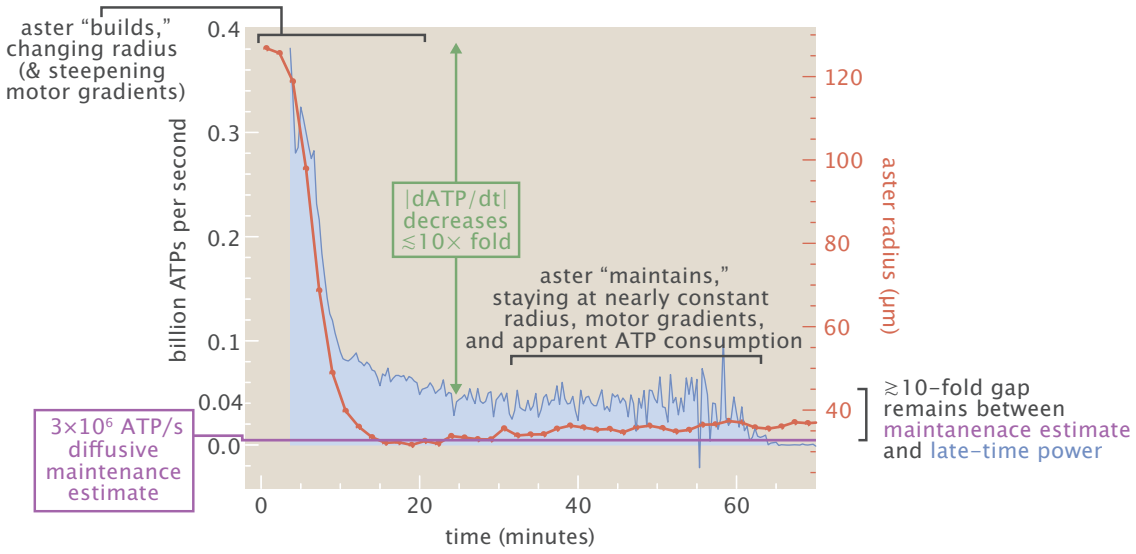


Figure 7.3: Experimental data motivating the question of how much power it takes to build versus maintain an aster's concentration gradients. The x-axis is time and the y-axis is the power, making it evident that the power consumption changes over time as the microtubule-motor system evolves towards its nonequilibrium aster steady state.

7.0.3 How much power does it take to *build* a gradient compared to that to *maintain* it?

As seen in Figure 7.3, representative aster footage suggests that ATP is spent almost ten times faster at early times than at late times in the dynamics of aster formation after light-induced motor dimerization occurs. This widens the gap between our largest measured dissipations and our best quantitative estimates for significant origins of power. As a result, in the current section, we refine our earlier estimates by examining the transient power in the earlier stages of aster formation.

Why does power change like this so much over time? One simple explanation is ATP runs out at late times and so trivially slows the hydrolysis rate by motors (since this rate scales with the ATP concentration). Another—much more interesting!—explanation is that the formative physics of asters at early times fundamentally demand separate, and greater, nonequilibrium expenditures than those required to just maintain an aster's structure once it has formed at late times. A precise synthesis of both specific experiments modulating global

ATP levels and field theory is needed to adjudicate these hypotheses.

To bear on these questions and learn about what they say about the precise destinies of ATP expenditure in pattern formation, here we calculate the free energy cost per time required to both build a concentration gradient in time, and to maintain it at a given state. Combining these calculations with representative phenomenology of motor distributions, we find much support for the latter idea that the power required to assemble a nonuniform profile can be hundreds of times larger than the power cost merely to maintain it (during some stages of aster formation). However, remarkably, the relative importance of building power and maintenance power often switches over the course of an aster's trajectory, and more generally can vary over space, time, and values of biophysical parameters. We introduce an amusing and natural "dissipative Peclét number" definable locally or globally in spacetime that neatly organizes and characterizes this competition between dissipative origins.

Building versus maintaining *arbitrary* (spherically-symmetric) gradients

Organisms build and depend on exquisite patterns of chemicals in space and time. Precisely how much energy must these systems pay to incite these patterns, and then to defy their decay towards equilibrium homogeneity? Does building or maintaining structures demand greater biochemical energy and power expenditures?

As simple steps towards these enduring questions, consider some active agent spending energy to maintain—or even grow—a concentration gradient over space. This activity, administering some material flux $\mathbf{J}_A$, must defy the spontaneous diffusive flux $\mathbf{J}_D$ of the substance that tends to relax the concentration gradient to uniformity. For a gradient to persist or build, these diffusive and active fluxes operate in opposing directions, making the concentration evolve according to a net flux $\mathbf{\Delta J} \equiv \mathbf{J}_A + \mathbf{J}_D$.

For concreteness, when the concentration profile is spherically-symmetric (varying only

a radial coordinate r), and without loss of generality under the convention that more substance is found at small r than at large r (namely $\frac{\partial c}{\partial r} \leq 0$), the agent's flux J_A has a negative sign, $J_A \leq 0$, and the diffusive flux is positive, $J_D = -D\frac{\partial c}{\partial r} \geq 0$. Calling the net flux $J_r = \Delta J$ since it is in the radial direction, these conventions cast the net flux as $J_r = J_D - |J_A|$. (If the gradient is merely maintained, neither decaying nor steepening, then the net flux is zero and $|J_A| = |J_D|$: the active agent moves just enough material thus spending just enough energy to just counteract the spontaneous diffusive flux that would result if it were not acting. While building a gradient, the agent's total active flux is thus partitioned into a maintenance flux matching diffusion, plus any attained nonzero net flux $|\Delta J| > 0$ steepening the gradient,

$$|J_A(r,t)| = |J_D(r,t)| + |J_r(r,t)|. \tag{7.36}$$

We can examine the mathematical underpinnings of this question as follows. Let the total free energy density at position $\mathbf{r}$ and time t be $f(\mathbf{r},t)$. Note that the free energy density should depend on time and space only via the concentration field $c(\mathbf{r},t)$, that is $f \equiv f(c(\mathbf{r},t))$. Then the total free energy $F^{\text{tot}}(t)$ of the system is its integral over all space,

$$F^{\text{tot}}(t) = \int d^3\mathbf{r}\; f(c(\mathbf{r},t)), \tag{7.37}$$

and the total power is the time derivative of this total free energy, giving,

$$\mathrm{P}^{\text{tot}}(t) = \frac{dF^{\text{tot}}}{dt} \tag{7.38}$$

$$= \frac{\partial}{\partial t}\int d^3\mathbf{r}\; f(c(\mathbf{r},t)) = \int d^3\mathbf{r}\; \frac{\partial f(c(\mathbf{r},t))}{\partial t} \tag{7.39}$$

$$= \int d^3\mathbf{r} \frac{\partial f(c)}{\partial c}\frac{\partial c(\mathbf{r},t)}{\partial t}. \tag{7.40}$$

Note that the chemical potential is by definition the change in free energy with respect to a change in particle number, or (dividing a local volume element by a volume) equivalently the change in free energy density with respect to concentration, $\mu(\mathbf{r}, t) \equiv \frac{\partial f}{\partial c}$. In addition, note that conservation of mass gives the standard continuity equation in the concentration field: the local rate at which concentration changes is the divergence of the net flux $\mathbf{J}(\mathbf{r}, t)$ without any homogeneous local sources,

$$\frac{\partial c}{\partial t} + \nabla \cdot \mathbf{J}_{\text{net}} = 0, \tag{7.41}$$

so we can express the $\frac{\partial c}{\partial t}$ in the integrand of Eq. 7.40 in terms of the net flux's divergence, $\frac{\partial c}{\partial t} = -\nabla \cdot \mathbf{J}_{\text{net}}$, yielding,

$$\mathrm{P}^{\text{tot}}(t) = \int d^3\mathbf{r} \frac{\partial f(c)}{\partial c} \frac{\partial c(\mathbf{r}, t)}{\partial t} \tag{7.42}$$

$$= -\int d^3\mathbf{r}\, \mu(\mathbf{r}, t)\, \nabla \cdot \mathbf{J}_{\text{net}}. \tag{7.43}$$

Eq. 7.43 is a volume integral of a scalar field (the chemical potential μ) times the divergence of a vector field (the net flux J_{net}). This permits us to invoke a consequence of the divergence theorem, which (see https://en.wikipedia.org/wiki/Divergence_theorem#Corollaries) delivers the vector identity that,

$$\int dV \left[\mathbf{F} \cdot (\nabla g) + g\,(\nabla \cdot \mathbf{F})\right] = \oint d\partial V \ \left(g\mathbf{F} \cdot \mathbf{n}\right) \tag{7.44}$$

$$\rightarrow \int dV\, g\,(\nabla \cdot \mathbf{F}) = \oint d\partial V \ \left(g\mathbf{F} \cdot \mathbf{n}\right) - \int dV \mathbf{F} \cdot (\nabla g)\,. \tag{7.45}$$

Identifying g as μ and $\mathbf{F}$ as $\mathbf{J}_{\text{net}}$ in this vector identity, we see that the power can be written as a volume integral of a gradient in the chemical potential coupled with the flux, minus a

surface integral,

$$\mathrm{P}^{\mathrm{tot}}(t) = \int d^3\mathbf{r}\, \mathbf{J}_{\mathrm{net}} \cdot \big(\nabla \cdot \mu\big) - \oint dA\; \mu(\mathbf{J}_{\mathrm{net}} \cdot \mathbf{n}). \tag{7.46}$$

For a system with no explicit dissipative action at infinity, the surface integral should vanish. This gives,

$$\mathrm{P}^{\mathrm{tot}}(t) = \int d^3\mathbf{r}\, \mathbf{J}_{\mathrm{net}} \cdot \big(\nabla \cdot \mu\big)\,. \tag{7.47}$$

Given a chemical potential $\mu(r) = \mu_0 + k_B T \ln c(r,t)/c_0$ varying in space (for a standard state concentration c_0 and standard potential μ_0), the local change in chemical potential imposed by moving a particle of substance an infinitesimal distance dr is, [1]

$$d\mu = \boxed{\frac{\partial \mu}{\partial r} dr} \tag{7.48}$$

$$= dr\; k_B T\; \frac{1}{c(r,t)} \frac{\partial c}{\partial r}. \tag{7.49}$$

This free energy cost couples with the earlier fluxes to define the areal power densities σ of maintaining and building the gradient,

$$\underbrace{\sigma_A(r,t)}_{\mathrm{W\ m^{-2}\ s^{-1}}} \equiv |J_A(r,t)| \frac{\partial \mu}{\partial r} dr = \underbrace{|J_D(r,t)| \frac{\partial \mu}{\partial r}}_{\equiv \rho_D, \mathrm{W\ m^{-3}\ s^{-1}}}\, dr + \underbrace{|J_r(r,t)| \frac{\partial \mu}{\partial r}}_{\equiv \rho_r, \mathrm{W\ m^{-3}\ s^{-1}}}\, dr. \tag{7.50}$$

[1] We invoke the continuum description of μ directly at the start of all calculations, instead of by a Taylor approximation of a differential $\Delta\mu$. Namely, define $\mu(r) \equiv \mu_0 + k_B T \ln c(r)/c_0$, already in its continuous non-Δ form. (In an ideal solution with composition c we can generally write $\mu = \mu_0 + k_B T \ln c/c_0$. See e.g. Atkins & de Paula, *Physical Chemistry*, 9th Ed., pg. 164, eq 5.19 [1] for review of this basic ideal logarithmic variation. If concentration varies in space as $c(r)$, we thus write $\mu(r) = \mu_0 + k_B T \ln c(r)/c_0$.) Then, $\frac{d\mu}{dr} = k_B T \frac{1}{c(r)} \frac{\partial c}{\partial r}$ is an automatic exact result (well, up to the validity of the logarithmic variation in the first place), requiring no Taylor expansion approximation. This directly gives the continuous analog of $\Delta\mu = \Delta r\; k_B T\; \frac{1}{c(r)} \frac{\partial c}{\partial r}$, which becomes $d\mu = dr \frac{d\mu}{dr} = dr\; k_B T\; \frac{1}{c(r)} \frac{\partial c}{\partial r}$.

Here we also defined the volumetric power densities $\rho_D(r,t)$ and $\rho_r(r,t)$; their values $4\pi r^2\rho(r,t)dr$ give the power attributable to a shell of radial thickness dr around the radial position r, and when integrated radially over space give the total power for these processes.

Integrating this density over all space attacks the question of whether the total power to build a gradient, or that to maintain it against diffusion, dominates. We can express the magnitude of the total power spent by a source as

$$P_{\text{tot}}(t) = \int \left(\frac{\partial \mu}{\partial r} dr\right) \underbrace{4\pi r^2}_{\text{area through which flux passes}} (|J_D| + |J_r|)\,. \tag{7.51}$$

This allows us to explicitly quantify the partitioning of total power into the power P_{build} attributable to building and power P_{maintain} to maintain the gradient, namely,

$$P_{\text{tot}}(t) = \boxed{\underbrace{\int dr\, 4\pi r^2 \frac{\partial \mu}{\partial r} |J_D(r,t)|}_{\equiv P_{\text{maintain}}(t)}} + \boxed{\underbrace{\int dr\, 4\pi r^2 \frac{\partial \mu}{\partial r} |J_r(r,t)|}_{\equiv P_{\text{build}}(t)}}. \tag{7.52}$$

To compute the relative magnitudes of these terms, we must specify the governing fluxes.

How does a specified time evolution of a concentration field $c(r,t)$ set the fluxes governing these dissipative expressions? Recall that by continuity, a flux has a divergence that sets the rate of change of concentration; if the flux $\mathbf{J} \equiv [J_r, J_\theta, J_\phi]$ is spherically-symmetric (making all angular derivatives vanish, $\partial_\theta, \partial_\phi \to 0$), then,

$$\frac{\partial c(r,t)}{\partial t} = -\vec{\nabla} \cdot \mathbf{J} \tag{7.53}$$

$$= -\frac{1}{r^2}\frac{\partial}{\partial r}\left[r^2 J_r\right]. \tag{7.54}$$

This means that

$$-r^2 \frac{\partial c}{\partial t} = \frac{\partial}{\partial r}\left[r^2 J_r\right], \tag{7.55}$$

or integrating with respect to space from $r = 0$ to $r = r$,

$$-\int_0^r dr\, r^2\, \frac{\partial c}{\partial t} = r^2 J_r\big|_{r=r} - r^2 J_r\big|_{r=0}. \tag{7.56}$$

Since we expect the flux $J_r(r = 0, t)$ to vanish at the origin (since there is no point source or sink there), the latter term vanishes, establishing that the radial flux is given by the radially-integrated rate-of-change of concentration,

$$J_r(r,t) = \boxed{-\frac{1}{r^2}\int_{r=0}^{r} dr\, r^2 \frac{\partial c}{\partial t}}. \tag{7.57}$$

On the maintenance side, Fick's first law tells us the flux required to compensate for diffusion is given by,

$$|J_D(r,t)| = \left|-D\frac{\partial c}{\partial r}\right| \tag{7.58}$$

$$= \boxed{D\left|\frac{\partial c}{\partial r}\right|}. \tag{7.59}$$

Taken together, Equations 7.49, 7.57, and 7.59 give precise answers to the question asked by Eq. 7.52 of how the powers to maintain or build any prescribed concentration profile $c(r,t)$ compare. To gain further insight, we now examine the consequences and lessons of these expressions by adopting specific concentration profiles $c(r,t)$.

Number-conserving, steepening, exponential concentration gradients

Below, we consider an example concentration profile $c(r,t)$. The key difference from the examples of section 7.0.2 is that now we consider the case in which the spatial extent of the gradient is varying over time (i.e. $\lambda(t)$). This model is physically-plausible and—it transpires—analytically-convenient. This model is motivated by two stylized facts. First, motors do exhibit approximately exponential distributions in space at some time points and over relevant spatial regions. See the conceptual description of such distributions in Figures 7.4(A) and (B) as well as exemplary phenomenology of some empirical motor distributions in Figure 7.5.

Second, physically, we believe there really should be conservation of total motors. Consider a spherically-symmetric time varying concentration gradient given by,

$$c(r,t) = c_0(t)\exp\left[-\frac{r}{\lambda(t)}\right], \tag{7.60}$$

where $c_0(t)$ is the concentration at the radial origin of the gradient ($r = 0$), namely the motor peak, and $\lambda(t)$ is a time-varying length scale of the gradient characterizing the spatial evolution of the field. The total number of motors is given by

$$N(t) = \int_{r=0}^{\infty} dr\, 4\pi r^2 c(r,t) = 8\pi\lambda(t)^3 c_0(t). \tag{7.61}$$

Now, to impose conservation of motors, we demand this integral amount to a constant value N, setting the amplitude prefactor $c_0(t)$ as,

$$c_0(t) \equiv \frac{N}{8\pi}\frac{1}{\lambda(t)^3}. \tag{7.62}$$

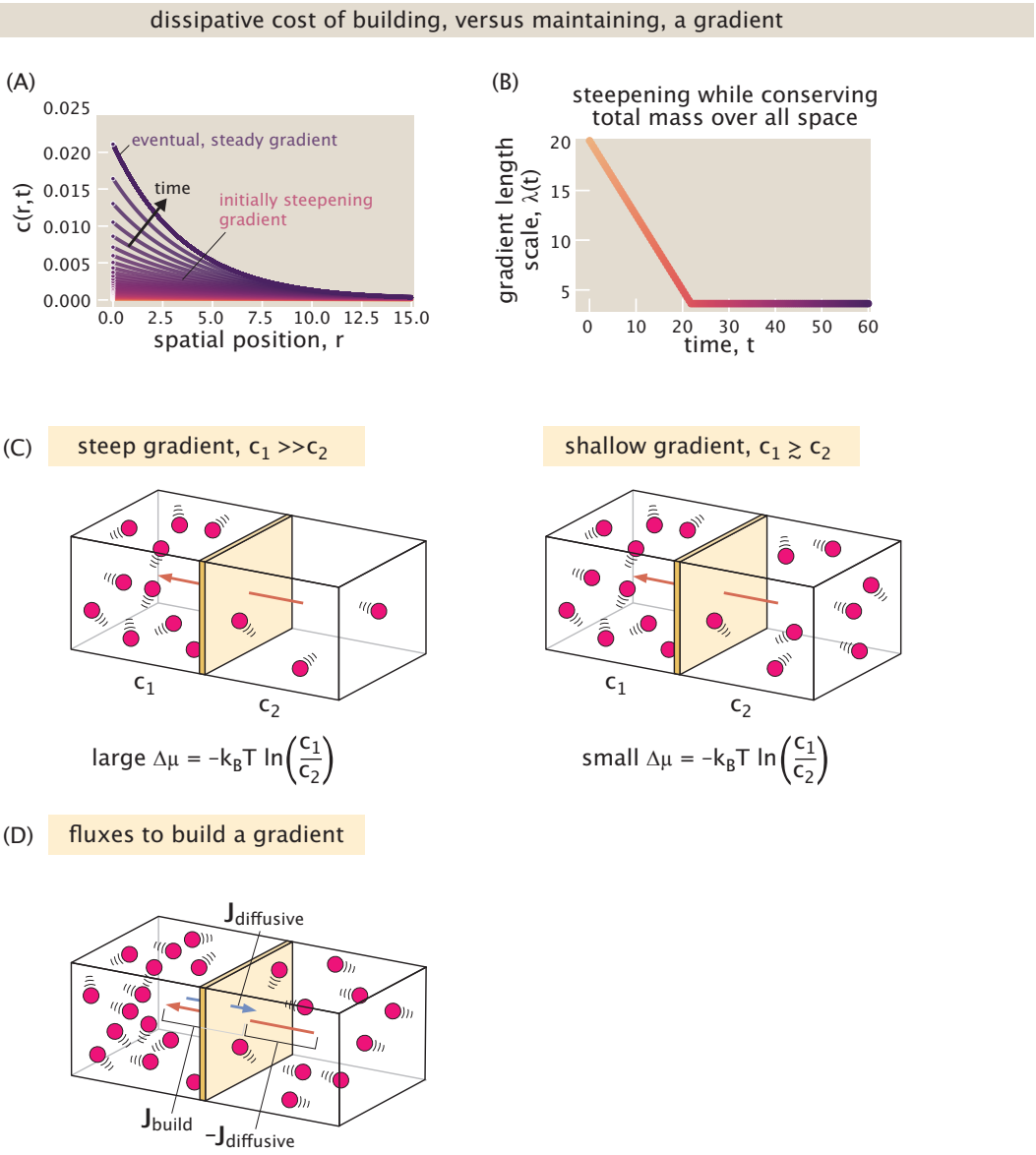


Figure 7.4: The dynamics of building a gradient. (A) The concentration profile as a function of time. (B) The length scale of the exponential gradient over time. (C) Schematic of the free energy cost of a steep and a shallow gradient. (D) The fluxes associated with building a gradient.

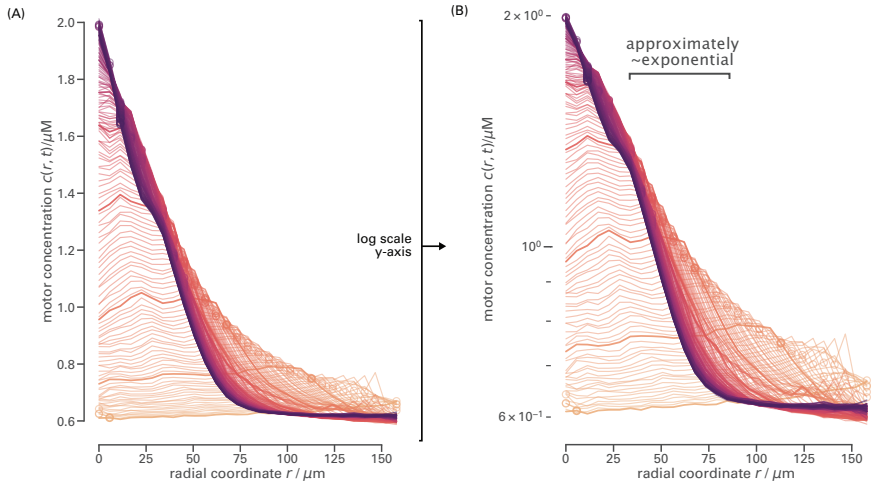


Figure 7.5: Example motor distributions over space, angularly averaged to show variation of motor concentration over radial coordinate r. Different lines (in distinct colors) show progression in time. (A) Motors steepen their gradients in time, as visible on linearly-scaled a y-axis. (B) Some regions of the aster are acceptably described as approximately exponentially declining with radial position r away from the center, as demonstrated by acceptably straight line on a semilog plot of the same data in (A).

This sets the spacetime evolution as,

$$\boxed{c(r,t) = \frac{N}{8\pi}\frac{1}{\lambda(t)^3}\exp\left[-\frac{r}{\lambda(t)}\right]}. \tag{7.63}$$

Given such a time-dependent concentration profile, we are poised to examine the power associated with both building and maintaining the gradient.

Resulting (net and diffusive) fluxes of steepening, exponential concentration gradients

Now, we substitute the conservative time evolution profile $c(r,t)$ of Eq. 7.63 into the expression for the net flux derived earlier using the continuity equation. First, we compute

the relevant corresponding time derivative $\frac{\partial c}{\partial t}$ as

$$\frac{\partial c}{\partial t} = \frac{N}{8\pi}\left(-3\frac{1}{\lambda(t)^4}\partial_t\lambda\exp\left[-\frac{r}{\lambda(t)}\right] + \frac{1}{\lambda(t)^3}\exp\left[-\frac{r}{\lambda(t)}\right]\frac{r\partial_t\lambda}{\lambda(t)^2}\right) \tag{7.64}$$

$$= \frac{N}{8\pi}\left(\frac{\partial_t\lambda}{\lambda(t)^4}\exp\left[-\frac{r}{\lambda(t)}\right]\left(\frac{r}{\lambda(t)} - 3\right)\right) \tag{7.65}$$

$$= c(r,t)\frac{1}{\lambda(t)}\frac{\partial\lambda}{\partial t}\left(\frac{r}{\lambda(t)} - 3\right). \tag{7.66}$$

Substituting this time derivative into our flux expression Eq. 7.57 gives,

$$J_r(r,t) = -\frac{1}{r^2}\left[\int_{r=0}^{r} dr\; r^2\;\frac{N}{8\pi}\left(\frac{\partial_t\lambda}{\lambda(t)^4}\exp\left[-\frac{r}{\lambda(t)}\right]\left(\frac{r}{\lambda(t)} - 3\right)\right)\right] \tag{7.67}$$

$$= -\frac{N}{8\pi}\frac{\partial_t\lambda}{\lambda(t)^4}\frac{1}{r^2}\left[\frac{1}{\lambda(t)}\int_{r=0}^{r} dr\; r^3\;\exp\left[-\frac{r}{\lambda(t)}\right] - 3\int_{r=0}^{r} dr\; r^2\exp\left[-\frac{r}{\lambda(t)}\right]\right] \tag{7.68}$$

We evaluate each remaining integral in turn. The first integral $\int_{r=0}^{r} dr\; r^3\;\exp\left[-\frac{r}{\lambda(t)}\right]$ evaluates to $6\lambda(t)^4 - \exp\left[-\frac{r}{\lambda(t)}\right]\lambda(t)\left(r^3 + 3r^2\lambda(t) + 6r\lambda(t)^2 + 6\lambda(t)^3\right)$. The second integral $\int_{r=0}^{r} dr\; r^2\exp\left[-\frac{r}{\lambda(t)}\right]$ gives $2\lambda(t)^3 - \exp\left[-\frac{r}{\lambda(t)}\right]\lambda(t)\left(r^2 + 2r\lambda(t) + 2\lambda(t)^2\right)$. These give the remarkably-simple weighted-difference

$$\frac{1}{\lambda(t)}\int_{r=0}^{r} dr\; r^3\;\exp\left[-\frac{r}{\lambda(t)}\right] - 3\int_{r=0}^{r} dr\; r^2\exp\left[-\frac{r}{\lambda(t)}\right] = -\exp\left[-\frac{r}{\lambda(t)}\right]r^3. \tag{7.69}$$

In fact, this simplicity can be anticipated directly by an integration-by-parts result. First, let's start with an integral of the form $\int dr\; r^n e^{-r/\lambda}$. The derivative of the integrand $r^n e^{-r/\lambda}$ is just,

$$\frac{\partial}{\partial r}r^n e^{-r/\lambda} = nr^{n-1}e^{-r/\lambda} - r^n\frac{1}{\lambda}e^{-r/\lambda}. \tag{7.70}$$

Now, integrate each term: this gives $r^n e^{-r/\lambda}$ again on the left side, and two terms involving

similar integrals on different integrands on the right side:

$$\int dr \left[\frac{\partial}{\partial r} r^n e^{-r/\lambda}\right] = n \int dr\, r^{n-1} e^{-r/\lambda} - \frac{1}{\lambda} \int dr\, r^n e^{-r/\lambda} \tag{7.71}$$

$$\rightarrow r^n e^{-r/\lambda} = n \int dr\, r^{n-1} e^{-r/\lambda} - \frac{1}{\lambda} \int dr\, r^n e^{-r/\lambda}. \tag{7.72}$$

This result directly gives the net flux cancellation above.

