UNIVERSITY OF CALIFORNIA

Los Angeles

Ensemble Modeling for Pathway Engineering and Strain Design

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Chemical and Biomolecular Engineering

by

Jimmy Gerard Lafontaine Rivera

2015

ABSTRACT OF THE DISSERTATION

Ensemble Modeling for Pathway Engineering and Strain Design

by

Jimmy Gerard Lafontaine Rivera

Doctor of Philosophy in Chemical and Biomolecular Engineering

University of California, Los Angeles, 2015

Professor James C. Liao, Chair

As global oil demand approaches unsustainable levels, people have turned to living organisms as a possible alternative source for many petroleum-derived products. Great advances in the fields of synthetic biology and metabolic engineering have enabled microbial production of fuels and other chemical feedstock. Nevertheless, the production cost of many of these processes makes it hard to compete with their petroleum derived counterparts.

Due to their complexity, living organisms present a particular challenge to their use as chemical catalysts. Although metabolic engineering has greatly enhanced the catalytic capabilities of microbes, the minor alterations that have yielded such improvements seem to be unable to take the catalytic activity to the levels necessary to compete with petroleum. Novel efforts are trying to use major changes in bacterial metabolism in order to bypass the current limitations of biological systems. Although very promising, engineered pathways do not have the chance to be tuned through years of evolution; as such their incorporation into metabolism can lead to metabolic

imbalances, which will ultimately hamper production. Due to the complexity of these metabolic systems, mathematical tools would be instrumental to assessing such problems.

Currently, there are no metabolic modeling methods that can both consider metabolite concentrations explicitly and account for all annotated reactions for the particular organism. There are two main difficulties impeding the development of such method: (1) Kinetic parameter values are unavailable for most reactions (the parameter space grows exponentially with the number of reactions); and (2) computation time increases exponentially as the system becomes larger, making difficult the use of random sampling. Through the use Ensemble Modeling and parameter continuation we were able to overcome these hurdles and develop some specific tools for the mathematical analysis of metabolic systems.

If a system is not robust, small changes in expression levels may lead to system failure due to the disappearance of a stable steady state. Given a lack of regulatory mechanisms, this problem becomes especially important when designing synthetic pathways. Nevertheless, it is often difficult to identify flaws in the design which might lower robustness using intuition alone. To address this issue, we developed a method termed Ensemble Modeling for Robustness Analysis (EMRA), which combines parameter continuation with the EM approach for investigating the robustness issue of metabolic pathways. Using a large ensemble of reference models with different parameters, we can determine the effects of parameter drifting until a bifurcation point. This method gives us an unbiased analysis of the robustness issues of a particular pathway structure.

Similarly, in order to be able to maintain a viable phenotype under small changes in internal or external conditions (e.g. variations in enzyme levels or nutrient concentrations) cells must possess a robust metabolic network. Through evolution, native metabolic pathways have seemingly solved the robustness problem by selecting a robust network structure and regulatory mechanisms such

that the feasible range of each parameter is sufficiently large. Although this fact is well accepted, it has never been used during the creation of kinetic models of native metabolism. By developing a quantitative and scalar index based on EMRA, we can add robustness as an objective during parameter-fitting. As kinetic models have a large number of parameters and the number of available data points is relatively low, we hypothesize that the additional robustness criterion will greatly improve the quality of fitted models.

By utilizing robustness, a cell-wide model of *Escherichia coli* was constructed. Cell-wide metabolic models incorporate many of the intrinsic complexities of living organisms. Although the utility of encompassing such a large portion of cellular metabolism is undeniable, the computational difficulties regarding such models have hampered their use. The use of entropy allowed for more intelligent model curation, eased the computational difficulties un-robust models present. Additionally, a novel application of parameter continuation was developed and used to successfully predict *in vivo* isobutanol production.

The dissertation of Jimmy Gerard Lafontaine Rivera is approved.

Yi Tang

Elliot Landaw

James C. Liao, Committee Chair

University of California, Los Angeles

2015

DEDICATION

No words can describe how amazing life has been to me. This work is dedicated to everybody that has made it that way.

1 TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my advisor Prof. James C. Liao for his mentorship. Additionally, I would like to thank the other modeling crew members during my tenure Dr. YiKun Tan and Dr. Yun Lee who contributed greatly to my professional development. I would also like to express my gratitude to the rest of the Liao lab for giving me the opportunity to learn and collaborate with them. I would like to thank all my friends for making graduate school fly by. And most specially, I would like to thank Brenda L. Gonzalez Flores, Olga N. Rivera Ayala, Jimmy Lafontaine Medina, Jessica M. Lafontaine Sanchez, Luis A. Lafontaine Rivera, Jean C. Lafontaine Rivera, Maria I. Ayala Rivera, Elias Rivera Serrano, God and the rest of my family because they were more than support; they were who pushed me forward.

The work in this dissertation was supported by National Science Foundation (MCB-1139318), UCLA-DOE Institute for Genomics and Proteomics, and the Office of Science (BER), U.S. Department of Energy (DE-SC0012384, DE-SC0001060 and DE-SC0008744).

Chapter 2 is a version of Lee, Yun, Jimmy G. Lafontaine Rivera, and James C. Liao. "Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways." Metabolic engineering 25 (2014): 63-71. Yun Lee and Jimmy G. Lafontaine Rivera created all figures and wrote the paper. James C. Liao served as the project director.

Chapter 3 is a version of Lafontaine Rivera, Jimmy G., Yun Lee, and James C. Liao. "An entropy-like index of bifurcational robustness for metabolic systems." Integrative Biology(2015). Yun Lee and Jimmy G. Lafontaine Rivera created all figures and wrote the paper. James C. Liao served as the project director.

VITA

Education

- 2009: B.S. in Chemical Engineering, Rensselaer Polytechnic Institute.
- 2009: B.S. in Biology, Rensselaer Polytechnic Institute.

Publications

- Lafontaine Rivera, J. G.*, Lee, Y.* (*co-first authors), & Liao, J. C. (2015). An entropy-like index of bifurcational robustness for metabolic systems. *Integrative Biology*.

- Lee, Y.*, Lafontaine Rivera, J. G.* (*co-first authors), & Liao, J. C. (2014). Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. *Metabolic engineering*, *25*, 63-71.

- Wernick, D. G., Choi, K. Y., Tat, C. A., Lafontaine Rivera, J. G., & Liao, J. C. (2013). Genome sequence of the extreme obligate alkaliphile Bacillus marmarensis strain DSM 21297. *Genome announcements*, *1*(6), e00967-13.

- Zomorrodi, A. R., Lafontaine Rivera, J. G., Liao, J. C. and Maranas, C. D. (2013), Optimization-driven identification of genetic perturbations accelerates the convergence of model parameters in ensemble modeling of metabolic networks. *Biotechnology Journal*, 8: 1090–1104. doi: 10.1002/biot.201200270.

- Huo, Y. X., Cho, K. M., Lafontaine Rivera, J. G., Monte, E., Shen, C. R., Yan, Y., & Liao, J. C. (2011). Conversion of proteins into biofuels by engineering nitrogen flux. *Nature biotechnology*, *29*(4), 346-351.

- Tan, Y., Lafontaine Rivera, J. G., Contador, C. A., Asenjo, J. A., & Liao, J. C. (2011). Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux. *Metabolic engineering*, *13*(1), 60-75.

1 Introduction

1.1 Background

Intracellular metabolite concentrations drive cellular metabolism through metabolic pathways catalyzed by enzymes. Thus, metabolite concentrations can be viewed as the driving force of flux after enzymes are expressed. Manipulation of these driving forces to favor the production of a desirable compound is a key to the success of metabolic engineering (18, 19). Therefore, it is of interest to predict what happens to intracellular metabolite concentrations upon introduction of an engineered pathway. For example, the introduction of a synthetic pathway, say, isobutanol pathway, into a host may deplete a precursor metabolite or a cofactor, which may eventually limit the flux through the desired pathway. Predicting such changes brought about by introducing a foreign pathway may help to design strategies for metabolic engineering.

Currently, only dynamic models of metabolism consider intracellular metabolite concentrations explicitly (12). These models represent balance of metabolite concentrations in terms of ordinary differential equations in which reaction fluxes are functions of metabolite, enzyme concentrations, and kinetic parameters (3). A growing amount of kinetic data and improvements in -omic fields has helped dynamic models increase in scope and accuracy. However, the general approach of dynamic modeling is still tedious and involves details that may or may not be necessary for our purpose here. For example, most efforts in dynamic model building have focused on parameter optimization and time-course predictions, which are useful but unnecessary if only the effect on steady-state metabolite concentration is of interest. However, owing to the large number of kinetic parameters and complexity of dynamic models, it is imperative that parameters are assigned adequately. Inappropriate assignment or purely random sampling of such parameters would lead

to infeasible models which have no steady state. Ensemble Modeling circumvents these limitations while minimizing the need for targeted experiments (20-24).

1.2 Previous Approach to Ensemble Modeling

Ensemble Modeling works by constructing an ensemble of models that all reach the same steady-state flux distribution, but span all the kinetic space allowed by this constraint (22). These models are dynamically different and will reach distinct steady-states upon enzyme tuning. This divergence upon enzyme tuning serves as a basis for screening and is used by the EM algorithm in order to fit kinetic parameters (24). Previous studies (21, 24) have shown using this scheme and screening with flux data, models will quickly reduce to a small set and this set becomes increasingly predictive as new data becomes available.

Anchoring models to the same reference steady-state allows Ensemble Modeling to greatly reduce the parameter space that needs to be sampled while still spanning all possible but feasible kinetics (22, 23). In this sense, Ensemble Modeling is unique to random sampling algorithms in that it utilizes available data as a sampling constraint. Sampling within this constrained space makes fitting the models to the data computationally. A schematic of this algorithm is shown in Figure 1.

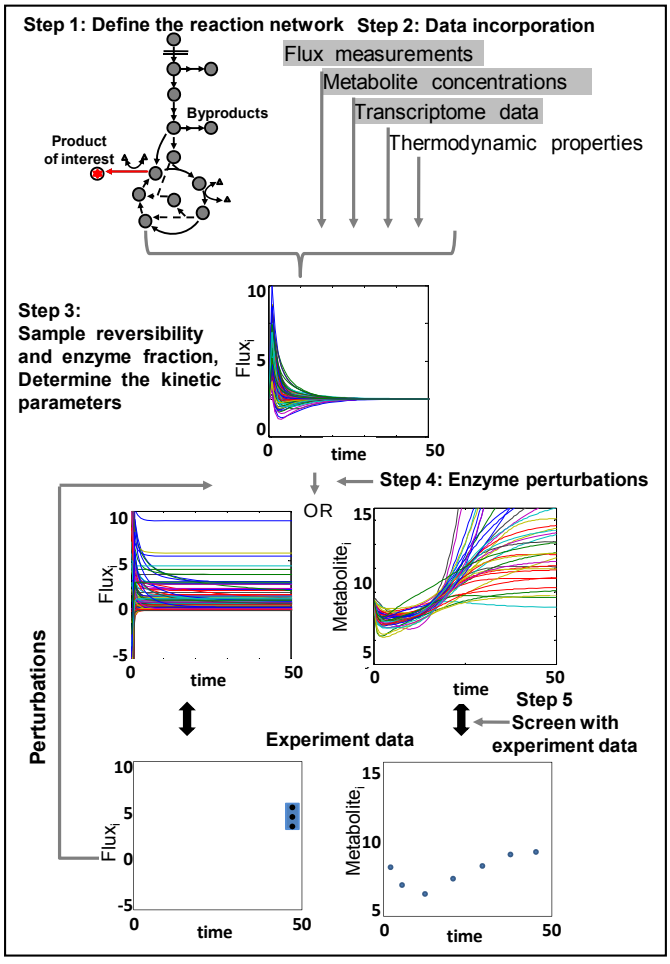

Figure 1-1. Schematic of the Ensemble Modeling algorithm

Construction of a dynamic model is based on network stoichiometry and enzyme kinetics. Current techniques to develop such models start from the selection of lumped enzyme kinetic expressions, search for parameters which might be available in literature and sample the remaining parameters within a large parameters space in order to fit transient metabolite concentration data (3). These approaches are limited by the fact that both lumped enzyme kinetic expressions and kinetic parameters are rarely available or inaccurate for the system being studied. In addition, transient metabolite concentration data is very difficult to measure and impossible to appropriately fit when sampling within such a large parameter space. Instead, Ensemble Modeling's sampling scheme reduces the sampling space as to not depend on available parameter information and is able to use readily available data (eg. external fluxes) in order to fit parameters.

Although not necessary, the EM approach most often uses elementary reaction kinetics to describe the reactions within the metabolic network (20-25). Elementary reaction kinetics were chosen for their non-linearity and consistency. In this scheme each enzymatic reaction is broken down into a series of uni-molecular or bi-molecular reactions including enzyme-substrate collisions, product formation and dissociation of free enzymes as demonstrated in the following scheme:

$$x_1 + E_i \underset{v_{i,2}}{\overset{v_{i,1}}{\rightleftarrows}} x_1E_i \underset{v_{i,4}}{\overset{v_{i,3}}{\rightleftarrows}} x_2E_i \underset{v_{i,6}}{\overset{v_{i,5}}{\rightleftarrows}} x_2 + E_i$$

This scheme can be used to represent any form of enzyme kinetics or regulation. The above mentioned reactions all follow mass action kinetics for each elementary step, but all exhibit non-linear behavior due to the bi-molecular interactions with enzymes. When using this form of

reaction kinetics, the ensemble of models can be sampled to reach a specified reference steady-state by sampling reversibility and enzyme fractions as shown below.

Modeling Objectives

Flux-oriented Objective:

To match model-predicted flux with data upon enzyme tuning

$$\min \sum_i \left(V_{i,ss}\left(\frac{E_{tot}+\Delta E_{tot}}{E_{tot}}\right) - J_{i,ss}^{measured}\left(\frac{E_{tot}+\Delta E_{tot}}{E_{tot}}\right)\right)^2$$

Pros:
- Objective function directly reflects metabolic engineering needs
- Flux readily measureable
- Effects of kinetics on flux accountable

OR

Metabolite-oriented Objective:

To match metabolite profiles with data

$$\min \sum_i \left(x_i(t) - x_i^{measured}(t)\right)^2$$

by changing K and E_{tot}, such that

$$\frac{dX}{dt} = F(X, K, E_{tot}) \qquad V_{ss} = G(X_{ss}, K, E_{tot}, \Delta E_{tot})$$

where ss represents steady state of the ODE, G represents the function of steady-state flux.

Pro:
- Direct test of ODE

Con:
- Not directly applicable to metabolic engineering

Ensemble Modeling Features

1) Use an additional constraint for the reference strain $(\Delta E_{tot} = 0)$

$$V_{i,ss} = G(X_{ss}, K, E_{tot}, \Delta E_{tot} = 0) = J_{i,ss}^{measured}$$

Pros:
- Anchor the ODE to the measured steady-state flux
- Reduce the parameter space significantly

2) Normalize the ODE so that the metabolite concentrations are dimensionless, but the fluxes are dimensional.

Pros:
- Metabolite concentrations are not needed if not available
- Fluxes carry the correct dimension.

3) Use elementary reactions as a general reaction framework.

Pros:
- Most fundamental and flexible.
- Can include regulation.

4) Convert K and E_{tot} to reversibility (R) and enzyme factions (e)

Pro:
- Bounded, easy sampling.

5) Parallel (ensemble) iteration

Pro:
- High throughput screening of parameters.

Figure 1-2. Mathematical summary of the Ensemble Modeling approach.

The reversibility is defined as:

$$R_{i,j} = \frac{\min\left(v_{i,2j-1}^{ref}, v_{i,2j}^{ref}\right)}{\max\left(v_{i,2j-1}^{ref}, v_{i,2j}^{ref}\right)} \quad \text{Eq. 1}$$

where the superscript indicates the values are for the reference state, i is used to differentiate between reactions and j indicates the step of the reaction. With this reversibility sampled between 0 and 1 (or thermodynamic constraints), the uni-directional($v_{i,2j-1}^{ref}, v_{i,2j}^{ref}$) fluxes can be calculated. In addition, all concentration variables are converted into dimensionless variables by dividing the metabolite concentration or enzyme complex concentration by the steady-state concentration or total enzyme concentration, respectively. Mass action kinetics with dimensionless concentration shown below:

$$v_{i,1} = \left(k_{i,1} E_{i,total}^{ref} X_i^{ss,ref}\right) \frac{[X_i]}{X_i^{ss.ref}} \frac{[E_i]}{E_{i,total}^{ref}}$$

$$= \tilde{K}_{i,1}^{ref} \tilde{X}_i \, \tilde{e}_{i,1} \quad \text{Eq. 2}$$

where $\tilde{K}_{i,1}^{ref}$ is the normalized rate constant. At the reference steady state $\tilde{X}_i$=1 and the sum of different dimensionless enzyme complex for a particular enzyme ($\sum_j \tilde{e}_{i,j}$) is 1. Therefore, when the enzyme fraction $\tilde{e}_{i,j}$ at reference state is sampled between 0 and 1 (subject to a conservation of the sum), the normalized rate constants can be calculated. Sampling kinetic parameters, through this algorithm guarantees that every model sampled will reach the reference steady-state yet have distinct dynamic behavior.

Anchoring to a reference steady-state makes EM such a powerful tool for the development of dynamic models. Anchoring to a particular steady-state not only limits sampling to models which

yield appropriate flux distributions for the reference state, but also sets a time scale for the system. These benefits make it possible to optimize kinetic parameters using appropriate objective functions depending on your data. A description of how EM can be used to optimize kinetic parameters and how it compares to traditional methods is shown in Figure 2.

1.3 Ensemble Modeling: Making informed predictions from minimal data

With the minimal amount of information necessary to run, Ensemble Modeling algorithm makes predictions using three levels of information. First, stoichiometric relationships (Figure 3, 1) provide the structure of the model and constrain the relationship of metabolites to one another through fluxes. As seen with constrain-based methods like Flux Balance Analysis, stoichiometric relationships already hold a significant amount of information about the system (26, 27). Using stoichiometric relationships alone does not allow us to solve for metabolite concentrations explicitly and therefore additional information is necessary. Second, kinetic relationships (Figure 3, 2) constrain how metabolite concentrations relate to fluxes. In kinetic models, fluxes are a

Ensemble Modeling predicts based on 3 levels of information:

Stoichiometric relationships:

1. $\frac{dx}{dt} = Sv$
   a. This is the most basic level of information.
   b. Holds the information on how fluxes affect metabolites
   c. FBA and other constrait-based methods make predictions based solely on stoichiometric relationships.

Kinetic relationships:

2. $v = f(k, x)$
   a. Although kinetic parameters may vary, the metabolites that affect a particular flux remain the same.
   b. The qualitative effect of a particular metabolite on flux tends to be constant.
      - Subtrates and activators have a positive effect.
      - Products and inhibitors have a negative effect.

Parameter relationships:

3. $k = f(V^{ref}, x^{ref})$
   a. Possible kinetic parameters are constrained by known steady-state fluxes.
   b. Normalization with respect to **x** allows EM to constrain parameter relationships without measurement or estimation of metabolite concentrations.
   c. Without constraining parameter relationships it is very difficult to model complex networks.

Figure 1-3. Sources of information in an unscreened ensemble of models.

function of metabolite concentrations and associated kinetic parameters. Unfortunately, kinetic parameter values are unknown for most reactions. Random sampling of these values leads to infeasible models or models that cannot be handled by standard computational tools. In order to bypass this problem, EM incorporates its third level of information (Figure 3, 3). By anchoring all sampled models to the same reference steady-state flux, EM constrains possible parameter relationships (28). Traditionally anchoring to a reference steady-state flux requires knowledge of metabolite concentrations (3), but by using non-dimensional concentration units, the EM algorithm can sample without the need to measure absolute metabolite concentrations for the reference state.

2 Ensemble Robustness Analysis for Engineering Non-Native Metabolic Pathways

Metabolic pathways in cells must be sufficiently robust to tolerate fluctuations in expression levels and changes in environmental conditions. Perturbations in expression levels may lead to system failure due to the disappearance of a stable steady state. Increasing evidence has suggested that biological networks have evolved such that they are intrinsically robust in their network structure. However, the robustness consideration has not played a major role in non-native metabolic pathway design. In this article, we presented Ensemble Robustness Analysis (ERA), which combines a continuation method with the Ensemble Modeling approach, for investigating the robustness issue of non-native pathways. ERA investigates a large ensemble of reference models with different parameters, and determines the effects of parameter drifting until a bifurcation point, beyond which a stable steady state disappears. We focus on the robustness of two synthetic central metabolic pathways that achieve carbon conservation: non-oxidative glycolysis and reverse glyoxylate cycle. With ERA, we determined the probability of system failure of each design and demonstrated that alternative designs of these pathways indeed display varying degrees of robustness. Furthermore, we demonstrated that target selection for flux improvement should consider the trade-offs between robustness and performance, which has not been previously appreciated.

2.1 Introduction

Metabolic engineering has advanced from minor alterations of existing pathways to significant re-routing of the metabolic path for better utilization of substrates (16, 29, 30) or formation of non-native products (15, 31-36). With a notable exception (37), all metabolic engineering efforts aim to achieve a steady state or quasi-steady state. Non-steady states in metabolic engineering typically result in accumulation or disappearance of intermediate metabolites. Sustained oscillation has not

found its application in practical metabolic engineering as yet. Since the engineered pathways are not tuned through evolution, the expression levels of the pathway genes may drift outside the working range as the physiological conditions change. Drifting of expression levels or other kinetic parameters may lead to gradual deterioration of performance or sudden system failure characterized by the disappearance of a stable steady state. While the deterioration of performance is undesirable, the occurrence of system failure could be catastrophic for the cell. Thus, a robust pathway design should first focus on avoiding system failure before attempting to improve performance.

Through evolution, native metabolic pathways apparently have solved the robustness problem by selecting a robust network structure such that the feasible range of each parameter is sufficiently large (38-41). In addition, various regulatory mechanisms are in place to dynamically control the kinetic parameters under various physiological conditions. In contrast, non-native or metabolically engineered native pathways are potentially prone to system failure, when a kinetic parameter moves away from the initially designed level, causing accumulation or depletion of metabolites and the disappearance of a stable steady state. Thus, robustness should be an important criterion for designing non-native pathways. Even with an artificial dynamic controller (42-44) designed for the non-native pathway, it is desirable to choose robust network configurations or parameter ranges that are inherently robust.

The robustness problem calls for a modeling approach that integrates kinetic parameters with systems performance. Kinetic parameters are perturbed in such models to examine the consequences of drifting. Unfortunately, key kinetic parameters (*e.g.*, V_{max}'s) are system-dependent and usually unknown. Previous efforts have addressed the uncertainty of metabolic parameters through the random sampling of parameters (24, 45, 46) to form an ensemble of

models. Various approaches are then used to extract useful information from the ensemble upon large parameter changes (21, 23), or infinitesimal perturbations that define control coefficients (45, 47, 48). Nevertheless, the probability of system failure due to the disappearance of a stable, desired steady state has not been addressed for an ensemble of models.

Here we combine a continuation method (49) with the previously developed ensemble modeling (EM) technique (23, 24) to evaluate the robustness of non-native metabolic pathways, particularly the probability of system failure. This hybrid approach, termed Ensemble Robustness Analysis (ERA), allows the investigation of large parameter changes in an ensemble of models. For each model, the continuation method enables rapid determination of the steady-state solution as a parameter of interest is altered up or down from the baseline. The continuation of steady-state solution proceeds along either direction until a bifurcation point, beyond which a stable steady state loses its stability. In metabolic systems, the disappearance of a stable steady state may or may not be accompanied by the emergence of sustained oscillations (50, 51). Instead, accumulation and depletion of intermediate metabolites are far more common, which lead to system failure. Even if metabolic oscillations are functional or beneficial in certain conditions, an oscillatory metabolic system is by no means ideal for maintaining a consistent supply of pathway products. Thus, metabolic design should avoid the bifurcation point, and use the ensemble analysis to estimate the probability of system failure for a given non-native pathway design.

In this article, we demonstrate the utility of ERA by comparing possible designs of two non-native pathways: non-oxidative glycolysis (NOG) (16) and reverse glyoxylate cycle (rGC) (17). In each design, we determined probabilities of system failure and identified targets that might improve performance. With both results, ERA allows the selection of targets for flux improvement by considering both performance and robustness.

2.2 Theory and Methods

2.2.1 Dynamic, kinetics-based model

A generic expression of kinetic parameter-based models for metabolic pathways is:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, \mathbf{p}) = \mathbf{S} \cdot \mathbf{v}(\mathbf{x}, \mathbf{p}) \qquad \text{Eq. 1}$$

Here, the time derivative of metabolite concentrations in vector form (**x**) is represented as the product of **S**, the stoichiometric matrix, and **v**, the vector of reaction fluxes. Since each reaction flux is also a function of kinetic parameters (**p**), such models are useful for studying the effect of parameter drifting on system performance. One way to accomplish this is to alter the parameter of interest and then solve the ordinary differential equation (ODE)-based model to a new steady state (Fig. 1a). This time-domain approach, although straightforward, is computationally expensive if one needs to analyze varying degrees of perturbation for a large number of models. Most importantly, the time-domain approach is inadequate for detecting the loss of a stable steady state.

