Phenomenological Models in Biological Physics: Cell Growth and Pluripotency Maintenance

by

Hyun Youk

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Physics

at the

Massachusetts Institute of Technology

June 2010

Signature of author: _____
Department of Physics
April 22, 2010

Certified by: _____
Alexander van Oudenaarden
Professor of Physics & Biology
Thesis Supervisor

Accepted by: _____
Krishna Rajagopal
Professor of Physics
Associate Department Head for Education

Phenomenological Models in Biological Physics: Cell Growth and Pluripotency Maintenance

by

Hyun Youk

submitted to the Department of Physics
on April 22, 2010, in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy in Physics

Abstract

A persistent challenge in quantitatively modeling a biological system is that the system often involves many components and just as dizzying number of interactions among those components. To further complicate matters, the parameters that characterize those interactions and components, like the rates of chemical reactions and concentrations of molecules inside the cell, have evaded detection by the conventional experimental tools. How does one model a system whose crucial parameters are unknown? And even if we know all the parameters inside the cell, there is an increasing uneasiness among many researchers that just writing down an equation for every interaction and components of the system is not practical. Crucially, it is not clear that such an extensive many-parameter model would always enhance our understanding of the complex biological system. A phenomenological model that involves just a few essential, easily measurable parameters that capture the essence of the complex biological system may provide insights that a many-parameter, large scale model may not provide. In this thesis, we describe our attempts at obtaining such a model for two complex biological systems: 1.) Cell growth as a result of glucose metabolism, and 2.) *in vitro* maintenance of the embryonic stem cell's pluripotency by a complex transcriptional network.

Thesis supervisor: Alexander van Oudenaarden
Title: Professor of Physics & Biology

Acknowledgements

First, I thank Alexander van Oudenaarden for allowing me to join his lab three and a half years ago even though as a physicist, I barely knew the difference between mRNA and DNA at the time. Alexander gave me a great deal of freedom during my time in his lab. He allowed me to design my own experiments and pursue them at my own pace, even though these experiments were often completely outside the scope of the lab's expertise, theme, or interests. When entering grad school, I wanted to find a lab in which I could grow as an independent researcher and Alexander's lab was just the perfect spot for me. Furthermore, Alexander was also generous in suggesting a new system to think about and setting up collaborations to explore it, leading to the work on embryonic stem cells described in this thesis. I thank Alexander for his advice, time and letting me realize how ridiculously fun it is to work at a lab bench.

I had the enormous pleasure and privilege to work with Professor Roland List at the University of Toronto all throughout my undergrad years. Roland hired me just after I had finished my freshmen year despite the fact that I didn't know anything about atmospheric physics, his research specialty, or about physics in general. In his seventies and as an emeritus professor, he still came into his office almost everyday with an unmatched enthusiasm and an optimistic outlook on life and his research. Since he did not have any graduate students or postdoctoral researchers to help him with his research, he gave me an enormous amount of attention and time. He believed that I should be treated like a graduate student. He taught me how to come up with my own research project, plan a method of attack once I had come up with a suitable problem, carry out the research independently, and then present and write a paper on the problem I had solved. But most importantly, through our nearly daily interactions, I realized that I wanted to be someone like Roland – someone who was naturally curious about everything and really loved something (science), and pursue it as a career. I thank Roland (who still comes into his office everyday, now in his eighties) for infecting me with his enthusiasm for science and making me realize that it was okay to be curious about so many things, and that I could do something to resolve those curiosities.

Arjun Raj has been a great mentor during my time in the lab. It amazed me that he could be so generous with his time despite his tight schedule. I learned how to ask the right questions, distinguish the trivial from the important, and how to plan and carry out experiments efficiently so that I learn something from every experiment I perform. When I was stuck on problems or had an exciting idea popping into my head, he was the "go-to" guy for advice and feedback. He was also a great person with whom to exchange quirky jokes and eccentricities that we both seemed to be abundant with. He is also credited as the first person who's experiment I have ruined thanks to his generosity in allowing me to squirt his valuable plasmids into the wrong tube when I was first learning how to use a pipette. I'm sure that his new lab is a welcoming environment that will serve as a fountain of original ideas in systems biology for years to come.

Jeff Gore has also been a great mentor during my time in the lab. Although Jeff and I did not always see eye-to-eye on science and politics, and we have very different styles of doing science, he has taught me to be a better critic of my own work as well as the work of others. He is also a great person to seek advice on career in science, life issues, and how to deal with the ups and downs of being a scientist. He let me take part in his project, resulting in my first foray into biological experiments. He is also credited as the person who taught me how to use pipettes when I joined the lab.

Scott Rifkin was a generous and a warm-hearted mentor, and a fun person to share a lab bench with. I was often struck by Scott's unmatched depth of knowledge and his scholarly approach to science. He also taught me how to write clearly and gave me support when things weren't going well on my side of the bench during my work on yeast. Scott also shared my almost irrational distaste for formalities and rigidities of life. It was a great fun to poke fun at those things with him during the countless late nights of pipetting at our bench while listening to political / historical podcasts streaming from his computer.

The past three and a half years in Cambridge have been a blast thanks to awesome friends outside the lab, particularly Alfonso Cecco, Joel Yuen, Tracy Washington, Solomon Bisker, Rachel Delano, Melva James, Nellie Nikolov, and Wasim Malik. I also thank the past and present members of the van Oudenaarden lab. In particular, I'm thankful to Shankar Mukherji, Bernardo Pando, Dale Muzzey, and Qiong Yang. Monica Wolf has ensured a smooth running of the van Oudenaarden lab during my last year of graduate school all while maintaining a great sense of humor and optimism that rubbed off on me. During the last year, I've been fortunate to collaborate with Lauren Surface in Laurie Boyer's lab on mESC work presented in this thesis. I'm thankful to Lauren and all the other members of the Boyer lab for welcoming me to their space, teaching me the basics of mammalian cell culturing, and dealing with all my silliness.

I also thank the Natural Sciences and Engineering Research Council of Canada (NSERC) for funding my work through its graduate fellowship (PGS-D2) program.

The most important person that I thank is my mother. She is an essential person in my life and has made me the person that I am today. For a large part of my life, she has selflessly worked long hours in odd jobs so that my brother and I could attend college, something that she could not afford for herself when she was growing up. When I doubted if I could attend graduate school, my mom was the first person to encourage me even though it meant that she had to sacrifice her health to work longer at these odd jobs because I would not be able to financially support her for several more years. Through the examples set by her over the years, she has taught me to believe in myself, to never give up when faced with challenges, give everything I've got in order to reach my goals, and get through challenges with humor. As these traits are responsible for getting me to this point in my developing career in science, it is only fitting that I dedicate this thesis to her.

Table of contents

Contents

1. Motivation

"Which end is nearer to God, if I may use a religious metaphor, beauty and hope, or the fundamental laws? I do not think either end is nearer to God. To stand at either end, and to walk off that end of the pier only, hoping that out in that direction is the complete understanding, is a mistake. And to stand with evil and beauty and hope, or to stand with the fundamental laws, hoping that way to get a deep understanding of the whole world, with that aspect alone, is a mistake."

Richard P. Feynman, "The Character of Physical Law". Chapter 5. (1965).

Quantitatively modeling biological systems: A challenge

A persistent challenge in quantitatively modeling a biological system is that the system often involves many components and just as dizzying number of interactions among those components. To further complicate matters, the parameters that characterize those interactions and components, like the rates of chemical reactions and concentrations of molecules inside the cell, have evaded detection by conventional experimental tools. How does one model a system whose crucial parameters are unknown? And even if we know all the parameters inside the cell, there is an increasing uneasiness among many researchers that just writing down a differential equation for every interaction and components in the system is not practical. Crucially, it is not clear that such an extensive many-parameter model would always enhance our understanding of the complex biological system. But this is not to say that individual chemical reactions and the interaction between the components that make up the system are not important. Richard Feynman's statement at the top of this page elegantly, and perhaps overly poetically, describes this tug of war between thoroughly modeling the individual interactions and modeling the collective behavior that emerges from a large number of those interactions without accounting for the individual interactions. Physicists know very well, particularly from their work in many-body physics, that even a system with many particles in which the interaction between every pair of particles is identical, can be an intractable by modern analytical techniques. But inside a cell, the interactions among the particles are different, spanning different timescales, concentrations, and spatial extents. This is one of the main challenges in quantitatively modeling biological

systems. Moreover, just as only a few types of particle-particle interaction in a solid-state system can lead to fascinating collective behaviors, the much larger number of interaction types inside a living cell produces fascinating array of collective phenomena. In this thesis, we explore two such behaviors: 1.) Cell growth as a result of myriad interactions, and 2.) *in vitro* maintenance of the embryonic stem cell's pluripotency by a complex transcriptional network.

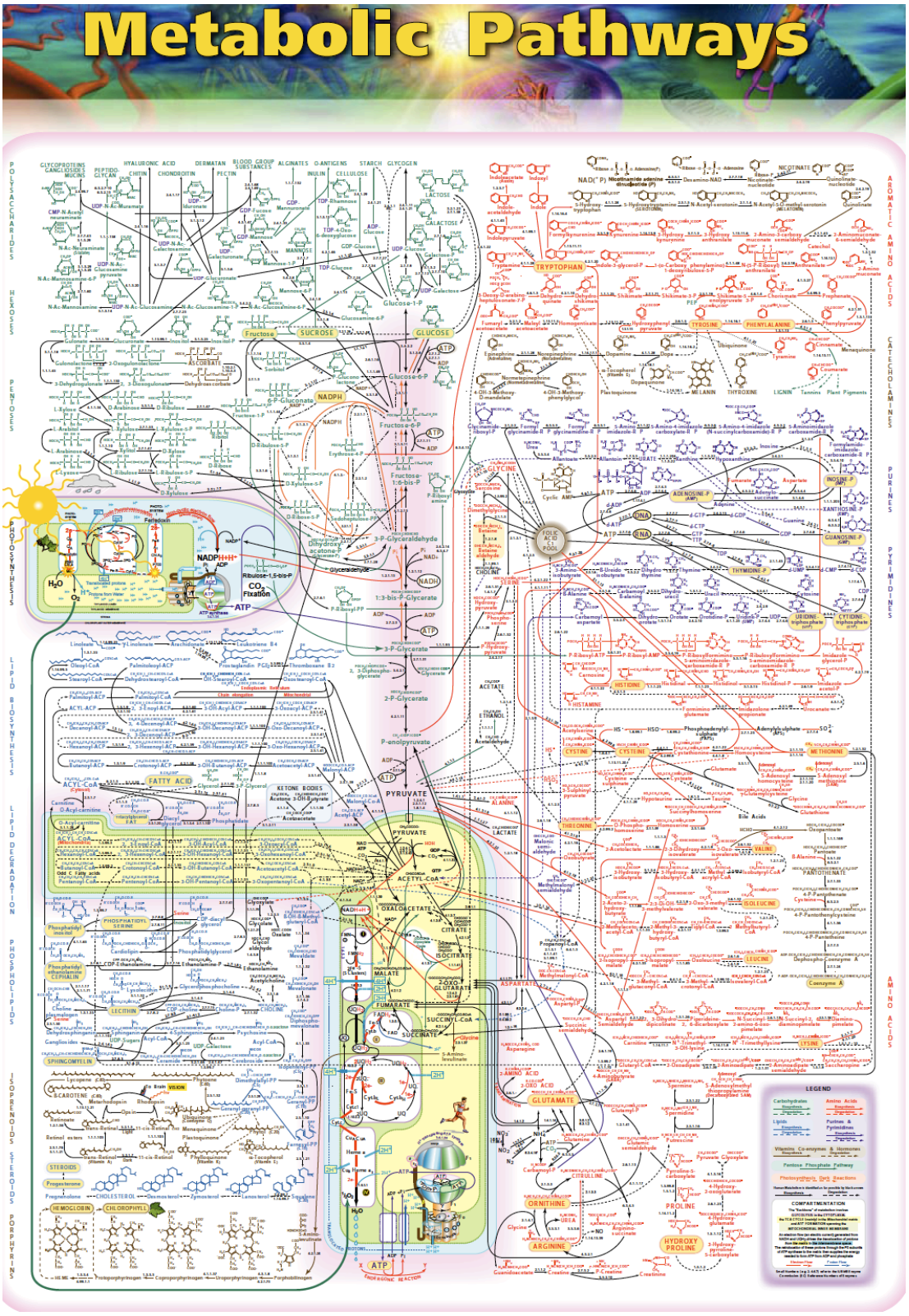


Figure 2.1. Complex network of chemical reactions is responsible for cell growth on glucose. (Source: IUBMB-Sigma-Nicholson metabolic pathways chart).

2. Quantitative models of passive transporters and metabolism: A brief overview

Chapter summary

This chapter provides a brief overview of simple mathematical models that are commonly used for studying passive transporters of glucose. It also surveys some conventional approaches to modeling sugar metabolism in cells, a complex network that underlies our experimental study in chapter 3.

2.1. How cells obtain energy from glucose

Every living organism must make all of its components from resources it gathers from the environment and use those components for survival and reproduction. To fuel these processes, all living organisms must extract energy from their surroundings. Among the different sources of energy, sugars play a central role in almost all organisms. Plants and primitive bacteria extract energy out of the photons from the Sun and use it to generate sugars for consumption - a process known as photosynthesis. Certain single-cell organisms like E. coli uptake ready-made sugars from their surroundings. Among the sugars that both single-celled and multicellular oganisms relie on, glucose is single-handedly the most preferred and important sugar molecule. This is partly seen from the fact that most other forms of simple sugar (monosaccharides) and complex sugar molecules (polysaccharides) are converted to glucose first, and then the cell digests the resulting glucose. In addition, organisms as diverse as yeast and humans can derive the other monosaccharides that are necessary to form polysaccharides for furnishing various parts of the cell (glycan biosynthesis) from glucose. Glucose is therefore a versatile molecule that is both a primary source of energy and an important building block of cellular components.

A striking result of evolution is that many organisms, from yeast to humans, use a vast network of very similar chemical reactions to extract energy for metabolizing a glucose molecule. Glucose metabolism involves extracting energy from glucose and storing it into the phosphates of Adenosine TriPhosphate (ATP). The cell uses ATP to transport the stored energy and release it for usage by breaking the phosphate bonds when needed. But a lot more happens than that once a glucose molecule enters the cell. The cell makes a diverse

set of enzymes to process the numerous bi-products that result in the process of glucose metabolism. As some of those bi-products are responsible for fueling other parts of cell such as amino-acid and glycan biosyntheses, the number of enzymes involved in processing each of the bi-products can be quite complicated. Indeed, the mind-boggling "wiring diagram" that shows just some of these chemical processes (Figure 2.1) gives a sense of how complex these processes are. Cell growth results from the concerted efforts of these chemical reactions (Figure 2.1) and many others not shown in this wiring diagram. Despite having identified and studied most of the enzymes involved in these processes, it has been a long standing challenge to build a widely applicable quantitative model that describes exactly how fast a cell will divide as a result of these processes. We will explore some conventional models that attempt to do so later in this chapter.

2.2. Glucose transporters

Transporters residing within the cell membrane are vehicles that move biomolecules, including glucose, between the extracellular and intracellular environments. Given the preference for glucose, many organisms suppress production of transporters and cellular machineries designed for digesting alternative forms of sugar when the cell senses glucose in its environment. There are three types of glucose transporters in nature. One of them is a passive transporter. Passive transporters carry glucose across the cell membrane, moving glucose from a region with higher concentration to a region with a lower concentration of glucose. Passive transporters in the budding yeast are known as HeXose Transporters (HXT) in the budding yeast and GLUcose Transporter (GLUT) in mammalian cells, including in some human cells. The other two types of transport are both energy-dependent. One of them involves hydrolysis of ATP to use the liberated chemical energy for glucose transport, while the other uses sodium ion gradient (SGLT in humans, which are found in epithelial and kidney cells). In this thesis, we deal with only the passive transport.

2.3. Quantitative model of passive transporters

Passive transporters are governed by thermal fluctuations and transport biomolecules from a region of high concentration to a region of low concentration (i.e. "down the concentration gradient"). The fact that such important transporters like GLUT2 (responsible for glucose transport in pancreatic beta cells which in turn release appropriate levels of insulin)

and HXTs (essential for uptake of glucose in S. Cerevisiae) are passive transporters that rely on seemingly unreliable thermal noise for their operation, is at first, alarming. There are at least two ways for a cell to cope with the fluctuations. One is that a single transporter binds to a substrate (that its designed to transport) many times within a given unit of time, thus the transporter has many chances to bring the substrate to the other side of the membrane. A single GLUT transporter, for instance, experiences greater than 100,000 events of glucose-binding per second on average. Not all these binding events lead to a successful transport of the glucose to the other side of the membrane due to the inefficiency of the bond that forms between the glucose and the transporter but by having a lot of attempts in a given amount of time, the transporter can increase its chance of a successful transport of the glucose molecule to the other side of the membrane. Another way to deal with the fluctuations is by having a large number of transporters present on a cell membrane at a given time. For example, there are typically some 100,000 or more HXTs covering the cell membrane of a budding yeast. From modeling perspective, the large number of transporters and the binding events allow us to avoid the Master equation approach. Instead, we can use a set of deterministic equations to look at the mean-behavior of the passive transporters.

A common starting point for studying passive transporters is the model known variously as the Alternating-conformer model, or the 3-state model which was first proposed by Widdas in 1950s. This model uses enzyme-substrate kinetics akin to Michaelis-Menten kinetics and states that a transporter can cycle through any one of the three discrete states (Figure 2). This model has been viewed as the gold standard for fitting measured transport parameters of various passive transporters. While fitting parameters of the model to the specific transporter is useful in illuminating the behavior of the particular transporter of interest, one could seek a global understanding of passive transporters by asking questions such as "What trends does one observe in the behavior of any passive transporter as one gradually changes the substrate binding affinity of the transporter?", instead of "What is the glucose binding affinity of HXT2?". Its the questions of the former type that we are concerned with in this section. After all, if there is a common belief that such a wide variety of passive transporters seem to operate in the substrate-enzyme type kinetics that the 3-state model describes, then this is an opportunity for us to study a biological mechanism in its generality, independent of the organism.

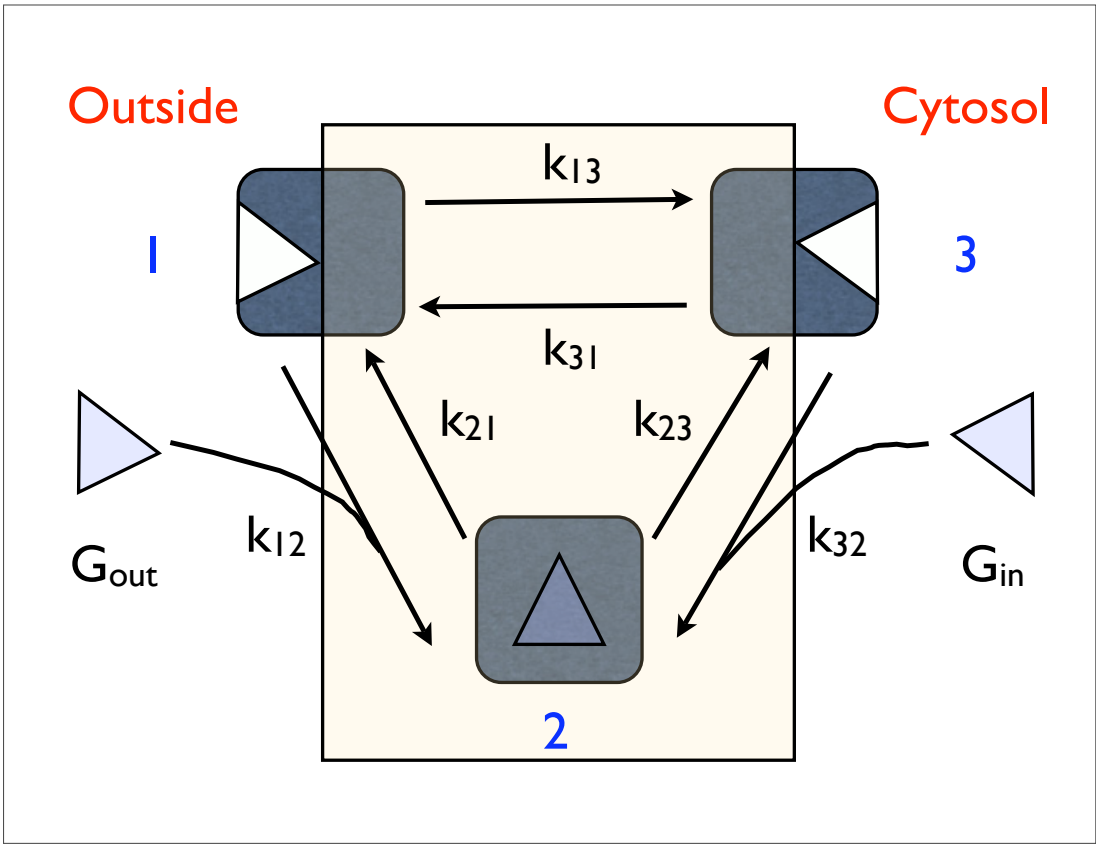


FIG. 2: **Three-state model of a passive transporter**

Here we implement the 3-state model of a generic passive transporter. The original model of Widdas actually involves at least four states, each representing the possible state of the transporter. They are: **1.** transporter facing outward, waiting to capture glucose, **2.** transporter facing outward with the glucose in its grip, **3.** transporter with glucose in its grip and facing the cytoplasm, and **4.** transporter facing inside the cytoplasm with glucose released from its grip. Our 3-state simplification comes from focusing on a time scale that is much longer than time needed for the conformational change from states 2 to 3 (and vice versa) to occur. Thus, our three states are: **1.** same as state 1 of the 4-state model, **2.** transporter with glucose in its grip (and can be facing either out or into the cytoplasm), and **3.** same as state 4 of the 4-state model. The cost of this simplification is that we no longer have information about the conformation of transporter with glucose in its grip.

The analysis of this model proceeds through the standard detailed balance calculations. Assuming that there is a source of glucose outside the cell and a sink (metabolism) inside the cell maintaining a constant gradient of glucose across the membrane, the steady-state flux of glucose can be obtained. The law of mass action puts a constraint on the activity constants $\{k\}$. Our system is then described by:

$$\begin{pmatrix} k_{31} \\ k_{32}[G_{in}] \end{pmatrix} = \begin{pmatrix} k_{12}[G_{out}] + k_{31} + k_{13} & k_{31} - k_{21} \\ k_{32}[G_{in}] - k_{12}[G_{out}] & k_{21} + k_{23} + k_{32}[G_{in}] \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{2.1}$$

where x_i and k_{ij} are the fraction of transporters in state i, and activity constant characterizing transporter transition from state i to j respectively. $[G_{out}]$ and $[G_{in}]$ are extracellular and intracellular glucose concentrations respectively.

The 2 x 2 matrix in Eqn. (2.1) is guaranteed to be invertible because it's determinant cannot be zero (due to activity constants and glucose concentrations always being non-negative). The system can thus always be solved by computing its inverse *regardless* of the values of $\{k_{ij}\}$.

I. Symmetric case

We first consider a symmetric passive transporter, one whose response to extracellular concentration of glucose is exactly the same as that of intracelluar concentration (i.e. $k_{ij} = k_{ji}$). We now have a reduction to three activity constants:

$$k_{12} = k_{32} \equiv k^+ \qquad \text{association} \tag{2.2a}$$

$$k_{23} = k_{21} \equiv k^- \qquad \text{dissociation} \tag{2.2b}$$

$$k_{13} = k_{31} \equiv k \qquad \text{flipping} \tag{2.2c}$$

The steady-state glucose influx R_{sym} into the cell is (modulo proper units):

$$R_{sym}(\alpha, \beta, k, [G_{out}], [G_{in}]) = k * F(\alpha, \beta, [G_{out}], [G_{in}]) \tag{2.3a}$$

$$F(\alpha, \beta, [G_{out}], [G_{in}]) = \frac{\alpha([G_{out}] - [G_{in}]))}{\alpha(\beta + 1)([G_{out}] - [G_{in}]) + \alpha^2[G_{out}][G_{in}] + 4\beta} \tag{2.3b}$$

where $\alpha = \frac{k^+}{k^-}$ and $\beta = \frac{k}{k^-}$. One reason for writing down the influx as such is that it facilitates the revelation of **scaling behavior**. Suppose we have a glucose influx (i.e. there's more glucose outside the cell). Now, suppose we double all the affinities but keep the concentration difference fixed. Then we'd expect the influx to double as well. This is clearly seen in Eqn. (2.3) since α and β remain invariant (and thus F is invariant under the doubling of afffinities), and only the k doubles. Thus, we see that in the symmetric model

$$R_{sym}(\alpha, \beta, 2k, [G_{out}], [G_{in}]) = 2k * F(\alpha, \beta, [G_{out}], [G_{in}]) = 2influx(k^+, k^-, k, [G_{out}], [G_{in}]) \tag{2.4}$$

This makes intuitive sense. It's interesting to note that R_{sym} doesn't just depend on concentration difference, but rather on their averages (both geometric and arithmetic) as well.

I. A. Symmetric : Michaelis-Menten

As expected, when either the extracellular or intracellular glucose concentration is zero, the influx rate (2.3) reduces to Michaelis-Menten influx (here, we've set [G$_{in}$]=0):

$$R_{sym}(k^+, k^-, k) = R_M \frac{[G_{out}]}{[G_{out}] + K_M} \tag{2.5a}$$

$$R_M = \frac{kk^-}{k + k^-} \tag{2.5b}$$

$$K_M = \frac{4V_M}{k^+} \tag{2.5c}$$

Thus, increasing affinity for glucose (larger k^+) accompanies decrease in K_M, which is in accordance with what we expect for a high affinity transporter. If the affinity is high, the extracellular glucose doesn't have to be so high to reach the maximum influx (i.e. K_M can be low), as is reflected by Eqn (2.5c). Also, it's interesting to note that the maximum flux of a symmetric transporter R_M is *independent* of association affinity k^+. This can be interpreted as follows: For a transporter to reach its full capability (transporting at a rate near R_M), a sufficiently large amount of extracellular glucose must be present (i.e. $[G_{out}] \gg K_M$). In this case, how much affinity the transporter has for glucose is irrelevant as not finding enough glucose to bind is no longer the issue. What determines R_M then is how

fast the other functions of the transporter (release glucose it has caught (k^-) then face back out (k) to catch the next glucose molecules) are carried out. The faster they are (large k^- and k), the larger the R_M. Of course, the caveat is that k^+ helps in setting how large is "sufficiently large".

I. B. Symmetric : Finite intracellular glucose concentration

Although by this point, the expressions already become burdensome to work with by hand, we can still draw analytical conclusions and meaningful interpretations. Suppose we now allow for a finite $[G_{in}]$, then we seek the **optimal affinity for glucose** (optimal value of k^+), which is found to be

$$k^+_{optimal} = 2\sqrt{\frac{kk^-}{[G_{in}][G_{out}]]}} \tag{2.6}$$

First, note of caution: above expression only holds for non-zero concentrations of glucose. If either one of the two concentrations is zero, then there is no optimal k^+ (we'd get Michaelis-Menten in which it is always desirable to make k^+ as large as possible).

We can interpret Eqn. (2.6) by noting that making k and k^- larger, leads to increasing the release of glucose by transporter and rate of conformational change (facing out to in, and vice versa). But we saw from expressions (2.3a) and (2.3b) that R_{sym} scales *linearly* with (k^+, k^-, k), which hints that to keep up with dissociation and flipping transitions, the affinity should also increase. Another way to look at this is to note that larger k and k^- means that there will be more *mistakes* made by the transporter: releasing glucose back out, spontaneously flipping to face the interior of cytoplasm *before* even grabbing glucose from outside. The relationship between influx R and k^+ is shown in Figure 3.

II. Asymmetric transport behavior

If the transporter is asymmetric, we can still draw some meaningful interpretations.

II. A. Asymmetric behavior: Michaelis-Menten

By assuming a zero intracellular glucose concentration, we once again obtain Michaelis-Menten influx R:

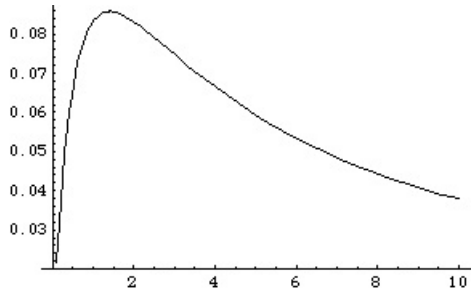


FIG. 3: **Influx of glucose as a function of glucose binding affinity** k^+: x-axis is k^+ and y-axis is the influx of glucose R. The values used here are (modulo proper units) k=1, k^-=1, $[G_{out}]$=2, and $[G_{in}]$=1. The maximal influx occurs at $k^+ = 2\sqrt{\frac{kk^-}{[G_{in}][G_{out}]]}}$.

$$R = \frac{R_M[G_{out}]}{[G_{out}] + K_M} \tag{2.7a}$$

$$R_M = (\frac{1}{k_{23}} + \frac{1}{k_{31}})^{-1} \tag{2.7b}$$

$$K_M = R_M(k_{21} + k_{23})(\frac{1}{k_{12}k_{23}} + \frac{1}{k_{32}k_{21}}) \tag{2.7c}$$

Again, the interpretations given for the symmetric case can be used to explain why R_M is independent of transporter's affinity for glucose association k^+.

II. B. Advantage of asymmetry : Michaelis-Menten

Suppose we assume that the intracellular glucose concentration is zero, and that there's a steady state influx of glucose from outside the cell. For a given amount of glucose, is it better for the transporter to be equipped with a symmetric response or an asymmetric response? If asymmetry is better, how much asymmetry is beneficial? We are now ready to answer these questions analytically.

The situation we envision here is described by Michaelis-Menten for asymmetric transporter (Eqn(2.7)). We first introduce dimensionless variables α , β, and γ, where

$$\alpha \equiv \frac{k_{12}}{k_{32}} \qquad \text{association comparison: } \frac{out \to in}{in \to out} \tag{2.8a}$$

$$\beta \equiv \frac{k_{31}}{k_{13}} \qquad \text{flipping comparison: } \frac{in \to out}{out \to in} \tag{2.8b}$$

$$\gamma \equiv \frac{k_{23}}{k_{21}} \qquad \text{dissociation comparison: } \frac{out \to in}{in \to out} \tag{2.8c}$$

Symmetric transporter corresponds to $\alpha = \beta = \gamma = 1$. Second law of thermodynamics imposes the constraint: $\alpha\beta\gamma = 1$ regardless of symmetry. Equation (2.7c) can be written with the aid of these variables as

$$K_M = \frac{R_M}{k_{12}}(1+\gamma)(\alpha + \frac{1}{\gamma}) \tag{2.9}$$

It may look like we haven't done much. But in fact, we have written K_M that facilitates seeing why asymmetry is better. Given a fixed $[G_{out}]$, one way to maximize influx is by making R_M large while making K_M small. Notice that

$$R_M = R_M(k_{23}, k_{31}) \tag{2.10a}$$

$$K_M = K_M(R_M, k_{12}, \alpha, \gamma) \tag{2.10b}$$

Here's a proof that asymmetry results in larger influx of glucose: Suppose we assign values k_{12}, k_{23} and k_{31}. Eqn. (2.10a) shows that this would fix R_M and what remains to be determined are α and γ. This corresponds to assigning specific values to the activity constants for only the *forward cycle* in the 3-state model. Notice that whatever value we assign to α (association comparison) and γ (dissociation comparison), we can always satisfy the second law of thermodynamics by choosing $\beta = \frac{1}{\alpha\gamma}$. Now, here comes the advantage of writing K_M as we did in Eqn (2.9); α and γ are factored out. In a symmetric transporter ($\alpha = \gamma = 1$), and hence

$$(1+\gamma)(\alpha + \frac{1}{\gamma}) = 4 \qquad \text{Symmetric transporter} \tag{2.11}$$

It's easy to see that K_M is smaller (thus larger influx) for asymmetric transporter. In particular, if we look at **slightly asymmetric** situation, in which $\alpha = 1 + \delta$ and $\gamma = 1 + \delta$ (where $|\delta| \ll 1$), we find

$$(1+\gamma)(\alpha+\frac{1}{\gamma}) \approx 4 + 2\delta + \mathcal{O}(\delta^2) < 4 \quad \text{(asymmetric transporter, } \delta < 0) \tag{2.12}$$

For small (i.e. to $\mathcal{O}(\delta)$) and negative δ, the asymmetric transporter thus results in a larger influx of glucose than the symmetric transporter, for the same amount of extracellular glucose concentration. Notice that we have investigated only one type of asymmetry here, namely that $\alpha = \gamma$, but this was sufficient for our purpose. In general, there is no simple way to determine what value of $\{k\}$ would result in maximal R_M. We'd have to resort to a brute force to locate the optimal set $\{k\}$. But what we have just extracted analytically is the fact that asymmetry **always** leads to a larger influx of glucose when $[G_{in}] = 0$ (and by extension, $[G_{in}] \approx 0$).

2.4. Design principles underlying passive transporters: Application to hexose transporters (HXTs) in yeast

Passive transport is the only method of glucose transport in the budding yeast. The budding yeast has eighteen hexose transporter (HXT) genes. But of these, only six of the HXT genes are expressed and the remaining genes remain silent (they are often called "pseudo genes"). The budding yeast uses a combination of these six Hxts (Hxt1-4, Hxt6, Hxt7), varying the amount of each type of Hxt made depending on how much glucose is available in its environment, to uptake glucose from its surrounding. Each of these Hxts have values of K_m and V_m that differ from each other, signifying their different uptake capabilities.

In the previous section, we found that in a given environment (fixed $[\mathrm{G}_{in}]$ and $[\mathrm{G}_{out}]$), a passive transporter whose half-saturation value K_m and maximal transport rate V_m with the following relationship has the largest net uptake rate

$$K_{m,optimal} = 2V_m\sqrt{\frac{[G_{in}][G_{out}]}{kk^-}} \tag{2.13}$$

A transporter with a "low" K_m is referred to as a high-affinity transporter. For example, those Hxts with high-affinity (Hxt2, Hxt6, Hxt7) are expressed highly in a "low" glucose environment ([glucose] $\sim 0.1\%$) whereas the low-affinity (those with larger K_m: Hxt1, Hxt3, Hxt4) are expressed highly in a glucose rich ([glucose] $> 1\%$) environment. Although the

magnitude of K_m does indeed measure the affinity of a transporter in this sense, it doesnt tell the whole story in light of Eqn. (2.13). For instance, given two transporters with similar V_m but with two distinct K_m, its possible that the one with the larger K_m (lower-affinity) has a larger uptake rate than the other one if it satisfies expression (2.13). But traditionally, the one that actually performs better would be known as having a lower affinity than the one that actually performs more poorly. So relative magnitudes of K_m of hexose transporters or any other passive transporters, for that matter, dont tell the whole story.

As an example of above argument, consider the two hexose transporters Hxt4 and Hxt7 in the budding yeast. They have similar values of V_m ($V_{m,Hxt4}$ = 12 nmol min^{-1} mg^{-1}, $V_{m,Hxt7}$ = 11.7 nmol min^{-1} mg^{-1}). If we assume that both have about the same k^- and k rates, and that both are symmetric transporters, then the net influx difference between the two is due to difference in their k^+. Under this assumption, from Reifenbergers estimates (1997), we estimate that k^+_{Hxt4} is about 1/9 of k^+_{Hxt7}. While we cannot directly estimate any of the rate constants, we can demonstrate some design principles underlying above argument by assigning some values to the rate constants. Figures 4 and 5 show these ideas graphically.

Before studying asymmetric transporters, we briefly note that Eqn. (2.13) can be written as

$$K_{m,optimal} = 2\sqrt{[G_{in}][G_{out}]}\frac{\sqrt{\sigma}}{1+\sigma}, \tag{2.14}$$

where $\sigma = k \ / \ k^-$. σ is a measure of how much rate limiting the flipping transition is compared to the substrate dissociation rate. As we'll state later, it is possible to estimate σ by measuring the iso-counter transport of radioisotope-tagged substrates. In fact, it will turn out that σ is a natural parameter for use in understanding the behavior of asymmetric transporters.

Symmetric transporters are known to exist in various cells. For example, certain transporters in liver cells in humans are known to have the symmetric behavior and there is some speculation that some of the primary Hxts may indeed be nearly symmetric.

A. Expression levels of HXTs: Symmetric model

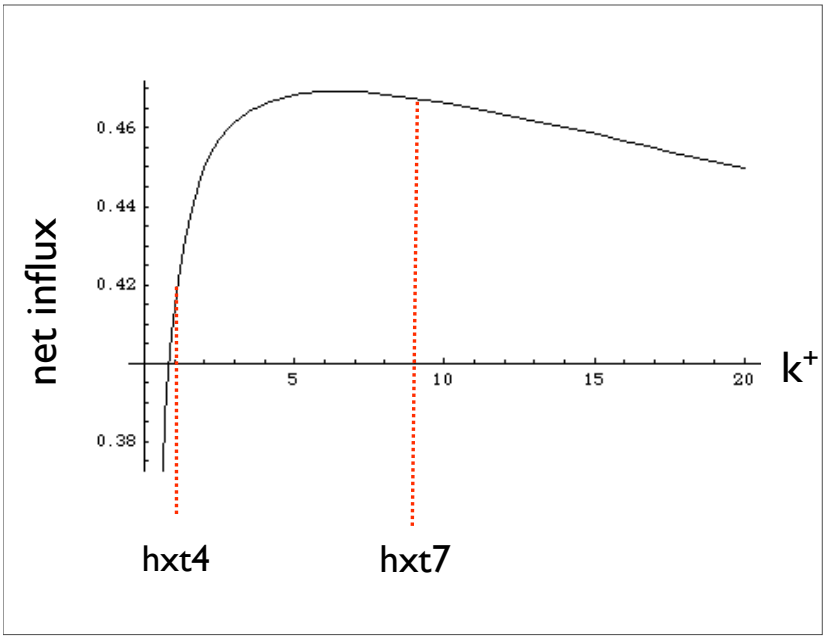


FIG. 4: By increasing the extracellular glucose concentration to $[G_{out}] = 60$ mM and keeping intracellular glucose concentration fixed at $[G_{in}] = 0.01$ mM, Hxt4 now achieves a higher uptake rate than Hxt7 despite its larger K_m.

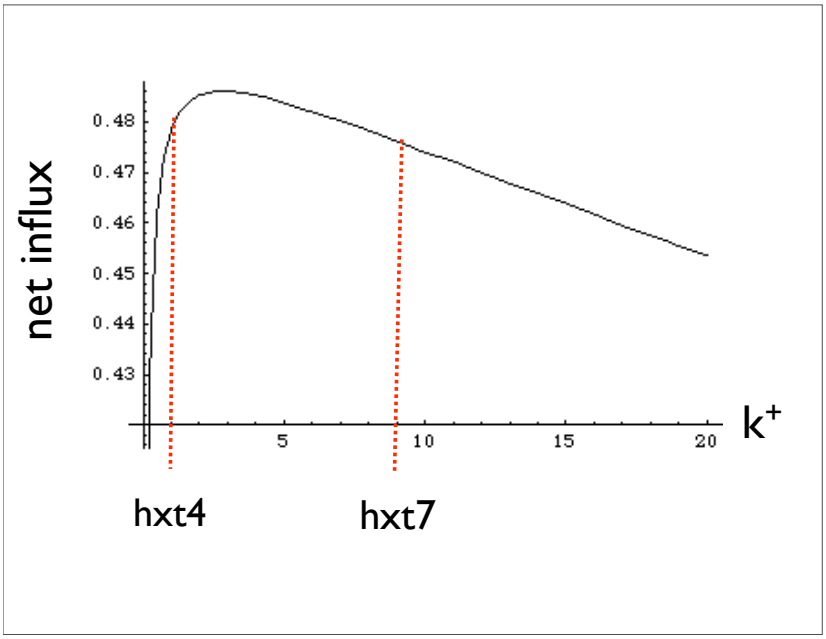


FIG. 5: At $[G_{out}] = 10$ mM, $[G_{in}] = 0.01$ mM, $k = k^- = 1$ (for both Hxt4 and Hxt7), and k^+=1, Hxt4 has lower influx than Hxt7.

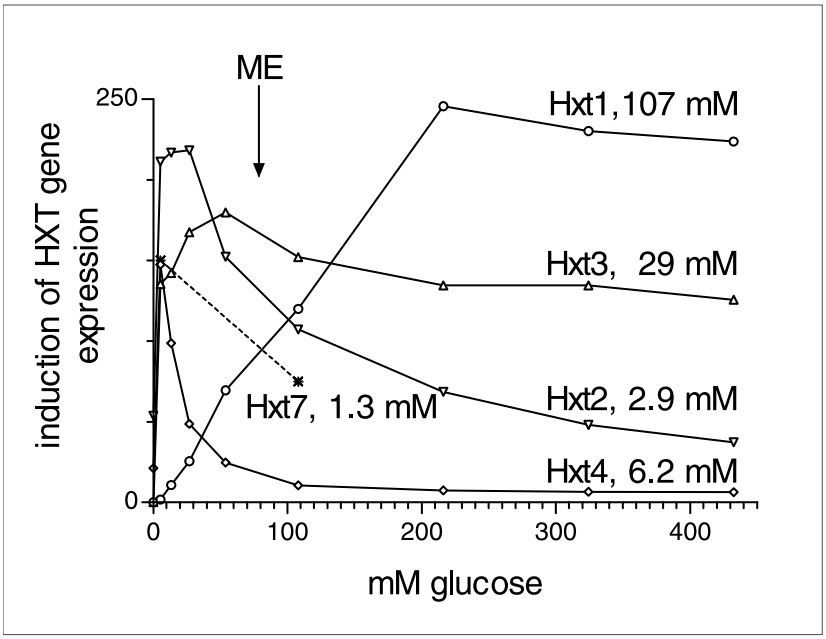


FIG. 6: **Expression levels of the primary HXT genes in the budding yeast measured by β-galactosidase assay (Taken from Maier et. al. (2002))**

We can try to understand the expression levels of HXT1-7 using the symmetric model. While there is no direct proof yet that any of the Hxts are symmetric transporters, we can extract some simple design principles by assuming they are. First, note the following expression of HXTs as a function of external glucose concentration (taken from Fig. 6 of Maier et. al.(2002)):

The values of K_m and V_m were obtained by two different methods (1. counter transport using radioisotope tagging, and 2. Typical initial uptake measurement during the first five seconds). Both have yielded values of V_m and K_m that match each other closely. These values are summarized in Table 1.

	Hxt1	Hxt2	Hxt3	Hxt4	Hxt6	Hxt7
K_m	107	2.9	29	6.2	0.9	1.3
V_m	50.9	15	18.5	12	11.4	11.7

TABLE I: V_m and K_m of Hxts in budding yeast: K_m is measured in mM. V_m is measured in nmol min^{-1} mg^{-1}.

Four main observations can be made:
1. With the exception of Hxt1, the rest of the essential HXTs have a similar V_m.
2. We can roughly lump the Hxts into three sets {Hxt1}, {Hxt3}, and {Hxt2, Hxt6, HXT7} according to the relative magnitudes of their K_m.
3. Looking at the expression levels of HXTs (Figure 6), we see that those Hxts with relatively low values of K_m tend to be suppressed more rapidly as extracellular glucose increases than those Hxts with larger values of K_m. For example, HXT2, HXT4, and HXT7 are suppressed more rapidly as $[G_{out}]$ increases (all have $K_m \sim 3$). Expression of HXT3 is suppressed more slowly than HXT2, HXT3, and HXT7 ($K_m \sim 30$). Finally, suppression of HXT1 is slowest of them all as G_{out} increases.
4. Lower K_ms are clustered together and there is a lone high-K_m transporter (HXT1).