Accordingly, the net (building) material flux adopts the charmingly-concise final form,

$$J_r(r,t) = -\frac{N}{8\pi} \frac{\partial_t \lambda}{\lambda(t)^4} \frac{1}{r^2} \left[-\exp\left[-\frac{r}{\lambda(t)}\right] r^3\right] \tag{7.73}$$

$$= r \frac{\partial_t \lambda}{\lambda(t)} \frac{N}{8\pi} \frac{1}{\lambda(t)^3} \exp\left[-\frac{r}{\lambda(t)}\right] \tag{7.74}$$

$$= \boxed{r \frac{\partial_t \lambda}{\lambda(t)} c(r,t)}. \tag{7.75}$$

This building flux expression exhibits some crucial properties.

First, we check this expression's units: $\frac{\partial_t \lambda}{\lambda}$ gives units of per time; r gives units of length; and $c(r,t)$ gives units of inverse cubic length. Together these quantities give per length squared per time, as required of any flux. Second, consider the sign of this net flux. A gradient that is steepening has a shrinking characteristic length scale $\lambda(t)$ in time, making $\partial_t \lambda < 0$; this makes $J_r(r,t) < 0$ correspondingly negative (net inwards flux). This sign of the net flux is physically correct: for an extant concentration gradient with greatest concentration in the center at small r and smaller concentration at larger r, we have $\frac{\partial c}{\partial r} < 0$, so a source flux that steepens this imbalance will be indeed pointed inwards, and the opposing diffusive flux $J_D = -D\frac{\partial c}{\partial r} > 0$ will be positive (pointing outwards).

Third, note that the net flux is proportional to temporal rate of change of the length scale λ divided by its current value, which is equivalent to $\frac{\partial_t \lambda}{\lambda(t)} = \frac{\partial \ln \lambda/\lambda_0}{\partial t}$ for some reference length λ_0. That is, the magnitude of the net flux is larger when the governing length scale changes

its value faster relative to/*in units of* its current value. If the governing length scale does not change at all in time, $\partial_t \lambda = 0$, then the net flux also vanishes.

Third, note that the flux scales with the radial position r. This can counteract the tendency for concentrations to be smaller at larger r. Specifically, at a fixed time (e.g. setting the value of $\frac{\partial_t \lambda}{\lambda}$ to some fixed constant) the spatial position r where the magnitude of this net building flux J_r is largest is peaked (nonmonotonic) in r at some optimum r_*, since there is a geometric competition between the $\sim r$ term and the exponentially declining gradient in r. The position r_* of this maximum net building flux satisfies

$$0 := \frac{\partial J_r}{\partial r} = \frac{\partial_t \lambda}{\lambda} \left(r \underbrace{\frac{\partial c}{\partial r}}_{\leq 0} + c(r,t) \right) \tag{7.76}$$

$$\rightarrow r_* = \frac{-c(r,t)}{\frac{\partial c}{\partial r}} \tag{7.77}$$

$$\rightarrow r_* = \lambda(t). \tag{7.78}$$

Here, we have learned the following interesting, elegant, nonobvious fact: the maximal material net flux occurs at the radial position which is exactly the current governing exponential length scale. This spatially-maximized flux is $J_r(\lambda, t) = \partial_t \lambda \; c(\lambda, t) = \partial_t \lambda \; \frac{N}{8\pi} \frac{1}{\lambda(t)^3} \frac{1}{e}$. Fourth, note that the flux scales with the absolute local concentration $c(r,t)$: scaling the concentration everywhere by some multiplicative factor α increases the corresponding flux by the same amount. Many of these properties are illustrated for the example of a linear ramp in time of the governing length scale $\lambda(t)$) in Figure 7.6.

For comparison to this net building flux, we also compute the corresponding maintenance (diffusive) flux under this simple conservative model of concentration $c(r,t)$ of Eq. 7.63,

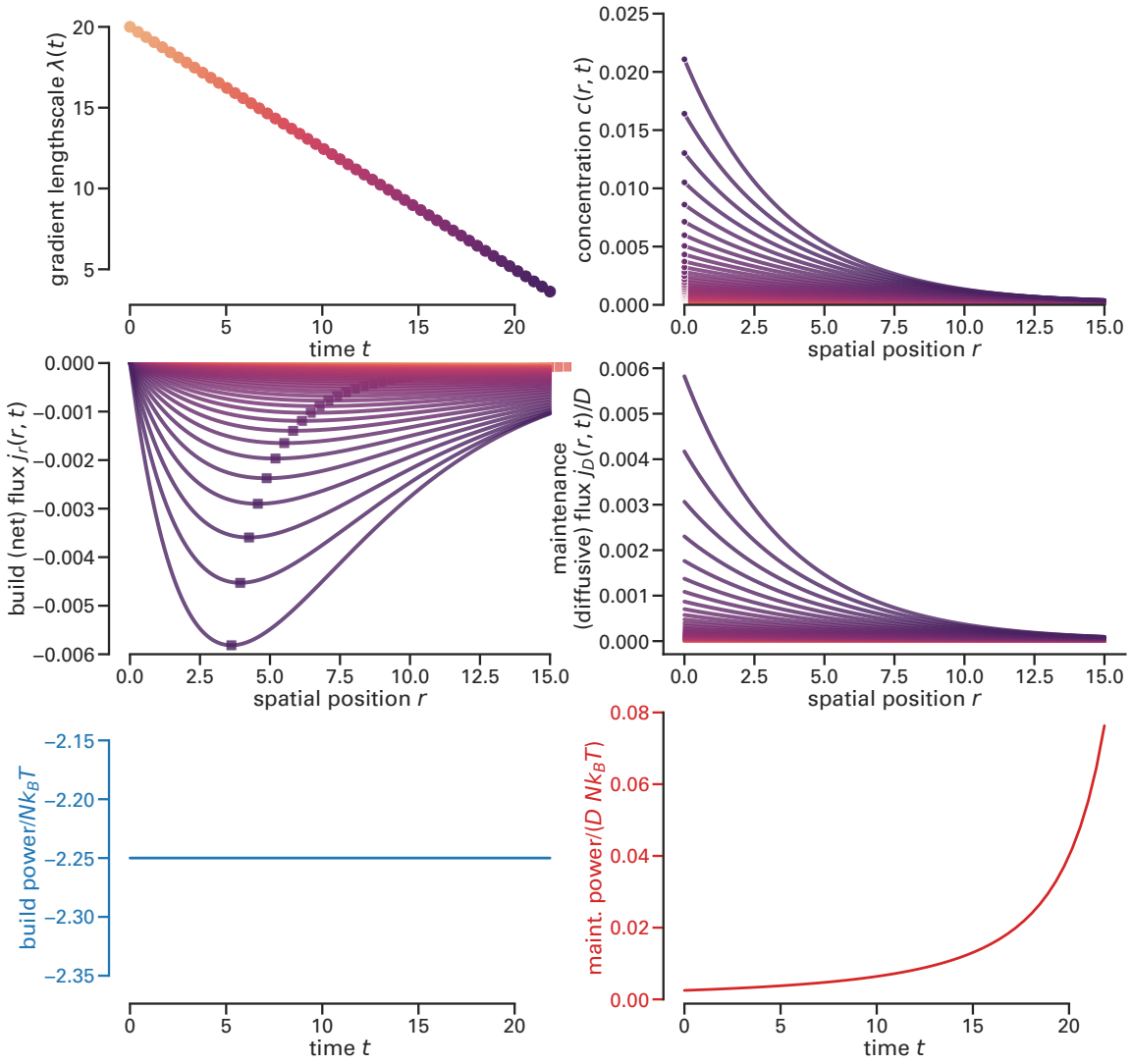


Figure 7.6: Example of the dissipative behaviors of a steepening (number-preserving) exponential gradient in time, where the characteristic length scale $\lambda(t)$ decreases as a linear ramp in time, $\lambda(t) = \lambda_0 - \gamma t$. (A) The illustrative linear ramp of the length scale in time, $\lambda(t) = 20 - 0.75t$, for arbitrary units of λ, t. (B) The corresponding concentration profile $c(r,t)$ over space and time (the latter indicated by distinct colors). (C) The resulting net material flux $J_r(r,t)$ required to build the gradient in this prescribed manner. Squares emphasize that the maximum flux occurs at the radial coordinate $r_* = \lambda(t)$. (D) The material flux J_D required to maintain the gradient at its current state. (E) The corresponding power $P_{\text{build}}(t)$ required to enact the steepening trajectory of the concentration gradient. This linear ramp shows a particularly simple (boring) form as $\partial_t \lambda$ is a constant for this example. (F) The maintenance (diffusive) power P_{maintain} increases with the gradient steepness, so also increases in time. (In all these traces, we measured concentration in units of the prefactor $N/8\pi$ and took the diffusion constant to be $D = 1$.)

finding,

$$J_D(r,t) = -D\frac{\partial c}{\partial r} \tag{7.79}$$

$$= -D\left(-\frac{1}{\lambda(t)}\frac{N}{8\pi}\frac{1}{\lambda(t)^3}\exp\left[-\frac{r}{\lambda(t)}\right]\right) \tag{7.80}$$

$$= D\frac{1}{\lambda(t)}c(r,t). \tag{7.81}$$

Comparing building (net) and maintenance (diffusive) *material* fluxes

Now, we compare the magnitudes of the building and maintenance fluxes given by Eqs. 7.75 and 7.81 by computing their ratio. Importantly, this local ratio of the material transport flux to maintain the gradient, versus that to build the gradient, is also exactly interpretable as the ratio of power densities accomplishing these processes at each point in spacetime. This follows since the volumetric power densities of both processes are each proportional to $\frac{\partial \mu}{\partial r}\Big|_{(r,t)}$ as specified by Eq. 7.50, namely,

$$\frac{\rho_r(r,t)}{\rho_D(r,t)} = \frac{|J_r(r,t)|\frac{\partial \mu}{\partial r}}{|J_D(r,t)|\frac{\partial \mu}{\partial r}} \tag{7.82}$$

$$= \frac{|J_r(r,t)|}{|J_D(r,t)|}. \tag{7.83}$$

Proceeding to compute this ratio of the net flux $J_r(r,t)$ to the maintenance diffusive flux $J_D(r,t)$ (and recalling $|J_r| = -J_r$, we see that

$$\mathrm{Pe}^{\circ}(r,t) \equiv \frac{|J_r(r,t)|}{J_D(r,t)} = \frac{-r\frac{\partial_t \lambda}{\lambda(t)} c(r,t)}{D\frac{1}{\lambda(t)} c(r,t)} \tag{7.84}$$

$$= \boxed{\frac{r(-\partial_t \lambda)}{D}}. \tag{7.85}$$

$$= \frac{r^2/D}{r/(-\partial_t \lambda)} = \frac{\text{diffusion time}}{\text{“advection” time}}. \tag{7.86}$$

$$\tag{7.87}$$

Thus, under this model, we derive the profoundly surprising result that the relative importance of the net flux J_r needed to *build* the gradient, compared to the flux needed to *maintain* the gradient, depends *only* on whether $r\frac{\partial \lambda}{\partial t} \overset{?}{>} D$. **Interestingly, this ratio can be interpreted as a sort of local material “Péclet number” for each spatial coordinate r and (global) “advective velocity” $\partial_t \lambda$ that compete with diffusion D.** We denote this dimensionless value $\mathrm{Pe}^{\circ}(r,t)$, where the superscript recalls this is a material (and local) value. When this ratio is smaller than one, the diffusive flux dominates; when the ratio exceeds one, the building flux dominates.

Accordingly, in our exponential concentration gradient setting, for any nonequilibrium control scheme accomplishing a time-varying characteristic length scale $\lambda(t)$ (yielding some value of rate-of-change $\partial_t \lambda \neq 0$), there exists some radial position r past which the magnitude of the flux J_r needed to build the gradient will always exceed the magnitude of the flux J_D needed to just maintain the gradient.

Generality of the local form of this ratio being interpreted as a Péclet number. How special to the details of our chosen illustrative example $c(r,t)$ is the fact that we can precisely interpret the ratio of a net flux to a diffusive flux as some governing effective Péclet number? In fact, not special at all; *any* concentration profile or flux will admit such a description, as we now

show. Dimensionally speaking, such a ratio is recognizable as,

$$\left[\frac{J_r}{J_D}\right] = \frac{\frac{1}{\text{area}} \times \frac{1}{\text{time}}}{D\frac{\text{concentration}}{\text{length}}} \tag{7.88}$$

$$= \frac{\frac{1}{\text{length}^2} \times \frac{1}{\text{time}}}{D\frac{1}{\text{length}}\frac{1}{\text{length}^3}} \tag{7.89}$$

$$= \frac{\text{length}^2 \times \frac{1}{\text{time}}}{D} \tag{7.90}$$

$$= \frac{\text{length} \times \frac{\text{length}}{\text{time}}}{D} \tag{7.91}$$

$$= \frac{\text{length} \times \text{velocity}}{D}, \tag{7.92}$$

which is exactly the form of a Péclet number.

Building and maintenance power densities are maximized at *distinct* parts of space

The radial dependence of the local Péclet number in Eq. 7.87 hints that the relative importance of building and maintenance can change over space. To understand this spatial dependence, we now explore further how the underlying power in a shell at a radial coordinate r varies in space quite distinctly for these two mechanisms. We preview these contrasting behaviors in Fig. 7.7.

The power attributable to a shell of radius r and thickness dr presenting a volumetric power density $\rho(r,t)$ is

$$\text{power}(r,t) = \rho(r,t)\, 4\pi r^2\, dr. \tag{7.93}$$

Therefore, the factor $4\pi\rho(r,t)r^2$—a power per radial length increment—controls the spatial distribution of dissipation. Accordingly, we now consider the shape of this radial power

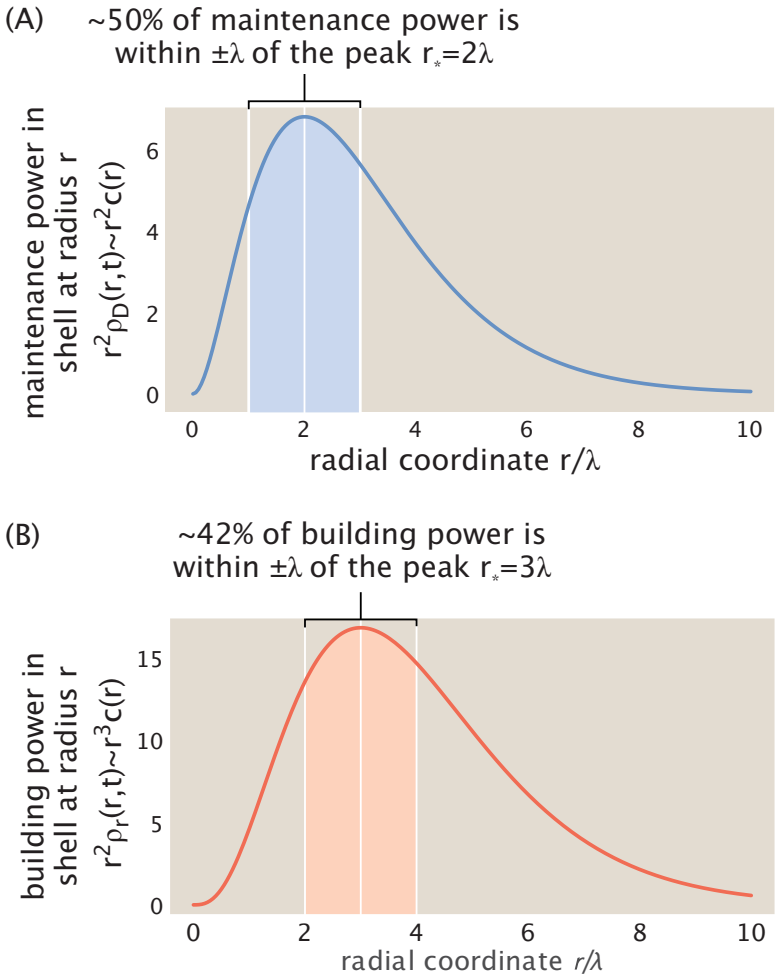


Figure 7.7: Contrasting the locations in space where distinct dissipative mechanisms are expected to be maximal.

density for both maintenance and building power. Substituting the fluxes captured by Eqs. 7.75 and 7.81 and chemical-potential variation (Eq. 7.111) appropriate for our exponential gradient gives the maintenance radial power density as,

$$r^2\rho_D(r,t) = r^2\left(|J_D(r,t)|\frac{\partial\mu}{\partial r}\right) \tag{7.94}$$

$$= r^2\left(D\frac{1}{\lambda(t)}c(r,t)\right)\left(-k_BT\frac{1}{\lambda(t)}\right) \tag{7.95}$$

$$= r^2c(r,t)\left(-\frac{D}{\lambda(t)^2}\,k_BT\right) \tag{7.96}$$

$$= r^2\left(\frac{N}{8\pi}\frac{1}{\lambda(t)^3}\exp\left[-\frac{r}{\lambda(t)}\right]\right)\left(-\frac{D}{\lambda(t)^2}\,k_BT\right) \tag{7.97}$$

$$= r^2\exp\left[-\frac{r}{\lambda(t)}\right]\left(-\frac{ND}{8\pi\lambda(t)^5}\,k_BT\right) \tag{7.98}$$

$$\boxed{\sim r^2\exp\left[-\frac{r}{\lambda(t)}\right]}. \tag{7.99}$$

Similarly, the building radial power density is given as,

$$r^2\rho_r(r,t) = r^2\left(|J_r(r,t)|\frac{\partial\mu}{\partial r}\right) \tag{7.100}$$

$$= r^2\left(r\frac{\partial_t\lambda}{\lambda(t)}c(r,t)\right)\left(-k_BT\frac{1}{\lambda(t)}\right) \tag{7.101}$$

$$= r^3c(r,t)\left(-\frac{1}{\lambda(t)}\frac{\partial_t\lambda}{\lambda(t)}\,k_BT\right) \tag{7.102}$$

$$= r^3\left(\frac{N}{8\pi}\frac{1}{\lambda(t)^3}\exp\left[-\frac{r}{\lambda(t)}\right]\right)\left(-\frac{1}{\lambda(t)}\frac{\partial_t\lambda}{\lambda(t)}\,k_BT\right) \tag{7.103}$$

$$= r^3\exp\left[-\frac{r}{\lambda(t)}\right]\left(-\frac{N}{8\pi}\frac{\partial_t\lambda}{\lambda(t)^5}k_BT\right) \tag{7.104}$$

$$\boxed{\sim r^3\exp\left[-\frac{r}{\lambda(t)}\right]}. \tag{7.105}$$

Since the exponentially-decreasing concentration $c(r,t)$ in space competes with the geometric prefactors of r^2 or r^3, respectively, in Eqs. 7.105 and 7.105, both of these power

densities have local maxima in space; these maxima occur when the power laws just start losing to the exponentially decreasing concentration field. However, notably, the spatial positions where these maxima occur at distinct radial optima. Specifically, the maximum of the radial maintenance power density occurs at $r_* = 2\lambda$, whereas the maximum of the radial building power density occurs at $r_* = 3\lambda$.

The majority of either type of power occurs within a moderately close vicinity of the respective maxima. To quantify just how much power is concentrated around these positions in space, we compute the fraction of this power found within a tolerance (radial distance) $\pm\delta$ of the maximum r_*, namely what we call the *localization fraction*,

$$f(\delta) \equiv \frac{\int_{r_*-\delta}^{r_*+\delta} dr\, r^2 \rho(r,t)}{\int_0^\infty dr\, r^2 \rho(r,t)}. \tag{7.106}$$

We find that the maintenance power density shows the localization fraction (about its maximum $r_* = 2\lambda$),

$$f_D(\delta) = \frac{\int_{r_*-\delta}^{r_*+\delta} dr\, r^2 \exp\left[-\frac{r}{\lambda(t)}\right]}{\int_0^\infty dr\, r^2 \exp\left[-\frac{r}{\lambda(t)}\right]} \tag{7.107}$$

$$= \frac{\frac{1}{e^2} 2\lambda \left(\left(\delta^2 + 10\lambda^2\right) \sinh\left[\frac{\delta}{\lambda}\right] - 6\delta\lambda \cosh\left[\frac{\delta}{\lambda}\right]\right)}{2\lambda^3}. \tag{7.108}$$

To get an intuitive sense of the scale at which this maintenance power is localized in space, we numerically evaluate this localization fraction with $\delta = \lambda$, which asks for the fraction of power in a shell of total radial thickness 2λ about its maximum $r_* = 2\lambda$. We find $f_D(\lambda) \approx 0.4965$, namely about $\boxed{\approx 50\%}$ of all the maintenance power over all space is localized to this radial region.

Next, we contrast this localization behavior to that of the building power, about its distinct maximum $r_* = 3\lambda$. We see that,

$$f_r(\delta) = \frac{\int_{r_*-\delta}^{r_*+\delta} dr\, r^3 \exp\left[-\frac{r}{\lambda(t)}\right]}{\int_0^{\infty} dr\, r^2 \exp\left[-\frac{r}{\lambda(t)}\right]} \tag{7.109}$$

$$= \frac{\frac{1}{e^3} 2\lambda \left(6\lambda \left(2\delta^2 + 13\lambda^2\right) \sinh\left[\frac{\delta}{\lambda}\right] - \delta(\delta^2 + 51\lambda^2) \cosh\left[\frac{\delta}{\lambda}\right]\right)}{6\lambda^4}. \tag{7.110}$$

Again numerically evaluating the localization fraction with a tolerance given by the characteristic length scale, $\delta = \lambda$, we see that $f_r(\lambda) = 0.4237$: about $\boxed{\approx 42\%}$ of the power to build the gradient is localized in this region around its distinct spatial maximum.

Interestingly, these power densities make clear a significant difference. The maintenance density is identical to the distribution of concentration in the first: that is, the fraction of all dissipation found within a shell at the maximum radius is exactly the same as the fraction of all substance found within the same shell. This is in substantial contrast to the distinct weighting of the building power, whose integrand enjoys an extra radial $\sim r^{+1}$ weight factor. Figure 7.7 summarizes and illustrates these dissipative contrasts. Figure 7.8 gives another high-level schematic overview of these spatial relationships.

Comparing building (net) and maintenance (diffusive) *powers*

For our specific exponential gradient, we observe the unusual fact that $\frac{\partial \mu}{\partial r}$ is independent of space r, namely,

$$\frac{\partial \mu}{\partial r} = k_B T \frac{\frac{\partial c}{\partial r}}{c(r,t)} = -k_B T \frac{1}{\lambda(t)}. \tag{7.111}$$

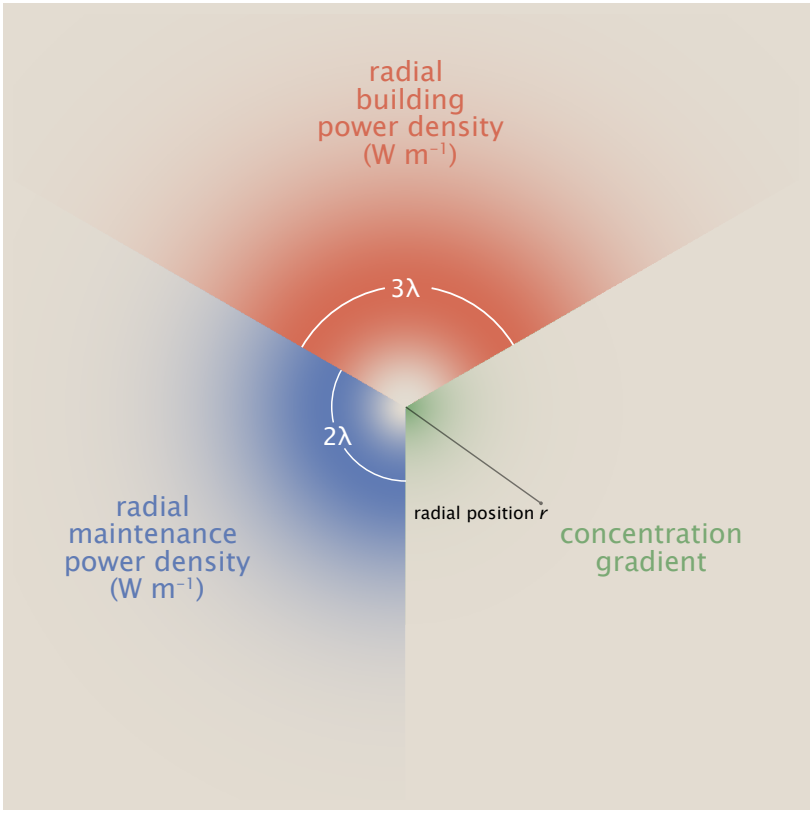


Figure 7.8: Visualization of the distinct spatial localizations of maintenance power (blue, lower left); building power (red, top); and concentration (green, lower right).