2.2.2 Continuation method

Here we adopt a computationally efficient and scalable continuation method (49) to investigate the effect of parameter drifting. This method aims to find a connected path of steady-state solutions ($\mathbf{x}_{SS}$) to the following equation:

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}_{SS}, \mathbf{p}) = \mathbf{0} \qquad \text{Eq. 2}$$

.

Since $\mathbf{F}(\mathbf{x}_{SS},\mathbf{p})$ is equal to zero, it follows that the total derivative of $\mathbf{F}(\mathbf{x}_{SS},\mathbf{p})$ with respect to **p** is also zero:

$$\frac{d\mathbf{F}(\mathbf{x}_{SS}, \mathbf{p})}{d\mathbf{p}} = \frac{\partial \mathbf{F}}{\partial \mathbf{x}_{SS}} \frac{d\mathbf{x}_{SS}}{d\mathbf{p}} + \frac{\partial \mathbf{F}}{\partial \mathbf{p}} = \mathbf{0}. \qquad \text{Eq. 3}$$

Further rearrangement of Eq. 3 yields Eq. 4:

$$\frac{d\mathbf{x}_{SS}}{d\mathbf{p}} = -\left(\frac{\partial \mathbf{F}}{\partial \mathbf{x}_{SS}} \right)^{-1} \frac{\partial \mathbf{F}}{\partial \mathbf{p}}, \qquad \text{Eq. 4}$$

which specifies the derivatives of steady-state concentrations with respect to kinetic parameters. Starting from a stable steady state, we can solve Eq. 4 along the direction where a kinetic parameter (usually the activity of a particular enzyme) is drifted up or down from the baseline (Fig. 1b). The corresponding solution, which traces a trajectory in the $\mathbf{x}_{SS}$-$\mathbf{p}$ space, will then characterize how the steady state responds to the drifting of one particular parameter.

Given that the solution of Eq. 4 requires the inversion of the Jacobian matrix, it is only valid in the range where the Jacobian is non-singular. In the continuation method, the point where the Jacobian becomes singular is considered a bifurcation point. Additionally, points where the Jacobian is ill-conditioned are assumed to be near-singular and therefore within tolerance limits of the bifurcation point. Given the nature of numerical integration, it is easy to miss the small region where the Jacobian is near-singular, to account for this, we also check the eigenvalues of the Jacobian matrix ($\partial \mathbf{F} / \partial \mathbf{x}_{SS}$) at each step to determine whether a bifurcation occurs. Since a stable steady state is required for proper functioning but nonetheless disappears beyond the bifurcation point, any parameter drifting that crosses the bifurcation point can be considered as entering the region of system failure (Fig. 1b). Thus, the continuation method allows both the investigation of parametric sensitivity and the detection of system failure due to disappearance of a stable steady state.

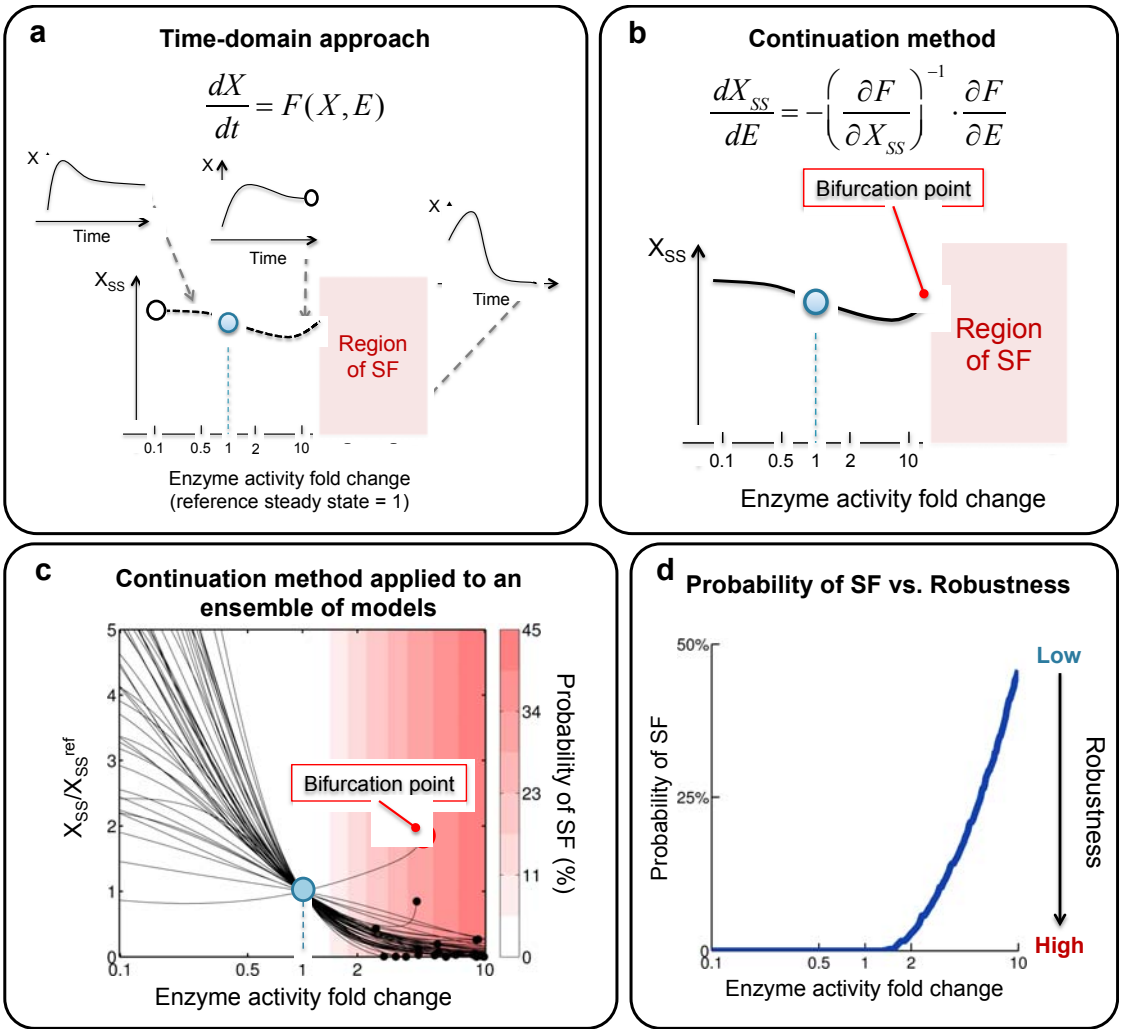

Figure 2-1. Continuation method and its application to ensemble of models. (**a**) At any given level of enzyme activity (defined as the fold change relative to the activity in the reference state), one can solve the ODEs until steady state (SS) to obtain X_{SS}, the steady-state value of X. (**b**) Alternatively, one can utilize the continuation method to trace the trajectory of X_{SS} as it varies according to the enzyme activity level. Perturbation of some enzymes may lead to the disappearance of a stable steady state, in which case the continuation method will detect and terminate at the bifurcation point. (**c**) The continuation method can be efficiently applied to an ensemble of models. Since each model has different kinetic parameters, which are sampled from feasible ranges if not known *a priori*, some models may undergo bifurcation while other models do not. The probability of system failure (SF) is defined as the percentage of models that bifurcate before a given level of perturbation. (**d**) Probability of SF can be used as a "reverse index" for robustness. That is, low probability of SF indicates high robustness, and vice versa.

2.2.3 Removing Jacobian singularity caused by conserved pools

In addition to bifurcation, the Jacobian matrix also becomes singular if conserved pools exist. Examples include the NAD(H) pool, AT(D/M)P pool, or cyclic pathways like TCA cycle without a net input or output in the model. These conserved pools will lead to linearly dependent rows in the stoichiometric matrix, thereby making it rank-deficient and singular. To remove this type of singularity, we identify all the metabolites subject to the conservation constraint and substitute one of them for an algebraic expression involving all the other conserved metabolites and the size of the conserved pool. For example, if [NADH] and [NAD] are subject to the conservation constraint

$$[\mathrm{NADH}] + [\mathrm{NAD}] = C, \qquad \text{Eq. 5}$$

where C is a constant, then we can remove the singularity by replacing every instance of [NAD] in the model with the algebraic expression C – [NADH]. Such substitution removes the linear dependency among rows of the Jacobian matrix and allows us to detect bifurcation without changing any properties of the system.

2.2.4 Ensemble Modeling

The construction of dynamic metabolic models in the form of Eq. 1 is based on both network stoichiometry, which is usually well known, and enzyme kinetic rate laws and parameters, which are system-dependent and often unavailable. Compared to kinetics data, the steady-state flux distributions are relatively easy to measure or estimate. Ensemble Modeling (EM) (23, 24) was recently introduced to take advantage of these data. It aims to construct a library (or ensemble) of models with unique parameters such that all models in the ensemble are anchored to the same steady-state flux distribution. This constraint on reaching the same steady-state flux distribution

significantly reduces the parameter space and enables effective sampling of parameters that make biological sense (22).

2.2.5 Normalization reduces the number of parameters

Consider an enzymatic reaction that obeys the canonical reversible Michaelis-Menten rate law:

$$V = \frac{V_{\max}\left[1-\left(\frac{P}{S}\right)\left(\frac{1}{K_{eq}}\right)\right]\frac{S}{K_{m,S}}}{1+\frac{S}{K_{m,S}}+\frac{P}{K_{m,P}}} \qquad \text{Eq. 6}$$

By fixing V at a value measured or estimated at a reference steady state (or simply *reference state*), we can randomly draw values of $K_{m,S}$, $K_{m,P}$ and K_{eq} from biologically reasonable distributions and calculate V_{max} if P and S are known at the same steady state. Unfortunately, the absolute values of metabolite concentrations (*e.g.*, P and S in Eq. 6) are usually system-dependent and unavailable, particularly for non-native pathways. As a result, it is common in the practice of EM (22, 24) to normalize the metabolite concentrations and kinetic parameters such that the normalized concentrations become 1 at the reference state. After normalization, the absolute values of metabolite concentrations become optional, and the now-dimensionless concentration variables reflect only the change relative to the reference state.

As an example, Eq. 6 can be rewritten in the normalized form:

$$V = \frac{V_{\max}\left[1-\left(\frac{\tilde{P}}{\tilde{S}}\right)\left(\frac{1}{\tilde{K}_{eq}}\right)\right]\frac{\tilde{S}}{\tilde{K}_{m,S}}}{1+\frac{\tilde{S}}{\tilde{K}_{m,S}}+\frac{\tilde{P}}{\tilde{K}_{m,P}}}, \qquad \text{Eq. 7}$$

where $\tilde{P} = \frac{P}{P^{ref}}, \tilde{S} = \frac{S}{S^{ref}}$, $\tilde{K}_{m,S} = \frac{K_{m,S}}{S^{ref}}$, $\tilde{K}_{m,P} = \frac{K_{m,P}}{P^{ref}}$ and $\tilde{K}_{eq} = \frac{S^{ref}}{P^{ref} \cdot K_{eq}}$ are dimensionless parameters and the superscript *ref* denotes the reference state. Note that the flux is still dimensional, while the normalized metabolite concentrations and kinetic parameters are dimensionless. Thus, if the desired pathway flux is known, then the absolute flux values can be used. Otherwise, the flux values can be relative as well.

Although the aforementioned normalization can substitute the metabolite concentrations (**x**) and kinetic parameters (**p**) with their normalized equivalents $\tilde{\mathbf{x}}$ and $\tilde{\mathbf{p}}$, respectively, such that

$$\mathbf{v}(\mathbf{x}, \mathbf{p}) = \mathbf{v}(\tilde{\mathbf{x}}, \tilde{\mathbf{p}}), \qquad \text{Eq. 8}$$

the time derivatives of normalized concentrations ($d\tilde{\mathbf{x}}/dt$) are still dependent on the steady-state concentrations at the reference state:

$$\begin{aligned} \frac{d\tilde{\mathbf{x}}}{dt} &= diag\left(1/x_1^{ref}, ..., 1/x_n^{ref}\right) \cdot \frac{d\mathbf{x}}{dt} \\ &= diag\left(1/x_1^{ref}, ..., 1/x_n^{ref}\right) \cdot \mathbf{F}(\mathbf{x}, \mathbf{p}) \\ &= diag\left(1/x_1^{ref}, ..., 1/x_n^{ref}\right) \cdot \mathbf{S} \cdot \mathbf{v}(\mathbf{x}, \mathbf{p}) \\ &= diag\left(1/x_1^{ref}, ..., 1/x_n^{ref}\right) \cdot \mathbf{S} \cdot \mathbf{v}(\tilde{\mathbf{x}}, \tilde{\mathbf{p}}) \\ &= diag\left(1/x_1^{ref}, ..., 1/x_n^{ref}\right) \cdot \mathbf{F}(\tilde{\mathbf{x}}, \tilde{\mathbf{p}}) \end{aligned} \qquad \text{Eq. 9}$$

Here, $diag\left(1/x_1^{ref}, ..., 1/x_n^{ref}\right)$ refers to a diagonal matrix whose diagonal entries are the inverses of steady-state concentrations at reference state.

According to Eq. 9, it is clear that the dynamic behavior of the normalized system is still affected by the concentrations at reference state even after normalization. However, if one is focused on steady-state analysis only, then the steady-state equations of the normalized system

$$diag\left(1/x_1^{ref},...,1/x_n^{ref}\right)\cdot\mathbf{F}(\tilde{\mathbf{x}}_{SS},\tilde{\mathbf{p}})=0 \qquad \text{Eq. 10}$$

can be reduced to

$$\mathbf{F}(\tilde{\mathbf{x}}_{SS},\tilde{\mathbf{p}})=0 \qquad \text{Eq. 11}$$

,

because $diag\left(1/x_1^{ref},...,1/x_n^{ref}\right)$ is a non-singular matrix (all diagonal entries are positive numbers representing inverses of concentrations) and therefore invertible. From Eq. 11 we can similarly derive (*cf.* Eqs. 2-4) the parametric-domain integration formula for the normalized system:

$$\frac{d\tilde{\mathbf{x}}_{SS}}{d\tilde{\mathbf{p}}}=-\left(\frac{\partial\mathbf{F}}{\partial\tilde{\mathbf{x}}_{SS}}\right)^{-1}\frac{\partial\mathbf{F}}{\partial\tilde{\mathbf{p}}} \qquad \text{Eq. 12}$$

Taken together, Equations 9-11 clearly indicate that the absolute values of the steady-state concentrations at reference state affect neither the eigenvalue characteristics of the normalized Jacobian, $\partial\mathbf{F}/\partial\tilde{\mathbf{x}}_{ss}$, nor the steady-state bifurcation behavior.

2.2.6 Ensemble Robustness Analysis

Since most engineered pathways have no built-in mechanisms to control expression levels or other kinetic parameters, it is particularly important to design pathways that are structurally robust against moderate perturbations in kinetic parameters, particularly in enzyme expression levels. However, pathways at the design stage do not have all the kinetic parameters. Thus, the designer has to evaluate robustness for a large number of parameter sets (or models). To this end, we

develop the Ensemble Robustness Analysis (ERA) via a combination of two techniques. First, we use the EM approach to construct an ensemble of models that span the feasible parameter space constrained by the desired reference flux. We then apply the continuation method to examine the effects of parameter drifting and detect the bifurcation point of each model (parameter set) in the ensemble (Fig. 1c). For any given perturbation, the percentage of models that are beyond the bifurcation point (loss of stable steady state) represents the probability of system failure (Fig. 1d) corresponding to the network structure and the assumed kinetic rate laws. This use of probability of system failure to evaluate robustness of a network structure is a defining feature of ERA and a unique addition to the existing toolkit of robustness measures (52-54).

2.2.7 Constructing ensembles for NOG and rGC

In this work, we assume that the kinetics of every enzymatic reaction in non-oxidative glycolysis (NOG) and reverse glyoxylate cycle (rGC) obeys the Michaelis-Menten rate law (see Table 1 for the list of rate laws). Since we are designing a non-native pathway without considering other interacting pathways, the absolute flux has no meaning. Hence, the input flux producing OAA in NOG (and PEP in rGC) is taken as 1, whereas all the other fluxes can be determined by solving the steady-state equation $\mathbf{S}\cdot\mathbf{v} = \mathbf{0}$.

To construct an ensemble of normalized models (no concentration data are available) for NOG or rGC, we draw values of $\tilde{K}_m$'s and $\tilde{K}_{eq}$'s uniformly from [0.1 10] and [1 10], respectively; set all the normalized concentrations to 1; and calculate $\tilde{V}_{max}$'s such that each reaction rate V is equal to the flux at the reference state. Once all the kinetic parameters are known, we check the eigenvalues of the normalized Jacobian to ensure that every sampled model is locally stable (*i.e.*, all eigenvalues have a negative real part) at the reference state.

Kinetic Rate Law	# of Substrates	# of Products	Reversible	Rate Expression
Michaelis-Menten	1	1	Yes	$\frac{V_{max}\left(S-\frac{P}{K_{eq}}\right)\left(\frac{1}{K_{ms}}\right)}{1+\frac{S}{K_{ms}}+\frac{P}{K_{mp}}}$
Michaelis-Menten	1	1	No	$\frac{V_{max}}{1+\frac{K_m}{S}}$
Random Bi-Bi	2	2	Yes	$\frac{V_{max}\left(S_1S_2-\frac{P_1P_2}{K_{eq}}\right)\left(\frac{1}{K_{m1}K_{m2}}\right)}{1+\frac{S_1}{K_{m1}}+\frac{S_2}{K_{m2}}+\frac{S_1S_2}{K_{m1}K_{m2}}+\frac{P_1}{K_{m3}}+\frac{P_2}{K_{m3}}+\frac{P_1P_2}{K_{m3}K_{m4}}}$
Michaelis-Menten	2	2	No	$\frac{V_{max}}{1+\frac{K_{m1}}{S_1}+\frac{K_{m2}}{S_2}+\frac{K_{m1}K_{m2}}{S_1S_2}}$
Ordered Uni-Bi	1	2	Yes	$\frac{V_{max}\left(S-\frac{P_1P_2}{K_{eq}}\right)\left(\frac{1}{K_{ms}}\right)}{1+\frac{S}{K_{ms}}+\frac{P_1}{K_{mP_1}}+\frac{P_2}{K_{mP_2}}+\frac{P_1P_2}{K_{mP_1}K_{mP_2}}}$
Ordered Bi-Uni	2	1	Yes	$\frac{V_{max}\left(S_1S_2-\frac{P_1}{K_{eq}}\right)\left(\frac{1}{K_{mS_1}K_{mS_2}}\right)}{1+\frac{S_1}{K_{mS_1}}+\frac{S_2}{K_{mS_2}}+\frac{S_1S_2}{K_{mS_1}K_{mS_2}}+\frac{P_1}{K_{mP_1}}}$

Table 2-1. Michaelis-Menten kinetic rate laws used in this work.

2.3 Applications

2.3.1 Robustness of synthetic non-oxidative glycolysis

ERA is particularly useful for examining the robustness of systems where only the network structure is known, but not the parameters. The first example is the synthetic NOG. Naturally occurring glycolytic pathways such as the Embden-Meyerhof-Parnes (EMP) pathway oxidize glucose to form pyruvate, which is then decarboxylated by the pyruvate dehydrogenase complex to form acetyl-coenzyme A (CoA), a precursor to various cellular constituents and products such as ethanol, 1-butanol, isoprenoids and fatty acids (55). However, one carbon equivalent is lost in the decarboxylation of pyruvate, thereby limiting the theoretical carbon yield to only 66% (Fig.

2a). For an organism that is not naturally equipped to recycle the wasted CO_2, this inherent carbon loss leads to a significant decrease in the carbon yields of fuels and chemicals. A pathway for complete carbon conservation was previously proposed but incompletely demonstrated to exist in nature (56).

Recently, a non-oxidative, synthetic version, termed non-oxidative glycolysis (NOG), was demonstrated conclusively both *in vivo* and *in vitro* to produce stoichiometric amounts of two-carbon (C2) molecule from hexose, pentose and triose phosphates with a 100% carbon yield (16). The net reaction of this pathway results in the formation of three acetyl phosphate (AcP) molecules from one fructose 6-phosphate (F6P) in a redox-neutral manner (Fig. 2a). Phosphoketolase, which was identified many decades earlier (57, 58), is the key enzyme because it is highly irreversible and provides a driving force for NOG. This enzyme, however, is known to react either with F6P to form one AcP and one erythrose 4-phosphate (E4P), or with xylulose 5-phosphate (X5P) to form one AcP and one glyceraldehyde 3-phosphate (G3P). Different homologues of phosphoketolase exhibited different relative activities towards F6P and X5P. This dual activity towards F6P (termed Fpk) or X5P (termed Xpk) has engendered two configurations that form a "basis" to all possible combinations: (1) NOG using only Fpk (Figs. 2b, left panel, and 2c); (2) NOG using only Xpk (Figs. 2b, right panel, and 2d).

The existence of alternative configurations of NOG raises an intriguing question: Are these configurations equally robust? Since at the design level we do not know the kinetic parameters, the ERA approach becomes particularly suitable. In constructing the network model, we did not include interacting pathways in the cell, as we are primarily interested in investigating the difference between the two configurations. Thus, the desired steady-state flux is taken as 1. Such models also reflect the situation in the *in vitro* NOG (16).

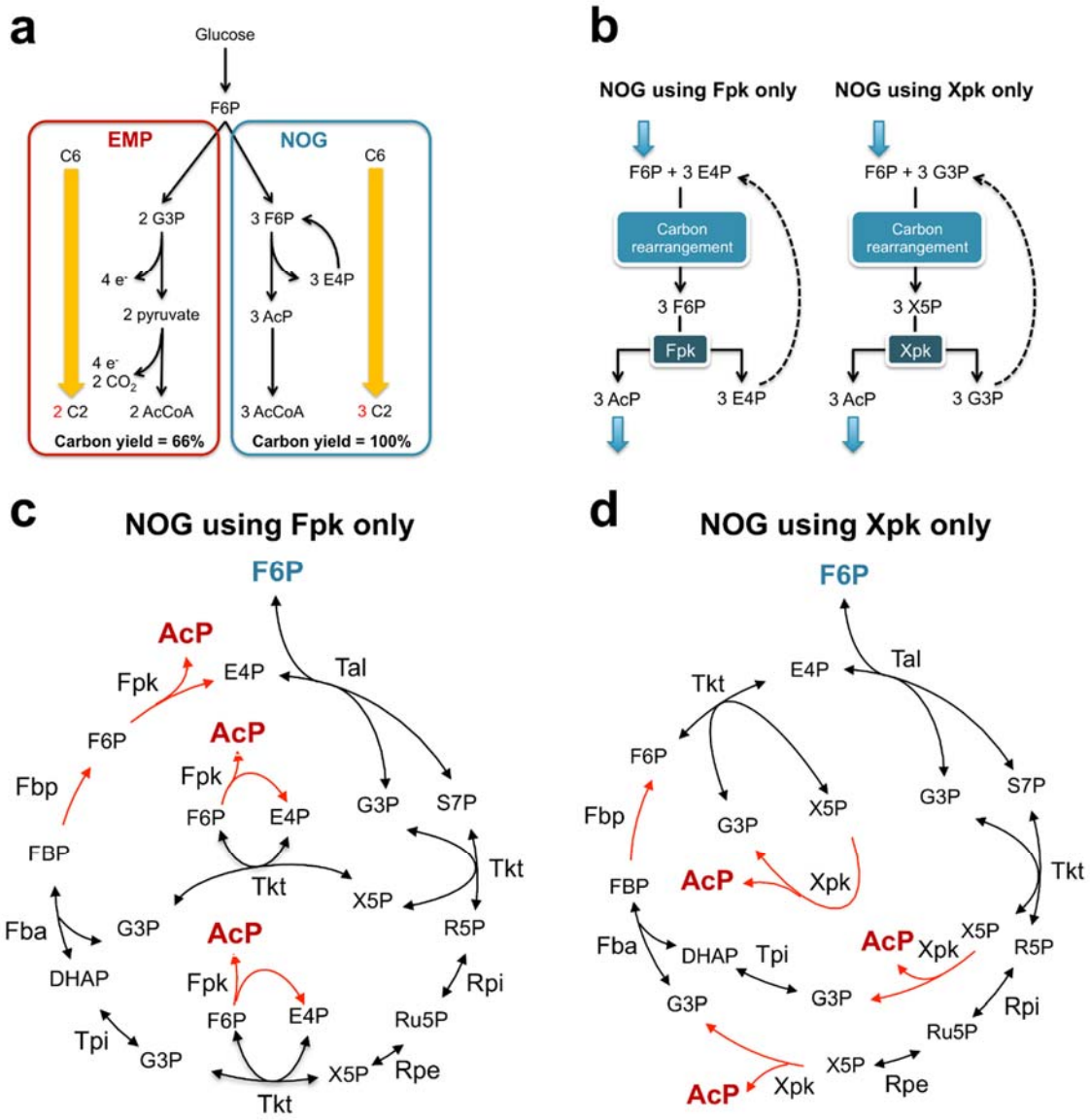

Figure 2-2. Non-oxidative glycolysis. **(a)** The theoretical carbon yield of NOG is 100%, while that of the Embden-Meyerhof-Parnes (EMP) pathway is only 66%. **(b)** The metabolic logic of NOG when only Fpk (left panel) or Xpk (right panel) is involved **(c)** Network configuration of NOG using Fpk only**. (d)** Network configuration of NOG using Xpk only. The red arrows denote the irreversible reactions that drive the pathway. Tal, transaldolase; Tkt, transketolase; Rpi, ribose 5-phosphate isomerase; Rpe, ribuloase 5-phosphate epimerase; Tpi, triose phosphate isomerase; Fba, FBP aldolase; DHAP, dihyroxyacetone phosphate; R5P, ribose 5-phosphate; Ru5P, ribulose 5-phosphate; S7P, sedoheptulose 7-phosphate. Figures are adapted from (16).

