Title
Stability as a Criterion for Metabolic Systems

Permalink
https://escholarship.org/uc/item/0ss674hh

Author
Theisen, Matthew

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Stability as a Criterion for Metabolic Systems

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Bioengineering

by

Matthew Theisen

2016

ABSTRACT OF THE DISSERTATION

Stability as a Criterion for Metabolic Systems

by

Matthew Theisen

Doctor of Philosophy in Bioengineering

University of California, Los Angeles, 2016

Professor James C. Liao, Chair

Simulation of metabolic systems with kinetic models requires a large number of

parameter values, which are either difficult or impossible to obtain experimentally. Network

information, such as stoichiometry, reversibility and steady state flux, can be used to determine

mechanistically realistic rate laws, and these can be used to constrain the parameter space to only

those values which satisfy the constraints. Until now, stability has been overlooked when

considering kinetic metabolic models. However, dynamical stability and robustness to

perturbation are important qualities for living organisms, since they may encounter changing

environments or stochastic variation across time or within populations. Considering stability can

both provide constraints on the parameter space and be used to interpret the response of the

model to queries about the performance of the metabolic system under various perturbations. I

have used stability analysis to predict the performance of many metabolic systems, with an

emphasis on providing guidance for experimental efforts and uncovering biological significance.

The uses of stability analysis have encompassed several projects. Optimization of a novel methanol condensation cycle (MCC) was accomplished by tuning the amount of an irreversible phosphoketolase enzyme to a local productivity and stability maximum, as predicted by stability analysis and confirmed by *in vitro* experimentation. Several other *in vitro* enzymatic were subjected to stability analysis, and predictions matched previously published experimental results.

Stability analysis was also applied to several microbial systems to maximize production of a desired compound: *n*-butanol in *Escherichia coli*, isobutanol in *Clostridium thermocellum* and lipids in *Yarrowia lipolytica*. In these systems, production simulations matched the observations and predictions for further production improvements were made.

Stability analysis was also applied to gain biological understanding of the significance of structural features of the Calvin-Bassham-Benson (CBB) pathway in plants. The phosphate/glyceraldehyde-3-phosphate translocator was identified as more important for stability than a proposed glucose-6-phosphate shunt. Further, productivity was increased after overexpression of sedoheptulose-1,7-biosphosphatase, but not RuBiSCO, in agreement with previous experimental reports.

The importance of stability in analysis of metabolic systems is affirmed by this work, and the techniques demonstrated here pave the way for even further explorations.

The dissertation of Matthew Theisen is approved.

Yi Tang

Beth Lazazzera

James C. Liao, Committee Chair

University of California, Los Angeles

2016

<u>**Table of Contents**</u>

Biographical Sketch:

PROFESSIONAL EXPERIENCE

Graduate Student Researcher, *UCLA,* Los Angeles, CA

MCAT Tutor, *Various agencies and self-employed*, Los Angeles, CA

Engineering Technician, *Medtronic Diabetes*, Los Angeles, CA

EDUCATION

University of California, Los Angeles, *PhD in Bioengineering,* Los Angeles, CA, 2016 (expected)

University of Minnesota, *BS in Chemical Engineering,* Minneapolis, MN, 2011

SCHOLARLY ACTIVITY:

Peer-Reviewed Publications

- Bogorad I, Chen CT, <u>Theisen M</u>, Wu TY, Lam A, Schlenz A, Liao J. "Building carbon–carbon bonds using a biocatalytic methanol condensation cycle". *PNAS*. 2014. (**111**)45:15928-33

- <u>Theisen M</u>, Rivera JL, Liao J. "Stability predicts productivity of enzymatic systems". *PLoS Computational Biology*. In revision.

Conference Talks

- <u>Theisen M</u>, Rivera JL, Liao J. "The State of Flux: Stability As a Criterion for Model Building and Prediction". *AICHe 2015.*

- Khodayari A, Dash S, <u>Theisen M</u>, Islam MA, Zheng Y, Liao JC, Stephanopoulos G, Maranas CD. "Development of kinetic models of metabolism for microbial production organisms". *AICHe 2015.*

- Lee Y, Rivera JGL, <u>Theisen M</u>, Liao JC. "Stability as a Criterion in Metabolic Design". *International Symposia on Chemical Reaction Engineering.* 2016. (Upcoming)

Posters

- Theisen M, Chen CT, Bogorad I, Wu TY, Lam A, Schlenz A, Liao J. "Methanol Bioconversion using the Ribulose Monophosphate and Non-oxidative Glycolysis Pathways". *2014 UCLA Tech Forum.*

- Lafontaine J, Lee Y, Theisen M, Stephanopolous G, Liao JC. "Ensemble Modeling for Increasing Lipid Production in *Yarrowia lipolytica*". *DoE Contractor-Grantee Meeting 2014.*

1. Introduction

1.1 Background

Increasing availability of natural gas has spurred interest in the conversion of single-carbon (C1) compounds, with particular interest in methanol.[1] Methane, the main component of natural gas, can be converted to methanol by a variety of methods including chemical conversion at the wellhead[2], while upgrading methanol to multi-carbon compounds remains difficult.[3,4] Concerns over the climate effects of carbon dioxide have spurred action to minimize its release into the atmosphere. Thus, more carbon-efficient methods for the utilization of methanol represents a goal which would address strong societal needs in these areas. As a reference point, natural enzymatic pathways for methanol processing would allow for a maximum of 67% carbon efficiency from methanol to higher alcohols, due to decarboxylation during the pyruvate dehydrogenase step.[5] To address this need, we have developed the methanol condensation cycle (MCC).[6] MCC converts one-carbon methanol to longer carbon-chain alcohols like *n*-butanol with 100% theoretical carbon efficiency.

Metabolic simulation tools have great promise to guide the development of new enzymatic pathways like MCC. One of the most popular tools developed so far is flux balance analysis (FBA). FBA uses stoichiometric information about a pathway to predict characteristics of the pathway like theoretical yield.[7] However, FBA is unable to utilize kinetic information about a pathway or organism. In general, functions (i.e. Michaelis-Menten) which relate enzymatic-reaction rate to substrate concentrations and parameters (V_{max}, K_m, etc.) are available.[8] However, a major difficulty in using kinetic models is obtaining accurate parameters.[9] A relatively new method for simulation of metabolic systems called ensemble modelling allows for kinetic simulation without *a priori* knowledge of parameters.[10,11] This is accomplished by using network

1

information, like steady state enzyme fluxes to constrain guesses about enzyme parameters to realistic values.