Armed with all the ingredients we need, we now compute each power term in turn as

$$P_{\text{build}}(t) = -4\pi \int dr\, r^2 \, \frac{\partial \mu}{\partial r} \left(-\frac{1}{r^2} \int_{r=0}^{r} dr\, r^2 \frac{\partial c}{\partial t} \right) \quad \text{(by Eq. 7.57)}, \tag{7.112}$$

and

$$P_{\text{maintain}}(t) = 4\pi \int dr\, r^2 \, \frac{\partial \mu}{\partial r} \left(-D \frac{\partial c}{\partial r} \right) \quad \text{(by Fick)}. \tag{7.113}$$

Substituting this and the remaining ingredients in our particular exponential gradient scenario, we see,

$$P_{\text{build}}(t) = -4\pi \frac{\partial \mu}{\partial r} \int_0^\infty dr\, r^2 \left(r \frac{\partial_t \lambda}{\lambda(t)} c(r,t) \right) \tag{7.114}$$

$$= -\frac{\partial \mu}{\partial r} \frac{\partial_t \lambda}{\lambda(t)} \int_0^\infty dr\, 4\pi r^2\, r c(r,t) \tag{7.115}$$

$$= -4\pi \frac{\partial \mu}{\partial r} \int dr\, r^2 \left(r \frac{\partial_t \lambda}{\lambda(t)} \frac{N}{8\pi} \frac{1}{\lambda(t)^3} \exp\left[-\frac{r}{\lambda(t)} \right] \right) \tag{7.116}$$

$$= -\frac{\partial \mu}{\partial r} \frac{4\pi}{8\pi} N \frac{\partial_t \lambda}{\lambda(t)} \frac{1}{\lambda(t)^3} \int_0^\infty dr\, r^3 \exp\left[-\frac{r}{\lambda(t)} \right]; \text{ (Mathematica gives } \int_0^\infty dr\, r^3 e^{-r/\lambda} = 6\lambda^4), \tag{7.117}$$

$$= -\frac{\partial \mu}{\partial r} \frac{N}{2} (\partial_t \lambda) \frac{1}{\lambda^4} 6\lambda^4. \tag{7.118}$$

$$= \boxed{\frac{\partial \mu}{\partial r} 3N(-\partial_t \lambda)}. \tag{7.119}$$

$$= -3N k_B T \frac{(-\partial_t \lambda)}{\lambda(t)} \tag{7.120}$$

$$= 3N k_B T \frac{\partial \ln \lambda(t)/\lambda_0}{\partial t} \tag{7.121}$$

(Here we could also just recall that for integer n, the value of $\int_0^\infty dr\, r^n e^{-r/\lambda}$ is $n!\, \lambda^{n+1}$.)

The maintenance power is

$$P_{\text{maintain}}(t) = \frac{\partial \mu}{\partial r} \int dr\, 4\pi\, r^2 \left(D \frac{1}{\lambda(t)} c(r,t) \right) \tag{7.122}$$

$$= \frac{\partial \mu}{\partial r} D \frac{1}{\lambda(t)} \underbrace{\int dr\, 4\pi\, r^2\, c(r,t)}_{=N, \text{ by Eq. } 7.61.} \tag{7.123}$$

$$= \boxed{\frac{\partial \mu}{\partial r} D \frac{1}{\lambda(t)} N}. \tag{7.124}$$

$$= -N k_B T D \frac{1}{\lambda(t)^2} \tag{7.125}$$

Thus the ratios of these two powers is

$$\text{Pe}^{\lightning}(t) \equiv \frac{P_{\text{build}}(t)}{P_{\text{maintain}}(t)} = \frac{\frac{\partial \mu}{\partial r} 3N(-\partial_t \lambda)}{\frac{\partial \mu}{\partial r} D \frac{1}{\lambda(t)} N} \tag{7.126}$$

$$= \boxed{\frac{3\lambda\, (-\partial_t \lambda)}{D}}, \tag{7.127}$$

where we have denoted this ratio as $\text{Pe}^{\lightning}(t)$ to emphasize its nature as a dissipative (and spatially-global) Péclet number. Is this unitless? The numerator is length squared per second, consistent with the denominator, so yes.

(Interestingly, we see that taking a spatial coordinate of $r = 3\lambda$ in the local flux ratio of $\text{Pe}^{\circ}(r,t)$ in Eq. 7.87 makes that local flux ratio there equal to this total radially-integrated power ratio $\text{Pe}^{\lightning}(t)$ of Eq. 7.127.)

How does the relative importance of building versus maintenance powers change over time, as gradients steepen?

Our foregoing analysis was mathematically general, applying to any exponential gradient regardless of the trajectory of the characteristic length scale $\lambda(t)$. To proceed

further, however, we must now specify some possible candidate time-evolutions $\lambda(t)$. How will the relative importance of maintenance versus building power be conserved or vary across diffrent $\lambda(t)$ profiles? Consider these three examples, which hint at the diversity of attainable behaviors.

Figure 7.6 illustrates characteristic behaviors of a concentration gradient steepening under a linearly-decreasing gradient length scale $\lambda(t)$ in time.

1. For a simple linear gradient in time, $\lambda(t) = \lambda_0 - \gamma t$ (for $t \leq \frac{\lambda_0}{\gamma}$), we see that $(\partial_t \lambda)\lambda(t) = -\gamma(\lambda_0 - \gamma t)$. This magnitude also decreases in time. In other words, the relative importance of building power to maintenance power must monotonically decrease over time. This is consistent with the fact that the building power is a constant (while maintenance power increases) in this case: see Eq. 7.120. See Figure 7.6.
2. For a power law shrinkage of the exponential gradient length scale in time, $\lambda(t) = Bt^{-\alpha}$ (with $\alpha \geq 0$ to assure that the gradient steepens), observe that

$$\partial_t \lambda = -B\alpha T^{-\alpha-1} = -\alpha \frac{\lambda(t)}{t}.$$

   Therefore, $(\partial_t \lambda)\lambda(t) = -\alpha \frac{\lambda(t)^2}{t}$, whose magnitude decreases with time t (since $\lambda(t)$ declines monotonically). In other words, the relative importance of building power to maintenance power must monotonically decrease over time.
3. For a more unusual (e.g. somewhat contrived) profile like $\lambda(t) = (1-t)^\eta$ for $\eta \in [0, 1/2]$, and $t < 1$, we appear to find a profile where—in contrast to the earlier examples—the magnitude of $(\partial_t \lambda)\lambda(t)$ increases in time, making the building power *increasingly* important relative to the maintenance power.

Comparing building (net) and maintenance (diffusive) *energies*

Under our model, during a time τ, the total energy spent on building is the time integral of Eq. 7.120,

$$E_{\text{build}}(\tau) = \int_{t=0}^{\tau} dt\ P_{\text{build}}(t) \tag{7.128}$$

$$= \int_{0}^{\tau} dt\ \frac{\partial \mu}{\partial r} 3N \partial_t \lambda \tag{7.129}$$

$$= 3N \int_{0}^{\tau} dt\ \left(-k_B T \frac{1}{\lambda(t)}\right) \partial_t \lambda \tag{7.130}$$

$$= -3N k_B T \int_{0}^{\tau} dt\ \frac{\partial_t \lambda}{\lambda(t)} \tag{7.131}$$

$$= -3N k_B T \int_{0}^{\tau} dt\ \frac{\partial \ln \lambda/\lambda_0}{\partial t} \tag{7.132}$$

$$= \boxed{-3N k_B T \ln\left[\frac{\lambda(\tau)}{\lambda(0)}\right]}. \tag{7.133}$$

Remarkably, this energy to build is *independent* of the precise timecourse $\lambda(t)$, only depending on the starting and ending length scale of the gradient!

We note with amusement that this expression Eq. 7.133 adopts precisely the same form (including numerical prefactors!) as the calculation for the work needed to contract a (microtubule) gas from an initial volume V_i and a final volume V_f in Eq. **??** specifying $W = -N k_B T \ln \frac{V_f}{V_i}$, if we take these volumes to be of order the characteristic length scale cubed, $V_i \sim \lambda(0)^3$ and $V_f \sim \lambda(\tau)^3$, where the cubic relation delivers our factor of three in Eq. 7.133.

In light of our discussion and estimates in this section: the fact that the order-of-magnitude estimate $\langle P_{\text{build}} \rangle = E_{\text{build}}/T = \frac{\int_0^T dt\ P_{\text{build}}(t)}{T}$ (for an observation/construction time T) is much smaller than instantaneous measured power values does *NOT* in fact say that building power is small compared to actual total power expenditures. This discrepancy says much more about the fact that *instantaneous* power

values can differ a lot from the *average* of the same values, than that the *instantaneous* fundamental/theoretical building power is small compared to *instantaneous* measured total power values!

In contrast the energy to maintain is the time integral of Eq. 7.124, giving

$$E_{\text{maintain}}(\tau) = \int_{t=0}^{\tau} dt\ P_{\text{maintain}}(t) \tag{7.134}$$

$$= \int_{t=0}^{\tau} dt\ \frac{\partial \mu}{\partial r} D \frac{1}{\lambda(t)} N \tag{7.135}$$

$$= DN \int_{t=0}^{\tau} dt\ \left(-k_B T \frac{1}{\lambda(t)}\right) \frac{1}{\lambda(t)} \tag{7.136}$$

$$= \boxed{-DNk_B T \int_0^{\tau} dt\ \frac{1}{\lambda(t)^2}}. \tag{7.137}$$

Notice, as expected, that such a maintenance energy accumulates even when the length scale λ is *not* changing in time ($\partial_t \lambda \neq 0$)—and at any finite constant λ, E_{maintain} will grow with longer τ.

It is hard to bound this maintenance energy without stipulating a precise timecourse for $\lambda(t)$. But, to reach for a rough feel, say that we only know that $\lambda(t)$ decreases monotonically in time on the time interval of interest, such that $\lambda(t = \tau) \leq \lambda(t) \leq \lambda(t = 0)$. Then we can conclude the simple fact that,

$$\int_0^{\tau} dt\ \frac{1}{\lambda(0)^2} \leq \int_0^{\tau} dt\ \frac{1}{\lambda(t)^2} \leq \int_0^{\tau} dt\ \frac{1}{\lambda(\tau)^2} \tag{7.138}$$

$$\rightarrow \frac{\tau}{\lambda(0)^2} \leq \int_0^{\tau} dt\ \frac{1}{\lambda(t)^2} \leq \frac{\tau}{\lambda(\tau)^2}, \tag{7.139}$$

which means that

$$DNk_B T \frac{\tau}{\lambda(0)^2} \leq |E_{\text{maintain}}(\tau)| \leq DNk_B T \frac{\tau}{\lambda(\tau)^2}. \tag{7.140}$$

Accordingly, the ratio of these two energies is only constrained here to lie between,

$$\frac{3Nk_BT\ln\left[\frac{\lambda(\tau)}{\lambda(0)}\right]}{DNk_BT\frac{\tau}{\lambda(\tau)^2}} \leq \frac{|E_{\text{build}}(\tau)|}{|E_{\text{maintain}}(\tau)|} \leq \frac{3Nk_BT\ln\left[\frac{\lambda(\tau)}{\lambda(0)}\right]}{DNk_BT\frac{\tau}{\lambda(0)^2}} \tag{7.141}$$

$$\rightarrow \frac{3}{D}\frac{\ln\left[\frac{\lambda(\tau)}{\lambda(0)}\right]}{\frac{\tau}{\lambda(\tau)^2}} \leq \frac{|E_{\text{build}}(\tau)|}{|E_{\text{maintain}}(\tau)|} \leq \frac{3}{D}\frac{\ln\left[\frac{\lambda(\tau)}{\lambda(0)}\right]}{\frac{\tau}{\lambda(0)^2}}. \tag{7.142}$$

Empirical estimates for an example aster's motor gradients

We now consider how well the theoretical insights described above jibe with our experimental measurements. It appears that the major obvious trends—e.g. that the relative importance of building and maintenance powers can switch dramatically in time across trajectories—will be widely preserved across aster phenomenology. To get a very rough sense of scale for how motor gradients actually change their characteristic length scales $\lambda(t)$ in time, we considered the same motor distributions shown earlier in the data of Figure 7.5. At each time, we found the radial position r_s at which the motor distribution was steepest in space. Then, assuming (coarsely, but not unreasonably) that these gradients were indeed locally exponentially decaying at this steepest point, we computed the approximate effective exponential gradient length scale as $\hat{\lambda}(t) = -\frac{c(r_s,t)}{\frac{\partial c}{\partial r}}$. Figure 7.9 depicts how this aster changes its apparent length scale.

Next, we—just by eye, not rigorously—extracted an apparent power law fit to each of two distinct dynamical regimes visible in the length scale's trajectory. This allows us to substitute empirically estimated values into the dissipative Péclet number of Eq. 7.127. The first regime, marked by the solid black square in Figure 7.9, implies that the building power $P_{\text{build}}(t)$ exceeds the maintenance power P_{maintain} by approximately a multiplicative factor of order,

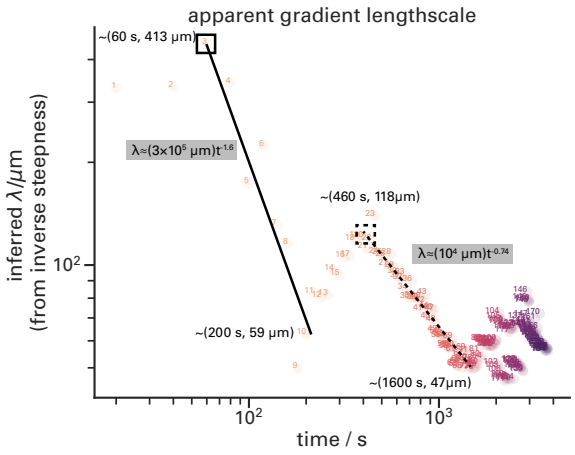


Figure 7.9: Temporal changes in apparent length scale of motor gradients of the data in Figure 7.5. At left, highlighted by a solid black line as a guide to the eye, an early-time steepening in the motors gives a steep approximate power law $\hat{\lambda} \approx (3 \times 10^6\ \mu\text{m})(t/1s)^{-1.6}$. The solid black square indicates a reference point $(t, \hat{\lambda}(t)) \approx (60\ s, 413\ \mu m)$ with an approximate rate-of-change of $\partial_t \hat{\lambda} \approx -11\ \mu m/s$ taken from the slope of the line. At right, a slower but still appreciable steepening resumes, marked by the dashed black line, with an approximate power law fit of $\hat{\lambda} \approx (10^4\ \mu m)(t/1s)^{-0.74}$. The dashed black square indicates a reference point at $(t, \hat{\lambda}) \approx (460\ s, 118\ \mu m)$ with an approximate rate of change of $-0.2\ \mu m/s$ taken from the slope of the dashed line.

$$\frac{P_{\text{build}}(t)}{P_{\text{maintain}}(t)} = \frac{3\lambda\ (-\partial_t \lambda)}{D} \tag{7.143}$$

$$\approx \frac{3(413\ \mu m)(-(-11\ \mu m/s))}{40\ \mu m^2/s} \tag{7.144}$$

$$\boxed{\approx 340\times}. \tag{7.145}$$

The second, more slowly-steepening, regime appears to register a more equal balance between building power and maintenance power, giving about

$$\frac{P_{\text{build}}(t)}{P_{\text{maintain}}(t)} = \frac{3\lambda\ (-\partial_t \lambda)}{D} \tag{7.146}$$

$$\approx \frac{3(118\ \mu m)(-(-0.2\ \mu m/s))}{40\ \mu m^2/s} \tag{7.147}$$

$$\boxed{\approx 1.8\times}. \tag{7.148}$$

These estimates, however coarse, argue quantitatively that **maintenance power could be dwarfed by building power to the tune of order a few hundred in the early times of an aster formation**, but show that the maintenance cost grows much more competitive at late times. □

Bibliography

[1] P. Atkins and J. de Paula. *Atkins' Physical Chemistry*. OUP Oxford, 9th edition, 2010. ISBN 9780199543373. URL https://books.google.com/books?id=BV6cAQAAQBAJ.

Part IV

Exploration out of equilibrium

Chapter 8

Biological Processes as Exploratory Dynamics

This chapter describes joint work with Jane Kondev, Marc Kirschner, Hernan G. Garcia, and Rob Phillips, who spearheaded the project. GS contributed calculations, helped with exploratory simulationss, and helped write the paper. It appears as a preprint on the arχiv.

Many biological processes can be thought of as the result of an underlying dynamics in which the system repeatedly undergoes distinct and abortive trajectories with the dynamical process only ending when some specific process, purpose, structure or function is achieved. A classic example is the way in which microtubules attach to kinetochores as a prerequisite for chromosome segregation and cell division. In this example, the dynamics is characterized by apparently futile time histories in which microtubules repeatedly grow and shrink without chromosomal attachment. We hypothesize that for biological processes for which it is not the initial conditions that matter, but rather the final state, this kind of exploratory dynamics is biology's unique and necessary solution to achieving these functions with high fidelity. This kind of cause and effect relationship can be contrasted to examples from physics and chemistry where the initial conditions determine the outcome.

In this paper, we examine the similarities of many biological processes that depend upon random trajectories starting from the initial state and the selection of subsets of these trajectories to achieve some desired functional final state. We begin by reviewing the long history of the principles of dynamics, first in the context of physics, and then in the context of the study of life. These ideas are then stacked up against the broad categories of biological phenomenology that exhibit exploratory dynamics. We then build on earlier work by making a quantitative examination of a succession of increasingly sophisticated models for exploratory dynamics, all of which share the common feature of being a series of repeated trials that ultimately end in a "winning" trajectory. We also explore the ways in which microscopic parameters can be tuned to alter exploratory dynamics as well as the energetic burden of performing such processes.

8.1 Theories of Inanimate and Animate Dynamics

Whatever level of phenomena we look at, small or large spatial scales, short or long time scales, low or high energies, these phenomena are usually dynamic. The quest to understand the dynamics of broader and broader classes of phenomena has repeatedly over the long arc of the history of science resulted in "new physics" [1, 2, 3, 4]. Indeed, as shown in Figure 8.1, there is a long and justly celebrated history of different approaches to working out the dynamics of processes in the world around us. The template for this subject both mathematically and physically was achieved by Newton standing on the shoulders of Galileo and others [5]. Newton saw that $\mathbf{F} = m\mathbf{a}$ provided an update rule that allowed us, using what we now know as calculus, to find the current state of the system by time stepping from previous states [6]. Though most of us learn about Newton's analysis of the planets in polar coordinates culminating in elliptical orbits, Newton's treatment of the

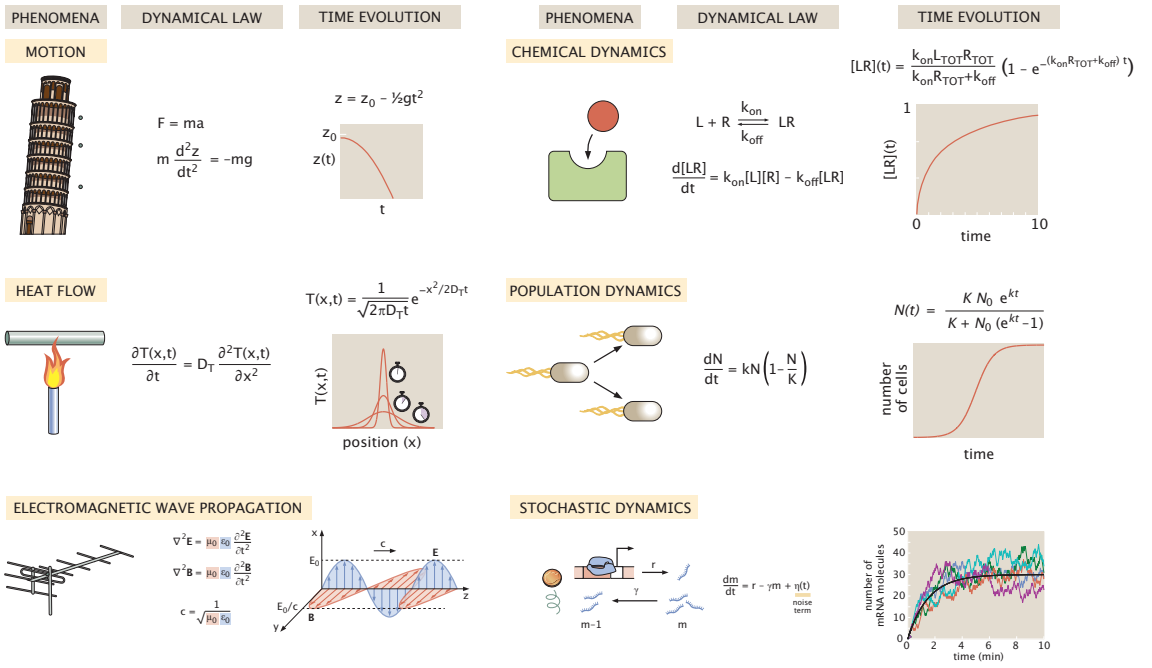


Figure 8.1: Gallery of dynamical equations. The defining example of dynamics in physics is the $F = ma$ dynamics of Newton that allows us to solve problems such as falling bodies or planetary motion. The dynamics of chemical reactions is another well established example of initial-condition driven dynamics, in this case showing the dynamics of ligand (red)-receptor (green) binding. The heat equation is a deterministic dynamical equation that tells us how an initial temperature distribution will evolve in space and time. There are a host of different dynamical equations describing the dynamics of populations, here we show the logistic equation. Maxwell's theory of is a quintessential example of a field theory which vastly expands the repertoire of classical physics. Langevin-like equations even allow us to work out the time evolution of stochastic chemical dynamics.

problem was as a series of straight line steps punctuated by impulses due to the force of gravitation resulting in an orbit that was a polygon [6]. We might summarize this and most subsequent dynamical laws as having the form

$$\text{state}(t + \Delta t) = \text{state}(t) + \text{update}(t)\Delta t, \tag{8.1}$$

the simplest embodiment of the forward-Euler method for solving differential equations, but which we think of differently as being the discrete representation of most versions of deterministic, initial-condition dictated, dynamics. The state now is determined by the state a time Δt earlier plus some update term.

In the centuries that followed Newton's creation of the "System of the World [6]," $F = ma$ dynamics was brilliantly generalized to continuous media with examples such as hydrodynamics and elastodynamics to describe the mechanics of fluids and solids [7, 8]. But other phenomena called for something beyond the mechanical world view [9]. As shown in Figure 8.1, new dynamical laws were articulated that made it possible to work out the time evolution of a myriad of processes. For example, the study of chemical transformations led to the emergence of the law of mass action and the rate equations of chemical kinetics, allowing for the appearance and disappearance of different chemical species [10, 11]. Another class of problems highlighted in Figure 8.1 arose as Fourier and others tackled the dynamics of temperature fields and heat flow in the form of a partial differential equation (the heat equation) showcased in the figure [12, 2]. In this case, a material subjected to some initial temperature distribution will evolve to some new temperature distribution [13, 14]. The dynamical equations describing the diffusion of matter and heat had and have a huge reach including to important problems such as the age of the Earth [15, 16, 17]. While physicists were busy working out the laws of inanimate matter, other kinds of scientists and mathematicians were also engaged in thinking about the dynamics of populations as exemplified by the logistic equation shown in the figure [18, 19, 20, 21]. In the middle of the 19th century, laws of dynamical evolution received a great boost in the form of Maxwell's theory of the electromagnetic field to describe the spacetime evolution of electromagnetic fields [1, 3, 4]. In addition to these exercises in deterministic dynamics, the discovery of Brownian motion inched scientists and mathematicians towards the consideration of random processes such as the jiggling motions of pollen grains observed by Brown, in that case using the Langevin equation for example [22, 23]. Many of these dynamical descriptions share the common feature that there is some initial state of the system, and the dynamical laws tell us how to compute the subsequent evolution of the system in time.