We now try to explain observations #2 - #4 using observation #1 and under the assumption that Hxts are symmetric transporters. In this process, we can glimpse into how the expression of family of symmetric transporters may be regulated.

Recall that the symmetric transporter is characterized by three rate constants, k, k^- and k^+ (where we denote $\sigma = k/k^-$). Lets assume that all the Hxts have about the same σ. Then since $V_m = k \ / \ (1+\sigma)$, this means that Hxt2-7 have about the same k value, according to observation #1. This, in turn, means that k^- is about the same for Hxt2-7. This would mean that these Hxts differ mainly in their k^+ (binding rate) values, which in turn directly tunes their K_m. We noted that among all the transporters with the same k, and k^-, there is $K_{m,optimal}$ (for fixed values of $[G_{out}]$ and $[G_{in}]$) that, if possessed by the transporter, results in the maximum net influx among them all. First, note that

$$K_{m,optimal} \sim \sqrt{[G_{in}][G_{out}]} \tag{2.15}$$

Teusink et al. (2001) has observed that for a range of $[G_{out}]$ from 10 mM to 250 mM, its corresponding internal glucose concentration $[G_{in}]$ ranges from 1.5 mM to 2.7 mM. The HXT expression curves shown in Fig. 6 sweeps $[G_{out}]$ in this range. Well keep $[G_{in}]$ fixed at 2 mM throughout the following analyses. Under these ranges of glucose levels, Eqn (2.15) states that the square root behavior results in the Hxts with lower $K_{m,optimal}$ values being more sensitive to changes in $[G_{out}]$ than those Hxts with higher $K_{m,optimal}$ values. For

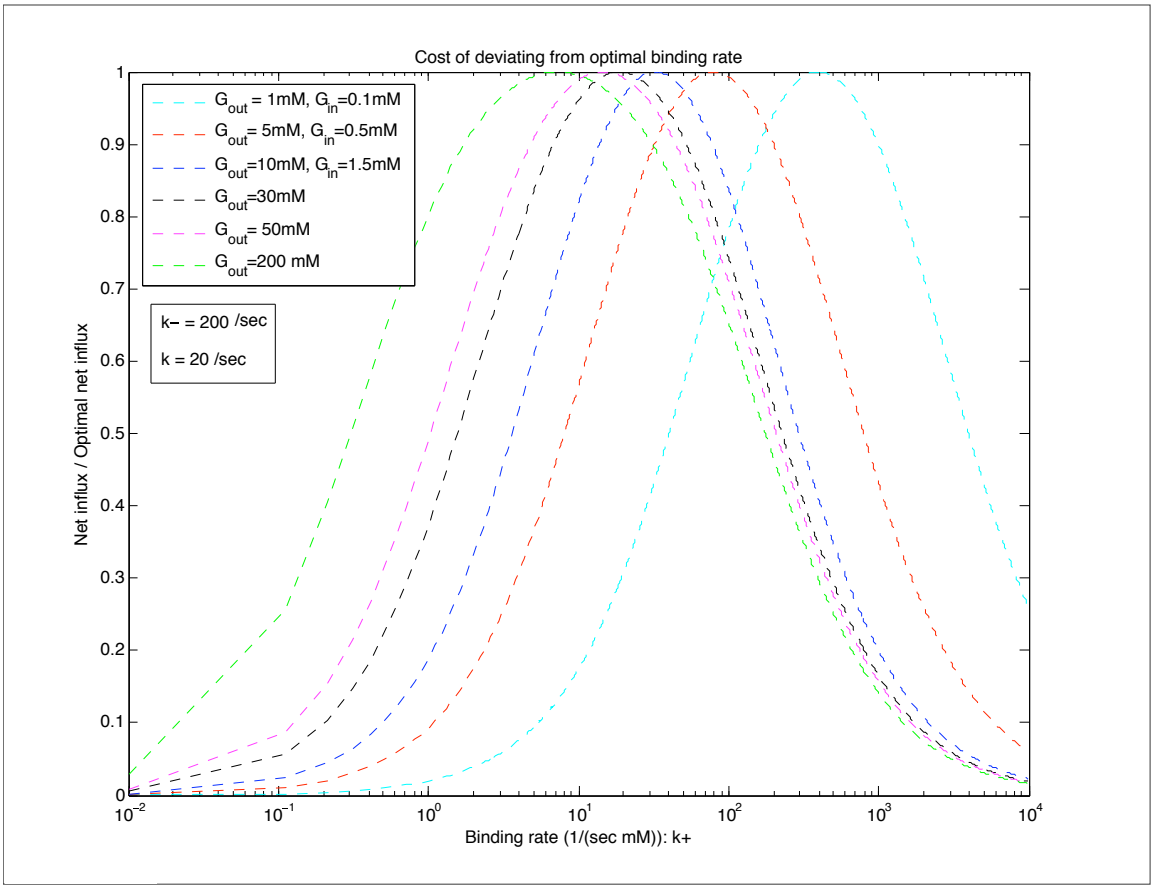


FIG. 7: **Fractional cost = net influx / maximum influx.** For each curve, the values of $[G_{out}]$ and $[G_{in}]$ are fixed and the symmetric binding rate k^+ is varied. Unless otherwise specified, $[G_{in}]$ is 1.5 mM. Note the broadening of width as $[G_{out}]$ / $[G_{in}]$ increases (i.e., for Hxt, this is equivalent to just increasing $[G_{out}]$ since $[G_{in}]$ remains constant around 2 mM). The values of $[G_{out}]$ shown here roughly correspond to the actual extracellular glucose concentrations the budding yeast is subjected to in Fig. 6.

instance, a few mM of $[G_{out}]$ fluctuation away from, say $[G_{out}] = 10$ mM results in $K_{m,optimal}$ being varied a lot more than when the same few mM fluctuation is around $[G_{out}]$ =100 mM. Figure 7 shows this argument graphically. It shows that the fractional cost in net influx when deviating from the $K_{m,optimal}$ at a given $[G_{out}]$.

In Figure 7, the cost of deviating from $K_{m,optimal}$ is shown for various representative values of $[G_{out}]$. For each Gout, the peak corresponds to the location of $K_{m,optimal}$. Notice

the overlap between $[G_{out}]$ = 1mM and $[G_{out}]$ =5 mM curves. If there was just a single transporter, say optimized for $[G_{out}]$=1mM but not for $[G_{out}]$=5 mM, then in an environment with $[G_{out}]$=5 mM, such a transporter would be able to uptake only about 60% of what it is capable of transporting had there been a transporter optimized for $[G_{out}]$=5 mM as well. Conversely, a similar remark can be made if the cell had a transporter optimized for $[G_{out}]$=5 mM but not for $[G_{out}]$ = 1 mM

The square root dependence of $K_{m,optimal}$ on $[G_{out}]$ means that the peak-to-peak distance is closer for smaller $[G_{out}]$ than for larger $[G_{out}]$. Also, as $[G_{out}]$ increases, due to the relatively stationary $[G_{in}]$, the width of peaks broaden. This means that theres more overlap and can thus explain why there are more Hxts with smaller K_m than larger K_m (Observation #4). This also gives a plausible argument for the Hxts with lower K_m being more rapidly suppressed than those with higher K_m as $[G_{out}]$ increases.

While we do not know how symmetric the Hxts are, above argument at least gives an insight into how to distribute symmetric transporters and some general principles behind how much to repress the transporter in a range of external substrate concentrations.

B. Understanding design principles underlying passive transporters

Our approach to studying passive carriers is based on asking the following question: "Given a passive transporter whose rate constants are measured, can we make sense of why the rate constants have the magnitudes that they do, and not some other values?". For instance, if the glucose binding affinity of human erythrocyte is larger outside the cell than in the cytosol, why must it be so?

From a designer's perspective, one can ask what functions a given passive transporter should be able to perform, and how to optimize it. The function is clear: shuttle biomolecules of interest into and out of the cell. But here's a dilemma: should a transporter optimize its efflux or influx ability? Is there a cost associated with optimizing one of the two? That is, to optimize its efflux, should its influx-ability be sacrificed to some extent?

Simultaneous increase in the efflux and influx of molecule is possible

Is it good to have a "balanced" transporter that is equipped with both respectable efflux and influx abilities? To make this concrete, let's consider the following: Any symmetric transporter has same net influx as net efflux in the following sense; If there are 100 mM of substrate X outside the cell and 1 mM of X in the cytosol, the net influx is the same as the net efflux that would occur if there were 100 mM of X in the cytosol and 1 mM outside the cell. But now, let's tweak the parameters of the transporter (binding rates, dissociation rates, and trans-membrane flipping rates). Is there a way to tweak the values in such a way that we increase both the net influx and net efflux rates of the transporter? To do so, we first propose to measure the "relative fitness" of a transporter as follows:

$$\text{Relative fitness} = (R_{influx,asym} - R_{influx,sym}) + (R_{efflux,asym} - R_{efflux,sym}) \tag{2.16}$$

"Relative fitness" is a measure of the overall performance of an asymmetric transporter relative to a symmetric transporter. To be more specific, given two concentrations $[G_{max}]$ and $[G_{min}]$ (where $[G_{max}] > [G_{min}]$) ("G" doesnt mean glucose. It means any substrate of interest to the transporter), and values of k^+ , k^- , and k (binding rate, dissociation rate, and flipping rate respectively), we take an asymmetric transporter with corresponding rates $k^+ + \delta^+$, $k^- + \delta^-$, $k + \delta$. We then compare the two using the metric (2.16). An example is shown in Figure 8:

With this metric (2.16), we can now figure out if increasing both influx and efflux is possible, starting out with a symmetric transporter.

It's interesting that to increase both the net influx and efflux rates, the rate constants should change in the same direction in both cases as shown in Figure 9. That is, if we had observed that increasing the net influx requires shifting the outside binding affinity to $k^+ + \delta^+$, then one may expect that increasing net efflux would require shifting the outside binding affinity to $k^+ - \delta^+$. But this is observed not to be the case. We have thus showed a rather non-intuitive result that increasing both the influx and efflux rates simultaneously is possible in some situations.

There are known passive carriers that need to effectively flush out substrates at times but also intake substrates at other times. For example, during periods of starvation, our liver cells need switch to efflux mode and synthesize glucose from amino acids, fatty acids and other small molecules, which are then expelled out to the blood stream. Our result is

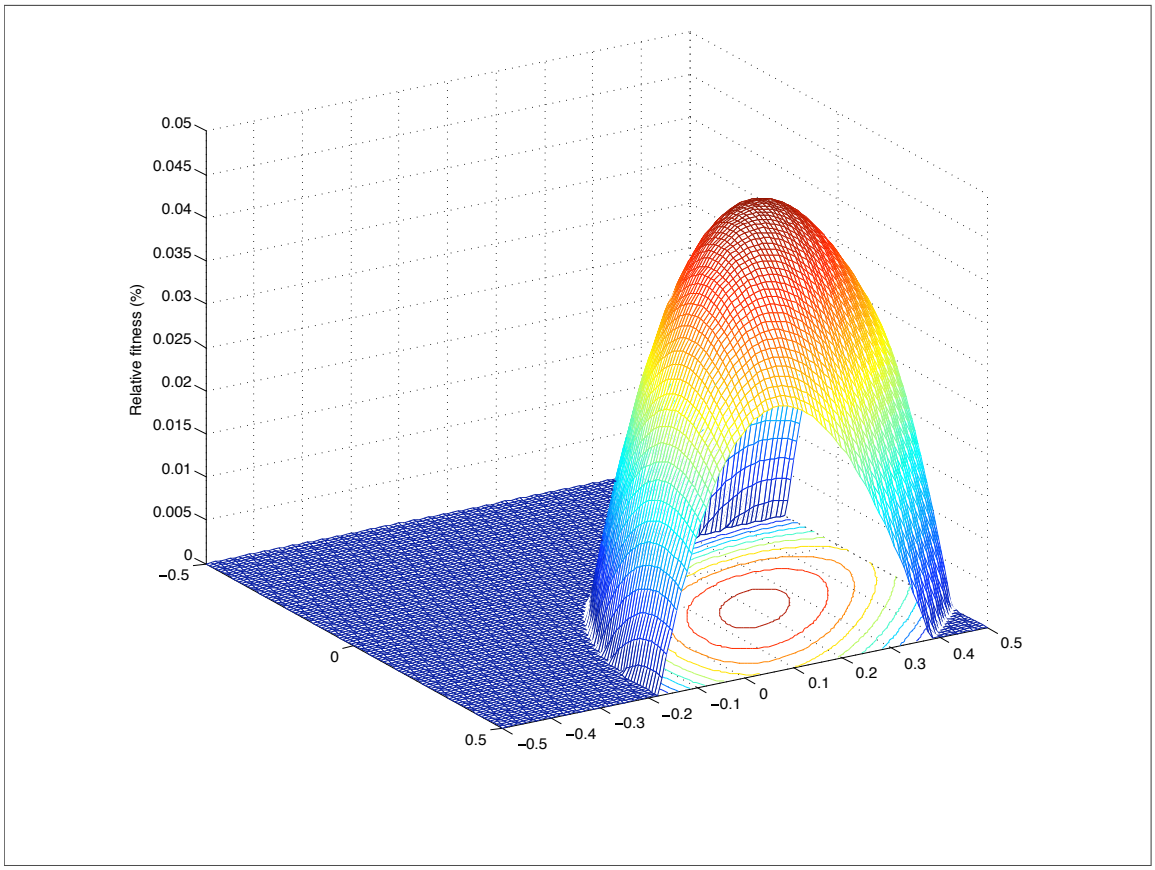


FIG. 8: **Relative fitness (Eqn (2.16) normalized by R_{sym}.** Only positive fitness is plotted by suppressing negative fitness (represented by flat region). The bulge represents asymmetric transporters that are more 'fit' than the symmetric transporter.

relevant for such transporters.

2.5. Many-parameter model of glucose metabolism in yeast

An example of an exhaustive, many-parameter model for glucose metabolism in yeast is known as the Metabolic Flux Analysis (MFA). Below, we outline some basic features of MFA and a related modeling framework known as the Metabolic Control Analysis (MCA).

Given a complex biological network, one would like to know how changing one node (enzyme, substrate, product, etc) in the network affects the other nodes that are either directly or indirectly connected to it. In particular, since during growth, cell's metabolic networks maintain near steady-state levels of fluxes, we can ask

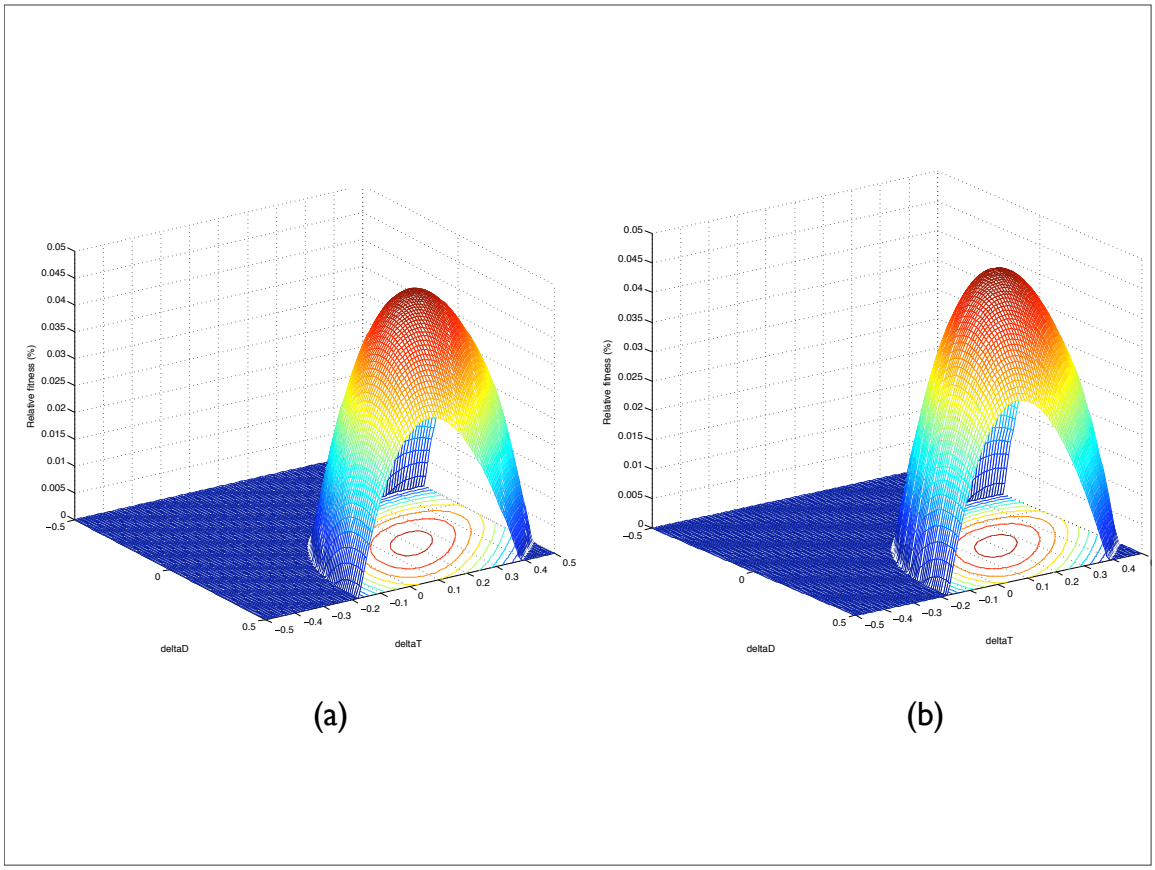


FIG. 9: (a) Relative net influx ability; (b) Relative net efflux ability. The appearance of peaks in the same region in (a) and (b) indicates increase in *both* influx and efflux abilities of the asymmetric transporter compared to the symmetric transporter.

• What is the steady-state distribution of fluxes in the metabolic network? Is it unique? If not, why does one set of distribution get expressed but not the other (**silent phenotypes**)?
• How does perturbation of a particular enzyme or metabolite from its steady-state value affect other parts of the network?

The first question is answered by Metabolic Flux Analysis (MFA) while the latter is answered by Metabolic Control Analysis (MCA).

I. METABOLIC CONTROL ANALYSIS (MCA)

A good way to understand MCA is by working through a simple example. Below, we study the simple metabolic chain and introduce the terminology used in MCA. We will highlight the MCA terms in red.

A. Toy model: Linear metabolic chain

A linear metabolic chain is defined as:

$$S \xrightarrow{V_1} X_1 \xrightarrow{V_2} X_2 \xrightarrow{V_3} X_3 \xrightarrow{V_4} X_4 \xrightarrow{V_5} P \tag{2.17}$$

X_i is a **metabolite**, S is an external substrate, and P is the final product of the metabolism. V_i is the amount of metabolite X_j generated per unit time, and is called the **flux**. MCA is a framework for quantifying how much change results in one part of network (say X_4) when you perturb another part of the network (say X_1). In MCA, these perturbations are always measured from the steady-state (in most cases, quasi-steady state) of network. These perturbations are usually limited to first order. As such, let's imagine that there is a source that maintains constant pool of S and a sink that maintains a constant pool of P in Eqn (2.17). Then everything in (2.17) is in steady-state with values (S_0, $X_{i,0}$). Now, we imagine perturbing the X_i's and S about their steady-state value. Taylor exapnding the flux balance equation for X_1 about steady-state values (S_0, $X_{1,0}$) due to the perturbation to first order, we obtain

$$\frac{d(X_1 - X_0)}{dt} = \frac{\partial V_1}{\partial S}\Big|_{S_0}(S - S_0) - \frac{\partial V_2}{\partial X_1}\Big|_{X_{1,0}}(X_1 - X_{1,0}), \tag{2.18}$$

and getting rid of units of concentration by normalizing leads to

$$\frac{d(X_1 - X_{1,0})/X_{1,0}}{dt} = \frac{V_{1,0}}{X_{1,0}}\frac{S_0}{V_{1,0}}\frac{\partial V_1}{\partial S}\Big|_{S_0}\frac{S - S_0}{S_0} - \frac{V_{2,0}}{X_{1,0}}\frac{X_{1,0}}{V_{2,0}}\frac{\partial V_2}{\partial X_1}\Big|_{X_{1,0}}\frac{X_1 - X_{1,0}}{X_{1,0}}. \tag{2.19}$$

The terms highlighted with red in above equation are called **elasticities** and are defined and denoted by

$$\varepsilon_S^{V_1} \equiv \varepsilon_S^1 \equiv \frac{S_0}{V_{1,0}}\frac{\partial V_1}{\partial S}\Big|_{S_0} = \frac{\partial ln(V_1)}{\partial ln(S)}\Big|_{S_0}$$
$$\varepsilon_{X_1}^{V_2} \equiv \varepsilon_1^2 \equiv \frac{X_{1,0}}{V_{2,0}}\frac{\partial V_2}{\partial X_1}\Big|_{X_{1,0}} = \frac{\partial ln(V_2)}{\partial ln(X_1)}\Big|_{X_{1,0}}$$

We have just introduced one of three main coefficients in MCA:

• Elasticity coefficient $\varepsilon_{X_{i-1}}^{V_i}$: Gives a *local* information about how much a flux V_i changes due to perturbation of metabolite X_{i-1} about its steady-state.

The main idea behind MCA is that using a set of these local elasticity coefficients, we want to *connect* these various local coefficients (via connection theorems) to make a global statement about the network and about rates and metabolites that are distantly connected to the metabolite X_{i-1} that's being perturbed.

Getting back to the example, using the following Taylor expansion to first order:

$$ln(X_i/X_{i,0}) = ln(X_{i,0}/X_{i,0}) + \frac{1}{X_{i,0}}(X_i - X_{i,0}) + \mathcal{O}((X_i - X_{i,0})^2), \tag{2.20}$$

equation (2.19) is rewritten as

$$\frac{dy_1}{dt} = \frac{V_{1,0}}{X_{1,0}}\varepsilon_S^1 y_0 - \frac{V_{2,0}}{X_{1,0}}\varepsilon_1^2 y_1, \tag{2.21}$$

where we defined

$$y_1 \equiv ln(X_1/X_{1,0})$$
$$y_0 \equiv ln(S/S_0).$$

As for the remaining metabolites, we have a similar expression for flux-balance:

$$\frac{dy_i}{dt} = \frac{V_{i,0}}{X_{i,0}}\varepsilon_{i-1}^i y_{i-1} - \frac{V_{i+1,0}}{X_{i,0}}\varepsilon_i^{i+1} y_i \tag{2.22}$$

with $y_i \equiv ln(X_i/X_{i,0})$.

At this point, we ask the following question: "How do the values of X_i change if we perturb the external substrate S flowing into the system?". We can obtain an approximate answer to this question by setting $dy_i/dt = 0$ in Eqn. (2.22) and solving for y_i for each i.

This is an approximation since we've ignored the higher order terms. Recursively solving for y_i, y_{i-1}, ... , y_1, we get

$$y_i = \frac{\varepsilon_S^1}{\varepsilon_i^{i+1}} y_0 \tag{2.23}$$

Above introduces us to our second coefficient of MCA, the concentration control coefficient (also called response coefficient) C_S^i defined as

$$C_S^i \equiv \frac{dln(X_i/X_{i,0})}{dln(S/S_0)}, \tag{2.24}$$

which in our linear metabolic chain is just $C_S^i = \frac{\varepsilon_S^1}{\varepsilon_i^{i+1}}$.

• concentration control coefficient C_S^i is a *non-local* dimensionless number that characterizes fractional change in metabolite X_i due to fractional change in the external substrate S about its equilibrium. This answers the question posed above.

One final coefficient of interest in MCA is a similarly defined control coefficient called flux control coefficient with respect to a metabolic parameter p, denoted $C_p^{V_i}$:

$$C_p^{V_i} \equiv \frac{dln(V_i)}{dln(p)} = \sum_{j=1}^{n} \varepsilon_j^i C_p^j + \varepsilon_p^i, \tag{2.25}$$

which, in this linear metabolic model is simply

$$C_S^{V_i} = \varepsilon_S^1. \tag{2.26}$$

This is a very simple result in this particular toy model. Physically, all these flux control coefficients are equal (and in particular, equal to C_S^1) since in a linear model, the steady-state fluxes through each of the enzymes must be equal.

If we know the elasticities and the flux distribution in the metabolic network, we can calculate the control coefficients of any metabolic system.

II. METABOLIC FLUX ANALYSIS (MFA)

In MFA, we represent every reaction that is taking place in our system of interest using a Stoichiometric matrix S). Stochiometric matrix S has integer entries, and transforms a

flux vector v(containing the reaction rates) into a vector that contains the time derivatives of the concentrations. In other words, we want to solve

$$\frac{d\vec{x}}{dt} = S\vec{v}, \tag{2.27}$$

in steady-state. To do so, MFA uses techniques from linear algebra (in particular Single-Value-Decomposition (SVD)). You can imagine that S is typically a very large matrix (e.g. 541 x 609 is the size of S for a soil bacterium *Geobacter sulfurreducens*), thus it typically admits many steady-state solutions v_{ss} to above equation. But by gathering all we know about, say the E.coli metabolism (e.g. Number of H^+ typically present in the cell (answer: ≈ 16), ratio of tRNAs to number of ribosomes (answer: about 10)) and so forth, dedicated large-scale modeling softwares can put constraints in the solution space, excising many of the v_{ss} that are mathematically correct, but are not biologically relevant. Sometimes, one is able to excise away enough biologically irrelevant solutions, leaving us with very few possibilities. From these, we can predict a set of fluxes that leads to optimal biological function (e.g. growth rate (bio-mass production)). This is often done using linear programming.

Here, we stay modest and restrict ourselves to construction of low dimensional Stochiometric matrices.

Example. Consider a system with just the following three reactions:

$$\begin{aligned} CP &\overset{v_1}{\leftrightarrow} PC, \\ C + P &\overset{v_2}{\leftrightarrow} CP, \\ C + AP &\overset{v_3}{\leftrightarrow} CP + A, \end{aligned} \tag{7}$$

and the corresponding stochiometric matrix is thus

$$S = \begin{array}{c} \begin{matrix} v_1 & v_2 & v_3 \end{matrix} \\ \begin{pmatrix} 0 & 0 & 1 \\ 0 & -1 & -1 \\ 0 & -1 & 0 \\ -1 & 1 & 1 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix} \end{array} \begin{matrix} A \\ C \\ P \\ CP \\ AP \\ PC \end{matrix} \tag{8}$$

where the **rows** correspond to the metabolites indicated and **columns** correspond to reactions v_1, v_2, v_3 respectively. Since the signs are taken care of in the stochiometric matrix S, the flux vector v must have entries that are **non-negative**. As we'll see, this leads to convex analysis being useful for analyzing the steady-state solution space. Notice that above S indeed leads to

$$\frac{d\vec{x}}{dt} = S\vec{v}, \tag{2.28}$$

where $\vec{x} =$ [[A], [C], [P], [CP], [AP], [PC]]^T. The goal of MFA is to solve above equation for steady-state, thus obtain a set of possible steady-state fluxes $\vec{v}_{ss}$.

■

Supplementing the stoichiometric matrix is an elemental matrix **E**. This matrix gives us the composition of all the chemical elements and compounds in a given network. Again, an example clarifies this notion:

Example. The following reaction:

$$2H_2 + O_2 \rightarrow 2H_2O,$$

has the following elemental matrix:

$$E = \begin{array}{c} \begin{array}{ccc} H_2 & O_2 & H_2O \end{array} \\ \begin{pmatrix} 0 & 2 & 1 \\ 2 & 0 & 2 \end{pmatrix} \end{array} \begin{array}{c} O \\ H \end{array} \tag{9}$$

■

Since no chemical element can be created or destroyed in any chemical reaction, the following conservation law holds (matrix multiplication between E and the stoichiometric matrix of the system S):

$$ES = 0 \tag{2.29}$$

This is a trivial example but it's precisely **constraints** of this sort (and there are many others, many of which are specific to a particular system being modeled) that will limit the solution space in steady-state. And if we have enough constraints to end up with a very small set of solutions, we can even simulate or experiment with those solutions and obtain the one that optimizes a biological function of interest such as growth rate.

2.6. A need for a phenomenological model of cell growth

In the previous section, we described two of the conventional quantitative models for describing the large-scale, complex network that forms glucose (and its bi-product) metabolism. But if our question is, "How fast does a yeast cell divide given a certain amount of glucose in its envrionment?", it seems that there should be a way to answer this question without caring about each individual chemical reaction that goes on inside the cell to break down a glucose molecule. One approach may be to identify a few 'rate limiting steps' in the complex network of reactions shown in Figure 2.1. This has been difficult as many of these pathways, if suppressed, are compensated by other intact pathways to break down a glucose molecule. In the next chapter, we provide a phenomenological model of yeast growth that does not require identification of any rate limiting steps.

Bibliography for Chapter 2

[1] IUBMB-Sigma-Nicholson metabolic pathways chart: http://www.iubmb-nicholson.org/chart.html

[2] Widdas, W. F., Kinetics of glucose transfer across the human erythrocyte membrane. *J. Physiol.* **120**, 23-24 (1953).

[3] Widdas, W. F., Facilitated transfer of hexoses across the human erythrocyte membrane. *J. Physiol.* **125**, 163-180 (1954).

[4] Maier, A., Volker, B., Boles, E., Fhrmann, G. F., Characterisation of glucose transport in *Saccharomyces cerevisiae* with plasma membrane vesicles (countertransport) and intact cells (initial uptake) with single Hxt1, Hxt2, Hxt3, hxt4, hxt6, Hxt7 or Gal2 transporters. *FEMS Yeast Research*, **2**, 539-550 (2002).

[5] Reijenga, K. A. *et. al.*, Control of glycolytic dynamics by hexose transport in *Saccharomyces cerevisiae*, *Biophysical Journal* **80**, 626-634 (2001).

[6] Fell, D. A. *Understanding the control of metabolism* (Portland, 1997).

[7] Savageau, M. A. *Biochemical systems analysis: A study of function and design in molecular biology* (Addison-Wesely, 1976).

[8] Palsson, B. O. *Systems biology: Properties of reconstructed networks* (Cambridge University, 2006).

3. Phenomenological model of cell growth

This work appeared in

Hyun Youk and Alexander van Oudenaarden, *Nature,* 462: 875-79 (2009).

Chapter summary

A central challenge in systems biology is describing how so many intracellular components work together to execute a biological function using a quantitative model that involves only a few essential parameters. By judiciously stripping away the complexity, such a model can highlight the key principles and provide deep insights that are either difficult to extract from or completely missed by large-scale (e.g. whole-genome) models. Providing a quantitative picture of microbial growth on nutrients, such as the growth of the budding yeast on the key carbohydrate glucose, is one of the oldest problems currently facing such a challenge. The work in this chapter tackles that challenge through quantitative growth experiments.

This chapter begins with simple growth experiments that show surprising features of budding yeast growth that go against predictions of many conventional growth models as well as some of our own intuitions. For instance, despite an increase in both the extracellular glucose level and cells' glucose import rate, cells do not just grow faster – they can grow slower and even approach a growth arrest. We find that this is because the cell's growth rate is determined by the convolution of two key mechanisms: glucose perception and import. Conventional growth models ignore the interplay between these two mechanisms. Experiments in this chapter lead us to an intuitive and quantitative model that demonstrates glucose perception and import are two separate modular components of cell growth that we can experimentally tune independently of each other. Because these two mechanisms trigger all the down-stream intracellular activities that are responsible for growth, our empirical model in fact incorporates the amalgamated effect of those vast intracellular activities into just two key terms that are easily measurable.

The work in this chapter shows that the approach commonly used in physics to extract the essence of a complex system can also be applied in tackling one of the oldest problems in biology.

3.1. Introduction

In 1942 Jacques Monod introduced his microbial growth model[1] that prompted quantitative studies of microbial metabolism[2-13]. This motivated a wealth of mathematical models describing the growth of budding yeast *Saccharomyces cerevisiae* on the key carbohydrate glucose[14]. These models mainly focus on the effect of glucose import on the growth rate. But in addition to importing glucose, yeast senses extracellular glucose through several glucose sensors. These two key events at the cell membrane – glucose sensing and import – then trigger many downstream intracellular molecular events (e.g. transcription, metabolic processes, post-transcriptional modifications) that collectively determine the growth rate[15]. Many conventional models overlook this collective effect by ignoring glucose sensing. Growth behaviors that are qualitatively very different from current models' descriptions may arise if glucose sensing and import are properly taken into account. One approach to addressing this deficiency is constructing detailed many-parameter models that attempt to explicitly track each of the vast molecular events involved in the yeast's glucose metabolism[8,9]. Such approach has provided detailed information on the flux of thousands of known metabolic reactions and new insights into yeast's growth on glucose. However, such an approach also conflates the effects of glucose sensing and import because it is not yet known how each of the vast molecular events are altered when glucose import rate is varied independently of the level of extracellular glucose sensed by the cell. The enormous number of metabolites and reactions involved makes experimentally determining each molecular change due to glucose sensing and import challenging. Indeed, a persistent challenge in obtaining a quantitative understanding of microbial growth on nutrients has been identifying just the few parameters that are necessary for extracting the central features from this complex cellular process. A phenomenological model that retains just those essential parameters may provide new insights and central design principles[16,17] underlying microbial growth. Motivated by these considerations, we sought to decouple and measure the separate effects of glucose sensing and import on cell growth, then provide a concise phenomenological model that elucidates how the interaction between the two determines the growth rate.

3.2. Dependence of growth rate on glucose concentration

To measure and separate out the effects of glucose perception and import on growth rate, we first decouple any control that glucose sensing has on glucose import. Such coupling primarily comes from the two glucose sensors (Snf3 & Rgt2)[18] that drive the transcriptional regulation of the six primary hexose transporters (Hxt1-4, 6 & 7)[19-23] which are responsible for glucose import (Figure 3.1). Our background strain lacks all the major and minor glucose transporter genes (*hxt1-17Δ, agt1Δ, stl1Δ, gal2Δ*)[24], thus no sensors affect the transcription level of any transporter genes including the *HXT*s. We made five "single-*HXT*" strains by introducing into the background strain only one of the five primary *HXT* genes (excluding *HXT7*) under the control of the inducible promoter P_{TET07}. Each of these strains contains just one type of *HXT* gene, and its expression level could be controlled by the inducer doxycycline independently of extracellular glucose (Figure 3.2). We measured the log-phase growth rate of the single-*HXT* strains in minimal media containing a range of different concentrations of doxycycline and glucose, whose concentrations were held constant during batch growth for each experiment. We found surprising behaviors in every single-*HXT* strain's growth rate (Figures 3.3 and 3.4). Since glucose no longer regulates the transcription of the sole *HXT* gene in our strains in a complicated manner, one would expect that an increase in extracellular glucose concentration would lead to a simple increase in the single-*HXT* strain's glucose uptake rate (when the doxycycline concentration is held constant). A typical conventional model[14] predicts that the growth rate should thus simply rise as the glucose level increases. Yet, depending on the initial glucose level, a further increase in the glucose level either increases or decreases the "Hxt1-only" strain's growth rate (Figure 3.3). This is also true for the growth rates of the "Hxt2-only" and "Hxt4-only" strains. Furthermore, despite growing as well as other strains at low glucose levels, the "*HXT3*-only" and "*HXT6*-only" strains even approach growth arrest for glucose level higher than 0.02% (Figure 3.3). Thus, we observed no systematic relationship between glucose level and growth rate. It is noteworthy that the wild-type strain, unlike these single-*HXT*s, simply grows faster when more glucose is present (Figure 3.3), a behavior we will consider more closely later.

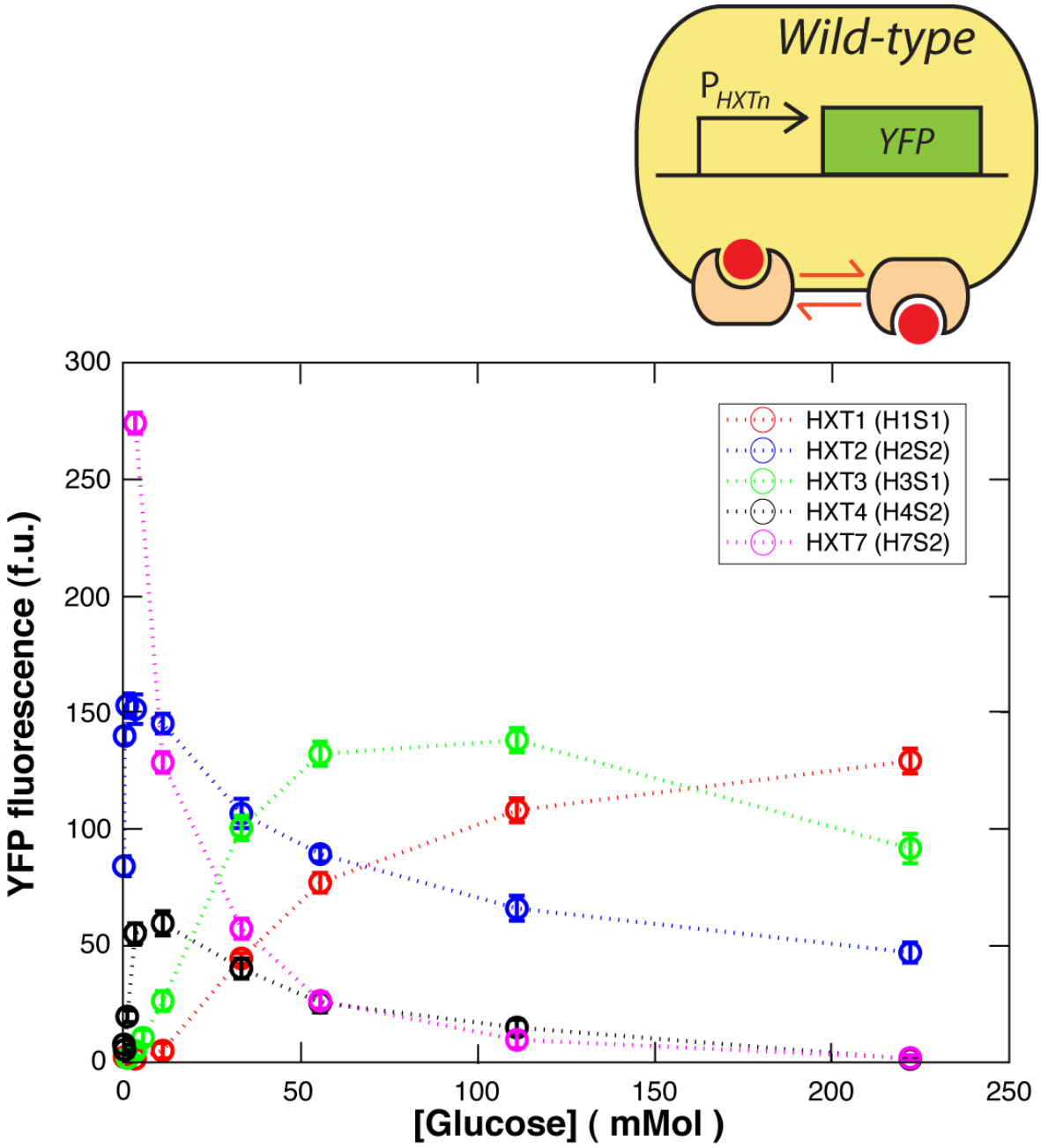


Figure 3.1. : Expression levels of *HXT* genes in wild-type strain (CEN.PK2-1C) measured using YFP reporters (used "Wild-type P_{HXTn}:YFP" strains; see "strain list"). As the extracellular glucose concentration varies, the expression level of each *HXT* gene in the wild-type strain changes. Two of the glucose sensors, Snf3 and Rgt2, initiate the signal transduction that results in these expression patterns. This result was originally reported in a previous work[3]. We have reproduced it here for the sake of completeness. Error bars, s.e.m.; *n*=3.

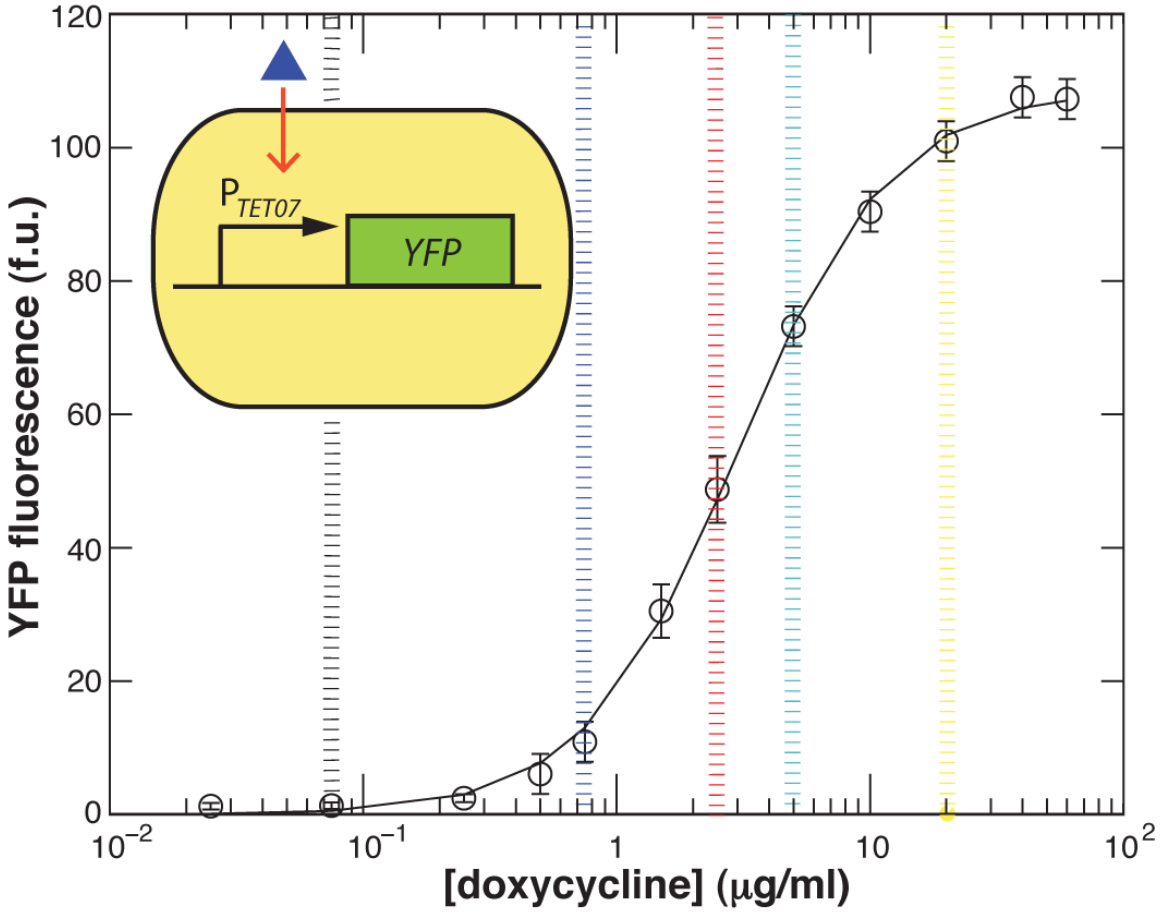


Figure 3.2. : Inducibility of P_{TET07} inferred from YFP fluorescence as a function of doxycycline concentration in the HY4DCal5 strain. HY4DCal5 and the single-*HXT* strains were all constructed using the same background strain HY4D1. Hence we can infer the inducibility of the promoter P_{TET07} in the single-*HXT* strains from this induction curve measured in the HY4DCal5 strain. The fluorescence data shown here were obtained while the cells were in log-phase growth in a standard synthetic media with 2% maltose and the indicated doxycycline concentration. The colored vertical dashed lines indicate the concentrations of doxycyline used in subsequent figures for characterizing the single-*HXT* strains. Error bars, s.e.m.; *n*=3.