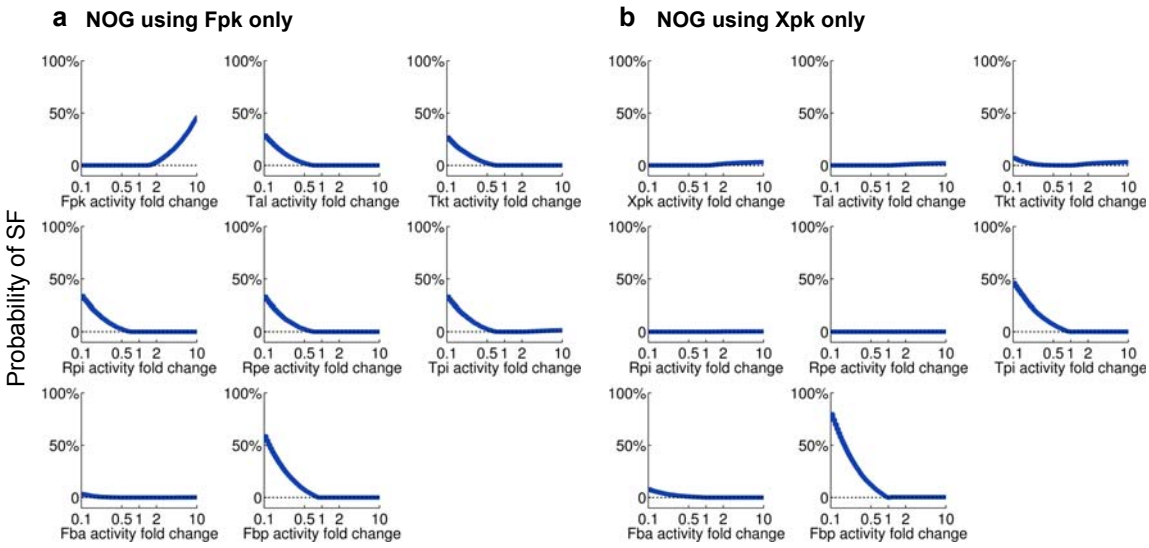

Figure 2-3. Robustness profiles of Fpk NOG (a) and Xpk NOG (b). The blue lines represent the probability of system failure (SF). Deviation of the blue line from the black dashed line (probability of SF equals zero) indicates a reduction of robustness. Enzyme activity "fold change" is defined as the ratio of the activity in the perturbed state over that in the reference state

We constructed an ensemble of 10,000 ODE models for each configuration in Fig. 2c & d using Michaelis-Menten kinetics. Other lumped kinetics or elementary reaction kinetics can be used in a similar manner, but they come with more parameters. After an ensemble was constructed, we examined each model's ability to maintain a non-trivial steady state under drifting in enzyme activity using the continuation method. Figure 3a & b show the robustness profiles of the two basis configurations in Fig. 2c & d, respectively. The results suggest that the configuration using Fpk (Fig. 2c) has an increased probability of system failure if the Fpk level is too high (Fig. 3a). In those cases, F6P will be depleted by Fpk, leaving no substrate for transaldolase and accumulation of E4P as a "kinetic trap" (Figure 4). This loss of steady state causes system failure both *in vivo* and *in vitro*. On the other hand, the Xpk system (Fig. 2d) is highly insensitive to the overexpression of Xpk (Fig. 3b). Overall, the Xpk configuration is considerably more robust against enzyme overexpression than the Fpk configuration (Fig. 3a & b). Thus, in this configuration, one can largely focus on overexpression of each individual enzyme in the pathway without compromising

robustness. On the down-regulation side, the Xpk system is also more robust than the Fpk system. These results demonstrate that robustness can be a useful criterion for comparing alternative designs of a synthetic pathway even when only stoichiometry is available.

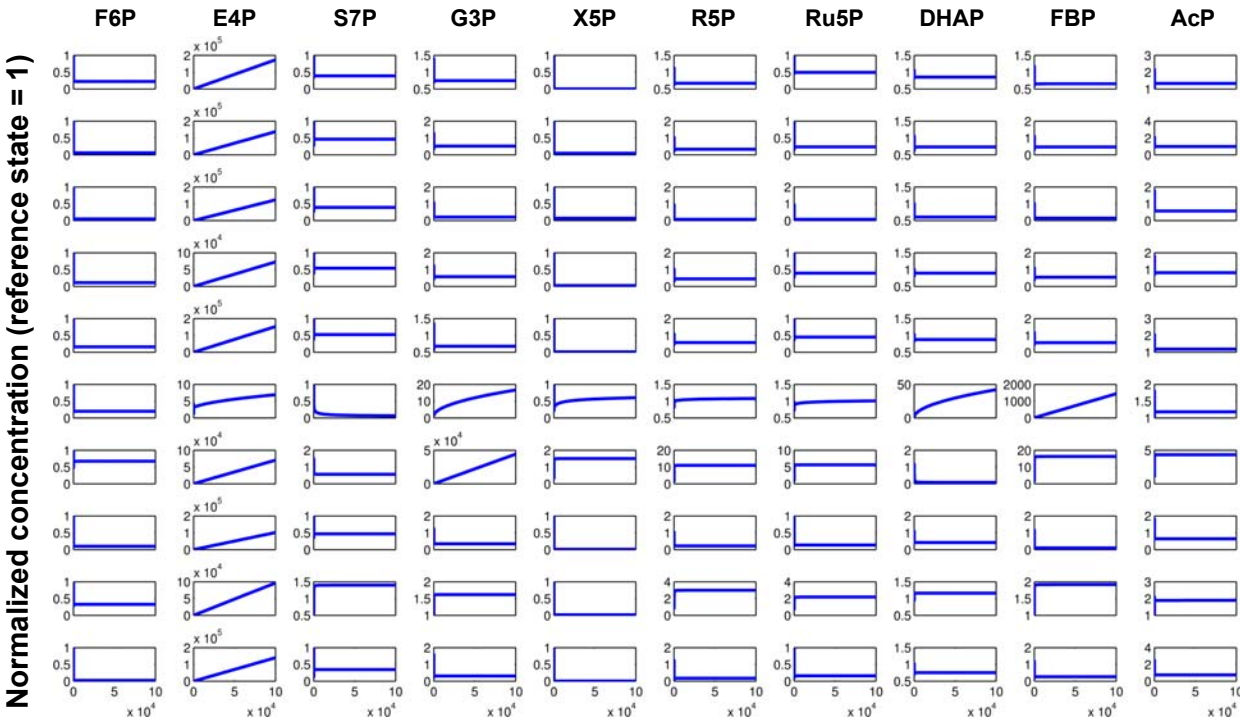

Figure 2-4. Representative response of Fpk NOG (Fig. 3c) to Fpk overexpression. The disappearance of a non-trivial steady state, which features accumulation of metabolites such as E4P, is found in 10 differently parameterized models of Fpk NOG. Since we focus on the demonstration of qualitative outcome when Fpk activity is too high (*e.g.*, 5-fold), the absolute values of the concentrations at reference state have no bearing and are taken as 1. F6P, fructose 6-phosphate; E4P, erythrose 4-phosphate; S7P, sedoheptulose 7-phosphate; G3P, glyceraldehyde 3-phosphate; X5P, xylulose 5-phosphate; R5P, ribose 5-phosphate; Ru5P, ribulose 5-phosphate; DHAP, dihydroxyacetone phosphate; FBP, fructose 1,6-bisphosphate; AcP, acetyl phosphate.

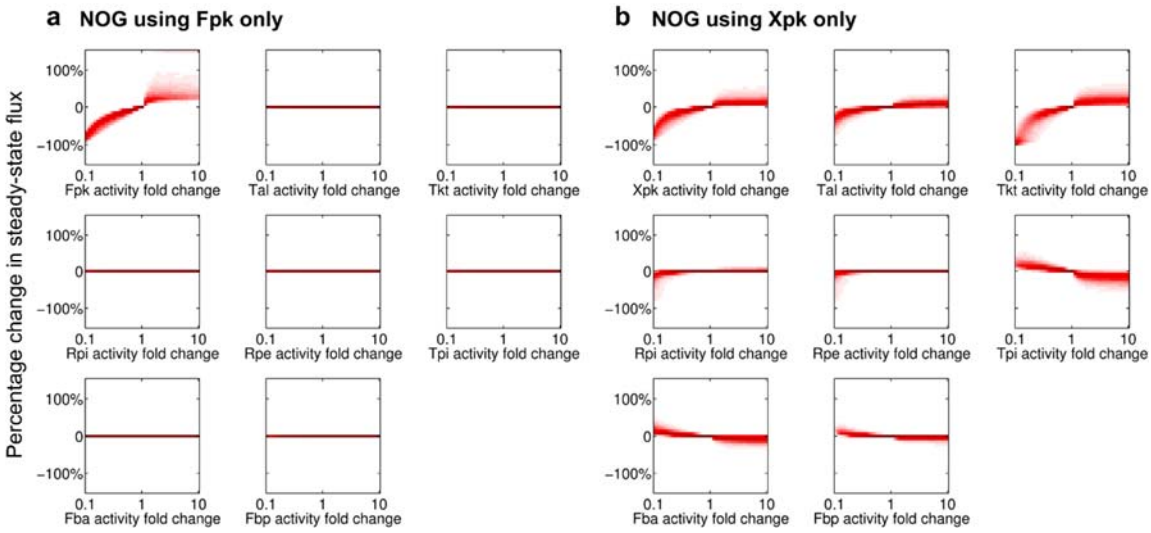

Figure 2-5. Effect of enzyme overexpression or knockdown on NOG flux. In each panel, the composite solution of 10,000 models at various levels of perturbation is shown. The intensity of color indicates the density of models, or the likelihood of occurrence. Enzyme activity "fold change" is defined as the ratio of the activity in the perturbed state over that in the reference state.

2.3.2 ERA identifies potential targets for improving pathway performance

The continuation method in ERA also yields the steady-state metabolite and flux solutions corresponding to kinetic parameter perturbations. The solutions of the ensemble can then guide the selection of potential targets for improving performance, based on the likelihood of improvement in the ensemble. Fig. 5 shows the composite solution of 10,000 models of NOG after different degrees of enzyme activity perturbation. Only the models with a stable steady state are recorded. The intensity of color in Fig. 5 indicates the density of models, or likelihood of results. For instance, in the case of Fpk NOG, overexpression of Fpk is the only single-perturbation option for increasing the production of AcP (Fig. 5a). However, this strategy is constrained by the trade-offs between performance and robustness, as a high Fpk activity also results in a high likelihood of system failure (Fig. 3a). In contrast, the Xpk NOG configuration has more targets for increasing

the AcP production flux (Fig. 5b) than the Fpk NOG configuration (Fig. 5a), and is more robust (Fig. 3b). Therefore, the Xpk NOG is less affected by the trade-offs between performance and robustness when considering single-enzyme perturbations. Together, these results demonstrate the ability of ERA to guide robust target selection for performance improvement. Of course, in vivo systems could use multiple but ERA

2.3.3 Robustness of a synthetic reverse glyoxylate cycle

In another example, we investigate the robustness of a pathway that can convert a four carbon (C4) metabolite, such as oxaloacetate (OAA), to two molecules of acetyl-coA (AcCoA) (17). In heterotrophs like *E. coli*, sugars can only be metabolized to AcCoA via decarboxylation of pyruvate through enzymes such as the pyruvate dehydrogenase complex, pyruvate oxidoreductase, or pyruvate formate lyase. Theoretically, it is possible to incorporate one more carbon to phosphoenolpyruvate or pyruvate to form OAA, and then split the C4 compound into two molecules of AcCoA, via a synthetic pathway that involves a reverse glyoxylate cycle (rGC) (Fig. 6a). This pathway can then be integrated with glycolysis via the native phosphoenolpyruvate

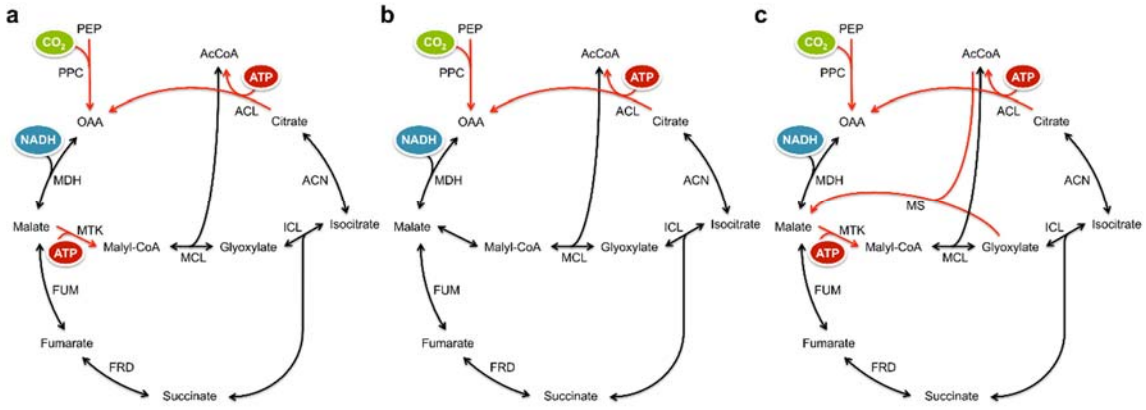

Figure 2-6. Reverse glyoxylate cycle (**a**) rGC converts one molecule of C4 metabolite such as OAA to two molecules of AcCoA. Black double arrows denote reversible reactions, whereas red arrows denote irreversible reactions. (**b-c**) rGC with modifications to achieve a hypothetical (**b**) or functionally (**c**) reversible conversion of malate to glyoxylate. PEP, phosphoenolpyruvate; MDH, malate dehydrogenase; ICL, isocitrate lyase; ACN, aconitase. Figures adapted from (17).

carboxylase (PPC), to permit the conversion of one molecule of glucose to four molecules of AcCoA, with the fixation of two CO_2 reduced by additional reducing power.

The key challenge of realizing rGC in *E. coli*, however, is to overcome the intrinsic thermodynamic barrier: both malate synthase (MS) and citrate synthase (CS) are known to be active only in the AcCoA assimilating direction, but not in the AcCoA synthesizing direction (59). To convert malate into glyoxylate and AcCoA, heterologous expression of a malate thiokinase (MTK) and a malyl-CoA lyase (MCL) provides the "reverse MS" activity. In addition, expression of a heterologous ATP-citrate lyase (ACL) allows for conversion of citrate to OAA and AcCoA (17). Since both MTK- and ACL-catalyzing reactions are coupled to ATP hydrolysis, they are irreversible and thus provide the necessary driving force for reversing the glyoxylate cycle in *E. coli*.

With these enzymes, we investigate whether rGC is robust to moderate drifting in enzyme activities. To do this, we similarly constructed an ensemble of 10,000 rGC models with the input flux equal to 1 and no interaction with other pathways. We then used the continuation method to determine how likely the drifting of an enzyme activity is to cause the loss of a steady sate and therefore system failure. Surprisingly, the basic rGC configuration (Fig. 6a) has severe robustness problems with respect to fumarase (FUM) and fumarate reductase (FRD) on the down-regulation side as well as MTK on the overexpression side (Fig. 7a). In those cases, increases in the activities of MTK or decreases in FUM or FRD activities lead to the accumulation of glyoxylate and malyl-CoA (Figure 8), which serves as another example of a kinetic trap leading to system failure.

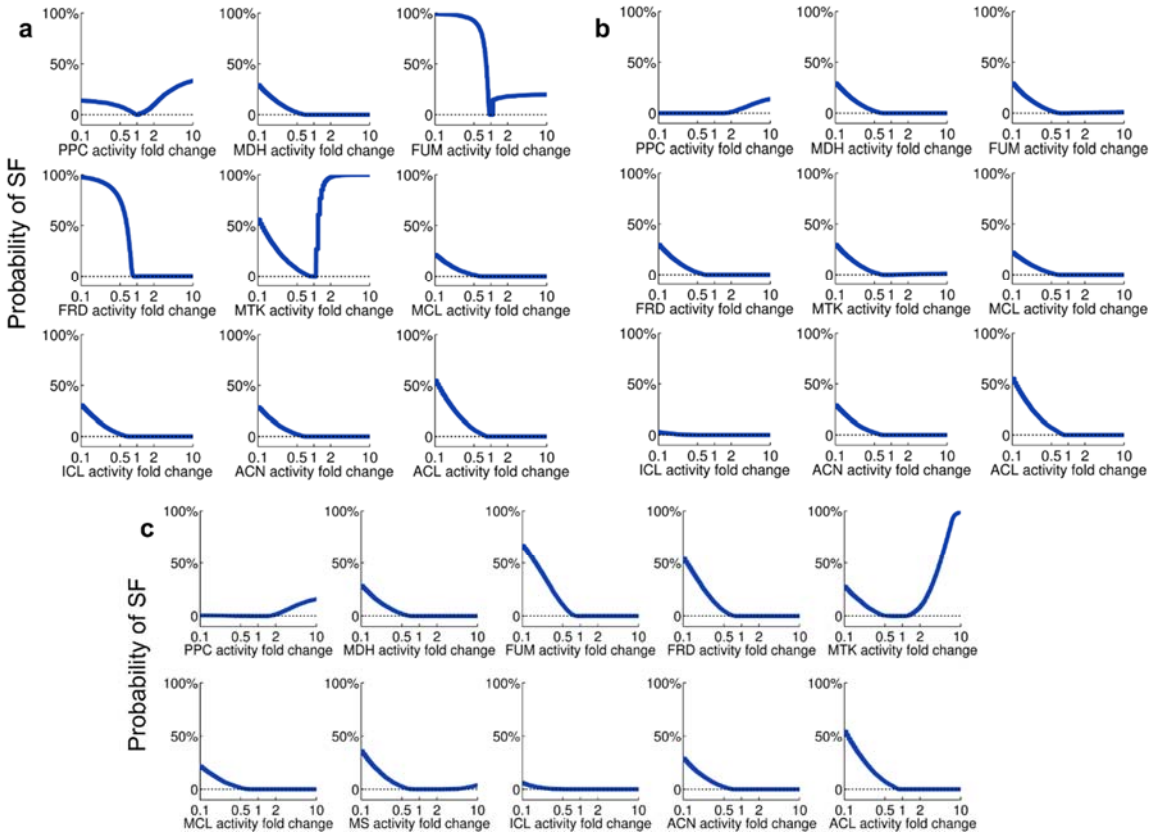

Figure 2-7. Robustness profiles of rGC. The robustness profiles of three rGC configurations in Fig. 6a-c are plotted, respectively in Fig. 7a-c, in the same fashion as Fig. 3.

Given the above results, we reasoned that the difficulty of rGC is the flux splitting at the node of malate, where the influx needs to be split *equally* at steady state. If the split is achieved through an irreversible enzyme (*e.g.*, MTK), then feasible kinetic parameters will be restricted, thereby depriving them from the flexibility to tolerate perturbations. To test this hypothesis, we examined (i) a rGC with a hypothetical reversible step converting malate to glyoxylate and AcCoA (Fig. 6b), and (ii) the native MS expressed along with MTK and MCL to achieve the functionally reversible conversion from malate to glyoxylate (Fig. 6c). For each configuration, we again constructed an ensemble of 10,000 models and used continuation method to evaluate their robustness.

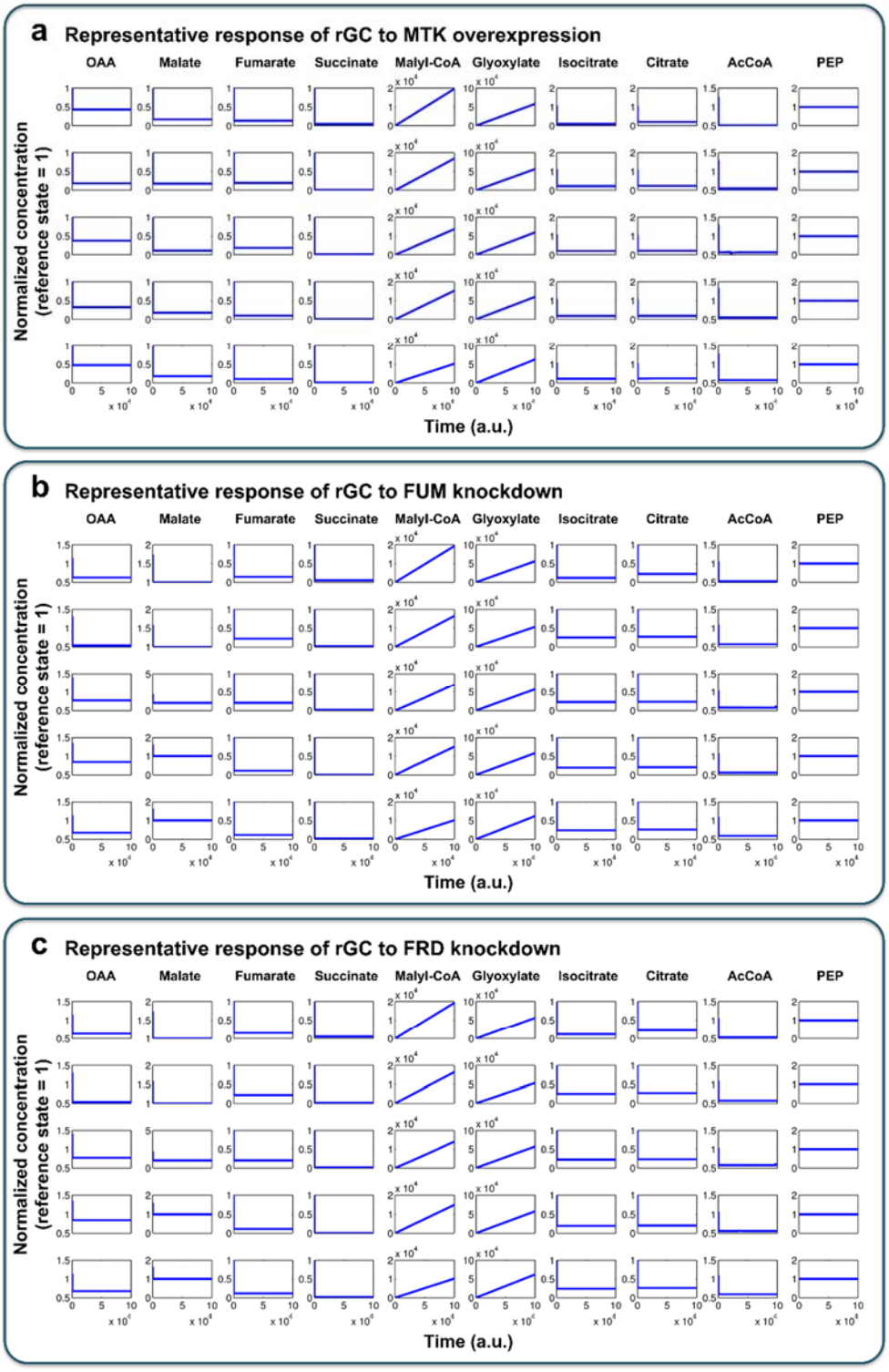

Figure 2-8. Representative response of rGC (Fig. 7a) to MTK overexpression (a), FUM knockdown (b), or FRD knockdown (c). In each case, the disappearance of a non-trivial steady state, which features accumulation of malyl-CoA and glyoxylate, is found in five differently parameterized rGC models. Since we focus on the demonstration of qualitative outcome when MTK activity is too high (*e.g.*, 5-fold) or FUM (FRD) activity is too low (*e.g.*, 0.2-fold), the absolute values of the concentrations at reference state have no bearing and are taken as 1. MTK, malate thiokinase; FUM, fumarase; FRD, fumarate reductase.

