Optimality, Robustness, and Noise in Metabolic Network Control

Muxing Chen
Gal Chechik
Daphne Koller

Department of Computer Science
Stanford University

May 18, 2007

Abstract

The existence of noise, or variability due to random events, in biological systems is a well-known phenomenon [Raser and O'Shea, 2005]. Here, we study the stochastic processes in a metabolic network controlling protein production which lead to variable protein levels across a culture of cells. Recent studies have shown that different enzymes exhibit differing levels of noise [Newman et al., 2006]. However, the sources, mechanisms, and regulation of this metabolic noise remains a largely unexplained question. To address these questions, we use optimization-based modeling to study the network response to enzyme variability in the metabolic network of *Saccharomyces cerevisiae*. Here, we use an optimization-based model of noise to construct a *compensation cost* measure of the effect of noise in reaction rates on the metabolic network. We then quantify the relationship between compensation cost and observed noise and discover that reactions which have a lower compensation cost, or smaller effect on the network under variation, also tend to have lower noise. This implies that the cell may optimize for some measure of protein "importance" in determining the noise tolerance for each enzyme. Finally, we explore possible mechanisms behind the regulation of noise, and find that burst frequency and burst size, the two main parameters controlling protein production [Cai et al., 2006] have strong inverse relationships to noise.

Contents

1 Introduction

1.1 Basics of Metabolic Networks

In order to survive, living organisms must convert energy from the surrounding environment into a form of usable energy. Specifically, the energy is adapted from the environment to achieve three major objectives: (1) to perform mechanical work in muscle contraction or other cellular movements, (2) to actively transport molecules and ions, and (3) to synthesize macromolecules and other bio-molecules from simple precursors [Berg et al., 2001]. To achieve these goals, cells use thousands of energy-transforming chemical reactions, collectively referred to as metabolism. We would like to study the metabolism of *S. cerevisiae* , or baker's yeast, to better understand the mechanisms behind the organisms survival.

1.1.1 Network Structure

Cellular metabolism can be viewed first from a structural standpoint. Each chemical reaction in metabolism converts a certain set of compounds, its substrates, into a set of different compounds, its products. These products may then be used as substrates by another reaction and converted into other compounds. This chain of events may continue until the desired end product is obtained. Such a sequence of reactions is known as a metabolic pathway. In cellular metabolism, these pathways are not independent. Rather, the metabolic pathways and chemical reactions in them are densely intertwined. The entire set of interconnected reactions is known as the metabolic network [Berg et al., 2001].

It is common to model the metabolic network *in silico* as a graph structure where each node represents a single compound and each edge represents a reaction converting one set of one or more compounds into a different set of one or more compounds [Varma and Palsson, 1994, Forster et al., 2003]. First, the *internal* metabolic reactions within the cell are modeled as in this graph structure. Then, we add *external* transport reactions, artificial reactions added to allow the cell to interact with the surrounding environment. These reactions, such as $CO_2xt \rightarrow CO_2$, represent substrates entering the cell or products leaving the cell. Finally, the growth of cellular biomass is represented as an output of the cell. We model the growth as an additional reaction which takes as its substrates a linear combination of the compounds

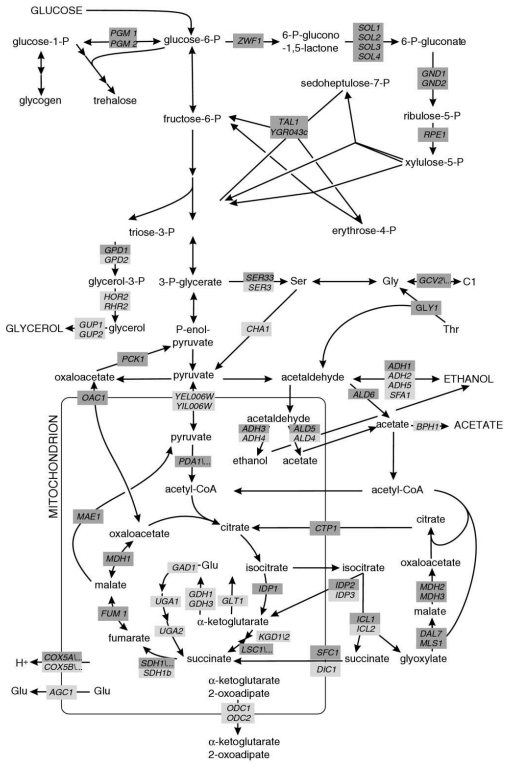


Figure 1: The subset of central metabolic reactions in the yeast metabolic network. Image was borrowed from [Blank et al., 2005].

needed for cellular growth and produces a single output, growth [Varma and Palsson, 1994, Edwards et al., 2001]. Later, we describe in detail the use of this network structure.

1.1.2 Network Function

The structure of a metabolic network only describes the set of achievable chemical reactions in the cell. However, at any given time, not all of these reactions are occurring. In fact, the cell often only "uses" a fraction of the available metabolic pathways [Papp and Pál, 2004, Giaever et al., 2002]. Thus, network structure is insufficient for studying metabolism. We must also consider the functional properties of the metabolic network.

A set of direct indicators of network activity are the reaction rates, or

fluxes, in the metabolic network. A reactions *flux* is defined as the rate of molecules per second at which the substrates are converted into the products [Varma and Palsson, 1994, Edwards et al., 2002]. Although it is possible for some reactions to occur spontaneously, in order to achieve significant reaction rates, the cell must produce special proteins called enzymes to catalyze, or speed up, reactions. Enzymes greatly accelerate the rate of reaction by binding to substrates and lowering the energy needed for the reaction [Berg et al., 2001]. A reaction may be catalyzed by one or more enzymes. Multiple enzymes catalyzing the same reaction are known as isozymes or isoenzymes. In simple terms, we can view a reactions flux as roughly dependent on the total number of enzymes catalyzing that reaction.

1.1.3 Network Regulation

The cells regulation of enzyme levels, and therefore fluxes, is controlled by three main factors: (1) the amounts of enzymes, (2) their catalytic activities, and (3) the accessibility of substrates. At the lowest levels, fluxes depend on the amount of substrates available. Substrates can be limited by both the availability of substrate molecules in the cells external living environment, as well as limits in the transfer of substrate molecules between different physical locations within the cell, such as between the cytosol and mitochondria. Fluxes also depend on the catalytic activities of enzymes, the amount by which it speeds up a reaction. Catalytic activity can be affected by feedback loops in the metabolic network. For example, in many biosynthetic pathways, feedback inhibition occurs, where the enzyme catalyzing the first reaction in the pathway is inhibited by the final product [Berg et al., 2001].

However, the main regulatory mechanism we will examine in this study is the high-level control pathway beginning from the cells DNA. The blueprint for constructing each enzyme, or protein, is stored by a gene in the organisms DNA. It takes two main steps to convert the blueprint into a working enzyme. First, the gene undergoes transcription, the process by which a DNA sequence is copied into mRNA, a type of information carrying molecule. Then, the gene undergoes translation, the process by which mRNA is decoded into a protein. The rate of transcription and translation of the genetic blueprint determines how much of the protein is produced. The amount of protein produced, in turn, largely determines the amount of flux in the catalyzed reactions [Berg et al., 2001].

What determines the rate of transcription and translation? Certainly,

the cell's environment plays a large role [Raser and O'Shea, 2005]. However, we are interested in the internal regulatory control: in a given environment, how is protein production determined? In this study, we interpret the genetic regulatory pathway as a control mechanism which the organism, or cell, "operates" by manipulating the rate of transcription and translation. We assume that the cell has developed as a result of evolutionary pressure some internal metabolic "objective function," a set of goals which maximize its chances of survival. It then attempts to "optimize" its objective function through flux control [Edwards et al., 2001]. By using this view of regulation, it will be possible to develop a computational model of cellular regulation.

1.2 Modeling Metabolic Networks

Generally, fluxes are very difficult to measure, especially on a genome-scale. Recently, a few labs have produced biological measurements of small subsets of fluxes in a few organisms; large scale flux measurements are generally not available [Blank et al., 2005]. Enzyme abundance and gene expression levels are somewhat easier to measure through biological experiments. Thus, our study of functional properties will use biological data at the level of expression and protein production, but in order to examine fluxes themselves, we must model the behavior of the flux network *in silico*.

1.2.1 Reconstructing the Network

Modeling a metabolic network first requires the reconstruction of the network structure. Recent advances in genome sequencing and large databases of known metabolic reactions have made possible full reconstructions of the metabolic network [Forster et al., 2003, Duarte et al., 2004]. First, a large reaction database is used to identify metabolic reactions and their stoichiometry: the substrates, products, and appropriate stoichiometric ratios of the compounds in the reaction. Then, the annotated genome is used to identify the genes, or open reading frames (ORFS), which encode the proteins that catalyze each reaction. The process is not trivial and for many reactions, the corresponding protein, gene, or both is still not known. After reconstruction, a complete network consisting of a list of reactions, the enzymes that catalyze them, and the genes that encode the enzymes is available for modeling [Forster et al., 2003, Duarte et al., 2004].