This simulation strategy may be helpful in the design of a methylotrophic *E. coli* strain. There are three critical enzymes for methylotrophy (growth on methanol): alcohol dehydrogenase (Adh), 3-hexulose-6-phosphate synthase (Hps) and 3-hexulose-6-phosphate isomerase (Phi). My preliminary results have shown that methanol consumption by *E. coli* is possible, and that it can be converted to biomass. Further, literature reports have identified methanol assimilation in *E. coli*.[12] However it is likely that levels of other enzymes, especially related to carbon rearrangement and glycolysis may require expression at different levels to accommodate the high carbon flux required for primary metabolism.

MCC bypasses the decarboxylating enzyme pyruvate dehydrogenase (Pdh) which generates key metabolic intermediate acetyl-CoA. Instead, it generates acetyl-CoA from acetyl-phosphate produced by phosphoketolase. The productivity of the pathway is hypothesized to have a local maximum relative to the concentration of phosphoketolase. Further, the pathway can be shown to be fully catalytic by analyzing ^{13}C labelling patterns of MCC products formed by reacting unlabeled sugar phosphates with ^{13}C labelled formaldehyde. *In silico* modelling of the cycle can be performed to confirm our understanding of experimental results.

Ensemble modelling and its extension ensemble modelling robustness analysis (EMRA) are relatively new methods for simulating metabolic systems.[10,13] In this work, EMRA and stability more broadly, will be demonstrated as useful tools for pathway development. Literature accounts have reported enzymatic systems with anomalous productivity characteristics.[6,14,15] For example, a glycolysis system showed decreasing productivity with increasing feed rate.[14] EMRA is a way to unify these anomalies with theoretical understanding.

Methylotrophic and non-methylotrophic organisms have fundamentally different metabolism. For example, the highly studied methylotroph *Bacillus methanolicus* MGA3 is not capable of growth on glucose and its only known multi-carbon growth source is mannitol.[16] There are very few reports of organisms which can utilize both glucose and methanol, with the most recent of such reports dating from the 1980s. These organisms have relatively long lag phases on both substrates.[17,18] Thus, it is possible that there are fundamental metabolic trade-offs that make simultaneous compatibility with methanol and glucose difficult. Elucidation of these fundamental differences will be attempted with ensemble modelling, yielding potential experimental insights. Further, overexpression and knockout targets will be identified by using ensemble modelling robustness analysis (EMRA).[13,19]

1.2 Importance of Cycles in Metabolism

The metabolism of life is a complex network of reactions. There are thousands of different reactions occurring even in a relatively simple life-form such as the bacterium *Escherichia coli*. However, many of these reactions occur with very little flux, and therefore it is often useful to break down metabolism into 'pathways', often focusing on the pathways that are highest in flux—the so-called primary metabolism. The exploration of these pathways and their variations is a major emphasis of this work.

Some pathways form a simple linear sequence (Fig 1-1), like glycolysis. Ignoring cofactors for the moment (NADH, ATP), glycolysis is a linear pathway, with each successive metabolite being converted to the next—without any complexities like branching or cycles. Glycolysis is one of the most important pathway for life, and is a highly conserved pathway found in almost all forms of life. It takes the energy locked up in the chemical bonds of glucose, and makes it available for the uses of life.

Simple Linear Pathway

S As starting substrate increases...

↓

↓ ...pathway flux increases...

↓

F ...and final product increases.
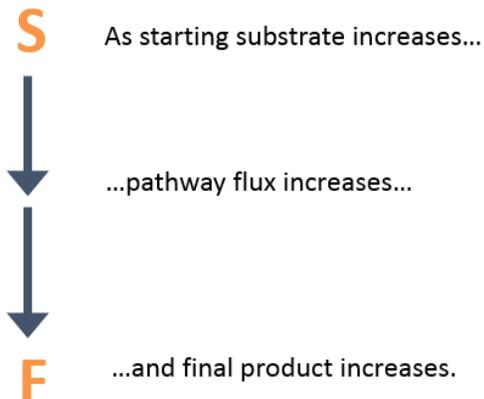
Fig. 1-1) Example of simple linear pathway. As substrate increases, pathway flux increases. Depletion of intermediates does not adversely affect pathway performance.

However, cyclic pathways have an equally, if not more important role in the chemistry of life. In general, the class of C1 biochemistry (single carbon compound chemistry), relies heavily on metabolic cycles, many of which are complex and have multiple branch points. These pathways include the Calvin-Bassham-Benson (CBB) pathway, used by plants for the fixation of carbon dioxide into sugars, and the ribulose monophosphate (RuMP) pathway of methanol assimilation.

Why is C1 biochemistry so important? Carbon dioxide is a low energy molecule, and the carbon contained in it can be used, along with some other source of energy (reducing power from NADPH, for instance), to form the organic molecules which make up life. To form carbon skeletons (sugars) in this way power the rest of life on this planet, one carbon dioxide needs to be bonded with other carbons, a feat which allows the rest of life to flourish.

Why are cycles so important to C1 biochemistry? Chemically it is infeasible to bond two C1 molecules directly to a C2. No enzymes are known which can accomplish this reaction, so instead, the C1 molecule must be incorporated into a longer molecule, as is the case in the CBB and RuMP pathways, where a C1 is incorporated into a C5 molecule to form a C6. The C6

4

molecule is then reshuffled back down to a C5, such that 5xC6 molecules shuffle down to 6xC5 molecules, resulting in a net gain fixed carbon (Fig 1-2).
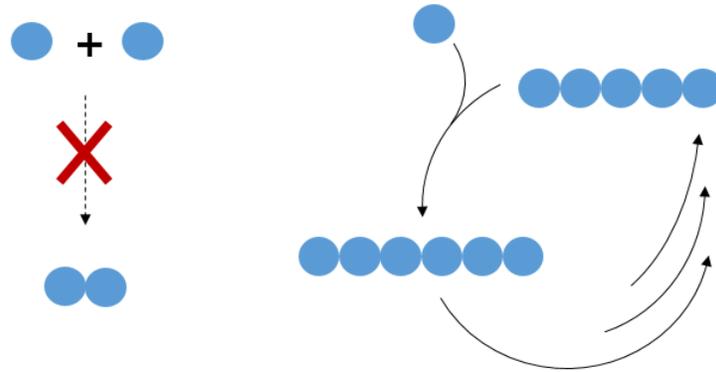
Fig. 1-2) Most C1 metabolism uses cycles. The reaction C1 + C1 = C2 is mechanistically unfeasible. Thus, C1 molecules like carbon dioxide and methanol must be incorporated into longer carbon chains (usually C5 sugars) to complete assimilation.