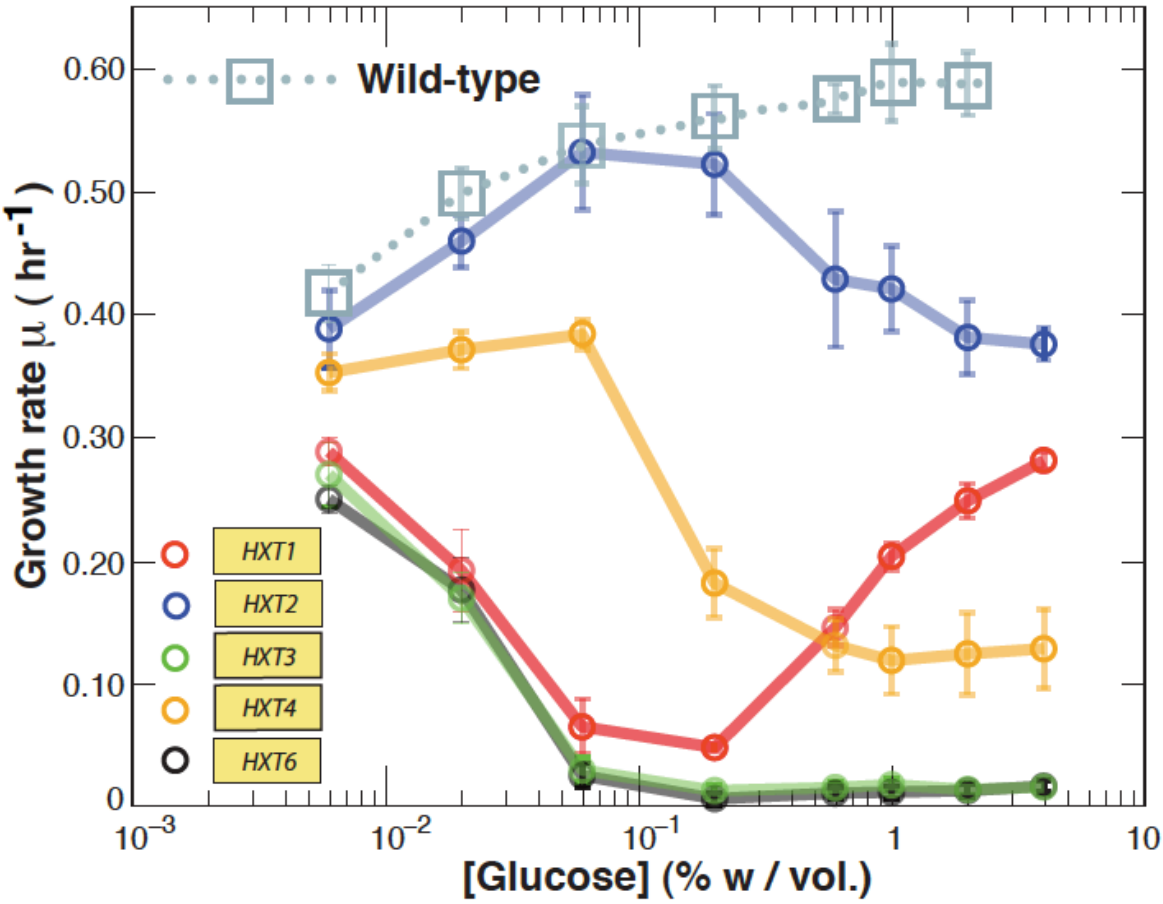


Figure 3.3. : Growth rates of "single-*HXT*" strains do not show any systematic trend with respect to glucose concentration. Log-phase growth rates of the wild-type strain (CEN.PK2-1C) and five single-*HXT* strains at varying [glucose] but constant [doxycycline] (0 μg/ml for wild-type and 2.5 μg/ml for single-*HXT* strains) are shown. The shape of each single-*HXT* strain's growth-rate curve is maintained over a wide range of doxycycline concentrations (Figure 3.4). The growth-rate curves of the "single-*HXT*" strains display stark differences from the wild-type's curve: single-*HXT* strains' growth rates can substantially decrease, and some strains even approach growth arrest, despite a monotonic increase in [glucose]. Error bars, s.e.m.; *n*=3.

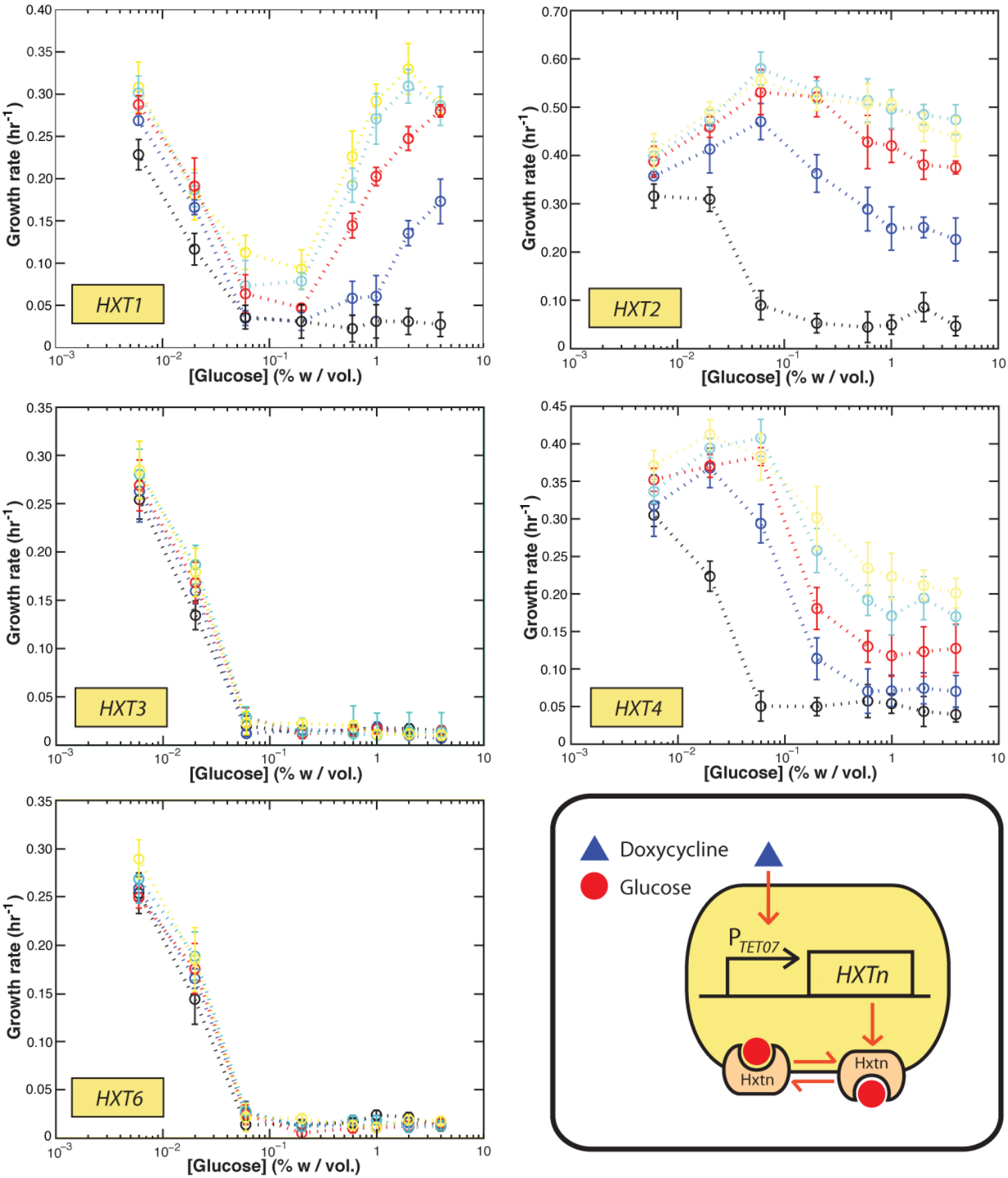


Figure 3.4. : (See next page for figure caption).

Figure 3.4. : Growth rate of single-*HXT* strains in various combinations of glucose and doxycycline concentrations. We measured the log-phase growth rate of the single-*HXT* strains in synthetic growth media containing doxycyline and glucose. These concentrations remained constant during each batch growth experiment. Each color corresponds to a particular value of doxycycline concentration as indicated in Figure 3.2. A curve of a given color shows how the growth rate changes as a function of the glucose concentration (at fixed doxycycline concentration). None of these strains' growth rates increase monotonically with an increase in the glucose level, unlike the parental wild-type strain (Figure 3.3). Depending on the initial glucose level, a further increase in the glucose level either increases or decreases the single-*HXT* strain's growth rate. Error bars, s.e.m.; *n*=3.

3.3. Dependence of growth rate on glucose uptake rate

Using our doxycycline inducible expression system, we were able to show that for every single-*HXT* strain at fixed doxycycline level, the glucose uptake rate increased as the glucose level increased (Figure 3.5). To measure glucose uptake rates, we fused yEGFP to the inducible *HXT* gene in each of the single-*HXT* strains (Figure 3.6). Measuring the average single-cell fluorescence in these strains gave us the relative number of Hxt proteins synthesized in these cells (Figure 3.7). Using the known Michaelis-Menten parameters of the Hxts[25,26], we calculated the cell's total glucose uptake rate. We also directly measured the cell's glucose uptake rate. The directly measured and calculated uptake rates were in good agreement (Figure 3.8): glucose uptake rate increased as the glucose concentration increased (at constant doxycycline concentration) (Figures 3.5 and 3.9). Hence despite a monotonic increase in both glucose uptake rate and extracellular glucose level, single-*HXT* strains at fixed doxycycline concentration can grow significantly faster, or slower, or even approach growth arrest as seen earlier (Figure 3.3), effects that no conventional growth model can either quantitatively or qualitatively describe.

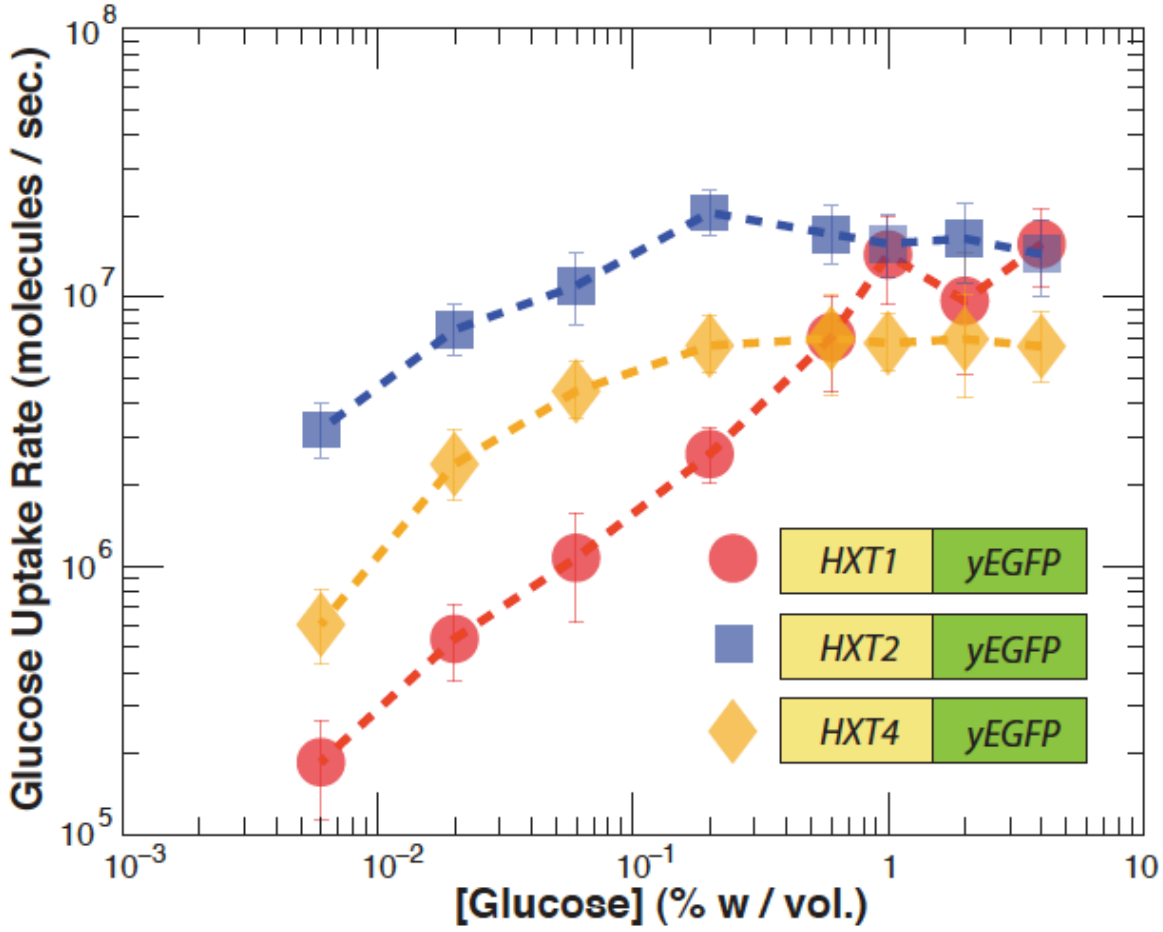


Figure 3.5. : A rise in [glucose] yields an increase in the uptake rate, but cells do not necessarily grow faster. To both measure and calculate glucose uptake rates, yEGFP was fused to the *HXT* gene in each "single-*HXT*" strain. These fluorescent single-*HXT* strains have the same growth-rate features as their non-fluorescent counterparts shown in Fig. 3.3 (Figure 3.6). The measured glucose uptake rates per cell for just three of these fluorescent single-*HXT* strains at [doxycycline] = 2.5 μg/ml are shown here. These fluorescent single-*HXT* strains' glucose uptake rates monotonically increase as [glucose] increases, despite the non-systematic behavior of their growth rates reflected in Fig. 3.3. Hence, a cell can grow faster, or slower, or approach growth arrest despite an increase in both its glucose uptake rate and [glucose]. Error bars, s.e.m.; *n*=3.

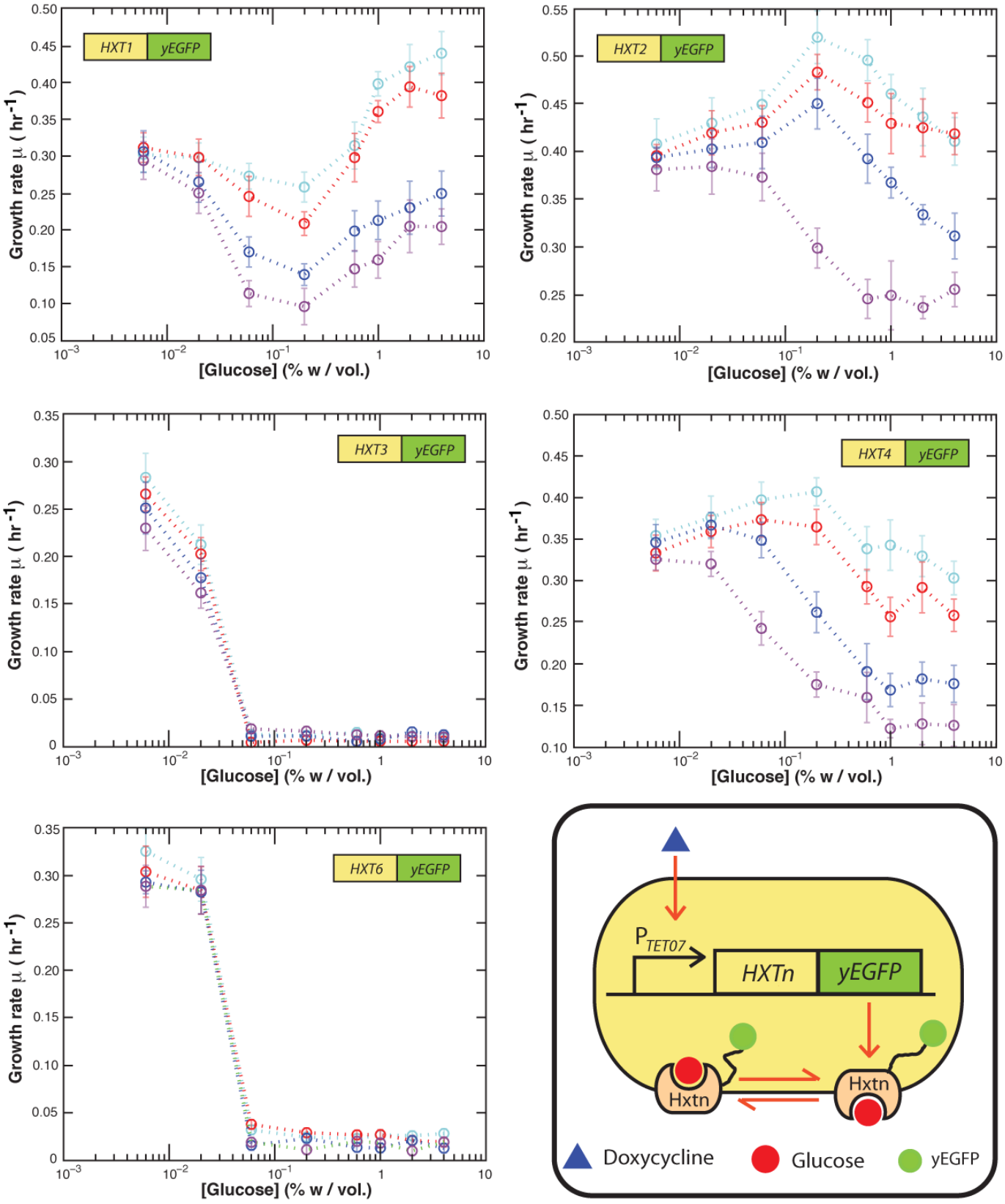


Figure 3.6 : (See next page for figure caption).

Figure 3.6. : Growth rate of fluorescent single-*HXT* strains (yEGFP fused to a *HXT* gene in each single-*HXT* strain) in various combinations of glucose and doxycycline concentrations. Each color corresponds to a particular value of doxycycline concentration as indicated in Figure 3.2 (purple represents [doxycycline] = 0.25 μg/ml). Fusing yEGFP to a *HXT* gene results in functional fluorescent single-*HXT* strains whose growth rates show the same key features that their non-fluorescent counterparts exhibit (Fig. 3.4). These fluorescent strains thus exhibit the same apparent non-systematic behavior in their growth rates as their non-fluorescent counterparts. Error bars, s.e.m.; *n*=3.

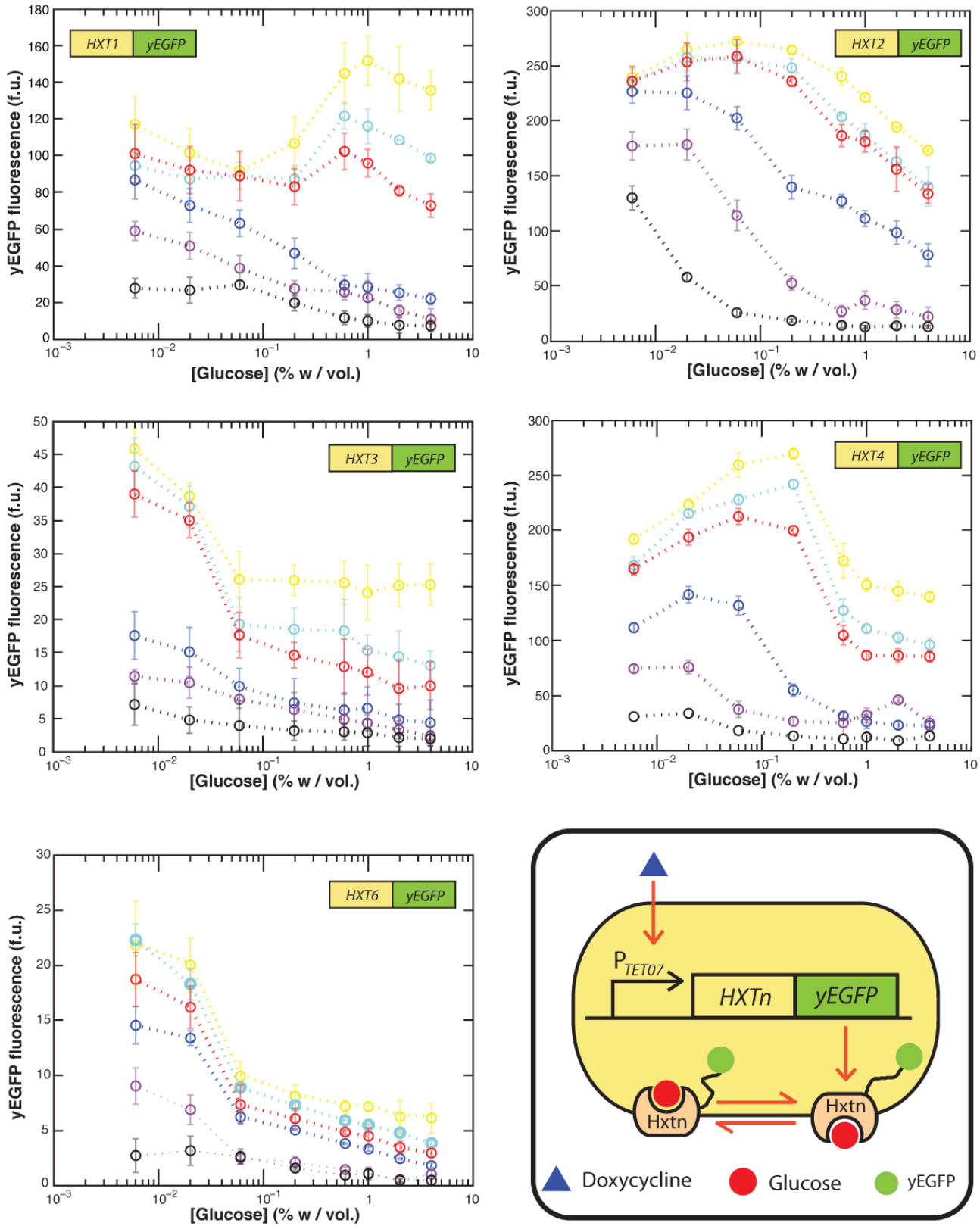


Figure 3.7 : (See next page for figure caption).

Figure 3.7. : Average, steady-state, single-cell yEGFP levels in the fluorescent single-*HXT* strains growing in various combinations of glucose and doxycycline concentrations. In the single-*HXT* strains with the yEGFP fused to the *HXT* gene, the relative number of Hxt proteins per cell was inferred from the average single-cell fluorescence measured using flow cytometer (See methods). Each color corresponds to a particular value of doxycycline concentration as indicated in Fig. 3.2 (purple represents [doxycycline] = 0.25 μg/ml). The fact that single-cell fluorescence value changes as the glucose level changes (at constant doxycycline level), indicates an existence of post-transcriptional regulations of the Hxts. Indeed, previous works have revealed some glucose dependent post-transcriptional regulations of Hxts (e.g., endocytosis of Hxt6 & Hxt7 at high glucose levels)[4] and other metabolic genes[2,5]. We took into account the effect that such regulation has on glucose uptake by directly measuring the uptake rates. The cell's glucose uptake rate is determined by the combination of two parameters – the amount of Hxt protein in the cell and the extracellular glucose concentration. By measuring the glucose uptake rates, we found that all the single-*HXT* strains' glucose uptake rates monotonically increase when the glucose concentration increases (at fixed doxycycline level; Figures 3.5 and 3.9). This is because even though the number of Hxt protein in a cell may decrease when the glucose level rises (at fixed doxycycline level), this decrease is feeble: it is more than compensated by the accompanying increase in the glucose level, resulting in net increase in glucose uptake rate. This is confirmed by both our measured and calculated glucose uptake rates (Figures 3.8 and 3.9). Error bars, s.e.m.; *n*=3.

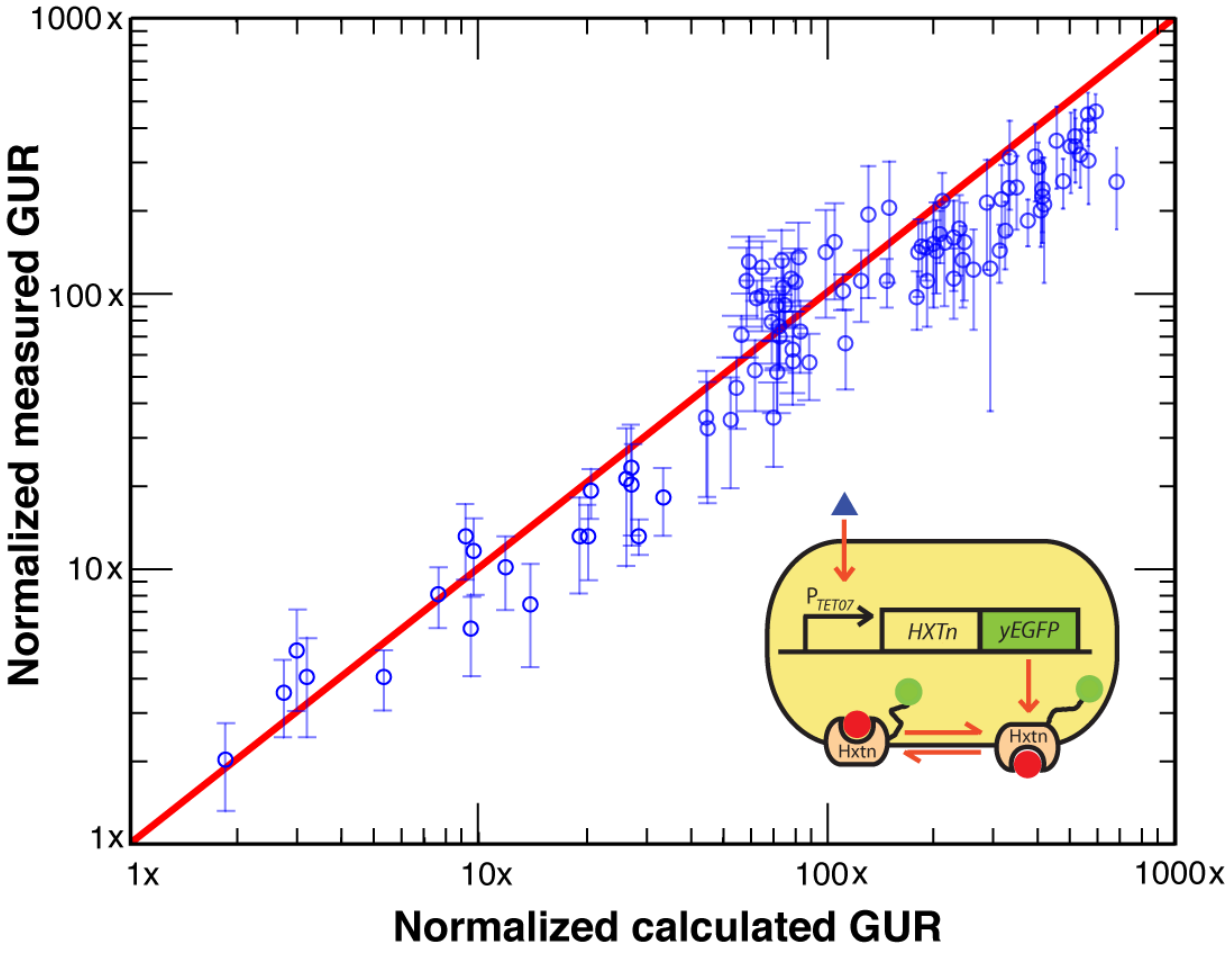


Figure 3.8. : Comparison of measured glucose uptake rate (GUR) with calculated GUR of the fluorescent single-*HXT* strains. The measured and calculated values of glucose uptake rate of all the fluorescent "single-*HXT*" strains are plotted together here. GURs are reported in normalized units to show that relative changes in both the measured and calculated GURs are in good agreement with each other. Error bars, s.e.m.; *n*=3.

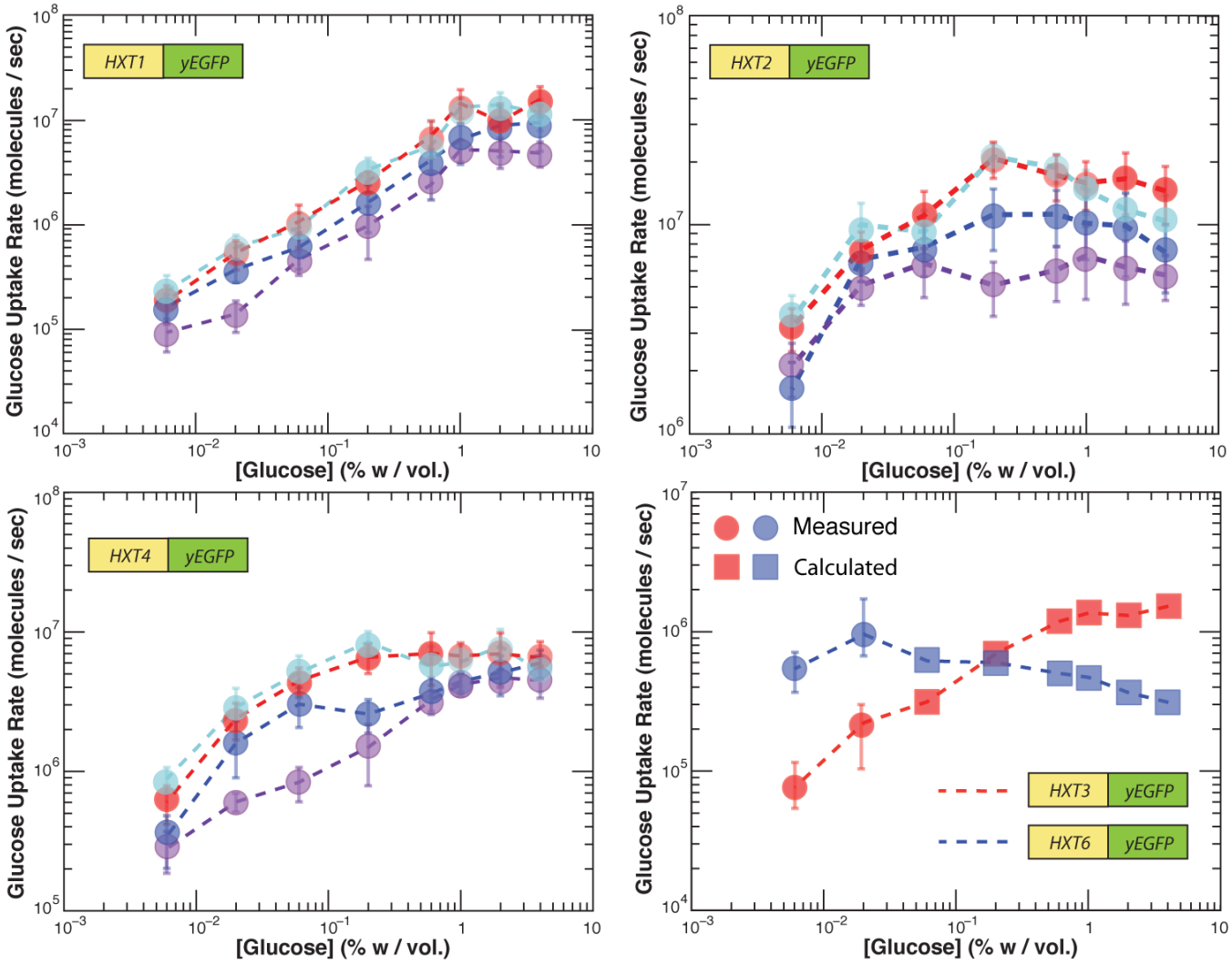


Figure 3.9. : Measured glucose uptake rates (GURs) of fluorescent single-*HXT* strains growing in various combinations of glucose and doxycycline concentrations. Each color corresponds to a particular doxycycline concentration as indicated in Figure 3.2 (purple represents [doxycycline] = 0.25 μg/ml). For "Hxt3-only" and "Hxt6-only" strains, only the GURs at [doxycycline] = 2.5 mg/ml are shown here for clarity. Since these two strains transiently approach near growth arrest for [glucose] > 0.02%, their GURs could not be measured using our method for [glucose] > 0.02%. Instead, we calculated their GURs as they transiently approached near growth arrest based on their yEGFP fluorescence (Figure 3.7). Close agreement between the measured and calculated GURs (Figure 3.8) gives us confidence in these calculated GURs. Error bars, s.e.m.; *n*=3.

3.4. Phenomenological model of growth

Plotting all five single-*HXT* strains' growth rates and uptake rates together resulted in a wide scatter of data points, where each data point is specified by two coordinates: uptake rate and growth rate (Figure 3.10). This plot reveals that uptake rate alone cannot specify the cell's growth rate. Specifying the glucose concentration by color-coding these data points (i.e., each data point now has three coordinates: (uptake rate, extracellular glucose concentration, growth rate)) causes a striking pattern to emerge (Figure 3.11). This analysis reveals that growth rate μ is determined by two independent variables: the glucose uptake rate r, and the extracellular glucose concentration g. Our full experimental data set of all five single-*HXT* strains over a wide range of glucose and doxycycline concentrations are described by a single equation

$$\mu(r,g) = P(g) \times \ln\left(\frac{r}{r_c}\right) + \mu_c \qquad [3.1]$$

where μ_c and r_c are constants specifying the point of convergence of the log-linear lines (Figure 3.11), and the function $P(g)$ describes the slope of the log-linear correlation between μ and r for each value of g. This equation does not depend on which Hxt the cell uses for glucose uptake. This slope $P(g)$ increases with increasing g, and in turn tends to decrease growth rate (when $r < r_c$). $P(g)$ quantifies the dramatic effect that the extracellular glucose has on growth rate independently of glucose import – the effect of glucose perception. Qualitatively, Eqn. [3.1] states that an increase in the extracellular glucose concentration may cause two counteracting effects: an increased glucose uptake rate r (which tends to increase growth rate), and an increased perception of extracellular glucose (which tends to decrease growth rate). The net result on growth rate (i.e., whether it rises or falls) is decided by the competition between these opposing effects of glucose perception and uptake. Which one of the two effects dominates depends on the actual values of g and r, in particular on the product $P(g)\ln(r/r_c)$ quantifying the interaction between glucose perception and import (Supplementary text).

The "growth landscape" in Figure 3.12, described by Eqn. [3.1], shows the full set of growth rates possible for a wide range of g and r. Because Eqn. [3.1] does not distinguish between the type and number of Hxts cells use for glucose import, it is applicable to cells with

any number of *HXT* genes, including the wild-type, as long as the cells achieve the uptake rate within the range we probed. The shape of this landscape allows for the unusual growth-rate behaviors observed, including the convex shaped growth rate of the "Hxt1-only" strain (orange path, Figure 3.12), the "Hxt6-only" strain's path towards growth arrest (red path, Figure 3.12) and the wild-type's hyperbolic growth rate (blue path, Figure 3.12). The wild-type strain is near the peak of this growth landscape yet its uptake rate is not much higher than those achieved by some single-*HXT* strains. The growth landscape shows that some values of (g, r) cannot sustain growth ($\mu = 0$). Indeed, for every g, there is a minimum uptake rate a cell needs to have in order for it to have any chance of growing in that particular glucose environment (Figure 3.13).

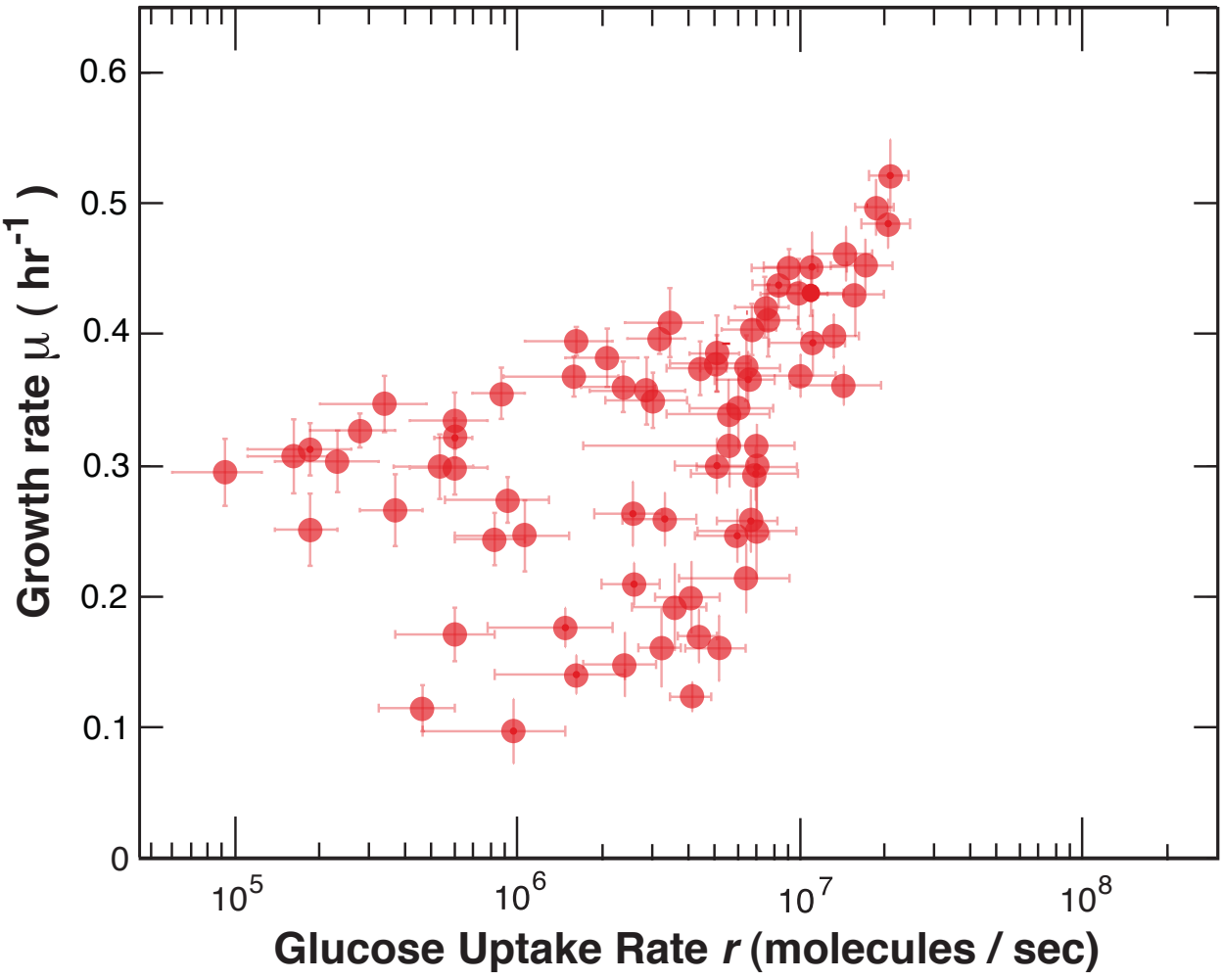


Figure 3.10. : Result of plotting the growth rate and the measured glucose uptake rate (GUR) of all the fluorescent single-*HXT* strains together. This shows that for a particular value of glucose uptake rate, multiple values of growth rates are possible. This means that glucose uptake rate alone cannot specify the cell's growth rate. But additionally coloring these data points according to the value of the extracellular glucose concentration leads to a striking pattern observed in Figure 3.11.

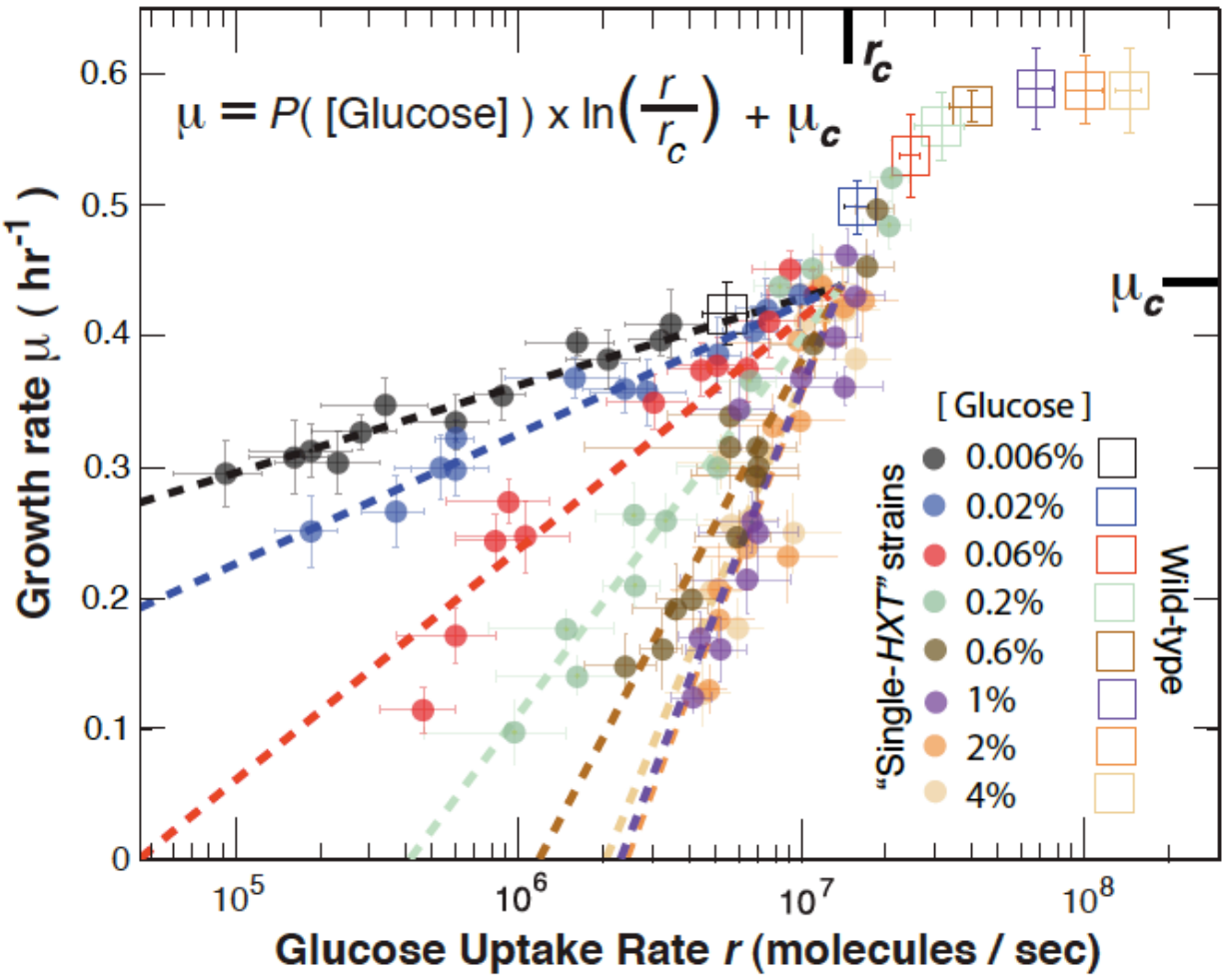


Figure 3.11.: Emergence of a concise growth model incorporating cell's perception and uptake rate of glucose. Plotting together all the measured growth rates and glucose uptake rates of the fluorescent single-*HXT* strains (Figure 3.10) then color-coding by extracellular glucose level reveals this striking pattern. This plot shows that extracellular glucose concentration *g* and glucose uptake rate *r* are two independent variables. Growth rate is concisely described by the fit function $\mu(r, g)$. *P(g)* is the slope of the log-linear correlation between growth rate and uptake rate for each *g*; we obtain *P(g)* by fitting. μ_c and r_c are constants specifying the point of convergence of the log-linear lines (μ_c = 0.44 hr^{-1} ,r_c = 1.4x10^7 molecules/s). Error bars, s.e.m.; n=3.