1.2.2 Flux Balance Analysis (FBA)

After the network structure is reconstructed, we can use a computational model to study fluxes. A widely used method of modeling genome-scale flux distributions is known as Flux Balance Analysis (FBA). FBA is used to model the metabolic network at a steady-state, when flux levels are in a dynamic equilibrium and therefore constant. FBA models metabolic fluxes by using a linear programming problem to solve for the set of feasible flux solutions [Edwards et al., 2002, Varma and Palsson, 1994, Schilling et al., 1999, Bonarius et al., 1997].

Linear programming, sometimes known as linear optimization, is the problem of maximizing (or minimizing) a linear function over a convex polyhedron specified by linear and non-negativity constraints [Schrijver, 1998]. The set of constraints describe a bounded polytope which represents the n dimensional space of possible solutions, where n is the number of variables in the solution space. The goal is to maximize the objective function within this constraint space. However, as is the case for flux networks, sometimes there is no unique maximum for the objective function [Edwards et al., 2002, Lee et al., 2000, Bonarius et al., 1997]. Nevertheless, linear programming provides a useful description of the flux capabilities of the metabolic network [Edwards et al., 2001].

FBA models the cell's regulatory system and manipulates the metabolic network *in silico* to optimize some metabolic objective function. Since the cell's true objective function is hidden from us, a simple objective function is used to approximate it. In FBA, we assume that the cell always optimizes its biomass growth. As explained in section 1.1.1, we can use the artificial biomass growth reaction as a linear objective function which represents growth [Varma and Palsson, 1994, Pramanik and Keasling, 1997]. Then, we can develop a set of linear constraints derived from reaction stoichiometry and environmental limitations. This allows us to solve a linear programming problem for the set of *optimal* flux solutions [Varma and Palsson, 1994, Edwards et al., 2002]. Note that the optimal solutions are the set of flux distributions which allow for the maximum possible level of growth given the constraints. The details of our implementation of FBA are described in section 3.1.1.

1.2.3 FBA Assumptions

The key assumption in FBA lies in the choice of the metabolic objective function [Varma and Palsson, 1994, Edwards et al., 2002]. The model assumes that the cell always attempts to maximize its biomass growth. In cells of micro-organisms such as *S. cerevisiae* which have undergone evolutionary pressures, it is reasonable to assume that the cell, or organism, will attempt to maximize growth. Furthermore, the accuracy of FBA in estimating growth and nutrient use has been experimentally validated against biological experiments [Edwards et al., 2001]. However, it has also been shown in some organisms such as *Bacillus Subtilis* that even micro-organisms subject to evolutionary pressures do not always maintain a flux distribution which optimizes growth [Fischer and Sauer, 2005]. In fact, Fischer et al. discovered that there exist knockout *B. Subtilis* strains (organisms with genes artificially removed) with higher growth levels than the wild-type strain. This implies that the wild-type *B. Subtilis* is not using growth as its only metabolic objective, since clearly higher growth is possible. What could be the motivation behind such metabolic behavior? In the next section, we describe an additional consideration for the metabolic objective function.

1.3 Noise in Metabolic Networks

One potential consideration behind the cell's metabolic regulation may be variation, or noise, in the control pathway. The control signals sent out by the cell from the highest levels may not be accurately received at the lowest levels of reaction flux. Rather, the signals are obscured by stochastic processes in each step of the regulatory pathway, from transcription to catalysis [Raser and O'Shea, 2005, Newman et al., 2006]. By the time a control signal to produce protein reaches the flux level, it has been altered by the noisy processes in the control path. Thus, a reaction's flux may experience random fluctuations as a result of this noise.

Here, we will refer to the variability in the metabolic network which results from stochastic processes in the control path as *metabolic noise*. There may also exist variation in network behavior which results from environmental differences between cells and differences in the internal states of cells [Raser and O'Shea, 2005]. However, studies have shown that noise in the control pathway is a significant source of variability in the network [Newman et al., 2006]. In this study, we will assume that the primary source of the observed

variation in the metabolic network is noise in the control pathway.

The variation in the metabolic flux network can be observed at several levels. Starting at the beginning of the regulatory pathway, we can observe variation in gene expression and mRNA expression. Then, we can observe variation in enzyme abundance [Raser and O'Shea, 2005, Newman et al., 2006]. Finally, although it is not currently measurable, variation in reaction fluxes themselves may be the most direct measurement of variation in metabolic functionality.

2 Problem Statement

2.1 Motivation

We would like to study the cells regulation of metabolic networks by examining the relationship between the effect of noise on the metabolic network and its control. Recent results have suggested that there may be biological motivations underlying the cells management of metabolic noise [Raser and O'Shea, 2005, Newman et al., 2006]. Newman et al. observe metabolic noise at the enzyme-level and quantify the variation in protein abundance between individual cells [Newman et al., 2006]. One important result by Newman et al. indicates that different enzymes exhibit different levels of noise, as measured by the coefficient of variation (CV) of enzyme abundance. Furthermore, Newman et al. demonstrate that an inverse relationship exists between enzyme noise and enzyme abundance. This inverse relationship between noise and abundance supports the idea that much of the observed noise may be due to stochastic variation in the regulatory pathway, and is not simply a result of environmental factors and differences internal states [Newman et al., 2006]. However, Newman et al. also observe that even when enzyme noise is normalized by abundance, we still see a distribution in the magnitude of *relative* noise (DM) in the set of metabolic enzymes (see section 3.3.2. In figure 2, we show the distribution of relative enzyme noise (DM) for all enzymes.

Why do different enzymes exhibit different degrees of relative noise? Organisms are generally known to evolve their survival strategies to an optimum due to evolutionary pressure [Varma and Palsson, 1994, Schilling et al., 1999, Edwards et al., 2001]. In the case of noise the existence of evolutionary pressure implies that the amount of noise in each enzyme may be "selected for" and optimized according to some underlying objective function. However,

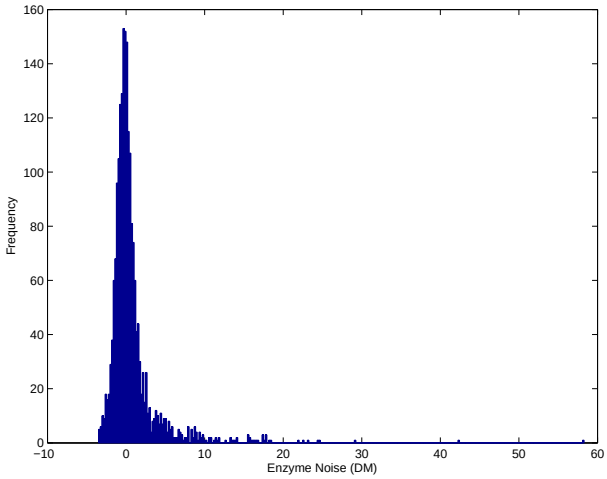


Figure 2: Distribution of relative enzyme noise (DM) in *S. cerevisiae* .

the motivation and mechanisms for optimizing metabolic noise remains an unexplained question.

Studies have examined this problem through different approaches. First, using a more biological approach, Newman et al. examine the role that protein functionality plays in determining the permitted level of variation [Newman et al., 2006]. They find that enzymes that allow the cell to adjust to environmental effects may have larger variation, due to both actual environmental differences and robustness to possible environmental change. On the other hand, enzymes which catalyze core functionality in the cell, for instance, protein production, may be more tightly regulated and therefore exhibit lower variation. Also, more generally, it may be advantageous for a population of organisms to maintain variation in certain enzymes to benefit the overall population [Newman et al., 2006, Raser and O'Shea, 2005]. For example, allowing greater variation in enzymes which respond to environmental shock would improve the entire populations probability of survival at the cost of individual cells growth.

In computational approaches, it has been demonstrated in FBA studies that organisms sometimes maintain suboptimal flux distributions and sacrifice growth for robustness [Fischer and Sauer, 2005]. Fischer et al. suggest that one component of the cell's metabolic objective may be to maintain robustness against environmental change. However, organisms may also main-

tain robustness against *internal* change, which may result from noise in the regulatory pathway. How does the flux network account for noise in regulation? In our study, we will use metabolic flux distributions to explore the "motivation" behind the cell's regulation of enzyme noise.

2.2 Approach

We assume that the cell has some "optimal" strategy which determines its regulation of metabolic noise. We propose to examine this strategy by relating the robustness of the metabolic network to noise in its reactions using a flux model. The classic approach to studying robustness is to use knockout analysis. This is done by conducting knockout experiments either *in vivo* or *in silico* where a gene (or reaction) is disabled, and the resulting phenotype is observed [Giaever et al., 2002, Papp and Pál, 2004]. We take a different approach and instead consider the effects of "non-disabling" reductions in enzyme or reaction functionality on the network.