But the dynamical laws described above left out one of the most important categories of dynamics in the natural world, namely, the strange dynamics in and of living organisms, processes including the long chains of energy-driven biochemical reactions, structure formation from macromolecular complexes to organelles, embryonic development and a myriad of other fascinating topics. It is fashionable to minimize the brilliant efforts of our predecessors when we later come to realize how "wrong" their ideas might have been. Within physics the 19th century idea of heat as a fluid, the famed caloric, or of the luminiferous ether, the medium within which electromagnetic disturbances were thought to propagate, are sometimes viewed as quaint and misguided. However, a more fair and nuanced view sees them instead as brilliant and creative examples of the continuing refinement that attends most scientific progress [1, 3, 4]. In the study of evolution, Lamarck's ideas are routinely dismissed as similarly misguided, though a more scholarly judgment is that like many others, Lamarck was engaged in thoughtful hypothesis making that was later learned to be inconsistent with observations and measurements. Perhaps no discredited topic draws such contempt as does that of vitalism. However, even here, a more generous interpretation of the 19th century vitalists is to acknowledge the obvious: living organisms seem different than their inanimate counterparts and clearly, the incomplete dynamical repertoire of classical physics might have something to gain by coming to terms with the purposeful dynamics of living organisms.

One way in which the study of the living might hold the secret to new principles is in the sense that the paradigm of Laplace's Demon [24], the strict unfolding of initial conditions through deterministic evolution, is often not respected by processes within living cells. Au contraire, cells often require a different kind of dynamics that is driven by the *final* conditions rather than the *initial* conditions and it is to this kind of processes that the remainder of our paper is dedicated. In particular, living organisms have developed ways of converting goals or functions into goal-oriented dynamical processes. In many everyday

human processes characterized by some purpose, goal or function, the standard paradigm is the construction of machines. We build machines to pump water, to cool materials [25], to move objects and the list goes on and on. By way of contrast, living organisms have used stochasticity in a novel way to project the future into the past, to enable random trajectories to "find" the goal or target in a process which we call exploratory dynamics [26]. Andrew Murray suggested to us an everyday occurrence that might help frame the uniqueness and weirdness of exploratory dynamics as shown in Figure 8.2. He asks us to imagine building a railroad line between New York City and Boston. We all know how this goes if done according to the strictures of modern civil engineering. There is a goal, there is a plan, the plan is implemented to achieve the goal. Surveyors do their work, steel is ordered, worked are deployed and paid to lay down track. However, if done according to the principles of biological exploratory dynamics, instead, construction workers will start in the two cities, randomly choosing a direction to start laying down track. If after a certain time, they have not encountered their counterparts coming from the other city, they remove the track and try again. That's all. For most of us, our immediate reaction is that this is woefully inefficient, perhaps even absurd. However, as seen in the examples of Figure 8.3, this kind of dynamics describes important biological processes over a dizzying array of spatial and temporal scales. Though this idea of exploratory dynamics was first clearly articulated in the context of the search and capture mechanism of chromosome segregation [27], in the time since a growing list of examples of this kind of dynamics has mounted [26]. As a result, we think we have reached a moment of timely rather than premature abstraction in which experience with these specific examples suggests that instead of being a particular mechanistic feature of chromosome search, exploratory dynamics is a much more general phenomenon, calling for its own abstractions, toy models and equations. In this paper, we examine one category of exploratory dynamics which features a series of abortive trajectories followed by a final successful trajectory.

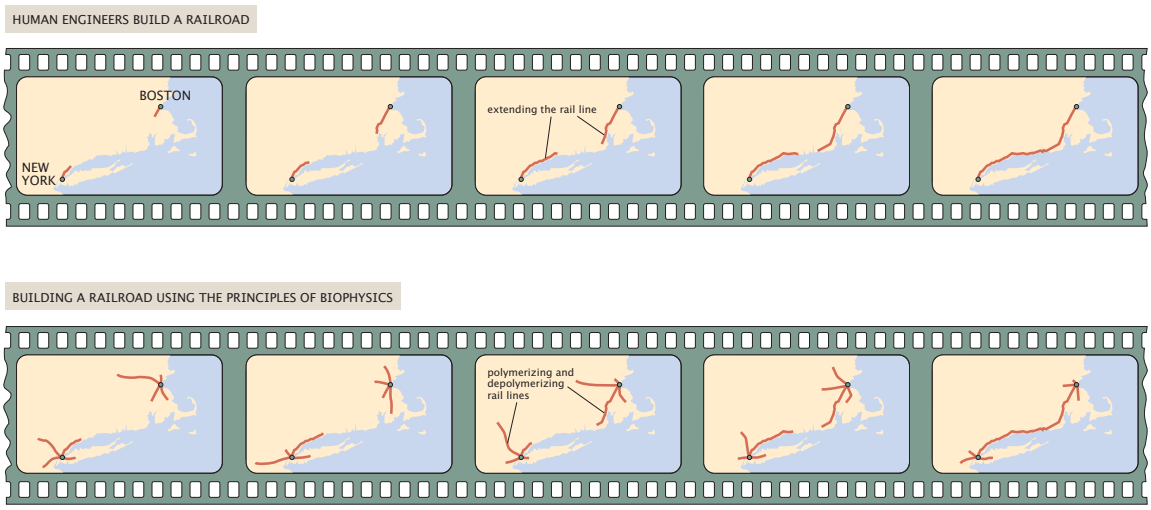


Figure 8.2: Comparison of human-scale dynamics and cellular-scale dynamics approaches to building a railroad between New York City and Boston. In both cases, there is a very specific goal. However, the strategy for achieving that goal is completely different in the human engineering context in comparison with the biological strategy of using stochastic variation and "selection" to achieve exploratory dynamics. The concept of the figure is due to a suggestion of Andrew Murray.

The organization of the paper is as follows. In section 8.2 we explain the concept of exploratory dynamics, first by describing biological processes that exemplify it and then by giving a cursory description of the kinds of theoretical ideas that have been put forth that might help explain such processes. Section 8.3 explains the statistical mechanics of exploratory trajectories using a simple zero-dimensional model. With those preliminaries in hand, section 8.4 describes the concrete problem of microtubules searching for chromosomes in the language of exploratory dynamics. Section 8.5 tackles a completely different biological phenomenom, namely, how transcription factors find their target site on DNA, revealing many mathematical similarities to the problem of the microtubule search problem. We note that in both of our case studies, we in no way pretend that we are the first to write down many of the models we explore, nor do we pretend our limited survey of the relevant literature to be comprehensive or encyclopedic. Rather, we spell them out in full detail in the hope of carefully illustrating the statistical mechanics of exploratory trajectories in a few limited examples in the hope that it might encourage

analysis of the full suite of examples shown in Figure 8.3, with the aim of more mathematical generality than the few examples considered here. We finish in section 8.6 with a high-level summary of the possible broader implications of exploratory dynamics and some final reflections on the career and influence of Prof. Erich Sackmann.

8.2 Exploratory Dynamics by Variation and Selection

In many cases, when we use or hear the words variation and selection, it is in the context of Darwinian evolution. However, the dynamics of variation and selection has much broader reach than in the context of traditional biological evolution. Perhaps surprisingly, there are many problems in molecular and cellular dynamics that can be couched in this same language. Examples range from the search of transcription factors for their binding sites [28, 29, 30, 31, 32]; microtubules being captured by kinetochores on chromosomes [33, 34, 35, 36, 37]; bacterial cells performing the biased random walks of chemotaxis [38, 39]; the formation of vasculature in developing embryos [40]; the formation of exquisitely precise neuronal connections [41]; the development of antibody repertoires [42, 43, 44]; animals forging for food sources [45, 46, 47, 48, 49, 50] and far beyond. In an earlier work, one of us (MK) in collaboration with John Gerhart wrote a book which considered living organisms in terms of some key principles that might describe broad swathes of biological phenomena [26]. Chapter 4 of that book entitled "The Exploratory Behavior of Biological Systems" described a series of case studies in biological phenomena over a wide range of spatial and temporal scales that could be viewed as being of an exploratory character. Those examples inspired the search for biophysical underpinnings to exploratory dynamics, and this paper is is our attempt to lay out several case studies in some mathematical detail.

Figure 8.3 provides a gallery of examples of biological processes that we think it might

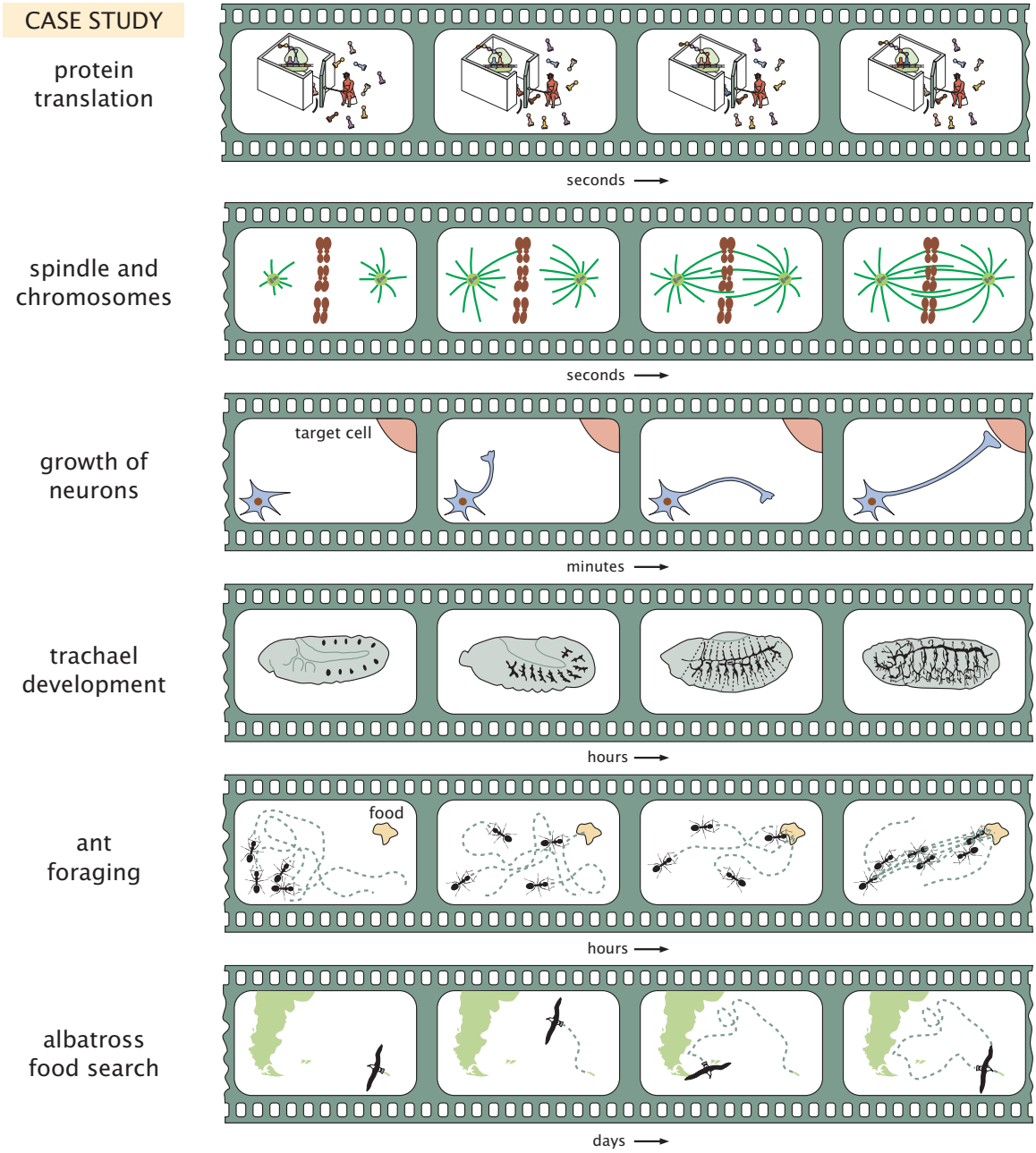


Figure 8.3: Gallery of exploratory dynamics. Examples from a wide variety of spatial and temporal scales illustrate the way in which the dynamics is characterized by a set of trajectories, all of which repeatedly fail to achieve their function before ultimately succeeding. Each example is explained in more detail in the text.

be fruitful to examine quantitatively from the perspective of exploratory dynamics. Though the formal definition of what constitutes exploratory dynamics might still be coming into focus, for the purpose of the present paper, we consider a process as exploratory if it is characterized by a well defined final state which is reached through a series of trials. In particular, we imagine a random process that produces a set of trajectories, a "standing variation" in trajectory space. A selection mechanism then picks out a subset of these trajectories, which are "functional." One can imagine refinements in which we ask whether the exploratory process requires energy investment such as in the case of chromosome search and capture, or not, as in the case of transcription factor search as considered later in the paper. In addition, one can also imagine situations in which there are cues that reinforce some trajectories that appear to be headed in the "right" direction such as is found in chemotaxis or the growth of vasculature.

To more concretely illustrate what we mean by exploratory dynamics, the first panel of Figure 8.3 reminds us that the process of protein translation is exploratory in the sense that repeatedly, the "wrong" amino acid is presented to the ribosome and only through this kind of biased trial-and-error dynamics does a protein get translated with high fidelity [51, 52]. In most cases in which a wrong tRNA arrives at the ribosome, there is a reset that restores the system back to its starting state. In this case, the "variation" refers to at least at 20:1 ratio of wrong to right tRNAs arriving at the ribosome, and the "selection" refers to the incorporation of the amino acid of interest into the nascent polypeptide chain. As shown in the second panel, one of the most compelling examples of exploratory dynamics is the dynamics of microtubules which precedes chromosome segregation in dividing cells [33, 34, 35, 36, 37]. Here, microtubules grow and shrink repeatedly from an organizing center known as the centrosome, and those microtubules that find the kinetochore on the chromosomes stay attached to their target. A reframing of the problem as one in exploratory dynamics is schematized in Figure 8.4. The third example shows the

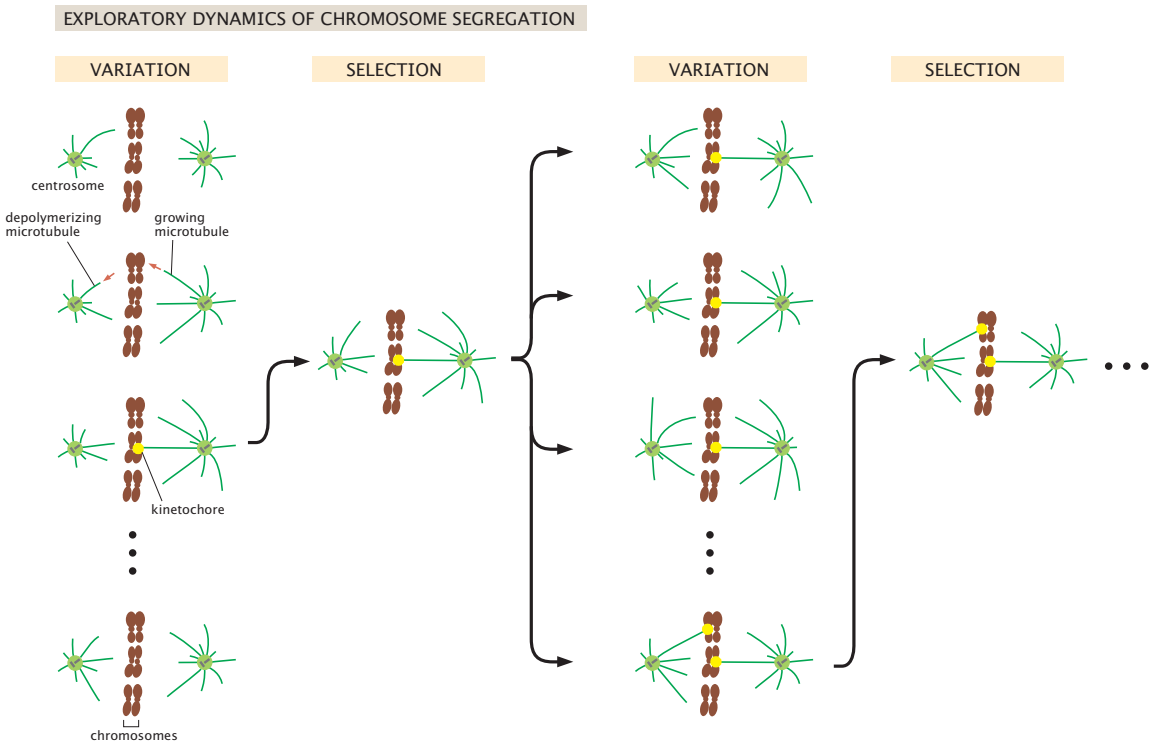


Figure 8.4: Microtubule search for kinetochores as exploratory dynamics characterized by stochastic variation in the dynamical trajectories of the system coupled with selection and stabilization of those trajectories that culminate in attachment to kinetochores.

growth of neurons, where connections come and go and are "selected" in some contexts as illustrated in the example of olfactory selection cited here [41]. In the developing fly embryo, though these animals don't have vasculature in the same sense as mammals, they do have a system of "pipes" that allow for the transport of oxygen to cells and the establishment of this spatial arrangement of pipes is itself exploratory [40]. More generally, the subject of branching morphogenesis provides a broad array of examples of exploratory dynamics [53, 54]. At much larger scales; animal foraging can be thought of as its own kind of exploratory dynamics [45, 46, 47, 48, 49, 50]. The main point of the figure and this discussion is simply to capture the reader's imagination with the hypothesis that perhaps exploratory dynamics is a new and important part of the dynamics repertoire that has thus far been understudied in general terms from the physics perspective.

Our discussion thus far primarily focused on the biological phenomenology of exploratory dynamics. Interestingly, in parallel, the worlds of both physics and mathematics had developments of their own aimed at solving problems in purposeful dynamics. Though it might seem far afield, the quest of the Allied Forces to search for German U-Boats gave rise to the inspiring and useful field of operations research and search theory [55, 56, 57]. Ironically, several of the 20th century's most famous biologists were key participants in the development of the mathematical and physical tools to treat search problems [58], including John Kendrew, noted for his resolution of the molecular structure of whale myoglobin [59]. Perhaps more surprisingly is that at the age of 67, CH Waddington, of Waddington landscape (and many other things) fame, wrote a profound text on the use of operations research during the Second World War of which he, along with Kendrew, was a key participant [60]. A parallel avenue of theoretical research that we think is quite relevant to exploratory dynamics has focused on random walks with resets [61, 62]. Yet another set of ideas were introduced that have been dubbed infotaxis [63, 64, 65]. In the remainder of the paper, we focus on a specific approach in which we describe exploratory dynamics as a statistical mechanics of a particular class of trajectories, but we are anxious to see what theorists will have to say about exploratory dynamics in coming years.

8.3 Zero-Dimensional Exploratory Dynamics

As noted above, one example of exploratory dynamics that is enormously consequential is the way in which microtubules pair up with chromosomes in preparation for the process of chromosome segregation as shown schematically in Figure 8.4. In this paper, we use this example as a specific case study that illustrates what are perhaps more general ideas of how to formulate exploratory dynamics as a statistical mechanics of trajectories. In this statistical

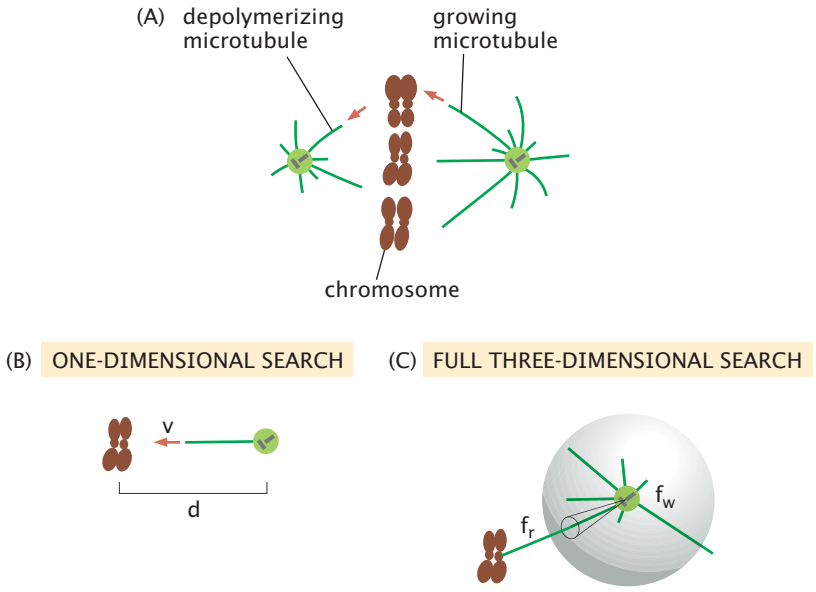


Figure 8.5: Schematic of the process of microtubules searching for chromosomes. (A) Schematic showing how microtubules grow out from the centrosome and then suffer catastrophes. (B) One-dimensional idealization of the process of microtubules searching for a chromosome. We imagine that the distance from the centrosome to the kinetochore is d and that the polymerizing microtubule grows with speed v. All chromosomes grow in the same direction and the exploratory character is in how long the microtubule survives before suffering a catastrophe. (C) Three-dimensional idealization of the process of microtubules searching for a chromosome. f_r is the probability of the microtubule growing in the right direction and $f_w = 1 - f_r$ is the probability of growing in the wrong direction. The right direction is defined by a cone of solid angle such that the microtubule will grow in a range of allowed directions that will contact the kinetochore.

mechanics of exploratory trajectories, all trajectories are characterized by the fact that they end when some "purpose" is fulfilled or some function is achieved. As we show below, in each case considered in this paper, this results in the use of the geometric distribution. We suspect that there are much more general ideas waiting to be discovered that can account for cases in which subsequent trajectories have a memory of those that have come before or in which signals reinforce trajectories that are headed in the right direction.

As a first model, we set up the mathematical language of the statistical mechanics of exploratory trajectories by considering trajectories built up of "wrong" and "right" moves. For the moment, we are noncommittal about the moves themselves since as we will see later in the paper, in different contexts, "move" will refer to the duration of the

polymerization process, the direction of polymerization or to changing between one-dimensional and three-dimensional diffusion in the case of transcription factors searching for their binding sites. All the trajectories have the same basic structure. A series of wrong moves followed by a final right move. At each time step, the wrong and right moves are each chosen with probability f_w and f_r, respectively, subject to the condition that $f_w + f_r = 1$. A given exploratory trajectory is assembled as a collection of i unsuccessful steps before a final successful step. The probability of such a trajectory is given by

$$p_i = \underbrace{f_w f_w \cdots f_w}_{\text{i}} f_r = f_w^i f_r. \tag{8.2}$$

We can rewrite this more intuitively as

$$p_i = (1 - f_r)^i f_r, \tag{8.3}$$

a probability distribution known more colloquially as the geometric distribution [66]. This fundamental equation has a structure that we will use over and over in thinking about the statistical mechanics of exploratory trajectories since it highlights the importance of the final state.

Since we will need to manipulate these distributions repeatedly, here we use it to compute the average waiting time until the exploration is complete. We begin by verifying that the distribution is normalized by computing

$$\sum_{i=0}^{\infty} p_i = \sum_{i=0}^{\infty} (1 - f_r)^i f_r = f_r \sum_{i=0}^{\infty} (1 - f_r)^i. \tag{8.4}$$

We recognize the remaining sum as a geometric series with the property that $\sum_{i=0}^{\infty} x^i =$

$1/(1-x)$ if $0 < x < 1$. Since we have $0 < 1 - f_r < 1$, our case yields

$$\sum_{i=0}^{\infty} p_i = f_r \sum_{i=0}^{\infty} (1-f_r)^i = \frac{f_r}{(1-(1-f_r))} = 1. \tag{8.5}$$

As expected, the distribution is normalized.

A more important question is how long does it take on average for the exploratory dynamics to terminate due to success in finding its target? The answer to that question is obtained by averaging over the time spent in all the possible trajectories and given by

$$\langle t \rangle = \sum_{i=0}^{\infty} (\underbrace{i\, t_{\text{step}}}_{\text{failures}} + \underbrace{t_{\text{step}}}_{\text{success}}) p_i, \tag{8.6}$$

where we have introduced the quantity t_{step} as the duration of each instance where the system makes an unsuccessful excursion (e.g. a microtubule polymerizes without finding the chromosome). Later in the paper, we will explicitly calculate t_{step} on the basis of the microscopic parameters associated with the microtubule polymerization problem or the transcription factor search problem. The second term in the sum immediately yields t_{step} since the distribution is normalized leaving us with

$$\langle t \rangle = \sum_{i=0}^{\infty} i\, t_{\text{step}}\, p_i + t_{\text{step}}. \tag{8.7}$$

Next, we need to evaluate

$$\langle t \rangle = t_{\text{step}} \sum_{i=0}^{\infty} i\, p_i + t_{\text{step}} = t_{\text{step}}\, f_r \sum_{i=0}^{\infty} i(1-f_r)^i + t_{\text{step}}, \tag{8.8}$$

which we recognize as having a term of the form

$$\sum_{i=0}^{\infty} i x^i = x \frac{d}{dx} \sum_{i=1}^{\infty} x^i = x \frac{d}{dx} \frac{1}{1-x} = \frac{x}{(1-x)^2}. \tag{8.9}$$

We now see that we can rewrite eqn. 8.8 as

$$\langle t \rangle = t_{\text{step}} f_r \sum_{i=0}^{\infty} i(1-f_r)^i + t_{\text{step}} = t_{\text{step}} (1-x) \sum_{i=0}^{\infty} i x^i + t_{\text{step}}, \tag{8.10}$$

where we have used the definition $x = 1 - f_r$. We now invoke the derivative trick introduced in eqn. 8.9 to arrive at

$$\langle t \rangle = t_{\text{step}} (1-x) \frac{x}{(1-x)^2} + t_{\text{step}} = t_{\text{step}} \frac{x}{(1-x)} + t_{\text{step}}. \tag{8.11}$$

If we now write this expression in terms of f_r, we find the very intuitive result that

$$\langle t \rangle = t_{\text{step}} \left(1 + \frac{1-f_r}{f_r} \right) = \frac{t_{\text{step}}}{f_r}. \tag{8.12}$$

The intuition for this result is best served by concretely imagining that $f_r = 1/M$ which means that 1 out of every M trials is successful. This implies in turn that $\langle t \rangle = M\, t_{\text{step}}$, stating that on average we need to carry out M trials for the exploration to be a success when the probability of success on a given trajectory is $f_r = 1/M$.