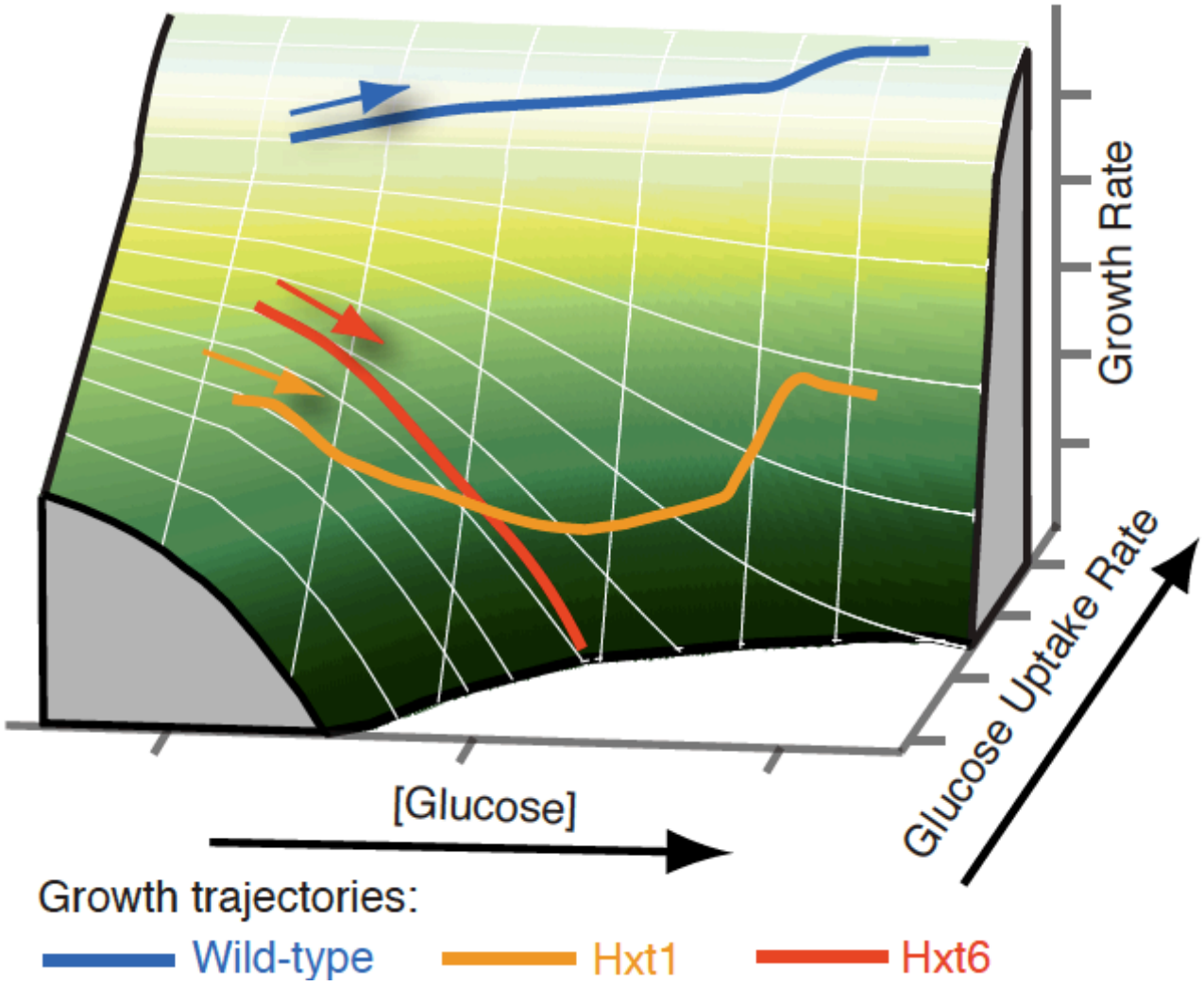


Figure 3.12. : The Growth landscape formed by perception and import of glucose in yeast. Full "growth landscape" of budding yeast: Three dimensional plot of the function $\mu(r, g)$. The "growth trajectories" followed by the parental wild-type (blue path, near the peak of this landscape), and fluorescent "Hxt1-only" and "Hxt6-only" strains (orange and red paths respectively) are shown. Colored arrows indicate the direction the cell travels on each path as *g* increases. The arrows along the two axes (along "[glucose]" and "Glucose uptake rate") point in the direction of increase.

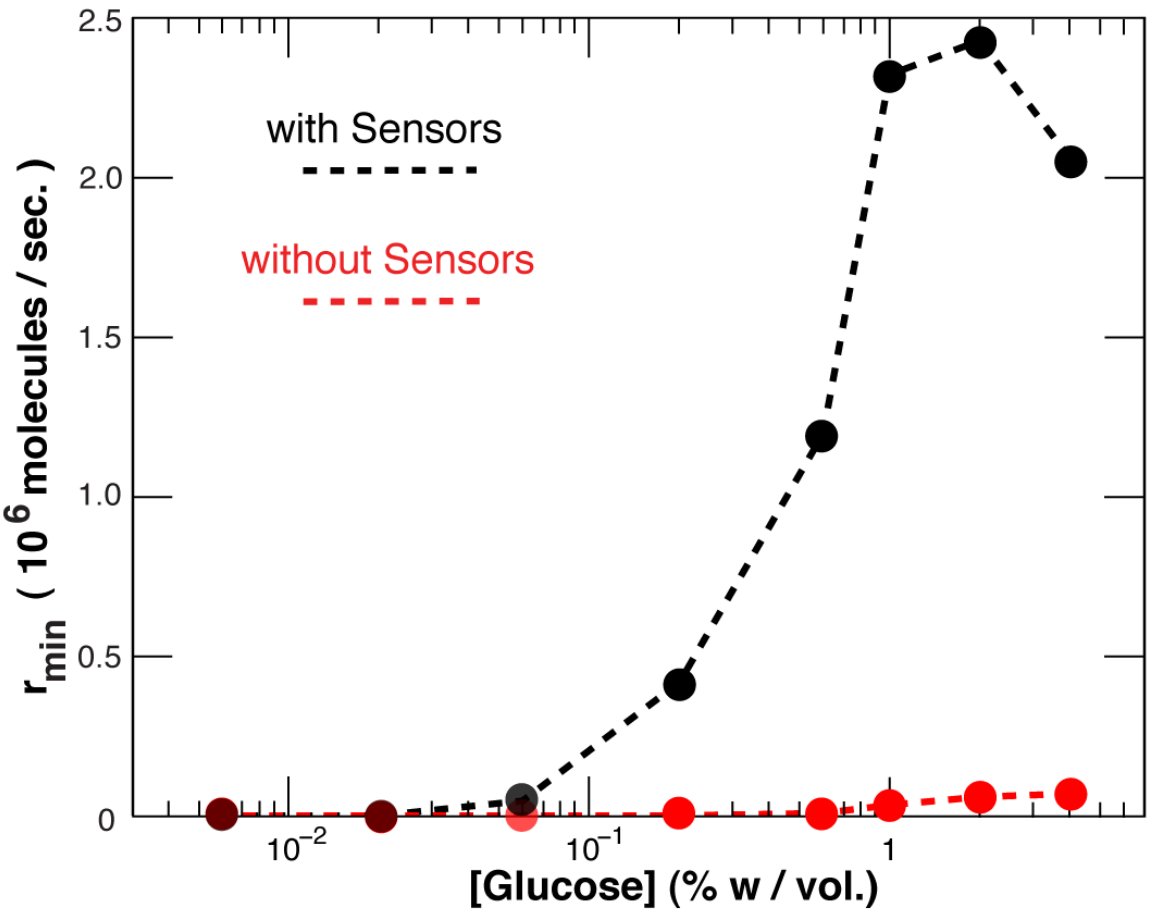


Figure 3.13.: Minimum glucose uptake rate r_{min} required for growth as a function of extracellular glucose *g*. $r_{\min} = r_c \exp\left(-\frac{\mu_c}{P(g)}\right)$ is the function describing these two curves, with *P(g)*, r_c, and μ_c fitted for a cell with the sensors (black line) and without the sensors (red line) (derived from Eqn [3.1]; see Figure 3.20 for "without sensor" strains). This shows that cells require a larger glucose uptake just to avoid growth arrest as more extracellular glucose is perceived.

3.5. Manipulation of glucose perception by sensors

Whereas the glucose uptake rate depends on the Hxts, glucose perception – captured by *P(g)* – should depend on mechanisms the cell uses to measure the level of extracellular glucose. Snf3 and Rgt2 are two glucose sensors primarily known for regulating transcription of both major and minor glucose transporter genes[18, 27] (*HXTs*, *GAL2*, *STL1, AGT1*). Since such regulation is disabled in our single-*HXT* strains, we could manipulate *P(g)* by knocking out these two glucose sensors without affecting the uptake rate *r*. We constructed a panel of single-*HXT* strains with these two sensors deleted (Figure 3.14). The relationship between growth rates and extracellular glucose concentration in these "sensorless" strains is strikingly different from that in strains with the two sensors intact (Figures 3.15 and 3.16). Growth rates now generally increase as the glucose level increases (at constant doxycycline level). Also, without the sensors the "Hxt3-only" and "Hxt6-only" strains no longer approach growth arrest as the glucose level increases. Because we deleted all minor glucose transporter genes and removed the glucose's control of the sole transporter expression in our single-*HXT* strains, changes in uptake rate were not the reason for the growth rescues we observed. For every combination of glucose and doxycycline concentrations, the uptake rate of the sensorless strains was nearly identical to the uptake rate of their sensor-containing counterparts (Figures 3.17 and 3.18).

In the sensorless strains, growth rate again explicitly depends on glucose concentration but with much reduced sensitivity (Figures 3.19 and 3.20). When Snf3 and Rgt2 are absent, a cell in 4% glucose acts as if it were in 0.06% glucose with intact sensors. Since the uptake rate remains virtually unchanged in the single-*HXT* strains when *SNF3* and *RGT2* are deleted, this reduced-sensitivity effect is due to a change in the perception function *P(g)*, not uptake rate *r* (Figure 3.20). The remaining dependence of the cell's growth rate on the glucose concentration even after Snf3 and Rgt2 have been deleted suggests that other sensors may contribute to the effect embodied in *P(g)*[28,29]. Nonetheless, our experiments show that Snf3 and Rgt2 are the key determinants of *P(g)* (as quantified in Figure 3.20).

The behavior depicted by Eqn. [3.1] should apply to the wild-type strain as well, as long as it achieves an uptake rate within the range probed with the single-*HXT* strains used to construct our growth landscape. We measured the wild-type's uptake rate and found that it was

below the critical uptake rate r_c for glucose concentrations smaller than 0.02% (Figures 3.11 and 3.24). For higher [glucose], the uptake rate exceeds r_c. When the wild-type cell's uptake rate is below r_c, its growth rate fits with the trend revealed in Figure 3.11. For higher glucose concentration, the effect of perception on the wild-type's growth rate disappears (Figure 3.11). One possible explanation is that as long as the glucose concentration is not too low, the wild-type escapes the seemingly detrimental effect of perception on growth rate by making enough hexose transporters to go beyond r_c. But for lower glucose level where its uptake is less than r_c, it properly tunes the interaction between glucose perception and uptake (quantified by the product $P(g)ln(r/r_c)$) such that its growth rate will increase when the cell perceives more extracellular glucose. Such tuning suggests that the transcriptional regulation of the *HXT* genes by Snf3 and Rgt2 is organized such that the wild-type always climbs uphill in the growth landscape (Figure 3.12) as it perceives an increase in the extracellular glucose concentration.

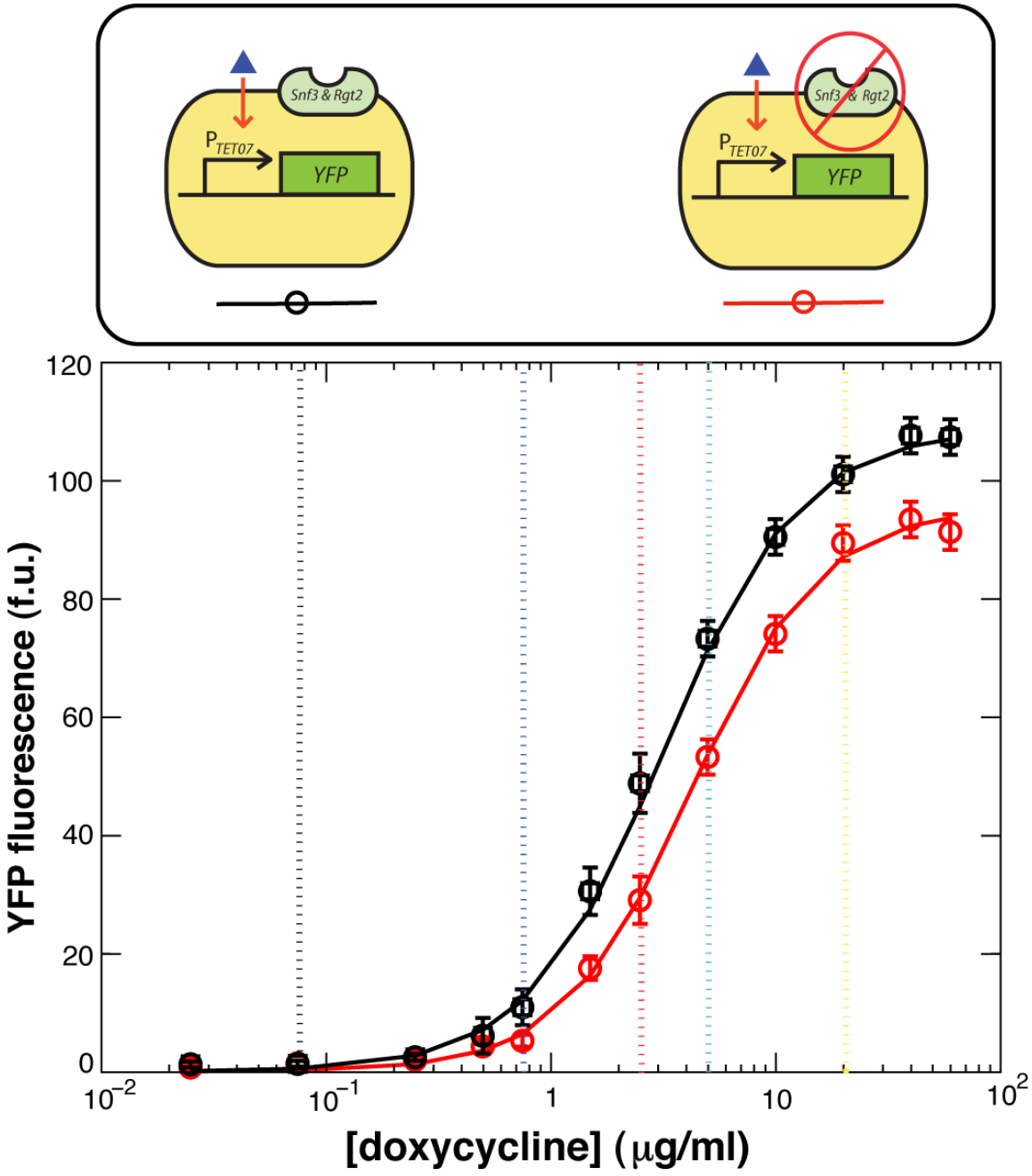


Figure 3.14.: Comparing the inducibility of P_{TET07} in HY4DCal5 (black line, with intact *SNF3* and *RGT2* genes) and in HY5FCal2 (red line, *snf3*Δ *rgt2*Δ). Single-cell fluorescence was measured using flow cytometer while HY4DCal5 and HY5FCal2 strains were in log-phase growth in standard synthetic media with 2% maltose and the indicated concentration of doxycycline. The colored vertical dashed lines indicate the concentrations of doxycyline used for characterizations of the "sensor-less" single-*HXT* strains (*snf3*Δ *rgt2*Δ) in subsequent supplementary figures. This plot shows that knocking out the two sensors makes only minor changes to the transcriptional activity of the promoter P_{TET07}. Error bars, s.e.m.; *n*=3.

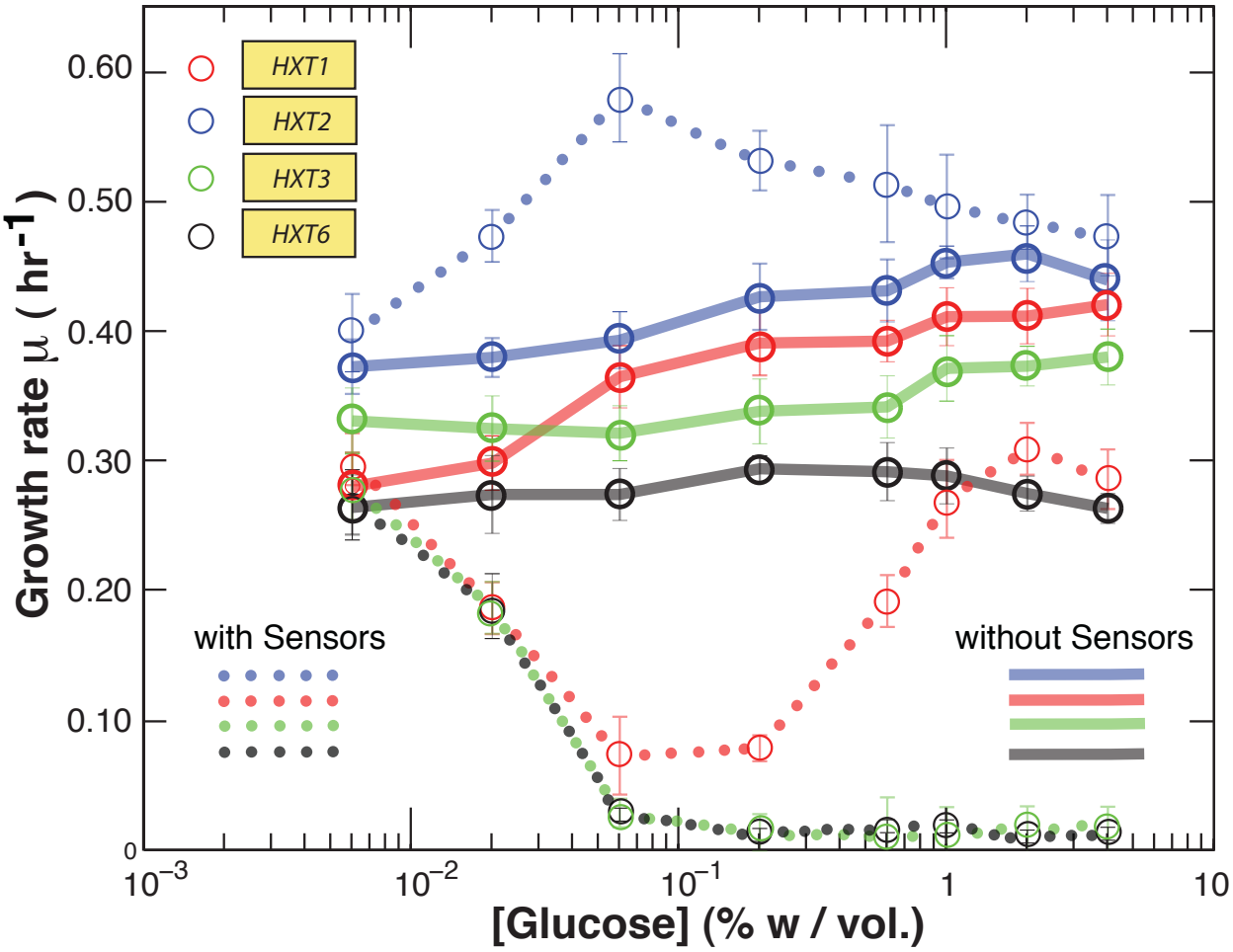
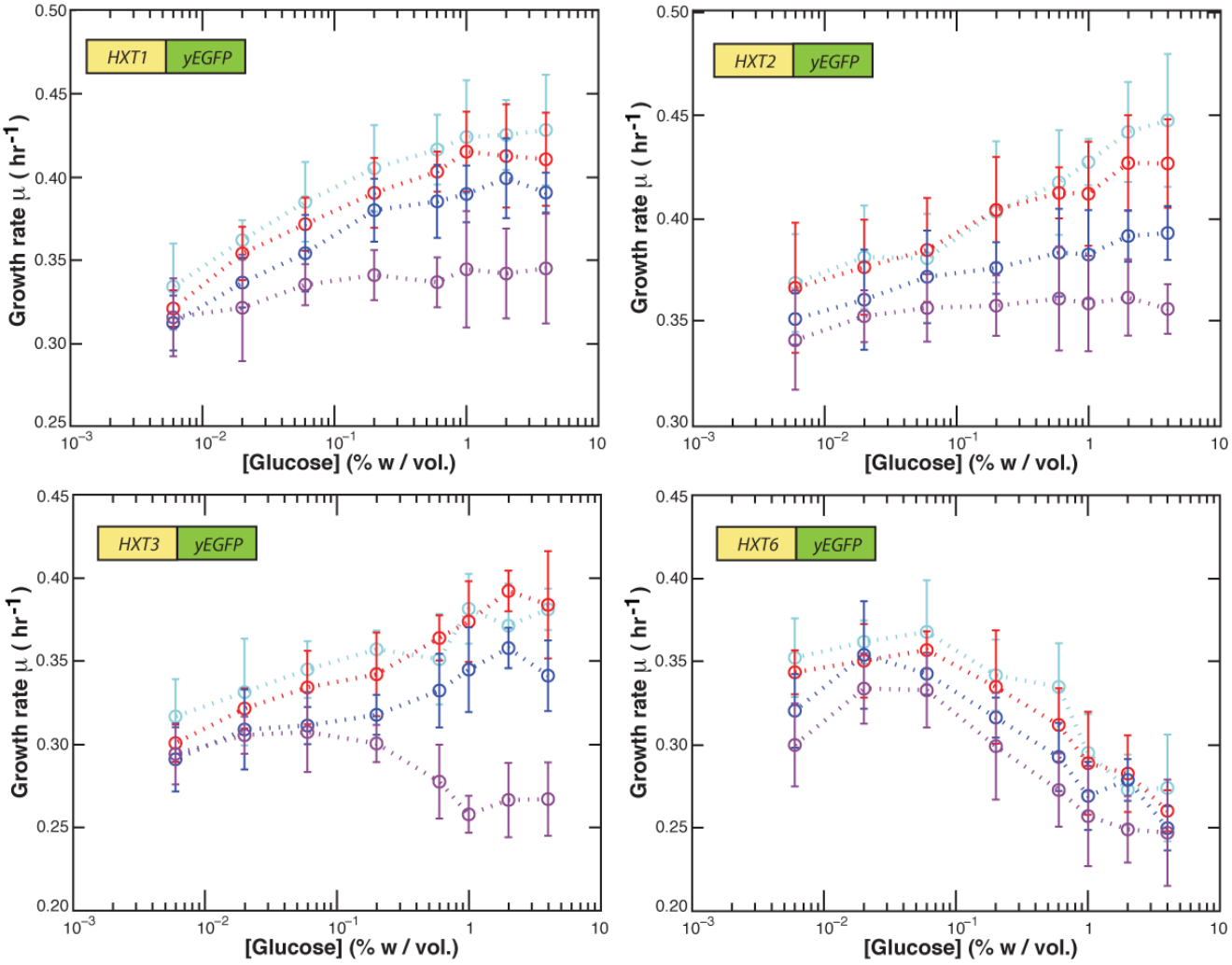


Figure 3.15. : Manipulation of the cell's perception of extracellular glucose yields significant growth-rate changes. Growth rates of single-*HXT* strains lacking two glucose sensors (*snf3*Δ*rgt2*Δ, bold lines) along with their counterparts with intact sensors (dotted lines) are shown for [doxycycline] = 5 μg/ml. Error bars, s.e.m.; n=3.



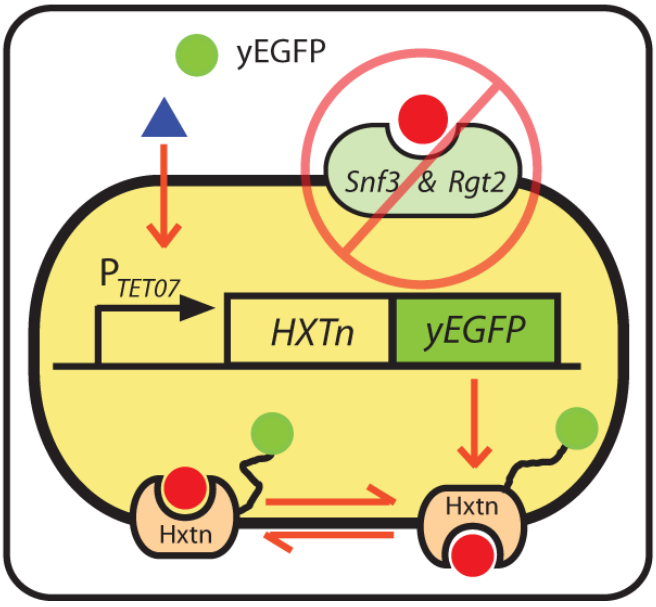


Figure 3.16. : Growth rates of fluorescent sensor-less single-*HXT* strains (*snf3*Δ *rgt2*Δ) in various combinations of glucose and doxycycline concentrations. Each color corresponds to a particular value of [doxycycline] indicated in Figure 3.14 (purple represents [doxycycline] = 0.25 μg/ml). These strains' growth rates behave in a starkly different manner from their sensor-intact counterparts (compare with Figure 3.6). "Hxt3-only" and "Hxt6-only" strains no longer approach near growth arrest when the two sensors are absent. Error, s.e.m. *n*=3.

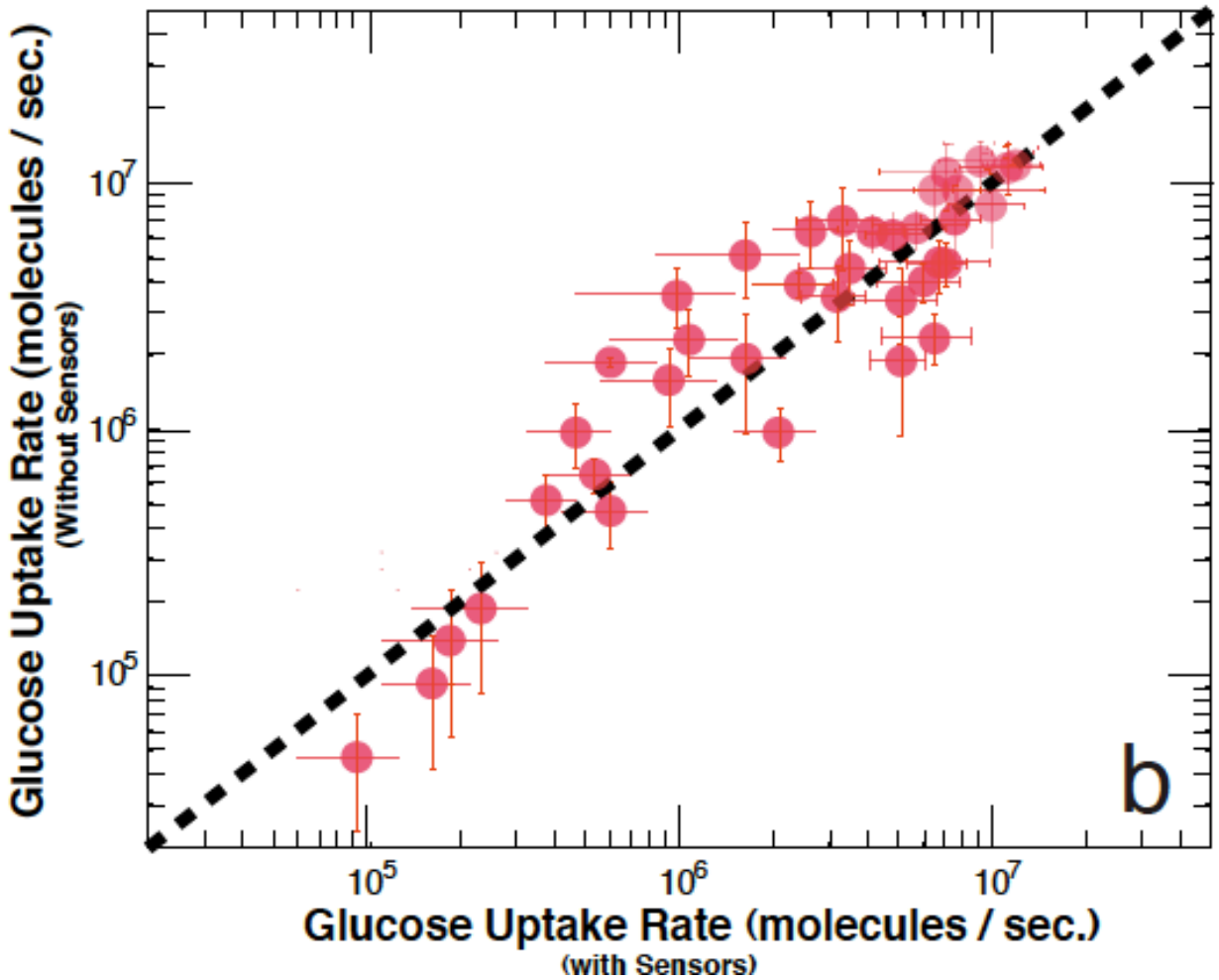
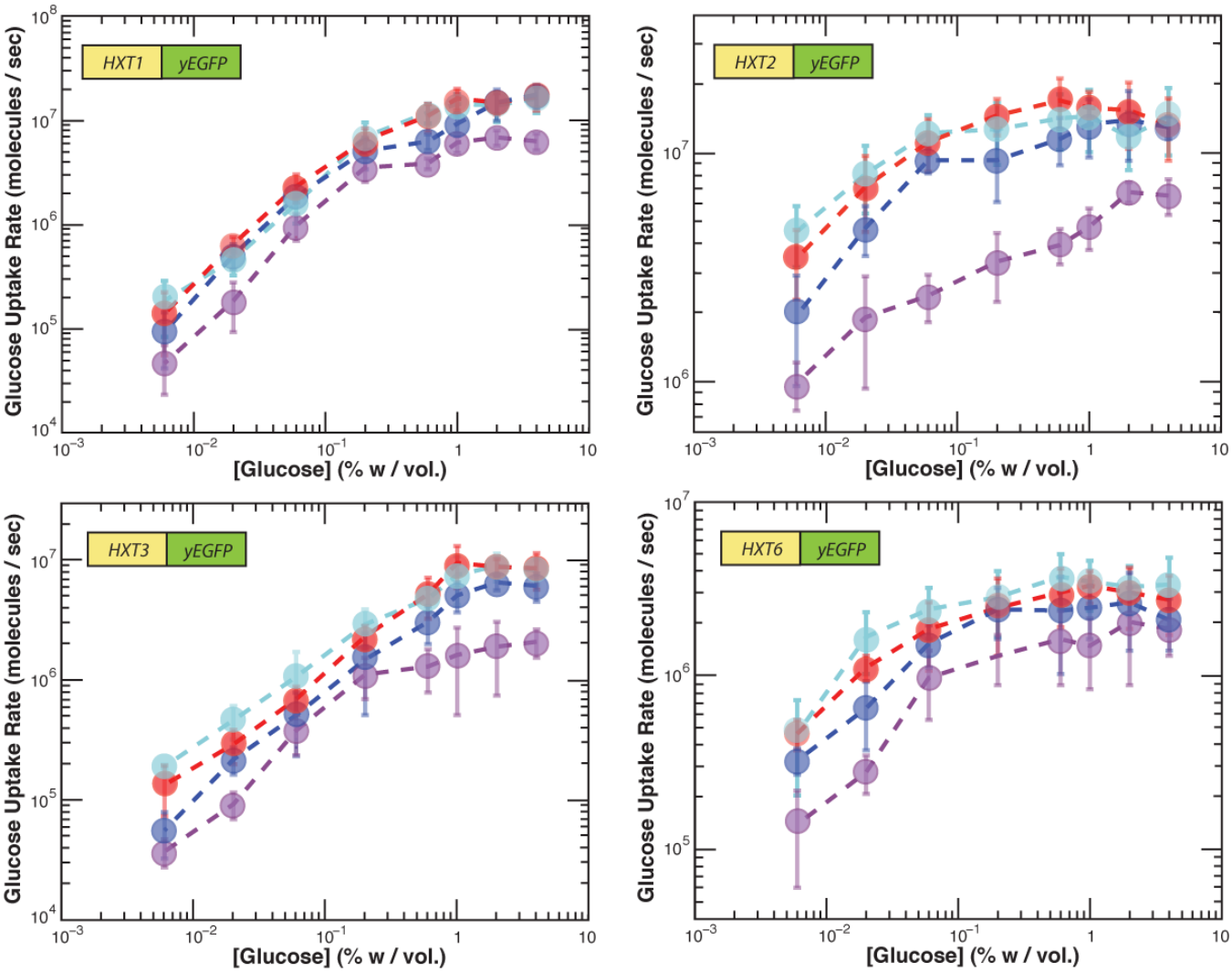


Figure 3.17. : Knocking out the two glucose sensors leaves the cell's glucose uptake rate virtually unperturbed. Just the "Hxt1-only" and "Hxt2-only" strains are shown here for simplicity (Figures 3.9 and 3.18 for others). Each data point represents a particular combination of glucose and doxycycline concentrations. Error bars, s.e.m.; n=3.



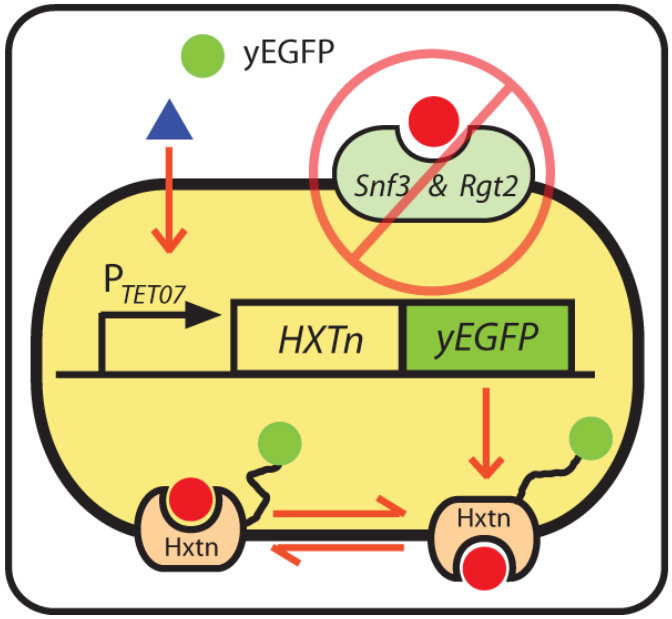


Figure 3.18.: Measured glucose uptake rates of sensor-less fluorescent single-*HXT* strains (*snf3*Δ *rgt2*Δ) in various combinations of glucose and doxycycline concentrations. Each color corresponds to a particular value of doxycycline concentration indicated in Figure 3.14 (purple represents [doxycycline] = 0.25 μg/ml). Knocking out the two glucose sensors hardly perturbs the glucose uptake rates of the single-*HXT* strains (compare with Fig. 3.9). However, the sensorless single-*HXT* strains' growth rates are qualitatively very different from those of their sensor-intact counterparts (can be seen by comparing Figures 3.16 and 3.6). This is due to the diminished sensing ability of the sensorless strains, as indicated by the significant decrease in *P([glucose])* (Figure 3.19). Having the sensors knocked out impairs the cell's ability to perceive how much extracellular glucose is surrounding the cell. In particular, the cell acts as if there is less glucose than there actually is (decrease in *P([glucose])*). Error bars, s.e.m.; *n*=3.

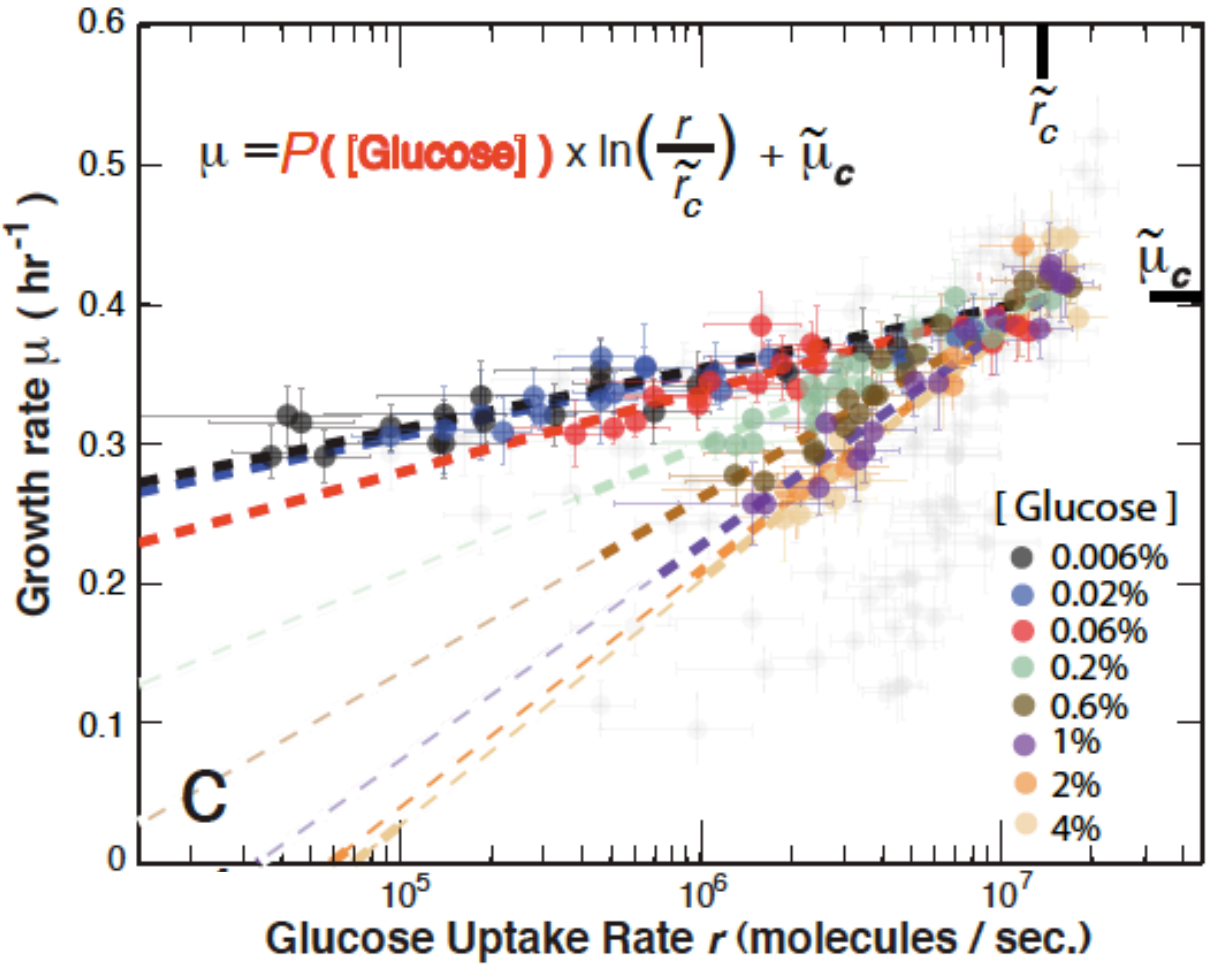


Figure 3.19. : Phenomenological model of yeast growth, in the absence of two primary glucose sensors. By yEGFP fusion, fluorescent sensor-less single-*HXT* strains were constructed for comparison with their sensor-intact counterparts. The features of growth rates seen in Fig. 3.15 were preserved after this fusion (Figure 3.16). Growth rates and glucose uptake rates of these strains were measured (Figures 3.18 and 3.21-23). For comparison, data for the sensor-intact single-*HXT* strains (from Fig. 3a) are shown in grey ($\tilde{\mu}_c$=0.40 hr^{-1}, $\tilde{r}_c$=1.4x10^7 molecules/s). Error bars, s.e.m.; n=3.