Given the importance and complexity of C1 metabolism, we need to understand how it can fail. A critical difference between linear pathways and cycles is that, in the case of cycles, if an intermediate is depleted, the pathway can no longer operate (Fig 1-3).

Pathway with Cycle

S Additional substrate will not be converted into product.

If intermediate metabolite is depleted...

Flux stops.

F
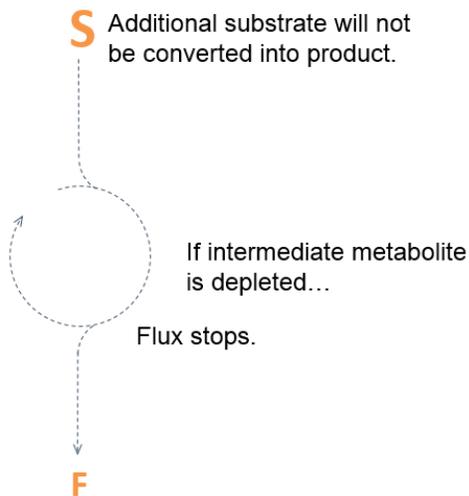
Fig. 1-3) Cyclic metabolism relies on the presence of intermediates for pathway function. For example, if the pathway cyclic intermediates are depleted, the addition of more starting substrate (S) will not result in more production.

1.3 Demonstrating Instability in a Simplified Metabolic Cycle

To demonstrate how we may begin to understand the stability of cyclic pathways, I defined a 'simplified' version of a cyclic pathway (Fig 1-4), and analyzed when it would fail, finding an analytical solution mathematically.
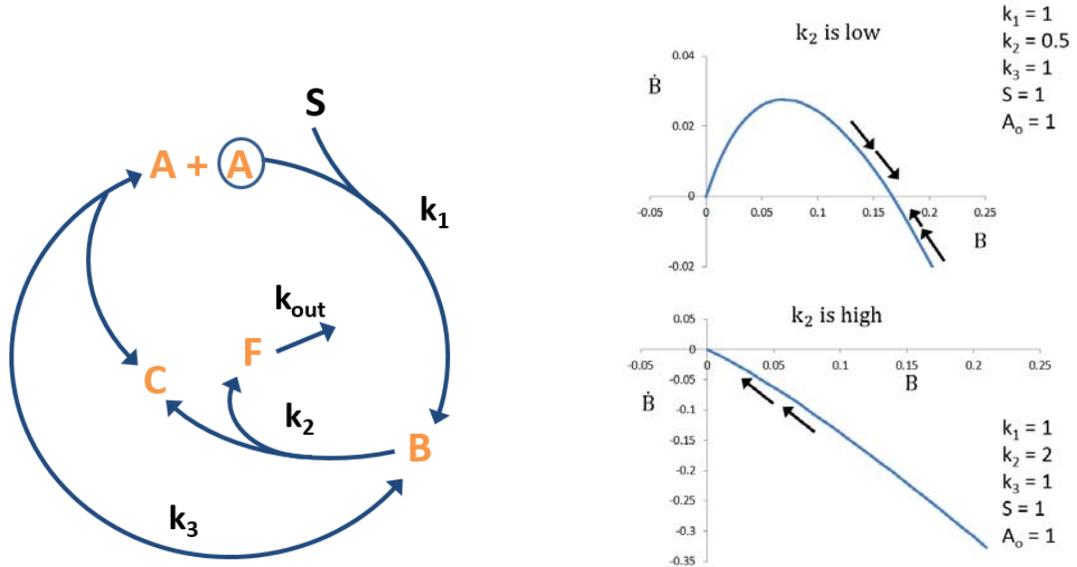
Fig. 1-4) Simplified version of MCC pathway. Graphs showing B' vs. B for different parameter values, showing that for some values, a non-zero stable steady state exists.

The system is kinetically defined as mass action with the following equations. B, C and F are assumed to have no initial concentration, only A.

$$\dot{A} = -k_1AS + 2k_3BC - 2k_3A^2$$

$$\dot{B} = k_1AS - k_2B - k_3BC + k_3A^2$$

$$\dot{C} = k_2B - k_3BC + k_3A^2$$

$$\dot{F} = k_2B - k_{out}F$$

$$B_o = C_o = F_o = 0$$

$$Eqns. \; 1\,a) - 1\,e)$$

Where k_i is rate constant i and capital letters indicate metabolite concentrations. The system is conserved, meaning the total amount of metabolites in the cycle (A, B, C) is constant. This can be found by observing the following:

$$A = A_o - \varepsilon_1 + 2\varepsilon_3$$

$$B = 0 + \varepsilon_1 - \varepsilon_2 - \varepsilon_3$$

$$C = 0 + \varepsilon_2 - \varepsilon_3$$

$$A + B + C = A_o$$

$$C = A_o - A - B$$

$$Eqns.\ 2\ a) - e)$$

Where ε_i indicates extent of reaction i. To find the nature of the steady state equation set Eqn. 1a) equal to zero. It is then possible to solve for A in terms of only A_o, B, S and k parameters using 2e). At this point, substituting this expression for A into Eqn. 1b) would yield an 'A-static' derivative of B—meaning that given A has stabilized, what is the time change of B? The result is the following:

$$\dot{B} = \frac{1}{2}Sk_1\sqrt{A_oB - \frac{3}{4}B^2 + \frac{k_1}{4k_3}BS + \frac{k_1^2}{16k_3^2}S^2 - Bk_2 - k_3\left(\frac{k_1^2}{8k_3^2}S^2 + \frac{k_1}{4k_3}BS\right)}$$

$$Eqn.\ 3$$

The roots of this equation would indicate the overall steady state—where A, B and C are not changing. The roots are:

7

$$B = 0; B = \frac{k_1^2}{k_3} \frac{S^2(k_3 A_o - k_2)}{\left(k_1^2 S^2 + 2k_2 k_1 S + 4k_2^2\right)}$$

Eqn. 4

The second root of B must be positive for a non-zero steady state to exist. Thus:

$$A_o > \frac{k_2}{k_3}$$

Eqn. 5

This is a beginning discussion on the importance of cycles in metabolism, and the role stability plays in performance of metabolic systems, both of which are major themes and will receive much more sophisticated treatments throughout.