Indeed, as shown in Fig. 7b, the previously seen robustness problems in rGC disappear when MTK is assumed to be reversible. Unfortunately, MTK cannot be engineered to become reversible because of thermodynamic constraints. Thus, one has to use a functionally reversible configuration (Figs. 6c, 7c), which is more robust than the completely irreversible configuration (Figs. 6b, 7b). Indeed, this configuration is likely to be the one that was experimentally demonstrated (17). Although this possibility was not specifically addressed by the authors (17), the native MS gene glcB was present in the chromosome of the strains tested, providing possible support of our analysis.

2.4 Conclusion

The robustness of non-native metabolic pathways is an important issue that needs to be addressed. Even in a biochemically feasible network with favorable thermodynamics, if inappropriate expression levels or other kinetic parameters are used, the pathway may enter a region where no stable steady state exists. This situation leads to system failure and is a more serious problem than just the drifting of performance. Therefore, in the design of modified or *de novo* metabolic networks, it is important to choose a network structure that is robust against system failure. Unfortunately, complete kinetic parameters are unavailable in the design phase. Thus, ERA uses an ensemble approach to examine the robustness based on perturbations of all feasible parameter sets. If the probability of system failure upon perturbation is low based on large sampling in the feasible kinetic space, then the network structure can be deemed robust. The use of a continuation method in ERA enables the detection of bifurcation points, as well as speedy examination of a large number of models. This capability of ERA in examining a large number of models for the probability of system failure is instrumental for the design of not only metabolic models, but also signal transduction pathways where ensemble methods have seen increased popularity (60).

Here we demonstrate the utility of the ERA approach in designing non-native central metabolic pathways that are important for carbon conservation. The two examples, NOG and rGC, are complex circular pathways that have not evolved naturally in the host. Therefore, the robustness issue is particularly important and demands further attention. Although only single-enzyme perturbations were considered in these examples, the perturbation of multiple enzymes could have an impact on the trade-off between performance and robustness and could also be examined by ERA. As the design and modeling of metabolic pathways gains importance, we expect that ERA will play an important role in such activities.

3 An entropy-like index of bifurcational robustness for metabolic systems

Natural and synthetic metabolic pathways need to retain stability when faced against random changes in gene expression levels and kinetic parameters. In the presence of large parameter changes, a robust system should specifically avoid moving to an unstable region, an event that would dramatically change system behavior. Here we present an entropy-like index, denoted as *S*, for quantifying the bifurcational robustness of metabolic systems against lost of stability. We show that *S* enables the optimization of a metabolic model with respect to both bifurcational robustness and experimental data. We then demonstrate how the coupling of Ensemble Modeling and *S* enables us to discriminate alternative designs of a synthetic pathway according to bifurcational robustness. Finally, we show that *S* enables the identification of a key enzyme contributing to the bifurcational robustness of yeast glycolysis. The different applications of *S* demonstrated illustrate the versatile role it can play in constructing better metabolic models and designing functional non-native pathways.

3.1 Introduction

The major role of most metabolic systems is to support cellular growth, maintenance, or adaptation without losing stability despite perturbations in the environment (41, 61). When the environmental or physiological conditions change, gene expression levels or kinetic parameters may drift outside their typical working ranges and lose stability. The stochastic nature of transcriptional and translational mechanisms (62) is one such source of noise and a robust system would maintain homeostasis despite the perturbations. Failure to retain a stability may lead to accumulation of toxic metabolites or depletion of essential intermediates. This detrimental state could cause growth arrest and has been linked to many diseases such as diabetes and cancer (63). In metabolic engineering, this situation leads to loss of production or cell death.

Non-linear system behavior changes qualitatively and dramatically when parameters cross a bifurcation point and exhibits instability or multiplicity of steady states. Thus, a necessary but insufficient requirement for stable or damped-oscillatory metabolic pathway design is to avoid crossing a bifurcation point in the presence of random perturbations in gene expression levels and environmental conditions that change kinetic parameters. For metabolic systems, a fixed-point bifurcation may cause the stable steady state (or fixed-point) to become unstable (64-66), or mark the emergence of undamped oscillations (50, 51) or multiple steady states (67) .

The distance away from an unstable region is defined as the bifurcational robustness (64), which measures the ability to return to a fixed point upon perturbation. Thus, building a theoretical foundation of robustness, and in particular defining a simple way to quantify it, represents a key challenge in systems biology (54). For small, local perturbations, stability criteria are well defined using linear stability analysis (68). For large perturbations, one must explore global properties of the system. It is important to make the distinction between bifurcational robustness, which pertains to stability, and local sensitivity (or robustness), which quantifies the changes in performance (flux, period of oscillation) as a function of small changes in parameters within the same dynamic regime.

Natural metabolic pathways are presumed to be at least bifurcationally robust against stochastic changes in protein expression levels. Thus models of natural metabolic pathways need to be similarly robust. However, there is no quantitative way to characterize bifurcational robustness in the presence of random parameter changes. Without a quantitative index, optimization of models for bifurcational robustness becomes difficult, if not impossible. Therefore, our goal here is to develop a quantitative index for bifurcational robustness, and show that such an index enables the optimization of the bifurcational robustness of metabolic models. The developed index is easy to

compute and applies to metabolic systems of various scale and complexity. Interestingly, the mathematical form of our robustness index resembles the definition of entropy in thermodynamics and information theory (69). We show that this entropy-like index, denoted as *S*, negatively correlates with bifurcational robustness. Metabolic systems with a small *S* are highly robust against bifurcation, and are more likely to retain a steady state under random perturbations affecting every enzyme than systems with a large *S*.

The utility of *S* was demonstrated through three examples. First, we show that the bifurcational robustness of a native pathway model can be significantly improved by applying a multi-objective optimization with *S* as an objective. Second, we show that the integration of *S* with EMRA (64) is able to discriminate, without any prior knowledge of kinetic parameters, the difference in bifurcational robustness between two configurations of a non-native metabolic pathway (16). Possible sources of pathway failure were also identified. Finally, by quantifying *S* in a series of yeast glycolysis models incorporating different features, we identified pyruvate decarboxylase as a key enzyme determining the robustness of yeast glycolysis, a finding consistent with earlier studies (70, 71). Together, our results demonstrate that *S* may serve as an unbiased standard by which the bifurcational robustness is judged.

3.2 Lack of robustness of existing metabolic models

In a survey of the robustness of existing metabolic models, we simulated natural perturbations and recorded the response of thirteen kinetic models of metabolic pathways with fitted parameters (Fig. 1). This *in silico* experiment was designed to mimic the real biological situation where protein expression levels, which affect kinetic parameters, vary randomly and non-specifically (62). Since natural metabolic pathways are presumed to be bifurcationally robust in such situations, the computational models of these metabolic systems need to be similarly robust. To our surprise, the

selected models displayed varying degrees of robustness against bifurcation. Some models are very robust and almost always retain stability after perturbation, even though their steady-state flux and metabolite concentrations are changed. Others respond poorly even to moderate perturbations (Fig. 1) where the system becomes unstable with some metabolites accumulating or vanishing, leading to system failure. If one accepts the assumption that natural systems are robust against enzyme expression perturbations, these models do not reflect this assertion. This unexpected result highlighted the need for the optimization of metabolic models with respect to bifurcational robustness, which in turn calls for the development of a quantitative robustness index.

3.3 Searching for a robustness index

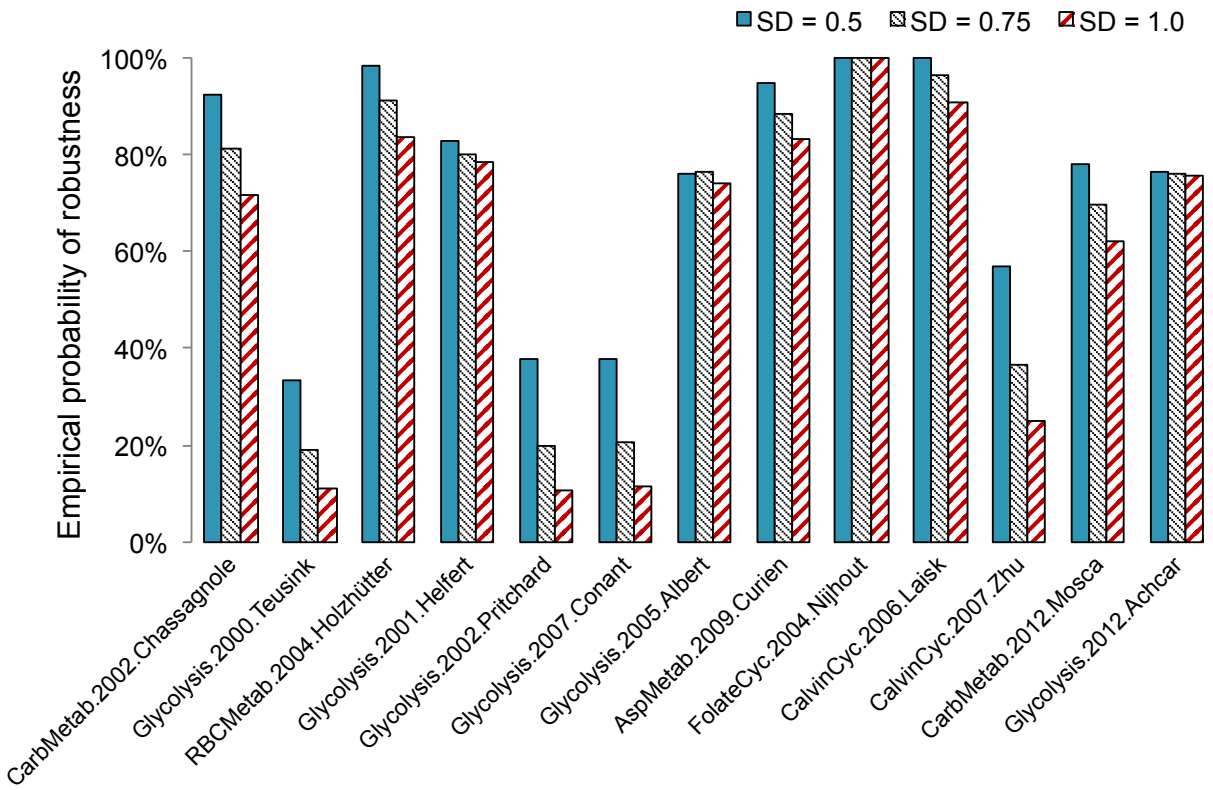

Figure 3-1. Existing models have robustness problems to different extent. The empirical probability of robustness, calculated as the fraction of 10,000 randomly perturbed models that retain a steady state, is shown for each of the 13 BioModels database models (1-13). SD, the standard deviation of log fold change. CarbMetab, carbohydrate metabolism; RBCMetab, red blood cell metabolism; AspMetab, aspartate metabolism; FolateCyc, folate cycle; CalvinCyc, calvin cycle.

The natural perturbation of environmental or physiological conditions often affects the expression levels of many genes, which in turn affect the kinetic parameters of all enzymes. Thus, an appropriate description of bifurcational robustness should focus on the probability that a metabolic system will not cross a bifurcation point and retain a stable steady state when every enzyme is randomly perturbed. Although this probability of retaining stability, denoted as P_{SS}, is a function of all enzymes, a simple approximation can be obtained under the assumption that there is no "crosstalk" between enzymes. That is, the probability of stability retention under the perturbation of a single enzyme is independent of other enzymes. In this simplest case, the expression for P_{SS} reduces to

$$P_{SS} \approx \prod_{i=1}^{n} p_i, \qquad \text{Eq. 1}$$

where p_i denotes the probability that the system will retain stability if only enzyme i is subject to variation.

When the environmental perturbation affects only one enzyme, the probability that the system will cross a bifurcation point, denoted as p_i, is determined by the area bounded by the bifurcation points and the probability density function (Fig. 2A & B). Fortunately, the bifurcation points can be readily determined based on the continuation method described previously (64). As an example, Figure 2C shows the bifurcation points with respect to each enzyme in Teusink *et al*'s yeast glycolysis model (12). Clearly, the model can tolerate wide variation in some enzymes, such as glucose 6-phosphate isomerase (PGI) and phosphoglycerate kinase (PGK), whereas a moderate perturbation in other enzymes, such as ATPase, can lead to a bifurcation and the loss of stability.

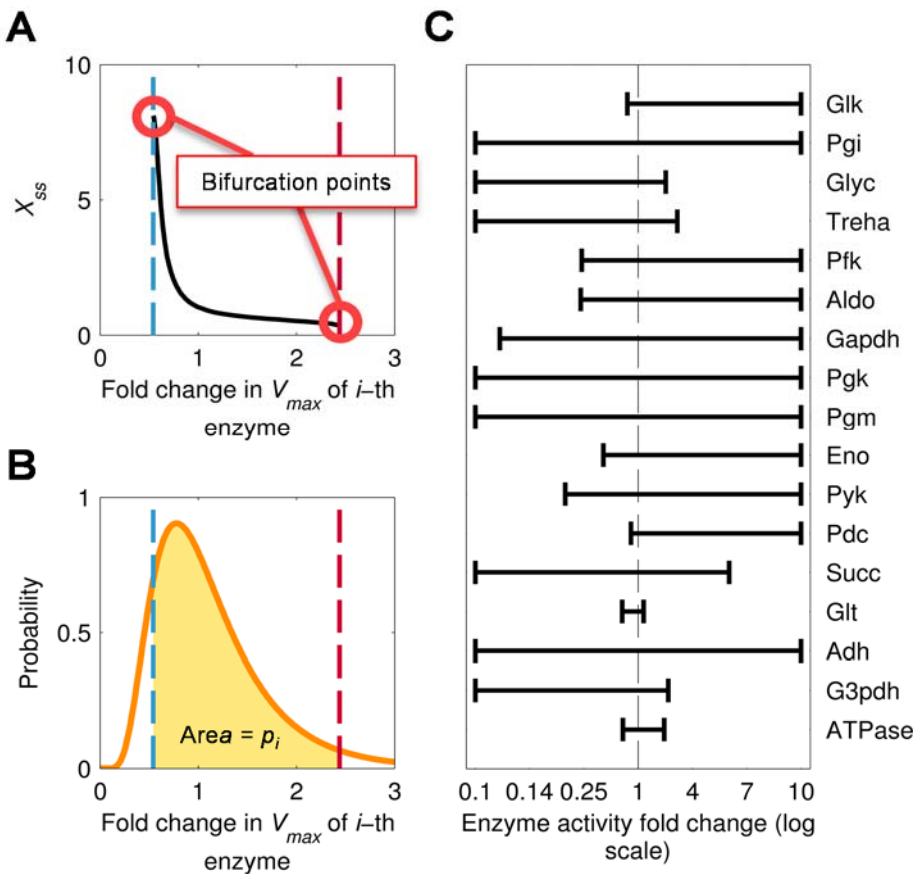

Figure 3-2. The continuation method enables the detection of bifurcation points. (**A**) Starting from a default steady state, the continuation method traces the trajectory of X_{SS} (the steady-state value of some metabolite concentration X) as it varies according to enzyme activity levels. In this example, the system loses stability when the enzyme activity is increased by over 2.5-fold or decreased by over 50%. (**B**) Given the probability density function (pdf; orange curve), we can calculate p_i (the probability of retaining a steady state when i-th enzyme is subject to a random perturbation) as the area under the pdf and between the bifurcation points (red and blue dashed lines). (**C**) The bifurcation points of V_{max} define the boundaries of single-enzyme perturbations. Abbreviations are defined at the end of the chapter.

For any metabolic model with all parameters specified or fitted, we can calculate p_i for every enzyme and use the formula in Eq. 1 to approximate P_{SS}. Figure 3 shows the performance of Eq. 1 in approximating the P_{SS}'s of the 13 metabolic models discussed above when compared to the Monte Carlo simulation. Clearly, the approximation yields a similar trend as the simulation results, but it tends to underestimate the true probability and may not be an appropriate index.

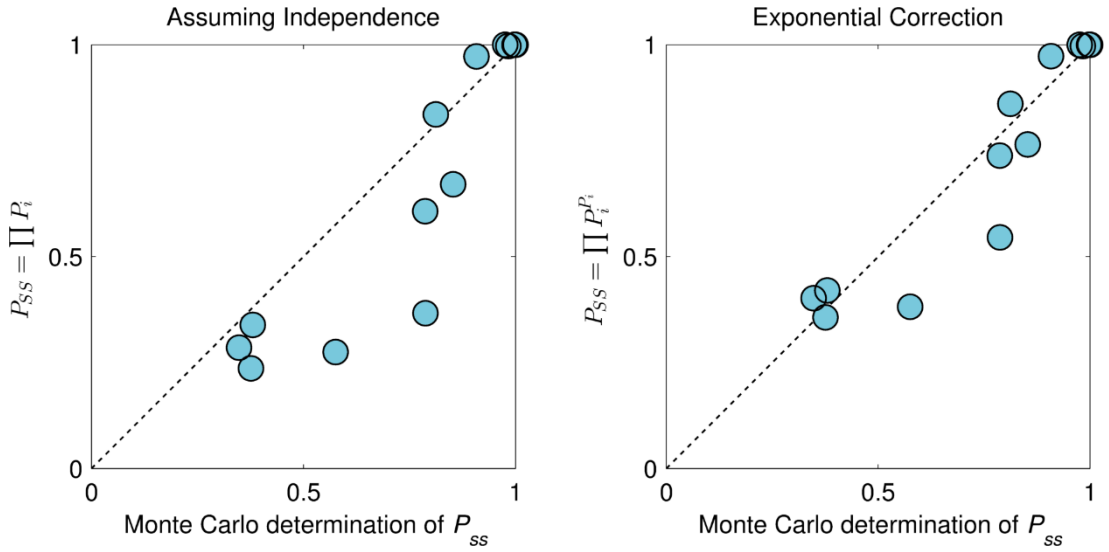

Figure 3-3. Comparison of the robustness (Probability of retaining stability, Pss) determined by in silico experiment and two approximations. (A) The first theoretical approximation was $P_{SS} \approx \prod p_i$. **(B)** The second theoretical approximation was $P_{SS} \approx \prod p_i^{p_i}$. In both cases the Monte Carlo approximation of P_{SS} was calculated by randomly perturbing1 every enzyme.

In an attempt to remedy the observed underestimation from the independence assumption, we applied an exponential correction factor to each individual probability:

$$P_{SS} \approx \prod_i p_i^{p_i} \,. \qquad \text{Eq. 2}$$

With an exponential correction, our rationale is to increase the value of p_i since the effect of crosstalk is more likely to be strong when p_i becomes smaller. Figure 3B shows the performance of this new approximation against the Monte Carlo simulation. Compared to the approximation under an independence assumption, Eq. 2 yields an improved correlation with the simulation results. Since determination of crosstalk between every enzyme is technically challenging, if not impossible, we believe that Eq. 2 provides a reasonable approximation without increasing computational costs. Interestingly, the mathematical form of Eq. 2 resembles the definition of entropy in thermodynamics and information theory (69), except that the exponent does not have a negative sign.

Here we propose an entropy-like robustness index, denoted as *S*, which corresponds to the negative of the logarithm of Eq. 2:

$$S = -\sum_i p_i \log(p_i). \qquad \text{Eq. 3}$$

Like entropy, *S* also enjoys the additive property as do thermodynamic and information theoretic entropy. That is, the system-level robustness *S* can be regarded as a simple sum of the enzyme-level robustness, $S_i = -p_i \cdot \log(p_i)$, which is determined solely by p_i. One difference between *S* and the thermodynamic entropy is that each pi is not mutually exclusive. Other thermodynamic properties or information theoretic properties of entropy do not easily carry over to the robustness index *S*. In robustness, the lower *S* the more robust the system is.

To test the performance of *S* as an index of bifurcational robustness, we calculated *S* for 13 metabolic models and compared those values to the Monte Carlo simulation of random perturbations (Figure 4). As expected, for the 13 models considered, *S* negatively correlates with bifurcational robustness. That is, models with a small *S* are highly robust against bifurcation, and are more likely to retain a steady state under random perturbations than models with a large *S*. In

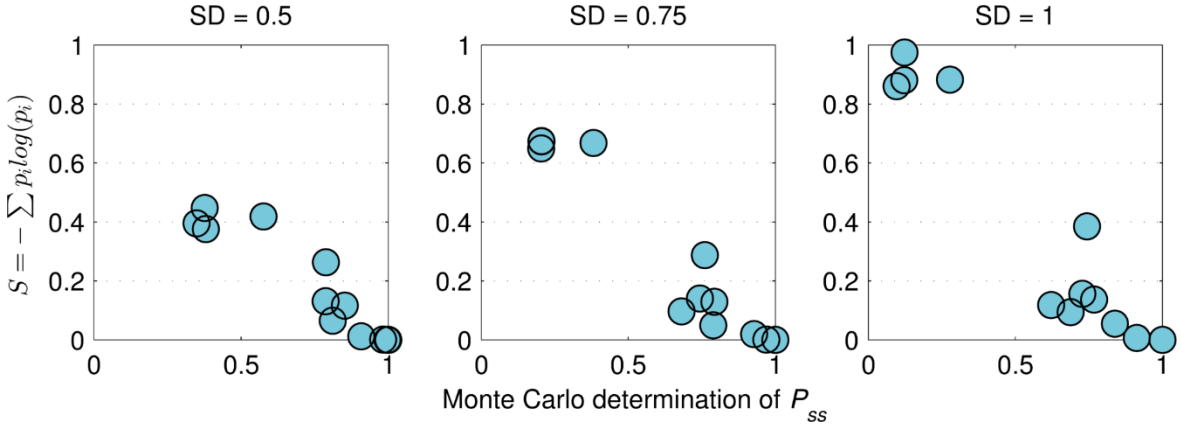

Figure 3-4. *S* is a proper index of bifurcational robustness for metabolic systems. For the 13 BioModels database models (blue filled circles; Methods), *S* decreases with increasing probability of stability retention (P_{SS}), which is determined by Monte Carlo simulation.

fact, the difference in *S* between robust and non-robust systems becomes more apparent when a higher perturbation level is tested. These results suggest that *S*, which can be calculated efficiently using the continuation method (64), may serve as an unbiased standard of bifurcational robustness. In the following sections, we will demonstrate how *S* can be used to: (i) optimize the bifurcational robustness of metabolic models with fitted parameters; (ii) compare the bifurcational robustness of alternative synthetic pathway designs when parameters are unknown; and (iii) identify key features determining the bifurcational robustness of a metabolic system.

3.4 Applications

3.4.1 Parameter optimization using S

Given its scalar nature, *S* can be readily incorporated into commonly used optimization algorithms for parameter fitting. To demonstrate this functionality, we re-fit the parameters of Teusink *et al*'s yeast glycolysis model (12) by using *S* as an optimization objective. In this particular case, we applied a multi-objective optimization so as to simultaneously (i) minimize the discrepancy between available data (metabolite concentrations and fluxes) and model predictions, and (ii) minimize the model's *S* value (See Methods). By incorporating *S* in the objective function, this algorithm was indeed able to significantly improve the bifurcational robustness of an otherwise non-robust model (Figure 5A). More importantly, neither metabolite concentrations nor fluxes required anything larger than a 2-fold change to accomplish this (Figure 5B). These results demonstrate that *S* enables the optimization of bifurcational robustness of a metabolic model and that such robustness optimization can be readily integrated into any parameter-fitting routine.

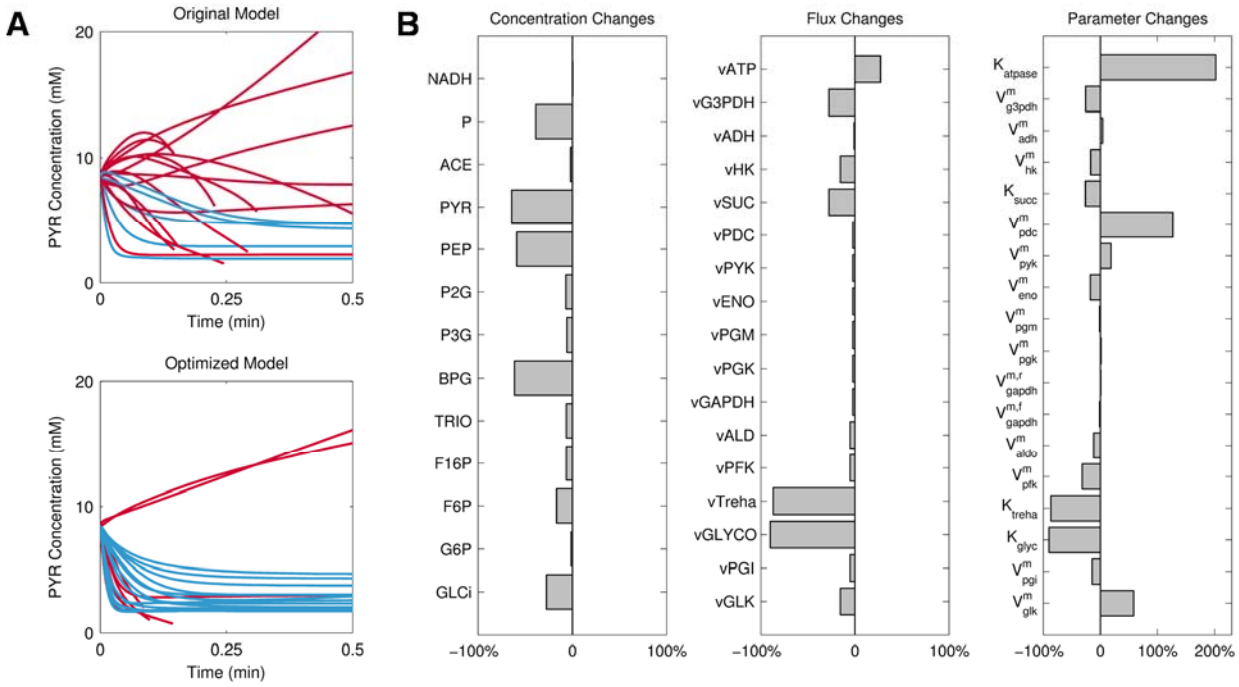

Figure 3-5. Optimization incorporating *S* returns a model with significantly improved robustness. **(A)** Comparison of the bifurcational robustness in the original and the optimized model by random perturbation. (—)Stability retained; (—) Stability lost **(B)** Percentage change in steady-state metabolite concentrations, steady-state fluxes and kinetic parameters between the best model returned by optimization and Teusink *et al.*'s model. Abbreviations are defined at the end of the chapter.