Our reason for going through this example is to showcase the structure of the result, since in all of our examples of exploratory dynamics, we will ultimately characterize the exploratory process as a statistical mechanics of trajectories, with all such trajectories sharing the feature that they involve repeated "failures" until they succeed on the final trajectory. Interestingly, for the examples considered throughout the paper, we will see many different realizations of the space of trajectories, but as noted above, they all are of the form $p_i = \underbrace{f_w f_w \cdots f_w}_{\text{i}} f_r$.

8.4 Chromosome Search Via Microtubule Reset Dynamics

8.4.1 One-Dimensional Chromosome Capture by Exploratory Dynamics

As our next case study, we consider the one-dimensional model summarized in Figure 8.5(B) where we idealize the growth and catastrophes as taking place in one dimension and ask for the time it takes for the microtubule that is subject to both growth and shrinkage to find its chromosomal target. This analysis largely imitates the beautiful work of Holy and Leibler, especially as developed in their Appendix B [34] as well as more recent work from Wollman et al. [36] and their thorough and insightful Supplementary Information. In the spirit of successively building up models of increasing sophistication, we begin with the unrealistic assumption that every microtubule grows in the right direction and ask only whether it survives long enough to reach the chromosome. One way to think about this model is that the exploratory behavior only exists in the sense that the different exploratory periods have different durations. Later, we will extend this framework to consider the case in which there is also a spatial component to the search process with different microtubules heading off in different directions, lending a second facet to the exploratory nature of the dynamics.

Before we can compute the time scale associated with the chromosomal search process, we must first consider the waiting time distribution that tells us how long a microtubule grows before it suffers a catastrophe. To answer that question, we turn once again to the geometric distribution. If the rate constant for a catastrophe is k_c, then the probability in a time step Δt that the microtubule will *not* suffer a catastrophe is $(1 - k_c \Delta t)$. Hence, if the microtubule is to pass through N time steps before having a catastrophe, the probability of that succession of events is

$$p_N = (1 - k_c \Delta t)^N k_c \Delta t. \tag{8.13}$$

Because of the beautiful properties of this distribution in the large N limit, namely

$$\lim_{N\to\infty}\left(1-\frac{x}{N}\right)^N = e^{-x}, \tag{8.14}$$

and using the fact that the total time elapsed is $t = N\Delta t$, we see that the probability that the microtubule will suffer a catastrophe in the time interval between t and $t + dt$ is given by

$$p_{\text{survive}}(t)dt = k_c e^{-k_c t} dt. \tag{8.15}$$

The most naive model of the average time before a microtubule suffers a catastrophe is given by $t_c = 1/k_c$ which we find by evaluating the average

$$t_c = \int_0^\infty t k_c e^{-k_c t} dt = \frac{1}{k_c}. \tag{8.16}$$

However, the average time before a catastrophe is actually shorter. Note that if the distance from the initial position of the microtubule to the chromosome is d and it grows at a rate v, this implies that if the microtubule grows for a time longer than $\tau = d/v$, the microtubule will have hit the chromosome, thus ending the exploratory set of trajectories. Hence, we need to be more careful in assigning a time scale to the failed trajectories. Instead, we should sum only over those trajectories that last a time shorter than τ. In particular, the average time of the failed trajectories is given by

$$t_f = \frac{\int_0^\tau e^{-k_c t} k_c t dt}{\int_0^\tau e^{-k_c t} k_c dt}, \tag{8.17}$$

where we only allow those trajectories that are shorter than τ and we have introduced the notation t_f with subscript f to denote "failed" trajectories. The denominator imposes

normalization on the distribution. These integrals can be evaluated to yield

$$t_f = \frac{1}{k_c} \frac{1 - (k_c\tau + 1)\, e^{-k_c\tau}}{1 - e^{-k_c\tau}}. \tag{8.18}$$

Note that this average time until a catastrophe is shorter than the naive estimate of $1/k_c$ derived above.

Given the survival distribution, the probability that the microtubule will survive long enough to reach the chromosome is given by

$$p_s = \int_\tau^\infty k_c e^{-k_c t}\, dt = e^{-k_c d/v}. \tag{8.19}$$

The integral computes the fraction of the probability distribution where the survival of the microtubule lasts longer than the time to reach the chromosome, $\tau = d/v$, where again, d is the distance to the microtubule and v is the velocity of microtubule growth. Our goal is to find the average time it takes for the growing and shrinking microtubule to reach the chromosome. The probability of success after i failures is once again given by the geometric distribution which we now write as

$$p_i = \left(1 - p_s\right)^i p_s \tag{8.20}$$

which reflects the fact that for i of the growth processes, the microtubule suffers a catastrophe before reaching the chromosome and only on the $i + 1^{th}$ try does it reach the chromosome with probability p_s. A variety of examples of these kinds of exploratory trajectories are shown in Figure 8.6. The time taken for this particular class of trajectories is given by $\tau + it_f$ since on average each of the failed growth and catastrophe events takes a time t_f and the final successful growth event takes a time τ.

The logic of our calculation of the average "exploration time" is that we need to sum over

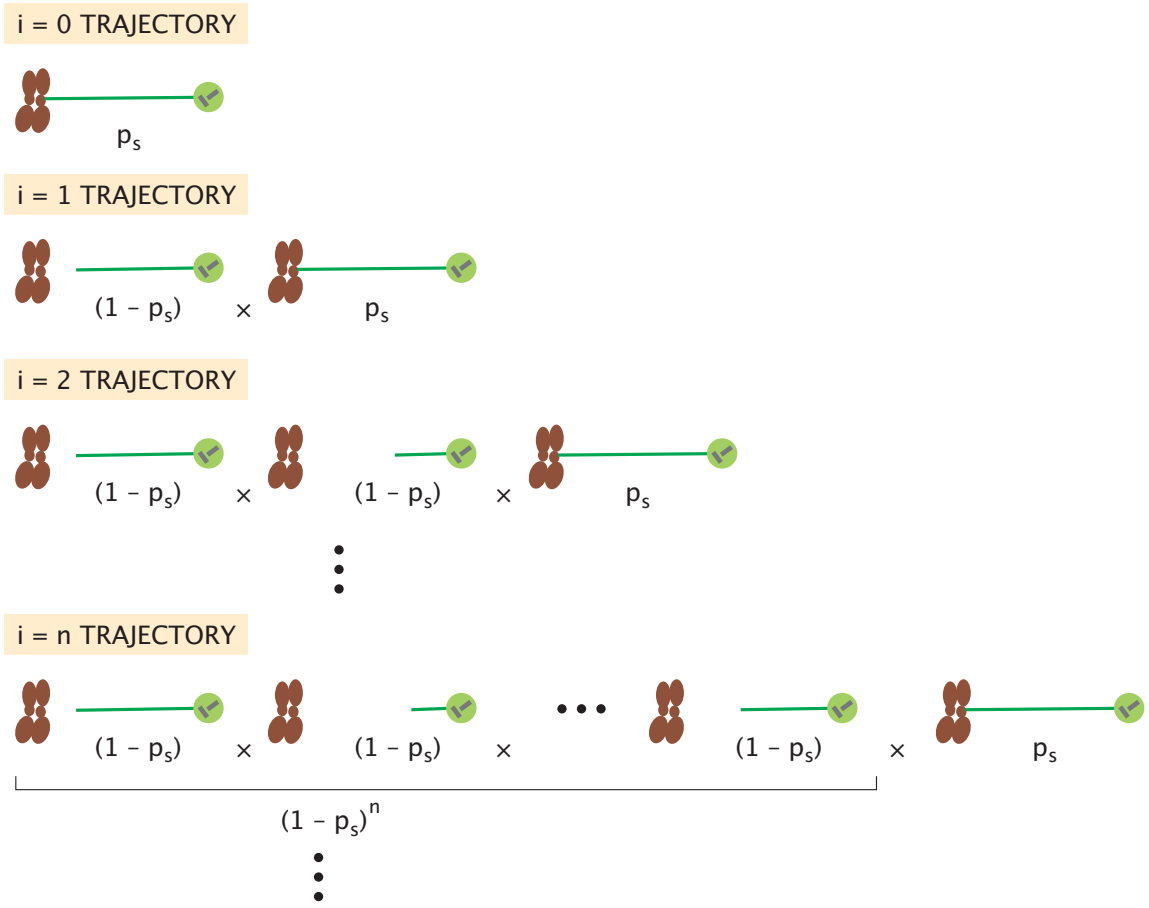


Figure 8.6: Trajectories and weights for one-dimensional model of chromosome search and capture. The index "*i*" labels the number of times that the microtubule *fails* to hit the chromosome before its eventual success. The "failed" trajectories are characterized by the fact that the microtubule suffers a catastrophe before reaching the chromosome. The failed trajectories show the maximum length of the microtubule before suffering a catastrophe.

all trajectories, every possible life history of the growing and shrinking microtubules that ends by ultimately finding the chromosome. In words, this means we have to sum over the case where the microtubule finds the chromosome on the first try, the second try, the third try and so on. Given this procedure, the average time to find the chromosome is given by

$$\langle t \rangle = \sum_{i=0}^{\infty} (\tau + i t_f) p_i. \tag{8.21}$$

Using the probability of success on the $(i+1)^{th}$ trial introduced in eqn. 8.20, we can rewrite the average exploration time as

$$\langle t \rangle = \sum_{i=0}^{\infty} (\tau + i t_f) \left(1 - p_s\right)^i p_s, \tag{8.22}$$

which is helpfully rewritten as

$$\langle t \rangle = \sum_{i=0}^{\infty} \tau \left(1 - p_s\right)^i p_s + \sum_{i=0}^{\infty} i t_f \left(1 - p_s\right)^i p_s. \tag{8.23}$$

The first of these sums yields τ since the geometric distribution is normalized. For the second sum, by exploiting the very useful derivative protocol for evaluating such sums that we already introduced in eqn. 8.9, we can write the second sum as

$$\sum_{i=0}^{\infty} i\, t_f \left(1 - p_s\right)^i p_s = t_f\, p_s \frac{(1 - p_s)}{p_s^2}. \tag{8.24}$$

Hence, the total search time is given by

$$\langle t \rangle = \tau + t_f \frac{(1 - p_s)}{p_s} \tag{8.25}$$

If we recall that $p_s = e^{-k_c d/v}$ and the result for t_f in eqn. 8.18, we find

$$\langle t \rangle = \tau + \frac{1}{k_c} \frac{1 - e^{-k_c \tau}(k_c \tau + 1)}{1 - e^{-k_c \tau}} \frac{\left(1 - e^{-k_c \tau}\right)}{e^{-k_c \tau}}. \tag{8.26}$$

Noting the cancellations of the terms $1 - e^{-k_c \tau}$ in the numerator and denominator, this simplifies to the lovely final result

$$\langle t \rangle = \frac{1}{k_c} \left(e^{k_c \tau} - 1 \right). \tag{8.27}$$

There are a variety of interesting limits to this result that are useful to consider. For example, in the case where the time to suffer a catastrophe is much larger than d/v, the microtubule will hit the chromosome on first try and the average time is d/v. To see this limit in practice when $k_c \tau \ll 1$, we Taylor expand the exponential resulting in

$$\langle t \rangle \approx \frac{1}{k_c} (1 + k_c \tau - 1) \approx \tau. \tag{8.28}$$

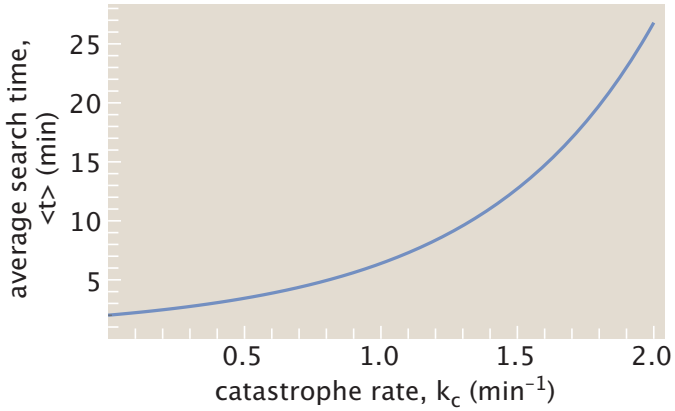


Figure 8.7: Search time as a function of the catastrophe rate for the one-dimensional model of chromosome search and capture. For this case $\tau = 2$ min.

Figure 8.7 shows the average search time as a function of the catastrophe rate k_c. The minimum in this curve is found when the catastrophe rate goes to zero. We can see this more formally by evaluating

$$\frac{d\langle t\rangle}{dk_c} = -\frac{1}{k_c^2}\left(e^{k_c\tau} - 1\right) + \frac{\tau}{k_c}e^{k_c\tau} = 0 \tag{8.29}$$

implying that the minimum occurs at

$$k_c = \frac{1}{\tau}\left(1 - e^{-k_c\tau}\right), \tag{8.30}$$

which is satisfied when $k_c\tau = 0$, meaning that the catastrophe rate itself is zero. The best strategy for this particularly simple model is to have no catastrophes. Let's now consider the more serious model that allows for a three-dimensional search [34, 36].

8.4.2 Three-Dimensional Chromosome Capture by Exploratory Dynamics

The calculation given above made the unrealistic assumption that every microtubule started out in the right direction, and thus we only needed to find out whether such

microtubules lasted long enough to reach the chromosome before suffering a catastrophe. We now examine the case in which every time a microtubule starts anew, it has a probability f_r of going in the right direction to meet the chromosome, and a probability $f_w = 1 - f_r$ of going in the wrong direction. This concept is illustrated schematically in Figure 8.5(C).

As with the previous examples, we need to write an expression for the probabilities of all the different kinds of trajectories as indicated in Figure 8.8. For concreteness, let's consider those trajectories that involve n unsuccessful polymerization events before a success on the $(n+1)^{th}$ try. However, there is a new twist relative to the calculation of the previous section. Now our trajectories can fail either because they are going in the right direction, but don't last long enough to reach the chromosome or alternatively, they are going in the wrong direction and no matter how long they last, they will not find the chromosome. As a result, we introduce the notation $p_i(n)$ for those trajectories that fail n times before a success, with i of those failures being in the right direction, but of insufficient duration, and $n - i$ of those failures going in the wrong direction. The probability of a trajectory in the right direction, but that does not last long enough to reach the chromosome is given by $f_r\left(1 - p_s\right)$ while the probability of a trajectory in the wrong direction is given by $\left(1 - f_r\right)$. Given these partial probabilities, we can now write the probability of $p_i(n)$ of this category of trajectories as

$$p_i(n) = \frac{n!}{i!\,(n-i)!} \underbrace{f_r^i \left(1 - p_s\right)^i}_{\text{right direction}} \underbrace{\left(1 - f_r\right)^{n-i}}_{\text{wrong direction}} f_r p_s, \tag{8.31}$$

where we see that out of the n unsuccessful trajectories, i of them went in the right direction and $n - i$ of them went in the wrong direction. The prefactor counts the number of ways that out of these n trajectories, i of them will be in the right direction and $n - i$ of them will be in the wrong direction.

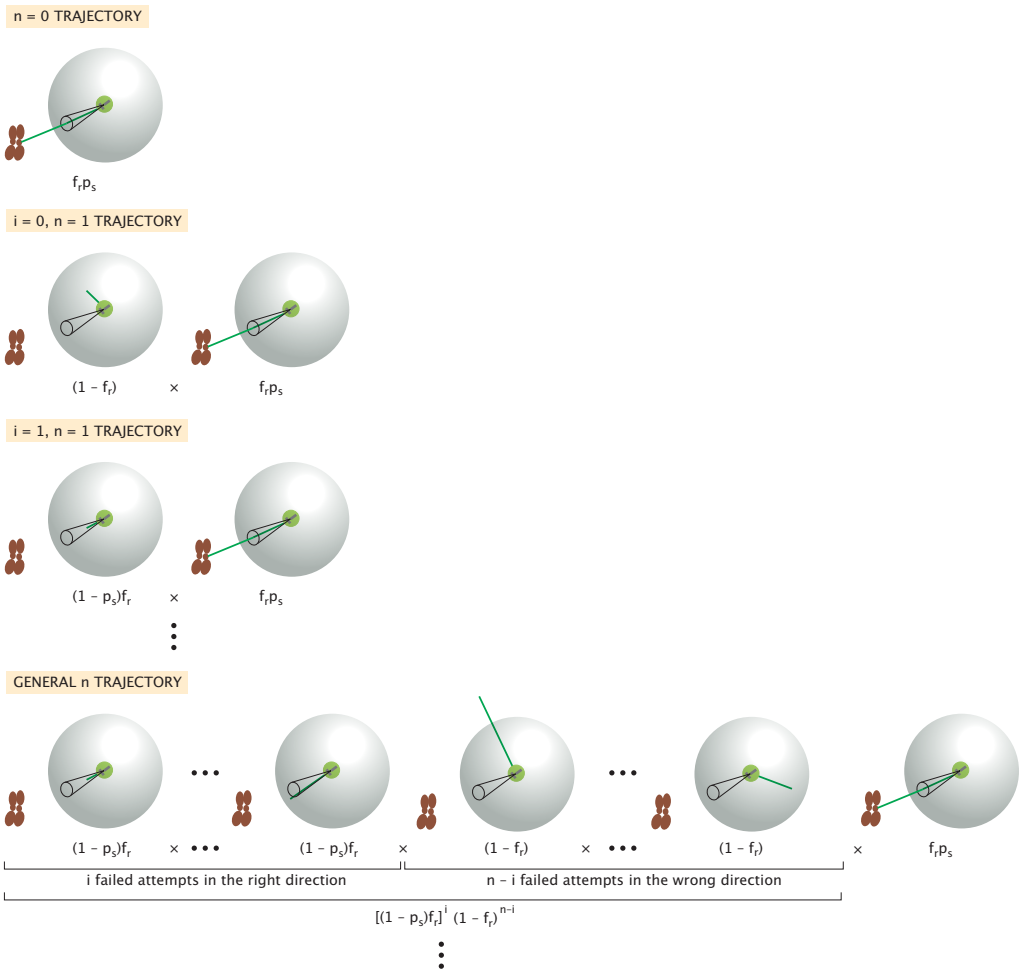


Figure 8.8: Trajectories and weights for three-dimensional model of chromosome search and capture. Trajectories are labeled by the label n which tells how many failed trajectories there were and the label i which tells us out of the n failed trajectories, how many of them were in the right direction. Each such trajectory has an associated probability and time until success and to find the average time we sum over the infinite set of trajectories as $n \longrightarrow \infty$.

The time associated with a trajectory involving n unsuccessful polymerization events before a final successful polymerization event is given by

$$t_i(n) = i\,t_r + (n-i)t_w + \tau. \tag{8.32}$$

Here we acknowledge that i of the unsuccessful trajectories go in the right direction and have an average lifetime t_r, $n-i$ of the unsuccessful trajectories go in the wrong direction and have an average lifetime t_w and the final successful trajectory has a time τ. As a reminder, recall that

$$t_w = \frac{1}{k_c}, \tag{8.33}$$

since these trajectories are of unrestricted length and t_r is given by the expression in eqn. 8.18 since these trajectories are constrained to be less than $\tau = d/v$.

Now we need to sum over all possible trajectories which we can organize successively by considering all trajectories with 0 failures, all with 1 failure, all with 2 failures $\cdots$ and all with n failures and so on, resulting in

$$\langle t \rangle = \sum_{n=0}^{\infty} \sum_{i=0}^{n} t_i(n) p_i(n). \tag{8.34}$$

More explicitly, this can be written as

$$\langle t \rangle = \sum_{n=0}^{\infty} \sum_{i=0}^{n} \frac{n!}{i!\,(n-i)!} f_r^i \left(1-p_s\right)^i \left(1-f_r\right)^{n-i} f_r p_s \times \left[i t_r + (n-i)t_w + \tau\right]. \tag{8.35}$$

Using all of the tools already introduced throughout the paper, all of these sums are of the form $\sum_{i=0}^{\infty} i x^i$ which we can evaluate to yield

$$\langle t \rangle = \tau + \frac{1}{p_s f_r}\left(t_r\left(1-p_s\right) f_r + t_w\left(1-f_r\right)\right) \tag{8.36}$$

which simplifies to the remarkably compact result

$$\langle t \rangle = \frac{1}{k_c}\left(\frac{e^{k_c\tau}}{f_r} - 1\right), \tag{8.37}$$

nearly identical to the earlier result of eqn. 8.27. This expression is plotted in Figure 8.9(A) where we see that there is an optimal catastrophe rate to achieve finding the chromosomes in the least time.

It is of interest to explore this expression in various limits to see if it jibes with our intuition. We note that in the case when $f_r = 1$, we recover precisely the same result we found earlier as eqn. 8.27, meaning that in the case where every microtubule starts off in the right direction, the present expression reduces to the earlier equation based upon the simplified model of Figure 8.5(B). As we did before, we can examine this equation for the search time to see if we can understand where it takes its minimum with respect to the catastrophe rate. To that end, we evaluate

$$\frac{d\langle t \rangle}{dk_c} = -\frac{1}{k_c^2}\left[\frac{e^{k_c\tau}}{f_r} - 1\right] + \frac{1}{k_c}\left[\frac{\tau}{f_r}e^{k_c\tau}\right] = 0 \tag{8.38}$$

which simplifies to the condition

$$\frac{1}{k_c^2}\left[\frac{e^{k_c\tau}}{f_r} - 1\right] = \frac{\tau}{k_c f_r}e^{k_c\tau}. \tag{8.39}$$

If we introduce the notation $x = k_c\tau$, then this condition simplifies further to

$$f_r = e^x(1 - x). \tag{8.40}$$

In Figure 8.9(B), we plot both sides of this equation and note that the solution corresponds to the point of intersection of the two curves, the horizontal line representing the left side, f_r = constant and the curve corresponding to the right side $e^x(1 - x)$. However, since we know that $f_r \ll 1$ because the microtubules polymerizing in the "right" direction is rare,

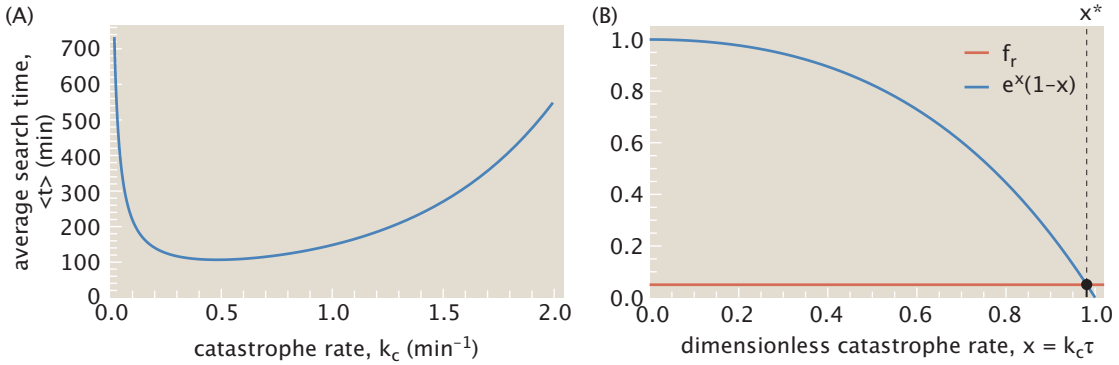


Figure 8.9: Optimal search time as a function of the catastrophe rate. (A) Time to capture the chromosome as a function of the catastrophe rate. (B) Plot of the two sides of the equation for the optimal search time (eqn. 8.40) as a function of the catastrophe rate k_c. The parameters used here are: $f_r = 0.05$, $d = 5$ μm, $v = 10$ μm/min implying $\tau = v/d = 2$ min.

this licenses defining the small parameter ε via $x = 1 - \varepsilon$ resulting in

$$f_r = e^{(1-\varepsilon)}(1 - (1 - \varepsilon)) \tag{8.41}$$

Now because of the smallness of the parameter ε, we can Taylor expand the exponential resulting in the very simple result

$$\varepsilon \approx \frac{f_r}{e}. \tag{8.42}$$

Recalling that $x = k_c\tau = 1 - \varepsilon$, we now have an approximate value for the optimal catastrophe rate given by

$$k_c \approx \frac{1}{\tau}\left(1 - \frac{f_r}{e}\right). \tag{8.43}$$

Given the fact that nearly all directions are "wrong" this implies that $f_r \ll 1$ (e.g. we used $f_r = 0.05$ in our plots), we end with the very simple result that the optimal catastrophe rate corresponds to $k_c \approx 1/\tau$.

Note that our treatment of the dynamics thus far has considered trajectories in which the lifetimes associated with the various processes were all average lifetimes. A more careful treatment of the exploratory dynamics would sum over *all* possible lifetimes for each and

every step. This would mean that instead of considering every right or wrong trajectory as having a lifetime, t_r or t_w, respectively, we would sum over *every* allowed lifetime. This would turn our sums over trajectories into integrals over trajectories, but ultimately would not change the results obtained above.