2. Methanol Condensation Cycle & Methanol Growth

2.1 Methanol Condensation Cycle (MCC)

2.1.1 MCC Introduction

This chapter discusses work on developing the methanol condensation cycle (MCC) and attempts to engineer *E. coli* for growth on methanol. MCC (Fig. 2-1) uses methanol dehydrogenase to convert methanol to formaldehyde. From there, hexulose-phosphate synthase (Hps) us used to incorporate formaldehyde into ribulose-5-phosphate (Ru5P). Phosphoketolase then cleaves fructose-6-phosphate (F6P) to erythrulose-4-phosphate and acetyl-phosphate. Carbon is then rearranged to regenerate Ru5P.
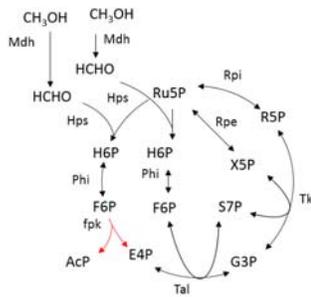
Fig. 2-1. Full version of MCC pathway. (Enzyme and compound names in Appendix B)

One concern in the implementation of MCC is the presence of a 'kinetic trap' whereby phosphoketolase completely consumes five- and six-carbon sugar phosphates (F6P, X5P, R5P, Ru5P). This leaves only G3P and/or E4P which cannot react with each other in the presence of only MCC enzymes. Phosphoketolase is known to be irreversible[20] and its thermodynamic properties (ΔG = -40 kJ/mol) support this interpretation. In contrast, other enzymes of the cycle (Rpe, Rpi, Tkt, Tal, Phi) are highly reversible.

MCC can be approximately modeled by a simple metabolic cycle of the form shown in Chapter 1. Due to the simple nature of this cycle, the conditions under which it will reach a stable steady state can be solved for analytically. The solution requires the irreversible enzyme rate (represents phosphoketolase) to be slower than the reversible enzyme rate (stands for Rpe, Rpe, Tkt, Tal). Additionally, the total amount of cycle intermediate must be higher than the ratio of the two rates. This requirement for maintenance of cycle intermediates is reminiscent of the well-known feature of the TCA cycle, cataplerosis.[21] TCA intermediates are replenished via pyruvate carboxylase or phosphoenolpyruvate carboxylase in response to acetyl-CoA flux which allows the cycle to continue functioning.[22]
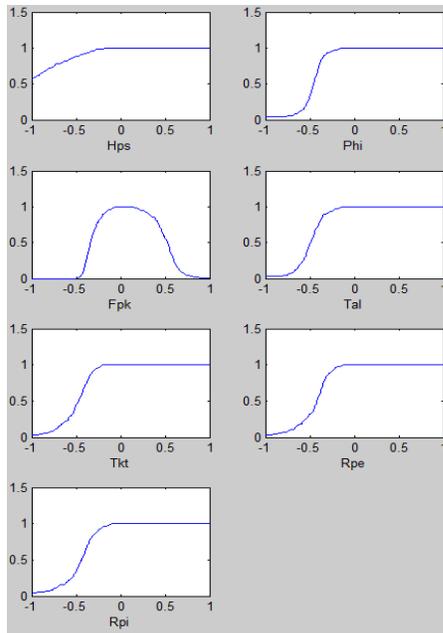
Fig. 2-2. Ensemble Modelling Robustness Analysis (EMRA) of the core MCC pathway. Y-axis represents the stability of the system. Specifically, it is the fraction of parameter sets that have a stable steady state. X-axis represents the log10 fold change of enzyme amount. Phosphoketolase (Fpk) shows a local maximum in robustness relative to enzyme amount while other enzymes show only increasing robustness for increasing amounts.

2.1.2 Ensemble Robustness Analysis of MCC *in vitro*

Ensemble Model Robustness Analysis (EMRA) is a method which analyzes the robustness of metabolic systems using only knowledge of the stoichiometry of the system and a pre-determined reference steady state[13]. Values for enzymatic parameters like K_m can be randomly drawn such that the reference steady state is maintained. Then, each set of parameters can be inspected for robustness by 'perturbing' the system by altering the amount of various enzymes. Some sets of parameters may be observed to become unstable at various levels of perturbation. This instability is detected by noting the sign of the real part of the eigenvalues of the Jacobian.[23]

A model of an *in vitro* MCC system (not including other cell metabolism) was developed which included the stoichiometry of the system, the reversibilities of the enzymes and the reference steady state. Using the ensemble modelling robustness analysis method, the robustness of the pathway to changes in each enzyme was analyzed. (Fig. 2-2) The reversible carbon rearrangement

10

enzymes (Rpe, Rpi, Tkt, Tal) must be present at a high level for maximum robustness and productivity. The irreversible enzyme, phosphoketolase (Fpk), shows a local maximum for robustness with both too little and too much being harmful for productivity, while all other enzymes have only increasing robustness with increasing amount.