3.4.2 Robustness index in the design of non-native pathways

Besides the optimization of the bifurcational robustness of existing models, *S* can be useful in non-native pathway design even when the kinetic parameters are unavailable. To address the uncertainty of kinetic parameters, we applied the calculation of *S* to an ensemble of models representing the feasible kinetic space (22-24). Such an ensemble approach has recently been adopted to evaluate the bifurcational robustness of non-native pathways and to identify configurations that are more likely to be functional (64).

Here we demonstrate the utility of *S* in non-native pathway design using two configurations of a synthetic non-oxidative glycolysis (NOG)(16): Fpk-NOG (Fig. 6A) and Xpk-NOG (Fig. 6B). Fpk-NOG contains a specific homolog of phosphoketolase (termed Fpk) that only reacts with fructose 6-phosphate (F6P), whereas the phosphoketolase in Xpk-NOG (termed Xpk) only reacts with

xylulose 5-phosphate (X5P). For each configuration, we constructed an ensemble of 10,000 models by random sampling (Methods) and calculated the distributions of S_i. As shown in Fig. 6C, both Fpk-NOG and Xpk-NOG are quite sensitive to the changes in the activity of triose phosphate isomerase (Tpi) and fructose 1,6-bisphosphatase (Fbp) (Fig. 6C). Nevertheless, the Fpk-NOG is considerably less robust than Xpk-NOG as it is also sensitive to many other enzymes (Fig. 6C; left column). This conclusion is also confirmed by Fig. 6D, where the average S of each configuration is visualized as the stacked contributions of average S_i. Although Xpk-NOG as a whole has a lower average S than Fpk-NOG, the high average S_i of Tpi and Fbp might indicate a potential problem during strain construction. As this example illustrates, the coupling of Ensemble Modeling with the calculation of S allows us to assess the robustness of a pathway design and identify possible causes of failure.

3.4.3 Determining the cause of non-robustness

Another possible use of S is in the identification of key features contributing to the bifurcational robustness of a metabolic system. To demonstrate this utility, we used van Heerden *et al.'s* model of yeast glycolysis (14) as an example. This model is adapted from the model developed by Teusink *et al.* (12) with five major changes:

1. Hexokinase (HK) inhibition by glucose 6-phosphate (G6P)
2. Consideration of phosphate as a free variable
3. Activation of pyruvate kinase (PYK) by fructose 1,6-bisphosphate (FBP)
4. A 6.1 fold rise in the V_{max} of pyruvate decarboxylase (PDC)
5. Trehalose and glycogen fluxes were considered as functions of G6P

Although these modifications are seemingly minor, the new model has a significantly lower S than the original model of Teusink *et al.* (Figure 7, inset). This is particularly interesting because other adaptations of Teusink *et al.*'s model, such as Pritchard *et al.*'s (11) and Conant *et al.*'s (4) glycolysis models, do not show a similar improvement in bifurcational robustness (Fig. 1). This

result suggests that some modifications made by van Heerden *et al.* are particularly important for robustness improvement.

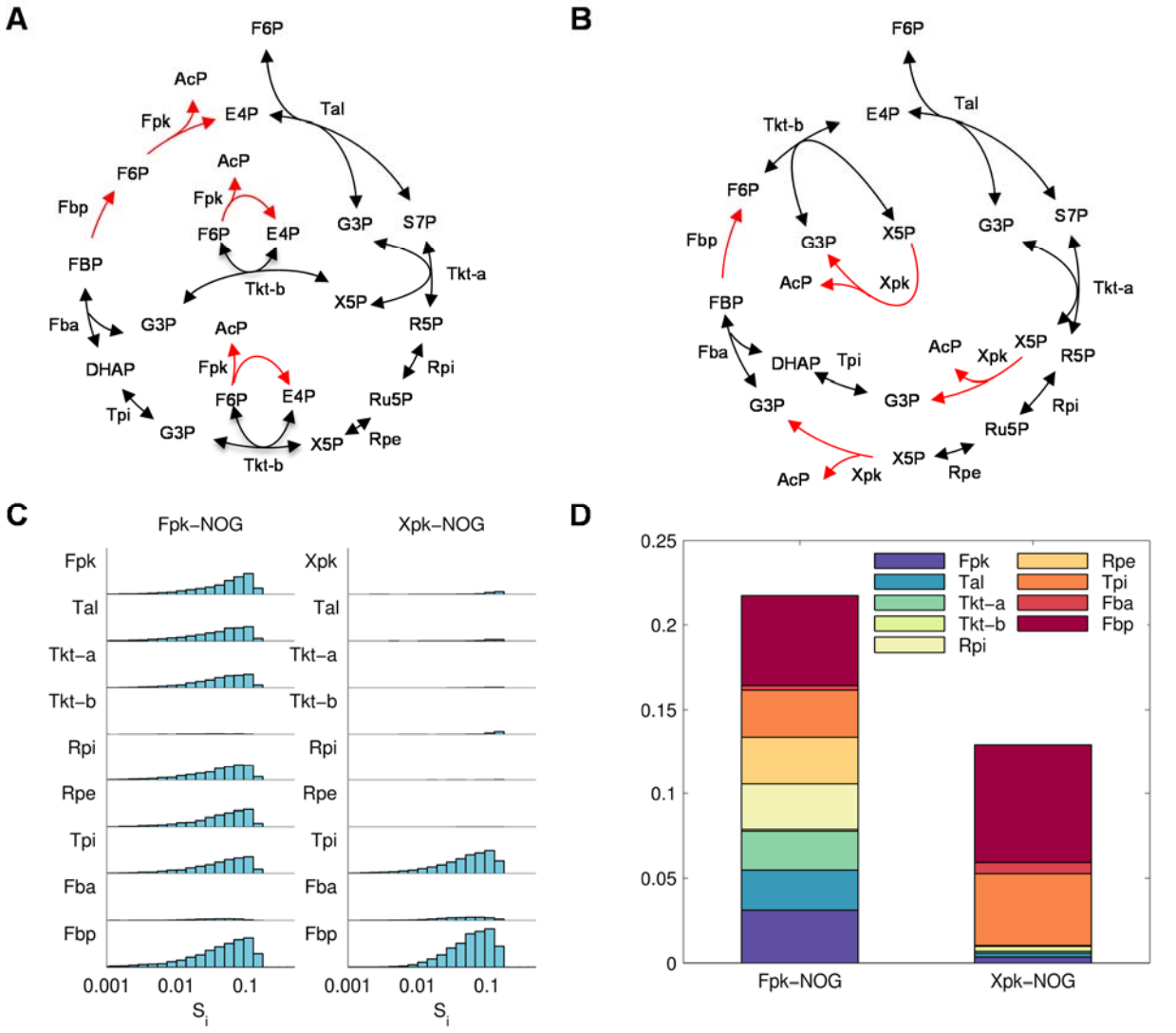

Figure 3-6. Robustness analysis of two configurations of a non-oxidative glycolysis (NOG) using *S*. **(A-B)** Two configurations of NOG where black double arrows denote reversible reactions and red arrows denote irreversible reactions. **(A)** In Fpk-NOG, phosphoketolase exhibits enzymatic activity only towards F6P and is denoted Fpk **(B)** In Xpk-NOG, phosphoketolase exhibits enzymatic activity only towards X5P and is denoted Xpk **(C)** Histogram of Si throughout the ensemble for every enzyme in each configuration. **(D)** The average *S* for each configuration is calculated as the sum of the average S_i of every enzyme.

To identify the key determinant of robustness in van Heerden's model, we constructed 16 alternative models by reversing all possible combinations of the first four major changes and calculated their respective *S* (Figure 7). As expected, the case where all four major changes were reversed (Figure 7, blue bar) is lost the robustness and exhibits a 43-fold increase in *S* when compared to van Heerden's model (Figure 7, orange bar). Interestingly, we found that high V_{PDC}^{m} reduced S (increased robustness) 25- to 100-fold when compared to models with a lower V_{PDC}^{m}. Thus, the activity of PDC appears to be the most important factor in the robustness of the model. In addition, the inhibition of HK by glucose 6-phosphate also contributes to the reduction of S and thus increase in robustness.

The finding that a high V_{PDC}^{m} is critical for the overall robustness of yeast glycolysis is consistent with several observations. First, a pyruvate-decarboxylase-negative (Pdc-) *Saccharomyces cerevisiae* mutant lacking all three PDC genes (PDC1, PDC5 and PDC6) not only exhibited a three-fold lower growth rate in rich medium containing glucose than the isogenic wild-type strain, but was also unable to grow in minimal medium with glucose as the sole carbon source (70). Second, the PDC6 gene, whose expression is either very low or absent in wild-type *S. cerevisiae*, was highly induced in the presence of excess sugars (71), suggesting that extra Pdc activity is beneficial for growth under high-sugar stress conditions. In these tests, the changes to trehalose and glycogen production kinetics were not reverted because the alternative systems would seldom reach a default steady state. These results demonstrate the utility of *S* in identifying key features that are essential for the global robustness of a metabolic system.

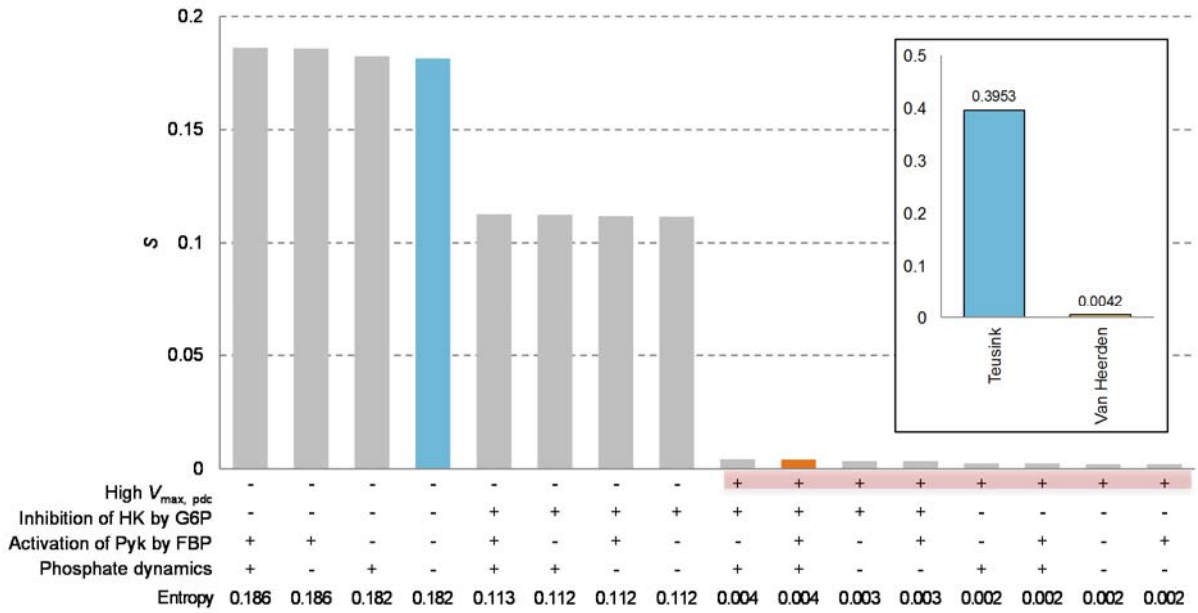

Figure 3-7. *S* helps identify pyruvate decarboxylase activity as a key parameter for the bifurcational robustness of yeast glycolysis. The inset shows the large difference in *S* between Teusink *et al.*'s [18] (blue bar) and van Heerden *et al.*'s (14) glycolysis models. To identify which of the first four major changes incorporated in van Heerden *et al.*'s model accounts for the robustness improvement, we constructed 16 alternative models by reversing every combination of the four changes and calculated each model's *S*. The orange bar corresponds to the original model, whereas the blue bar corresponds to the extreme case where all four changes were removed.

3.5 Conclusions

Robustness is an inherent property of biological systems to maintain desired function when faced with perturbations in environmental or physiological conditions. However, the intrinsic nonlinearity of metabolic systems suggests that a system can suddenly lose a stable steady state in the presence of random perturbations if a bifurcation point is crossed. Therefore, a bifurcationally robust system is needed to tolerate large changes in gene expression levels or kinetic parameters without crossing a bifurcation point. It is necessary to note that the robustness against bifurcation is different from local sensitivity (45, 72-76), which concerns the quantitative change of system properties against small perturbations, but is equally important.

Here we develop an index for bifurcational robustness, denoted as S, and show that it negatively correlates with a system's robustness against bifurcation in the presence of random parameter (or enzyme) changes. Interestingly, the definition of S is mathematically similar to the entropy in thermodynamics and information theory (69). As a result, S is also an extensive property as is the thermodynamic or information theoretic entropy. That is, the robustness of a metabolic system against random enzyme changes (S) is a sum of the robustness with respect to random changes in individual enzymes (S_i). This additive property gives us a tool for identifying the possible obstacles to the in vivo observability of a non-native pathway. However, other entropy properties in thermodynamics and information theory do not readily apply.

We demonstrated the utility of S using three examples. First, we show that S enables the optimization of bifurcational robustness of a yeast glycolysis model (12). Given that experimental data used for fitting is generally sparse, adding S as an additional optimization objective can further reduce the uncertainty of parameter values that are otherwise loosely constrained (77, 78). Indeed, the parameter changes returned by the optimization not only improve the bifurcational robustness dramatically, but also correspond with the upgrades made to the model in a subsequent modelling effort. This first demonstration sheds light on the important role that parameter optimization considering robustness can play in building better metabolic models.

The calculation of S calls for a model with all parameters specified or fitted, which is not always possible. For example, non-native pathway design normally starts with a list of enzymatic reactions that constitute the pathway, but the key parameters (e.g. V_{max}'s) are strain and condition dependent and usually unknown. To address the uncertainty of kinetic parameters, we show that S can be readily integrated with Ensemble Modeling for Robustness Analysis (64) to enable the design of robust non-native pathways. Depending on the goal of a metabolic pathway designer, one can

either investigate the distributions of S_i to identify potential sources of pathway failure, or simply use S to differentiate robust designs from non-robust designs (Fig. 6D). Given the versatility of S, we expect its combination with EMRA will bring unique value to the design of viable non-native pathways.

Finally, we show that S enables the comparison of a series of related models based on their bifurcational robustness. Since each model incorporates distinct features, a comparison of these models should elucidate the feature(s) that improve robustness the most. Indeed, by comparing the 16 different models of yeast glycolysis (14), we found that a high pyruvate decarboxylase activity is necessary for the robustness of this pathway. This interesting finding not only is supported by physiological data (70, 71), but it may also offer an indirect explanation to the large number of PDC genes in the *S. cerevisiae* genome (70). Even though the models we used consider only enzyme kinetics and regulations at the kinetic level, other regulatory mechanisms, such as transcriptional and post-transcriptional regulations, can also be included in the model, and the computation can proceed identically.

3.6 Methods

3.6.1 Models of metabolic systems with known parameters

To investigate the bifurcational robustness of metabolic systems, we examined thirteen ordinary differential equation (ODE)-based kinetic models of various metabolic systems from the BioModels Database (79). We selected these models because: (i) they are based on rate expressions exhibiting saturation kinetics, which is essential for discussing large perturbations; (ii) the model descriptions are sufficient for simulation; and (iii) each model reaches a non-trivial steady state (also called the default steady state) when simulating with the default parameters. Lumped mass

action and power-law models were excluded because they represent local behavior and do not exhibit saturation kinetics in large perturbations.

3.6.2 Bifurcation detection with the continuation method

Here we use a continuation method (49, 65, 66) as a computationally cheap and scalable alternative to study the steady-state response to parameter perturbations. In general, this method aims to find a connected path of steady-state solutions ($\mathbf{x}_{SS}$) as follows:

$$\frac{d\mathbf{x}}{\mathrm{dt}} = \mathbf{F}(\boldsymbol{x}_{SS}, \mathbf{p}) = \mathbf{0} \qquad \text{Eq. 4}$$

Since $\mathbf{F}(\mathbf{x}_{SS},\mathbf{p})$ is always equal to zero, it follows that the total derivative of $F(\mathbf{x}_{SS},\mathbf{p})$ with respect to p is also zero:

$$\frac{d\mathbf{F}(\boldsymbol{x}_{SS}, \mathbf{p})}{\mathrm{d}\mathbf{p}} = \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}_{SS}} \frac{d\boldsymbol{x}_{SS}}{d\boldsymbol{p}} + \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{p}} = \mathbf{0} \qquad \text{Eq. 5}$$

Rearranging Eq. 5 then yields Eq. 6:

$$\frac{d\boldsymbol{x}_{SS}}{d\boldsymbol{p}} = -\left(\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}_{SS}}\right)^{-1} \frac{\partial \boldsymbol{F}}{\partial \boldsymbol{p}} = \mathbf{0} \qquad \text{Eq. 6}$$

which defines the derivatives of steady-state concentrations with respect to kinetic parameters and sets the ground for parameter continuation. Starting from the set of parameters that characterize the reference steady state, the integration of Eq. 6 can proceed in the direction where a specific parameter (*e.g.*, V_{max}) is increased or decreased (Fig. 2A). The corresponding solution, which traces a trajectory in the $\mathbf{x}_{SS}$-$\mathbf{p}$ space, will then characterize how the steady state changes according to the parameter of interest. It should be noted that Eq. 6 is mathematically equivalent to the steady-state first-order sensitivity equations. Therefore, as the algorithm (*i.e.*, the differential equation

solver) proceeds in the parametric domain, the (local) sensitivity profile of metabolite concentrations with respect to parameters will be updated simultaneously as the steady state moves along the $\mathbf{x}_{SS}$-$\mathbf{p}$ trajectory.

Given that Eq. 6 is ill-defined when the Jacobian matrix ($\partial\mathbf{F}/\partial\mathbf{x}_{SS}$) is not invertible, it is important to detect the point where the Jacobian matrix becomes singular. This boundary point is also known as the bifurcation point (65, 66). Therefore, the bifurcation point defines the parameter range where a system is functional with a stable steady state. In practice, the Jacobian almost always becomes badly conditioned when the system is near a bifurcation point. When this happens, we stop the solver and declare such an edge case a bifurcation point. Additionally, due to the nature of numerical integration, it is possible to "jump" over the exact point of singularity. To account for this, we always check if any of the negative eigenvalues of the Jacobian matrix becomes positive in case a bifurcation point has escaped the detection.

3.6.3 Calculation of S

Unless otherwise specified, parameters were assumed to vary according to a log-normal distribution with the median equal to the parameter's value at the reference steady state and a standard deviation of $\log(X/X_{original})$ equal to .5. Such distribution has a standard deviation of .36, which is conservative compared to the variations reported in Newman et al. (62) where the average ratio of standard deviation in protein level to protein level was found to be around 1.

3.6.4 Parameter optimization utilizing *S*

Given that Teusink *et al.* model of yeast glycolysis (12) has a known robustness issue (Supplementary Information), we seek to modify the parameters so that the tuned model can satisfy

experimental data and exhibit high robustness. To accomplish this, maximal rate constants ($\boldsymbol{V}^{m}$) were subject to optimization with the following objective function:

$$\min_{\boldsymbol{V}^{m}} \left(\sum_{i} \frac{\left| V_i - V_i^{org} \right|}{V_i^{org}} + \sum_{j} \frac{\left| X_j - X_j^{org} \right|}{X_j^{org}} + 100 \cdot S \right)$$

Here, V_i represents the steady-state flux of reaction i and X_j represents the concentration of metabolite j. The superscript *org* refers to the particular values in the Teusink *et al.* model. In this study, a weighting of 100 was used to over-emphasize S. Optimization was performed using a simulated annealing algorithm in MATLAB (The MathWorks, Inc., USA). We allowed parameters to vary within 10-fold of their original values, and reported the results with the lowest objective value among 20 independent runs.

3.6.5 Calculation of *S* for models with unknown parameters

To demonstrate the utility of S in quantifying the robustness of non-native pathways, we constructed an ensemble of 10,000 models for each of the two configurations of a non-oxidative glycolysis (NOG) pathway as described previously (64). Given that this is a heterologously expressed pathway, expression levels are assumed to have much higher uncertainty. Therefore, the probabilities of steady state retention under single-enzyme perturbation (p_i) were calculated assuming a log-uniform distribution between 0.1-fold and 10-fold of the default activity.

3.6.6 Reverting Changes to van Heerden model

The yeast glycolysis model by van Heerden *et al.* (14) is derived from the model by Teusink *et al.*(12) with five major modifications. (These have been mentioned in the text)

1. Inhibition of hexokinase (HK) by glucose 6-phosphate (G6P)
2. Consideration of phosphate as a free variable

3. Activation of pyruvate kinase (PYK) by fructose 1,6-bisphosphate (FBP)
4. A 6.1 fold increase in the V_{max} of pyruvate decarboxylase (PDC)
5. Trehalose and glycogen fluxes became functions of G6P

To study the degree by which each of these modifications improves the bifurcational robustness, we created new models by reverting the main traits of these changes. In general, if the modifications required changes to the rate laws (such as the first three modifications), we adjusted the corresponding maximal rate so as to maintain the original steady-state flux. However, such adjustments are not always possible. The fourth modification, for example, required an increase of the maximal rate of Pdc, meaning that reverting this modification will lead to a change in the steady state, in particular the pyruvate concentration.

In this study, only the first four modifications were reverted. We did not revert the fifth modification because it causes very erratic behavior where many models don't even reach a default steady state. For the four modifications considered, we reverted all possible combinations of these changes.

3.7 Abbreviations

Abbreviation	*Enzyme Name/Function*
Glk	Gluco kinase
Pgi	Phosphogluco isomerase
Glyc	Glycogen production
Treha	Trehalose production
Pfk	Phosphofructokinase
Aldo	Fructose-1,6-bisphosphate aldolase
Gapdh	D-glyceraldehyde-3-phosphate dehydrogenase
Pgk	Phosphoglycerate kinase
Pgm	Phophoglycerate mutase
Eno	Phosphopyruvate hydratase
Pyk	Pyruvate kinase
Pdc	Pyruvate decarboxylase
Succ	Succinate production
Glt	Glucose transport
Adh	Alcohol dehydrogenase
G3pdh	Glycerol 3-phosphate dehydrogenase
ATPase	Adenosine triphosphatase

Table 3-1. Abbreviations used in Figure 1C and Figure 4B.

Abbreviation	*Metabolite Name*
NADH	Reduced nicotinamide adenine dinucleotide
P	High energy phosphate pool (2[*ATP*] + [*ADP*])
ACE	Acetaldehyde
PYR	Pyruvate
PEP	Phosphoenolpyruvate
P2G	2-phospho-D-glycerate
P3G	3-phospho-D-glycerate
BPG	1,3-biphospho-D-glycerate
TRIO	Dihydroxyacetone phosphate + glyceraldehyde-3-phosphate
F16P	Fructose-1-6-biphosphate
F6P	Fructose-6-phosphate
G6P	Glucose-6-phosphate
GLCi	Internal glucose

Table 3-2. Abbreviations used in Figure 4B.

4 Stability as a criterion for cell-wide model building and identifying the kinetically attainable flux

4.1 Introduction

As societies seek alternatives to traditional petrochemical synthesis, microbial synthesis of chemicals from renewable feedstocks has become increasingly important (15-18, 30, 32, 80). However, the development and optimization of suitably efficient strains is a lengthy and costly endeavor. Thus, methods for speeding this development process will prove helpful for the success of metabolic engineering (81, 82).