In particular, we focus on exploring how the metabolic network responds to changes such as those which can be caused by noise. How large of an effect does variation in an enzyme or a reaction's functionality have on the network? High noise levels may often cause downward fluctuations in a reaction's flux. We want to develop a measurement of how these fluctuations may affect the network. We then use this measure to judge how"important" each reaction is to an organisms survival. We hypothesize that the importance of a reaction to an organisms survival will relate to the cell's regulation of that reaction's noise. Since downward variation can significantly impede a reaction's functionality, we propose that more important reactions should have lower variation.

To explore the relationship between noise and reaction importance, we create a model of metabolism and "observe" the cell's fluxes *in silico*. We then experiment with the network's robustness to change by exposing each reaction in the network to downward "noise." We quantify the network's response to these flux reductions, and construct a measure of the importance of reactions. We can then compare these measurements to observed reaction noise. Finally, we examine additional mechanisms underlying the relationship between robustness and noise.

3 Experiments and Results

3.1 Modeling Network Fluxes

In order to examine the effect of noise on cellular metabolism in *S. cerevisiae* , we must first establish the cell's original, "intended" flux distribution. The first step is to create a model of fluxes *in silico*. Since biological measurements of the full flux distribution do not exist, we must make several simplifying assumptions about the cell's fluxes in order to estimate them. First, we assume, as in FBA [Varma and Palsson, 1994, Edwards et al., 2002], that the cell chooses a flux distribution which maximizes growth. We also assume that the cell will use a minimal flux distribution to achieve maximum growth [Kuepfer et al., 2005, Blank et al., 2005]. In section 3.2.4, we will consider the effects of these assumptions on our results. Using these assumptions, we can compute a flux distribution to represent the cell's intended target flux distribution.

3.1.1 FBA Implementation

We first describe the implementation of the basic FBA model in *S. cerevisiae* which forms the basis for the later models that we will construct. In FBA, we attempt to maximize biomass growth subject to a set of linear mass balance and environmental constraints [Edwards et al., 2002, Varma and Palsson, 1994].

First, we view growth as an additional reaction v_{growth} which takes as its substrates a linear combination of k components of growth in the necessary ratios and generates a single product, biomass [Forster et al., 2003]. We add the following reaction to the network, where α_i is the coefficient of compound c_i:

$$\alpha_1 c_1 + \alpha_2 c_2 + \ldots + \alpha_k c_k \longrightarrow biomass \tag{1}$$

Since we assume that the cell wants to maximize its growth, we simply use the flux of the biomass growth reaction as the objective function to maximize. We then form several sets of linear constraints on the fluxes. We add one mass balance constraint for each node (compound) X in our network which requires that the amount of X produced by reactions involving X must equal

the amount of X consumed [Forster et al., 2003]. Thus, for each compound X, we have the equation:

$$v_{in} - v_{out} = \frac{dX}{dt} = 0 \tag{2}$$

We can summarize the entire set of mass balance equations by a single matrix equation. We represent the set of all reactions by a $m \times n$ matrix S, where each row represents a compound and each column in the matrix represents a reaction. Thus, entry $S_{i,j}$ represents the stoichiometric coefficient of compound i in reaction j. We can then write the set of mass balance constraints from equation 2 as:

$$S \bullet \vec{v} = \frac{dX}{dt} = 0 \tag{3}$$

where $\vec{v}$ is the $n \times 1$ vector of all reaction fluxes. Then, to model the availability of nutrients in the environment, we add additional artificial reactions to allow compounds to enter and leave the cell. The input reactions are then bounded by inequality constraints based on known physical data about the cell's environment. Additional information about the cell's physical environment or metabolic capabilities is similarly added as constraints. For instance, non-reversible reaction fluxes are constrained to be greater than or equal to 0. Thus, combining 3 and 1, our final optimization problem is constructed as follows [Varma and Palsson, 1994, Forster et al., 2003]:

$$\begin{aligned} &\textit{maximize} && v_{growth} = \alpha_1 c_1 + \alpha_2 c_2 + \ldots + \alpha_3 c_3 \\ &\textit{subject to} && S \bullet \vec{v} = \vec{0} \\ & && \vec{v}_{input} \leq \vec{ub}_{input} \\ & && \vec{v}_{non-rev} \geq \vec{0} \end{aligned} \tag{4}$$

where $\vec{ub}_{input}$ represents the vector of upper bounds on the input fluxes. We may then solve this linear optimization problem using any of a number of well-researched algorithms. In our work, we construct an FBA problem and solve it using the simplex algorithm implemented by the CPLEX optimization toolbox. The output of the optimization problem is a set of fluxes which achieve maximal growth. However, the flux solution may not be unique; there may be several different flux distributions which can be used to achieve the same levels of growth. Thus, FBA cannot be used to determine a unique flux solution [Bonarius et al., 1997, Lee et al., 2000].

3.1.2 Incorporating Biological Information

FBA is not sufficient for determining a unique biologically meaningful distribution of fluxes [Bonarius et al., 1997, Lee et al., 2000]. However, variants of FBA have been developed which can be adapted to produce more meaningful flux solution [Segre et al., 2002, Shlomi et al., 2005].

One variant of FBA which was developed to analyze knockout mutations is the minimization of metabolic adjustment (MOMA) method [Segre et al., 2002]. The principal behind MOMA is that organisms which have been genetically engineered with knocked-out genes would have not been subject to the same evolutionary pressures as the wild-type organism. Thus, it would be unreasonable to assume that knockout mutants will have adapted a flux distribution which maximizes their growth. Rather, Segre et al. argue that a knockout mutant will undergo a minimal flux redistribution with respect to the flux configuration of the original wild-type organism. Thus, given a set of fluxes $\vec{v}^{wt}$ for the wild-type, they compute the fluxes $\vec{v}$ for a knockout by minimizing the Euclidean distance between the knockout and the wild type fluxes:

$$\begin{aligned} &minimize \quad D(wt, ko) = \sqrt{\sum_{i=1}^{n}(v_i - v_i^{wt})^2} \\ &subject\,to \qquad S \bullet \vec{v} = \vec{0} \\ &\qquad\qquad \vec{v}_{input} \leq \vec{ub}_{input} \\ &\qquad\qquad \vec{v}_{non-rev} \geq \vec{0} \\ &\qquad\qquad v_{knockout} = 0 \end{aligned} \tag{5}$$

This method has been shown to produce more accurate flux results than FBA for knockout mutants. We can adapt this principal to compute a more accurate flux distribution for the wild-type as well. MOMA attempts to minimize the distance between the mutant fluxes and some known flux distribution for the wild-type. Since no known biological measurements of the flux distribution for the wild-type exist, MOMA uses FBA estimates as the "known" flux distribution. However, recently biologists have been able to measure small subsets of flux distributions *in vivo*. Thus, studies by Kuepfer et al have adapted the MOMA strategy in order to compute a more accurate wild-type flux distribution relative to the subset of measured fluxes [Kuepfer

et al., 2005]. We use a subset of 30 reactions whose fluxes have been measured *in vivo* as our known fluxes and compute the wild-type flux distribution such that the fluxes in the measured subset of reactions undergo a minimal flux redistribution relative to the measured fluxes [Kuepfer et al., 2005].

The issue with this approach is that the subset of 30 measured fluxes v_{meas} is not large enough to determine all 672 fluxes in the network. We must decide what to do with the remaining flexibility in the flux solution space. Here, we follow the assumption made by Kuepfer et al. that the cell will use the minimal fluxes necessary to achieve maximum growth [Kuepfer et al., 2005]. This assumption is needed to remove futile cycles from the network. There are cycles in the network where a loop involving two or more reactions of the form $v_1 : c_1 \rightarrow c_2$ and $v_2 : c_2 \rightarrow c_1$ develops. Here, unless fluxes are minimized, v_1 and v_2 could assume an arbitrarily large amount of flux without violating any constraints. Thus, we set such cycles at the minimum flux level needed to maintain growth. Although this assumption is not accurate, we show in section 3.2.4 that it is reasonable enough for our purposes of studying robustness.

We thus calculate the wild-type flux distribution by computing the minimal fluxes needed to both achieve the maximum growth and to maintain the measured fluxes v_{meas} within one percent of their measurements. We adapt this as shown by Kuepfer et al. into an L1 (linear) optimization problem by using substitution variables to minimize the sum of the absolute values of the fluxes [Kuepfer et al., 2005]. Thus, we write:

$$\begin{aligned} minimize \quad & \sum_{i=1}^{n}(x_i + y_i) \\ subject\,to \quad & x_i - y_i = v_i - f_i \\ & S \bullet \vec{v} = \vec{0} \\ & \vec{v}_{non-rev} \geq \vec{0} \\ & (1-\epsilon) \times \vec{f}_{meas} \leq \vec{v}_{meas} \leq \vec{f}_{meas} \times (1+\epsilon) \\ & where\,\epsilon = 0.01 \end{aligned} \tag{6}$$

where $\vec{f}_{meas}$ is the vector of biological flux measurements in which any flux which has no flux measurement is set to 0. Using this minimization, we obtain a flux solution which conforms to the known biological measurements, optimizes growth, and minimizes extraneous fluxes. Although this flux solution will be generally biased downwards due to the minimization of fluxes,

we can interpret it as a biased estimate of the cell's observed fluxes. Using this "observation," we can now examine how the cell responds to changes in its fluxes, which may result from either environmental shocks or internal noise.