2.1.3 Demonstration of catalytic MCC

Demonstrating the recycling feature of MCC is key to proving its overall function. This is because product can be formed using carbon only from the sugar phosphate intermediates, without incorporating significant amounts of foromaldehyde. An assay tested the core of MCC, going from formaldehyde to acetate. ^{13}C labelled formaldehyde was added to unlabeled ribose-5-phosphate (R5P).
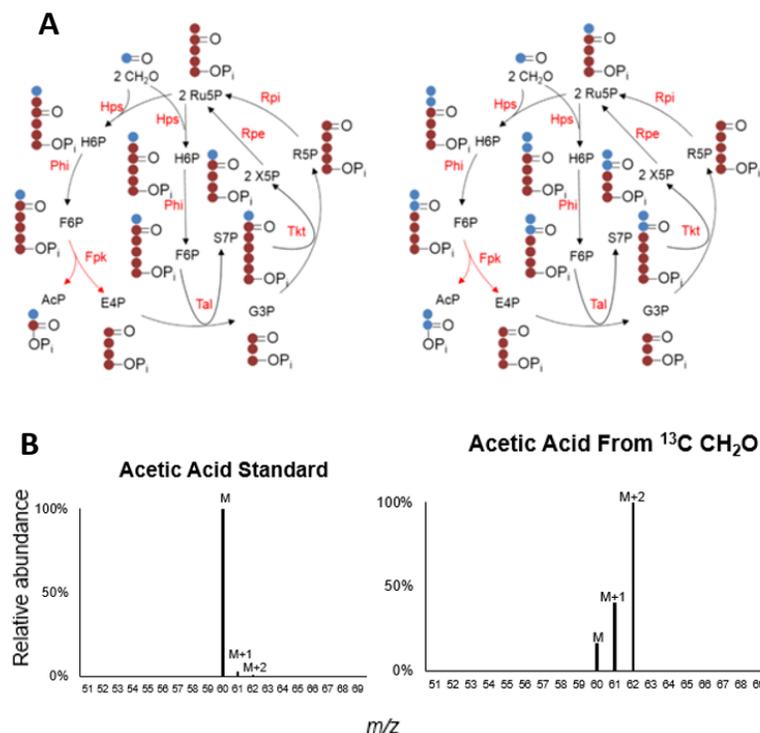
Fig. 2-3. A) Tracing 13C labelled carbon as it works through the cycle, generation of isotope 62 (M+2) requires a fully catalytic cycle where carbon is successfully regenerated. **B)** Acetic acid standard mass scatter showing m/z = 60 is the dominant peak. In an in vitro assay using purified enzymes, m/z = 62 was identified as the dominant peak, indicating fully catalytic cycle.

11

There are three distinguishable isotopes of acetic acid: the unlabeled 60-isotope (M), the single-labeled 61-isotope (M+1) and the double-labelled 62-isotope (M+2). The 62 (M+2) isotope is only generated in a fully catalytic process (Fig 2-3A), where carbon is turned over through the cycle at least twice. In fact, 62 is the dominant isotope detected in the labelling experiment (Fig 2-3B).

For this experiment, acetic acid was quantified on GC/MS. A standard curve for acetate was generated with relative standard deviation (RSD) of 3-7%. Quantifying acetate and its isotopes has previously been reported using the 60/61/62 m/z peaks as the 'quant' ions.[24] Using 60/61/62 as a quant ion is beneficial since in unlabeled acetate, 60 is a lone peak without surrounding peaks (Fig 2-3B).[25] Thus, it is very straightforward to quantify each isotope of acetic acid present when using 13C substrates for production.

2.1.4 Demonstration of an *in vitro* kinetic trap in MCC

To confirm the existence of a kinetic trap in MCC, *in vitro* assays of MCC using purified enzymes were carried out. These assays tested the core of MCC, going from formaldehyde to acetate. [13]C labelled formaldehyde was added to unlabeled ribose-5-phosphate (R5P). Varying the amount of phosphoketolase (Fpk) enzyme used in the reaction mixture was shown to cause a local maximum in the amount of acetic acid produced (Fig 2-4, bars) in the reaction mixture. This confirms the algebraic cycle analysis (Appendix A) as well as the EMRA analysis from Fig. 3.
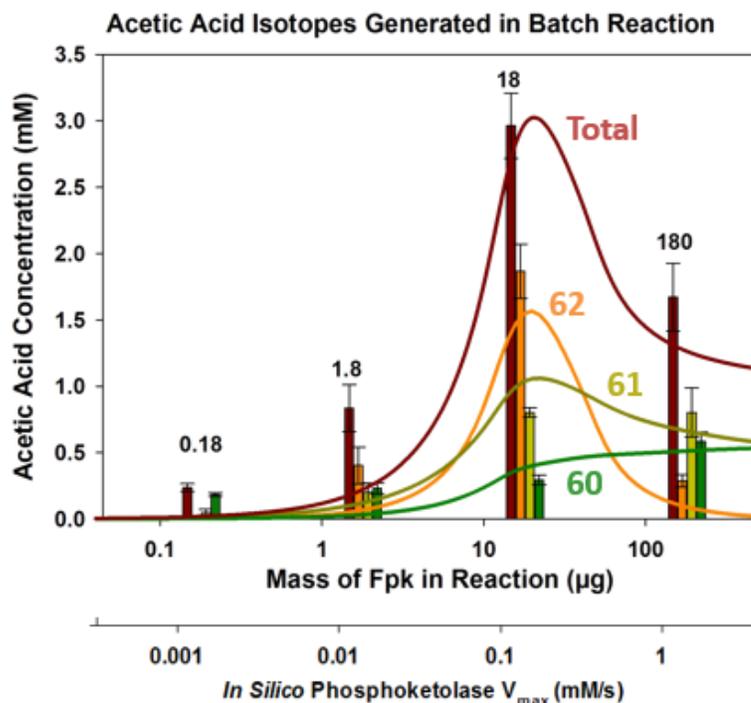
Fig. 2-4. Results of in vitro (bars, N=3, SSD shown) and in silico (lines) experiments for batch conversion of formaldehyde to acetic acid. Total acetic acid and isotope distributions shown.

Additionally, the distribution of acetic acid isotopes generated at each Fpk level was measured. A kinetic model of the system including 13C labelling effects was constructed (Fig. 2-4, lines). Time domain simulation of the experiment also showed a similar trend for overall acetic acid production and the distribution of acetic acid isotopes generated by the ^{13}C formaldehyde. The simulation used a kinetic Michaelis-Menten model using Km values randomly selected over a 10-fold range and V_{max} values (besides Fpk) randomly selected over a 3-fold range. The average of ten scenarios was taken.