8.4.3 Molecular Control of Exploratory Dynamics

The description of exploratory dynamics considered thus far in the paper is decidedly incomplete. There is a whole pageant of accessory proteins that are part of the process of chromosomal search and capture that we have not mentioned at all since the primary focus of our paper is to focus on general ideas about exploratory dynamics. Several excellent reviews give a scholarly assessment of the complexities of the process of chromosomal search [37, 67]. Further, several earlier theoretical approaches examined precisely these kinds of elaborations on the most naive version of the search and capture model as exploratory dynamics [34, 36]. The way we think about the many molecular interventions implicated in the low error rate construction of the spindle of dividing cells [68] is that the parameters k_c and v used throughout our discussion are not fixed, but rather are subject to molecular control. For example, we know that there are a variety of proteins that can either speed up or hinder the polymerization of cytoskeletal filaments [67], which the model accounts for by having the rate v be a function of concentration and state of modification of these regulatory proteins.

This brief section will focus on how parameters such as the catastrophe rate k_c and the microtubule growth rate v might be controlled by specific molecular interventions. One reason for our interest in this topic is because we wonder how the search time depends upon the size of the cell itself, a topic that was examined carefully by Wollman et al. [36]. In particular, motivated by the idea that in some organisms during embryonic development there are reductive cell divisions which continually reduce the size of the cells. This then has the effect of changing the distance between chromosomes and the centrosome, which

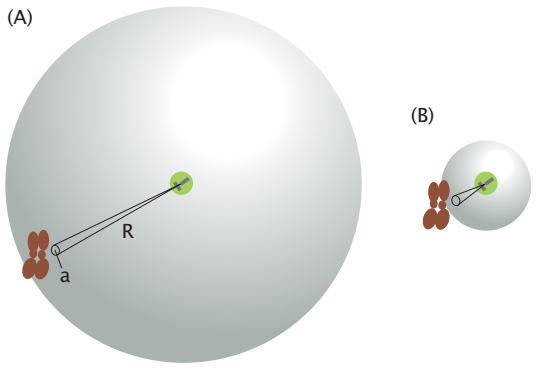


Figure 8.10: The geometry of chromosome search and cell size. (A) For a large cell size, the cone of successful directions is characterized by the probability $f_r = a/4\pi R^2$. (B) For a smaller size, f_r is larger.

suggests that the search time will change as well, being shorter in smaller cells. Figure 8.10 illustrates this idea by showing how the solid angle cone that characterizes growth in the "right" direction depends upon cell size. As cells get smaller, the area of the correct solid angle becomes a larger fraction of the total area.

To estimate how cell size, and hence distance between centrosome and kinetochore, alters the search time, we begin by considering that the size of the kinetochore can be measured by an area a (independent of the cell size) as shown in Figure 8.10. Hence, the fraction of trajectories which are starting in the right direction is given by

$$f_r = \frac{a}{4\pi R^2}, \tag{8.44}$$

where R is a measure of the cell size or the distance from centrosome to kinetochore. In light of this parameter, we can rewrite the search time given in eqn. 8.37 as

$$\langle t \rangle = \frac{1}{k_c}\left(\frac{4\pi R^2}{a} e^{R/\lambda} - 1\right), \tag{8.45}$$

where we have introduced the length scale $\lambda = v/k_c$ which emerges naturally as a key parameter in the problem. Specifically, λ enjoys the interpretation of the typical length

grown by a microtubule before catastrophizing, so R/λ is a dimensionless ratio characterizing how large the cell size is compared to typical growth excursions of the microtubule. An immediate and simple consequence of this model is that the search time is exponentially sensitive to changes in cell size, assuming that R/λ is large. In the opposite case, when R/λ is small, the exponential is roughly equal to one. In this case the search time grows quadratically with cell size. Either way, the simple model for kinetochore search we have introduced predicts that the search time will depend sensitively on the cell size.

Our examination of the simple model of kinetochore search is motivated primarily by the goal of revealing a larger set of problems that we feel are usefully described by the ideas of exploratory dynamics. In the specific context of chromosome capture the inadequacies of this model have been examined previously [36]. In particular, the model described above predicts search times that are significantly longer than those measured experimentally [36]. We now review the approaches that have been taken to account for discrepancies with what is known experimentally. Here to demonstrate the utility of even this simple a model we start from the large body of literature, very well reviewed in [37, 67], which explores the spatial and temporal scaling of the spindles of dividing cells. As we will touch on below, the simplest overarching observation is that the sensitive dependence of search time on cell size described above is not consistent with the observed behavior.

Remarkably, in a variety of experiments on developing embryos and encapsulated *Xenopus* egg extracts, the search time is roughly constant (or more precisely, the spindle assembly time), independent of cell size [69, 70]. For a concrete example, see Figure 6 of [69]. In the embryo, different cell sizes result from reductive cell divisions, while in the synthetic system the experimentalist controls the size of the encapsulating vesicle, which plays the role of a synthetic cell. These experimental observations on the lack of a cell-size dependence to the search time are at odds with the result of eqn. 8.45. This fact invites us

in the remainder of this section to consider how the two rate parameters, v and k_c, should scale with the cell size [71, 34, 36, 67] so that the search time will be independent of R. As cogently described by Belmont et al. [71], the dynamic instability model features four rate parameters: the rates of polymerization and depolymerization, and the rates of catastrophe and rescue. In the highly schematized model described here, we only preserved two of those four rates, namely, the catastrophe rate k_c and the polymerization rate associated with the speed of growth v, since we ignored rescues altogether.

Based on eqn. 8.45, to prevent the search time from growing exponentially with cell size (when the cell is big compared to typical microtubule excursions so the dimensionless parameter R/λ is large), the typical length the microtubules reach before collapsing must scale linearly with cell size. That is, $\lambda = v/k_c \propto R$, where the $\propto$ symbol represents the idea that if the quantity on the right hand side (i.e. R) were to be doubled or halved, the same fate would befall the quantity on the left hand side (i.e. v/k_c). We are still left with the term that is quadratic in R and in our simple model we would further require $k_c \propto R^2$ in order for the search time to be independent of R. It is interesting to note that the only way to satisfy both scaling requirements is for the speed of polymerization to be proportional to R^3, which is the same as saying that it is proportional to the cell volume. This scaling is exactly what was found experimentally in [69], and corresponds to a constant speed of polymerization per cell volume.

Another parameter that we have thus far not touched upon is the number of microtubules. Thus far, the entirety of our discussion has knowingly focused on the single microtubule case. But of course, as is immediately obvious in the most cursory examination of dividing cells, there are many microtubules present in the spindle [72]! More recently a set of beautiful experiments revealed that the scaling of spindle size with cell size during embryonic development is due to a finely-regulated number of microtubules in the spindle [70]. While our simple model does not consider this

parameter, the presence of multiple microtubules will decrease the search time in proportion to the number of microtubules, assuming that each one searches independently; this picture results in a revised search time for at least one microtubule to reach the kinetochore of,

$$\langle t \rangle = \frac{1}{N k_c} \left(\frac{4\pi R^2}{a} e^{R/\lambda} - 1 \right), \tag{8.46}$$

where N is the number of microtubules. Interestingly, this new twist allows us to return to the problem of how the search time can be independent of cell size. If the number of microtubules N scales with the cell's surface area ($N \propto R^2$), the quadratic prefactor's dependence on R drops out. Indeed, careful measurements have revealed precisely this kind of scaling [70]. The whole prefactor would lose all cell size dependence if the catastrophe rate k_c were constant with cell size, $k_c =$ constant. Next, to further abolish the exponential factor's variation with cell size in eqn. 8.46, either the cell is small (giving $R/\lambda \ll 1$), or—as discussed previously—$\lambda = v/k_c \propto R$, which (with constant catastrophe rate) compels the microtubule growth rate v to vary linearly with cell size R, namely $v \propto R$. These distinct biophysical relationships make precise predictions amenable to experimental scrutiny.

While we find these evocative comparisons between experimental data and the simple model of exploratory dynamics intriguing [34, 36], much more work needs to be done to specifically measure how the key parameters of models such as these scale with cell size (and other variables of physiological state). Here we only wish to point out how a simple model such as this points to fresh and specific questions about regulation of assembly of microtubules, which can be addressed by experiments.

8.4.4 Exploratory Power

At first glance, the analogy shown in Figure 8.2 in which a railroad between Boston and New York is constructed by randomly laying down track seems completely ridiculous. When viewed at human engineering scales, exploratory dynamics leads immediately to questions of efficiency, costs and benefits, words that are likely too imprecise for discussing the molecular and cellular scale events that use this mechanism in living organisms. A fascinating question that arises in the context of these problems in exploratory dynamics is how much they cost energetically. Post-translational modifications imply a steady-state energy flux [73]; adaptation of the chemotaxis receptors a steady-state energy flux in the form of methylation [74]; and the repeated cycles of growth and shrinkage in microtubule systems imply costs both in the form of cytoskeletal polymerization and of the motors that attend this process [75]. Recent calorimetry measurements using microtubule-motor systems found a remarkably large heat dissipation in comparison with the larger scale mechanical motions of these collectives [76]. In the remainder of this section, we perform a succession of calculations as shown in Figure 8.11 and summarized in Figure 8.12, beginning with the energetics of the exploratory dynamics of a single microtubule, followed by crude estimates of the cost of all of the microtubules in a spindle and then a series of comparisons of these powers and energies to other key processes in the dynamics of a cell.

For the exploratory dynamics of microtubule search and capture described in previous sections, we can make an elementary estimate of how many ATPs or GTPs have been consumed. The simplest estimate asks us to examine the energetic cost of repeated cycles of microtubule polymerization and depolymerization, ignoring for now the attendant costs associated with the many molecular motors that are part of the process as well. The starting point of the estimate is the recognition that every tubulin dimer added to the growing microtubule costs a GTP as shown in Figure 8.11(A) and is labeled as ΔG_{search}.

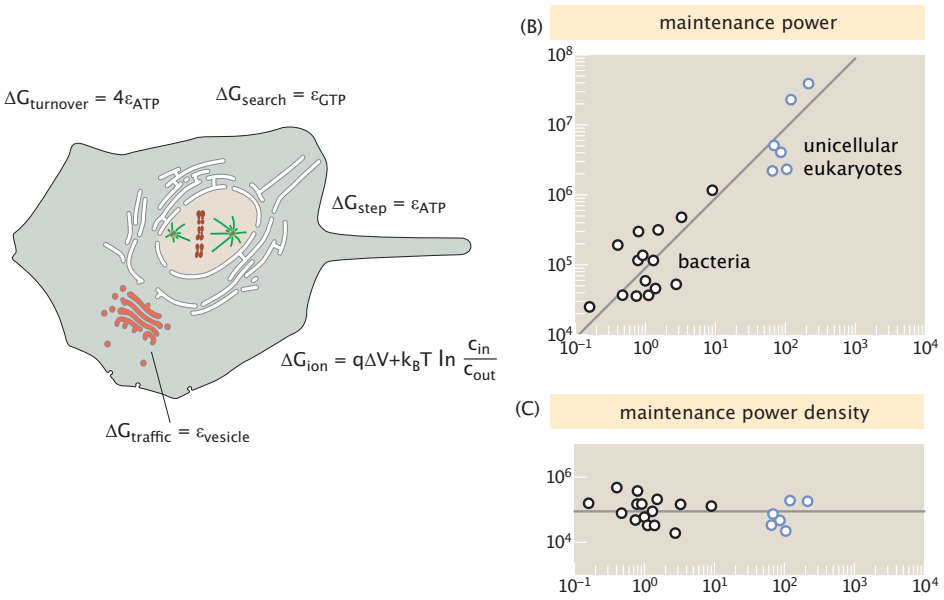


Figure 8.11: The power of biological processes. (A) Comparison of the bioenergetics of exploratory dynamics in chromosome search and capture with other key cellular processes. In each case, the ΔG shows the free energy cost of a unit process such as the addition of a tubulin monomer to a growing microtubule or the cost of transporting a single ion across the cell membrane. The power is given in turn by $J_{\text{process}}\Delta G_{\text{process}}$, where J_{process} is the flux of the process of interest. (B) Power to maintain cellular processes, as inferred by continuous chemostat measurements, as a function of cell volume; proportional scaling (acceptably capturing these data) is denoted by the gray line [77]. (C) Power densities per volume ensuing from the data in (B), reflecting a highly typical power density of $\langle\rho\rangle \approx 9 \times 10^5$ ATP/(μm^{3}s) for overall cellular maintenance.

The total power is given as $J_{\text{search}}\Delta G_{\text{search}}$, where J_{search} is the number of tubulin additions per unit time. As a result, the rate of GTP hydrolysis can be estimated as

$$J_{\text{search}} = \frac{n_{\text{proto}} v}{l_{\text{mono}}} \tag{8.47}$$

where n_{proto} is the number of protofilaments in a microtubule, v is the speed of microtubule growth in units of length/time and l_{mono} is the length of a tubulin dimer. If we assign a characteristic energy of $\varepsilon_{GTP} = 20\ k_B$T per GTP hydrolysis, this tells us that the power is given by

$$\text{power} \approx J_{\text{search}}\varepsilon_{GTP} \approx \varepsilon_{GTP}\frac{n_{\text{proto}}\, v}{l_{\text{mono}}}. \tag{8.48}$$

If we approximate the relevant numbers such as $n_{\text{proto}} \approx 10$, $v \approx 10$ μm/min and $l_{\text{mono}} \approx 5\ nm$, we find that the power for a single microtubule to perform exploratory behavior is roughly 500 ATP/s.

The results given above examine the energetics of a single microtubule doing repeated instances of exploratory search and capture. But as any microscopy image will reveal, this process involves hundreds to thousands of microtubules [36, 78] and hence our estimates both for the time scale of search and capture, and for the energetics and power need to be amended to account for this effect. As a result, we estimate the power expended during the search and capture process is between 5×10^3 and 5×10^4 ATP/s.

With this equation in hand, we can also then estimate the total energy expended in the search process given by the power times the elapsed time. If we now recall eqn. 8.37 which tells us on average how long the search process lasts, we find that the total energy is given by,

$$\text{energy} = \text{power} \times \langle t \rangle = \varepsilon_{GTP} \times J_{\text{search}} \times \langle t \rangle = \frac{n_{\text{proto}}\, v\, \varepsilon_{GTP}}{l_{\text{mono}}} \frac{1}{k_c} \left[\frac{e^{k_c \tau}}{f_r} - 1 \right]. \tag{8.49}$$

Given these various estimates, it is of great interest to compare them to other estimates of the energy budgets of cells. A critical question that we come back to over and over in the context of biological problems is how to judge whether a given process is "costly" or not. The starting point for answering that question is to take stock of the entire energy budget of a cell. A systematic study was made of the "biosynthetic cost of a gene" which attempted to work out the energetics of all of the key processes that take place within cells of all types [77]. In their Figure 1 whose data we reproduce here as Figure 8.11(B), they report on the maintenance cost with a value of roughly $f \times 10^9$ ATP/(μm^3 hr), which translates into roughly 10^6 ATP/(μm^3 s). This characteristic power density, inferred from extrapolated chemostat measurements on bacteria and amoebozoa, also aligns with typical metabolic rates of various human tissue cells, ranging from the slower metabolism of

resting erythrocytes ($\approx 8.5 \times 10^{-17}\mathrm{W}/\mu\mathrm{m}^3$) to the rapid metabolism of T-cell lymphocytes in antigen responses ($\approx\ 6.5 \times 10^{-13}\mathrm{W}/\mu\mathrm{m}^3$) [79], where we invoked the characteristic scale that one ATP hydrolysis liberates 10^{-19} J. Accordingly, a power density of about a million ATPs consumed per second per cubic micron serves as a very useful baseline power for the entirety of the processes associated with cellular maintenance. To develop specific intuition, we consider several particular processes that are constantly in play in cells of different types and that will give us a sense of the scale of the cost of the search and capture process.

Cells constantly remake proteins to compensate for their active or spontaneous degradation. The energetic cost to resynthesize proteins can be formidable, as a simple estimate reveals. Specifically—assuming that the cellular abundance of proteins is maintained at a steady-state—the stable copy number N, overall resynthesis rate q, and degradation rate k, and typical lifetime τ can be related by Little's theorem in queuing theory [80], $N/\tau = q = k$. Since mammalian cells typically have concentrations $c \approx$ a few million proteins per cubic micron [81]; typical protein turnover rates in bacteria [82] and eukaryotes [83] are both of order $k \approx f \times 10^{-5}\ \mathrm{s}^{-1}$ (consistent with measured typical half lives of a few hours and where we remind the reader that we adopt the convention that $f =$ few [84]); each protein is $\ell \approx$ a few hundred amino acids long and each peptide bond requires $\eta \approx 4$ ATP equivalents to polymerize [85], protein resynthesis to maintain the standing proteome pool should demand of order $ck\ell\eta \approx 10^5$ ATP/($\mu\mathrm{m}^3$ s). This significant cost can be within a factor few of the typical total power maintenance density of the cell we referred to above and is clearly much larger than the cost of the microtubule search and capture dynamics. Proteins have energetic demands in other ways as well. The post-translational modification of proteins can also demand modest to appreciable energy expenditures in a cell. Classic measurements and modeling of cyclic (de)phosphorylation cascades [86] report that typical operating conditions for these cascades demand at most of

order 1 mM ATP/min, corresponding to at most a few $\times 10^5$ ATP/μm^3 s.

Another fascinating, crucial, energetically-costly process that must be carried out at all times is the maintenance of the membrane potential as seen in Figure 8.11(A). In particular, the difference in ionic concentrations across membranes will be lost in the absence of active processes to maintain those gradients. Specifically, despite a very small permeability to such ions, there is a leak current of ions across lipid bilayer membranes. In particular, if we assume a typical permeability of $p = 10^{-5}$ nm/s [87] and a concentration difference across the membrane of 100 mM, then the flux is given by

$$J_{ion} = p\Delta c \approx 1 \frac{\text{ion}}{\mu\text{m}^2\,\text{s}}. \tag{8.50}$$

For a eukaryotic cell with a surface area of $f \times 10^3$ μm^2, this corresponds to $f \times 10^3$ ions/s. The cost to transport one such ion up its gradient is roughly $\Delta G_{ion} = q\Delta V + k_B T \ln(c_{in}/c_{out}) \approx 1$ ATP as seen in Figure 8.11(A), indicating that the power of this process is about 10^3 ATP/s, a tiny fraction of the overall power of the cell, but quite comparable to the power of a single microtubule engaging in the search and capture process.

Membrane potentials are not the only gradients within cells. Protein gradients of various kinds are present in cells and embryos of all types. A particularly well studied example that might give us intuition for power and energy scales of such gradient formation is offered by the eukaryotic flagellum. Here the process of intraflagellar transport (IFT) is characterized by molecular motors that carry cargo to the tip of the flagellum [88] as seen in Figure 8.11(A) with the cost of a single motor step given by $\Delta G_{step} = \varepsilon_{ATP}$. This process results in a linear concentration gradient of molecular motors within the flagellum, some large fraction of which are attached to microtubules and thereby consuming ATP [89]. Rough estimates

yield a power of the form

$$\text{power} = \frac{c_0 + c_{\text{tip}}}{2} L \left(\frac{v}{\delta}\right) \varepsilon_{ATP}, \tag{8.51}$$

where L is the length of the flagellum separating base from tip; c_0 and c_{tip} are respectively the concentrations of motors at the base and tip; and δ is the step size of the motor. Using parameters reported by the experiments of ref. [89] yields 5×10^5 ATP/s, or 5×10^4 ATP/μm^3 s (on the basis of flagellar volume).

The process of membrane trafficking is constantly rearranging the internal and external membranes of cells in a highly energetically costly way. As seen in Figure 8.11(A), every time a vesicle is created, a simple estimate implies a cost of $\Delta G_{\text{traffic}} \approx$ 10-20 ATPs corresponding to the 250-500 $k_B T$ cost of each such vesicle [90]. As was noted in a review on vesicles and trafficking: "A fibroblast kept in resting conditions in a tissue culture plate internalizes an amount of membrane equivalent to the whole surface area of the cell in one hour" [91]. If we estimate the area of such a cell as $f \times 10^3$ μm^2 and consider 100 nm vesicles, the area of each such vesicle is roughly $f \times 10^4$ nm^2, meaning roughly 6×10^4 vesicles are created every hour, or roughly 20 every second. At a cost of 10-20 ATP per vesicle, this amounts to a power of roughly $f \times 10^2$ ATP/s.

All of these estimates—collectively summarized in Figure 8.12—form a useful and fascinating backdrop for trying to better understand the energy budgets of cells and specifically, the costs of exploratory dynamics. For a process such as microtubule search and capture which results in the high-fidelity segregation of chromosomes, even in light of such estimates, it is not clear how to decide if a given power (or energy) is costly since chromosome segregation with errors can be deadly, the highest cost of all.

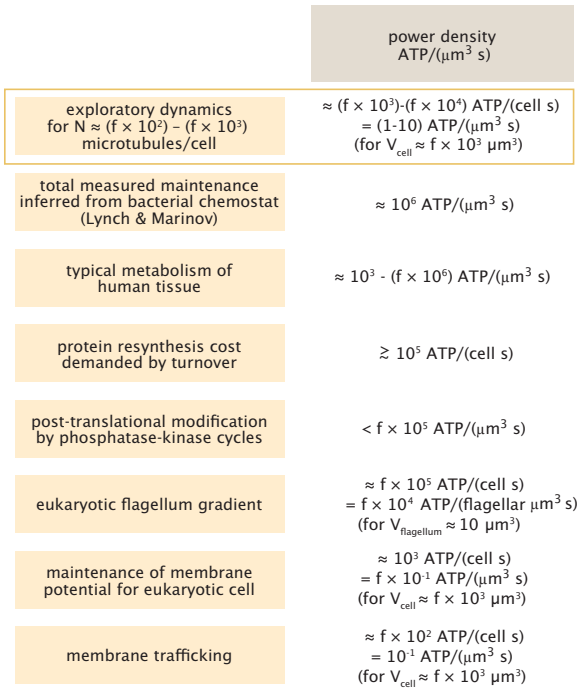


Figure 8.12: Summary of the power required for various cellular processes per volume. The goal of the figure is to compare the cost of microtubule search and capture by exploratory dynamics with the cost of other cellular processes. Note that we use the symbol f for "few" with the arithmetic rule that $f \times f \approx 10$ [84].

8.5 Transcription Factor Search for Binding Sites as Exploratory Dynamics

The problem of how transcription factors search and find their binding sites provides another illuminating example of exploratory dynamics. When we write simple chemical reaction schemes for describing the binding and unbinding of transcription factors to their cognate sites as is shown in Figure 8.13(A) they mask a much more complicated underlying exploratory reality. As seen in Figure 8.13(B), for a transcription factor to find its binding site and bind to it, the protein needs to traverse the complex intracellular environment including the complex polymeric geometry of the chromosomal DNA itself. Mechanistically, this exploratory dynamics takes place as shown schematically in Figure 8.14, where a combination of 1D diffusion along the DNA punctuated by episodes of 3D diffusion result in a search time that, in principle, is faster than either approach by itself. We also note as was said above, that this mechanism shares many features of chromosome search and capture, but unlike that case, does not involve any energy expenditure.

Our approach to this problem essentially imitates a beautiful treatment of the problem by Hachmo and Amir [32] though it is important to note that a broad swath of theoretical works have shed light on this fascinating topic of "facilitated diffusion" [28, 29, 30, 31]. We invoke a model where a transcription factor can "slide" along the DNA and also jump between genomically distant but spatially proximal parts of the genome. We note that, in addition to now classic in vitro bulk studies [29, 30], there have been a variety of beautiful in vitro studies using the tools of single-molecule biophysics to watch proteins as they bind to and translocate along DNA [92, 93, 94, 95]. Similarly, a new generation of single-molecule experiments has made it possible to query this same search process in the much more complex in vivo environment of living cells [96, 97]. For our purposes, all of this work provides a rich canvas for further elaborating the general idea of exploratory

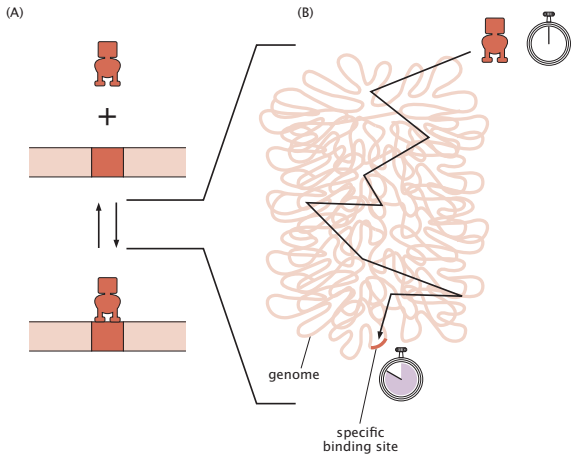


Figure 8.13: Searching for a binding site on the chromosome. (A) A particular binding site for a transcription factor controls the expression of some gene of interest and our goal is to replace the simple schematized reaction on the left with a picture of the dynamics that acknowledges how the kinetics depends upon the complex geometry of the chromosome. (B) The transcription factor has to "search" throughout the entire genome.

dynamics.

In Figure 8.14 we examine this canonical model of "facilitated diffusion" that combines 3D diffusion and 1D sliding along the DNA. In this model, a transcription factor diffuses along the DNA with a diffusion constant D_{1D} until it falls off the DNA after a time τ_{1D}. Once in the cytoplasm, the transcription factor undergoes 3D diffusion for a time τ_{3D} until it lands again on a different part of the DNA. Our goal is to estimate the search time under this composite exploratory dynamics which can be thought of as one-dimensional localized search for the binding site coupled with a "reset" mediated by three-dimensional diffusion and indicated schematically in Figure 8.15 using the evocative language of variation (diffusion) and selection (binding).