3.2 Modeling Network Robustness

To explore the cell's strategy for controlling levels of noise in metabolic networks, we would like to explore the effect of noise in each reaction and catalysts (enzymes) in the network. First, we use the FBA assumption that cellular growth is the main metabolic objective. Thus, we are interested in the robustness of the network's optimal growth to fluctuations in reaction fluxes: how important is each reaction to maintaining an optimal solution to the cell's objective function, or growth? Past studies have answered this question by using hard constraints which knockout the flux completely [Giaever et al., 2002, Blank et al., 2005]. We instead develop a "soft" measure of the effect of downward variation in a reaction on the network, which we call the *compensation cost* of a reaction. We then examine how each reaction's compensation cost compares to its noise.

3.2.1 Measuring Compensation Distance

To compute the measure of compensation cost for each reaction, we first construct a "compensation curve" plotting the cost of maintaining optimal growth as a reaction is varied. We note that we are only concerned with downward variation, because increasing flux in a reaction will never make it more "difficult" to maintain optimal growth. We begin by computing an "observed" flux distribution by using our flux minimization estimate of fluxes (section 3.1.2). Then, we take each flux v_i and vary it below it's observed level. We then measure how much the rest of the network must compensate in order to maintain a high percentage (90%) of the optimal growth level.

To construct the compensation curves, we solve a set of optimization problems for each reaction r of interest. To compute the compensation curve for a target reaction r^* with flux v^*, we enforce an upper bound constraint on v^* and reduce the upper bound in increments. For each level of $v*$, we then solve for the flux distribution which maintains near-optimal growth and is the closest (using Euclidean distance) to the original "observed" distribution.

Specifically, we add the following constraints to each optimization problem. First, the other internal fluxes are not bounded, since we assume that the cell will be able to re-route fluxes in response to perturbations in one flux (not always biologically valid). The inputs to the network are bounded between zero and their observed flux levels (from previous section) in order to simulate a static external environment, since we only want to observe internal metabolic changes. Finally, the growth flux is lower bounded to 90% of the optimal growth to require that the cell remain in a near-optimal metabolic state. We also maintain the mass balance and reversibility constraints. Thus, for each target flux $v^* \, \epsilon \, \vec{v}$, we write the following set of optimization problems for the set of reduction factors $\delta \, \epsilon \, D = \{1, 0.95, 0.9, \ldots, 0\}$:

$$
\begin{aligned}
&foreach \; \delta \, \epsilon \, D: \\
&\quad minimize && \sum_{i=1}^{n} (v_i - f_i)^2 \\
&\quad subject\,to && S \bullet \vec{v} = \vec{0} \\
& && \vec{v}_{non-rev} \geq \vec{0} \\
& && \vec{v}_{input} \leq \vec{ub}_{input} \\
& && v_{growth} \geq 0.9 \times f_{growth} \\
& && v^* \leq \delta \times f^*
\end{aligned}
\tag{7}
$$

where v^* is the target reaction for which we are computing the compensation cost, $\vec{ub}_{input}$ is the vector of upper bounds on the input, $\vec{f}$ is the vector of "observed" fluxes from the minimization estimate, and f_{growth} is the optimal growth flux. Thus, we obtain a set of minimum flux distances $Dist = d_1, d_2, \ldots, d_m$ where d_j corresponds to the minimized distance d computed for $\delta_j \, \epsilon \, D$ and m is the number of δ increments. Importantly, we must now normalize each distance d_i by the original flux of the target reaction f^*. This is done to account for the magnitude of the original flux f^* of the target reaction r. Clearly, a larger f^* would require a larger amount of compensation in the network. However, we only want to consider the relative levels of network compensation between reactions.

Finally, given a set of distances, we can plot the normalized network compensation distance $dist_j = \frac{d_j}{f^*}$ for $j = 1 \ldots m$ as a function of the downward variation $(1 - \delta_j)$ for each target reaction. This gives us a compensation cost curve for each reaction. We focus on analyzing the compensation cost curves

for target reactions in the set of 30 measured reactions. These reactions are less affected by the minimization bias, since the "starting point" f^* of the target reaction is not computed from the flux minimization. We show three examples of the types of curves we obtain in figure 3. From these curves, we will calculate a measure of each reaction's *compensation cost*.

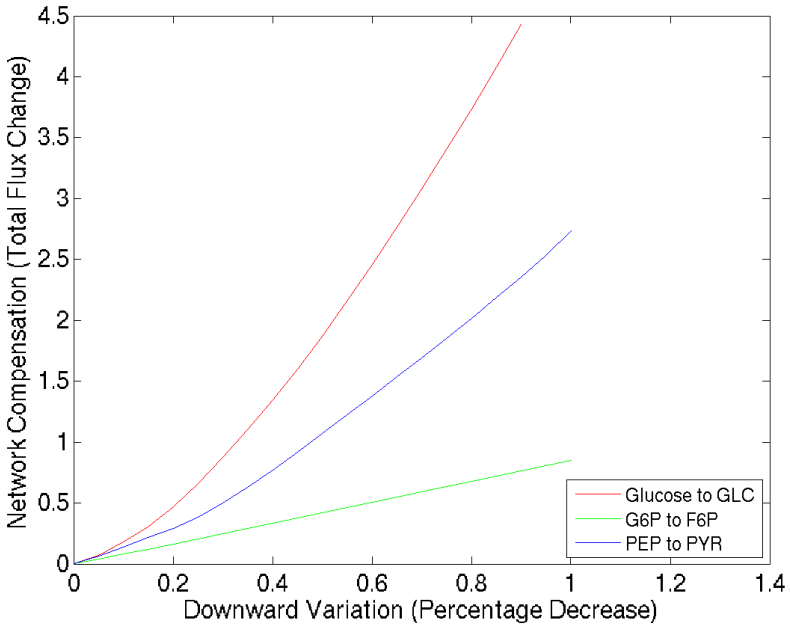


Figure 3: Network compensation distance curves for three reactions in *S. cerevisiae* .

3.2.2 Mechanisms of Compensation Distance

In general, when a target reaction r^*'s flux v^* is varied below its observed level, the network must re-route fluxes through alternative pathways to produce the compounds which r^* originally supplied [Blank et al., 2005, Segre et al., 2002]. We develop a measure of the magnitude of the flux re-routing, the *compensation distance*. In general, the curve of compensation distance as a function of downward variation in v^* becomes a function with linear segments of increasing slope, as seen in figure 3.

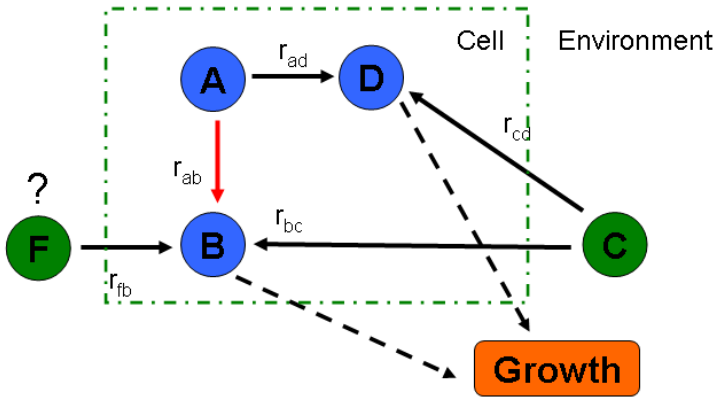


Figure 4: Diagram of a hypothetical section of the metabolic network. Nodes represent compounds and arrows represent pathways of one or more reactions.

Individual segments of the compensation distance curve often reveal a linear relationship between compensation distance and downward variation. This may occur when there is a direct redundant pathway between r^*'s substrates and products, since every decrease in v^* will cause a corresponding increase in the redundant pathway's flux. However, for many reactions, the slope of the line also changes at many points along the x axis. This may be explained by the network's use of additional alternative pathways to compensate for r^* when the original pathway becomes insufficient. We consider a simplified example of network compensation in figure 4. Here, each node represents a set of one or more compounds, and each arrow represents a pathway of one or more reactions converting the compounds. We also ignore the stoichiometric ratios between compounds and assume that each set of compounds is required in a 1:1 ratio.

We consider the case when our target reaction $r^* = r_{ab}$. When reaction r_{ab}'s flux is decreased, the network may first compensate for the reduced production of compound B by increasing the flux in r_{cb} to convert C to B. The "extra" A may the be used to produce the amount of D originally produced by C. However, C is a compound which is taken from the environment. Thus, if we decrease r_{ab} further, we may need to produce more B than we can produce using C, since the environment may not supply enough of com-

pound C. In our model, the constraint on the amount of substrate D is represented by an upper bound on the external inputs to the network.