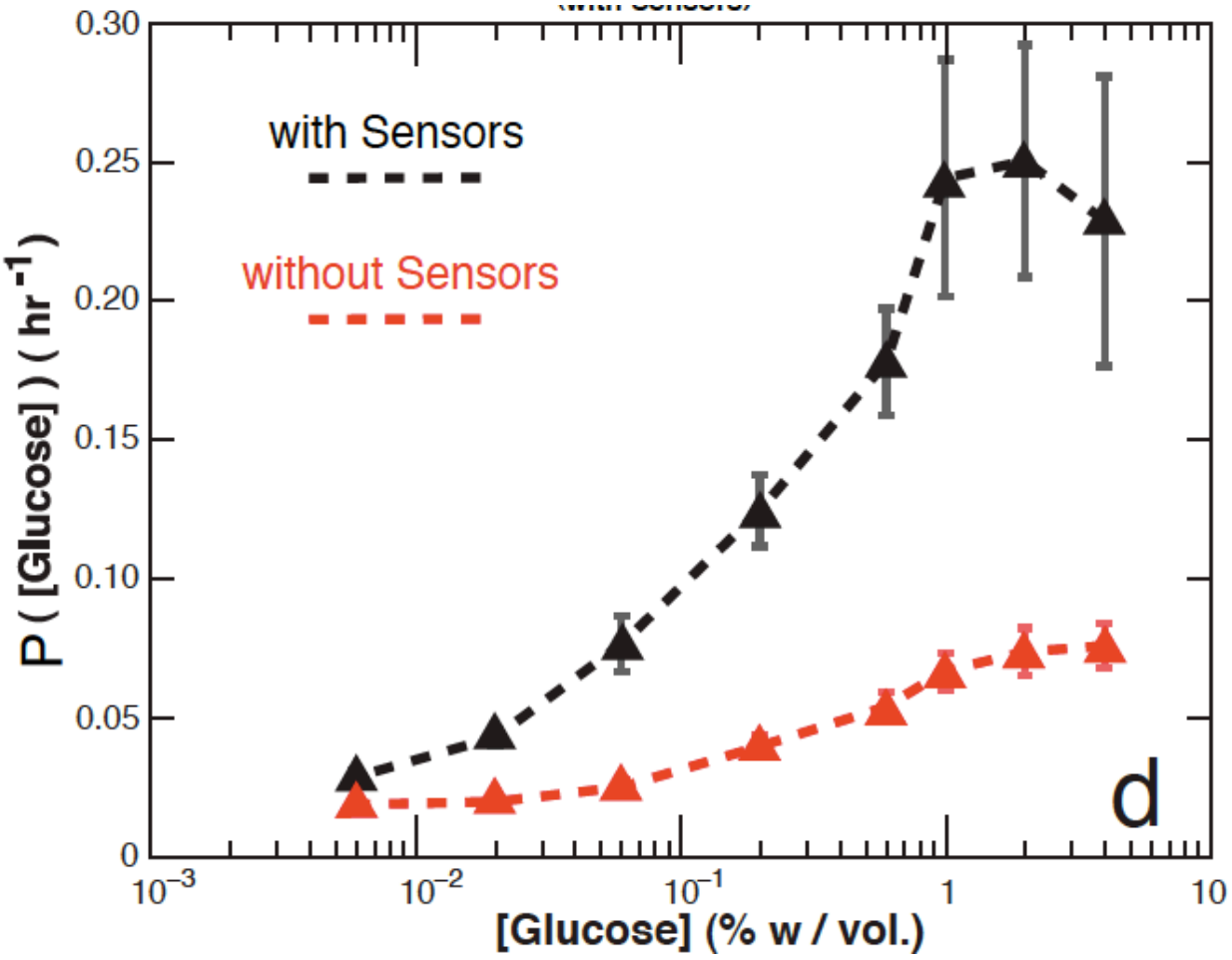
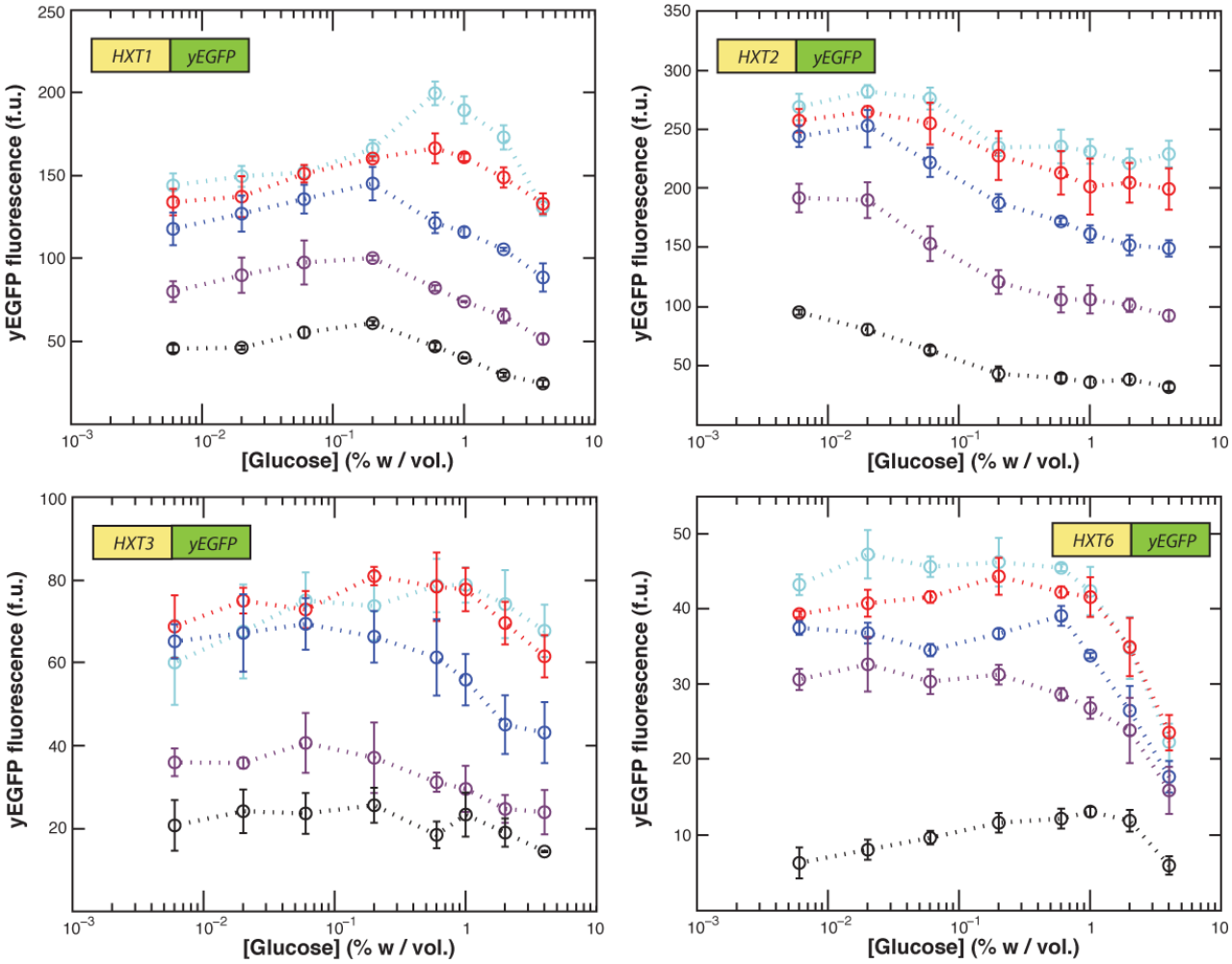


Figure 3.20.: Sensitivity to glucose is reduced in the absence of the two primary glucose sensors Snf3 and Rgt2. The sensitivity function *P(g)*, calculated from fitting the data in Figures 3.10 and 3.19 is shown for strains with intact sensors (black) and *snf3*Δ*rgt2*Δ strains (red). Error bars indicate 95% confidence interval in these fits.



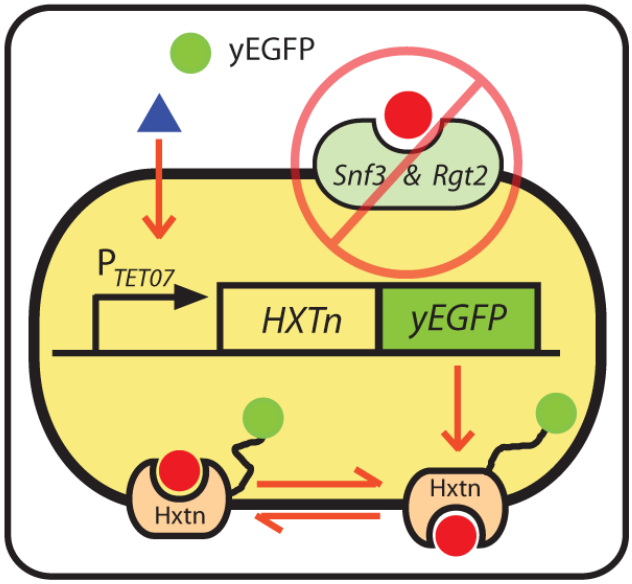


Figure 3.21. : Average single-cell, steady-state yEGFP fluorescence in the sensor-less single-*HXT* strains (*snf3*Δ, *rgt2*Δ) in various combinations of glucose and doxycycline concentrations. Each color corresponds to a particular value of doxycycline concentration indicated in Fig. 3.14 (purple represents [doxycycline] = 0.25 μg/ml). The relative number of Hxt proteins per cell was inferred from the average single-cell fluorescence measured using flow cytometer (See methods). Post-transcriptional regulations of Hxts are observed in these "sensor-less" strains, just as we observed such regulations in the "single-*HXT*" strains with the sensors (See Fig. 3.7). We took into account the effect of these regulations on the glucose uptake by directly measuring the glucose uptake rates. As in the sensor-intact single-*HXT* strains (Fig. 3.7), our measured glucose uptake rates of all the single-*HXT* strains monotonically increased as the glucose level increased (Fig. 3.18). The reason for this is identical to the one given in the figure caption for Fig. 3.7. Error bars, s.e.m.; *n*=3.

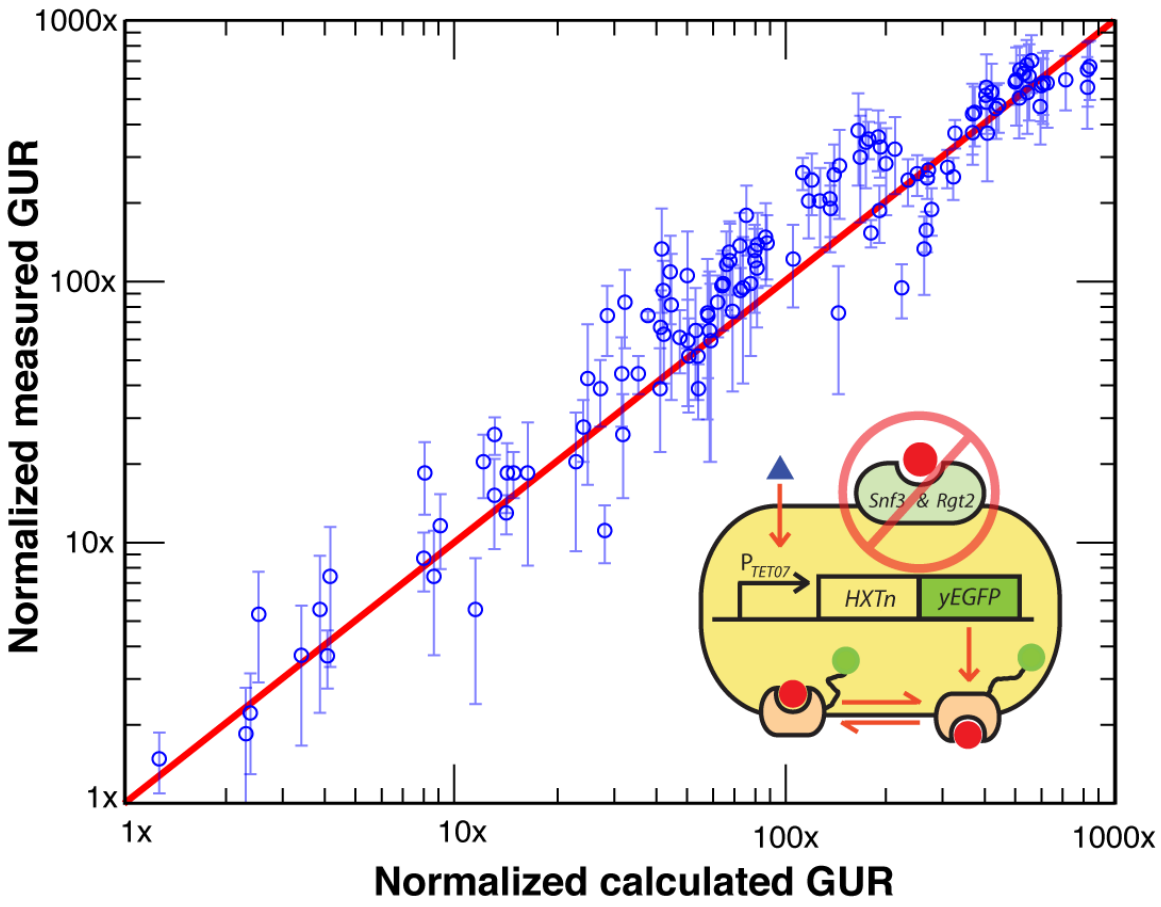
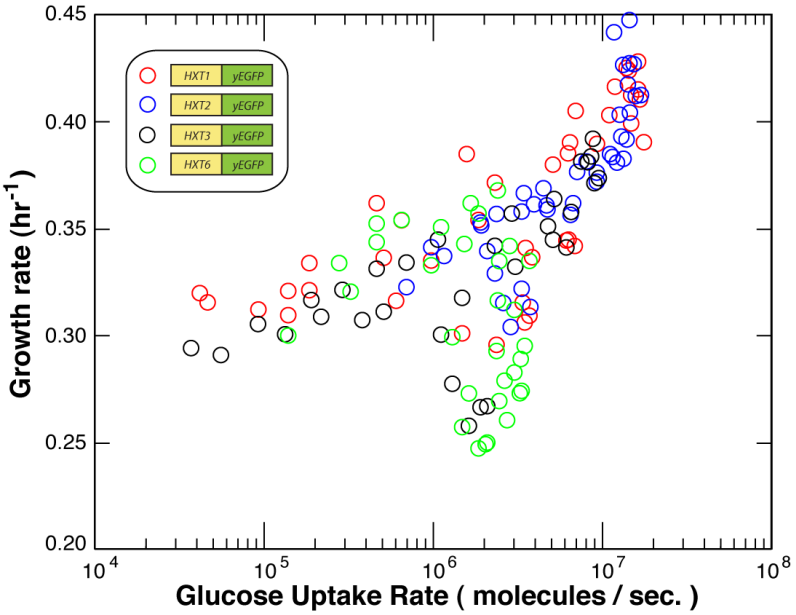


Figure 3.22.: Comparison of measured glucose uptake rate (GUR) and calculated GUR of the sensor-less fluorescent single-*HXT* strains (*snf3*Δ*rgt2*Δ). The measured and calculated values of glucose uptake rates of all the fluorescent single-*HXT* strains without the two glucose sensors (*snf3*Δ *rgt2*Δ) are compared here. GURs are reported in normalized units to show that the relative changes in both the measured and calculated GURs are in good agreement with each other. Error bars, s.e.m.; *n*=3.

a.



b.

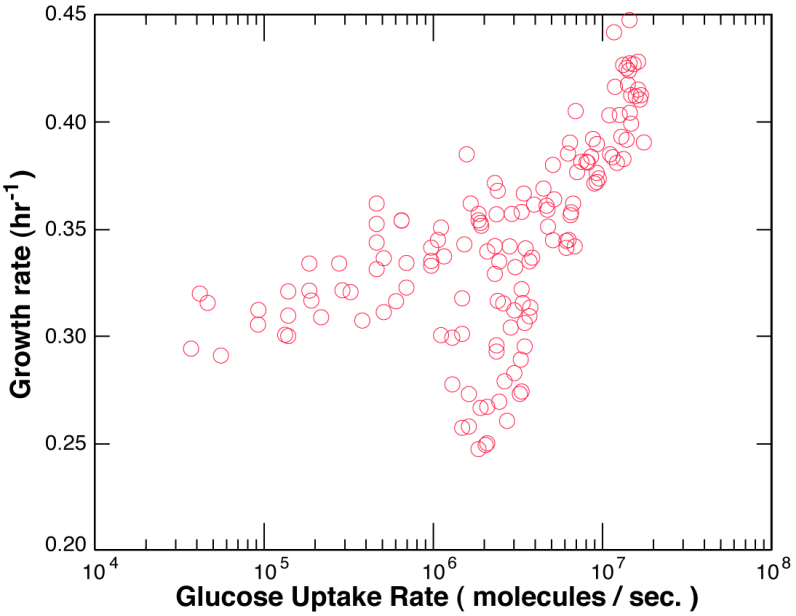


Figure 3.23. : (See next page for figure caption).

Figure 3.23.: Result of plotting the growth rate and the measured glucose uptake rate (GUR) of all the sensorless fluorescent single-*HXT* strains together. a. The color scheme represents the particular single-*HXT* strain (without *SNF3* and *RGT2*) to which the data points belong. The overlap of data points belonging to different single-*HXT* strains but at the same GUR and glucose concentration, along with the pattern emerged in Figure 3.19, together demonstrate that only the value of GUR but not which Hxt was responsible for the glucose import, is a factor in determining the growth rate. **b.** Obtained by removing the colors from the data points shown in (a). This shows that for a particular value of glucose uptake rate, multiple growth rates are possible. This means that glucose uptake rate alone cannot specify the growth rate. But additionally coloring these data points according to the value of the extracellular glucose concentration leads to a striking pattern observed in Figure 3.19.

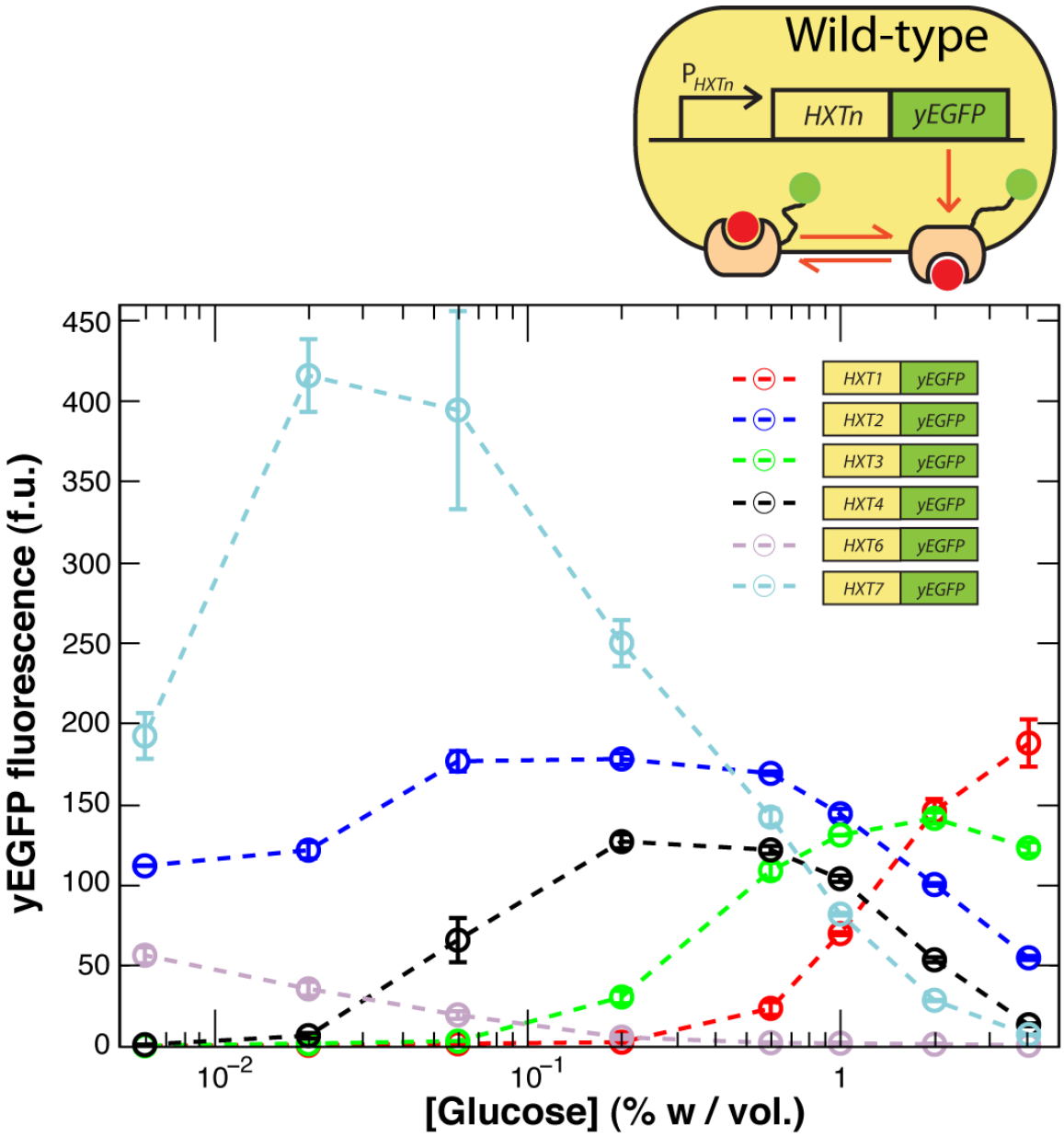


Figure 3.24.: Average single-cell fluorescence in the 'wild-type' strain (CEN.PK2-1C) with yEGFP fused to each *HXT* gene. The relative number of each Hxt protein present in the wild-type strain was inferred from these fluorescence levels. As in the single-*HXT* strains, the wild-type's glucose uptake rate that was calculated using these fluorescence values (See methods) was in close agreement with the directly measured glucose uptake rate (Figures 3.8 and 3.11). Error bars, s.e.m.; *n*=3.

3.6. Biological nature of the critical point (μ_c, r_c)

The critical point (μ_c, r_c) may represent a region of phase transition in the cell's growth and metabolism. The cell dramatically increases its ethanol production rate as its uptake rate increases above the critical rate r_c (Figure 3.25). This suggests that when its uptake rate is below r_c, the cell metabolizes glucose largely through respiration, but then switches to a largely fermentative metabolism as the uptake rate exceeds r_c. A key rate limiting step in fermentation is import of glucose and therefore the cell only redirects its glucose flux from respiration to fermentation when its glucose uptake rate is sufficiently high[30,31]. Our results suggest that this major redistribution of flux occurs around r_c.

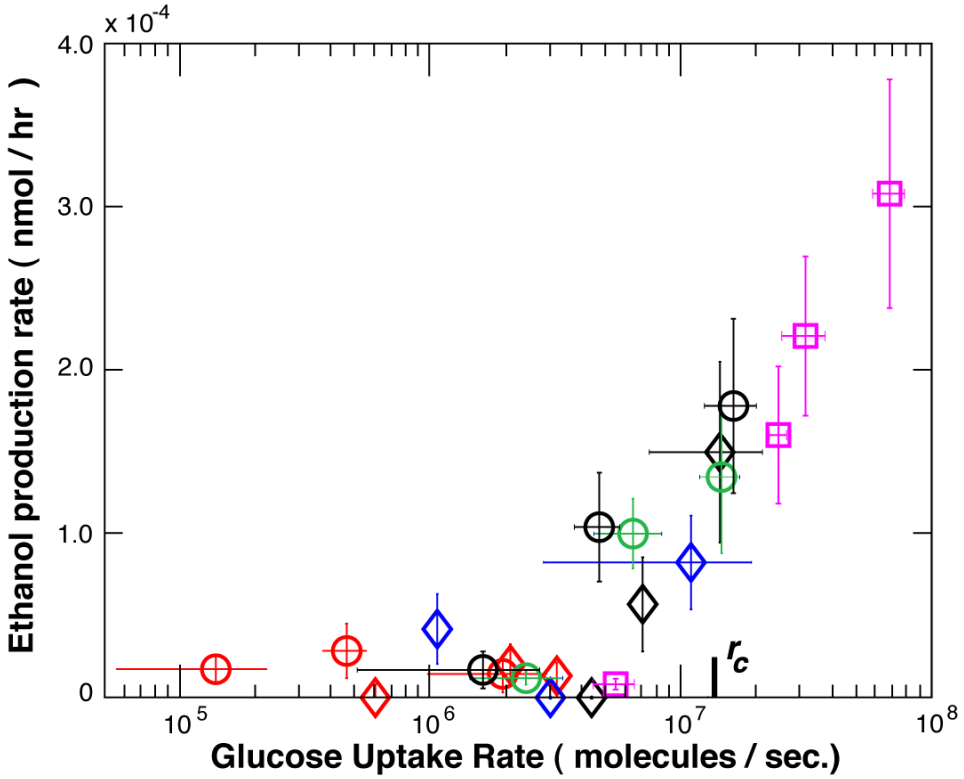


Figure 3.25.: Rate of ethanol production per cell suggests a shift from respiration to fermentation near the critical uptake rate r_c. Following a procedure essentially identical to the one used in measuring the glucose uptake rate (See methods), the average rate of ethanol production per cell was measured using a commerical ethanol assay kit (BioVision cat.#K620). Shown here are the ethanol production rates of the single-*HXT* strains with (diamonds) and without (circles) Snf3 and Rgt2 grown in various [glucose] (0.006% (red), 0.06% (blue), 0.2% (green), and 1% (black)). The ethanol production rates of the wild-type strain (squares) grown in these four values of [glucose] are also shown. The sharp increase in ethanol production rate indicates a shift from largely respirative to fermentative metabolism near the critical uptake rate r_c. Error bars, s.e.m.; n=3.

3.7. Growth model sets a constraint on faster growth of yeast

When the cell's glucose uptake rate r is lower than r_c, an increase in the extracellular glucose concentration causes two counteracting effects (Figure 3.11). First, since the Hxt is a passive transporter, it leads to an increased glucose uptake rate (which tends to increase the growth rate). Second, it causes the cell to perceive a higher amount of glucose (which tends to decrease the growth rate). This counteracting interaction between the two mechanisms – glucose perception and uptake – determines how the growth rate changes (i.e., whether it increases or decreases) as a result of an increase in the extracellular glucose level. The effect of this interaction on the growth rate is quantified by $P(g)\ln\left(\frac{r}{r_c}\right)$ which couples the two mechanisms. In particular, if g_0 and r_0 are glucose concentration and uptake rate of a cell in growth environment 'A', while g_1 and r_1 are for growth environment 'B', then the cell grows faster in environment 'B' than in environment 'A' if the following inequality is satisfied:

$$P(g_0)\ln\left(\frac{r_0}{r_c}\right) < P(g_1)\ln\left(\frac{r_1}{r_c}\right) \qquad \textbf{[3.2]}$$

Due to the counteracting nature of the two mechanisms, a higher uptake rate ($r_0 < r_1$) and a higher glucose level ($g_0 < g_1$) do not guarantee that above inequality will hold. This can be visualized in the growth landscape (Figure 3.12).

3.8. Balancing glucose perception and import.

While an increase in the amount of glucose in the cell's surrounding causes the Hxt to transport glucose faster due to the passive nature of the Hxt, it also results in the cell perceiving a higher amount of glucose that can decrease the cell's growth rate as seen Figures 3.11 and 3.12 (unless the cell's uptake rate is larger than r_c). If the cell wishes to prevent its growth rate from decreasing in this situation, it has to not only increase its uptake rate but do so by at least a certain minimum amount. The cell can achieve this by changing both the number and type of Hxt it makes as a function of glucose. To formalize this notion, let $N_{HXTn}(g)$ be the number of Hxt type "n" the cell makes when it senses a particular concentration of glucose g in its surrounding. Then the cell's total uptake rate r is a function $r = r(\{N_{HXTn}(g)\}, g)$ where $\{N_{HXTn}(g)\}$ is the set of all types of Hxts made by the cell. The growth rate $\mu(\{N_{HXTn}(g)\}, g))$ then is a curve parameterized by g: it is a particular "growth trajectory" in the space of all possible growth rates

(Fig. 3c). Hence the particular set $\{N_{HXTn}(g)\}$, hard-wired into the cell by transcriptional regulation of the *HXT* genes, determines the particular "growth trajectory". Using the expression for growth rate $\mu(r,g) = P(g)\ln\left(\frac{r}{r_c}\right) + \mu_c$ obtained in Figure 3.11, the requirement that the cell's growth rate never decreases whenever *g* increases ($d\mu/dg \geq 0$) means that the following inequality has to be satisfied at all points on its growth trajectory:

$$\text{(Perception)} \quad \frac{dP/dg}{P} \leq \frac{d\ln(r/r_c)/dg}{\ln(r_c/r)} \quad \text{(Import).} \qquad \textbf{[3.3]}$$

Those parts of the growth trajectory where above inequality is not met correspond to the cell's growth rate decreasing despite an increase in *g*. Notice the left hand side of Eqn.[3.3] involves only the effect of glucose perception while the right hand side involves only the glucose import. Above inequality represents the balance of fractional changes due to these two separate effects. Any synthetic rewiring of the transcriptional regulation of the *HXT* genes, leading to changing the set $\{N_{HXTi}(g)\}$ from the wild-type values, should be done mindful of above inequality: a lesson learned from the "single-*HXT*" strains.

3.9. Possible molecular mechanisms underlying the effects of glucose perception and import

Glucose metabolism, involving thousands of chemical reactions and numerous intracellular events (gene regulations, post-transcriptional modifications, etc.), is a complex process. But the equation for growth rate $\mu(r,g) = P(g)\ln\left(\frac{r}{r_c}\right) + \mu_c$ obtained in Figure 3.11 shows us that such a complex set of components can work in concert to yield a simple description. Since glucose metabolism involves thousands of intracellular activities ranging from metabolic reactions, transcription of many genes, and post-transcriptional modifications, it is difficult to pinpoint to a single, or most likely, many correlated molecular events that underlie the phenomenological growth model revealed in our study. As a case example, a recent microarray study[1] has shown that hundreds of genes involved in ribosomal biogenesis, which are energetically very costly due to their relatively short half-lives, are up-regulated by many

decade-folds as the extracellular glucose concentration is increased from 0.01%, to 0.1%, and then to 1% w/vol. Future studies that look at the global expression levels while varying the glucose perception and import independently of each other may help in distinguishing what fraction of these expression level changes are due to (1) changes in the perception of extracellular glucose level as opposed to (2) changes in the glucose import rate. Such a study would shine light into the transcriptional regulations that may be responsible for the growth effects embodied in our growth model. In addition, a large-scale study that measures the changes in the metabolites (using mass spectrometry, for instance) while the cell's perception and import of glucose are varied independently of each other over a wide range will further elucidate what is likely to be a vast number of molecular mechanisms underlying the phenomenological growth model uncovered in our study. It is well known that glucose-mediated post-transcriptional modifications exist, especially of metabolic proteins[2]. Measuring how these events are affected separately by glucose perception and import on a global scale would be difficult but worthwhile. Decupling the glucose perception from glucose import in large-scale studies will yield valuable insights in understanding the vast molecular events that are likely working in concert to produce the phenomenological growth model revealed in our study.

3.10. Why would the phenomenological model dictate yeast's growth?

Glucose perception and import are two separable modules that each affects the growth rate. But it is the interaction between the two modules that ultimately determines the cell's growth rate, and that interaction can be both precisely altered and measured. But why would it make sense that yeast grows according to Eqn. [3.1], which allows for a possible detrimental growth if the interaction between the perception and import modules is not properly tuned? One explanation may be that yeast has no way to directly 'measure' its glucose import rate in real-time. Indeed, there is no known 'flux sensor' that the yeast uses to measure its glucose import rate in real-time and then adjust the production level of Hxts to change the glucose import rate if the yeast senses that the flux is too low. In fact, Hxt expression levels are primarily set by the extracellular glucose concentration[32] (Figure 3.11). While yeast certainly can measure the extracellular glucose level directly and the intracellular glucose level indirectly (for example, through the catabolite-repressor Mig1 which uses intracellular glucose as its substrate)[33-35], knowing the two

glucose levels is not sufficient for the yeast to infer what its glucose import rate is. This is because a given steady-state glucose concentration gradient can be maintained by a combination of wide ranges of glucose import rate and intracellular glucose breakdown rate. Since the cell has no direct way to measure the breakdown rate (there is no known 'rate sensor' measuring intracellular glucose breakdown), the cell cannot infer what the glucose import rate is in real-time just from the difference between extracellular and intracellular glucose. Given the engineering difficulty of building 'flux sensors', the yeast may have solved the problem by evolving glucose sensors such as Snf3 and Rgt2 to measure the extracellular glucose level, then anticipate a certain glucose import rate would be achieved, set up intracellular activities to process glucose being imported at the anticipated rate, and make sure that such an import rate is indeed achieved by putting its *HXT* genes under the control of those glucose sensors (Figure 3.11).

On-going efforts at large-scale modeling of glucose metabolism, gene regulation[36] and cellular signaling must decouple and consider how the cell's response varies when glucose uptake rate is varied independently of extracellular glucose level. For instance, microarray studies have shown that hundreds of genes involved in ribosomal biogenesis, which are energetically very costly, are up-regulated by many folds as the yeast is subjected to ever increasing levels of glucose[37]. In these studies, as the level of glucose is increased, so does the glucose import rate. These observed large-scale changes are thus due to the conflated effects of glucose perception and import. It would be interesting to measure which of these changes are due to glucose perception and import separately by decoupling the two effects. We hope that our model, as well as the framework used to extract some key principles from the complexity underlying yeast growth, will assist ongoing efforts to rationally engineer[38-40] and understand microbial metabolism at the systems-level[41-48].

Summary of experimental methods used

Strain background and construction. A list of strains with diagrams summarizing their key features are provided in the supplementary information section. All strains were derived from the haploid strain CEN.PK2-1C (*MAT* α, kind gift from E. Boles)[24], referred to as the "wild-type" in our study. Both EBY.VW4000 and EBY.VW5000 are deficient in hexose transport due to deletions of all *HXT* genes as well as genes encoding transporters with minor glucose uptake capabilities (*agt1Δ ydl247wΔ yjr160Δ*)[24]. HY4D1 and HY5F1 each contain rtTA protein expressed constitutively by the *MYO2* promoter (inserted into EBY.VW4000 and EBY.VW5000 respectively using plasmid pDH18 (EUROSCARF) containing *HIS5* gene) and CFP constitutively expressed by P_{TEF1}. *Xho*I-P_{TET07}-*Bam*HI, *Bam*HI-*HXTn*-*Not*I fragments were cloned into pRS305 (EUROSCARF) backbone containing *LEU2* gene (*n*= 1-4, 6). Integrating these plasmids into defective *LEU2* locus (leu2-3) in HY4D1 by linearizing the plasmids with *Nar*I, the "single-*HXT*" strains were constructed. To construct fluorescent "single-*HXT*" strains, *yEGFP*-T_{ADH1}-*Kan* fragment was amplified from pKT127 plasmid (EUROSCARF) and was fused to C-terminus of *HXTn* ORF in each of the single-*HXT* strains via standard PCR integration[49]. This fragment was also fused to C-terminus of *HXTn* ORF (*n*=1-4, 6, 7) in CEN.PK2-1C, thus resulting in six fluorescent wild-type strains (Figure 3.24). The "sensorless" versions of single-*HXT* strains (*snf3Δ rgt2Δ)* were constructed in the same way as their sensor-intact counterparts mentioned above by using HY5F1 instead of HY4D1. To probe the wild-type's transcriptional regulation of each of the *HXT* genes (Figure 3.1), *Xho*I-P_{HXTn}-*Bam*HI, *Bam*HI-*YFP*-*Not*I fragments were cloned into pRS305 backbone containing *LEU2* gene (*n*=1-4, 7) and was integrated into the defective *LEU2* locus (leu2-3) in CEN.PK2-1C by linearizing the plasmid with either *Nar*I (for *n*=1) or *Cla*I (for all other *n*), resulting in five strains. The P_{HXT1}, P_{HXT2}, P_{HXT3}, P_{HXT4}, and P_{HXT7} promoter sequences refer to 1941-, 850-, 1996-, 1544-, 2042-bp upstream of the start codon of the respective genes. These sequences include all the known binding sites of transcription factors for the respective genes[50].

Growth rate measurements. All growth rates reported in our study were measured while the cells were in log-phase growth in 5 mL batch cultures at 30 ^{o}C, in a standard synthetic media with various combinations of glucose and doxycycline concentrations. To bring the cells into

log-phase, the single-*HXT* strains were first grown in a standard synthetic media containing 2% maltose and the desired concentration of doxycycline until the cells have been in log-phase for roughly 12 hours. This procedure ensured that the cells were already making Hxts needed to initiate glucose uptake immediately after being transferred to glucose media. Then these cells were diluted into the standard synthetic media with the same amount of doxycycline, but this time containing glucose instead of maltose. These dilutions were done such that by the time the density of cells in the batch culture reached level detectable by our spectrophotometer (Hitachi U-1800) (roughly 15 hours after dilution), the cells had adjusted to the glucose media and were in log-phase growth. Hence, the transient growth rate change associated with maltose to glucose media transfer did not enter into our growth rate measurements. In a separate experiment, we confirmed this was indeed the case by further diluting these cultures into an identical glucose media, which showed that having the cells pre-grown in maltose before did not affect the growth rates reported in our study. By measuring the Optical Density (OD_{600} : Absorbance at $\lambda = 600$ nm) of these batch cultures over time, we extracted the growth rate of the cells. Strains that approached growth arrest also went through the same procedure as above. Upon transfer to glucose media from maltose media, these cells' growth rates transiently decreased to nearly zero during a period of roughly 24 hours. By looking at the cells under the microscope, no abnormal cell morphologies were detected, thus indicating normal growth (i.e. no pseudohyphal or filamentous growth was detected).

Fluorescence measurements. The average single-cell fluorescence due to yEGFP fused to C-terminus of *HXT* genes in both the wild-type and the single-*HXT* strains was measured using a Becton Dickinson FACScan flow cytometer with excitation laser at 488nm. Emission filter FL1 (530/30) was used to detect the yEGFP fluorescence levels as well as the YFP for determining the P_{TET07} induction curves in the calibration strains HY4DCal5 and HY5FCal2. Before observation using FACscan, the strains were grown using the protocol outlined in "growth rate measurements" section. The mean fluorescence values reported in our study represent the steady-state levels of Hxt proteins in single cells, since no appreciable changes in fluorescence was detected while the cells were growing in log-phase.

Glucose uptake rate measurements and calculations. Glucose uptake rates of cells were determined by measuring the rate of glucose depletion in the growth medium while the cells were in log-phase growth. First, the reasoning behind this procedure is as follows: If the cell's growth rate at glucose concentration G_0 is μ, $G(t)$ is the concentration of glucose in the growth medium at time t, $r(G(t))$ is the uptake rate per OD of the cells as a function of extracellular glucose, and r_0 is the OD of cells at t = 0, then the decrease in glucose concentration in the growth medium over time t is

$$G_0 - G(t) = \int_0^t r(G(\tau))\rho_0 \exp(\mu\tau)d\tau . \qquad \textbf{[3.4]}$$

If this change in glucose concentration is sufficiently small, but large enough to be detectable by our chemical assay (described below), then we can approximate $r(G(t)) \approx r(G_0)$ and μ as a constant during the time interval t. Then above equation can be solved for $r(G_0)$:

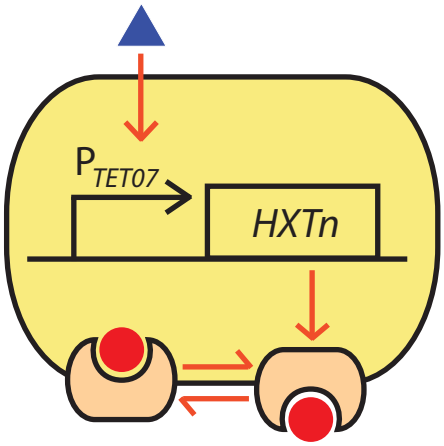
$$r(G_0) \approx \mu \frac{(G_0 - G(t))}{\rho(t) - \rho_0}, \qquad \textbf{[3.5]}$$

where $r(G_0)$ is the uptake rate per OD_{600}, measured in units of mM hr^{-1} OD_{600}^{-1}. This was then converted into molecules sec^{-1} $cell^{-1}$ using conversion factor 1.7 x 10^7 cells ml^{-1} OD_{600}^{-1}. $\rho(t) - \rho_0$ is the change in OD_{600} of the cells measured using the spectrophotometer (Hitachi U-1800), and μ is the growth rate determined by the method mentioned previously. The change in glucose concentration $G_0 - G(t)$ was measured using the standard commercial glucose assay kit (Sigma G3293) based on conversion of glucose through hexokinase and $NADP^+$ dependent glucose-6-phosphate-dehydrogenase. We compared the measured glucose uptake rates with the uptake rates calculated using an independent method for the fluorescent single-*HXT* and wild-type strains. We calculated the glucose uptake rates by using the known Michaelis-Menten parameters (V_m and K_m) of Hxts[26] and the relative number of Hxt proteins per cell inferred from measuring the average single-cell yEGFP fluorescence (Supplementary Information). These comparisons showed a close agreement between our measured and calculated uptake rates (Figures 3.8 and 3.22).

Strains used in this study:

Doxycycline Glucose yEGFP

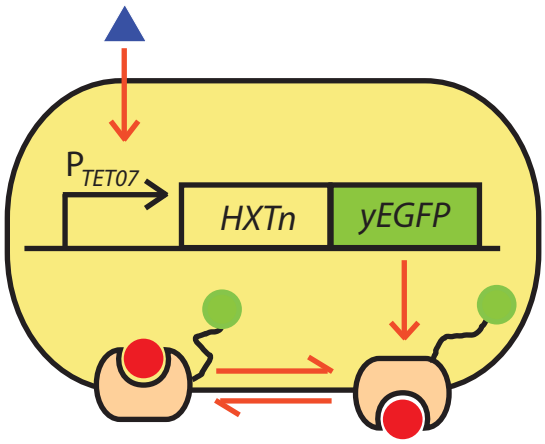
"Single-*HXT*" strains



HY4D1 *LEU2* P_{TET07}: *HXTn*

	Strain name
HXT1	H1C3
HXT2	H2C1
HXT3	H3C1
HXT4	H4C9
HXT6	H6C4

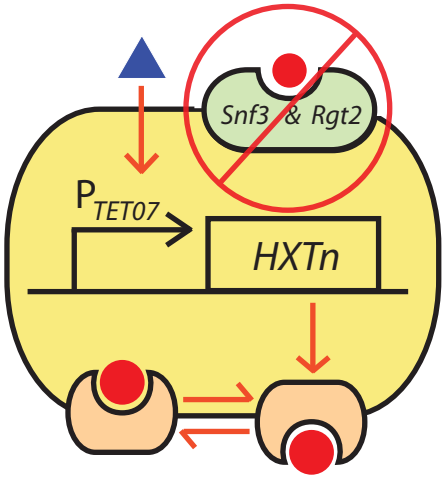
"Single-*HXT*::yEGFP" strains



"Single-*HXT*" strain *Kan*^{R} *HXTn*::*yEGFP*

	Strain name
HXT1	H1C3Fus14
HXT2	H2C1Fus18
HXT3	H3C1Fus32
HXT4	H4C9Fus1
HXT6	H6C4Fus20

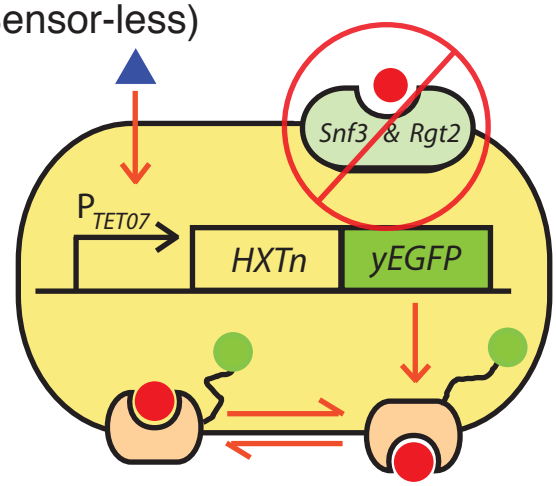
"Single-*HXT*" strains (Sensor-less)



HY5F1 *LEU2* P_{TET07}: *HXTn*

	Strain name
HXT1	H1NS2
HXT2	H2NS5
HXT3	H3NS5
HXT4	H4NS2
HXT6	H6NS2

"Single-*HXT*::yEGFP" strains (Sensor-less)

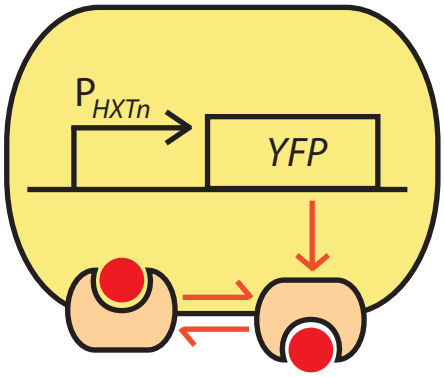


"Single-*HXT*" strain *Kan*^{R} *HXTn*::*yEGFP* (Sensor-less)

	Strain name
HXT1	H1NS2Fus30
HXT2	H2NS5Fus4
HXT3	H3NS5Fus2
HXT4	could not be made
HXT6	H6NS2Fus22

Strains used in this study (Continued):

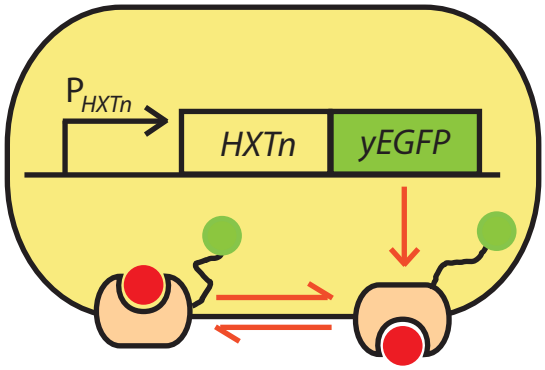
Wild-type P_{HXTn}:*YFP*



CEN.PK2-1C *LEU2* P_{HXTn} :*HXTn*

	Strain name
HXT1	H1S1
HXT2	H2S2
HXT3	H3S1
HXT4	H4S2
HXT7	H7S2

Wild-type *HXTn*::yEGFP



CEN.PK2-1C *Kan*^{R} *HXTn*::*yEGFP*

	Strain name
HXT1	CenH1Fus10
HXT2	CenH2Fus5
HXT3	CenH3Fus4
HXT4	CenH4Fus10
HXT6	CenH6Fus10
HXT7	CenH7Fus15

Strains for studying P_{TET07} induction:

Strain name		Genotype	Notes
HY4DCal5	EBY.VW4000	*HIS5 LEU2* P_{MYO2} *:rtTA* P_{TET07}*:YFP*	Used in Fig. S2.
HY5FCal2	EBY.VW5000	*HIS5 LEU2* P_{MYO2} *:rtTA* P_{TET07}*:YFP*	Used in Fig. S10.