To better capture the effects of kinetic-altering manipulations, a model incorporating kinetic information is necessary (10, 81, 83). Kinetic information is generally presented as rate equations which relate enzymatic flux to substrate concentrations and enzyme parameters. Although countless kinetic models have been generated for many particular applications (1, 3, 12, 20, 73, 84-86), this process still remains tedious and the scale of models is generally limited to a small set of pathways depending on the application. This limitation comes from the difficulty in determining enzyme parameters (45, 46, 87), especially V_{max} which depends on enzyme concentration and is thus very dependent on the growth stage and nutrients available to the microbe. The inconvenience of obtaining enzyme parameters for *in vivo* conditions represents a profound challenge for kinetic metabolic simulation.

Ensemble modeling (EM) (21-25, 64, 88)is a method developed to circumvent the need for *a priori* knowledge of enzyme parameters while still considering kinetic relationships. Using stoichiometry, enzymatic rate equations, reference steady state fluxes and stability analysis, EM makes enhanced use of known network information to generate many feasible parameter sets (i.e. an ensemble) (22). By using a collection of randomly assigned yet feasible models, EM is able

generate deep biological insight (e.g. identification of potential over-expression candidates) without the need for specialized experiments for the development of a kinetic model.

The analysis of a high number of parameter sets, however, comes at a cost. In order to determine the effect of a change in kinetic parameters it is necessary to integrate hundreds or thousands of systems of ODEs through time. Given the stiffness of large biochemical reaction models (89, 90), especially those with randomly assigned parameters, such computation may also become infeasible as the size of the model increases. Recently, parameter continuation has been applied to the EM algorithm (64, 91) greatly increasing the computational efficiency of the analysis. This new application in EM not only allows us to greatly increase the size of the model, but can also be used in the validation of the constructed model (91).

Here, we present an ensemble model of E. coli consisting of 193 reactions which encompass glycolysis, the TCA cycle, the synthesis of all 20 canonical amino acids as well as nucleotides. Starting from a computer generated model using stoichiometric and regulatory data from the Ecocyc database (92-94), the model was analyzed using entropy as a measure of bifurcational robustness. Using robustness analysis (64, 91) and on the premise that adequate models of cellular metabolism should be sufficiently robust (38-41, 54, 61, 95), essential regulatory mechanisms that were initially missing or wrongly incorporated in the model were identified and corrected. The revised model was then used to predict *in vivo* isobutanol production using a special application of the continuation method. The high accuracy of predicted isobutanol production from minimal data is not only evidence of the value of the prediction method but more importantly highlights the power of considering bifurcational robustness as a model building tool.

4.2 Results

4.2.1 Generation of an E. coli model

A model of *E. coli* under micro-aerobic growth condition was constructed to study isobutanol production. Given the large overlap between the isobutanol production pathway and the valine synthesis pathway, a cell-wide model incorporating amino acid biosynthesis was chosen for this study. Given the importance of cofactors NADH and NADPH in the production of isobutanol, nucleotide biosynthesis pathways were also included in the model. An automated script (See Methods) was used to construct the network representation by extracting reaction information from the EcoCyc database (92-94) for every pathway considered (Figure 1a). The extracted information included the reaction stoichiometry, reversibility and significant biochemical regulators for each reaction. Additionally, outlet reactions for each of the biomass components within the scope of the model were included. Each output was represented as the independent secretion of a particular metabolite with the stoichiometry corresponding to the millimoles of the metabolite per gram dry weight of biomass as reported in Feist et al. (96). In total, the model is composed of 193 reactions (Figure 1b). Up to this point, the scope of the model (deciding which pathways to include) is the only human input.

Having obtained an adequate model stoichiometry and under the assumption of specific rate laws depending on the number substrates and products of the reaction (See Methods) we only need a reference steady state to generate an ensemble of enzyme kinetic parameter sets representing our model organism. Given that we are studying the response of wild type E. coli to the isobutanol production pathway under micro-aerobic conditions, we chose to use production and growth data (Table 1) from Fischer et al. (97) for the determination of the reference state flux distribution. Under these constraints and to be consistent with the assumption of low oxygen, linear programing

was used to determine the reference state flux distribution by minimizing the flux going through citrate synthase. The optimized fluxes, although not necessarily accurate, provide a starting point for model building. One can also use multiple reference states representing variations in each of the key flux splitting points, and build multiple models to examine effects of the split ratios.

	Specific Rate
Glucose Uptake	7.1 $mmol*g^{-1}*h^{-1}$
Acetate Secretion	6.6 $mmol*g^{-1}*h^{-1}$
Growth Rate	0.52 $g*g^{-1}*h^{-1}$

Table 4-1. Growth parameters from Fischer et al. The external fluxes reported in Fischer et al. were used to approximate the internal fluxes of the model.

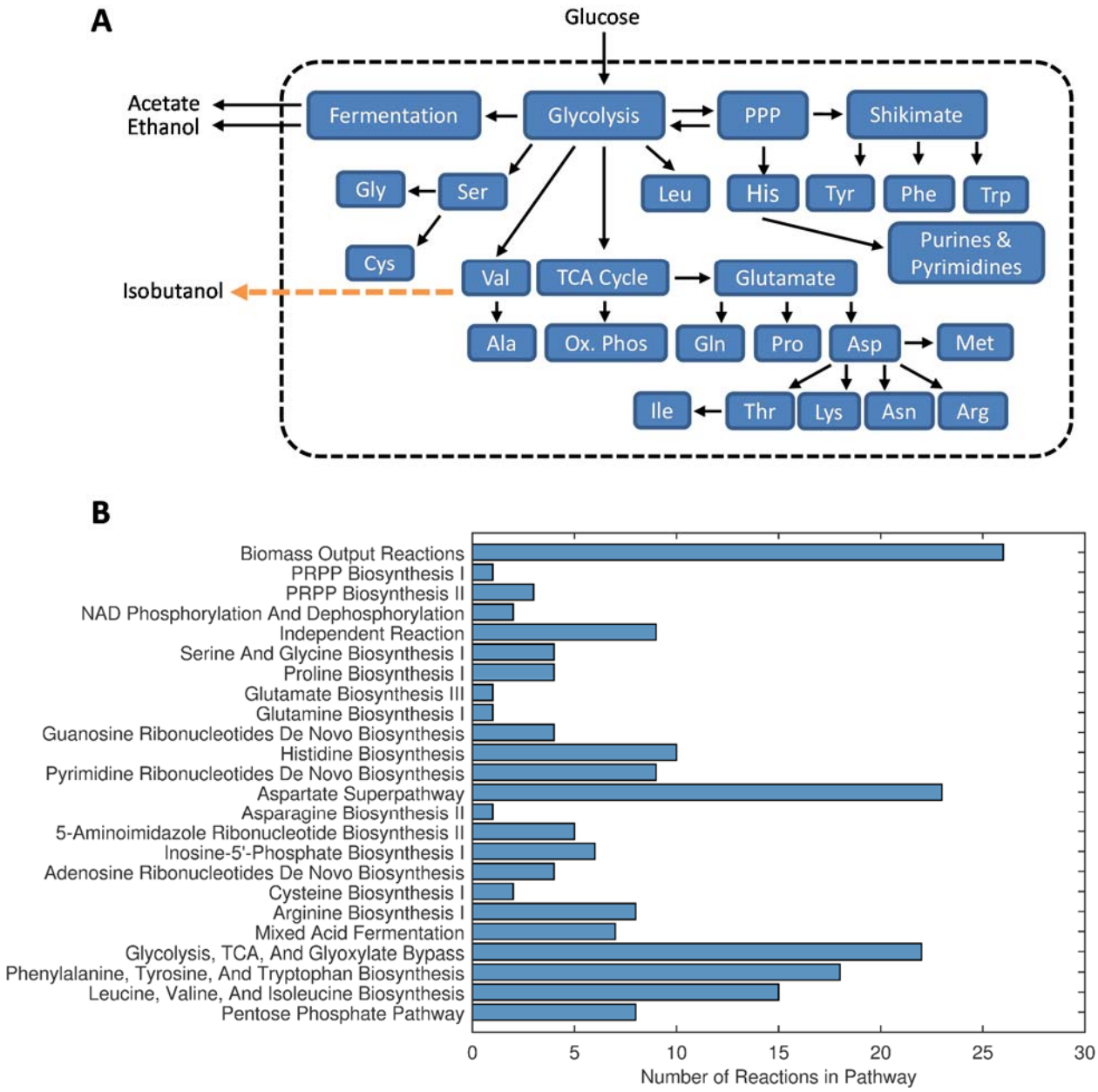

Figure 4-1. Scope of the model. (A) Overview of the pathways incorporated into the model. Biosynthesis pathways for the 20 natural amino acids, the nucleotides and the fermentation pathways were included. **(B)** Overview of the reactions in the model. A total of 193 reactions encompassing the aforementioned pathways were included in the model.

4.2.2 Entropy as an index to improve instability

The ability to maintain stability despite random changes in internal or external conditions is a necessary characteristic for survival (38-41). As such, models of the metabolic system need to reflect similar robustness against instability. Biological systems have evolved complex regulatory mechanisms to help maintain this stability (53, 98), including regulation at the enzyme kinetic level (5). One of the advantages of using kinetic models is their ability to incorporate these regulatory mechanisms to better represent the system. On the other hand, examining robustness against instability may help to identify either possible regulatory mechanisms that are omitted in the model, or mistakes in the model-building process.

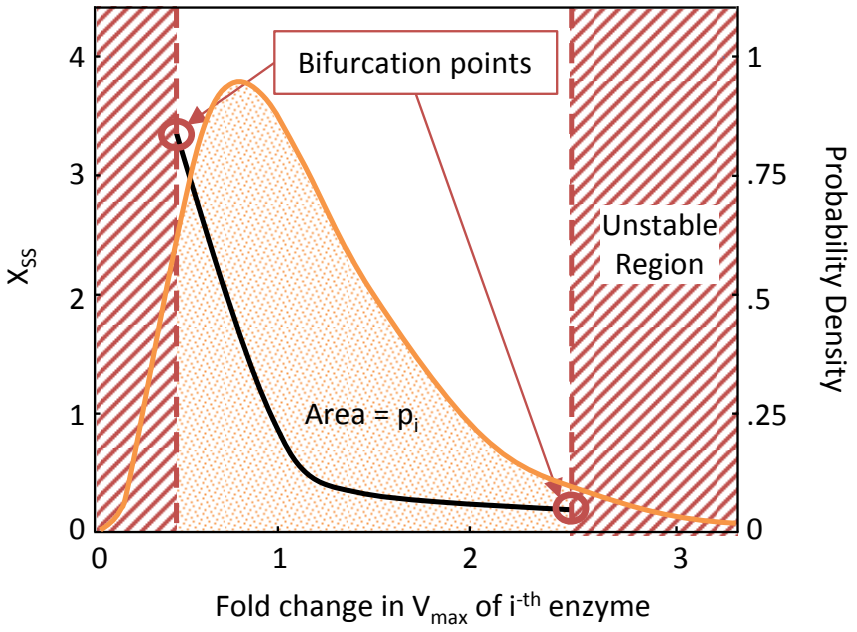

Figure 4-2. Calculating the probability of the system remaining stable after a random perturbation to enzyme *i*. Using parameter continuation on $V_{max,i}$ we can trace the steady-state solutions (black line) of the system as Vmax,I deviates from it reference value (1) to find the bifurcation points (red circles). Assuming the enzyme naturally varies according to a log-normal distribution (orange curve), we can calculate the probability of remaining stable as the area under the curve between the bifurcation points (orange dotted area).

In an attempt to pinpoint possible flaws in the model, the generated model was subject to robustness analysis as described in previous publications (64, 91). Stochastic variations in enzyme levels can be represented as a probability density function of the fold change in Vmax. Additionally, by perturbing a single enzyme, it is possible to determine the smallest and largest value of Vmax through which the system will remain stable as the bifurcation points of the corresponding parameter. Under such a model of variation and with knowledge of the bifurcation points for the Vmax of enzyme i, the probability (p_i) that the system will remain stable following a random perturbation on the level of enzyme i alone is given by the area bounded by the probability distribution function between the bifurcation points (Fig. 2). The "entropy" (S_i) contributed to the model by enzyme i is defined as (91):

$$S_i = -p_i \log(p_i) \tag{1}$$

Thus, entropy roughly correlates with the probability that a system becomes unstable when changing the activity of that particular enzyme (91). The total entropy (S) of the system is then given by:

$$S = \sum_i S_i \tag{2}$$

Given that entropy can both give a quantitative index of bifurcational robustness for the model as a whole and also discriminate between enzymes as sources of instability, it was the index of choice for aiding in the improvement of the model.

Figure 3A shows the distribution of entropy among enzymes in the model. As expected, most enzymes have very small entropies (<0.01) and the number of enzymes decreases as the entropy increases. This result indicates that only a few enzymes contribute to the lack of robustness of the whole model. Given the complex relationship between enzymes, a particular misrepresentation in one enzyme may cause robustness issues in other enzymes. With this in mind, we focused the investigation on enzymes with higher entropy values (>0.08) for possible causes of instability. Additionally, enzymes were grouped by their corresponding pathway (Figure 3B). Under such grouping, 4 pathways which are prone to instability were found: (a) the pentose phosphate pathway, (b) the homoserine biosynthesis pathway, (c) adenosine ribonucleotide biosynthesis pathways and (d) glycolysis and the tricarboxylic acid cycle. Upon investigation of these reactions and pathways, three possible flaws were identified.

First, the specification of homoserine as an activator of aspartate kinase in the EcoCyc database may be questionable. Homoserine is one of the main products of the pathway and having it be an activator of the first irreversible step of the pathway might pose a strong robustness liability. Such robustness liability comes from the inevitable accumulation of product, when the rate of the reaction increases as its product becomes more abundant. Literature reports (99, 100) indicate the inhibition of aspartate kinase by lysine and threonine; these are included in the model. Nevertheless, to our knowledge, no literature reports homoserine as an activator. Therefore, the rate law of aspartate kinase must be corrected to not include activation by homoserine.

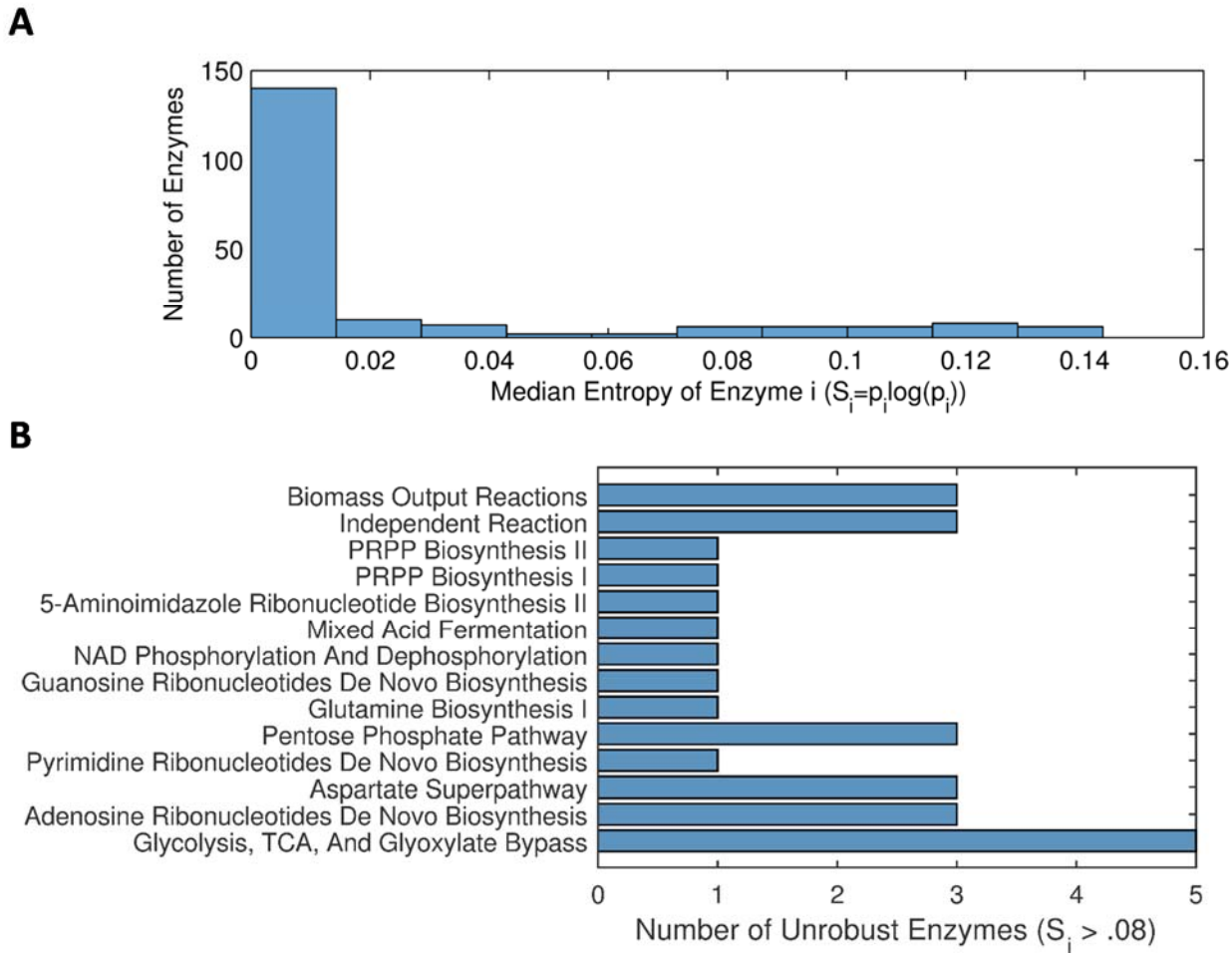

Figure 4-3. Robustness analysis of the automated model. (A) Distribution of entropy among enzymes in the model. Most enzymes have very low values of entropy and the number of models decreases as the entropy increases. At an entropy value of around .08, there is a resurgence of models. **(B)** Number of unrobust ($S_i > .08$) enzymes in each pathway. The pentose-phosphate pathway, the aspartate biosynthesis pathway, adenosine ribonucleotide biosynthesis pathway and the glycolysis/TCA stand out as the most unrobust pathways in the model.

Second, the glucose transport and glycolysis may be a source of instability. Glycolysis is a very well-studied and fairly straightforward pathway. Upon finding no apparent flaws in the regulation of glycolytic enzymes, we turned to the phosphotransferase system (PTS) as possible source of instability. In our original model formulation, a summarized reaction for the PTS was entered manually and was not assigned any regulatory mechanisms. Additionally, the original rate law

used for PTS exhibited Michaelis-Menten type kinetics for the overall reaction. Such naïve implementation may present a problem because the PTS has been recognized to be regulated by the PEP/pyruvate ratio (3, 101, 102). Additionally, this system is known to be inhibited by glucose-6-phosphate (3, 103, 104). With this in mind, the model was modified to use the following rate law for PTS:

$$V_{PTS} = \frac{V_{m,PTS}(X_{PEP}/X_{PYR})X_{Glc}}{\left(k_{m,1} + k_{m,2}(X_{PEP}/X_{PYR}) + k_{m,3}X_{Glc} + X_{Glc}(X_{PEP}/X_{PYR})\right)\left(1 + X_{G6P}/k_{i,G6P}\right)} \quad (3)$$

where variables X represent the concentrations of metabolites phosphoenolpyruvate (PEP), pyruvate (PYR), glucose (Glc) and glucose-6-phosphate (G6P), $V_{m,PTS}$ is the maximal rate of the reaction and parameters k represent the affinity coefficients. This rate equation can still be normalized with respect to metabolite concentrations and parameter sampling can be performed as with the rest of the model.

Third, the pentose phosphate pathway may contain some error. Upon inspection we observed that the model had no regulation associated with 6-phosphogluconate dehydrogenase. As one of the irreversible steps, the second NADPH forming and the only carbon releasing step in the pentose phosphate pathway, 6-phosphogluconate dehydrogenase should be well regulated. Indeed, this enzyme is a well-known regulatory point for the pentose phosphate pathway (105-108) with its activity significantly inhibited by ATP (3, 108), NADPH (3, 107) and fructose-1-6-diphosphate (107, 108). Given the complex role of ATP in the glycolytic process, the regulation of 6-phosphogluconate by ATP may be of particular importance for model stability. In order to fix this,

the model was modified to include the inhibition of 6-phosphogluconate dehydrogenase by ATP, NADPH and fructose-1-6-diphosphate.

No apparent flaws were found in the adenosine ribonucleotides biosynthesis pathway. The relatively high entropy values might suggest other regulatory mechanisms (eg. transcriptional regulation) as responsible for the stability of this pathway. This would not be unexpected as ATP is a molecule of extreme importance to the cell and might be an interesting point for subsequent research.

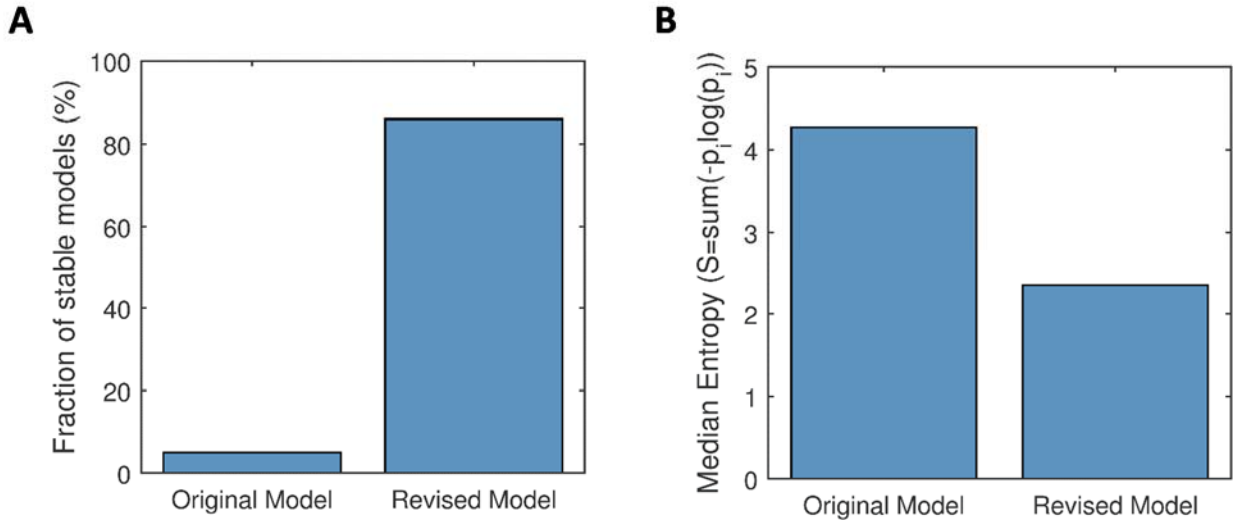

Figure 4-4. The revised model is more stable and robust. (A) Upon revision, the fraction of stable models at the reference is greatly increased. **(B)** The entropy of the model is significantly reduced upon revision of the three identified flaws.

By applying the aforementioned corrections to the three identified flaws, a revised model was produced. This model contains the same reactions and reference state flux distribution as the original model but has changes in the regulation of aspartate kinase, 6-phosphogluconate dehydrogenase and PTS as well as a change in the functional form of PTS. Figure 4A shows the

dramatic difference in the percentage of sampled models which are stable between the revised model and the original model.

Although parameters are sampled constrained to a steady state, the probability of obtaining a parameter set for which that steady state was stable was very low without a few essential regulatory mechanisms (Figure 4A). Only about 5% of the parameter sets sampled yield a stable steady state. Others generate an unstable system that cannot be physically realizable, despite the existence of the fixed point (steady state). After revising the model, the stability of the structure improved significantly (Figure 4A), as indicated by the high percentage of randomly sampled parameter sets that generate a stable system. Although it is easy to sample around the instability problem, such an approach would completely miss inadequacies in the model.

Similarly, the median entropy was also greatly reduced (Figure 4B). Although at first sight the impact to entropy does not look as dramatic as the change in the fraction of stable models, it is important to recall that entropy is a logarithmic measure of probability and a unit change in entropy corresponds to an order of magnitude in probability. Although highly related, the fraction of stable models and the median entropy are very distinct parameters which can vary independently. As such, it is important to consider both during model development.

4.2.3 Determining kinetically attainable flux

E. coli does not naturally produce isobutanol and as such there are no reactions in our reference model capable of producing it. Given the kinetic structure and the reference state (e.g. the wild-type strain) of the host metabolic system, it would be interesting to determine the maximum flux that can be drawn from the cell upon introducing a non-native pathway.

Starting from the model definition at the reference steady state:

$$\frac{dX}{dt} = F(X,k) = Sv(X,k) \tag{3}$$

for metabolite concentrations *X*, kinetic parameters k and stoichiometry S, we wish to incorporate information for a new pathway for which kinetics are unknown. Although the new pathway may introduce new metabolites in the system, we consider how the new pathway flux might affect the kinetics of the host metabolic system. As such, we can represent the new model as:

$$\frac{dX}{dt} = F(X,k,\varphi) = Sv(X,k) + S_{new}\varphi \tag{4}$$

where S_{new} represents the stoichiometry of the new pathway and φ represents the flux going through the pathway. By its definition, parameter φ is 0 at the reference state and the sampling of parameters k is unaffected by the new formulation.