At this point, if there is another compound F which can supply B in the same way, the network will "switch" and use this alternative pathway. At this point, the slope of the distance function must increase, since we are increasing the flux in a second alternative pathway. Note that the second pathway must be longer than the first, since we are always minimizing distance. Thus, we eventually end up with a compensation distance curve with linear segments of increasing slope.

In the above example, if there did not exist a second redundant pathway through some F in the network, the network would become unable to sustain optimal growth for further decreases in r. In plots of these reactions, the curve becomes infeasible after some level of decrease. This is shown in reaction $Glucose \rightarrow GLC$ in figure 3, where a reduction of the reaction beyond approximately 90% makes it impossible for the network to maintain optimal growth.

3.2.3 Measuring Compensation Cost

Using the compensation distance curves, we can quantify the magnitude of the effect of a reaction's variation on the network. We can measure how "difficult" it is for a network to compensate for a decrease in the target reaction. The compensation distance is related to the number of fluxes which must be altered (usually increased) to compensate for variance. Thus, a larger compensation distance implies that variance in a reaction is more costly to a cell in terms of both energy use and feasibility We quantify the cost of a reaction's variance on the network using a *compensation cost* measure, which we measure as the average gradient across the entire curve (equation 8). A larger gradient indicates that the reduction in a reaction's flux requires more changes in the network to maintain optimality, while a smaller gradient suggests that a reaction has shorter alternative pathways available.

$$compensation\,cost = \frac{dist_{last} - dist_{first}}{\delta_{last} - \delta_{first}} \tag{8}$$

The compensation cost can be viewed as a measure of how noise in each reaction would affect the metabolic network. In particular, it measures how "important" each reaction is to the network by quantifying how "difficult" it is for the cell to maintain near-optimal growth when that reaction experiences

downward noise. We note that the compensation cost is independent of the target flux's magnitude, since it is normalized by the amount of flux decrease. As a result, it depends on the structure of the network and the magnitudes of the other fluxes in the network. In the next section, we consider the effect of our flux estimation on the compensation cost. Then, we will examine how compensation cost is related to reaction noise.

3.2.4 Effect of Flux Estimation Bias

Since we are estimating our flux distribution by using the minimization described in section 3.1.2, we expect that our estimated fluxes will have a downward bias [Kuepfer et al., 2005, Blank et al., 2005]. However, we note that we do not minimize *each* flux independently; rather, we minimize the sum of the fluxes. Thus, we expect the overall distribution to be lower than the true flux distribution, but we do not consider each individual estimated flux to be a minimum; rather, it is simply a downward biased estimator of the true flux.

First, we verify the downward bias in our estimation by using a leave-one-out (LOO) cross-validation on our subset of 30 measured reactions. To do so, we compute the flux minimization estimate of the full flux distribution, but leave one flux v_i of the 30 measured fluxes *out* of the measurement $\vec{f}$ constraint. We instead obtain an estimate of v_i from the estimated flux distribution, and compare it with its true measurement. We repeat this for all $v_i \, \epsilon \, \vec{v}_{meas}$. We thus obtain the plot in figure 5, which compares the true fluxes of the measured subset of reactions against their estimated values. The plot shows a strong correlation ($r = 0.92$ and $p < 10^{15}$) between the true and estimated fluxes, but also demonstrates the expected downward bias in the estimated fluxes. The estimated fluxes have a median downward bias of approximately 44% below their true values.

We would like to know what impact the estimation bias has on the final compensation costs. To observe the effect of having a higher true flux distribution, we perform a rough simulation as follows. We assume, based on empirical results, that the distribution for the true flux for each reaction is centered 60% above the estimated flux, with a standard deviation of 60%. We then sample a flux for each reaction (independently) from the distribution of "true" fluxes to obtain a full set of "permuted" flux estimates. Using this set of permuted fluxes, we then re-compute the compensation cost measure for each target reaction in the measured subset. We repeat this process for

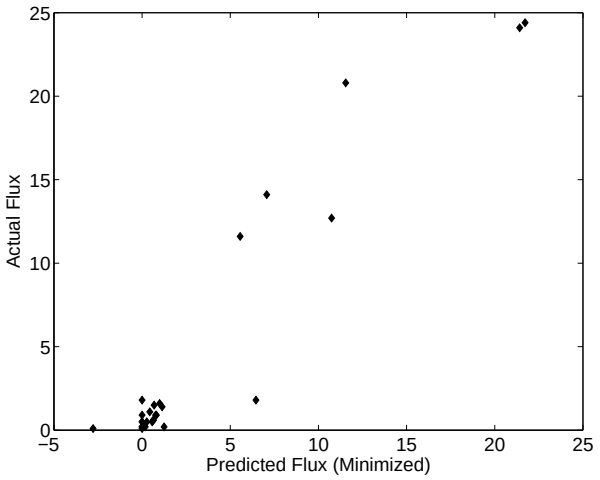


Figure 5: Flux minimization estimates vs. flux measurements from a LOO cross validation.

$n = 100$ samples, and obtain a set of permuted compensation costs. We then compare these compensation costs to the costs computed from the original flux distribution and observe their correlation in figure 6.

We see that, aside from the three outliers in the lower right, there is a fairly clear positive relationship between the compensation costs calculated from the flux minimization estimates and the compensation costs calculated from the permuted flux distributions. The correlation coefficient between the average compensation cost computed from the permuted samples and the original compensation cost is $r = 0.5856$, $p = 0.0021$. We note that the permuted flux samples tend to have lower compensation cost because the permuted fluxes are on average higher than the minimized fluxes; thus, there is already "extra" flux available for compensation in the permuted distributions. This implies that the compensation costs calculated using the "true" flux distribution would have a strong correlation with the estimated compensation costs. Thus, although a better flux estimation would clearly benefit our analysis, the biased minimized fluxes may still provide fairly realistic estimations of the compensation cost.

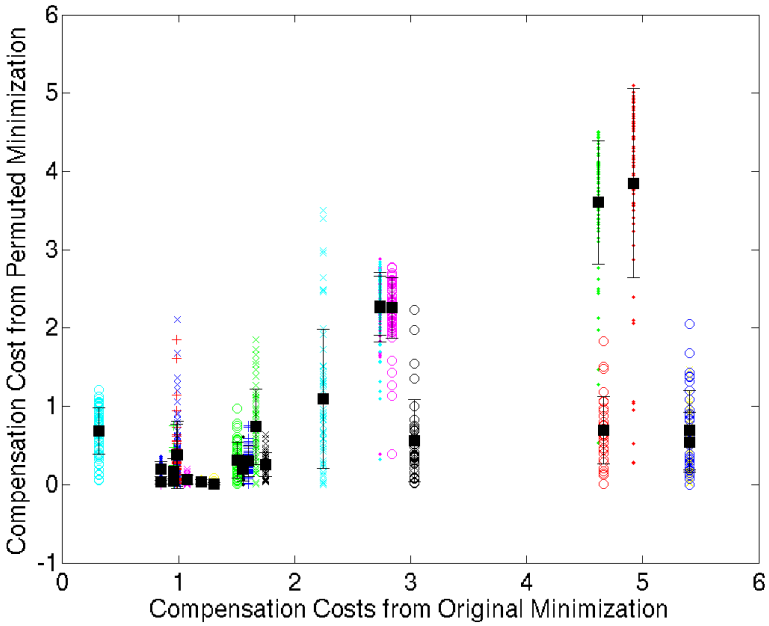


Figure 6: Compensation costs from original minimized flux estimates vs. sample compensation costs from permuted flux distributions. Here, each set of colored symbols represent a set of permuted compensation cost samples for a single target reaction. Each square represents the mean permuted compensation cost of the given target reaction and the bars represent the standard deviation of the permuted compensation cost samples.

3.3 Computing Reaction Noise

The compensation cost gives us a measure of how the network is affected by the noise in each reaction. Now, we would like to relate that to the observed noise in *S. cerevisiae* cells. Recent studies have used single-cell proteomic analysis on *S. cerevisiae* to measure variation in single cell enzyme abundance levels [Newman et al., 2006]. As is explained in section 1.3, Newman et al. hypothesize that much of the noise in enzyme abundance levels may be explained by internal stochastic processes. Thus, we can interpret enzyme-level noise as noise which is regulated by the cell's metabolic control. Thus, we will use an enzyme-level measurement of variability [Newman et al., 2006]

to compute an estimated noise level for each reaction. We can then compare the cell's regulation of noise against the network's compensation costs.

3.3.1 Enzyme Noise

We first review Newman et al.'s computation of enzyme noise, and then extend it to compute noise at the reaction level. First, they compute the coefficient of variation (CV) using abundance and variation measurements for each enzyme. For enzyme e, they compute $CV(e)$ as:

$$CV(e) = \frac{\sigma_e}{\mu_e} \tag{9}$$

where σ_e is the standard deviation of e's abundance and μ_e is the mean. Newman et. al. show that the $CV(e)$ of an enzyme is shown to be inversely related to its abundance μ_e. This is a typical characteristic of stochastic processes internal to a call such as the production and destruction of protein molecules. However, we would like to observe relative noise levels between different enzymes; thus, we use Newman's computation of distance to the median (DM) as a measure of noise normalized by abundance. To compute DM, we first plot $\frac{1}{\mu}$ vs. CV and compute a set of running medians of the CV's across this plot using a window of size 60. Then, we fit a line to the plot of $\frac{1}{\mu}$ vs $median(CV)$ and compute each DM as follows:

$$DM(e) = CV(e) - median(CV(E)) \tag{10}$$

where E is the set of 60 fluxes which are centered around e by $\frac{1}{\mu}$. The DM output for each enzyme represents its relative noise level normalized for abundance.