We begin by examining the probability of the transcription factor finding its target solely through 1D diffusion. The key idea in formulating the estimate is that during the time the transcription factor engages in 1D diffusion, it covers a distance given approximately by

$$L_{1D} = \sqrt{D_{1D}\tau_{1D}}. \tag{8.52}$$

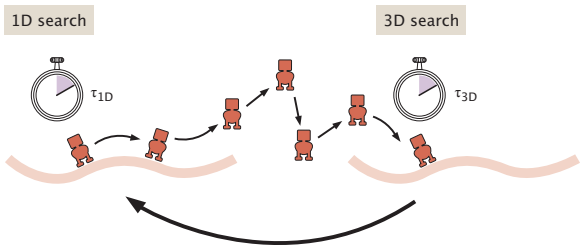


Figure 8.14: Simple model of combined 1D and 3D transcription factor search for a specific binding site. A transcription factor diffuses along the DNA for an approximate time τ_{1D} after which it falls of the DNA and performs three-dimensional diffusion for a time scale τ_{3D}. The cycle is then repeated until the transcription factor reaches its binding site.

If the total length of DNA to be searched is L_{genome} and the transcription factor lands on a random DNA position, then the probability p that the transcription factor will find its specific binding site during an episode of 1D diffusion that lasts a time τ_{1D} is given by the fraction of the genome explored during a sliding period. This argument leads to the estimate

$$p = \frac{L_{1D}}{L_{\text{genome}}} = \sqrt{\frac{D_{1D}\tau_{1D}}{L^2_{\text{genome}}}}. \tag{8.53}$$

Now we want to compute the probability of finding the site during repeated cycles of 1D and 3D diffusion. First, as shown in eqn. 8.53, the probability of finding the binding site after the first sliding period, corresponding to zero 3D hops, is given by

$$p_0 = p = \sqrt{\frac{D_{1D}\tau_{1D}}{L^2_{\text{genome}}}}. \tag{8.54}$$

This takes approximately a time

$$T_0 = \tau_{1D}. \tag{8.55}$$

If no site is found during the first sliding interval, there is a chance that the site will be found during the next sliding period. This probability is given by

$$p_1 = (1-p)p, \tag{8.56}$$

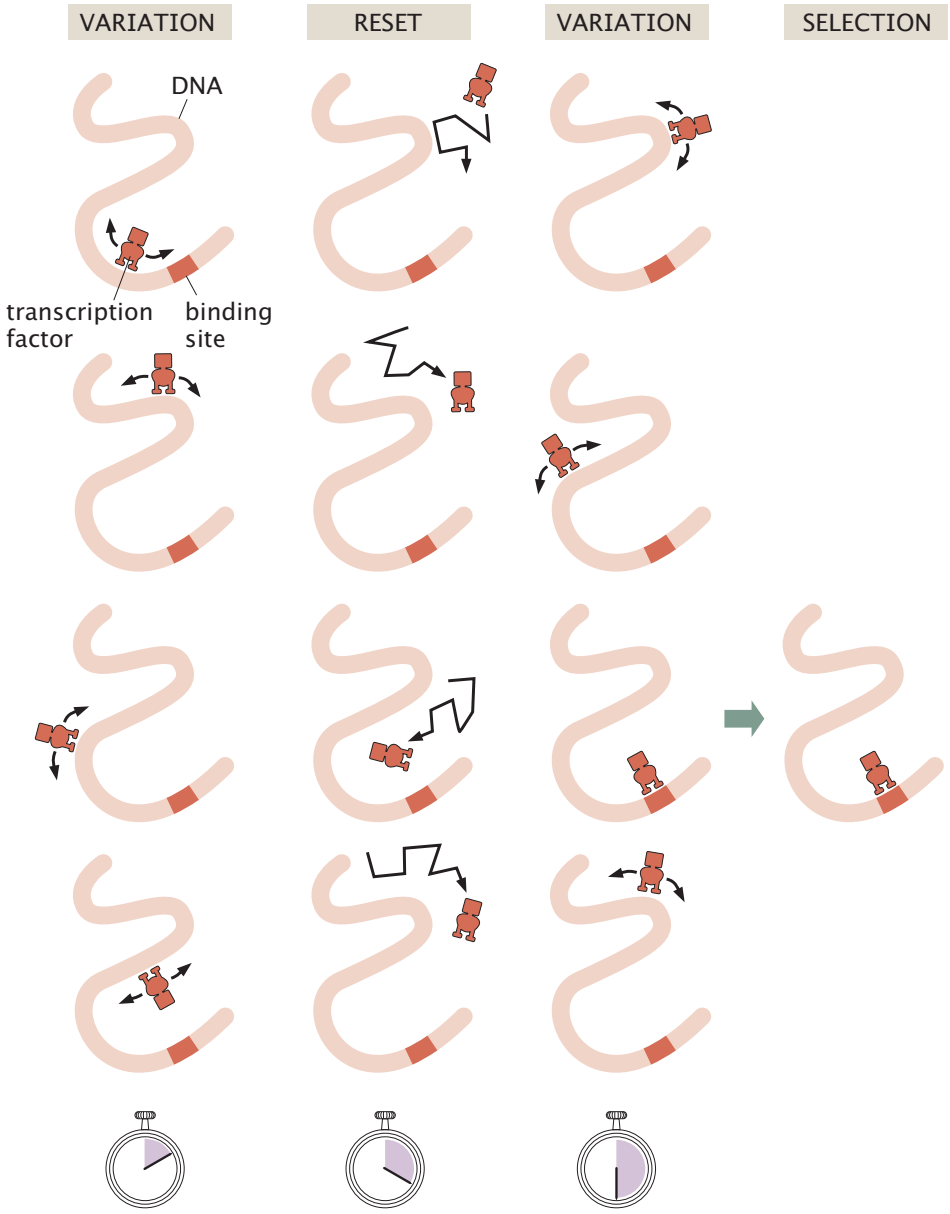


Figure 8.15: Exploratory dynamics of transcription factor search as an example of variation and selection. We can think of transcription factor search as a succession of search events along the DNA, punctuated by reset events of three dimensional diffusion. "Selection" occurs when the transcription factor binds to its target site.

where $1-p$ is the probability of the first 1D search being unsuccessful, and p the probability of succeeding in the next 1D search. The time it takes for a binding site to be found in this second attempt, after one 3D hop, is given by

$$T_1 = \tau_{1D} + \tau_{3D} + \tau_{1D}, \tag{8.57}$$

where the first term corresponds to the time of the first 1D search, the second term to the time of the subsequent 3D search, and the last term to the time of the second 1D search, when the site was found. Similarly, if the binding site is found after the second 3D hop, the transcription factor will have to have failed in its search during the the previous two 1D searches such that

$$p_2 = (1-p)(1-p)p \tag{8.58}$$

and

$$T_2 = \tau_{1D} + \tau_{3D} + \tau_{1D} + \tau_{3D} + \tau_{1D}. \tag{8.59}$$

We see a clear pattern emerge, namely, for the case involving i episodes of 3D search the probability of finding the site is given by

$$p_i = (1-p)^i p \tag{8.60}$$

and the corresponding search time is

$$T_i = (i+1)\tau_{1D} + i\tau_{3D}. \tag{8.61}$$

As a result, we can rewrite the average time for the transcription factor to find its binding site as

$$T = \sum_{i=0}^{\infty} T_i p_i, \tag{8.62}$$

where p_i is the probability of the transcription factor finding its site after i 3D hops, and T_i is the time it takes to perform the trajectory with i episodes of 3D diffusion. We can write this all out explicitly as

$$T = \sum_{i=0}^{\infty} T_i p_i = \sum_{i=0}^{\infty} (1-p)^i p \left((i+1)\tau_{1\text{D}} + i\tau_{3\text{D}} \right). \tag{8.63}$$

We note that the mathematical structure of this problem is essentially identical to our calculations in the previous section where we also found that our exploratory trajectories have probabilities described by the geometric distribution. This means that we can exploit much of the same algebra we did in the previous section.

To start, we expand the terms in parentheses in eqn. 8.63 to obtain

$$T = \tau_{1\text{D}}\, p \sum_{i=0}^{\infty} (1-p)^i i + \tau_{1\text{D}}\, p \sum_{i=0}^{\infty} (1-p)^i + \tau_{3\text{D}}\, p \sum_{i=0}^{\infty} (1-p)^i i. \tag{8.64}$$

We now regroup the terms and arrive at

$$T = (\tau_{1\text{D}} + \tau_{3\text{D}}) p \sum_{i=0}^{\infty} (1-p)^i i + \tau_{1\text{D}}\, p \sum_{i=0}^{\infty} (1-p)^i. \tag{8.65}$$

We note that the second sum is of the form of a geometric series which can be summed once again to yield

$$\sum_{i=0}^{\infty} (1-p)^i = \frac{1}{1-(1-p)} = \frac{1}{p} \tag{8.66}$$

As a result, we have

$$T = (\tau_{1\text{D}} + \tau_{3\text{D}}) p \sum_{i=0}^{\infty} (1-p)^i i + \tau_{1\text{D}}\, p \frac{1}{p} = (\tau_{1\text{D}} + \tau_{3\text{D}}) p \sum_{i=0}^{\infty} (1-p)^i i + \tau_{1\text{D}}. \tag{8.67}$$

To carry out the remaining sum, we invoke the use of the derivative trick we originally

introduced in eqn. 8.9. Namely, we write the sum as

$$\sum_{i=0}^{\infty}(1-p)^i i = -(1-p)\frac{d}{dp}\left(\sum_{i=0}^{\infty}(1-p)^i\right), \tag{8.68}$$

where the extra $-(1-p)$ term takes care of the fact that, after taking the derivative, the $(1-p)^i$ term inside the summation would otherwise become $-(1-p)^{i-1}$. We again have the familiar geometric sum that can be carried out to yield

$$-(1-p)\frac{d}{dp}\left(\sum_{i=0}^{\infty}(1-p)^i\right) = -(1-p)\frac{d}{dp}\frac{1}{p} = -(1-p)\frac{(-1)}{p^2} = \frac{1-p}{p^2}. \tag{8.69}$$

As a result, eqn. 8.65 becomes

$$T = (\tau_{1\mathrm{D}} + \tau_{3\mathrm{D}})p\frac{1-p}{p^2} + \tau_{1\mathrm{D}}\,p\frac{1}{p} = (\tau_{1\mathrm{D}} + \tau_{3\mathrm{D}})\frac{1-p}{p} + \tau_{1\mathrm{D}}. \tag{8.70}$$

This can be simplified to the form

$$T = \frac{\tau_{1\mathrm{D}}}{p} + \frac{\tau_{3\mathrm{D}}}{p} - \tau_{3\mathrm{D}}. \tag{8.71}$$

Now, we assume that $p \ll 1$, meaning that it will always take many cycles of 1D and 3D exploration to find the binding site. Mathematically, this means $\tau_{3\mathrm{D}}/p \gg \tau_{3\mathrm{D}}$ such that we can ignore the last $\tau_{3\mathrm{D}}$ term in the expression. In light of this approximation, we find a total search time of the form

$$T = \frac{\tau_{1\mathrm{D}}}{p} + \frac{\tau_{3\mathrm{D}}}{p}. \tag{8.72}$$

To understand the implications of the calculation that led to the total search time shown in eqn. 8.72, we recall the definition of p introduced in eqn. 8.54. As a result of this definition,

we can use the formula for p to rewrite our result for the search time as

$$T = \sqrt{\frac{\tau_{1D} L_{genome}^2}{D_{1D}}} + \tau_{3D}\sqrt{\frac{L_{genome}^2}{D_{1D}\tau_{1D}}}. \tag{8.73}$$

Further, we will non-dimensionalize the equation by measuring the search time in units of the 3D exploration time τ_{3D}, leading to

$$\frac{T}{\tau_{3D}} = \bar{T} = \sqrt{\frac{\tau_{1D}}{\tau_{3D}}\frac{L_{genome}^2}{D_{1D}\tau_{3D}}} + \sqrt{\frac{L_{genome}^2}{D_{1D}\tau_{1D}}}, \tag{8.74}$$

where we have defined $\bar{T} = T/\tau_{3D}$. In addition, we define the non-dimensional 1D search time $\bar{\tau}_{1D} = \tau_{1D}/\tau_{3D}$ measured in units of the 3D search time, to arrive at

$$\bar{T} = \sqrt{\bar{\tau}_{1D}\frac{L_{genome}^2}{D_{1D}\tau_{3D}}} + \sqrt{\frac{L_{genome}^2}{D_{1D}\tau_{3D}\bar{\tau}_{1D}}}. \tag{8.75}$$

We see that both square roots in the equation have a factor

$$\alpha = \sqrt{\frac{D_{1D}\tau_{3D}}{L_{genome}^2}}, \tag{8.76}$$

corresponding to the fraction of the genome the transcription factor would have explored by 1D diffusion during the time it was actually diffusing in 3D. As a result, the non-dimensionalized time can be written as

$$\bar{T} = \frac{1}{\alpha}\left(\sqrt{\bar{\tau}_{1D}} + \sqrt{\frac{1}{\bar{\tau}_{1D}}}\right). \tag{8.77}$$

This expression for the search time makes the striking prediction that there is an optimal balance between the time the transcription factor spends searching in one dimension and three dimensions. Specifically, when the one-dimensional search time is relatively long

($\bar{\tau}_{1\mathrm{D}} \gg 1$), the first term $\sqrt{\bar{\tau}_{1\mathrm{D}}}$ dominates the overall search time, implying that the overall search time grows as $\bar{T} \sim \bar{\tau}_{1\mathrm{D}}^{1/2}$. Otherwise (for short one-dimensional search times $\bar{\tau}_{1\mathrm{D}} \ll 1$), the second term $1/\bar{\tau}_{1\mathrm{D}} \gg 1$ dominates, implying in turn that the overall search time shrinks as $\bar{T} \sim \bar{\tau}_{1\mathrm{D}}^{-1/2}$. The crossover between these behaviors gives a minimum overall search time, when these terms are comparable.

We can quickly foresee that the value of the one-dimensional search time $\bar{\tau}_{1\mathrm{D}}$ that minimizes the overall search time is located at $\bar{\tau}_{1\mathrm{D}} = 1$ by remembering the arithmetic mean-geometric mean inequality, which states that for two nonnegative values a, b, $\frac{a+b}{2} \geq \sqrt{ab}$; taking $b = 1/a$ gives $a + \frac{1}{a} \geq 2$, where the optimal value of equality requires $a = 1/a$, or $a = 1$. (Amusingly, the same optimization problem emerges when asking for the side lengths of a rectangle of unit area that has the smallest perimeter.) Letting a be $\sqrt{\bar{\tau}_{1\mathrm{D}}}$ translates this result to mean that smallest overall search time occurs when $\sqrt{\bar{\tau}_{1\mathrm{D}}} = 1$. This optimum may also be seen by taking the derivative of $\bar{T}$ with respect to $\bar{\tau}_{1\mathrm{D}}$

$$\frac{d\bar{T}}{d\bar{\tau}_{1\mathrm{D}}} = \frac{1}{\alpha}\frac{d}{d\bar{\tau}_{1\mathrm{D}}}\left(\sqrt{\bar{\tau}_{1\mathrm{D}}} + \sqrt{\frac{1}{\bar{\tau}_{1\mathrm{D}}}}\right) = \frac{1}{\alpha}\left(\frac{1}{2}\frac{1}{\bar{\tau}_{1\mathrm{D}}^{1/2}} - \frac{1}{2}\frac{1}{\bar{\tau}_{1\mathrm{D}}^{3/2}}\right) = 0, \tag{8.78}$$

which is satisfied if $\bar{\tau}_{1\mathrm{D}} = 1$, regardless of the numerical value of the parameter α. This result carries the physical meaning that $\tau_{1\mathrm{D}} = \tau_{3\mathrm{D}}$: the shortest overall search time is reached when the times spent by the transcription factor exploring in one and three dimensions are equal. Equivalently, this optimum corresponds to the case where the ratio of the distance traveled in 1D (L_{1D}) and that in 3D ($L_{3D} \equiv \sqrt{D_{3D}\tau_{3D}}$) are matched to an appropriate balance of the respective diffusion coefficients, $\frac{L_{1D}}{L_{3D}} = \sqrt{\frac{D_{1D}}{D_{3D}}}$.

Armed with this analytical understanding of the tradeoff between one dimensional searches and three dimensional resets, we now explore where typical cells operate parametrically. Before we can make progress on exploring all of these results numerically, we first need to estimate some of the key parameters that are present in the model. In

particular, we need an estimate of τ_{3D}, the time scale for 3D exploration between successive episodes of 1D diffusion along the DNA. Figure 8.16 offers a naive way to make such an estimate. In Figure 8.16(B), we consider a transcription factor in a "gas" of binding sites. When the transcription factor is not bound to the DNA, we imagine these binding sites as freely diffusing with diffusion constant D_{bs} in the vicinity of the transcription factor, and the transcription factor itself diffuses with a diffusion constant D_{TF}. In this simple model, the time τ_{3D} corresponds to the average time for one of these binding sites to reach the transcription factor. To estimate the concentration of this gas of binding sites, we use the fact that there are of order 10^6 such binding sites in a bacterial genome in a volume of 1 μm^3, each one corresponding to a base pair in the genome. We can use this concentration in conjunction with the diffusion limited on-rate [38, 98] to estimate the diffusion time as

$$\tau_{3D} = \frac{1}{4\pi(D_{bs} + D_{TF})a\, c_0}. \tag{8.79}$$

We note, however, that base pairs do not diffuse independently of each other. Because base pairs are part of the huge polymer that makes up the chromosome, their effective diffusion coefficient is expected to be much lower than that of a free base pair. Indeed, recent in vivo studies in a broad swath of organisms have reported an effective diffusion coefficient of chromosomal loci of the order of 10^{-2} μm^2/s [99, 100, 101]. As a result, if we take the diffusion constant of a base pair as $D_{bs} \approx 10^{-2}$ μm$^2/s$ we see that it is much smaller than the diffusion constant of the protein itself ($D_{TF} \approx 10$ μm$^2/s$) and hence we can safely ignore the contribution of base pair diffusion. If we now impose the size of the target site to be $a \approx f$ nm and the concentration of base pairs is approximated by $c_0 = 10^6/\mu$m$^3 \approx 1$mM, we arrive at $\tau_{3D} \approx f \times 10^{-6}$ s. Finally, in order to estimate the numerical value of α defined in eqn. 8.76, we must also determine the 1D diffusion constant describing the sliding of the transcription factor along the DNA. Single-molecule studies have found 1D diffusion constants of the order of $D_{1D} \approx 10^5 - 10^6$ bp^2/s $\approx 10^{-2} - 10^{-1}$ $\mu m^2/s$ [92, 93]. In light of

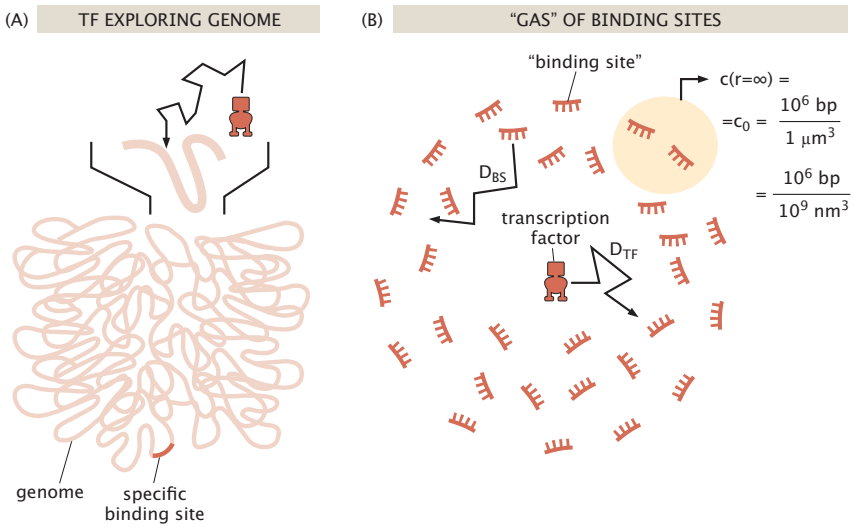


Figure 8.16: Estimating time scales of encounters between a transcription factor and the genome. (A) During episodes of 3D diffusion, the transcription factor diffuses through the cellular environment before encountering the genome. (B) Estimating the time scale of diffusion by using the diffusion-limited on rate idea by pretending that the binding sites themselves are diffusing around. (C) Estimating the time scale of diffusion by pretending that the diffusion takes place within a discrete lattice of binding sites in 3D space.

all of these estimated parameters, we find $\alpha \approx f \times 10^{-7} - 10^{-6}$.

Using the parameters we have estimated, Figure 8.17 shows the total search time as a function of the time spent searching in 1D in units of the time spent searching in 3D. The figure affirms the appearance and location of the optimal 1D search time, namely, when the time spent searching in 1D is equal to the time searching in 3D.

Despite how crude our estimates were, we can now examine whether the transcription factor binding site search dynamics takes a time consonant with the optimal search time which is found when the 1D and 3D search times are equal. To make that determination, we need the 1D exploration time which we can then compare to our estimate of the 3D exploration time that we performed earlier on in this section. In vivo experiments in bacteria have estimated the sliding length of a transcription factor to be of order $L_{1D} \approx 100$ bp [96]. Given our estimated 1D diffusion constant, this translates into a 1D search time of roughly $\tau_{1D} \approx 10^{-2} - 10^{-1}$ s. This value for τ_{1D} differs by several orders of magnitude

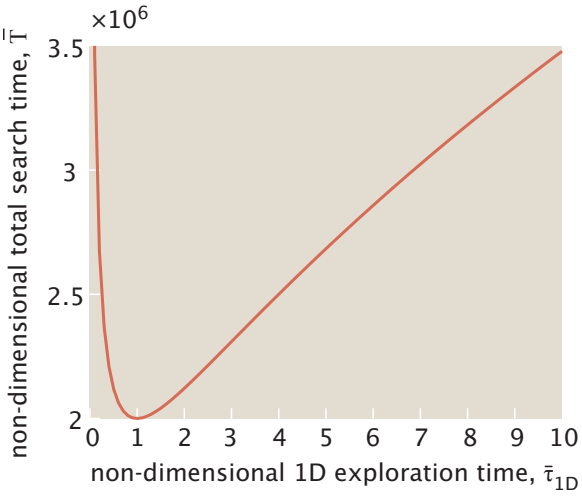


Figure 8.17: Time for a transcription factor to find its binding site in a model of combined 1D and 3D search. The search time is measured in units of the time spent searching in 3D as a function of the relative time searching in 1D to 3D. The parameter α was chosen to have the value $\alpha = 10^{-6}$ corresponding to the fraction of the genome that would have been explored by 1D diffusion during the time of a 3D exploration. Note that as the time spent exploring in 1D approaches zero, $\bar{\tau}_{1D} \to 0$, clearly the total overall search time $\bar{T}$ diverges, as encountering the correct binding site requires at least some time spent exploring in 1D.

from our estimate of τ_{3D} shown in Figure 8.16 (B). However, it is important to note that neither the 1D nor the 3D exploration times have been directly measured in vivo and that our numerical estimates depend upon a succession of poorly known parameters. Indeed, recent experiments in mammalian cells paint a different picture. These experiments estimated that a transcription factor spends a similar amount of time in the 1D and 3D exploration modes (75% of the time doing 3D exploration, to be more precise) [97].

A second question we can explore guided by the estimate resulting in Figure 8.17 is to ask whether this facilitated diffusion mechanism results in faster search times of transcription factors for their specific binding sites than explorations based on a 1D or 3D search alone. To make this possible, we first assign absolute units to the optimal non-dimensional total search time $\bar{T}$ shown in the figure. Specifically, we compute

$$T_{\text{min}} = \bar{T}_{\text{min}} \tau_{3D} \approx f \text{ s}. \tag{8.80}$$

Given in vitro measurements of the 1D diffusion constant of $D_{1D} \approx 10^5 - 10^6$ bp^2/s already described above, we can estimate a search time for a specific binding site through a genome composed of $L_{\text{genome}} = 10^6$ bp as

$$T_{1D} \approx \frac{L_{\text{genome}}^2}{D_{1D}} \approx 10^6 - 10^7 \text{ s}, \tag{8.81}$$

which is clearly several orders of magnitude slower than the search time associated with the facilitated diffusion mechanism. Finally, we estimate the time for a search mechanism based on pure 3D diffusion as

$$T_{3D} = \frac{1}{4\pi D_{\text{TF}}^{\text{eff}} a c_0}, \tag{8.82}$$

where $c_0 \approx 1$ nM, the concentration of an individual base pair in a 1μm^3 nuclear volume, and where $a \approx f$ nm. To make this estimate, we also acknowledge that, in this case, $D_{\text{TF}}^{\text{eff}}$ is an effective diffusion constant that needs to account for the interactions between the transcription factor and the milieu of the nucleus such as those leading to crowding and to non-specific DNA binding during the search process. Measurements and estimates of this effective diffusion constant remain an open subject. We choose values of $D_{\text{TF}}^{\text{eff}} = 10^{-1} - 1\ \mu$m^2/s based on recent in vivo measurements [97]. As a result, we get

$$T_{3D} = \frac{1}{4\pi D_{\text{TF}}^{\text{eff}} a c_0} \approx f(10 - 10^2) \text{ s}. \tag{8.83}$$

This search time is up to two orders of magnitude slower than the optimum search time due to facilitated diffusion T_{min}, suggesting that facilitated diffusion could be at play to speed up the transcription factor binding search process. Regardless of the specific numerical values, which are clearly still a subject of active investigation in the field, we posit that the model shown in this section is yet another intriguing example of the power of exploratory dynamics—in this case in the form of facilitated diffusion—to accomplish

and potentially speed up specific biological functions.