Using this new formulation, we are now able to determine the effect of the addition of a nonnative pathway to the model by performing parameter continuation along parameter φ. Although without prior experimental efforts we do not know what value of φ to integrate to, there is one point that always carries significance, the bifurcation point. In essence, the bifurcation point of this flux represents the maximum amount of flux that the system can hold before losing stability–in other words, the kinetically attainable flux. Figure 5A shows the result of integrating through flux using φ (Figure 5A): at the bifurcation point, the system suddenly loses stability. This bifurcation point corresponds to the maximum flux that can be drawn out of the host kinetic system. Assuming the kinetic parameters for the same reaction are known, integration through V_{max} could have two different behaviors depending on the system. In the first case, the system is not robust against increases in Vmax (Figure 5B). In this case, the bifurcation point of φ corresponds to the flux going through the enzyme at the bifurcation point (Figure 5 A vs B). In the second case, the system

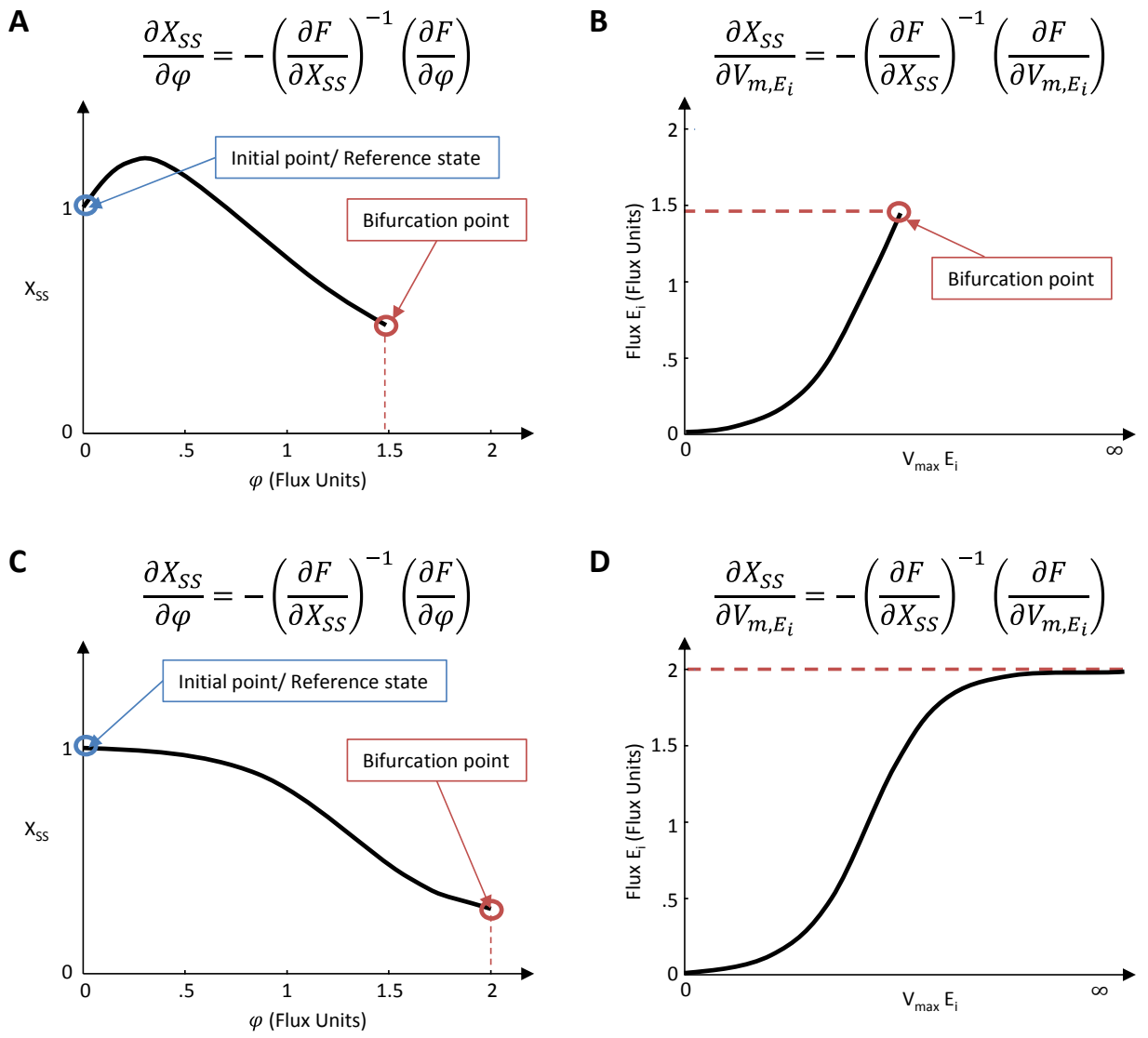

Figure 4-5. Understanding the kinetically attainable flux. (A) Like integrating through kinetic parameters, integration through flux (φ) yields a trajectory in the x-φ space. **(B)** If the same reaction but with known kinetics is incorporated into the model, the flux at the bifurcation point of the integration through $V_{max,i}$ will correspond the bifurcation point of φ. **(C)** Integration though flux for a different reaction yields a different trajectory in the x-φ space. **(D)** If the same reaction but with known kinetics is incorporated and the system is robust to infinite levels of enzyme *i*, the bifurcation point in (C) will correspond to the asymptote of flux as $V_{max,i}$ goes to infinity.

is fully robust against increases in V_{max} (Figure 5D). In this case, the bifurcation point of φ corresponds to the flux asymptote of the integration with known kinetics as V_{max} approaches infinity (Figure 5 C vs D). In other words, even if the integration using kinetics never bifurcates, the bifurcation point of φ still corresponds to its maximum flux. If an enzyme is heterologously expressed with a strong promoter in a high copy plasmid, its rate is expected to be limited by the underlying host metabolism. An enzyme under such conditions is expected to operate near the kinetically attainable flux. Under this assumption, an estimate of production can be calculated for each model in the ensemble starting from a wild type model.

4.2.4 Isobutanol production as an example

The isobutanol production from several E. coli strains with different genotypes was reported by Atsumi et al. (15). In their work, Atsumi et al. started from a wild-type E. coli strain and overexpressed isobutanol producing genes to achieve high isobutanol production under semi-aerobic conditions. Similarly, we can use the previously developed model for wild-type E. coli under semi-aerobic growth as a reference for calculating the kinetically attainable isobutanol flux for the various genotypes presented.

The first production strain reported contained the KivD (from *Lactococcus lactis*) and Adh2 genes which together are capable of generating isobutanol from the valine precursor 2-ketoisovalerate (2-KIV). The effect of this genotype was modeled by calculating the kinetically attainable flux from 2-KIV to isobutanol. The model of this genotype predicted 8% of the maximum theoretical yield from glucose while the experiment produced slightly more, 10% (Figure 6). The next genotype reported by Atsumi et al. included overexpression of the *ilvIHCD* genes to increase availability of 2-KIV by catalyzing its production from pyruvate. The effects of this overexpression were modeled by calculating the kinetically attainable flux directly from pyruvate

to isobutanol to mimic the effect of the heterologous over-expression of the entire valine-isobutanol pathway. Integration along this pathway flux resulted in an increased predicted yield of 26% vs. an experimentally determined yield 29% (Figure 6). The final genotype reported included a knockout of fermentative pathways. This was modeled by simulating a 90% reduction in the activity of alcohol dehydrogenase, lactate dehydrogenase and phosphotransacetylase using parameter continuation, and then starting from that state to calculate the kinetically attainable flux with respect to the entire valine-isobutanol pathway. Prediction resulted in an increased yield of 43% vs. 34% for the experimental production (Figure 6).

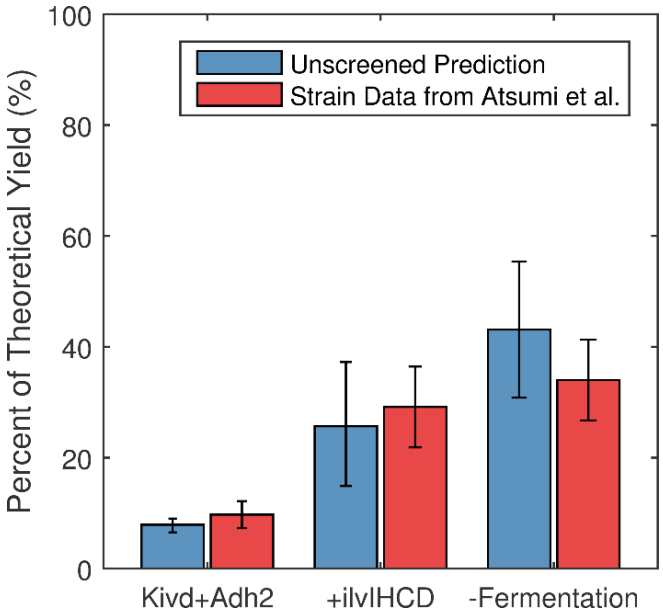

Figure 4-6. Isobutanol production in the three genotypes reported in Asumi et al. (15). In silico predictions (blue) closely mirros experimental results (red).

4.3 Discussion & Conclusions

Performing kinetic simulation on cellular metabolism to predict the outcomes of metabolic engineering remains an alluring goal. Ensemble Modeling has the dual benefits of not requiring a priori parameter values and also not resorting to simplified rate laws like mass-action kinetics.

Instead, Ensemble Modeling allows the use of realistic rate laws that exhibit both saturation phenomenon and kinetic regulation. This comes with the trade-off of requiring metabolic fluxes for a reference steady state in order to efficiently sample parameters. However, this requirement is perhaps less a handicap than it first appears. Although optimally metabolic fluxes would be determined through targeted experiments (109, 110), approximation using available data and stoichiometric methods (111, 112) can still yield realistic models. In this work, the reference flux distribution was constructed using data from the literature and predictions closely mirrored experimental results.

Having constructed such a large model using automated methods, it is important to validate the model before making predictions. Based on the premise that cellular metabolism must be robust against the stochastic nature of gene expression and this must be well-represented in metabolic models, the computer-generated model was tested using bifurcational robustness. Using entropy (91) to quantify bifurcational robustness, a small subset of enzymes were identified as the main sources of instability in the model. Upon manual investigation of these enzymes several mistakes, including erroneously added or missing regulation and inappropriate rate equations, were identified. Upon revising the model, there was a dramatic increase in the percentage of sampled models that were stable at the reference state. Similarly, the bifurcational robustness of the model was greatly increased. Overall, entropy showed to be a great tool for the identification of missing regulatory relationships and could perhaps be used for the unraveling of novel regulatory relationships not yet discovered.

Finally, the revised model was used for predicting *in vivo* isobutanol production through the calculation of the kinetically attainable flux. The kinetically attainable flux calculates the maximum flux that can be achieved through a reaction before the system loses stability. In

comparison to a reaction with known kinetics, this flux corresponds to either the flux at the bifurcation point as V_{max} increases from 0 or the asymptote of flux as V_{max} goes to infinity if there is no bifurcation point. Simulating the three isobutanol producing genotypes presented in Atsumi et al. (15) showed the adequacy of predicting using the kinetically attainable flux and the ability to create realistic models using Ensemble Modeling; curation using robustness analysis was crucial to achieving a realistic model. Both of these finding together showcase the use of Ensemble Modeling for two very distinct but closely related purposes: (a) identifying regulatory mechanisms essential to robustness and (b) for predicting the outcome of metabolic engineering.

4.4 Methods

4.4.1 Automated model generation

An automated script was generated in MATLAB (The MathWorks, Inc., USA) to parse Attribute-Value format files of the EcoCyc v17.1 database (93) and extract reaction information for a specified set of pathways (Table 2). For every reaction, the stoichiometry, reversibility and regulators were recorded whilst neglecting vitamins, ions, salts, inorganic compounds and tetrahydrofolate derivatives. Using the generated model stoichiometry, a reference state flux distribution was obtained by using linear programming constrained to the data in Table 2, constraining parallel reactions using biological insight (*cf.* Section 4.5) and minimizing the flux going through citrate synthase to represent micro-aerobic conditions.

EcoCyc IDs of the Pathways Included in the Model
1. GLYCOLYSIS-TCA-GLYOX-BYPASS
2. MALIC-NADP-RXN
3. PEPCARBOX-RXN
4. PENTOSE-P-PWY
5. GLUTSYNIII-PWY
6. GLNSYN-PWY
7. PWY0-781
8. ASPARAGINESYN-PWY
9. PROSYN-PWY
10. SER-GLYSYN-PWY
11. CYSTSYN-PWY
12. BRANCHED-CHAIN-AA-SYN-PWY
13. VALINE-PYRUVATE-AMINOTRANSFER-RXN
14. PWY0-661
15. PWY0-662
16. PWY0-162
17. HISTSYN-PWY
18. PWY-6123
19. PWY-6122
20. PWY-7221
21. PWY-7219
22. COMPLETE-ARO-PWY
23. ARGSYN-PWY
24. NADPHOS-DEPHOS-PWY
25. FERMENTATION-PWY
26. RXN-8899

Table 4-2. EcoCyc IDs of the pathways included in the model.

4.4.2 Rate equations

With the exception of the phosphotransferase system in the revised model (See Results), all rate expressions in the model were assigned strictly depending on the number of substrates, products, reversibility and regulation type. Michaelis-Menten type kinetics were used for all reactions having up to two substrates and two products with non-allosteric regulation (113). Reactions with one substrate and two products or vice versa were modeled with ordered mechanism with the order following alphabetical order. Reactions with two substrates and two products were modeled with

a random order rate equation. Reactions having more than three substrates or products or those exhibiting allosteric regulation were modeled using the modular rate law described by Liebermeister, Uhlendorf and Klipp (114). Outlet reactions were modeled using first order kinetics.

4.4.3 Calculation of S

Parameters were assumed to vary according to a log-normal distribution with the median equal to the parameter's value at the reference steady state and a standard deviation of $\log(X/X_{\text{original}})$ equal to 0.5. This log-normal distribution has a standard deviation of 0.36, which is within the rage reported Newman *et al.* (62) where the average ratio of standard deviation in protein level to protein level was found to be around 1.

4.5 Automatic model generation: Matlab Code

To further clarify the model building process, the following sections show the MATLAB code for the different model generation steps. Every assumption made is outlined in the code.

4.5.1 Define the pathways or reactions to be considered

```
Pathways = {
   'GLYCOLYSIS-TCA-GLYOX-BYPASS';
   'MALIC-NADP-RXN';
   'PEPCARBOX-RXN';
   'PENTOSE-P-PWY';
   'GLUTSYNIII-PWY';
   'GLNSYN-PWY';
   'PWY0-781';
   'ASPARAGINESYN-PWY';
   'PROSYN-PWY';
   'SER-GLYSYN-PWY';
   'CYSTSYN-PWY';
   'BRANCHED-CHAIN-AA-SYN-PWY';
   'VALINE-PYRUVATE-AMINOTRANSFER-RXN';
   'PWY0-661';
   'PWY0-662';
   'PWY0-162';
   'HISTSYN-PWY';
   'PWY-6123';
   'PWY-6122';
```

```
    'PWY-7221';
    'PWY-7219';
    'COMPLETE-ARO-PWY';
    'ARGSYN-PWY';
    'NADPHOS-DEPHOS-PWY';
    'FERMENTATION-PWY';
    'RXN-8899';
    };
```

4.5.2 Recursively obtain the list of reactions

```
Rxns = {};
PathwayNames = {};
Rxn2Pathway = [];
for n=1:length(Pathways),
    [PathwayNames{n}, SubRxns] = ParsePathway(Pathways{n});
    if ~isempty(SubRxns),
        Rxns = [Rxns; SubRxns];
        Rxn2Pathway = [Rxn2Pathway; n.*ones(length(SubRxns),1)];
    else
        Rxns = [Rxns; Pathways{n}];
        PathwayNames{n} = '';
        Rxn2Pathway = [Rxn2Pathway; n];
    end
end

[Rxns, ind,~] = unique(Rxns);
Rxn2Pathway = Rxn2Pathway(ind);
```

4.5.3 Obtain reaction information for every reaction

```
Reactions = repmat(ReactionInformation, length(Rxns),1);
Metabolites = {};
for  n=1:length(Rxns),
    [ Reactions(n).dG0, Reactions(n).Reversible, Reactions(n).Metabolites, Reactions(n).Stoichiometry, ...
        Reactions(n).EnzymeNames, Reactions(n).Regulators, Reactions(n).RegulationType,
Reactions(n).CommonName ] = ReactionInfo( Rxns{n} );
    Reactions(n).Pathway = PathwayNames{Rxn2Pathway(n)};
    Metabolites = [Metabolites; Reactions(n).Metabolites];
end
Metabolites = unique(Metabolites);
```

4.5.4 Manually add reactions

```
%Add PTS reaction
Reactions(end+1).Reversible = 0;
Reactions(end).Metabolites = {'PHOSPHO-ENOL-PYRUVATE';'GLC-6-P'; 'PYRUVATE'};
Reactions(end).Stoichiometry = [-1; 1; 1];
Reactions(end).EnzymeNames = {'PTS'};
Reactions(end).CommonName = 'PTS';
```

```
Rxns = [Rxns; 'PTS'];

%Add glk reaction
Reactions(end+1).Reversible = 0;
Reactions(end).Metabolites = {'ATP';'GLC-6-P'; 'ADP'};
Reactions(end).Stoichiometry = [-1; 1; 1];
Reactions(end).EnzymeNames = {'glk'};
Reactions(end).CommonName = 'glk';
Rxns = [Rxns; 'glk'];

%Add Ethanol transport reaction
Reactions(end+1).Reversible = 0;
Reactions(end).Metabolites = {'ETOH'};
Reactions(end).Stoichiometry = [-1];
Reactions(end).EnzymeNames = {'EtOH_out'};
Reactions(end).CommonName = 'Ethanol Output';
Rxns = [Rxns; 'EtOH_out'];

%Add Acetate transport reaction
Reactions(end+1).Reversible = 0;
Reactions(end).Metabolites = {'ACET'};
Reactions(end).Stoichiometry = [-1];
Reactions(end).EnzymeNames = {'Acetate_out'};
Reactions(end).CommonName = 'Acetate Output';
Rxns = [Rxns; 'Acetate_out'];

%Add Lactate transport reaction
Reactions(end+1).Reversible = 0;
Reactions(end).Metabolites = {'D-LACTATE'};
Reactions(end).Stoichiometry = [-1];
Reactions(end).EnzymeNames = {'Lactate_out'};
Reactions(end).CommonName = 'Lactate Output';
Rxns = [Rxns; 'Lactate_out'];

%Respiration
Reactions(end+1).Reversible = 0;
Reactions(end).Metabolites = {'NADH'; 'ADP'; 'NAD'; 'ATP'};
Reactions(end).Stoichiometry = [-1; -2.5; 1; 2.5];
Reactions(end).EnzymeNames = {'CollapsedRespiration'};
Reactions(end).CommonName = 'Summarized Respiration';
Rxns = [Rxns; 'CollapsedRespiration'];

%Independent Outlet For Biomass: From Feist et al. 2007
BiomassComponents = {{'GLT'}; {'GLN'}; {'L-ASPARTATE'}; {'ASN'}; {'PRO'}; {'SER'}; {'GLY'};
{'CYS'}; {'LEU'}; {'VAL'}; {'L-ALPHA-ALANINE'}; ...
   {'THR'}; {'ILE'}; {'HIS'}; {'MET'}; {'LYS'}; {'TYR'}; {'PHE'}; {'TRP'}; {'ARG'}; {'UTP'}; {'CTP'};
{'GTP'}; {'ATP'}; {'NAD'}; {'NADP'}};
Coefficients = [-0.263; -0.263;-0.241;-0.241;-0.221;-0.216;-0.612;-0.092;-0.451;-0.423;-0.514;-0.254;-
0.291;-0.095;-0.154;-0.343;-0.138;-0.185;-0.057;-0.296;-0.144;-0.134;-0.215;-0.174;-0.001831;-
0.000447;];
for n=1:length(BiomassComponents),
   Reactions(end+1).Reversible = 0;
   Reactions(end).Metabolites = {BiomassComponents{n}};
   Reactions(end).Stoichiometry = [Coefficients(n)];
   Reactions(end).EnzymeNames = {strcat('Biomass', num2str(n))};
   Reactions(end).CommonName = strjoin(BiomassComponents{n}, 'to Biomass');
```

```
   Reactions(end).Pathway = 'Biomass Output Reactions';
   Rxns = [Rxns; strcat('Biomass', num2str(n))];
end
```

4.5.5 Prepare input variables for Ensemble Modeling

```
S = [];
Sreg = [];
Reversibilities = [];
EnzymeNames = Rxns;
for n=1:length(Reactions),
   S = [S zeros(length(Metabolites),1)];
   Sreg = [Sreg zeros(length(Metabolites),1)];
   Reversibilities = [Reversibilities; Reactions(n).Reversible];
   for m=1:length(Reactions(n).Metabolites),
      S((strcmp(Reactions(n).Metabolites{m}, Metabolites)),end) = Reactions(n).Stoichiometry(m);
   end
   for nn=1:length(Reactions(n).EnzymeNames),
      for m=1:length(Reactions(n).Regulators{nn}),
         Sreg((strcmp(Reactions(n).Regulators{nn}{m}, Metabolites)),end) =
Reactions(n).RegulationType{nn}(m);
      end
   end

end

Net.S = S;
Net.Sreg = Sreg;
Net.MetabName = Metabolites;
Net.EnzName = EnzymeNames;
```

4.5.6 Modify the model according to the scope

```
%For simplicity, quinones can be represented as 1.5 ATP
%For simplicity, thioredoxin can be substituted by NADPH
%For simplicity NAD-P-OR-NOP choose NADP. Later this could be expanded to use both substrates
Substitutions = {{1} {'|Menaquinones|'} {'='} {1.5} {'ADP'}; ...
            {1} {'|Menaquinols|'} {'='} {1.5} {'ATP'}; ...
            %{1} {'|Oxidized-NrdH-Proteins|'} {'='} {1.5} {'ADP'}; ...
            %{1} {'|Reduced-NrdH-Proteins|'} {'='} {1.5} {'ATP'}; ...
            %{1} {'|Oxidized-flavodoxins|'} {'='} {1.5} {'ADP'}; ...
            %{1} {'|Reduced-flavodoxins|'} {'='} {1.5} {'ATP'}; ...
            {1} {'|Quinones|'} {'='} {1.5} {'ADP'}; ...
            {1} {'|Reduced-Quinones|'} {'='} {1.5} {'ATP'}; ...
            {1} {'|Ubiquinones|'} {'='} {1.5} {'ADP'}; ...
            {1} {'|Ubiquinols|'} {'='} {1.5} {'ATP'}; ...
            %{1} {'|Ox-Thioredoxin|'} {'='} {1.5} {'NADP'}; ...
            %{1} {'|Red-Thioredoxin|'} {'='} {1.5} {'NADPH'}; ...
            {1} {'NAD-P-OR-NOP'} {'='} {1} {'NADP'}; ...
            {1} {'NADH-P-OR-NOP'} {'='} {1} {'NADPH'}; ...
%              {1} {'|Acceptor|'} {'='} {1} {'NADP'}; ...
%              {1} {'|Donor-H2|'} {'='} {1} {'NADPH'}; ...
```

```
            };
for n=1:size(Substitutions,1),
   Met1 = find(strcmp(Substitutions{n,2}, Net.MetabName));
   Met2 = find(strcmp(Substitutions{n,5}, Net.MetabName));
   Net.S(Met2,:) = Net.S(Met2,:)+(Substitutions{n,4}{1}/Substitutions{n,1}{1}).*Net.S(Met1,:);
   Net.S(Met1,:) = [];
   Net.Sreg(Met1,:) = [];
   Net.MetabName(Met1) = [];
end


%Remove negligible metabolites
VetoedClasses = {'Cofactors';
            'Vitamins';
            'Ions';
            'Salts';
            'Inorganic-Compounds';
            'THF-Derivatives';
            };
NegligibleMetabolites = [];
for n=1:length(Net.MetabName),
   [BadClass, CarbonCount, NitrogenCount] = CompoundChecker(strrep(Net.MetabName{n}, '|', ''),
VetoedClasses); % There seems to be an incosistency in the database where -GLU-N does not appear in
compounds
   if BadClass==1 || CarbonCount==0 || (CarbonCount==1&&NitrogenCount==0),
      NegligibleMetabolites = [NegligibleMetabolites; n];
   end
end

Net.S(NegligibleMetabolites,:) = [];
Net.MetabName(NegligibleMetabolites) = [];
Net.Sreg(NegligibleMetabolites,:) = [];

% Remove empty or repeated rxns
EmptyRxns =  find(sum(abs(Net.S),1) == 0);
Net.S(:,EmptyRxns) = [];
Net.Sreg(:,EmptyRxns) = [];
Net.EnzName(EmptyRxns) = [];
Reactions(EmptyRxns) = [];
Reversibilities(EmptyRxns) = [];
[~, Unq, ~] = unique(Net.S', 'rows');
RptRxns = logical(ones(size(Net.S,2),1));
RptRxns(Unq) = 0;
Net.S(:,RptRxns) = [];
Net.Sreg(:,RptRxns) = [];
Net.EnzName(RptRxns) = [];
Reversibilities(RptRxns) = [];
Reactions(RptRxns) = [];
Net.Reversibilities = Reversibilities;
```