3.3.2 Reaction Noise

Since multiple enzymes can catalyze a reaction (see section 1.1.2), a reaction's noise may be influenced by variation in more than one enzyme [Berg et al., 2001]. Thus, we must combine the enzyme-level noise measurements into appropriate reaction-level noise estimates. To do so, we will combine the abundance and variation measurements for the set of enzymes E catalyzing a reaction r_i to compute the coefficient of variation for that reaction $CV(r_i)$. Then, given a set of $CV(r_i)$'s for all reactions, we can repeat the above DM

computation on the reaction-level. We first compute reaction-level CV's as follows:

$$CV(r) = \frac{\sqrt{\sum_{e \,\epsilon\, E} \sigma_e^2}}{\sum_{e \,\epsilon\, E} \mu_e} \tag{11}$$

We then repeat the DM computation as above for reaction CV's by sorting by $\frac{1}{\mu}$, fitting a line to the median of the CV's, and computing the distances to the median for each reaction: $DM(r) = CV(r) - median(CV(r))$. Our final output is a set of noise measurements $DM(r)$ for all reactions r. Each reaction's noise measurement represents the level of noise regulated for that reaction by the cell's metabolic control.

3.4 Robustness and Noise

We can now compare the effect of each reaction's noise on the network with the cell's control of that reaction's noise. We use the compensation cost as a measurement of the network response to a reaction's noise, and the reaction noise ($DM(r)$) computed in section 3.3.2 as an estimate of the noise permitted by the cell. We consider the question: are reactions with a higher compensation cost regulated more strictly? To answer this question, we plot the compensation cost (CC) of each reaction against its noise (DM) in figure 7.

3.4.1 Results

We first focus on comparing the compensation cost and noise for the subset of measured reactions r_{meas}. Since the target flux measurements for the measured reactions were not estimated through the flux minimization (see section 3.1.2), they represent a less biased measurement of the flux "starting points." However, we note that even these compensation costs are still affected by the minimization, since they still depend on the changes in the full network.

In figure 7, we see a significant inverse relationship between CC and DM, with a Pearson correlation score of $r = -0.525$ and $p = 0.0142$. Thus, we observe that reactions whose downward variation causes larger changes in the network tend to have lower noise.

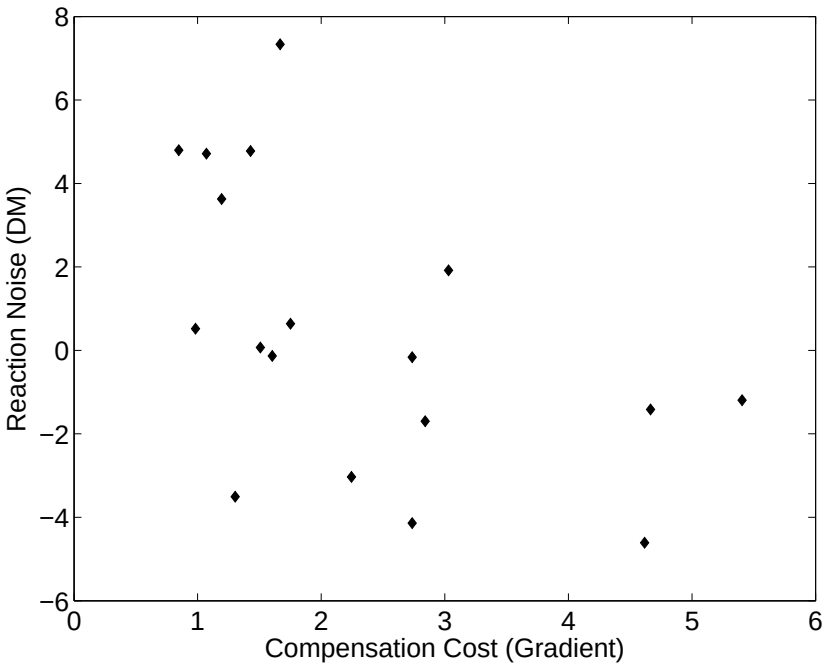
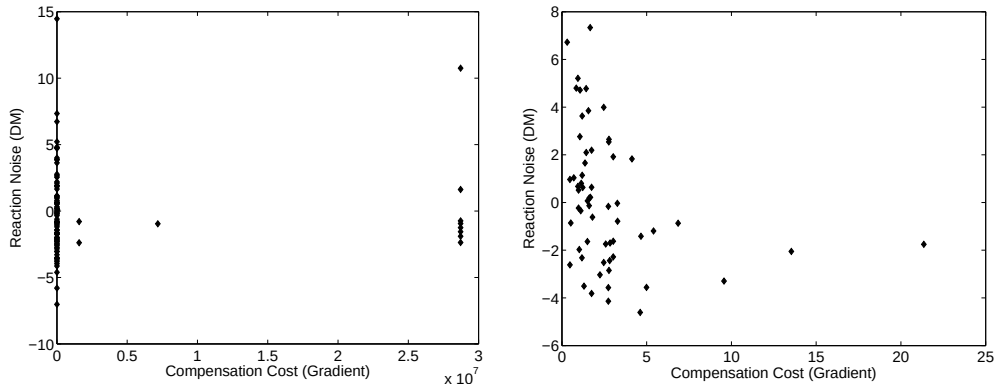


Figure 7: Compensation cost vs. reaction-level noise (DM) for the subset of measured reactions.

We then consider the same comparison across all fluxes. Since our flux "starting points" in this case are obtained through the flux minimization, we know that they will be biased downwards. This bias may affect the resulting compensation cost. First, we see in figure 8(a) that there is no apparent relationship between the compensation cost and the reaction noise ($p = 0.646$) across the full network. However, the reactions on the graph appear to be clearly separable into two distinct sets: those with extremely high compensation cost ($> 10^7$), and those with "normal" compensation costs. Upon closer examination, the reactions with extremely high compensation cost all become "un-compensate-able": at some point, the network can no longer compensate for reductions in the fluxes of these reactions. The compensation distance of these reactions becomes extremely high before reductions in their flux cause optimal growth to become infeasible. Thus, we would expect that these reactions have high compensation cost. Here, we will refer to these reactions as "essential" reactions.



(a) Compensation cost vs. reaction-level noise (DM) for all reactions.

(b) Compensation cost vs. reaction-level noise (DM) for all reactions high compensation cost outliers.

Figure 8: Compensation cost vs. reaction-level noise.

Set of Reactions	Correlation Coeff (r)	P-value
Measured reactions only	-0.525	0.0142
All reactions	0.0485	0.6463
All non-essential reactions	-0.3029	0.0186
All essential reactions	0.1343	0.4637

Table 1: Summary of correlations between compensation cost and reaction-level noise (DM).

We would like to examine these sets of reactions separately. Thus, we filter out essential reactions and compare the compensation cost and noise of the remaining reactions in figure 8(b). Here, we see that among non-essential reactions, there is a significant negative relationship between compensation cost and reaction noise. This is confirmed by the correlation coefficient where $r = -0.3029$ and $p = 0.0186$. Finally, we note that there is no significant relationship between compensation cost and reaction noise in essential reactions. This is unsurprising, since these reducing the flux of these reactions essentially causes an uniformly infinitely large compensation distance. We summarize the correlations in table 1.

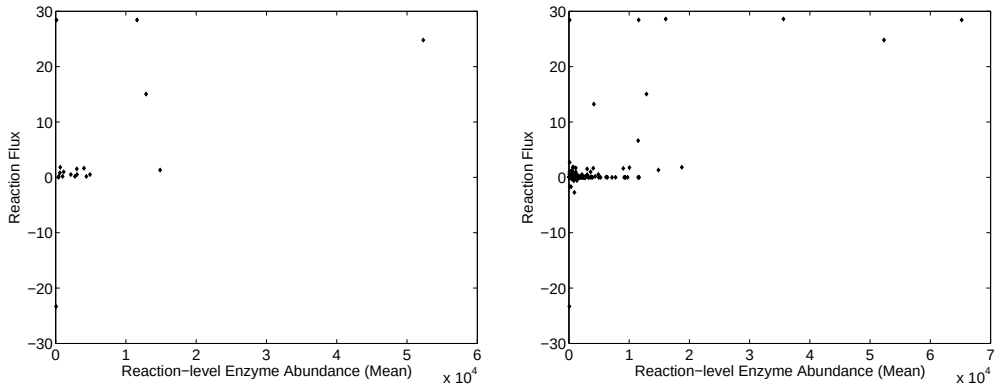
3.4.2 Interpretation

We can interpret the relationship between compensation cost and noise in several different ways. First, we can consider compensation cost as a measure of the "importance" of a reaction to maintaining optimal growth. A reaction with many redundant pathways, such as the reaction $G6P \rightarrow F6P$ in figure 3, will have a low compensation cost, as it will be easy to duplicate that reaction's functionality [Kuepfer et al., 2005]. Thus, such a reaction is not uniquely essential to maintaining the network's optimality. However, a reaction with a high compensation cost, such the reaction $Glucose \rightarrow GLC$ in figure 3, may have fewer direct redundant pathways. Instead, to compensate for the variability, the network must re-route fluxes through longer alternative pathway(s) change the flux distribution in many different pathways. This may be disruptive to cellular function, and may not even be possible *in vivo* [Segre et al., 2002, Shlomi et al., 2005]. Thus, the flux in reactions with high compensation cost may be considered more "important" to achieving metabolic objectives such as growth. Therefore, it may not be surprising that the cell allows more noise in less "important" reactions, since variability in these reactions will most likely not affect growth, while strictly regulating the noise in more "important" reactions to prevent the disruption of growth.