Others:

Strain name	Genotype	Notes
HY4D1	EBY.VW4000 *HIS5* P_{MYO2}*:rtTA*	Common parent for all "single-*HXT*" strains (with sensors)
HY5F1	EBY.VW5000 *HIS5* P_{MYO2}*:rtTA*	Common parent for all "single-*HXT*" strains (*snf3*Δ *rgt2*Δ)
EBY.VW4000	(See Reference*)	*hxt1-17*Δ *RGT2* and *SNF3* intact. Gift from E. Boles.
EBY.VW5000	(See Reference*)	*hxt1-17*Δ *rgt2*Δ *snf3*Δ. Gift from E. Boles.
CEN.PK2-1C	(See Reference*)	"Wild-type" used in this study. Gift from E. Boles.

Note: EBY.VW4000 and EBY.VW5000 are both unable to grow on glucose since all HXTs as well as genes for transporters with minor glucose uptake capability had been deleted (hxt1-17Δ agt1Δ stl1Δ gal2Δ).

* R. Wieczorke, S. Krampe, T. Weierstall, K. Freidel, C. Hollenberg, and E. Boles. "Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*", *FEBS Letters*, **464** (3), 123-128 (1999).

Bibliography for Chapter 3

[1.] Monod, J. *Recherches sur la croissance des cultures bacteriennes*. (Hermann et Cie, 1942).

[2.] Bennet, M. R. *et al.* Metabolic gene regulation in a dynamically changing environment. *Nature* **454**, 1119-1122 (2008).

[3.] Zaslaver, A. *et al.* Just-in-time transcription program in metabolic pathways. *Nature Genetics* **36**, 486-491 (2004).

[4.] Airoldi, E. *et al.* Predicting cellular growth from gene expression signatures. *PLoS Comp. Biol.* **5**, e1000257 (2009).

[5.] Dekel, E., & Alon, U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588-592 (2005).

[6.] Krishna, S., Semssey, S., & Sneppen, K. Combinatorics of feedback in cellular uptake and metabolism of small molecules. *Proc. Natl Acad. Sci. USA.* **104**, 20815-20819. (2007).

[7.] Ihmels, J., Levy, R., & Barkai, N. Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nature Biotechnology.* **22**, 86-92 (2003).

[8.] Famili, I., Forster, J., Nielsen, J., & Palsson B.O. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proc. Natl Acad. Sci. USA.* **100**, 13134-13139 (2003).

[9.] Bilu, Y., Shlomi, T., Barkai, N., & Ruppin, E. Conservation of expression and sequence of metabolic genes is reflected by activity across metabolic states. *PLoS Comp. Biol.* **2**, e106.

[10.] Levine, E. & Hwa, T. Stochastic fluctuations in metabolic pathways. *Proc. Natl Acad. Sci. USA.* **104**, 9224-9229. (2007).

[11.] Fell, D. A. Understanding the control of metabolism. (Portland, London) (1997).

[12.] Savageau, M.A., Biochemical systems analysis: A study of function and design in molecular biology (Addison-Wesely, Reading, MA 1976).

[13.] Goyal, S. & Wingreen, N.S. Growth-induced instability in metabolic networks. *Phys. Rev. Lett.* **98**, 138105:1-4. (2007).

[14.] Nielsen, J., Villadsen, J., & Liden, G. *Bioreaction Engineering Principles* 235-311 (Springer, 2003).

[15.] Dickinson, J. R. & Schweizer, M. *The Metabolism and Molecular Physiology of Saccharomyces cerevisiae* 42-103 (CRC, 2004).

[16.] Alon, U. Simplicity in biology. *Nature* **446**, 497 (2007).

[17.] Mallavarapu, A. Thomson, M. Ullian, B. & Gunawardena, J. Programming with models: modularity and abstraction provide powerful capabilities for systems biology. *J. R. Soc. Interface*, **6**, 257-270 (2009).

[18.] Moriya, H. & Johnston, M. Glucose sensing and signaling in *Saccharomyces cerevisiae* through the Rgt2 glucose sensor and casein kinase I. *Proc. Natl Acad. Sci. USA.* **101**, 1572-1577 (2004).

[19.] Boles, E., Hollenberg, C.P. The molecular genetics of hexose transport in yeasts. *FEMS Microbiol. Rev.* **21**, 85-111 (1997)

[20.] Reifenberger, E., Freidel, K, & Ciriacy, M. Identification of novel *HXT* genes in *Saccharomyces cerevisiae* reveals the impact of individual hexose transporters on glycolytic flux. *Mol. Microbiol.* **16**, 157-167 (1995).

[21.] Bisson, L.F., Coons, D.M., Kruckeberg, A.L., & Lewis, D.A., Yeast sugar transporters. *Crit. Rev. Biochem. Mol. Biol.* **28**, 259-308 (1993).

[22.] Ozcan, S., & Johnston, M., Three different regulatory mechanisms enable yeast hexose transporter (HXT) genes to be induced by different levels of glucose. *Microbiol. Mol. Biol. Rev.* **63**, 554-569 (1999).

[23.] Pao, S.S., Paulsen, I.T., & Saier Jr., M.H., Major facilitator superfamily, *Microbiol. Mol. Biol. Rev.* **62**, 1-34 (1998).

[24.] Wieczorke, R. *et al.* Concurrent knock-out of at least 20 transporter genes is required to block uptake of hexoses in *Saccharomyces cerevisiae*. *FEBS Lett.*, **464**, 123-128 (1999).

[25.] Reifenberger, E., Boles, E.,& Ciriacy, M. Kinetic characterization of individual hexose transporters of *Saccharomyces cerevisiae* reveals the impact of individual hexose transporters on glycolytic flux. *Eur. J. Biochem.* **245**, 324-333 (1997).

[26.] Maier, A., Volker, B., Boles, E., & Fuhrmann, G. F. Characterisation of glucose transport in *Saccharomyces cerevisiae* with plasma membrane vesicles (countertransport) and intact cells (initial uptake) with single Hxt1, Hxt2, Hxt3, Hxt4, Hxt6, Hxt7 or Gal2 transporters. *FEMS Yeast Res.*, **2**, 539-550 (2002).

[27.] Walsh, M C., Scholte, M., Valkier, J., Smits, H.P., & van Dam, K. Glucose sensing and signaling properties in *Saccharomyces cerevisiae* require the presence of at least two members of the glucose transporter family. *J. Bacteriol.* **170**, 2593-2597 (1996).

[28.] Jiang, Y., Davis, C., & Broach, J., Efficient transition to growth on fermentable carbon sources in *Saccharomyces cerevisiae* requires signaling through the Ras pathway. *EMBO J.* **17**, 6942-6951 (1998).

[29.] Boer, V.M., Amini, S., Botstein, D., Influence of genotype and nutrition on survival and metabolism of starving yeast. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6930-6935 (2008).

[30.] van Hoek, P., van Dijken, J., & Pronk, J. Effects of specific growth rate on fermentative capacity of baker's yeast. *App. Env. Microbiol.* **64**, 4226-4233 (1998).

[31.] Reijenga, K.A. *et al.* Control of glycolytic dynamics by hexose transport in *Saccharomyces cerevisiae*. *Biophys. J.* **80**, 626-634 (2001).

[32.] Kaniak, A., Xue, Z., Macool, D., Kim, J.H., & Johnston, M. Regulatory network connecting two glucose signal transduction pathways in *Saccharomyces cerevisiae*. *Eukaryotic Cell*. **3**, 221-231 (2004).

[33.] Gancedo, J.M. The early steps of glucose signaling in yeast. *FEMS Microbio. Rev.* **32**, 673-704 (2008).

[34.] Kim, J.H., & Johnston, M. Two glucose-sensing pathways converge on Rgt1 to regulate expression of glucose transporter genes in *Saccharomyces cerevsiae*. *J. Biol. Chem.* **281**, 26144-26149 (2006).

[35.] Santangelo, G.M. Glucose signaling in *Saccharomyces cerevisiae*. *Micro. Mol. Biol. Rev.* **70**, 253-282 (2006).

[36.] Levy, S. *et al.* Strategy of transcription regulation in the budding yeast. *PLoS One*. **2**, e250:1-10 (2007)

[37.] Yin, Z. *et al.* Glucose triggers different global responses in yeast, depending on the strength of the signal, and transiently stabilizes ribosomal protein mRNAs. *Mol. Microbiol.* **48**, 713-724 (2003).

[38.] Stephanopoulos, G. Challenges in engineering microbes for biofuels production. *Science* **315**, 801 (2007).

[39.] Lorenz, D.R., Cantor, C.R., & Collins, J.J. A network biology approach to aging in yeast. *Proc. Natl Acad. Sci. USA.* **106**, 1145-1150 (2009).

[40.] Ostergaard, S., Olsson, L., & Nielsen, J., Metabolic engineering of *Saccharomyces cerevisiae*. *Micro. Mol. Bio. Rev.* **64**, 34-50 (2000).

[41.] Kell, D. B., Metabolomics and systems biology: making sense of the soup. *Curr. Opin. Microbiol.* **7**, 296-307 (2004).

[42.] Savageau, M.A., Coelho, P., Fasani, R., Tolla, D., & Salvador, A. Phenotypes and tolerances in the design space of biochemical systems, *Proc. Natl Acad. Sci. USA*. **106**, 6435-6440. (2009).

[43.] Ihmels, J. *et al.* Rewiring of the yeast transcriptional network through the evolution of motif usage. *Science* **309**, 938-940. (2005)

[44.] Klumpp, S., & Hwa, T. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc. Natl. Acad. Sci. USA*. **105**, 20245-20250 (2008)

[45.] Duarte, N.C., Palsson, B.O., & Fu, P. Integrated analysis of metabolic phenotypes in *Saccharomyces cerevisiae*. *BMC Genomics* **5**, 1471-2164 (2004).

[46.] Daran-Lapujade, P. *et al.* The fluxes through glycolytic enzymes in *Saccharomyces cereivisiae* are predominantly regulated at posttranscriptional levels. *Proc. Natl. Acad. Sci. USA*. **104**, 15753-15758 (2007).

[47.] Castrillo, J.I. *et al.* Growth control of the eukaryote cell: a systems biology study in yeast. *J. Biol.* **6**:4, 1-25 (2007).

[48.] Stelling, J. Mathematical models in microbial systems biology. *Curr. Opin. Microbiol.* **7**, 513-518 (2004).

[49.] Sheff, M. & Thorn, K. Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast* **21**, 661-670 (2004).

[50.] Kim, J.H., Polish, J., & Johnston, M. Specificity and regulation of DNA binding by the yeast glucose transporter gene repressor Rgt1. *Mol. Cell. Biol.* **23**, 5208-5216 (2003).

[51.] Wieczorke, R. *et al.* Concurrent knock-out of at least 20 transporter genes is required to block uptake of heoxes in *Saccharomyces* cerevisiae. *FEBS Lett*. **464**, 123-128 (1999).

[52.] Sheff, M. & Thorn, K. Optimized cassettes for fluorescent protein tagging in

Saccharomyces cerevisiae. Yeast, **21**, 661-670 (2004).

[53.] Kim, J.H., Polish J., & Johnston, M. Specificity and regulation of DNA binding by the yeast glucose transporter gene repressor Rgt1. *Mol. Cell. Biol*. **23**, 5208-5216 (2003).

[54.] Maier, A., Volker, B., Boles, E., & Fuhrmann, G.F., Characterisation of glucose transport in *Saccharomyces cerevisiae* with plasma membrane vesicles (countertransport) and intact cells (initial uptake) with single Hxt1, Hxt2, Hxt3, Hxt4, Hxt6, Hxt7 or Gal2 transporters. *FEMS Yeast Research* **2**, 539-550 (2002).

[55.] Ozcan, S. & Johnston, M. Three different regulatory mechanisms enable yeast hexose transporter (HXT) genes to be induced by different levels of glucose. *Mol. Cell. Biol.* **15**, 1564-1572 (1995).

[56.] Krampe, S., Stamm, O., Hollenberg, C.P., & Boles, E. Catabolite inactivation of the high-affinity hexose transporters Hxt6 and Hxt7 of *Saccharomyces cerevisiae* occurs in the vacuole after internalization by endocytosis. *FEBS Letters*, **441**, 343-347 (1998).

[57.] Yin, Z., Hatton, L., & Brown, A.J.P. Differential post-transcriptional regulation of yeast mRNAs in response to high and low glucose concentrations. *Molec. Microbiol.* **35**, 553-565 (2000).

4. Embryonic stem cells: A brief overview

Chapter summary

In this chapter, we introduce some basic facts about mouse embryonic stem cells. We outline historical developments, common experimental techniques, and important discoveries. This chapter provides the necessary background to understand our experiments in chapter 5.

4.1. Historical overview

In the early 1960s, groundbreaking experiments by biophysicists James Till and Ernest McCulloch, then working at the University of Toronto, established the existence of stem cells for the first time. At the time, they were studying leukemia and the formation of normal blood cells. By using a novel quantitative method for identifying monoclonal population of cells, the two researchers injected bone marrow cells into irradiated mice. A small swelling due to an aggregate of cells, called 'nodules', formed in the spleens of the mice as a result of their injection [1, 2]. They showed that the number of nodules formed was proportional to the number of bone marrow cells that were injected into the mouse. They called these nodules 'spleen colonies' and hypothesized that each nodule arose from a single marrow cell. They hypothesized that a single marrow cell could self-renew and differentiate into different cell types, giving rise to the nodules. Together with their graduate student Andy Becker, they demonstrated that each nodule indeed arose from a single cell [3]. Later work with Lou Siminovitch showed that these cells could self-renew [4]. These works demonstrated that a stem cell (called 'mesenchymal stem cell') exists within the bone marrow, as self-renewal and multipotency are the hallmarks of stem cell. Specifically, Till and McCulloch had found a tissue-specific, adult stem cell. These can form either one (unipotent adult stem cell) or a few specific types of cells (multipotent adult stem cell) in the body.

Embryonic stem cells are pluripotent: They can differentiate into almost all the major cell types that make up the body. An early (< 5 days for mice and humans) embryo exists as a small (about 0.1 mm in diameter) sac of fluid and cells (Figure 4.1). A layer of cells, known as trophectoderm, shields the inner cell mass (ICM) and fluid inside the blastocyst.

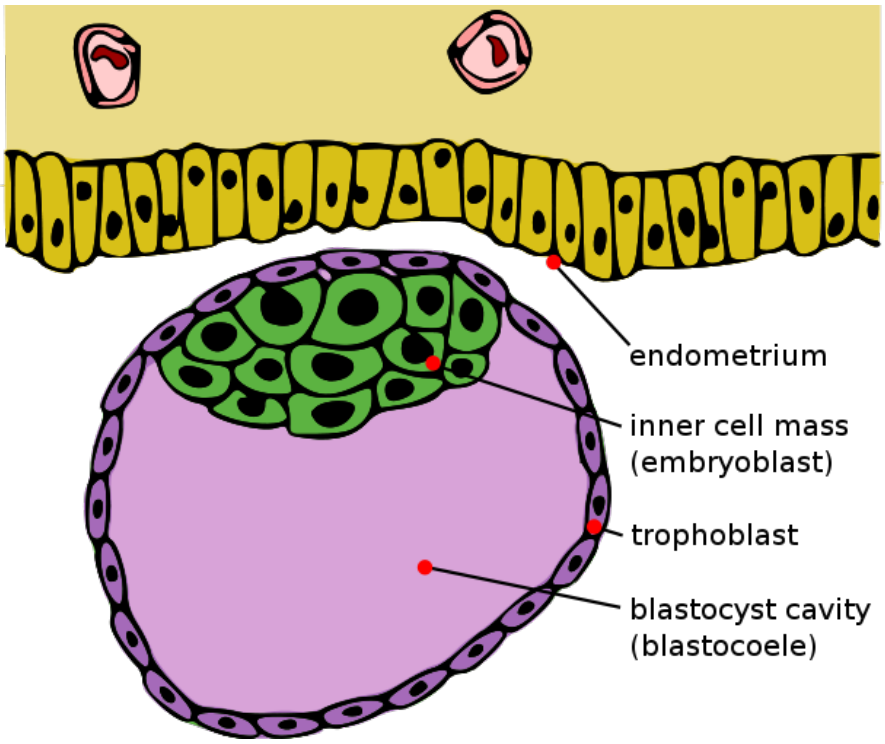


Figure 4.1: Blastocyst during early mammalian embryogenesis. The inner cell mass can be extracted from the mouse blastocyst and cultured as self-renewing, pluripotent cells in a petri dish. These cells are known as embryonic stem cells. (From Wikimedia commons).

The trophectoderm will eventually form the placenta. The ICM, when extracted from the blastocyst, can be cultured in petri dish as self-renewing, pluripotent cells known as embryonic stem cells. Martin Evans, Matthew Kaufman, and Gail Martin have discovered such a method in 1981 [5,6]. They succeeded in isolating embryonic stem cells from mouse blastocysts and growing them in cell culture. One of the first applications of embryonic stem cells cultured this way was in making 'knockout' mice. In these mice, specific gene has been knocked out of the mouse genome. This is typically done by mutating or excising a gene of interest in the embryonic stem cell, then injecting the cell into a mouse blastocyst. This blastocyst eventually gives rise to a mouse with the targeted gene knocked out. These genetically altered mice have since been widely used as models for studying progression of wide range of human diseases such as cancer, diabetes, cystic fibrosis, and arthritis. Human embryonic stem cells have not been

isolated until 1998, when the groups of James Thomson and John Gearhart independently discovered a method for isolating the ESCs from human embryos [7,8]. Although there are differences between human and mouse embryonic stem cells, they share many primary features. This allows researchers to use mouse embryonic stem cells to draw meaningful conclusions about human embryonic stem cells, which are more difficult to work with for technical and ethical reasons.

4.2. Conventional experimental toolkits for studying pluripotency transcriptional network

An embryonic stem cell is a precursor cell that can grow into every kind of mammalian tissue; including muscle, bone, and brain. At the moment, only mice and human embryonic stem cells can be isolated and studied. It is difficult to pinpoint exactly what network of proteins are responsible for maintaining a stem cell in this uncommitted state. Genome wide studies have revealed the immense complexity of the transcriptional state of the pluripotent stem cell. Below, we detail current genome-wide experimental techniques and what they have found. We will then explain what single-cell level studies of transcripts can tell us that such genome scale approaches may overlook.

First we introduce some basic experimental tools that are used for measuring the genome wide transcriptional state of a population of cells.

(i). DNA Microarray

DNA microarray is a device that measures the average expression level of a large number of genes. In this device, a population of cells is first lysed, from which all the mRNA molecules within the population of cells are gathered. Each mRNA is then reverse transcribed into its cDNA (complementary DNA), which would exclude all the introns since the mRNA results after splicing). This cDNA is then introduced to a small well, at the bottom of which is a small (about one picomole) amount of fluorescently tagged ‘probe’ DNA that would complementarily bind the cDNA of interest. The intensity of florescence in the well is proportional to the number of cDNA that is bound to the probe DNA in the well. This florescence thus quantifies, on average, how much abundant the mRNA of a particular gene within the population of cells. Nowadays, DNA microarray ‘chip’ can have many wells, enough to measure expression levels of the entire

genome of some organisms (~6,000 genes for the budding yeast), and is a routine tool for researchers.

(ii.) Chromatin ImmunoPrecipitation (ChIP)

Chromatin ImmunoPrecipitation (ChIP) allows researchers to identify proteins that interact with a particular sequence of DNA within a cell such as transcription factors. In this procedure, one first crosslinks all the polypeptides within a cell by using a fixative like formaldehyde. This is analogous to hard-boiling an egg, in which the inside contents of the egg are preserved, with its yolk fixed in its place. The fixed cells are then lysed and only the chromatins and any proteins that are bound to any region of the chromatin are extracted from this lysate. These proteins come down along with the chromatin due to the fixation step mentioned above. By sonication, the chromatin is fragmented into smaller pieces with the proteins still bound to them. To check which region of DNA a specific protein of interest is bound, antibody specific to that protein is conjugated to the protein. An additional purification step purifies only those DNA sequences that are bound to the antibody-protein complex for analysis.

(iii.) ChIP-chip

ChIP-chip (also known as ChIP-on-chip) combines ChIP with the DNA microarray technology. This is a high-throughput version ChIP-Seq, in which all putative segments of DNA within the cell's genome can be obtained by ChIP, then sequenced using a whole-genome microarray chip. Since DNA microarray can quantify the relative amounts of a particular DNA sequence, it allows gives some quantitative information about frequency with which a particular protein binds to the sequence within a population of cells.

(iv.) ChIP-Sequencing (ChIP-Seq)

Combining ChIP with high-throughput nucleotide sequencing gives rise ot ChIP. By sequencing the pieces of DNA that is pulled down with the antibody-protein complex at the end of the ChIP procedure, one can identify regions of DNA that a given protein has a measurable affinity for binding. This technique thus helps with unraveling a putative transcriptional network through identifying possible transcription factors and their binding locations in the genome. This method, however, does not discriminate between proteins that are bound to a promoter region

while the gene controlled by that promoter is transcriptionally inactive, from those proteins that are actually bound to the promoter during the gene is transcriptionally active.

4.3. Current understanding of the pluripotency transcriptional network

If there is one finding about pluripotency transcriptional network that almost all researchers will agree on, it is that the network involves many genes, potential transcription factors, and intricate chromatin states that we do not yet fully understand. Aside from the shear complexity of the transcriptional network that maintains the stem cell's pluripotency, many conflicting reports and round-about experiments have caused much confusion and disagreements among the researchers in the field. Even the 'pluripotency transcriptional network' is not well defined due to these complications. A simplified network of important transcription factors still contains a fairly large number of unknown connections and genes (Figure 4.2) [9]. In almost all the models of transcription networks that researchers propose, Oct4, Sox2, and Nanog form a core 'trinity' of regulators. For this reason, our work in chapter 5 will study these three players in detail.

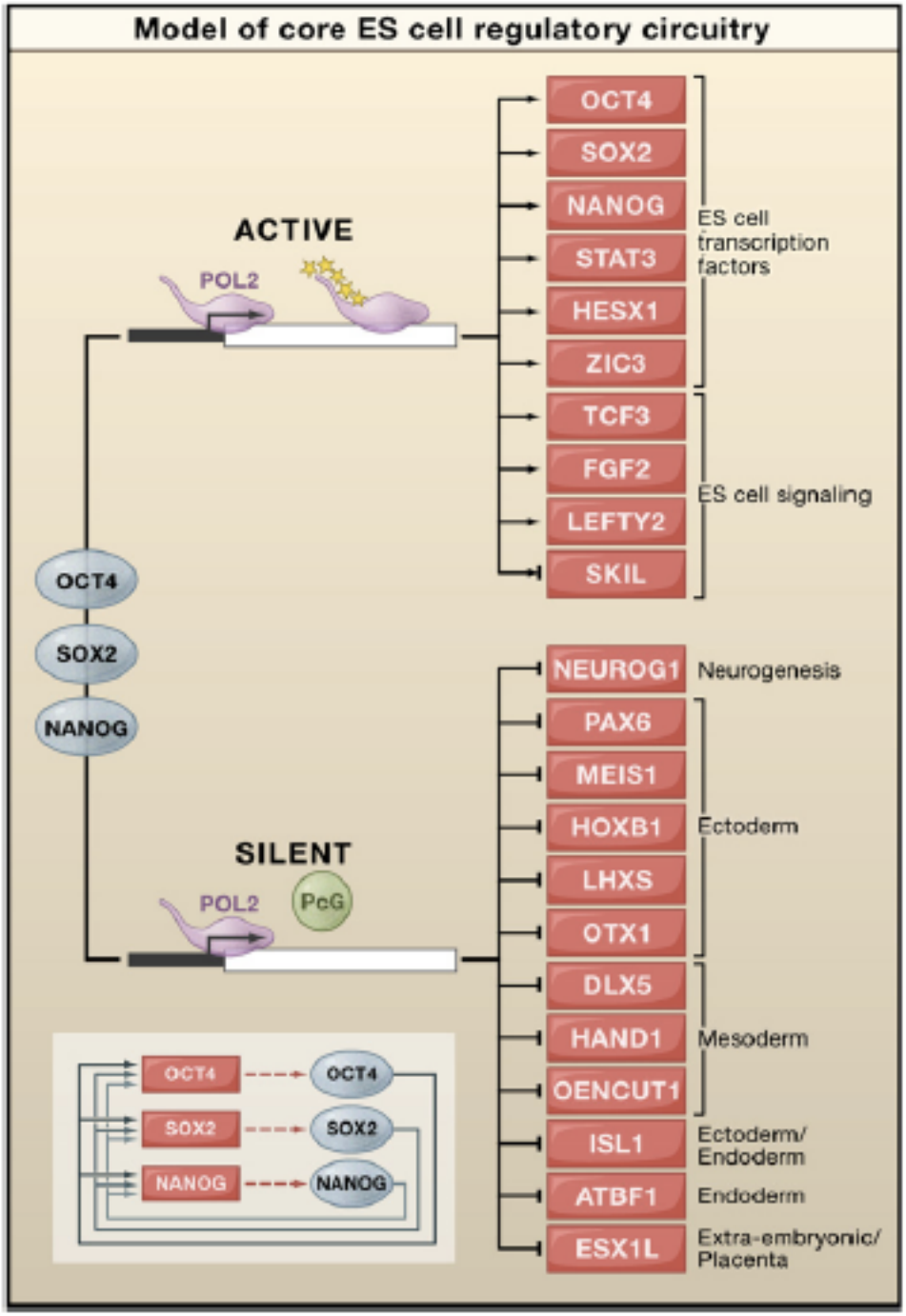


Figure 4.2.: One model of a 'core' embryonic stem cell regulatory circuitry. Depending on the choice of literature, other players may be added or removed from the list of genes shown here. But many agree that Oct4, Sox2, and Nanog form a core 'trinity' of regulators. (Adapted from Jaenisch and Young, *Cell* **132**, 567-582 (2008) [9]).

Primary transcription factors maintain pluripotency

A growing list of studies now pinpoint to three nuclear 'master' regulators, Oct4, Sox2, and Nanog as essential for maintaining pluripotency in vivo and in vitro [10]. From ChIP-Seq studies, researchers found that these three proteins show binding affinity for a large set of promoter regions, potentially affecting the their transcription. And since many of their target genes code for transcription factors that may also bind to the promoters of Oct4, Nanog, and Sox2, researchers believe that a complex transcriptional network governed by many transcription factors are responsible for maintaining the pluripotent state of an embryonic stem cell. In addition, Oct4, Nanog, and Sox2 each binds to its own promoter, thus adding complexity to an already intricate network.

Role of Nanog in embryonic stem cells:

Recent single-cell level studies of embryonic stem cells, particularly by the group led by Austin Smith, have shown that expression level of Nanog varies highly from cell-to-cell within a population of cells growing in the same petri dish. A stem cell can persist in culture when both copies of Nanog alleles are deleted from the cell's genome. This cell, however, shows a markedly reduced self-renewal efficiency and an increased likelihood to differentiate into primitive endoderm-like cells [11]. By injecting Nanog -/- cells are injected into the early mouse embryo, researchers found that primordial germ cells failed to mature. The role of Nanog is now believed to be primarily in constructing the inner cell mass and germ cells. At the other extreme, researchers found that overexpressing Nanog results in inhibition of differentiation into certain cell types. These findings, together establish that Nanog is indispensible in the mouse embryonic development and is important in maintenance of pluripotency of embryonic stem cells. Crucially, these findings indicate that a finely maintained expression level of Nanog may be important in a stem cell.

Role of Oct4 in embryonic stem cells:

Oct4 is an oncogene. This means that its overexpression aids in turning a normal cell into a tumor cell. This means that its overexpression may potentially interrupt stem cells function. At the same time, its expression level is necessary for the stem cell to remain pluripotent. This trade-off is a theme that we will come back to in the next chapter.

An ectopic overexpression of Oct4 in a mouse embryo inhibits proper differentiation into epithelial tissues [12]. By overexpressing Oct4 in somatic tissue cells in adult mice, researchers showed that abnormal, dysplastic growth results in the affected tissues, reminiscent of tumorigenesis. During differentiation into endoderm [12] and neurectoderm of the embryo, Oct4 expression is gradually shut down. Researchers have also shown that a loss-of-function mutation for Oct4 results in early embryonic lethality due to inappropriate differentiation of pluripotent epiblast cells into trophectoderm [12]. Turning down the level of Oct4 expression in mouse embryonic stem cells results in in trophectoderm differentiation, where as overexpression induces differentiation into exraembryonic mesoderm and endoderm. Thus varying the dosage of Oct4 expression corresponds to varying the cell fate of the embryonic stem cell. After Oct4 is turned off, by turning it on in certain somatic cells, researchers have seen dedifferentiation of the somatic cell. A two-fold increase in Oct4 expression level causes embryonic stem cells to differentiate into endoderm and mesoderm, whereas knockdown of Oct4 causes the embryonic stem cells to enter into trophectoderm-like cells [13].

Role of Sox2 in embryonic stem cells:

Accumulating evidence now pinpoint Sox2 to be a master regulator that controls the expression levels of developmentally important genes such as Oct4, Nanog, nestin, and FGF4. Sox2 also binds to its own promoter region, suggesting that it autoregulates itself along with other transcription factors that bind there. These studies indicate that perturbing Sox2 expression level would result in potentially a large number of genes being affected. A loss-of-function mutation in Sox2 results in loss of pluripotency in embryonic stem cells. A recent study showed that ectopically overexpressing Sox2 gene results in differentiation of embryonic stem cells into cells that display markers for a variety of different cell types (neuroectoderm, mesoderm, and trophectoderm) but not including those associated with endoderm. Interestingly, a recent work showed that ectopically overexpressing Sox2, in a related type of cell known as embryonic carcinoma cells, results in downregulation of the endogenous Sox2 expression level. Since both embryonic carcinoma cell (F9 EC) and embryonic stem cells share the same Sox2:Oct3-4 target genes and their expression levels turn off when both differentiate, this result may be apply in the embryonic stem cells as well [14]. We will show how to test this prediction in chapter 5. As with

Nanog and Oct4, these results together indicate that Sox2 expression level may need to remain within a certain range in order for the stem cell to remain pluripotent.

4.4. A need for a phenomenological understanding of pluripotency circuitry

There is a mounting list of novel mechanisms and more detailed understanding of the already known molecular events that are important for transcriptional maintenance of pluripotency. But an alternative approach, one in which less attention is paid to the exact molecular details and pays attention to a few parameters that capture the essence of the system instead, may improve our understanding. In addition, a non-invasive technique for measuring the transcriptional state of the cell, one that does not require genetic manipulation or lysing of many cells, may be important for studying stem cells given the intricacies that we know are lost during invasive techniques such as fluorescent protein fusion and cell lysis. Chapter 5 describes our early work towards reaching such a phenomenological understanding of the pluripotency transcriptional circuitry using a non-invasive technique.s

Bibliography for Chapter 4

[1.] McCulloch, E. A., and Till, J. E. The radiation sensitivity of normal mouse bone marrow cells, determined by quantitative marrow transplantation into irradiated mice. *Radiation Research* **13**, 115-125 (1960).

[2.] Till, J. E., and McCulloch, E. A. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiation Research* **14**, 213-222 (1961).

[3.] Becker, A. J., McCulloch, E. A., and Till, J. E., Cytological demonstration of the clonal nature of spleen colonies derived from transplanted mouse marrow cells. *Nature* **197**, 452-454 (1963).

[4.] Till, J. E., McCulloch, E. A., and Siminovitch, L. A stochastic model of stem cell proliferation, based on the growth of spleen colony forming cells. *Proc Natl Acad Sci.* **51**, 29-36 (1964).

[5.] Evans, M., and Kaufman, M. Establishment in culture of pluripotent cells from mouse embryos. *Nature* **292**, 154-156 (1981).

[6.] Martin, G. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci* **78**, 7634-7638 (1981).

[7.] Thomson, J. A. *et al.* Embryonic stem cell lines derived from human blastocysts. *Science* **282**, 1145-1147 (1998).

[8.] Shamblott, M. J. *et al.* Derivation of pluripotent stem cells from cultured human primordial germ cells. *Proc Natl Acad Sci* **95**, 13726-13731 (1998).

[9.] Jaenisch, R., and Young R., Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567-582 (2008).

[10.] Niwa, H. How is pluripotency determined and maintained? *Development* **134**, 635-646 (2007).

[11.] Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230 -1234 (2007).

[12.] Hochedlinger, K., Yamada, Y., Beard, C., and Jaenisch, R. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* **121** 465-477 (2005).

[13.] Niwa, H., Miyazaki, J.-I., and Smith A. G., Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genetics* **24**, 372-376

(2000).

[14.] Boer, B. *et. al.* Elevating the levels of Sox2 in embryonal carcinoma cells and embryonic stem cells inhibits the expression of Sox2:Oct-3/4 target genes. *Nucleic Acids Research* **35**, 1773-1786 (2007).

5. Towards a phenomenological model of transcriptional maintenance of pluripotency

Chapter summary

A major conceptual question in developmental biology is how an embryonic stem cell maintains a pluripotent state that affords the stem cell the flexibility to turn into any major cell type in the body. We may obtain insights into this question by systematically perturbing the key players of the transcriptional network that is responsible for maintaining pluripotency. Using a nascent technique for measuring integer counts of individual mRNA molecules within single cells *in situ*, we can examine cell-to-cell variability in transcriptional response to such perturbations in mouse embryonic stem cells. This method also allows for measurement of the endogenous transcript levels without the invasive genetic engineering of the embryonic cells. This chapter describes our early work towards understanding such transcriptional responses in single embryonic stem cells. In particular, we show that increasing the expression level of Nanog can have an intricate transcriptional response by Oct4, Sox2 and Nanog itself at the single-cell level. In particular, we demonstrate that if Nanog expression level is increased above a certain threshold, as yet unknown mechanism is activated in which the level of Nanog expression level is quickly shut down. When the Nanog expression level is increased above the endogenous level but below the threshold, both Oct4 and Sox2 transcript levels increase in proportion to Nanog. We hypothesize possible mechanisms to explain these observations and suggest future experiments to test them.

5.1. Introduction

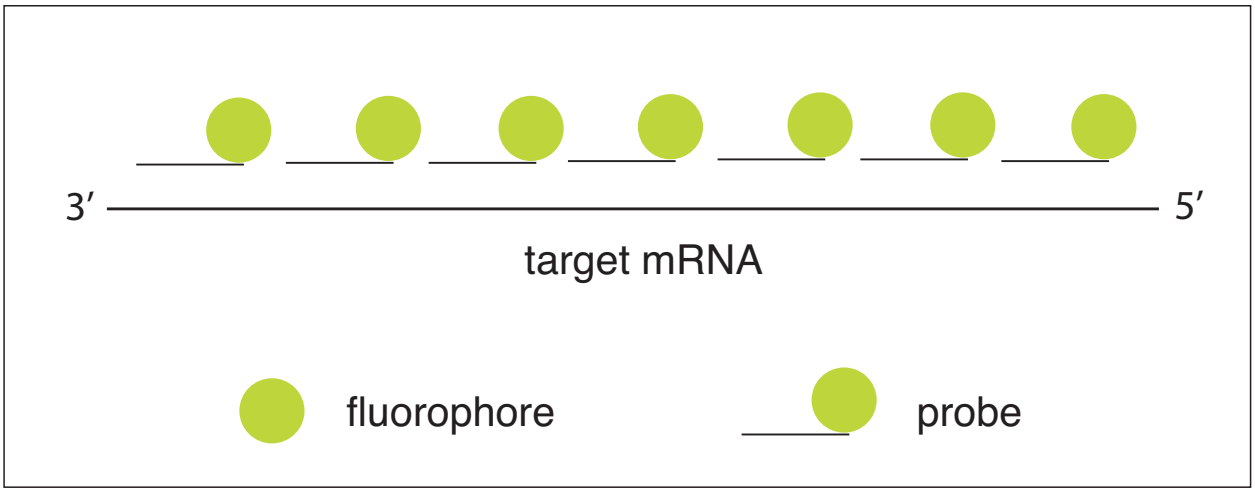
Pluripotency, the ability of a cell to turn into any of the cell types that make up the organism, is one of the hallmarks of embryonic stem cells. A major conceptual question is how the complex network of transcriptional circuits maintains the pluripotent state that affords the stem cell the flexibility to turn into any major cell type in the body. To acquire a better understanding of this question, we can systematically perturb the levels of some key proteins in the transcriptional circuitry involved in pluripotency maintenance. We can then measure the cell's transcriptional response by measuring the integer counts of key transcripts, in single mouse embryonic stem cells. To do this, we use a nascent technique for directly visualizing individual mRNA molecules *in stiu* within single cells, by fluorescence microscopy (known as "RNA FISH"). In doing so, we may discover interesting patterns of cell-to-cell variability in RNA levels of key genes in the pluripotency network (notably Nanog, Sox2, and Oct4). Furthermore, we can reveal relationships between the transcriptional response of each key gene to perturbations of the others and the role of cell-to-cell variability in this response. Our goal is to develop a quantitative model that will explain these observations. Below, we describe our early work towards achieving this goal.

5.2. RNA FISH

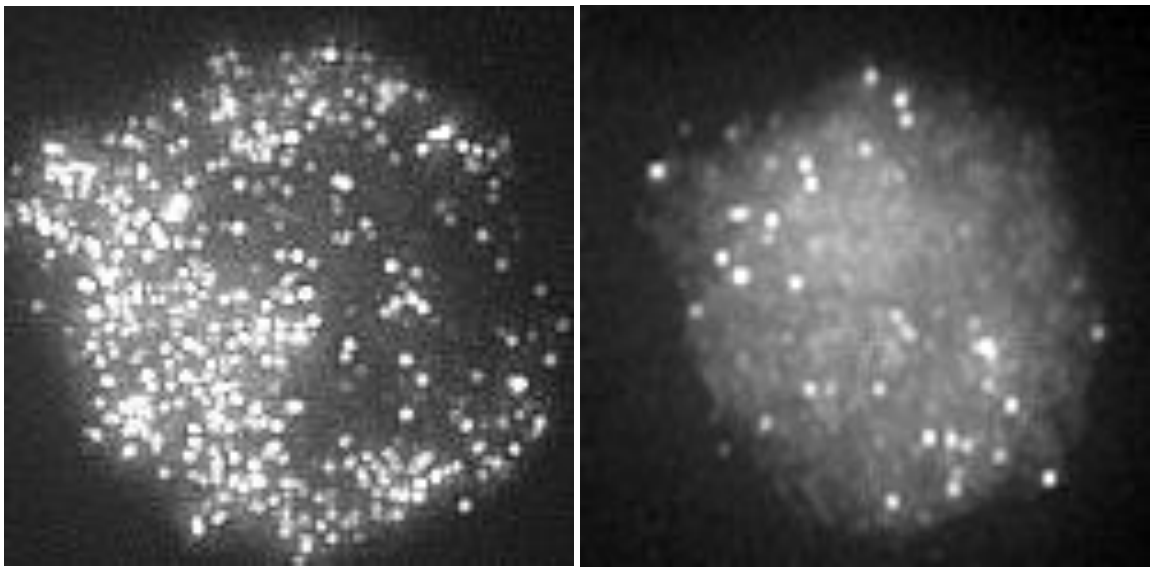
Fluorescence In Situ Hybridization (FISH) is a method that has allowed researchers to look at both the endogenous mRNA (RNA FISH) and DNA (DNA FISH) (Gall, 1968; Levsky and Singer 2003). FISH has been applied to a wide range of organisms, from yeast to human cells. It first requires the cells to be fixed, in which all the polypeptides within the cell are crosslinked using a fixative such as formaldehyde. A good analogy to fixing a cell is boiling an egg, in which the inside contents of the egg are spatially preserved. We then design fluorescently labeled single-stranded oligonucleotide sequence that will complementarily bind to the single-stranded RNA or DNA of interest. By fluorescence microscopy, one can then look at the spatial localization and total counts of the fluorescence. DNA FISH has been widely used for looking at chromosomal abnormalities, species identification, and chromosomal substructures such as looping between two loci. RNA FISH has been used for observing spatial localization of mRNA within single cells, and quantifying the absolute counts of mRNA molecules.

The particular version of RNA FISH we use in this chapter is a modification of the RNA FISH method described by Robert Singer and his coworkers (Femino et al, 1998). In this method, we use a large collection (at least 30) of single-stranded oligonucleotides, each labeled with a single fluorophore, that binds along the length of the target mRNA (Figure 5.1a). The binding of so many fluorophores to a single mRNA results in a signal that is bright enough to be detectable with an epi-fluorescence microscope as a diffraction-limited spot.

We have designed RNA FISH probes that target Nanog, Sox2, and Oct4 endogenous mRNA molecules. Each transcript has forty-eight probes, each of which are attached to a single fluorophore molecule. We used three different fluorophores (Cy5, Alexa594, and TMR) whose excitation and emission spectral peaks do not overlap, to label the probes for each gene with a different color. We could thus observe the three different kinds of mRNA molecules (Nanog, Sox2, and Oct4) within the same single cell and distinguish them from each other (Figures 5.1b and 5.1c). After computationally identifying all the diffraction-limited spots in a field of view under a microscope, we determine which cells each spot belongs to by identifying cell boundaries by hand (Figure 5.2). We could thus obtain the absolute counts of mRNA molecules in individual stem cells.



(a)



(b) (c)

Figure 5.1: RNA FISH applied to mouse embryonic stem cells. (a). Each mRNA molecule is targeted by 30 or more probes. Each probe is attached to a single fluorophore molecule and consists of about 20 oligonucleotide bases. These 30 or more probes then bind along the stretch of the single target mRNA molecule. The bindingof so many fluorophores to a single mRNA results in a signal that is bright enough to be seen as a diffraction-limited spot using an epi-fluorescence microscope. **(b).** Oct4 mRNA molecules are seen as diffraction limited spots within a single mouse embryonic stem cell. **(c).** Nanog mRNA molecules are seen as diffraction limited spots within the same single cell seen in (b). Having Nanog and Oct4 probes labeled with two different fluorophores that do not cross-talk, we can distinguish mRNA molecules of Nanog from those of Oct4.