4.5.7 Prepare flux distribution

```
LB = -1000*Net.Reversibilities;
UB = 1000*ones(size(Net.S,2),1);
```

```
% Fluxes taken from Fischer et al. 2004
GlucoseUptake = 7.1;
LB(RxnIndex('PTS', Net.EnzName)) = .9*GlucoseUptake;     % Assumine 90% of the glucose uptake goes
through PTS
UB(RxnIndex('PTS', Net.EnzName)) = .9*GlucoseUptake;
LB(RxnIndex('glk', Net.EnzName)) = .1*GlucoseUptake;          % The other 10% goes through glk
UB(RxnIndex('glk', Net.EnzName)) = .1*GlucoseUptake;
LB(RxnIndex('EtOH_out', Net.EnzName)) = .001;          % Ethanol production
UB(RxnIndex('EtOH_out', Net.EnzName)) = .001;
LB(RxnIndex('Acetate_out', Net.EnzName)) = 6.6;          % Acetate production
UB(RxnIndex('Acetate_out', Net.EnzName)) = 6.6;
LB(RxnIndex('Lactate_out', Net.EnzName)) = .001;          % Lactate production
UB(RxnIndex('Lactate_out', Net.EnzName)) = .001;
LB(RxnIndex('Biomass1', Net.EnzName):RxnIndex('Biomass26', Net.EnzName)) =
GlucoseUptake*.180*.41;       % Growth rate  = GlucoseUptake*GlucoseMW*BiomassYield
UB(RxnIndex('Biomass1', Net.EnzName):RxnIndex('Biomass26', Net.EnzName)) =
GlucoseUptake*.180*.41;
LB(RxnIndex('RXN-5822', Net.EnzName)) = 0;                    %NADP Phosphatase can be neglected
UB(RxnIndex('RXN-5822', Net.EnzName)) = 0;
LB(RxnIndex('ISOCIT-CLEAV-RXN', Net.EnzName)) = 0;                    %The glyoxylate bypass was
found inactive
UB(RxnIndex('ISOCIT-CLEAV-RXN', Net.EnzName)) = 0;

Aeq = zeros(9,length(LB));
Beq = zeros(9,1);

% Reverse reactions
Aeq(3,[RxnIndex('F16BDEPHOS-RXN', Net.EnzName), RxnIndex('6PFRUCTPHOS-RXN',
Net.EnzName)]) = [95 -5];          % fbp:pfk = 5:95
Aeq(4,[RxnIndex('PEPDEPHOS-RXN', Net.EnzName), RxnIndex('PEPSYNTH-RXN', Net.EnzName)]) =
[5 95];        % pyk:pps = 95:5
Aeq(7,[RxnIndex('R601-RXN', Net.EnzName), RxnIndex('SUCCINATE-DEHYDROGENASE-
UBIQUINONE-RXN', Net.EnzName)]) = [-2 1];         %frd:sdh = 1:2

% Parallel reactions
Aeq(5,[RxnIndex('MALATE-DEH-RXN', Net.EnzName), RxnIndex('MALATE-DEHYDROGENASE-
ACCEPTOR-RXN', Net.EnzName)]) = [1 -1];          % mdh:mqo = 1:1
Aeq(6,[RxnIndex('PPENTOMUT-RXN', Net.EnzName), RxnIndex('PRPPSYN-RXN', Net.EnzName)]) =
[1 1];         % prs:deoB (In His Pwy ATP->AMP:2ATP->2ADP) = 1:1     PRPPSYN
Aeq(8,[RxnIndex('NAD-SYNTH-GLN-RXN', Net.EnzName), RxnIndex('NAD-SYNTH-NH3-RXN',
Net.EnzName)]) = [5 -95];        %nadE(Glu):nadE(NH4) = 95:5
Aeq(9,[RxnIndex('PYRUVDEH-RXN', Net.EnzName), RxnIndex('PYRUVFORMLY-RXN',
Net.EnzName)]) = [5 -95];        %pdh:pfl = 95:5

Aeq = [Aeq; Net.S];

Beq = [Beq; zeros(size(Net.S,1),1)];

% Minimize repiration for semi-aerobic condition
f = zeros(length(LB), 1);
f(RxnIndex('CITSYN-RXN', Net.EnzName)) = 1;
x = linprog(f,[],[],Aeq,Beq,LB,UB,[],optimoptions('linprog','MaxIter',10000));
Net.Vref = x;
```

4.5.8 Remove reactions without flux

```
Net.S(:,abs(Net.Vref)<1e-10) = [];
Net.Sreg(:,abs(Net.Vref)<1e-10) = [];
Net.EnzName(abs(Net.Vref)<1e-10) = [];
Net.Reversibilities(abs(Net.Vref)<1e-10) = [];
Reactions(abs(Net.Vref)<1e-10) = [];
Net.Vref(abs(Net.Vref)<1e-10) = [];
```

4.5.9 Remove metabolites without reactions

```
EmptyMets =  find(sum(abs(Net.S),2) == 0);
Net.S(EmptyMets,:) = [];
Net.Sreg(EmptyMets,:) = [];
Net.MetabName(EmptyMets) = [];
```

4.5.10 Specify inlet reactions

```
InletRxns = {'PTS'; 'glk'; 'PTS_r'};
Net.Vin_index = [];
for n=1:length(InletRxns),
   Net.Vin_index = [Net.Vin_index; RxnIndex(InletRxns{n}, Net.EnzName)];
end
```

4.5.11 Specify outlet reactions

```
OutletRxns = {'EtOH_out'; 'Acetate_out'; 'Lactate_out'};
for n=1:26,
   OutletRxns = [OutletRxns; strcat('Biomass', num2str(n))];
end
Net.Vout_index= [];
for n=1:length(OutletRxns),
   Net.Vout_index = [Net.Vout_index; RxnIndex(OutletRxns{n}, Net.EnzName)];
end
```

4.5.12 Export the model

```
save('SampleCellWideModel', 'Net', 'Reactions')
```

5 References

1. Achcar F, Kerkhoven EJ, Bakker BM, Barrett MP, & Breitling R (2012) Dynamic modelling under uncertainty: the case of *Trypanosoma brucei* energy metabolism. *PLoS computational biology* 8(1):e1002352.

2. Albert MA*, et al.* (2005) Experimental and *in silico* analyses of glycolytic flux control in bloodstream form *Trypanosoma brucei*. *The Journal of biological chemistry* 280(31):28306-28315.

3. Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, & Reuss M (2002) Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnology and bioengineering* 79(1):53-73.

4. Conant GC & Wolfe KH (2007) Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Molecular systems biology* 3:129.

5. Curien G*, et al.* (2009) Understanding the regulation of aspartate metabolism using a model based on measured kinetic parameters. *Molecular systems biology* 5:271.

6. Helfert S, Estevez AM, Bakker B, Michels P, & Clayton C (2001) Roles of triosephosphate isomerase and aerobic metabolism in *Trypanosoma brucei*. *The Biochemical journal* 357(Pt 1):117-125.

7. Holzhutter HG (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *European journal of biochemistry / FEBS* 271(14):2905-2922.

8. Laisk A, Eichelmann H, & Oja V (2006) C3 photosynthesis *in silico*. *Photosynth Res* 90(1):45-66.

9. Mosca E*, et al.* (2012) Computational modeling of the metabolic States regulated by the kinase akt. *Frontiers in physiology* 3:418.

10. Nijhout HF, Reed MC, Budu P, & Ulrich CM (2004) A mathematical model of the folate cycle: new insights into folate homeostasis. *The Journal of biological chemistry* 279(53):55008-55016.

11. Pritchard L & Kell DB (2002) Schemes of flux control in a model of *Saccharomyces cerevisiae* glycolysis. *European journal of biochemistry / FEBS* 269(16):3894-3904.

12. Teusink B*, et al.* (2000) Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry. *European journal of biochemistry / FEBS* 267(17):5313-5329.

13. Zhu X-G, de Sturler E, & Long SP (2007) Optimizing the distribution of resources between enzymes of carbon metabolism can dramatically increase photosynthetic rate: a numerical simulation using an evolutionary algorithm. *Plant Physiol* 145(2):513-526.

14. van Heerden JH*, et al.* (2014) Lost in Transition: Start-Up of Glycolysis Yields Subpopulations of Nongrowing Cells. *Science (New York, N.Y.)* 343(6174):1245114.

15. Atsumi S, Hanai T, & Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* 451(7174):86-89.

16. Bogorad IW, Lin TS, & Liao JC (2013) Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature* 502:693-697.

17. Mainguet SE, Gronenberg LS, Wong SS, & Liao JC (2013) A reverse glyoxylate shunt to build a non-native route from C4 to C2 in Escherichia coli. *Metabolic engineering* 19:116-127.

18. Shen CR*, et al.* (2011) Driving forces enable high-titer anaerobic 1-butanol synthesis in Escherichia coli. *Applied and environmental microbiology* 77(9):2905-2915.

19. Shen CR & Liao JC (2013) Synergy as design principle for metabolic engineering of 1-propanol production in Escherichia coli. *Metabolic engineering* 17:12-22.

20. Contador CA, Rizk ML, Asenjo JA, & Liao JC (2009) Ensemble modeling for strain development of L-lysine-producing Escherichia coli. *Metabolic engineering* 11(4-5):221-233.

21. Rizk ML & Liao JC (2009) Ensemble modeling for aromatic production in Escherichia coli. *PloS one* 4(9):e6903.

22. Tan Y, Lafontaine Rivera JG, Contador CA, Asenjo JA, & Liao JC (2011) Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux. *Metabolic engineering* 13(1):60-75.

23. Tan Y & Liao JC (2012) Metabolic ensemble modeling for strain engineers. *Biotechnology journal* 7(3):343-353.

24. Tran LM, Rizk ML, & Liao JC (2008) Ensemble modeling of metabolic networks. *Biophysical journal* 95(12):5606-5617.

25. Rizk ML, Laguna R, Smith KM, Tabita FR, & Liao JC (2011) Redox homeostasis phenotypes in RubisCO-deficient Rhodobacter sphaeroides via ensemble modeling. *Biotechnology progress* 27(1):15-22.

26. Oberhardt MA, Palsson BO, & Papin JA (2009) Applications of genome-scale metabolic reconstructions. *Mol Syst Biol* 5:320.

27. Puchalka J*, et al.* (2008) Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology. *PLoS Comput Biol* 4(10):e1000210.

28. Tan Y, Rivera JG, Contador CA, Asenjo JA, & Liao JC (2011) Reducing the allowable kinetic space by constructing ensemble of dynamic models with the same steady-state flux. *Metabolic engineering* 13(1):60-75.

29. Zhang M, Eddy C, Deanda K, Finkelstein M, & Picataggio S (1995) Metabolic Engineering of a Pentose Metabolism Pathway in Ethanologenic Zymomonas mobilis. *Science (New York, N.Y.)* 267(5195):240-243.

30. Huo YX*, et al.* (2011) Conversion of proteins into biofuels by engineering nitrogen flux. *Nature biotechnology* 29(4):346-351.

31. Ingram LO, Conway T, Clark DP, Sewell GW, & Preston JF (1987) Genetic engineering of ethanol production in Escherichia coli. *Applied and environmental microbiology* 53(10):2420-2425.

32. Zhang K, Sawaya MR, Eisenberg DS, & Liao JC (2008) Expanding metabolism for biosynthesis of nonnatural alcohols. *Proceedings of the National Academy of Sciences of the United States of America* 105(52):20653-20658.

33. Steen EJ*, et al.* (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 463(7280):559-562.

34. Zhang K, Li H, Cho KM, & Liao JC (2010) Expanding metabolism for total biosynthesis of the nonnatural amino acid L-homoalanine. *Proceedings of the National Academy of Sciences of the United States of America* 107(14):6234-6239.

35. Dellomonaco C, Clomburg JM, Miller EN, & Gonzalez R (2011) Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals. *Nature* 476(7360):355-359.

36. Choi YJ & Lee SY (2013) Microbial production of short-chain alkanes. *Nature* 502(7472):571-574.

37. Fung E*, et al.* (2005) A synthetic gene–metabolic oscillator. *Nature* 435(7038):118-122.

38. Barkai N & Leibler S (1997) Robustness in simple biochemical networks. *Nature* 387(6636):913-917.

39. Alon U, Surette MG, Barkai N, & Leibler S (1999) Robustness in bacterial chemotaxis. *Nature* 397(6715):168-171.

40. Stelling J, Gilles ED, & Doyle FJ, 3rd (2004) Robustness properties of circadian clock architectures. *Proceedings of the National Academy of Sciences of the United States of America* 101(36):13210-13215.

41. Stelling J, Sauer U, Szallasi Z, Doyle FJ, 3rd, & Doyle J (2004) Robustness of cellular functions. *Cell* 118(6):675-685.

42. Farmer WR & Liao JC (2000) Improving lycopene production in Escherichia coli by engineering metabolic control. *Nature biotechnology* 18(5):533-537.

43. Zhang F, Carothers JM, & Keasling JD (2012) Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature biotechnology* 30(4):354-359.

44. Dahl RH*, et al.* (2013) Engineering dynamic pathway regulation using stress-response promoters. *Nature biotechnology* 31(11):1039-1046.

45. Wang L, Birol I, & Hatzimanikatis V (2004) Metabolic control analysis under uncertainty: framework development and case studies. *Biophysical journal* 87(6):3750-3763.

46. Steuer R, Gross T, Selbig J, & Blasius B (2006) Structural kinetic modeling of metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America* 103(32):11868-11873.

47. Alves R & Savageau MA (2000) Systemic properties of ensembles of metabolic networks: application of graphical and statistical methods to simple unbranched pathways. *Bioinformatics (Oxford, England)* 16(6):534-547.

48. Schwacke JH & Voit EO (2004) Improved methods for the mathematically controlled comparison of biochemical systems. *Theoretical biology & medical modelling* 1:1.

49. Allgower EEL & Georg K (2003) *Introduction to numerical continuation methods* (Society for Industrial and Applied Mathematics, Philadelphia, PA).

50. Danø S, Sørensen PG, & Hynne F (1999) Sustained oscillations in living cells. *Nature* 402(6759):320-322.

51. Chandra FA, Buzi G, & Doyle JC (2011) Glycolytic oscillations and limits on robust efficiency. *Science (New York, N.Y.)* 333(6039):187-192.

52. Ma L & Iglesias PA (2002) Quantifying robustness of biochemical network models. *BMC bioinformatics* 3:38.

53. Stelling J, Klamt S, Bettenbrock K, Schuster S, & Gilles ED (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420(6912):190-193.

54. Kitano H (2007) Towards a theory of biological robustness. *Molecular systems biology* 3:137.

55. Gronenberg LS, Marcheschi RJ, & Liao JC (2013) Next generation biofuel engineering in prokaryotes. *Current opinion in chemical biology* 17(3):462-471.

56. Schramm M & Racker E (1957) Formation of erythrose-4-phosphate and acetyl phosphate by a phosphorolytic cleavage of fructose-6-phosphate. *Nature* 179(4574):1349-1350.

57. Heath EC, Hurwitz J, Horecker BL, & Ginsburg A (1958) Pentose fermentation by Lactobacillus plantarum. I. The cleavage of xylulose 5-phosphate by phosphoketolase. *The Journal of biological chemistry* 231(2):1009-1029.

58. Schramm M, Klybas V, & Racker E (1958) Phosphorolytic cleavage of fructose-6-phosphate by fructose-6-phosphate phosphoketolase from Acetobacter xylinum. *The Journal of biological chemistry* 233(6):1283-1288.

59. Flamholz A, Noor E, Bar-Even A, & Milo R (2012) eQuilibrator--the biochemical thermodynamics calculator. *Nucleic acids research* 40(Database issue):D770-775.

60. Kuepfer L, Peter M, Sauer U, & Stelling J (2007) Ensemble modeling for analysis of cell signaling dynamics. *Nature biotechnology* 25(9):1001-1006.

61. Kitano H (2004) Biological robustness. *Nature reviews. Genetics* 5(11):826-837.

62. Newman JR*, et al.* (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* 441(7095):840-846.

63. DeBerardinis RJ & Thompson CB (2012) Cellular metabolism and disease: what do metabolic outliers teach us? *Cell* 148(6):1132-1144.

64. Lee Y, Lafontaine Rivera JG, & Liao JC (2014) Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. *Metabolic engineering* 25:63-71.

65. Chow SN, Hale JK, & Malletparet J (1975) Applications of Generic Bifurcation .1. *Arch Ration Mech An* 59(2):159-188.

66. Guckenheimer J & Holmes P (1983) Local Bifurcations. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, (Springer), pp 117-165.

67. Chung JD & Stephanopoulos G (1996) On physiological multiplicity and population heterogeneity of biological systems. *Chemical Engineering Science* 51(9):1509-1521.

68. Lyapunov AM (1892) The general problem of motion stability. *Annals of Mathematics Studies* 17.

69. Shannon CE (1948) A Mathematical Theory of Communication. *The Bell System Technical Journal* 27:379-423, 623-656.

70. Flikweert MT*, et al.* (1996) Pyruvate Decarboxylase: An Indispensable Enzyme for Growth of *Saccharomyces cerevisiae* on Glucose. *Yeast* 12(3):247-257.

71. Erasmus DJ, van der Merwe GK, & van Vuuren HJ (2003) Genome-wide expression analyses: Metabolic adaptation of *Saccharomyces cerevisiae* to high sugar stress. *FEMS Yeast Res* 3(4):375-399.

72. Savageau MA (1971) Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. *Nature* 229(5286):542-544.

73. Heinrich R & Rapoport TA (1974) A linear steady-state treatment of enzymatic chains. General properties, control and effector strength. *European journal of biochemistry / FEBS* 42(1):89-95.

74. Wang L & Hatzimanikatis V (2006) Metabolic engineering under uncertainty. I: framework development. *Metabolic engineering* 8(2):133-141.

75. Wang L & Hatzimanikatis V (2006) Metabolic engineering under uncertainty--II: analysis of yeast metabolism. *Metabolic engineering* 8(2):142-159.

76. Voit EO (2000) *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists* (Cambridge University Press).

77. Gutenkunst RN*, et al.* (2007) Universally sloppy parameter sensitivities in systems biology models. *PLoS computational biology* 3(10):e189.

78. Jaqaman K & Danuser G (2006) Linking data to models: data regression. *Nature reviews. Molecular cell biology* 7(11):813-819.

79. Le Novere N*, et al.* (2006) BioModels Database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic acids research* 34(Database issue):D689-691.

80. Atsumi S, Higashide W, & Liao JC (2009) Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nature biotechnology* 27(12):1177-1180.

81. Matsuoka Y & Shimizu K (2015) Current status and future perspectives of kinetic modeling for the cell metabolism with incorporation of the metabolic regulation mechanism. *Bioresources and Bioprocessing* 2(1):1-19.

82. Miskovic L & Hatzimanikatis V (2010) Production of biofuels and biochemicals: in need of an ORACLE. *Trends in biotechnology* 28(8):391-397.

83. Kacser H & Burns JA (1973) The control of flux. *Symposia of the Society for Experimental Biology* 27:65-104.

84. Curto R, Voit EO, Sorribas A, & Cascante M (1998) Mathematical models of purine metabolism in man. *Mathematical biosciences* 151(1):1-49.

85. Visser D & Heijnen JJ (2003) Dynamic simulation and metabolic re-design of a branched pathway using linlog kinetics. *Metabolic engineering* 5(3):164-176.

86. Dean JT, Rizk ML, Tan Y, Dipple KM, & Liao JC (2010) Ensemble modeling of hepatic fatty acid metabolism with a synthetic glyoxylate shunt. *Biophysical journal* 98(8):1385-1395.

87. Delgado JP & Liao JC (1991) Identifying rate-controlling enzymes in metabolic pathways without kinetic parameters. *Biotechnology progress* 7(1):15-20.

88. Zomorrodi AR, Lafontaine Rivera JG, Liao JC, & Maranas CD (2013) Optimization-driven identification of genetic perturbations accelerates the convergence of model parameters in ensemble modeling of metabolic networks. *Biotechnology journal* 8(9):1090-1104.

89. Hoops S*, et al.* (2006) COPASI- A COmplex PAthway SImulator. *Bioinformatics (Oxford, England)* 22(24):3067-3074.

90. Sahle S*, et al.* (2006) Simulation of biochemical networks using COPASI - A complex pathway simulator. *Proceedings of the 2006 Winter Simulation Conference, Vols 1-5*:1698-1706.

91. Rivera JGL, Lee Y, & Liao JC (2015) An entropy-like index of bifurcational robustness for metabolic systems. *Integrative Biology*.

92. Keseler IM*, et al.* (2011) EcoCyc: a comprehensive database of Escherichia coli biology. *Nucleic acids research* 39(Database issue):D583-590.

93. Keseler IM*, et al.* (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic acids research* 41(Database issue):D605-612.

94. Karp PD*, et al.* (2014) The EcoCyc Database. *EcoSal Plus* 2014.

95. Batchelor E & Goulian M (2003) Robustness and the cycle of phosphorylation and dephosphorylation in a two-component regulatory system. *Proceedings of the National Academy of Sciences of the United States of America* 100(2):691-696.

96. Feist AM*, et al.* (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular systems biology* 3:121.

97. Fischer E & Sauer U (2003) Metabolic flux profiling of Escherichia coli mutants in central carbon metabolism using GC-MS. *European journal of biochemistry / FEBS* 270(5):880-891.

98. Heinrich R & Schuster S (1996) *The Regulation of Cellular Systems* (Springer).

99. Chassagnole C, Rais B, Quentin E, Fell DA, & Mazat JP (2001) An integrated study of threonine-pathway enzyme kinetics in Escherichia coli. *Biochemical Journal* 356:415-423.

100. Stadtman E, Cohen G, LeBras G, & de Robichon-Szulmajster H (1961) Feed-back inhibition and repression of aspartokinase activity in Escherichia coli and Saccharomyces cerevisiae. *Journal of Biological Chemistry* 236(7):2033-2038.

101. Patnaik R, Roof WD, Young RF, & Liao JC (1992) Stimulation of Glucose Catabolism in Escherichia-Coli by a Potential Futile Cycle. *J Bacteriol* 174(23):7527-7532.

102. Deutscher J (2008) The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol* 11(2):87-93.

103. Clark B & Holms WH (1976) Control of Sequential Utilization of Glucose and Fructose by Escherichia-Coli. *J Gen Microbiol* 95(Aug):191-201.

104. Kaback HR (1969) Regulation of Sugar Transport in Isolated Bacterial Membrane Preparations from Escherichia Coli. *Proceedings of the National Academy of Sciences of the United States of America* 63(3):724-&.

105. Moritz B, Striegel K, de Graaf AA, & Sahm H (2000) Kinetic properties of the glucose-6-phosphate and 6-phosphogluconate dehydrogenases from Corynebacterium glutamicum and their application for predicting pentose phosphate pathway flux in vivo. *European Journal of Biochemistry* 267(12):3442-3452.

106. Sprenger GA (1995) Genetics of pentose-phosphate pathway enzymes of Escherichia coli K-12. *Arch Microbiol* 164(5):324-330.

107. Veronese FM, Boccu E, & Fontana A (1976) Isolation and Properties of 6-Phosphogluconate Dehydrogenase from Escherichia-Coli - Some Comparisons with Thermophilic Enzyme from Bacillus-Stearothermophilus. *Biochemistry-Us* 15(18):4026-4033.

108. Desilva AO & Fraenkel DG (1979) 6-Phosphogluconate Dehydrogenase Reaction in Escherichia-Coli. *Journal of Biological Chemistry* 254(20):237-242.

109. Antoniewicz MR, *et al.* (2007) Metabolic flux analysis in a nonstationary system: Fed-batch fermentation of a high yielding strain of E. coli producing 1,3-propanediol. *Metabolic engineering* 9(3):277-292.

110. Wasylenko TM, Ahn WS, & Stephanopoulos G (2015) The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in Yarrowia lipolytica. *Metabolic engineering*.

111. Wiback SJ, Mahadevan R, & Palsson BO (2004) Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: The Escherichia coli spectrum. *Biotechnology and bioengineering* 86(3):317-331.

112. Schellenberger J*, et al.* (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* 6(9):1290-1307.

113. Cornish-Bowden A (2013) *Fundamentals of enzyme kinetics* (John Wiley & Sons).

114. Liebermeister W, Uhlendorf J, & Klipp E (2010) Modular rate laws for enzymatic reactions: thermodynamics, elasticities and implementation. *Bioinformatics (Oxford, England)* 26(12):1528-1534.