We can also consider the implications of the relationship between compensation cost and noise in terms of energy. Compensation cost is primarily related to the number of reactions whose flux must change in order to maintain optimal growth (since we normalize for the magnitude of the flux). Since we are decreasing flux in the varied reaction r, we can assume that most of the flux changes will be flux increases. Thus, the cell will be forced to increase reaction rates in the alternative pathway(s) to compensate for variability, which requires energy ?cite. Reactions with larger compensation costs will require increases in more reaction rates and therefore have a higher energy cost of compensation. Thus, downward variability in high compensation cost reactions would require considerably more energy to prevent disruption of growth. Furthermore, we note that *in vivo* the cell may not always be able to increase reaction rates in any given pathway [Segre et al., 2002, Shlomi et al., 2005]. Thus, we would expect reactions with a higher compensation cost to have lower noise.

Variables	Correlation Coeff (r)	P-value
Flux vs. abundance (measured reactions)	0.5093	0.0259
Compensation cost vs. Flux (measured reactions)	0.2695	0.1927
Noise (DM) vs. abundance (all reactions)	-0.1213	0.0565
Noise (DM) vs. abundance (measured reactions)	-0.3708	0.118

Table 2: Summary of correlations between confounding variables.



(a) Flux vs. abundance in the measured reactions.

(b) Flux vs. abundance in all estimated fluxes.

Figure 9: Flux vs reaction-level enzyme abundance.

3.5 Confounding Variables

We would like to further examine the relationship between compensation cost and noise. First, we consider the possibility of confounding variables such as protein abundance or flux affecting the relationship between compensation cost and noise. The correlation coefficients and p-values between the different variables are summarized in table 2.

We first account for possible confounding variables in the relationship between compensation cost and reaction noise. First, since fluxes are naturally related to enzyme abundance in the cell [Berg et al., 2001], it is possible that this relationship is a factor in the relationship between compensation cost, which is based on flux, and reaction noise, which is based on enzyme variation.

First, we examine the relationship between flux and enzyme abundance

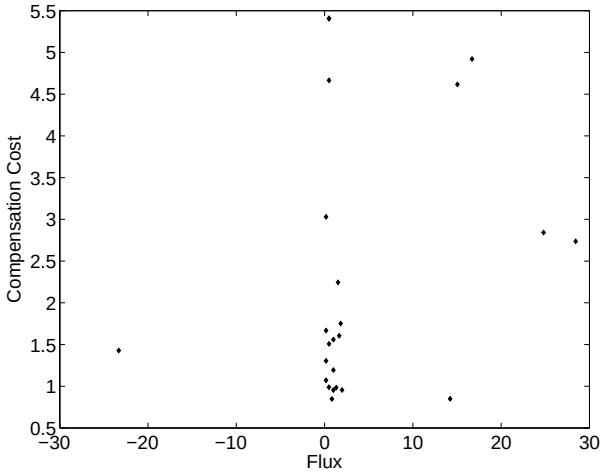
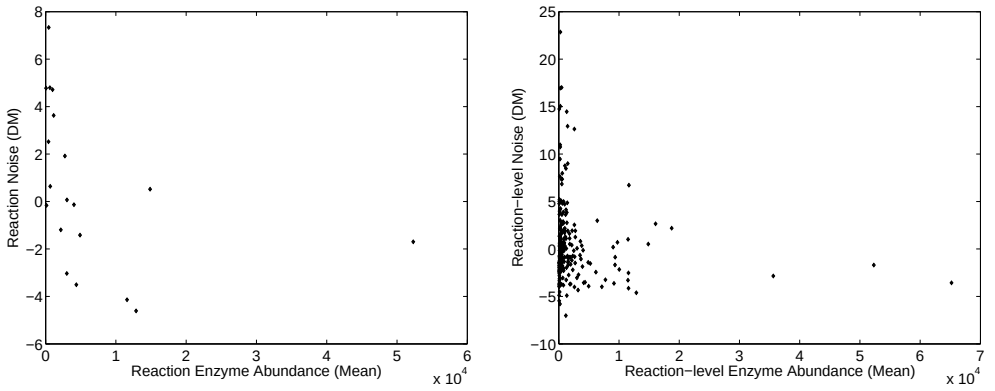


Figure 10: Compensation cost vs. flux for measured reactions.

for the measured subset of reactions in figure 9(a). The relationship has a strong positive correlation with $r = 0.5093$ and $p = 0.0259$. However, if we remove the single outlier in the upper right of the plot, the p-value increases to $p = 0.1131$, and if we remove the other two outliers in the upper right, the correlation disappears. A similar relationship exists for the set of minimized fluxes and enzyme abundance, as shown in figure 9(b). Thus, there is a strong relationship between flux and enzyme abundance, but much of it may be explained by a few reactions with large flux and high abundance.

We then check if compensation cost may be related to flux for the measured subset of reactions (since these fluxes are less biased). As explained in section 3.2.1, we divide the compensation distances by the original flux of the target reaction in order to normalize the compensation costs. Thus, we do not expect to see a relationship between flux and compensation cost. This is reflected in the plot of compensation cost vs. flux in figure 10, where we see that the correlation between compensation cost and flux is not significant ($r = 0.2695$ and $p = 0.1927$).

Finally, we want to verify that the reaction-level noise measurement, the distance to the median (DM) is not related to enzyme abundance. We computed the DM as a "relative" noise measure, relative to the median noise in proteins of similar abundance levels [Newman et al., 2006]. Again, we do not expect to see a significant correlation between DM and protein abun-



(a) Noise vs. abundance in measured reactions.

(b) Noise vs. abundance in all estimated fluxes.

Figure 11: Reaction-level noise vs. reaction-level enzyme abundance.

dance. However, when we plot DM vs. abundance (figure 11), we notice a moderately negative relationship where $r = -0.1213$ and $p = 0.0565$ for all reactions, and a slightly weaker relationship for the measured subset where $r = -0.3708$ and $p = 0.118$.

Between the compensation cost, flux, abundance, and reaction noise, we see that there are weak relationships between compensation cost and flux, as well as between flux and abundance. However, it is unlikely that these relationship fully explain the correlation between compensation cost and reaction noise.

3.6 Enzyme Production as a Gamma Distribution

Recently, a study by Cai et al. measured protein (enzyme) expression levels at the single molecule level in live cells [Cai et al., 2006]. They demonstrated that in *Escherichia coli* cells, stochastic protein production processes can be modeled using a gamma distribution where the key parameters of the distribution correspond to the factors controlling protein expression. The two key parameters in protein expression consist of: the average frequency of expression bursts per cell cycle, a; and the average number of protein molecules per burst, b. These factors combine to determine the level of steady-state protein abundance in the cell. Xie et al. showed that the distribution of steady-state protein abundance levels in a cell can be fitted to a gamma distribution. The

parameters a and b in the gamma distribution correspond to the average frequency and burst size, respectively. Thus, the probability distribution of enzyme abundance $p(x)$ can be modeled as:

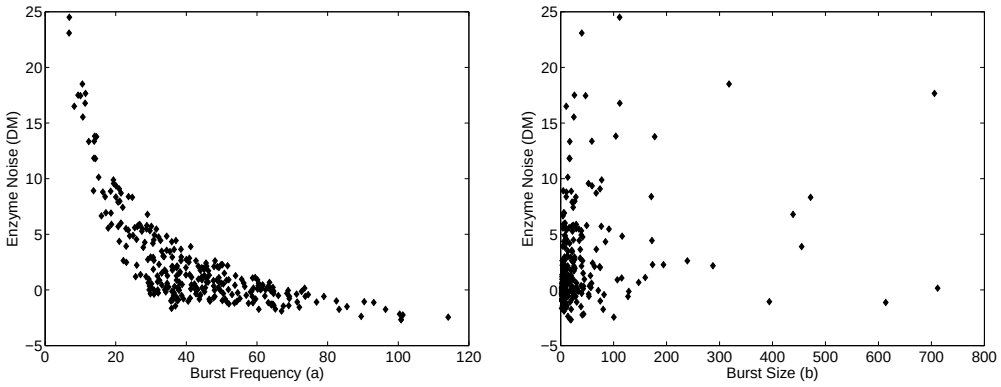
$$p(x) = \frac{x^{a-1}e^{-x/b}}{b^a\Gamma(a)} \tag{12}$$