8.6 Discussion

The work of Prof. Erich Sackmann, partly to whom this work honors, was characterized by a principled approach to understanding the huge diversity of biological processes. In the spirit of his many successes, we used this opportunity to playfully engage with the idea that many biological processes can be thought of as case studies in exploratory dynamics. In contrast with the conventional physics paradigms for dynamics in which, given some initial conditions, the equations of dynamics are used to find the future evolution of the system, we have argued that biology's unique and necessary "solution" to achieving some target function is through a quite distinct alternative: exploratory dynamics. From this point of view, the classic ideas of variation and selection in the context of biological evolution is hypothesized to have a much broader reach.

To provide some quantitative substance to the hypothesis of exploratory dynamics, we considered several important case studies that have been studied rigorously and successfully for decades, namely, chromosome search and capture, and the search of transcription factors for their target DNA binding sites. Though our paper focused primarily on these particular case studies, we argue that the idea of exploratory dynamics has much broader biological reach, both in terms of the phenomena of living organisms and in the diversity of possible theoretical approaches to these problems.

Ultimately, from a theoretical perspective, we think of questions in exploratory dynamics as falling within the broad province of statistical physics. More specifically, there are a growing number of examples of different approaches to considering the statistical mechanics of microscopic trajectories as opposed to the more traditional statistical mechanics of states. The underlying statistical mechanics described here forced us to

consider a number of problems involving the geometric distribution. However, we suspect that there are other mathematical paradigms that will be useful, or even necessary, for considering the broad class of problems in exploratory dynamics. For example, there is a growing literature on the subject of random walks with reset in which the random walker is forced to return to the origin [62].

Clearly, molecular structures, cells, and organisms are often steered by external cues that vary in time, space, and history—cues that themselves are often sculpted by feedbacks from exploratory trajectories, as with ants leaving odor trails. Further, exploratory success probabilities (setting the "life expectancies" of exploratory trajectories) may manifest rich memory effects, for instance when having explored for a while itself increases or decreases the probability of impending success. These complexities urge new mathematical frameworks to be developed and applied that accurately capture the exquisite and complex resulting tradeoffs and consequences of exploration. While it may prove productive to build metaphors between the biological phenomena showing these complexities and algorithmic phenomena such as reinforcement learning or stochastic gradient descent, our central contention is that the unique ingredients, constraints, and goals of biology demand altogether new classes of exploratory behavior whose structure and contrasting examples may also shed light on these more traditional domains.

In summary, we view a fascinating dichotomy between the many distinct and impressive dynamical laws that have formed the backdrop of physics and mathematics for centuries and the unique and necessary exploratory dynamics adopted by many biological processes. Though we often adopt perspectives in which systems move deterministically from some initial condition, from the examples of exploratory dynamics, perhaps one learns that the best plan can be no plan at all.

Author Contributions

JK, MK, HGG, GLS, and RP designed research; JK, MK, HGG, GLS, and RP performed research; and JK, MK, HGG, GLS, and RP wrote the paper.

Declaration of Interests

All authors declare no conflicts of interest.

Acknowledgments

We are grateful to the NIH for support through award numbers DP1OD000217 (Director's Pioneer Award) and NIH MIRA 1R35 GM118043-01 to RP. MK acknowledges support from the NIH under grants R35-GM145248 and R01 AG07334. JK was supported by the Simons Foundation and NSF grants DMR-1610737 and MRSEC DMR-2011846. HGG was supported by NIH R01 Awards R01GM139913 and R01GM152815, by the Koret-UC Berkeley-Tel Aviv University Initiative in Computational Biology and Bioinformatics, by a Winkler Scholar Faculty Award, and by the Chan Zuckerberg Initiative Grant CZIF2024-010479. HGG is also a Chan Zuckerberg Biohub Investigator (Biohub–San Francisco). GLS thanks the US NSF Graduate Research Fellowship Program under Grant DGE-1745301 and Caltech's Center for Environmental Microbial Interactions for support. We are all deeply grateful to the CZI Theory Institute Without Walls for the support that makes our exploratory studies of exploratory dynamics possible. We have benefited greatly from conversations with Xingbo Yang, Peter Foster, Tim Mitchison, Natalia Orlovsky, Lena Koslover, Yuhai Tu and Milo Lin. Ana Duarte and Sara Mahdavi provided helpful discussion as well as numerical simulations. We are also grateful

to Ariel Amir for very helpful discussions and insights and for suggesting the approach for transcription factor search.

Bibliography

[1] Albert Einstein and Leopold Infeld. *The Evolution of Physics: The Growth of Ideas from Early Concepts to Relativity and Quanta*. Simon and Schuster, New York,, 1961.

[2] SG Brush. *The Kind of Motion We Call Heat*. North Holland, Amsterdam, 1986.

[3] Emilio Segré. *From Falling Bodies to Radio Waves: Classical Physicists and their Discoveries*. Dover Publications, Mineola, N.Y., dover edition, 2007.

[4] Malcolm S. Longair. *Theoretical Concepts in Physics: an Alternative View of Theoretical Reasoning in Physics*. Cambridge University Press,, 2020.

[5] I. Bernard Cohen. *The Birth of a New Physics*. Greenwood Press, Westport, Conn., 1981.

[6] Colin Pask. *Magnificent Principia: Exploring Isaac Newton's Masterpiece*. Prometheus Books, 2019.

[7] Kip S. Thorne and Roger D. Blandford. *Modern Classical Physics*. Princeton University Press, Princeton, New Jersey, 2021.

[8] C. L. Hueschen and R. Phillips. *The Restless Cell: Continuum Theories of Living Matter*. Princeton University Press, Princeton, 1st edition, 2024.

[9] M Klein. Mechanical Explanation at the End of the Nineteenth Century. *Centaurus*, 17:58–82, 1973.

[10] A. B. Kudryavtsev, Reginald F. Jameson, and W. Linert. *The Law of Mass Action*. Springer, Berlin; New York, 2001.

[11] Martin Feinberg. *Foundations of Chemical Reaction Network Theory*. Springer International Publishing: Imprint: Springer, 2019.

[12] Jean-Baptiste-Joseph Fourier. *The Analytical Theory of Heat*. Dover Publications, Mineola, N.Y., 2003.

[13] John Crank. *The Mathematics of Diffusion*. Clarendon Press, Oxford, Eng, 2d edition, 1975.

[14] H. S. Carslaw and J. C. Jaeger. *Conduction of Heat in Solids*. Clarendon Press; Oxford University Press, Oxford Oxfordshire New York, 2nd edition, 1986.

[15] F. M. Richter. Kelvin and the Age of the Earth. *Journal of Geology*, 94(3):395–401, 1986.

[16] G. Brent Dalrymple. *The Age of the Earth*. Stanford University Press, Stanford, California, 1991.

[17] G. Brent Dalrymple. *Ancient Earth, Ancient Skies: The Age of Earth and its Cosmic Surroundings*. Stanford University Press, Stanford, California, 2004.

[18] Alfred J. Lotka. *Elements of Mathematical Miology*. Dover Publications, New York„ 1956.

[19] Ted J. Case. *An Illustrated Guide to Theoretical Ecology*. Oxford University Press, New York ; Oxford, 2000.

[20] William B. Provine. *The Origins of Theoretical Population Genetics*. University of Chicago Press, Chicago, 2nd edition, 2001.

[21] Nicolas Bacaër. *A Short History of Mathematical Population Dynamics*. Springer Verlag London Ltd., London, 2011.

[22] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Reviews of Modern Physics*, 15(1):0001–0089, 1943.

[23] D. S. Lemons. *An Introduction to Stochastic Processes in Physics*. The Johns Hopkins University Press, Baltimore, Maryland, 2002.

[24] Jimena Canales. *Bedeviled: A Shadow History of Demons in Science*. Princeton University Press, Princeton ; Oxford, 2020.

[25] Nicola Twilley. *Frostbite: How Refrigeration Changed our Food, our Planet, and Ourselves*. Penguin Press, New York, 2024.

[26] Gerhart J. and Kirschner M. *Cells, Embryos and Evolution*. Blackwell Science, Malden, Mass., 1997.

[27] M. Kirschner and T. Mitchison. Beyond Self-Assembly - from Microtubules to Morphogenesis. *Cell*, 45(3):329–342, 1986.

[28] Peter H. von Hippel. *On the Molecular Bases of the Specificity of Interaction of Transcriptional Proteins with Genome DNA*, pages 279–347. Springer US, Boston, MA, 1979. ISBN 978-1-4684-3417-0. doi: 10.1007/978-1-4684-3417-0_8. URL https://doi.org/10.1007/978-1-4684-3417-0_8.

[29] R. B. Winter and P. H. von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 2. The *Escherichia coli* repressor–operator interaction: equilibrium measurements. *Biochemistry*, 20(24):6948–60, 1981.

[30] R. B. Winter, O. G. Berg, and P. H. von Hippel. Diffusion-driven mechanisms of protein translocation on nucleic acids. 3. The *Escherichia coli lac* repressor–operator interaction: kinetic measurements and conclusions. *Biochemistry*, 20(24):6961–77, 1981.

[31] T. Hu, A. Y. Grosberg, and B. I. Shklovskii. How proteins search for their specific sites on DNA: the role of DNA conformation. *Biophys J*, 90(8):2731–44, 2006.

[32] O. Hachmo and A. Amir. Conditional probability as found in nature: Facilitated diffusion. *American Journal of Physics*, 91(8):653–658, 2023.

[33] T. J. Mitchison and M. W. Kirschner. Properties of the Kinetochore In vitro. II. Microtubule Capture and ATP-Dependent Translocation. *Journal of Cell Biology*, 101(3):766–777, 1985.

[34] T. E. Holy and S. Leibler. Dynamic Instability of Microtubules as an Efficient Way to Search in-Space. *Proc. Natl. Acad. Sci. U.S.A*, 91(12):5682–5685, 1994.

[35] G. G. Gundersen. Evolutionary conservation of microtubule-capture mechanisms. *Nature Reviews Molecular Cell Biology*, 3(4):296–304, 2002.

[36] R. Wollman, E. N. Cytrynbaum, J. T. Jones, T. Meyer, J. M. Scholey, and A. Mogilner. Efficient chromosome capture requires a bias in the 'search-and-capture' process during mitotic-spindle assembly. *Current Biology*, 15(9):828–832, 2005.

[37] R. Heald and A. Khodjakov. Thirty years of search and capture: The complex simplicity of mitotic spindle assembly. *J Cell Biol*, 211(6):1103–11, 2015.

[38] H. C. Berg. *Random Walks in Biology*. Princeton University Press, Princeton, N.J., expanded edition, 1993.

[39] Howard C. Berg. *E. coli in Motion*. Springer, New York, 2004.

[40] A. Ghabrial, S. Luschnig, M. M. Metzstein, and M. A. Krasnow. Branching morphogenesis of the tracheal system. *Annual Review of Cell and Developmental Biology*, 19:623–647, 2003.

[41] T. C. Li, T. M. Fu, K. K. L. Wong, H. J. Li, Q. J. Xie, D. J. Luginbuhl, M. J. Wagner, E. Betzig, and L. Q. Luo. Cellular bases of olfactory circuit assembly revealed by systematic time-lapse imaging. *Cell*, 184(20):5107–5121.e14, 2021.

[42] S. Tonegawa. Somatic generation of antibody diversity. *Nature*, 302(5909):575–581, 1983.

[43] Scott H. Podolsky and Alfred I. Tauber. *The Generation of Diversity: Clonal Selection Theory and the Rise of Molecular Immunology*. Harvard University Press, Cambridge, Mass., 1997.

[44] Kenneth Murphy, Paul Travers, Mark Walport, and Charles Janeway. *Janeway's Immunobiology*. Garland Science, New York, 8th edition, 2012.

[45] P. A. Prince, A. G. Wood, T. Barton, and J. P. Croxall. Satellite Tracking of Wandering Albatrosses (Diomedea-Exulans) in the South-Atlantic. *Antarctic Science*, 4(1):31–36, 1992.

[46] B. A. Block, I. D. Jonsen, S. J. Jorgensen, A. J. Winship, S. A. Shaffer, S. J. Bograd, E. L. Hazen, D. G. Foley, G. A. Breed, A. L. Harrison, J. E. Ganong, A. Swithenbank, M. Castleton, H. Dewar, B. R. Mate, G. L. Shillinger, K. M. Schaefer, S. R. Benson, M. J. Weise, R. W. Henry, and D. P. Costa. Tracking apex marine predator movements in a dynamic ocean. *Nature*, 475(7354): 86–90, 2011.

[47] Gandhimohan M. Viswanathan. *The Physics of Foraging: An Introduction to Random Searches and Biological Encounters*. Cambridge University Press, Cambridge; New York, 2011.

[48] R. Kays, M. C. Crofoot, W. Jetz, and M. Wikelski. Terrestrial animal tracking as an eye on life and planet. *Science*, 348(6240), 2015.

[49] M. Antolos, S. A. Shaffer, H. Weimerskirch, Y. Tremblay, and D. P. Costa. Foraging Behavior and Energetics of Albatrosses in Contrasting Breeding Environments. *Frontiers in Marine Science*, 4, 2017.

[50] M. A. Hindell, R. R. Reisinger, Y. Ropert-Coudert, L. A. Hückstädt, P. N. Trathan, H. Bornemann, J. B. Charrassin, S. L. Chown, D. P. Costa, B. Danis, M. A. Lea, D. Thompson, L. G. Torres, A. P. Van de Putte, R. Alderman, V. Andrews-Goff, B. Arthur, G. Ballard, J. Bengtson, M. N. Bester, A. S. Blix, L. Boehme, C. A. Bost, P. Boveng, J. Cleeland, R. Constantine, S. Corney, R. J. M. Crawford, L. Dalla Rosa, P. J. N. de Bruyn, K. Delord, S. Descamps, M. Double, L. Emmerson, M. Fedak, A. Friedlaender, N. Gales, M. E. Goebel, K. T. Goetz, C. Guinet, S. D. Goldsworthy, R. Harcourt, J. T. Hinke, K. Jerosch, A. Kato, K. R. Kerry, R. Kirkwood, G. L. Kooyman, K. M. Kovacs, K. Lawton, A. D. Lowther, C. Lydersen, P. O. Lyver, A. B. Makhado, M. E. Márquez, B. McDonald, C. R. McMahon, M. Muelbert, D. Nachtsheim, K. W. Nicholls, E. S. Nordoy, S. Olmastroni, R. A. Phillips, P. Pistorius, J. Plötz, K. Pütz, N. Ratcliffe, P. G. Ryan, M. Santos, C. Southwell, I. Staniland, A. Takahashi, A. Tarroux, W. Trivelpiece, E. Wakefield, H. Weimerskirch, B. Wienecke, J. C. Xavier, S. Wotherspoon, I. D. Jonsen, and B. Raymond. Tracking of marine predators to protect Southern Ocean ecosystems. *Nature*, 580(7801):87–+, 2020.

[51] J. J. Hopfield. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci U S A*, 71(10):4135–9, 1974.

[52] J. Ninio. Kinetic amplification of enzyme discrimination. *Biochimie*, 57(5):587–95, 1975.

[53] E. Hannezo and B. D. Simons. Multiscale dynamics of branching morphogenesis. *Current Opinion in Cell Biology*, 60:99–105, 2019.

[54] A. Palavalli, N. Tizón-Escamilla, J. F. Rupprecht, and T. Lecuit. Deterministic and Stochastic Rules of Branching Govern Dendrite Morphogenesis of Sensory Neurons. *Current Biology*, 31 (3):459–472.e4, 2021.

[55] Bernard Osgood Koopman. *Search and Screening: General Principles with Historical Applications*. Pergamon Press, Elmsford, N.Y., 1980.

[56] Daniel H. Wagner, W. Charles Mylander, Thomas J. Sanders, and United States Naval Academy. Operations Analysis Study Group. *Naval Operations Analysis*. Naval Institute Press, Annapolis, Md., 3rd edition, 1999.

[57] E. T. Jaynes. Entropy and search theory. *preprint presented at the First Maximum Entropy Workshop*, 1981.

[58] Stephen Budiansky. *Blackett's War*. Alfred A. Knopf, New York, 1st edition, 2013.

[59] Horace Freeland Judson. *The Eighth Day of Creation*. Cold Spring Harbor Laboratory Press, New York, 1996.

[60] C. H. Waddington. *O.R. in World War 2: Operational Research Against the U-boat*. Elek, London„ 1973.

[61] P. C. Bressloff. Queueing theory of search processes with stochastic resetting. *Physical Review E*, 102(3), 2020.

[62] M. R. Evans, S. N. Majumdar, and G. Schehr. Stochastic resetting and applications. *J. Phys. A Math*, 53:193001, 2020.

[63] M. Vergassola, E. Villermaux, and B. I. Shraiman. 'Infotaxis' as a strategy for searching without gradients. *Nature*, 445(7126):406–9, 2007.

[64] A. Loisy and C. Eloy. Searching for a source without gradients: How good is infotaxis and how to beat it. *Proc. Roy. Soc. A*, 478:20220118, 2022.

[65] R. Monthiller, A. Loisy, M. A. R. Koehl, B. Favier, and C. Eloy. Surfing on Turbulence: A Strategy for Planktonic Navigation. *Phys Rev Lett*, 129(6):064502, 2022.

[66] Joseph K. Blitzstein and Jessica Hwang. *Introduction to Probability*. CRC Press/Taylor & Francis Group, 2019.

[67] B. Lacroix and J. Dumont. Spatial and temporal scaling of microtubules and mitotic spindles. *Cells*, 11(2), 2022.

[68] G. Ha, P. Dieterle, H. Shen, A. Amir, and D. J. Needleman. Measuring and modeling the dynamics of mitotic error correction. *Proc. Natl. Acad. Sci. U.S.A*, 121(25), 2024.

[69] B. Lacroix, G. Letort, L. Pitayu, J. Sallé, M. Stefanutti, G. Maton, A. M. Ladouceur, J. C. Canman, P. S. Maddox, A. S. Maddox, N. Minc, F. Nédélec, and J. Dumont. Microtubule Dynamics Scale with Cell Size to Set Spindle Length and Assembly Timing. *Developmental Cell*, 45(4):496–+, 2018.

[70] E. M. Rieckhoff, F. Berndt, M. Elsner, S. Golfier, F. Decker, K. Ishihara, and J. Brugues. Spindle scaling is governed by cell boundary regulation of microtubule nucleation. *Current Biology*, 30(24), 2020.

[71] L. D. Belmont, A. A. Hyman, K. E. Sawin, and T. J. Mitchison. Real-Time Visualization of Cell-Cycle Dependent Changes in Microtubule Dynamics in Cytoplasmic Extracts. *Cell*, 62 (3):579–589, 1990.

[72] J. Baumgart, M. Kirchner, S. Redemann, A. Bond, J. Woodruff, J. M. Verbavatz, F. J□licher, T. M□ller-Reichert, A. A. Hyman, and J. Brugu□s. Soluble tubulin is significantly enriched at mitotic centrosomes. *Journal of Cell Biology*, 218(12):3977–3985, 2019.

[73] H. Qian. Phosphorylation energy hypothesis: Open chemical systems and their biological functions. *Annual Review of Physical Chemistry*, 58:113–142, 2007.

[74] G. Lan, P. Sartori, S. Neumann, V. Sourjik, and Y. H. Tu. The energy-speed-accuracy trade-off in sensory adaptation. *Nature Physics*, 8(5):422–428, 2012.

[75] Terrell L Hill and Marc W Kirschner. Bioenergetics and kinetics of microtubule and actin filament assembly–disassembly. *International review of cytology*, 78:1–125, 1982.

[76] P. J. Foster, J. Bae, B. Lemma, J. J. Zheng, W. Ireland, P. Chandrakar, R. Boros, Z. Dogic, D. J. Needleman, and J. J. Vlassak. Dissipation and energy propagation across scales in an active cytoskeletal material. *Proc. Natl. Acad. Sci. U.S.A*, 120(14), 2023.

[77] M. Lynch and G. K. Marinov. The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences of the United States of America*, 112(51):15690–15695, 2015.

[78] R. Kiewisz, G. Fabig, W. Conway, D. Baum, D. Needleman, and T. M⊠ller-Reichert. Three-dimensional structure of kinetochore-fibers in human mitotic spindles. *Elife*, 11, 2022.

[79] RA Freitas. Nanomedicine: Basic Capabilities, Vol. 1. *Landes Bioscience: Georgetown, Texas*, 1999.

[80] John DC Little et al. A proof for the queuing formula $l = \lambda w$. *Operations research*, 9(3): 383–387, 1961.

[81] Ron Milo. What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays*, 35(12):1050–1055, 2013.

[82] Karthik P Jayapal, Siguang Sui, Robin J Philp, Yee-Jiun Kok, Miranda GS Yap, Timothy J Griffin, and Wei-Shou Hu. Multitagging proteomic strategy to estimate protein turnover rates in dynamic systems. *Journal of proteome research*, 9(5):2087–2097, 2010.

[83] Eran Eden, Naama Geva-Zatorsky, Irina Issaeva, Ariel Cohen, Erez Dekel, Tamar Danon, Lydia Cohen, Avi Mayo, and Uri Alon. Proteome half-life dynamics in living human cells. *Science*, 331(6018):764–768, 2011.

[84] Sanjoy Mahajan. *Street-Fighting Mathematics: The Art of Educated Guessing and Opportunistic Problem Solving*. MIT Press, Cambridge, Mass., 2010. ISBN 9780262514293 (pbk. alk. paper) 026251429X (pbk. alk. paper).

[85] George Stephanopoulos, Aristos A Aristidou, and Jens Nielsen. Metabolic engineering: Principles and methodologies. 1998.

[86] E Shacter, P Boon Chock, and ER Stadtman. Energy consumption in a cyclic phosphorylation/dephosphorylation cascade. *Journal of Biological Chemistry*, 259(19):12260–12264, 1984.

[87] R. Milo and R. Phillips. *Cell Biology by the Numbers*. Garland Press, New York, NY, 2016.

[88] H. Ishikawa and W. F. Marshall. Ciliogenesis: Building the cell's antenna. *Nature Reviews Molecular Cell Biology*, 12(4):222–234, 2011.

[89] A. Chien, S. M. Shih, R. Bower, D. Tritschler, M. E. Porter, and A. Yildiz. Dynamics of the IFT machinery at the ciliary tip. *Elife*, 6, 2017.

[90] R. Phillips, J. Kondev, J. Theriot, and H. G. Garcia. *Physical Biology of the Cell, 2nd Edition*. Garland Science, New York, 2013. (Illustrated by N. Orme).

[91] T. Kirchhausen. Three ways to make a vesicle (vol 1, pg 187, 2000). *Nature Reviews Molecular Cell Biology*, 2(3):216–216, 2001.

[92] Y. M. Wang, R. H. Austin, and E. C. Cox. Single molecule measurements of repressor protein 1D diffusion on DNA. *Physical Review Letters*, 97(4), 2006.

[93] P. C. Blainey, A. M. van Oijent, A. Banerjee, G. L. Verdine, and X. S. Xie. A base-excision DNA-repair protein finds intrahelical lesion bases by fast sliding in contact with DNA. *Proc. Natl. Acad. Sci. U.S.A*, 103(15):5752–5757, 2006.

[94] B. van den Broek, F. Vanzi, D. Normanno, F. S. Pavone, and G. J. Wuite. Real-time observation of DNA looping dynamics of Type IIE restriction enzymes NaeI and NarI. *Nucleic Acids Res*, 34(1):167–74, 2006.

[95] I. J. Finkelstein, M. L. Visnapuu, and E. C. Greene. Single-molecule imaging reveals mechanisms of protein disruption by a DNA translocase. *Nature*, 468(7326):983–987, 2010.

[96] P. Hammar, P. Leroy, A. Mahmutovic, E. G. Marklund, O. G. Berg, and J. Elf. The *lac* repressor displays facilitated diffusion in living cells. *Science*, 336(6088):1595–8, 2012.

[97] D. Normanno, L. Boudarene, C. Dugast-Darzacq, J. Chen, C. Richter, F. Proux, O. Benichou, R. Voituriez, X. Darzacq, and M. Dahan. Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat Commun*, 6:7357, 2015.

[98] R. Phillips, J. Kondev, J. Theriot, and H. G. Garcia. *Physical Biology of the Cell, 2nd Edition*. Garland Science, New York, 2013. (Illustrated by N. Orme).

[99] S. C. Weber, A. J. Spakowitz, and J. A. Theriot. Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm. *Phys Rev Lett*, 104(23):238102, 2010.

[100] M. Gabriele, H. B. Brandao, S. Grosse-Holz, A. Jha, G. M. Dailey, C. Cattoglio, T. S. Hsieh, L. Mirny, C. Zechner, and A. S. Hansen. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science*, 376(6592):496–501, 2022.

[101] D. B. Bruckner, H. Chen, L. Barinov, B. Zoller, and T. Gregor. Stochastic motion and transcriptional dynamics of pairs of distal DNA loci on a compacted chromosome. *Science*, 380(6652):1357–1362, 2023.