2.1.5 Formaldehyde feeding of MCC for higher production in cycle

To push production higher, formaldehyde feeding assays were performed. An excessive initial bolus of formaldehyde was found to be ineffective [Data not shown]. This may because of

formaldehyde enzyme toxicity. Formaldehyde is known to be reactive to proteins by causing cross-linking.[26]

To reduce the concentration in the reaction mixture formaldehyde feeding was performed at 5 mM/hr for 4 hours with an initial bolus of 5mM formaldehyde and 2 mM R5P. A control without transaldolase (Tal) was used to illustrate the impact of carbon recycling on cycle performance. Additionally, a no feed control was also used.
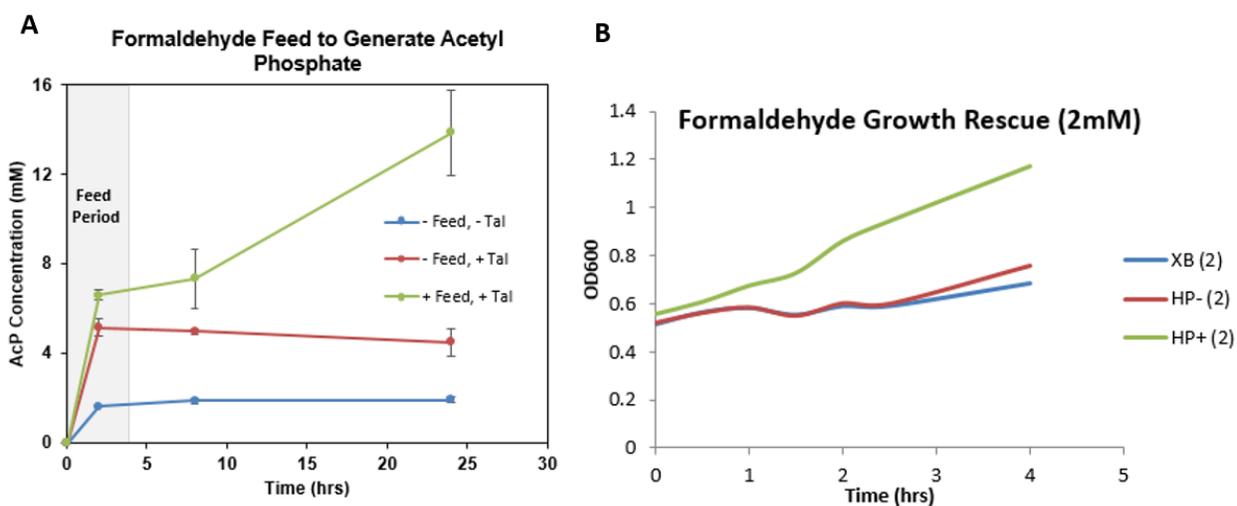
Fig. 2-5. A) *In vitro* conversion of formaldehyde to acetyl-phosphate. In the feed condition, formaldehyde was fed during the first 4 hours at 5 mM/hr. **B)** An Hps/Phi expressing strain of E. coli is resistant to formaldehyde up to 2 mM. XB strain is wild type (XL1 Blue). HP- is Hps/Phi expressing strain, no IPTG. HP+ is Hps/Phi expressing strain, with IPTG.

The cycle was found (Fig. 2-5A) to perform conversion to 100% (theoretical conversion would be $25/2 + 2 = 14.5$ mM) (within error). In other experiments, when formaldehyde was fed at a faster rate, or for a longer time period, overall production at 24 hours was found to be less than 100%.

2.2 In vivo methylotrophy & Formaldehyde Tolerance

2.2.1 Formaldehyde tolerance of *E. coli*

In vivo formaldehyde toxicity is likely to be an issue considered during the induction of methylotrophy in *E. coli* and MCC production in *E. coli*. Formaldehyde is known to be toxic to

cells.[27] One possible mechanism for this toxicity is DNA-protein crosslinking[27,28]. A formaldehyde-resistant strain of *E.coli* has been identified which uses plasmid-based expression of a formaldehyde dehydrogenase [29,30].

Genes for the enzymes 3-hexulose-6-phosphate synthase (Hps) and 3-hexulose-6-phosphate isomerase (Phi) were introduced into *E. coli* on a plasmid with an IPTG-inducible promoter to generate strain HP. These genes have been functionally expressed as a bifunctional fusion protein in *E. coli*, but have not previously been shown to confer formaldehyde resistance [31]. Cells were induced at OD600 0.4-0.6 with IPTG at 0.25 mM and allowed to grow overnight to saturation at 25 C. Then they were then diluted to OD_{600} ~0.5 with LB while maintaining concentrations of IPTG and kanamycin. Formaldehyde was added in three different concentrations, 0, 1 and 2 mM. Additionally, three conditions of cells were tested: XL1-Blue (Commerical cloning strain from Agilent Technologies, XB), HP+ (+IPTG) and HP- (-IPTG).

The results (Fig. 2-5B) show that at 2 mM formaldehyde, there is a significant difference between HP+ and the two other experimental types. Specifically, HP +IPTG continues growing from about OD 0.5 to 0.9 in four hours while without IPTG and without plasmid, they advance to only about 0.6. Formaldehyde tolerance may be an important part of allowing *E. coli* to become methylotrophic, since formaldehyde is a key intermediate in the consumption of methanol, and native methylotrophs have formaldehyde detoxification systems[32,33]. No difference in behavior was observed at 0 and 1 mM formaldehyde and growth was normal (not shown).

2.2.2 Methanol tolerance and consumption

Wild type *E. coli* is relatively tolerant to methanol, growing at up to 8% methanol in LB (Fig. 2-6A). Concentrations of up to 4% do not significantly affect *E. coli* growth. Cells were grown to saturation, then diluted at 1% and allowed to grow to OD600 of approximately 0.4-0.5.