We can apply this model to protein abundances in yeast in order to gain insight into the mechanisms behind noise control. In our experiments, we have a set of steady-state protein abundance measurements in *S. cerevisiae* where we know the mean and variance of the abundance distribution. Using the result from Xie et al., we can assume that these protein abundances can also be modeled as a gamma distribution. Thus, we can use the mean μ and variance σ^2 to compute the parameters a and b of the gamma distribution representing protein production. We can compute the parameters as follows:

$$a = \frac{\mu^2}{\sigma^2} \quad and \quad b = \frac{\sigma^2}{\mu} \tag{13}$$

We can use equation 13 to obtain the burst frequency and burst size of the enzyme production process in yeast. Thus, we can ask the question: is one or both of the parameters of protein production a factor behind enzyme noise? We can compare the parameters a and b against the relative enzyme noise DM (equation 10). We note that our noise measure DM is not the *variance* of the enzyme abundance distribution; rather it is a measure of the relative variation in enzyme levels normalized by the mean abundance (see section 3.3.1). We plot the two parameters against DM in figure 12.

In figure 12, we can see clearly that burst frequency has a strong inverse relationship with noise ($r = -0.7034$ and $p < 10^{-16}$), while burst size has a weaker negative relationship with noise. Thus, we see that manipulating burst frequency may be a significant factor behind controlling enzyme noise, while burst size is also a potentially important, albeit weaker, factor. We also note that the set of enzymes used in this portion of the study are not all metabolic enzymes. Due to a lack of data, we were not able to draw comparisons between metabolic enzyme production parameters and compensation costs. We summarize our comparisons in table 3. We note that we also see a relationship between the mean μ and burst frequency with correlation $r = 0.35$ and $p = 8.7e - 10$. However, the mean μ is not related to the noise DM. Thus, the mean does not explain the relationship between burst frequency and noise.



(a) Burst frequency (a) vs. noise (DM) (b) Burst size (b) vs. noise (DM)

Figure 12: Parameters (burst frequency and size) of a gamma model of enzyme expression plotted against enzyme noise (DM).

Variables	Correlation Coeff (r)	P-value
Burst frequency vs. enzyme noise	-0.7034	$< 10^{-16}$
Burst size vs. enzyme noise	-0.2376	3.87×10^{-5}
Burst frequency vs. burst size	0.102	0.0806
Mean abundance vs. enzyme noise	-0.0491	0.4019
Burst size vs. mean abundance	0.3479	8.7×10^{-10}

Table 3: Summary of correlations between parameters of enzyme production and other variables.

4 Discussion

In this study, we examine the motivations behind the regulation of metabolic network noise in *S. cerevisiae* . Recent research measuring variation in enzyme abundances shows that there is a distribution of noise in enzyme abundances which may largely be a result of stochastic processes in metabolic regulatory pathways [Newman et al., 2006]. It is generally thought that organisms are influenced by evolutionary pressures to adopt biologically optimal states [Edwards et al., 2001]. They do so to achieve a diverse set of objectives needed for survival, including but not limited to growth, maintenance, and robustness to the environment [Fischer and Sauer, 2005]. Thus,

we reason that the distribution of noise found in the metabolic network may also be "optimized" by the cell according to its underlying objective function. To explore the motivations behind this optimization, we use a system-wide optimization model inspired by Flux Balance Analysis (FBA) to model metabolic behavior [Edwards et al., 2002, Varma and Palsson, 1994]. We simplify the problem by assuming that the cell's underlying objective function is to maximize growth. Then, we use the network *in silico* to experiment with the robustness of the network's optimal growth to noise.

Our main result was to characterize a property of each reaction which can explain, to some degree, the differences in noise levels between reactions. This *compensation cost* quantifies the effect of a target reaction's (downward) variability on the network by measuring the degree of compensation required by the network to maintain near-optimal growth when the target reaction's flux is permuted downwards. One would expect that variability in reaction fluxes can inhibit the achievement of the cell's objective function [Papp and Pál, 2004]. Thus, we see that reactions with higher compensation cost whose variation has a larger effect on the network also tend to have lower noise.

We can interpret this relationship in several ways. First, the compensation cost can be seen as a measure of a reaction's "importance" to maintaining optimal growth. More important reactions have fewer accessible redundant pathways, and therefore it is difficult (and in a living organism may be impossible) to compensate for a reduction in their fluxes. Thus, in important reactions characterized by high compensation cost, it would benefit the cell to maintain a low level of reaction noise. Also, a higher compensation cost may require more energy to activate long alternative pathways. This would cost the cell a greater energy price, and may also be infeasible in a living organism. Thus, again it would be beneficial to the cell to maintain lower noise in reactions with high compensation cost.

Finally, we explore possible mechanisms of noise control by modeling the protein production process in *S. cerevisiae* as a gamma distribution with the parameters of burst frequency (a) and burst size (b) [Cai et al., 2006]. We show that both burst frequency and size are inversely related to noise; however, frequency has a much stronger negative relationship with noise. This implies that protein burst frequency may be an important mechanism for controlling enzyme noise levels. For instance, to catalyze a target reaction r, a given amount of enzymes can be produced using either a higher burst frequency and lower burst size, or vice versa. However, if r has a high compensation cost, the cell may want to maintain lower variance in the cat-

alyzed reaction's flux. In this case, it could regulate the noise by "choosing" to produce the enzymes by using higher burst frequency and lower burst size.

References

Jeremy M. Berg, John L. Tymoczko, and Lubert Stryer. *Biochemistry*. W. H. Freeman and Co., New York, New York, 5th edition, 2001.

Lars Blank, Lars Kuepfer, and Uwe Sauer. Large-scale 13c-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biology*, 6(6):R49, 2005. ISSN 1465-6906. doi: 10.1186/gb-2005-6-6-r49. URL `http://genomebiology.com/2005/6/6/R49`.

HPJ Bonarius, G. Schmid, and J. Tramper. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends in Biotechnology*, 15(8):308–314, 1997.

L. Cai, N. Friedman, and X.S. Xie. Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440(7082):358–362, 2006.

NC Duarte, MJ Herrgard, and BO Palsson. Reconstruction and validation of saccharomyces cerevisiae ind750, a fully compartmentalized genome-scale metabolic model. *Genome Res*, 14:1298–1309, 2004. doi: 10.1101/gr.2250904.

J.S. Edwards, R.U. Ibarra, and B.O. Palsson. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nature Biotechnology*, 19:125–130, 2001.

J.S. Edwards, M. Covert, and B. Palsson. Metabolic modelling of microbes: the flux-balance approach. *Environmental Microbiology*, 4(3):133–140, 2002.

E. Fischer and U. Sauer. Large-scale in vivo flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism. *Nature Genetics*, 37:636–640, 2005.

J Forster, I Famili, P Fu, BO Palsson, and J Nielsen. Genome-scale reconstruction of the saccharomyces cerevisiae metabolic network. *Genome Res*, 13:244–253, 2003. doi: 10.1101/gr.234503.

G. Giaever, A.M. Chu, L. Ni, C. Connelly, L. Riles, S. Véronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. André, et al. Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, 418:387–391, 2002.

L. Kuepfer, U. Sauer, and L.M. Blank. Metabolic functions of duplicate genes in Saccharomyces cerevisiae, 2005.

S. Lee, C. Palakornkule, MM Domach, and IE Grossmann. Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Computers and Chemical Engineering*, 24(2):711–716, 2000.

J.R.S. Newman, S. Ghaemmaghami, J. Ihmels, D.K. Breslow, M. Noble, J.L. DeRisi, and J.S. Weissman. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, 441:840–846, 2006.

B. Papp and C. Pál. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature*, 429:661–664, 2004.

J. Pramanik and JD Keasling. Stoichiometric model of Escherichia coli metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnology and Bioengineering*, 56(4):398–421, 1997.

J.M. Raser and E.K. O'Shea. Noise in Gene Expression: Origins, Consequences, and Control. *Science*, 309(5743):2010–2013, 2005.

C.H. Schilling, S. Schuster, B.O. Palsson, and R. Heinrich. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog*, 15(3):296–303, 1999.

Alexander Schrijver. *Theory of Linear and Integer Programming*. John Wiley and Sons, 1998.

D. Segre, D. Vitkup, and G.M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99(23):15112–15117, 2002.

T. Shlomi, O. Berkman, and E. Ruppin. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences*, 102(21):7695–7700, 2005.

A. Varma and B.O. Palsson. Metabolic flux balancing: Basic concepts, scientific and practical use. *Bio/Technology*, 12(10):994–998, 1994.