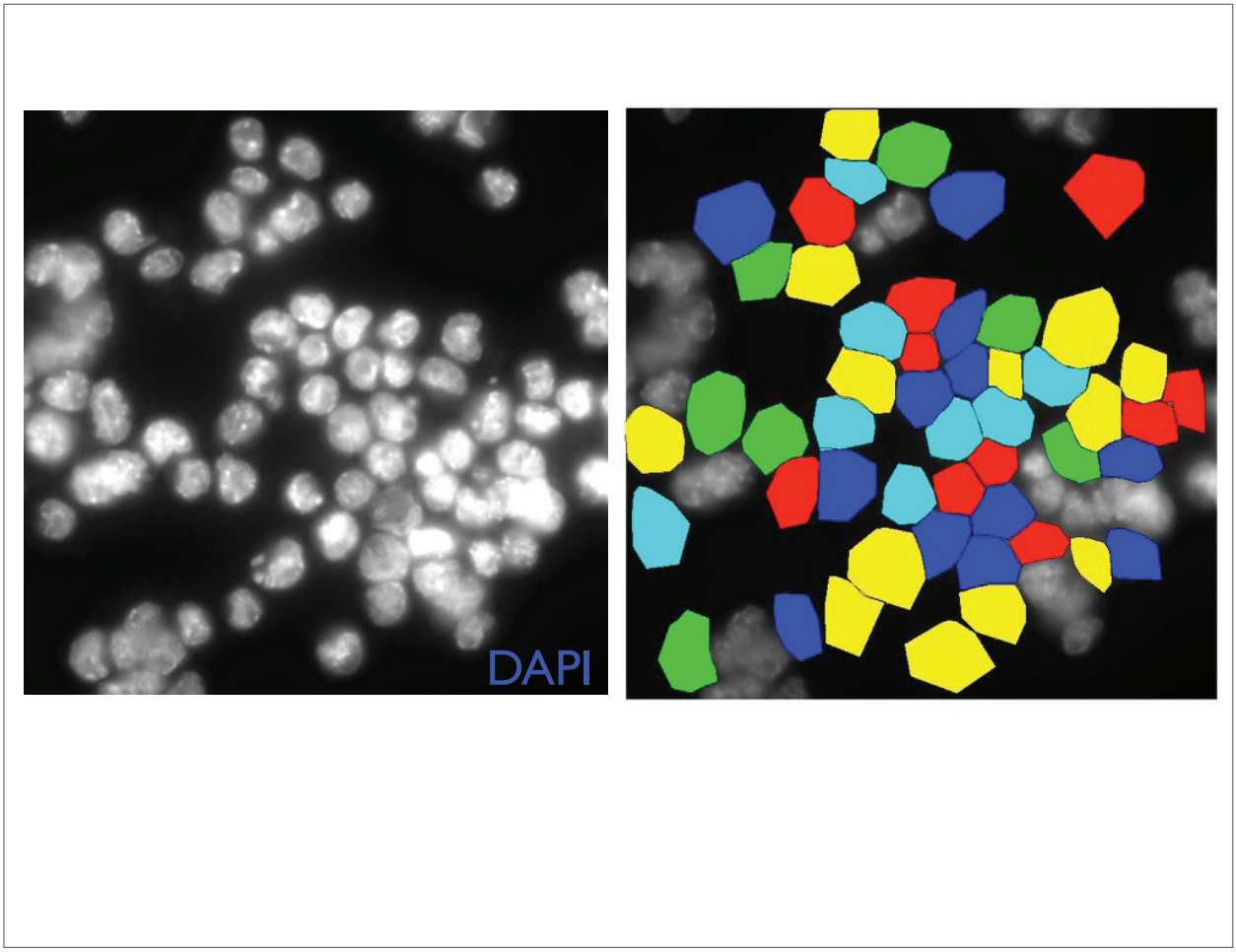


Figure 5.2: Cell segmentation by hand to identify each cell within a population of embryonic stem cells. (Left) DAPI staining of a cell's nucleus helps with identifying individual cells in the field of view. (Right) By hand, each cell's hypothetical boundary is drawn.

5.3. Single-cell distrbutions of Nanog, Oct4, and Sox2 transcript counts

By measuring the integer counts of the Nanog, Sox2 and Oct4 mRNA molecules within the same single cell, and doing this for every cell making up the population of the wild-type (V6.5) mouse embryonic stem cells, we obtained a single-cell histogram of each of the three types of transcripts (Figure 5.3). Nanog transcript is nearly exponentially distributed, with most of the cells within a population (~ 60%) having less than fifty counts. Oct4 transcript level could be fit well by a Gaussian distribution, with mean of about 250 transcripts per cell. Sox2 transcript distrubiton can be fit with a gamma distribution (in fact, so can Nanog and Oct4 since exponential and Gaussian distributions are special cases or limits of gamma distribution). Each of the three transcripts thus show their characteristic distributions that are distinct from each other. The fact that we can fit these single-cell transcript distributions is consistent with models for transcriptional bursting (Raj et al, 2006; Friedman et al. 2006). Although a detailed study that measures various rates relevant for transcriptional dynamics is an interesting study on its own, our focus is on coming up with simpler models that capture the relationships between these three transcripts that does not involve the level of molecular details that transcriptional bursting models invoke. Since we counted the number of Nanog, Sox2 and Oct4 transcripts all within the same single cell, we could look for any possible correlations between any two transcript levels. We found that given the cell-to-cell heterogeneity in the levels of all three transcripts, linear regression fits provided a poor description of the pluripotent transcriptional state of any one cell.

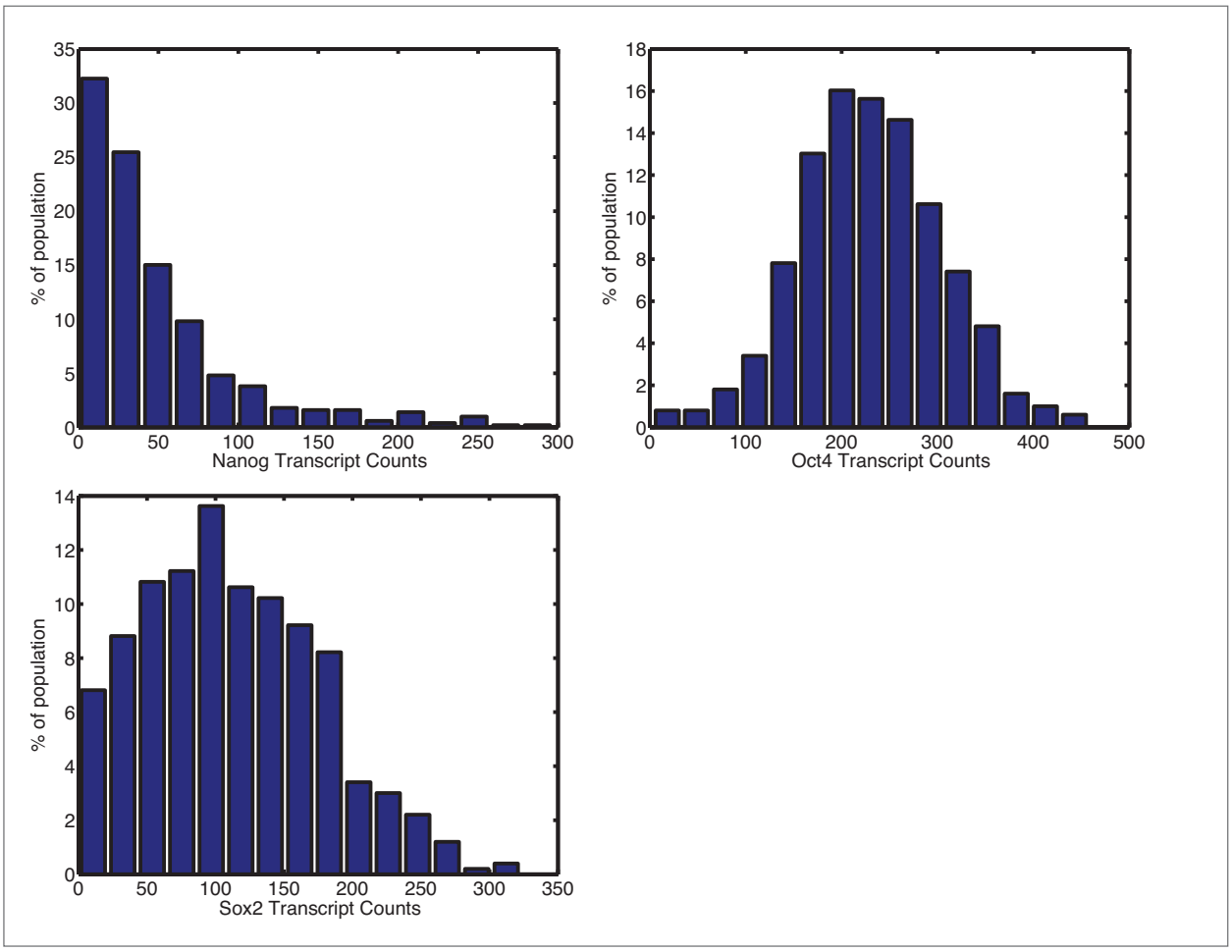


Figure 5.3: Heterogeneities in expression levels of Nanog, Sox2 and Oct4 in the wild-type embryonic stem cells (V6.5) are captured by measuring single mRNA molecules within single cells. We measured transcripts in 500 stem cells for these histograms.

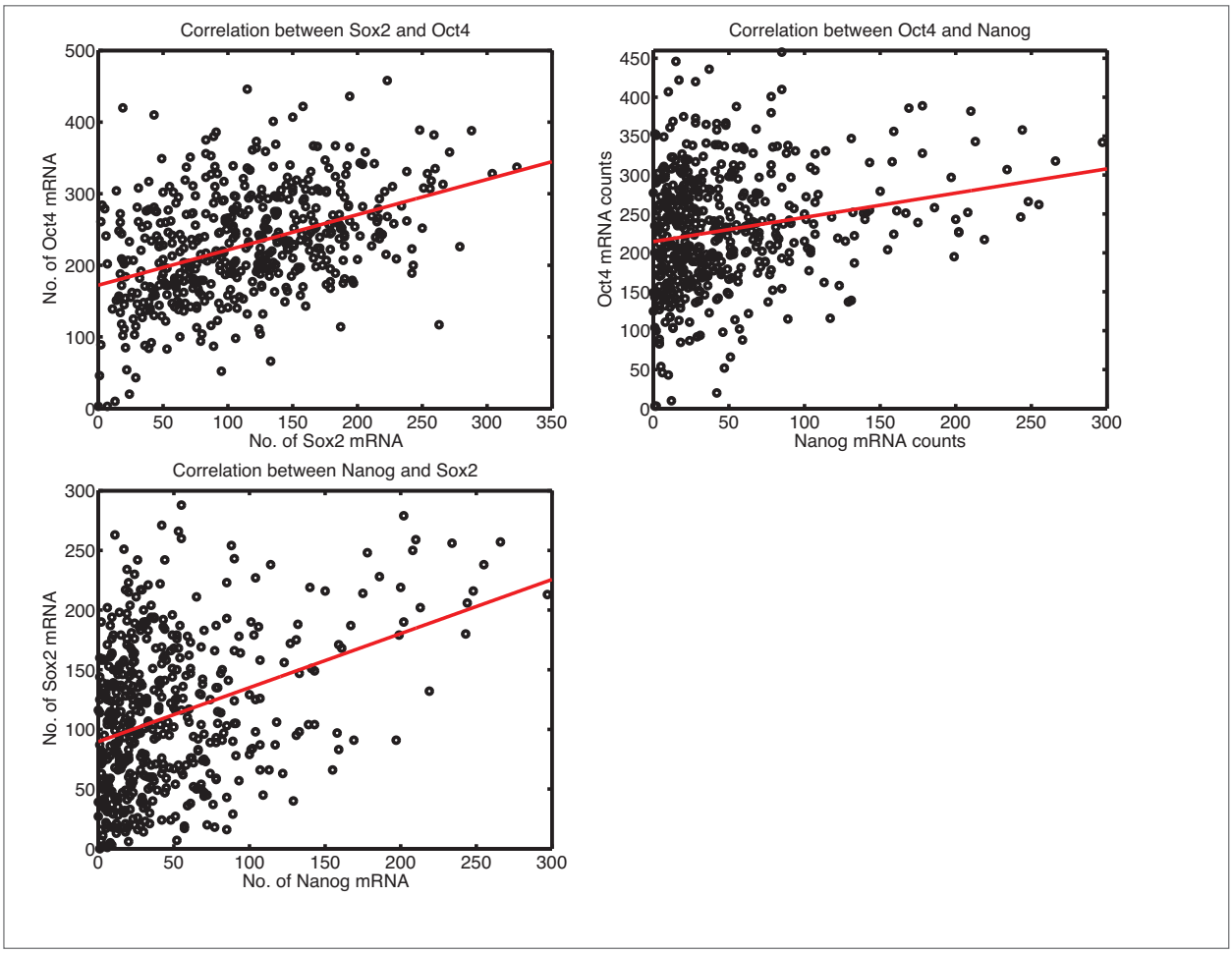


Figure 5.4. **Correlations between Nanog, Oct4, and Sox2 transcript levels in wild-type (V6.5) embryonic stem cells.** Red line is a linear regression fit. Given the cell-to-cell heterogeneity in all three transcript levels, linear regression fits clearly provide a poor description of the pluripotent transcriptional state of any one cell.

5.4. Perturbing pluripotency transcriptional network: Overexpression of Nanog

Perturbing the key parameters of a transcriptional network can reveal novel connections between genes within the network as well as the underlying design principles. In the case of the transcriptional network responsible for pluripotency maintenance, many of the genes that make up the network are known but not much is known about their connections. For example, while some studies have shown the effects of over and underexpressing Nanog on the change in the average expression level of other genes, our measurement of the wild-type's transcript distributions show that the average gene expression level does not tell the whole story. For example, consider all the cells within a population that have less than 50 Nanog transcripts. Some of these cells will have very high Oct4 transcript level (around 300 Oct4 mRNA), while some will have nearly half that many Oct4 mRNA (around 150 Oct4 mRNA) (Figure 5.4). This type of variability is completely missed by population level studies of transcription network perturbations.

Motivated by our observation that a correlation plot involving Nanog shows interesting, albeit a bit subtle, features (Figure 5.4), we began our network perturbation study with overexpresion of Nanog. A cell line has been constructed by Rudolf Jaenisch's lab, which allows for an ectopic overexpression of Nanog gene. In the V6.5 cell line, researchers have homologously recombined the FRT-hygro-polyA cassette downstream of the ColA1 (Collagen) locus. When the Flpase enzyme is present, the FRT sites recombine to add the gene in the targeting vector into where the FRT site is in the genome. These cells responded to an increase in the concentration of the inducer doxycycline by increasing their Nanog expression level. We cultured this inducible-nanog cell line in various concentrations of doxycycline ([dox] = 0.05, 0.5, 2, 6 ug/ml), and measured the changes in Nanog, Oct4, and Sox2 transcript levels in individual cells over five days. These time course experiments yielded intriguing findings. First, tracking the change in Nanog distribution over the five days of doxycycline induction (Figure 5.5) at a relatively low concentration of doxycycline ([dox] = 0.5 ug/ml) shows that the total Nanog distribution stabilizes around the third day of induction: minimal changes occur in the Nanog distribution between the third and fifth days of induction. In response to the increase in the average Nanog transcript level, Oct4 and Sox2 both increase slightly (Figure 5.6). Second, tracking the change in Nanog distributions (Figures 5.7-5.8) show that for relatively high

concentration of doxycycline ([dox] = 2 ug/ml and 6 ug/ml), the total Nanog transcript level is transiently increased and is then turned down after one day of constant doxycycline induction. Correspondingly, the average expression level of Oct4 and Sox2 are transiently increased in proportion to Nanog, and is then decreased together (Figures 5.9-5.10). By the fifth day of doxycycline induction, Sox2 and Oct4 single-cell transcript distributions show that they have noticeably changed compared to the zeroth day distribution (Figure 5.11): Cells that expressed relatively high levels of Sox2 and Oct4 transcripts no longer exist in the new distribution that the cells reached on the fifth day of constant doxycycline induction. We do not yet know a possible mechanism behind this change in distribution. Given the background information we presented in the previous chapter, it would not be surprising if there are intricate mechanisms such as RNAi that may be 'turned on' when the cell detects that the total Nanog transcript level goes above a certain threshold. Whatever the mechanism may be, our experiments clearly demonstrate that the cell somehow detects Nanog transcript or protein level going above a certain level. Future work may follow up on unraveling this mechanism.

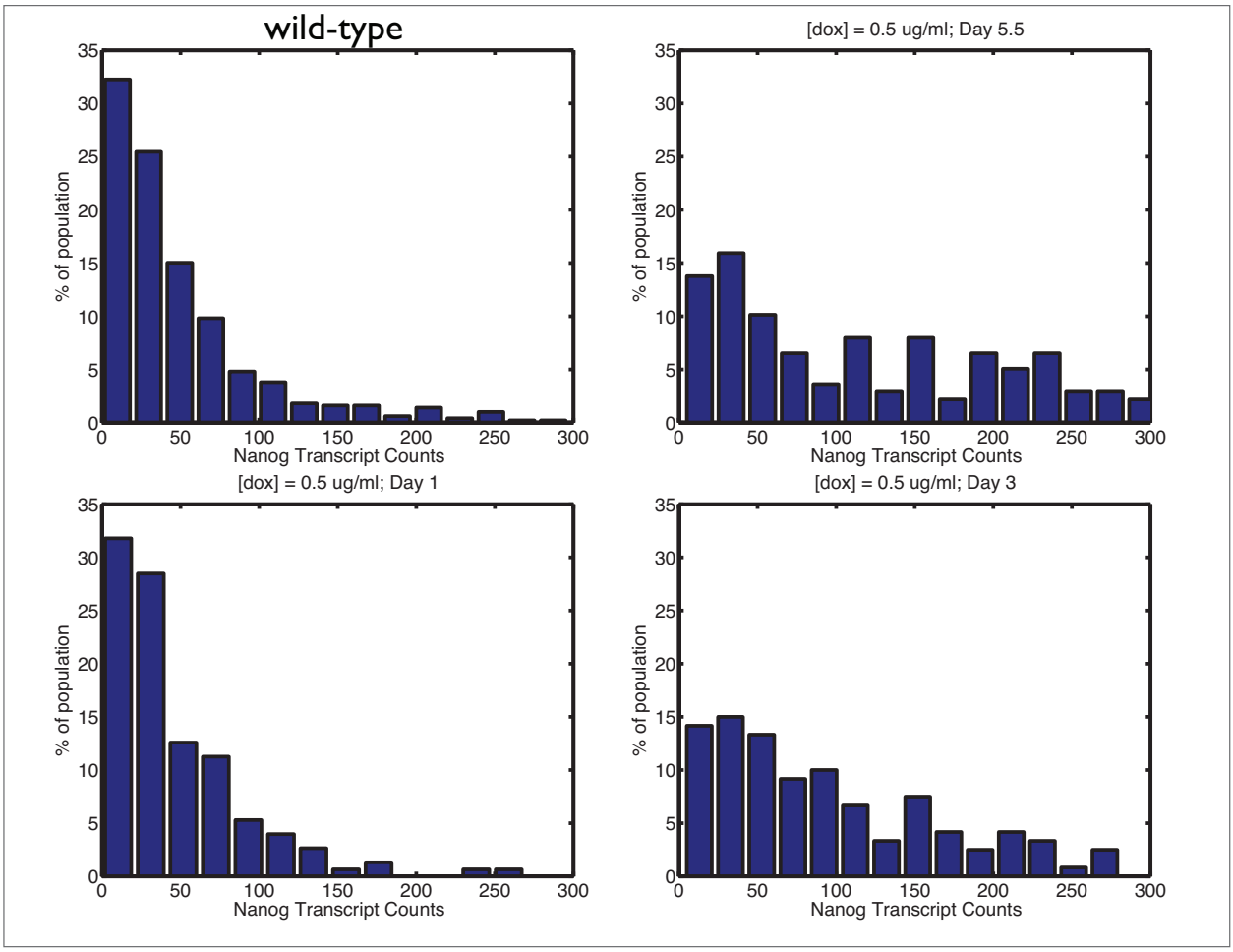


Figure 5.5: Changes in distribution of Nanog transcript in single stem cells after an ectopic overexpression of Nanog; ([doxyclicine] = 0.5 ug/ml).

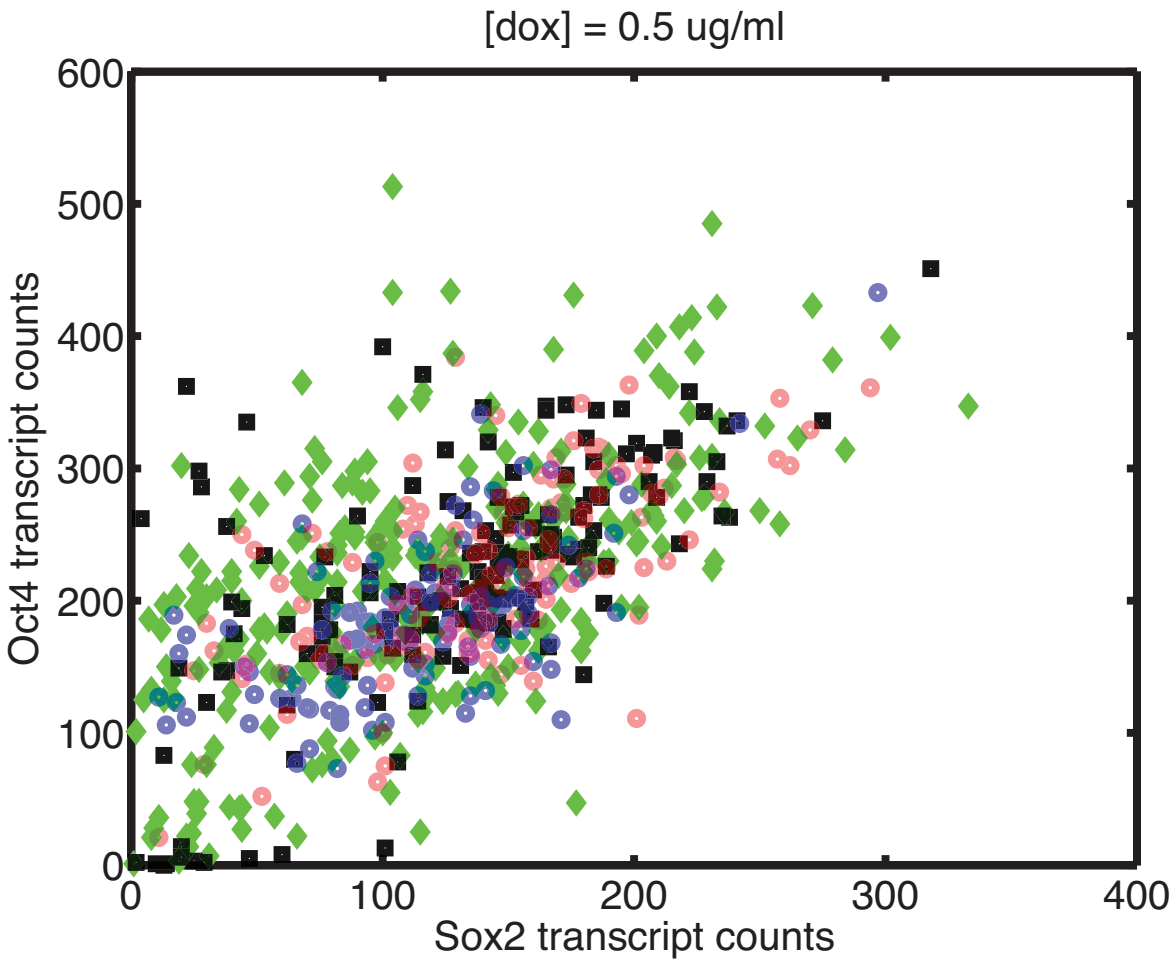


Figure 5.6: Changes in Oct4 and Sox2 as a result of the doxycycline-induced ectopic expression of Nanog. Different colors indicate 'snapshots' of different days during a constant doxycycline-induction: Day 0 (Green), Day 1 (Black), Day 3 (Blue), Day 5 (Red).

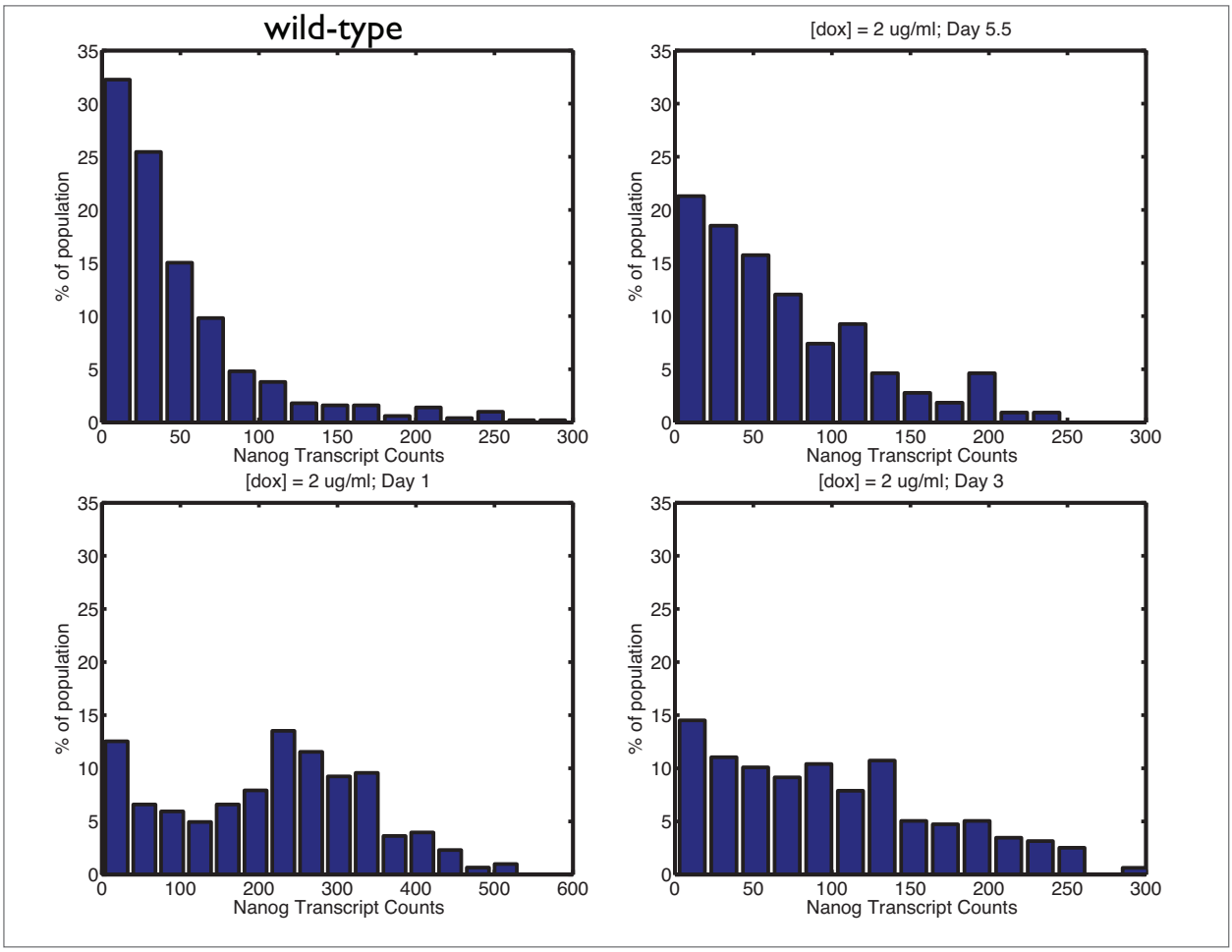


Figure 5.7: Changes in distribution of Nanog transcript in single stem cells after an ectopic overexpression of Nanog; ([doxyclicine] = 2 ug/ml).

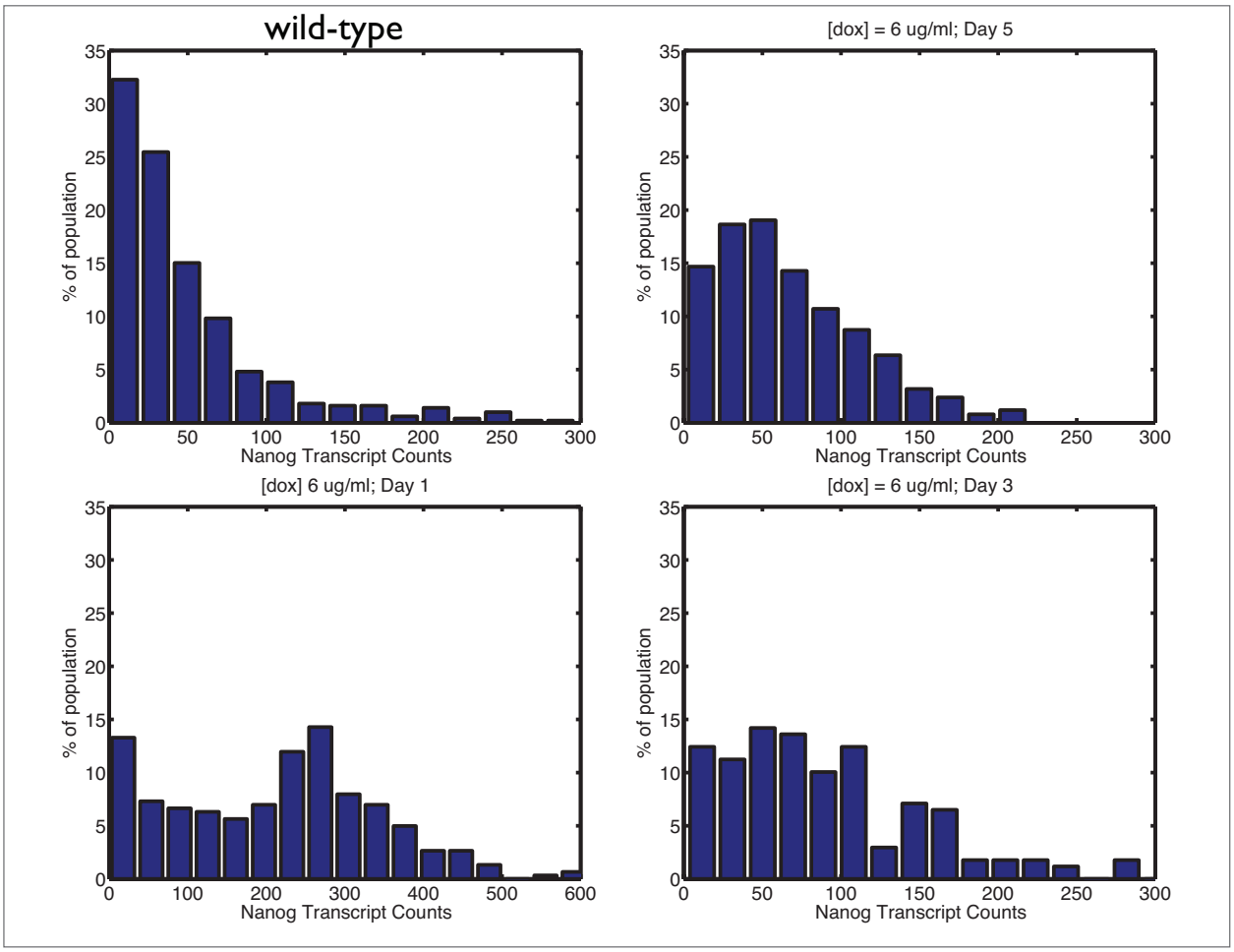


Figure 5.8: Changes in distribution of Nanog transcript in single stem cells after an ectopic overexpression of Nanog; ([doxyclicine] = 6 ug/ml).

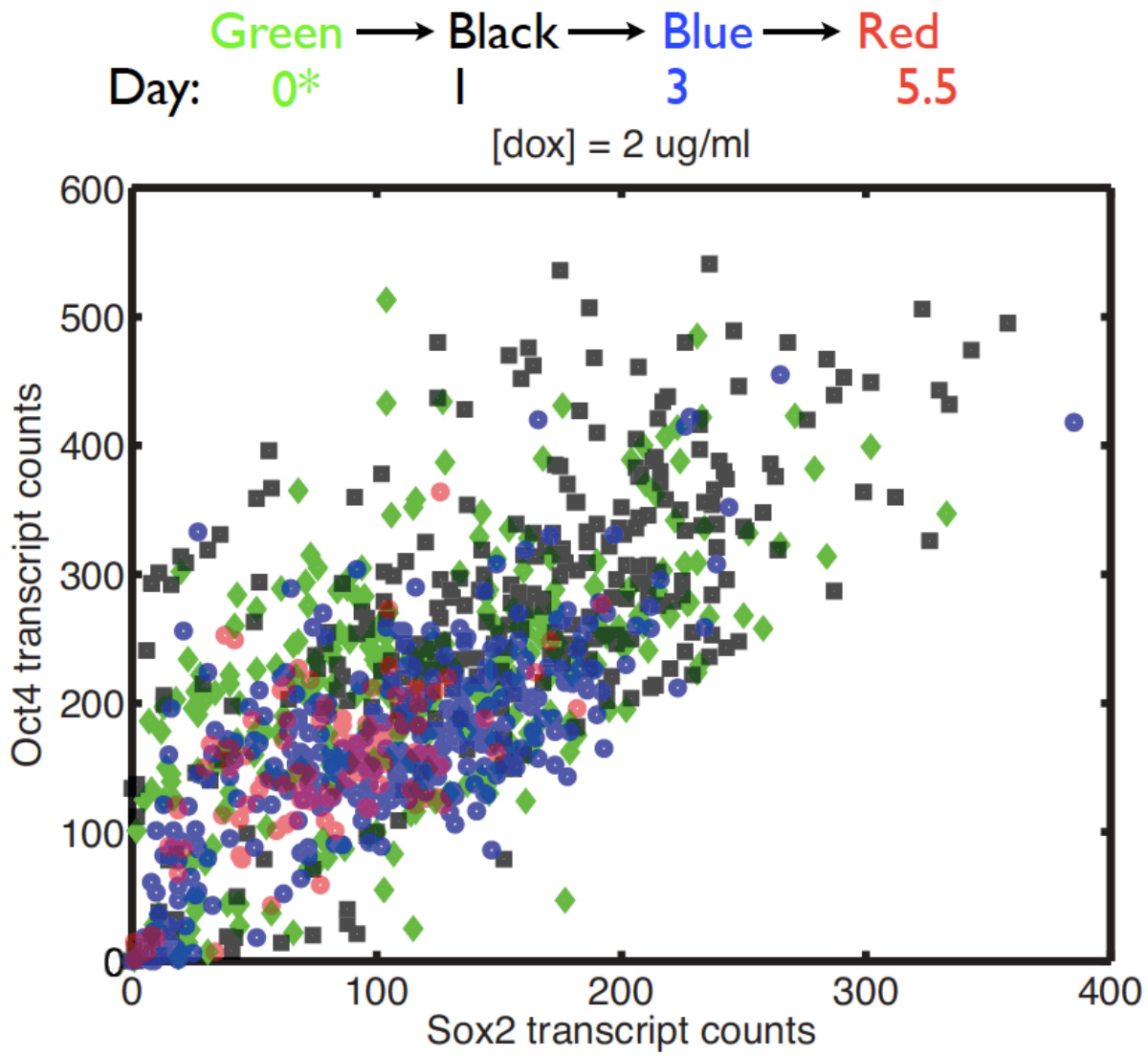


Figure 5.9: Changes in Oct4 and Sox2 as a result of the doxycycline-induced ectopic expression of Nanog; [dox] = 2 ug/ml. Different colors indicate 'snapshots' of different days during a constant doxycycline-induction: Day 0 (Green), Day 1 (Black), Day 3 (Blue), Day 5.5 (Red).

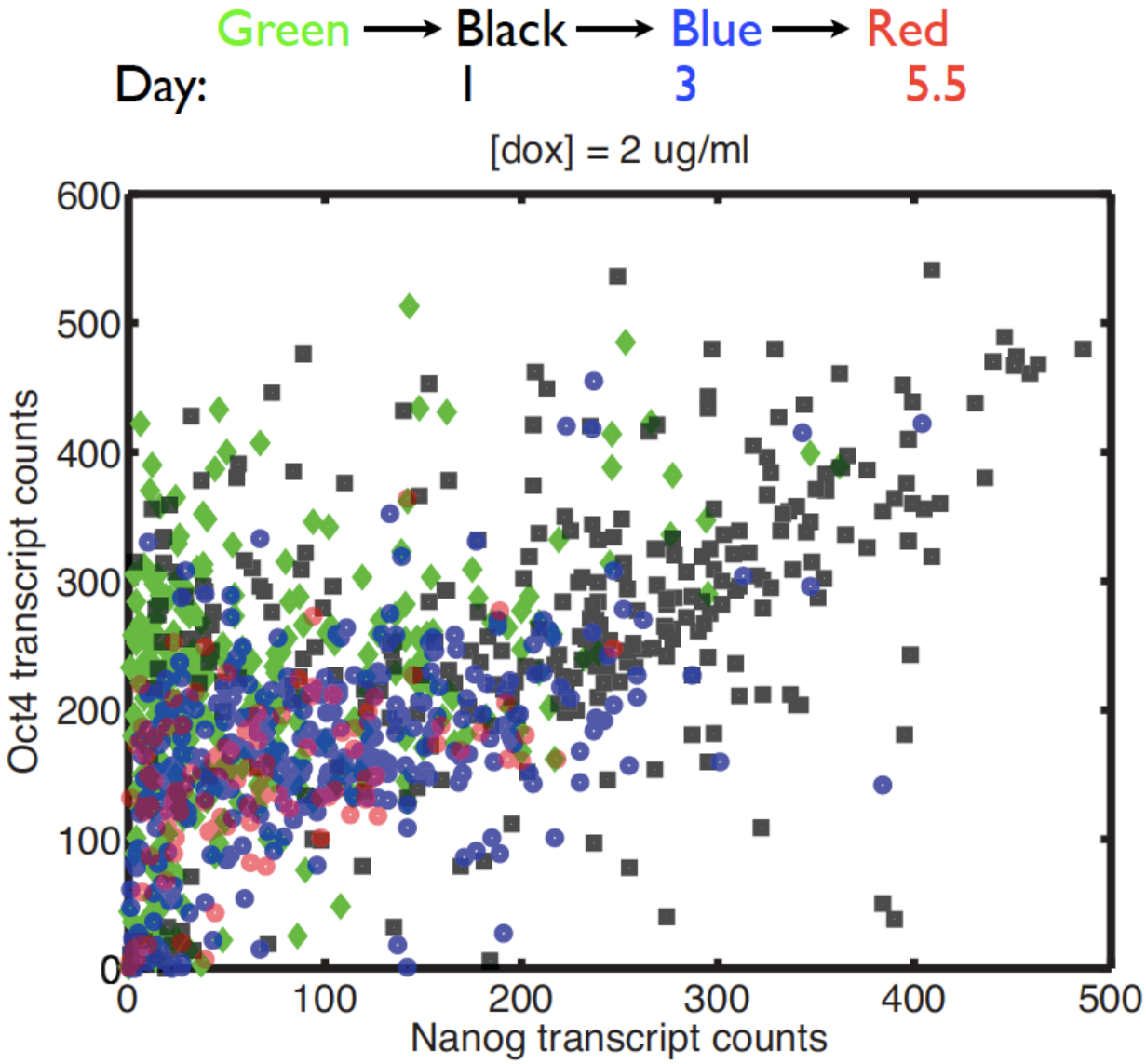


Figure 5.10: Changes in Oct4 as a result of the doxycycline-induced ectopic expression of Nanog; [dox] = 2 ug/ml. Different colors indicate 'snapshots' of different days during a constant doxycycline-induction: Day 0 (Green), Day 1 (Black), Day 3 (Blue), Day 5.5 (Red).

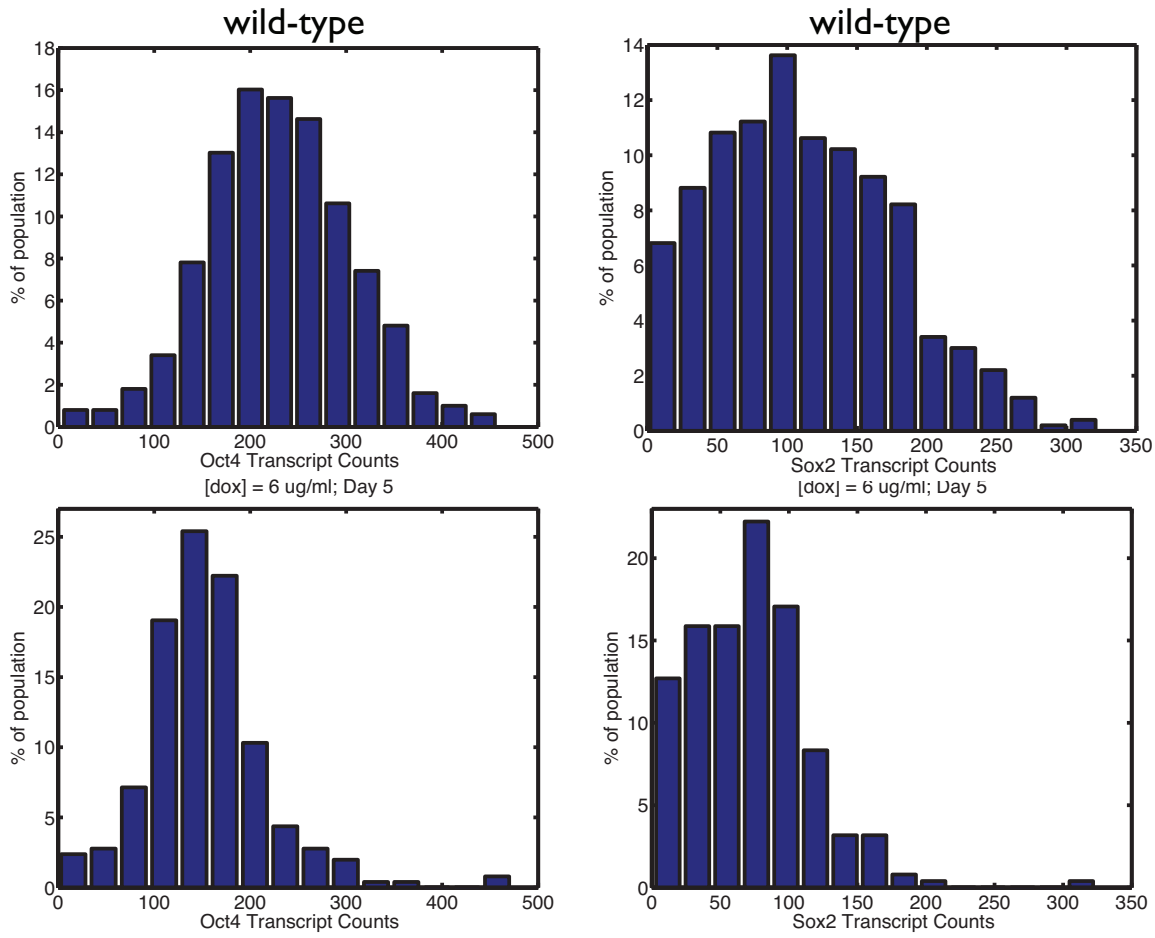


Figure 5.11: At fifth day of constant high level of doxycycline induction ([dox] = 6 ug/ml), a single-cell transcript distributions that are markedly different from the zeroth day ('wild-type') distributions are reached.