These exponential phase cells were then allowed to grow in LB with various methanol concentrations as shown. The result shows that at 8% methanol, growth is significantly reduced, but still present. This result is encouraging for bioprocessing of methanol with *E. coli*. Longer chain alcohols like isobutanol completely stop growth at levels below 1%[34], so the tolerance of up to 8% methanol is favorable for the use of methanol as a feedstock.
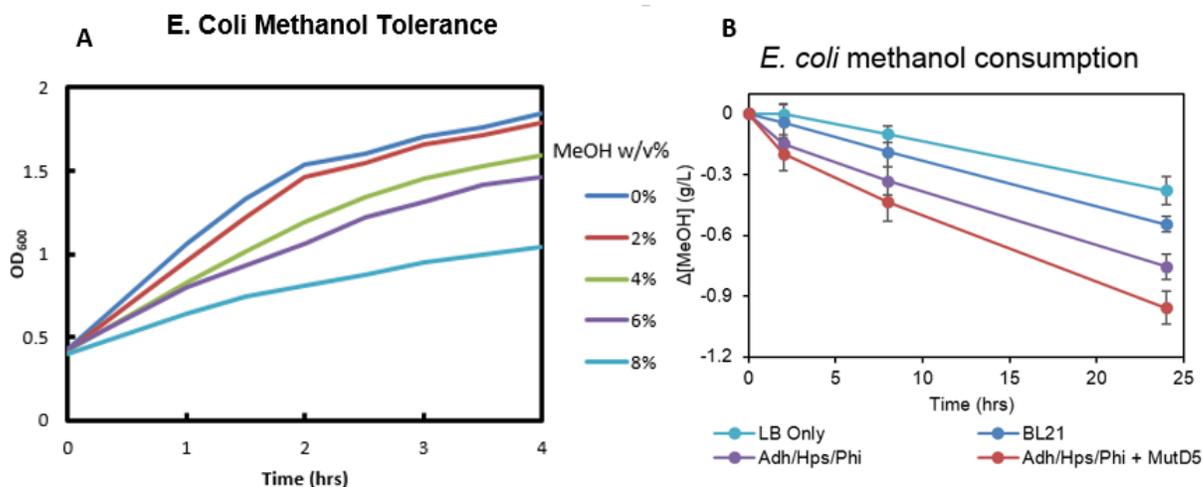
Fig. 2-6) A) Methanol tolerance of exponentially growing JCL16 *E. coli* in LB with varying methanol concentrations. **B)** Methanol consumption by transformed strains of *E. coli*. Media contained 2 g/L methanol. Adh/Hps/Phi are the three genes required for methylotrophy while MutD5 is a mutator allele.

Methanol consumption theoretically requires only three enzymes in *E. coli*: Adh, Hps and Phi, since the other enzymes of RuMP are already on the E. coli genome[5,35]. Initial tests have shown that these three enzymes are indeed sufficient for some methanol consumption. An inducible plasmid containing genes for all three enzymes was constructed and introduced to *E. coli*. Another strain also containing mutD5, a mutator allele, was also measured (Fig. 2-6B). The strains were then grown and induced overnight. The strains were then fed 50 mM MeOH (~2 g/L) and grown for 24 hours. Methanol was measured at four time points. The net change in methanol was compared to native *E. coli* strains which do not have methanol assimilation genes and an evaporation control (LB Only).

2.2.3 Partial methylotrophy by *E. coli*

16

Partial methylotrophy (conversion of methanol to biomass) by *E. coli* has been demonstrated using JCL16 E. coli expressing Adh, Hps & Phi—the three RuMP genes not natively expressed by *E. coli*. The cells were first induced in exponential phase and grown to saturation. After that, the cells were diluted with M9 and fed with 5% LB medium. The cells were well mixed to ensure homogeneous distribution of cells and nutrients, then separated into two groups, one to which methanol was added, and a no methanol control. The cells were also fed methanol daily (50 mM) throughout the experiment and IPTG was re-added on day 3. This experiment was also performed with JCL16 wild type.
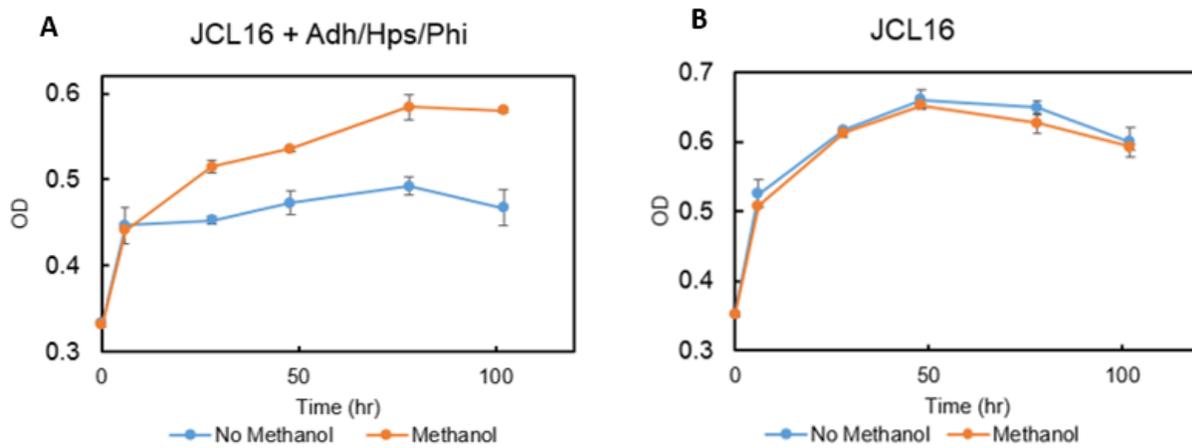
Fig. 2-7) Partial methylotrophy by E. coli **A)** JCL16 E. coli transformed with the genes essential for methylotrophy shows additional growth in the presence of 100mM methanol. The base medium was M9 with 3% LB. **B)** The wild type control (JCL16) showed no additional growth with methanol.

The wild type control showed no additional growth in the methanol condition (Fig. 2-7B), while the cells expressing Adh/Hps/Phi (Fig. 2-7A) showed significant additional growth with methanol.

3. Stability Predicts Productivity of *in vitro* Enzymatic Systems

3.1 Introduction

(Note: This chapter is from "Stability of Ensemble Models Predicts Prouctivity of Enzymatic Systems" in *PLoS Computational Biology*. Co-author Jimmy G Lafontaine Rivera provided valuable discussions and James Liao served as PI and corresponding author)[36]