5.5. Future directions

We have presented our initial studies in which we perturbed the expression level of Nanog, one of three key transcription factors that plays an essential role in the pluripotency transcriptional network. Specifically, we demonstrated that the stem cell is able to detect an increase in Nanog expression beyond a certain level. We showed that the cell responds by shutting down the total Nanog transcript level. We do not yet know by what mechanism the cell does this. It may be a partial silencing of the Col1A locus where the pTET-Nanog has been integrated, or perhaps a more interesting mechanism such as RNAi that gets triggered for silencing both endogenous and exogenous Nanog transcript. Certain miRNAs have been shown, in human embryonic stem cells, to target the coding region of key pluripotency genes. So this may be an intriguing possibility. Another future direction is to focus on a small, short-time scale perturbation of Nanog expression. This would be independent of the 'threshold' behavior we observed for the larger perturbation Nanog. A logical next step to take is building a careful, quantitative model that describes the shifts in the single-cell distributions that we observed as a result of the perturbation in Nanog. Overexpressing Oct4 and Sox2 using doxycycline is another logical follow-up study.

Our demonstration that some interesting features in Nanog, Sox2, and Oct4 transcripts within single stem cells exists sets up a stage for these future studies.

Bibliography for Chapter 5

[1.] Raj, A. *et al.* Imaging individual mRNA molecules using multiple singly labeled probes. *Nature methods* **5**, 877 (2008).

6. Experimental protocols

6.1. Protocols used for yeast work (Chapter 3)

1. Growing and storing plasmids:

Plasmids need to be grown inside *E. coli* cells. Before a plasmid is to be transformed into a yeast cell, the host *E.coli* needs to proliferate. To do so, we grow the *E.coli* cells in LB media (Luria Broth) at 37 C. Plasmids that have been extracted from E.coli cells can be stored in the 4C fridge. To grow up an appreciable number of *E. coli* cells, pick a single colony from a freshly streaked selective plate and inoculate a culture of 5 ml LB medium containing the selective antibiotic (usually ampicilin). Incubate for 12 hours at 37 C with vigorous shaking.

2. Extracting a plasmid from *E. coli*:

This protocol uses the Qiaprep spin miniprep kit. This method is not suitable for large (> 10 kbp) plasmids.

Grow 5 mL of monoclonal *E. coli* cells for about 15 hours in a rotator in 37C.

A. Harvest the bacterial cells by centrifugation at 8000 rpm for 3 min at room temperature.

B. Resuspend the pelleted bacterial cells in 250 microliters Buffer P1 and transfer to a microcentrifuge tube. Make sure no bacterial clumps remain. Ensure that RNase A has been added to buffer P1.

C. Add 250 microliters Buffer P2 and mix thoroughly by inverting the tube 4-6 times. DO NOT Vortex. Do not allow reaction to proceed more than 5 min.

D. Add 350 microliters Buffer N3 and mix immediately and thoroughly by inverting the tube 4-6 times. The solutin should become cloudy.

E. Centrifuge for 10 min at 13000 rpm.

F. Decant the supernatant into a QIAprep spin column by decanting or pipetting

G. Centrifuge at 13000 rpm for 30 to 60 seconds. Discard flow through.

H. Wash the QIAprep spin column by adding 0.5 ml Buffer PB and centrifuging 30-60 s. Discard flow through.

I. Wash spin column by adding 0.75 ml Buffer PE and centrifuging for 30-60 s.

J. Discard flow-through and centrifuge for additional 1 minute. This is important.

K. Place the QIAprep column in a clean 1.5 ml microcentrifuge tube. To elute DNA, add 50 microliter Buffer EB or water to the center fo each QIA prep spin column, let stand for 1 min, and then centrifuge for 1 min.

This completes the extraction of the plasmid from the E. Coli.

3. Linearizing plasmids (Digestion):

Before a plasmid DNA can be transformed into a yeast genome, it needs to be linearized. Digestion might also be necessary prior to ligation steps. The linearization is achieved by cutting the plasmid at the required site with a restriction enzyme. The following protocol is to be used, for a final volume of 20 ul:

A. Use 10 microliters plasmid

B. Add 2 microliters buffer (It is at 10X) - look on the wall chart to see which buffer is to be used with the enzyme being used.

C. Add 2 microliters of 10X BSA (if needed).

D. Add 1 microliter of the restriction enzyme (do not use more than 1 microliter per 20 microliter final volume).

E. Add 5 microliters of H2O (make up to 20 ul total).

F. Incubate the mixture at 37 C for at least 3 hours (sometimes even overnight depending on the enzyme).

This completes the linearization of the plasmid. Now the plasmid needs to be checked (purification).

4. Gel purification:

An agarose gel is used to perform electrophoresis to verify and extract the linearized plasmid.

A. Prepare 60 ml of 1% agarose gel using UltraPure Agarose. Dissolve the agarose in TAE buffer, not H2O.

B. Boil the agarose solution in a microwave.

C. Prepare the Gel Slab using the appropriate wells.

D. Add sybersafe - 6 microliters for 60 ml agarose solution. The cybersafe is at 10000X concentration.

E. Add load dye to the dna that is to be checked. typical volumes are 20 microliters of DNA and the load dye is 6X so about 4 microliters of load dye to 20 microliters dna should be good.

F. For the ladder, mix 19 TAE, 1 ladder and 4 load dye. The ladder is kept at -20C and the dye is kept at 4C.

G. Inject the dna and ladder and control into the gel wells and run for 40 minutes at 110 volts.

H. Check the DNA using the UV lamp

I. Excise the DNA fragment from the Agarose gel with a blade and place in an eppendorf tube.

J. Add three volumes of buffer QG to 1 volume of gel (1mg=1ul)

K. Incubate at 50 C for 10 min (or until gel is completely dissolved). To help dissolve, mix by vortexing every 2 to 3 min.

L. Check that the color of the dissolved solution is yellow. If the color is orange or violet, add 10 ul of 3M sodium acetate, ph 5.0 , and mix

M. Add 1 gel volume of isopropanol to sample and mix.

N. Place a QIAquick spin column in a provided 2 ml collection tube.

O. Apply the sample to the spin column and centrifuge for one minute (maximum volume is 800 ul). If more than 800 ul is needed, load and spin again.

P. Discard flow through and place the spin column in the collection tube.

Q. Add 0.75 ml of Buffer PE to the column and centrifuge for 1 min.

R. Discard flow through and centrifuge for additional 1 min.

S. Place column in a clean eppendorf tube.

T. Elute the dna by adding 50 ul of water to the centre of the membrane. LET IT STAND FOR 1 MIN. Centrifuge for 1 min to collect the dna.

5. **DNA Ligation:**

DNA ligation is the process used to splice dna pieces together. These pieces must be compatible which is assured if the ends being spliced were both cut by the same restriction enzyme.

A. The total volume for the reaction is 20 microliters

B. Mix in a tube:

- 1 ul backbone (the plasmid to which the fragment is being ligated)
- 2 ul buffer - T4 DNA ligase buffer
- 1 ul enzyme - T4 DNA ligase
- 5 ul of each DNA fragment
- Make up the rest of volume with water up to 2o ul.
- Also make a control tube with backbone only (no insert)

C. Vortex well, spin down and leave on bench for 1 hour

D. Transform the resulting plasmid into E.Coli.

6. Transforming plasmids into E.Coli:

Once a plasmid is ready, it must be introduced into E. Coli cells for proliferation. Competent E.Coli cells are required for this purpose. This protocol is for transforming 20 ul plasmid into E.Coli cells so first make ready an eppendorf tube containing 20 ul plasmid.

A. Thaw competent E. Coli cells on ice.

B. Add 75 ul competent E. coli cells to the plasmid DNA and pipette gently to mix.

C. Incubate on ice for 30 minutes.

D. Incubate for 50 seconds at 42 C.

E. Put back in ice for 2 minutes.

F. Add 0.9 ml LB media.

G. Incubate for 1 hour at 37 C on shaker.

H. Spin down and decant till only 100 ul is left in the tube.

I. Plate on an antibiotic plate.

7. Freezing *E.coli* and *S. cerevisiae*:

A. Take 1 ml cells from the bottom of the culture tube. Do not centrifuge. Put these cells in a freezing tube.

B. Add 0.5 ml of 80% glycerol.

C. Freeze in -80 C.

8. **PCR:**

PCR is a technique that is used to multiply a certain desired segment of a DNA sample. In order to use PCR one must first design primers. The primers must be such that one of the primers has an end homologous to one of the DNA strands at one end and the other primer must have an end homologous to the other strand at the other end. The other end of the primer should be designed to have the desired restriction site.

A. The primers come in solid form

B. Spin down the primer

C. Make 1mM stock solution of the primer. The number of moles present is written on the sheet that comes with the primer.

D. Vortex and spin down

E. Make a 5 uM solution of the primer by dilution.

F. To maximise the probability of success, use four different conditions in which to do the PCR. They are outlined in the table below (to make 100 ul final volume; all volumes in the table are in uL):

	1	2	3	4
water	74.5	59.5	64.5	49.5

10mM dNTPs	3	3	3	3
50mM $MgSO_4$	2	2	2	2
template DNA	0.5	0.5	0.5	0.5
5uM forward primer	4	4	4	4
5uM reverse primer	4	4	4	4
10x PCR enhancer	0	0	10	10
DNA polymerase (e.g. PFX)	2	2	2	2

G. Perform PCR in a thermocycler, with the appropriate melting temperature.

9. PCR Purification

Once the PCR process is complete the DNA needs to be collected in water. The process is as follows:

A. Add 5 volumes of Buffer PB to 1 volume of the PCR sample and mix.

B. Apply to a QIAquick column and centrifuge for 1 min. Discard flow through.

C. Add 0.75 ml Buffer PE to the QIAquick column and centrifuge for 1 min.

D. Discard flow through and centrifuge for additional 1 min.

E. Place the column in a clean tube. Elute dna by adding 50 ul water, letting stand for 1 min and then centrifuging for 1 min.

10. Colony PCR

A. 15 uL Zymolyase in a PCR tube.

B. Scrape off ~ 1/4 colony with a pipette tip, inoculate into zymolyase.

C. Run the tube in PCR w/ a program “GEN”. (~30 minutes).

D. Do PCR on this product.

6.2. Protocols used for stem cell work (Chapter 5)

1. Recipe for embryonic stem cell growth media

62.5 mL FBS (for final 12.5% FBS) (Fetal Bovine Serum, HyClone)

5 mL L-glutamine (Gibco)

5 mL pen/strep (Gibco)

5 mL non-essential amino acids (NEAA) (Gibco)

4 uL Beta-Mercaptoethanol (BME)

1 aliquot LIF ($7x10^7$ units)

fill to 500 mL knockout DMEM (Gibco)

2. Recipe for freezer media (for freezing down cell lines into vials)

For 20 mL total volume:

12 mL DMEM (Gibco)

4 mL CCS (Cosmic Calf Serum, HyClone)

4 mL DMSO (Invitrogen)

3. Recipe for trypsin

For 200 mL total volume:

20 mL 2.5% trypsin (Gibco)

432 uL 0.5M EDTA

fill to 200 mL with 1X PBS

Cell culture protocols

4. Feeding cells (for 15 cm plates. For 10cm plates, cut all volumes in half):

A. Put media in waterbath to warm ~ 10 minutes.

B. Flame pipet carefully and aspirate off old media.

C. Add 20 mL warmed media to plates near edge (adding to middle of plate can potentially life cells off the plate).

5. Splitting cells <u>(for 15 cm plates. For 10cm plates, cut all volumes in half):</u>

A. Aspirate old media out of plates the same way as in feeding cells.

B. Wash cells with 10 mL of 1x PBS, then aspirate out PBS.

C. Add 2mL 1x trypsin, make sure plate is coated, then place in incubator for ~ 5 minutes.

D. Fill a conical tube half way with fresh media while you wait.

E. Tap gently on the sides of plates to make sure all cells are lifted by the trypsin. Then flame a plugged Pasteru pipet, and gently pipet up and down the trypsinized cells, making sure there are no patches of cells still stuck, and that all the cells are individually suspended.

F. To deactivate the trypsin add the trypsinized cells directly into the conical tube with media.

G. Rinse plate once with fresh media and collect remaining media and cells in the conical tube. It is important here to not wash the cells over the plate, as they will re-stick, and you will lose cells.

H. Spin conical tube with media and cells for 1000 rpm for 3 minutes.

I. Aspirate out media, then resuspend in desired volume of media for plating (e.g. 1mL media per plate).

6. Freezing cells

A. Trypsinize cells according to above procedure.

B. Make fresh freezer media every time you freeze cells. Depending on what cells you are freezing, you my have to count the cells beforehand to determine appropriate aliquots and volume of freezer media needed.

C. Label cryo tubes with.

D. Aliquot 0.75 mL cells, making sure to carefully resuspend every time, to assure consistent suspension.

E. As quickly as possible, aliquot 0.75 mL freezer media, cap then place tubes in a Mr. Frosty filled with isopropanol, and place I nthe -80 freezer overnight.

F. Once frozen, place tubes in appropriate liquid nitrogen box.

7. Single molecule RNA FISH protocols

This material has been adapted from a similar protocol designed for yeast, which appeared in:

H. Youk, A. Raj, and A. van Oudenaarden, *Methods in Enzymology* 470, pp429-446 (2010).

A brief overview of our method is as follows. A set of short (between 17 to 22 bases long) oligonucleotide probes that bind to a desired target mRNA are designed and are coupled to a fluorophore (such that one oligonucleotide probe is bound to a single fluorophore) with desired spectral properties. After fixing the yeast cells, these probes are hybridized to the target mRNA molecule. This results in multiple (typically about 48) singly-labeled probes bound to a single mRNA molecule. In turn, the mRNA molecule can give off enough fluorescence to be detected as a diffraction-limited spot using a standard fluorescent microscope. Below we describe a step-by-step procedure that was implemented for the RNA FISH in embryonic stem cells, presented in Chapter 5.

(i) Designing oligonucleotides

The first step is the design of a collection of oligonucleotide probes that together are complimentary to a large part of the open read frame of the target mRNA (one can also utilize the untranslated regions of the mRNA if necessary). Each probe is between 17 to 22 bases long and we have generally found that 30 or more such probes are sufficient to give a detectable signal. We have also found that our signals are sometimes clearer when the GC content of each probe is close to 45%. We also leave a minimum of two bases as a spacer between two adjacent probes that cover the mRNA, although it is possible that one can relax this requirement without any adverse effects. A program that facilitates the designing of probes meeting the constraints mentioned above is available freely at http://www.singlemoleculefish.com. Sometimes it is not possible to design probes that meet all the constraints mentioned above, and these criteria should not be viewed as absolutes, but more as guidelines we try to adhere to when possible. After

designing the probes, we order them from companies with parallel synthesis capabilities (we use BioSearch Technologies based in Novato, California, USA) with 3'-amine modifications. Since the synthesis typically results in a much larger number of oligonucleotides than are necessary, one should have them synthesized on the smallest possible scale (we typically have them synthesized on the 10 nmole (delivered) scale). The 3'-amine then serves as a reactive group for the succinomidyl-ester coupling of the fluorophore described next.

(ii). Coupling fluorophores to oligonucleotides

The next step is the attachment of a fluorophore with desired spectral properties to the commercially synthesized oligonucleotides (we will describe which fluorophores we use in the later section entitled "Choice of fluorophore and appropriate filter sets"). We do this by pooling the oligonucleotides and coupling them *en masse*, thus reducing the labor involved. In all the steps we describe below, we use RNase free water (Ambion) to prepare our solutions and use filtered pipette tips to prevent aerosol contaminations.

Procedure:

1. From the commercially synthesized set of oligonucleotides, each at a concentration of 100 μM in RNase free water (we find this is a practical starting concentration to work with), pipette around 1 nmol/10 μL of each oligonucleotide probe into a single microcentrifuge tube (i.e. if there are 48 probes, then 1 nmol of each of the 48 probe solutions should be combined into a single tube with a final volume of 480 μL).

2. Add 0.11 volumes (vol./vol.) of 1 M sodium bicarbonate (prepared with RNase free water) to this probe mixture, resulting in a final sodium bicarbonate concentration of 0.1 M. If the total volume of the mixture at this stage is less than 0.3 mL, add enough 0.1 M sodium bicarbonate to bring the final volume of the mixture to 0.3 mL.

3. Dissolve roughly 0.2 mg of the desired fluorophore (functionalized with a succinimidyl ester group) separately into a tube containing 50 μl of 0.1 M sodium bicarbonate. If using TMR, first dissolve the TMR in about 5 μL of dimethyl sulfoxide (DMSO) and then add 50 μL of 0.1 M sodium bicarbonate to it. This is because TMR does not readily dissolve in aqueous solutions.

4. Add the dissolved fluorophore to the 0.3 mL of probe mixture, vortex, and cover this tube in aluminum foil to prevent photobleaching from unwanted exposure to ambient light. Leave the tube in the dark overnight.

5. Next day, precipitate the probes out of solution by adding 12% vol./vol. of sodium acetate at pH 5.2 followed by 2.5 volumes of ethanol (95% or 100%).

6. Place the tube at -70C for at least one hour, then spin the sample down at 16,000 RPM for at least 15 minutes at 4C.

7. A small colored pellet should have collected at the bottom of the tube at this stage. This pellet contains both the coupled and uncoupled oligonucleotides. The vast majority of the uncoupled fluorophore, however, remains in the supernatant, and so aspirate as much of this supernatant away as possible without disturbing the pellet (one should take care to aspirate soon after removal from the centrifuge, since oligonucleotides have a tendency to redissolve rapidly at room temperature.

   Note: Many precipitation protocols now call for another washing step in 70% ethanol. We have found this step unnecessary.

8. The pellet is stable and can be stored in -20C for up to one year. This concludes the coupling step.

Choice of fluorophore and appropriate filter sets:

In order to perform imaging of multiple different RNA species at the same time, one needs to select fluorophores with excitation and emission properties that can be distinguished by appropriately chosen bandpass filters; otherwise, the signal from one channel may potentially bleed into another channel. We describe here the fluorophore and filter set combination that we use for our microscopy. Other combinations are no doubt feasible as well.

The fluorophores we utilize are TMR (tetramethylrhodamine), Alexa 594 and Cy5. TMR has proven to be exceptionally photostable in our hands, and its excitation maximum of 550 nm aligns nicely with the excitation maxima of mercury and metal-halide light sources. Alexa 594 is also quite photostable, and while it's spectral properties are similar to those of TMR (absorption

at 594nm), we are able to distinguish its presence using appropriate filters. The third fluorophore we use is Cy5, which is rather bright and is spectrally separated from the other two fluorophores (Cy5 absorbs at 650nm). Cy5 does, however, suffer from photobleaching effects, thus requiring the use of a glucose oxidase oxygen scavenging system to make imaging feasible. We have not tried any dyes that are further redshifted than Cy5. However, we have experimented with Alexa 488, which absorbs at a lower wavelength than TMR. While we were sometimes able to detect signals, the higher cellular background at these lower wavelengths lead to weaker signals, so we generally avoid the use of fluorophores bluer than TMR.

The filters combinations we use are typical bandpass filter and dichroic sets mounted in cubes that the microscope can place in the fluorescence light path. For TMR, we use a standard XF204 filter from Omega Optical. For Alexa 594, we use a custom filter from Omega Optical with a 590DF10 excitation filter, a 610DRLP dichroic, and a 630DF30 emission filter. For Cy5, we use the 41023 filter from Chroma, which is designed for Cy5.5. It is likely that a filter more appropriate for Cy5 would work even better. These filters do a good job of preventing any signals from one fluorophore from being detected in another channel (Raj *et al.* 2008). Sometimes a very bright Alexa 594 signal can bleed somewhat into the TMR channel (we estimate the bleed through to be about 10%) but practically this bleedthrough is impossible to detect owing to the low signal intensities of the mRNA spots.

(iii). Purification of probes using HPLC:

We now describe a purification procedure for separating the coupled oligonucleotides from the uncoupled oligonucleotides. We purify the coupled oligonucleotides using HPLC (High Performance Liquid Chromatography): the addition of the fluorophore makes the normally hydrophilic oligonucleotide significantly more hydrophobic, allowing for separation by chromatography. The HPLC should be equipped with a dual wavelength detector for a simultaneous measurement of absorption by DNA (at 260 nm) and fluorophore (depends on the fluorophore: e.g. 555 nm for TMR and 594 nm for Alexa 594). In our lab, we have used an Agilent 1090 equipped with Chemstation software and a C18 column suitable for oligonucleotide purification (218TP104). The two buffers used for HPLC are: 0.1M triethylammonium acetate ("Buffer A") and acetonitrile ("Buffer B").

Procedure:

1. Before running the purification program on the HPLC, equilibrate the column by flowing 93% buffer A/7% Buffer B through for about 10 minutes; if the column is not equilibrated, then the oligonucleotides will simply flow straight through without any separation.

2. Re-suspend the oligonucleotide pellet in an appropriate volume of water (we use 115 µl) and then inject this into the HPLC inlet.

3. Run a HPLC program in which the percentage of buffer A varies from 7% to 30% over the course of about 45 minutes with a flow rate of 1 mL per minute. During the execution of the program, carefully monitor the two absorption curves, one for DNA (at 260 nm) and the other for the coupled fluorophore (e.g. 555 nm for TMR and 594 nm for Alexa 594). Generally speaking, one will observe two broad peaks over time. The first peak, containing the more hydrophilic material, consists of the uncoupled oligonucleotides and will only exhibit absorption in the 260 nm channel (Fig. 6.1A). This peak may appear relatively ragged due to the presence of multiple oligonucleotides, each of which has a slightly different retention time in the HPLC. The second peak, often narrower than the first, will appear some time after the first peak and contains the coupled oligonucleotides; thus, it will show absorption in both the 260 nm and the fluorescent (e.g., 555 nm) channels (Fig. 6.1B). The duration of time between the first and second peaks varies depending on the hydrophobicity of the fluorophore; we have found that oligonucleotides coupled to Cy5 have a long retention time of almost 20 minutes after the first peak, whereas TMR and Alexa 594 result in shorter retention shifts (Fig. 6.1B).

4. Collect the contents of this peak (in the flurophore absorption channel) manually into clean, RNase free tubes. It is important to collect all the solution that is coming out of the outlet, starting from the beginning of the left shoulder of this second peak and stopping the collection just at the tail-end of the right shoulder of this second peak (Fig. 6.1B), because the different coupled oligonucleotides will have slightly different retention times; do not just "collect the peak". This collection typically lasts around 3-7 minutes in our experience. With the volumes we mentioned for our HPLC setup above, we typically collect between 5 to 14 mL in this step with 0.5 mL per tube. The program we use then typically flows 70%

Buffer B throught the column for about 10 minutes. This step will "strip" the column of any impurities that may have stuck to the column and is especially important if you plan to purify additional probes. Be sure, however, to allow sufficient time for the column to re-equilibrate to 7% B/93% A before injecting another sample.

5. After collecting the solution of coupled probes, dry the collection in a speedvac rated for acetonitrile until the liquid is fully evaporated (about 3 to 5 hours). It is important to keep light out of the Speedvac to avoid photobleaching of dyes, especially for highly photolabile cyanine dyes such as Cy3 and especially Cy5.

6. Re-suspend the contents in a total of 50 to 100 μL of TE (10 mM Tris with HCl to adjust pH, 1 mM EDTA, Ambion) at pH 8.0. This final suspension solution is now the "probe stock".

7. From the "probe stock", create dilutions of 1:10, 1:20, 1:50, and 1:100 in TE to make "working stocks". This dilution series is used to determine which concentration of probes yields the best signals for RNA FISH.

8. Store these probes in dark at -20 C until sample is ready to be prepared. We found that the probes can be stored for years in this way.

(iv). Fixing embryonic stem cells

Having isolated the coupled probes, it is now time to fix the stem cells so that these probes can be hybridized to their target mRNAs in these cells.

Procedure:

1. Add 5 mL of 37% formaldehyde (i.e. 100 % formalin) directly to the growth media containing the cells and let it sit for 15 minutes at room temperature to fix the cells. One should take safety precautions when using the carcinogen, formaldehyde (i.e. use chemical fume hood, gloves, and long-sleeved protective clothing).

2. After spin down and washing with 1x PBS, add 1 mL of 70% ethanol (diluted in RNase free water) to the cells and leave them for an hour or even overnight at 4 C.

The stem cells have now been fixed and are ready for hybridization. These cells can be stored in ethanol for up to a week after fixation and perhaps even longer.

(v). Hybridizing probes to target mRNA

The hybridization step contains three key parameters that may be varied to optimize the FISH signal. These are the temperature at which hybridization takes place, the concentration of formamide used in the hybridization and wash, and the concentration of the probe. The first two parameters essentially set the stringency of the hybridization; i.e., the higher the temperature or the concentration of formamide, the lower the likelihood of nonspecific binding of the probes. Formamide is used in both DNA and RNA FISH for stabilizing single stranded, denatured DNA and for deionizing RNA. We usually elect to adjust the formamide concentration rather than temperature and thus perform all FISHs at 30 C. Typcially, we have found that hybridization and wash buffers containing 10% formamide work quite nicely for most probes, yielding a fairly low background while also producing clear particulate signals. However, when the GC content of the probes is relatively high (>55%), we have found that we sometimes have to employ formamide concentrations up to 20% or sometimes higher. However, care must be taken in these instances, since the use of higher formamide concentrations can sometimes lead to a greatly diminished signal. Generally, we try to obtain signals at a standard concentration of formamide, because this greatly facilitates the simultaneous detection of multiple mRNAs: if the hybridization conditions are the same, multiplex detection is simply a matter of mix and match.

The concentration of probe used is also very important in obtaining clear, low background signals. Typically, the optimal probe concentration must be found empirically, but we have found that concentrations can vary over roughly an order of magnitude and still produce satisfactory results. We typically start by using a 1:1000, 1:2500 and 1:5000 dilution of the original stock into hybridization buffer. One of these concentrations will usually yield good signals, but sometimes one must use drastically lower concentrations (100 fold lower) in order to obtain signals.

(v). a. Preparation of hybridization and wash buffers

The following procedure describes preparation of 10 mL of hybridization buffer with the desired formamide concentration. Be sure to adjust the volumes appropriately if you're preparing a different total volume of hybridization buffer.

Procedure:

1. Dissolve 1 g of high molecular weight dextran sulfate (>50,000) in approximately 5 mL of nuclear free water. Depending on the particular preparation of dextran sulfate used, the powder may dissolve quite rapidly with a bit of vortexing or may require rocking for several hours at room temperature. In the end, the solution should be clear and fairly viscous, although some preparations are far less viscous but still appear to work.

2. Add 10 mg of *E. coli* tRNA (Sigma, 83854), vortexing to dissolve.

3. Add 1mL of 20x SSC (RNase free, Ambion), 40 μL (to get 0.02% in 10 mL) of RNase free BSA (stock is 50 mg/mL = 5% solution from Ambion, AM261), 100 μL of 200mM vanadyl-ribonucleoside complex (NEB S1402S), formamide to the desired concentration (10%-30%), and then water to a final volume of 10mL. When using formamide, one must first warm the solution to room temperature before opening to avoid oxidation; also, care must be taken when using formamide (i.e., use in the hood, wear protection, etc.) because it is a suspected carcinogen and teratogen and is readily absorbed through the skin.

4. Once the solution is thoroughly mixed, filter the buffers into small aliquots; this removes any potential clumps that can yield a spotty background. We simply filter the solution in 500 μL aliquots using cartridge filters from Ambion.

5. Store the solution at -20C for later use; solution is typically good for several months to a year.

6. Prepare the wash buffer by combining 5 mL of 20x SSC (Ambion), 5 mL of formamide (to final concentration of 10% vol./vol.; this is adjusted if the hybridization buffer has a different formamide concentration), and 40 mL of RNase free water (Ambion) into one solution.

(v). b. Hybridizing probes to stem cells in solution

Procedure:

1. Warm the hybridization solution to room temperature before opening its cap to prevent oxidation of the formamide.

2. Add 1-3 μL of desired concentration of probes to 100 μL of the hybridization buffer. To determine what the desired concentration of probes is, we initially perform hybridizations with four dilutions of probes: 1:10, 1:20, 1:50 and 1:100 (mentioned in "Purification of probes using HPLC" section), and see which dilution gives the clearest signal.

3. Centrifuge the fixed sample and aspirate away the ethanol, then resuspend the fixed cells in a 1 mL wash buffer containing the same formamide concentration as the hybridization buffer.

4. Let the resuspension stand for about 2-5 minutes at room temperature.

5. Centrifuge the sample and aspirate the wash buffer. Then add the hybridization solution.

6. Incubate the sample overnight in the dark at 30 C.

7. Next morning, add 1 mL of wash buffer to this sample, vortex, centrifuge, then aspirate away the supernatant.

8. Resuspend in 1 mL of wash buffer, then incubate in 30 C for 30 minutes.

9. Repeat the wash in another 1 mL of wash buffer for another 30 minutes at 30 C, this time adding 1 μl of 5 mg/mL DAPI for a nuclear stain.

10. A : *If using photostable fluorophores such as TMR or Alexa 594*: then there is no need to add the GLOX solution. Just resuspend the sample in an appropriate volume (larger than 0.1 mL) of 2x SSC and proceed to imaging.

10 B: *If using a highly photolabile fluorophore such as Cy5*: resuspend the fixed cells in the GLOX buffer (used as an oxygen-scavenger that removes oxygen from the medium to prevent light-initiated fluorophore destroying-reactions; see section 2.5. C) without the enzymes and incubate it for about 2 minutes for equilibration (see next section for details). Then centrifuge, aspirate away the buffer and resuspend the cells in a 100 μL of GLOX buffer with the enzymes (glucose oxidase and catalase). These cells are now ready to be imaged.

We found that our samples (either with or without the anit-bleach solution) can be kept at 4C for a day's worth of imaging. Keeping the samples at 4C prevents the probe-target hybrids from dissociating and thus degrading the signals.

(v). c. Preparation of anti-bleach solution and enzymes ("GLOX solution")

During imaging, we typically take several vertical stacks ("z-stacks") of images through a cell in a field of view, causing a hybridized fluorophore in a fixed cell to be excited by intense light several times. More importantly, when more than one type of fluorophore is used for imaging two or three species of mRNA, such z-stacks must be repeated to excite each of the different fluorophores, leading to even more exposure of the fluorophores. In our experience, only TMR and Alexa 594 could withstand such repeated excitations, whereas Cy5 signal would rapidly degrade due to its especially high rate of photobleaching. To decrease the photolability of Cy5, we used an oxygen-scavenging system consisting of catalase, glucose oxidase, and glucose (GLOX solution) that is slightly modified from that used by (Yildiz *et al.* 2003). This GLOX solution acts as an oxygen-scavenger that removes oxygen from the medium. Since the light-initiated reactions that destroy fluorophores require oxygen, the GLOX buffer thus prohibits these reactions from taking place. Indeed, we found that Cy5 was able to withstand nearly 10 times more exposure with the GLOX solution than without it. The following is a procedure for preparing the GLOX solution.

Procedure:

1. Mix together 0.85 mL of RNase free water with 100 μL of 20x SSC, 40 μL of 10% glucose, and 5 μL of 2 M Tris CL (pH 8.0). This is the GLOX buffer (without glucose oxidase and catalase).

2. Vortex the mixture, and then aliquot 100 μL of it into another tube.

3. To this 100 μL aliquot of GLOX buffer (GLucose-OXygen scavenging solution without enzymes), add 1 μL of glucose oxidase (from 3.7mg/mL stock, dissolved in 50mM sodium acetate, pH 5.2, Sigma) and 1 μL of catalase (Sigma). Before pipetting the catalase, vortex it a bit, since the catalase is kept in suspension (also, care should be taken when handling the catalase, since it has a tendency to get contaminated). This 100 μL will be referred to as

"GLOX solution with enzymes". The GLOX solution without the enzyme will later be used as an equilibration buffer.

(v). d. Imaging Samples Using Fluorescent Microscope

The fixed cells with probes properly hybridized are now ready for imaging. Our microscopy system is relatively standard: we use a Nikon TE2000 inverted widefield epifluorescence microscope. It is important to use a fairly bright light source. For instance, a standard mercury lamp will suffice, although the newer metal-halide light sources (e.g., Lumen 200 from Prior) tend to produce a more intense and uniform illumination. Another important factor is the camera. It is important to use a cooled CCD camera that is optimized for low-light imaging rather than acquisition speed; we use a Pixis camera from Roper. Also, the camera should have a pixel size of 13 μm or less. We should point out that the signals from the newer EMCCD cameras are no better than these more standard (and cheaper) cooled CCD cameras. We typically use a 100x DIC objective. If one is interested in imaging with Cy5, one must be sure that the objective has sufficient light transmission at those longer wavelengths; this can sometimes require an IR coating. When mounting the cells, it is important to make sure that one uses #1 coverglass (18 x 18 mm, 1 ounce) and that the yeast are directly on the coverglass: do not adhere the yeast to the slide and then cover with coverglass. One can enhance the adherence of the yeast to the coverglass by coating the coverglass with poly-L-lysine (put fresh 1 mg/mL poly-L-lysine solution on the coverglass for 20 minutes, then suction off) or concanavalin A. It is also important to use #1 coverglass: we have found that even though most objectives are corrected for #1.5 coverglass, the mRNA spots are usually fuzzier and less distinct when imaged through #1.5 coverglass.

There are two somewhat standard procedures often employed during fluorescence microscopy that we have found interferes with our single mRNA signals. One of these is the use of commercial anti-fade mounting solutions, which tend to introduce a large background while also decreasing the fluorescent signals from target mRNAs. We recommend instead using the custom made GLOX solution or 2x SSC for imaging, being careful not to let the sample dry out. We also discourage using the standard practice of using a nail polish to seal the sample, as it introduces a background autofluorescence in the red channels that interferes with fluorescence from mRNA.

(vi). Image processing: Detecting diffraction limited mRNA spot

We have devised an algorithm that automates some fraction of the work involved in analyzing images obtained from the samples (Raj *et al.* 2008). The first step in our algorithm is applying a three-dimensional linear filter that is approximately a Gaussian convolved with a Laplacian to remove the non-uniform background while enhancing the signals from individual mRNA particles, thus enhancing the signal-to-noise ratio (SNR) (Fig. 6.2B). The full width at half maximum of this Gaussian corresponds to the optimal bandwidth of our filter, and depends on the size of the observed particle. This width is a fit parameter that we empirically adjust to maximize the SNR. However, even after filtering the images, they will contain some noise that requires thresholding to remove. In order to make a principled choice of threshold, we sweep over a range of possible values of the threshold, and plot the number of mRNAs detected at each value (Fig. 6.2C). Here, a single mRNA is defined as a collection of localized pixels (in the series of z-stacks) that form a connected component (Fig. 6.2D). We then typically find a plateau in this plot of the number of mRNAs counted as a function of the value of the threshold (i.e. increasing the threshold does not change the number of mRNAs counted) as seen in Fig. 6.2C. This implies that the signals from mRNAs are well separated from the background noise rather than a smooth "blending" in of the mRNA signals with the background noise. Indeed, the value of threshold chosen in this plateau range yielded mRNA counts nearly equal to the mRNA counts we obtained through an independent method in which we count by eye without the aid of automation. The software used for this purpose is available for download on *Nature method*'s supplementary information site for (Raj *et al.* 2008). One can also make measurements based on mRNA spot intensity, although we feel that great care must be taken in these situations. One issue is that the intensity depends on how precisely focused the spot is, although this can be ameliorated by taking a large number of closely spaced fluorescent stacks. Another problem with computing total or mean intensity is that the boundary of the mRNA is hard to define, and the ultimate intensity measurement will depend heavily on this somewhat arbitrary choice. One way to skirt the issue is to use the maximum intensity within a given spot, since this is independent of the size of the spot.

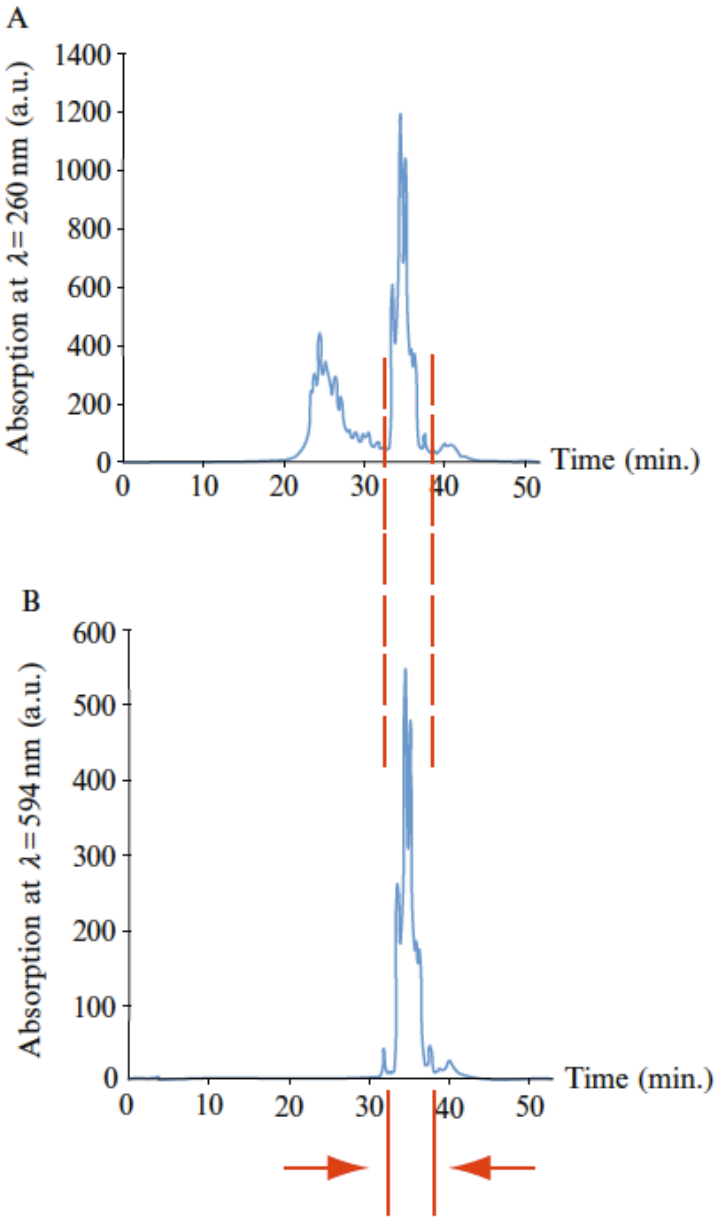


Figure 6.1: Chromatographs obtained during the HPLC purification of oligonucleotides coupled to the fluorophore (Alexa 594) from uncoupled oligonucleotides. (A) Absorption (at 260 nm, for DNA) curve as a function of time monitored during purification of probes coupled to Alexa 594 using HPLC. The first peak that appears between 20 and 30 minutes in this channel correspond to oligonucleotide probes that do not have Alexa 594 coupled to them. Eluate is not collected for the duration of this peak. (B) Absorption (at 594 nm, for Alexa 594) curve as a function of time. Both absorption curves (A) and (B) are obtained simultaneously for the duration of the HPLC run. Only one distinct peak appears in this channel, representing absorption by probes with Alexa 594 successfully coupled to them. This peak coincides with the second peak in the 260 nm channel shown in (A). Eluate is collected for the entire duration of this peak in the 594 nm channel. (Reprinted from H. Youk, A. Raj, and A. van Oudenaarden. *Methods in Enzymology* **470**, 429-446 (2010)).

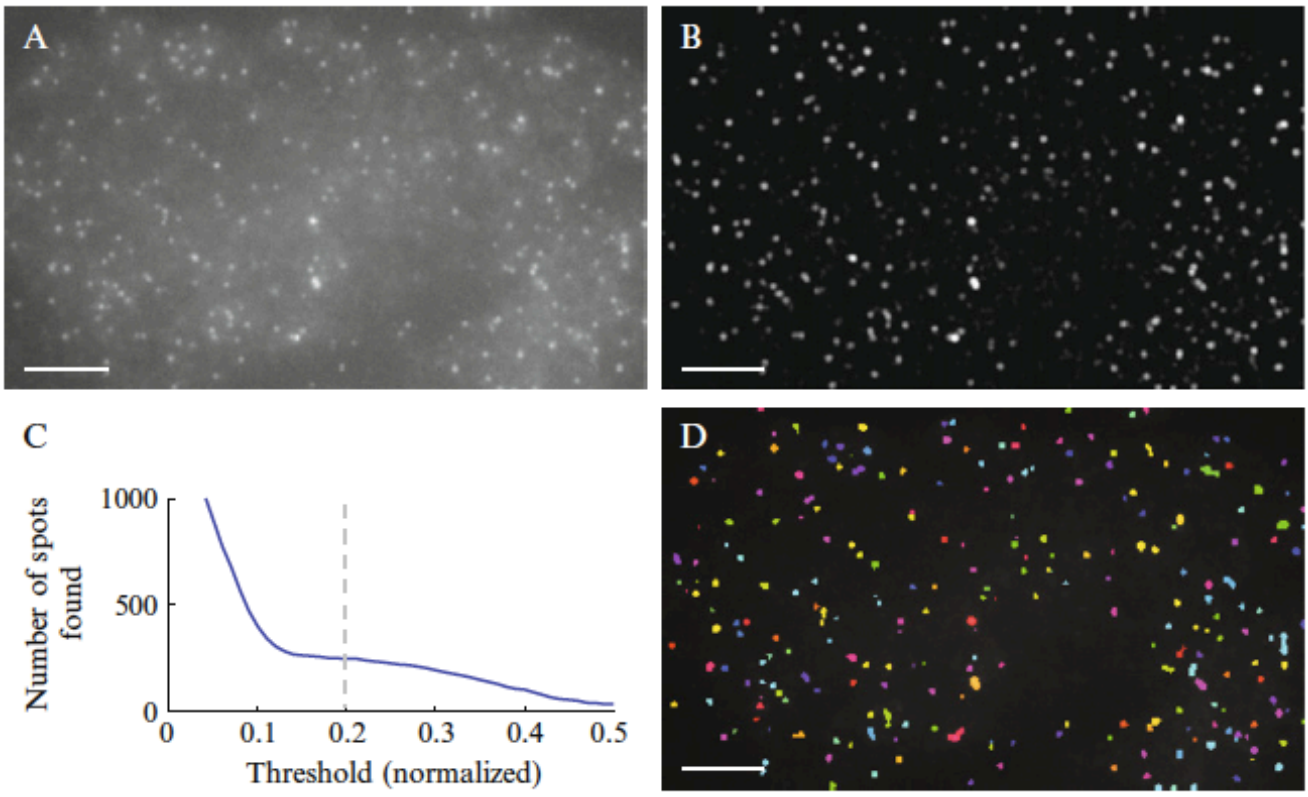


Figure 6.2: Example of mRNA spot detection algorithm applied to raw images of *FKBP5* mRNA particles in A549 cells induced with dexamethasone. (A) Raw image data showing *FKBP5* mRNA particles. (B) Upon applying a three-dimensional linear filter that is approximately a Gaussian convolved with a Laplacian to remove the non-uniform background while enhancing the signals from individual mRNA particles on the raw image shown in (A) the signal-to-noise ratio (SNR) is increased. (C) The number of spots counted as a function of the threshold value of the background after the application of the linear filter shows an existence of a plateau. This indicates a clear distinction between background fluorescence and actual mRNA spots. (D) Using the value of threshold shown as the grey line in (C), the raw image (A) has been transformed to an image in which each distinct computationally identified spot has been assigned a random color to facilitate visualization. (Reprinted from A. Raj, P. van den Bogaard, S. Rifkin, A. van Oudenaarden, and S. Tyagi. *Nature methods* **5**, 877-879 (2008)).

Bibliography for Chapter 6

[1.] Youk, H., Raj, A., and van Oudenaarden A. Imaging single mRNA molecule in yeast. *Methods in Enzymology* **470**, 429-446 (2010).

[2.] Raj, A. *et al.* Imaging individual mRNA molecules using multiple singly labeled probes. *Nature methods* **5**, 877 (2008).