Metabolic systems typically operate either under a stable steady state or an oscillatory mode. A non-oscillatory unstable system may result in multiple problems, including depletion of metabolites essential for growth, accumulation of toxic intermediates, or depletion of cofactors in the pathway—all ultimately leading to loss of production or cell death. While systems with stable steady states or sustained oscillation have been studied extensively [37–42], to our knowledge metabolic systems prone to instability have not been investigated as much. Both stable (Fig. 3-1a) or unstable (Fig 3-1b) system have a mathematical steady state (or fixed point), but the unstable steady state is not realizable in the physical world because any deviations from the steady state are amplified. Therefore, through evolution the unstable systems are selected against or stabilized by various levels of controls. However, the issue of stability is particularly important when engineering a novel pathway or altering an existing one.
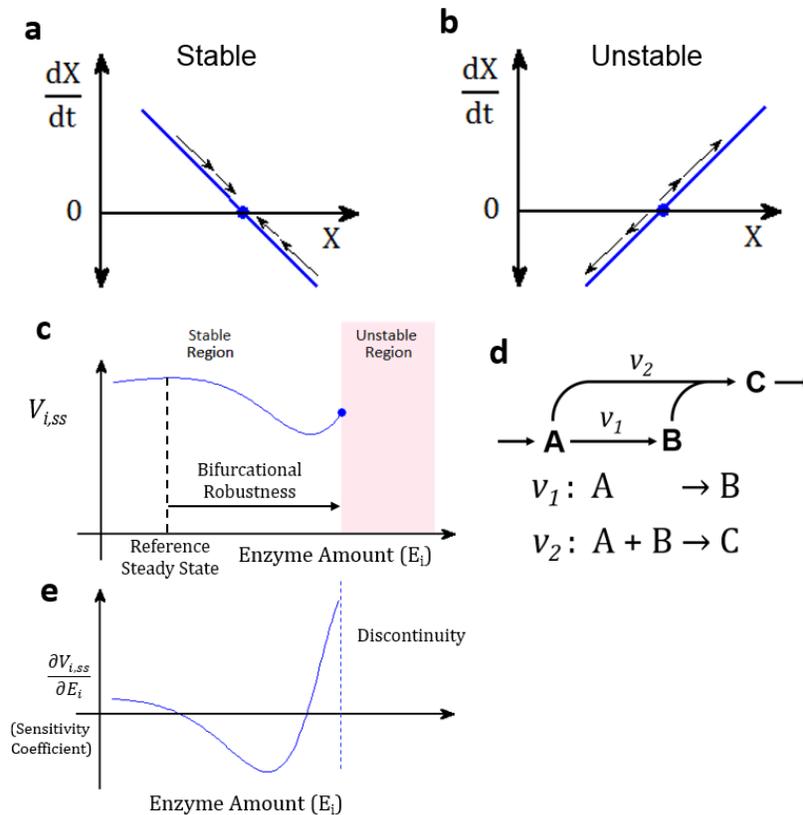
Fig. 3-1) Schematic figure showing how instability can occur and how it can cause lower production in batch experiments. **a-b)** In a one-dimensional dynamical system, the sign of $d\dot{x}/dx$ determines stability of a fixed point ($\dot{x} = 0$). If the sign of $d\dot{x}/dx$ is negative **(a)**, the system is stable to stochastic perturbations from the fixed point. In contrast, if $d\dot{x}/dx$ positive **(b)**, the fixed point is unstable. In a multivariate system, the analogous value is the maximum of the real parts of the eigenvalues of the Jacobian matrix. (i.e. if max(Re(Eig(Jac)))) is greater than 0 (the jacobian is singular), the fixed point will be unstable, if it is less than zero the fixed point will be stable unstable). **c)** (adapted from Lee, Rivera & Liao) Instability may be detected by using Ensemble Modelling Robustness Analysis. Bifurcational robustness investigates the distance between the reference steady state and the bifurcation point. **d)** A kinetic trap in which multiple reactions (v_1 & v_2) are competing for the same substrate (A). If the enzyme catalyzing v_1 increases greatly, it may cause instability by decreasing [A] so much that v_2 can not continue. **e)** Traditional sensitivity analysis calculates the sensitivity coefficient which represents the derivative of steady state production with respect to enzyme amount. However, sensitivity analysis doesn't investigate the likelihood of instability.

Furthermore, even starting from a stable steady state system, increasing an enzyme activity beyond a specific level may result in system failure (see Fig. 3-1c, adapted from Lee et al. 2014 [13]) because the system enters an unstable region, resulting in loss of a productive steady state. The likelihood of losing stability is characterized by bifurcational robustness using Ensemble Modeling for Robustness analysis (EMRA) [13]. Instability caused by enzyme perturbation has been predicted in proposed synthetic pathways and natural pathways in previous analyses[13,43]. One means of stability loss, among other possibilities, is a kinetic trap (Fig. 3-1d), resulting from a metabolic

branch point within a cyclic pathway. Upon perturbation, a kinetic trap may cause a sudden, unexpected, and qualitative change in dynamic behavior (Fig. 3-1c). Since cyclic pathways are common in metabolism, particularly when cofactor recycling are involved, such examples are copious. The bifurcational robustness is a measure of how far an enzyme amount must be perturbed before bifurcation occurs (Fig. 3-1c). Sudden system failure due to entering an unstable regime differs from the gradual deterioration of performance characterized by local sensitivity analysis. Sensitivity analysis, Biochemical Systems Theory [44–48], or metabolic control analysis (MCA) [49] is concerned with identifying the sensitivity coefficient (Fig 3-1e), which is the derivative of steady state production flux with respect to enzyme amount. In this work, we further examine the tendency for a metabolic system to be unstable based on their intrinsic network structure, which is determined by the network stoichiometry and kinetic rate laws. One way that this work builds on global sensitivity analysis is in that it focuses heavily on what we term the bifurcational robustness (Fig 3-1c), rather than the value of the sensitivity coefficient.

In previous uses of EMRA, unstable parameter sets found while constructing ensembles were discarded [13,43]. Here, we examine the intrinsic probability for a system to be unstable. This is fundamentally distinct from the tendency to bifurcate upon change from a stable steady state. In addition, previous EMRA simulations were applied to continuous processes. However, production experiments using enzymatic systems—whether *in vitro* or *in vivo*—are often carried out as a batch system due to practical considerations. Thus, it is unclear how simulations from a continuous mode can inform experimental strategies for new metabolic pathways which are investigated in batch or cell free experiments. Using four metabolic systems, we showed that the instability problem discussed above is indeed an issue, even with batch experiments. Interestingly, this type

of abrupt change is observed in common biological systems, including glycolysis. The results suggest that the stability issues may be more prevalent than previously appreciated.

3.2 Systems Description

We use the following enzymatic systems as examples for our investigation. Three of these systems have been described previously and some experimental data are available to validate our predictions. The other system (glucose to isoprene pathway) has not been experimentally investigated.
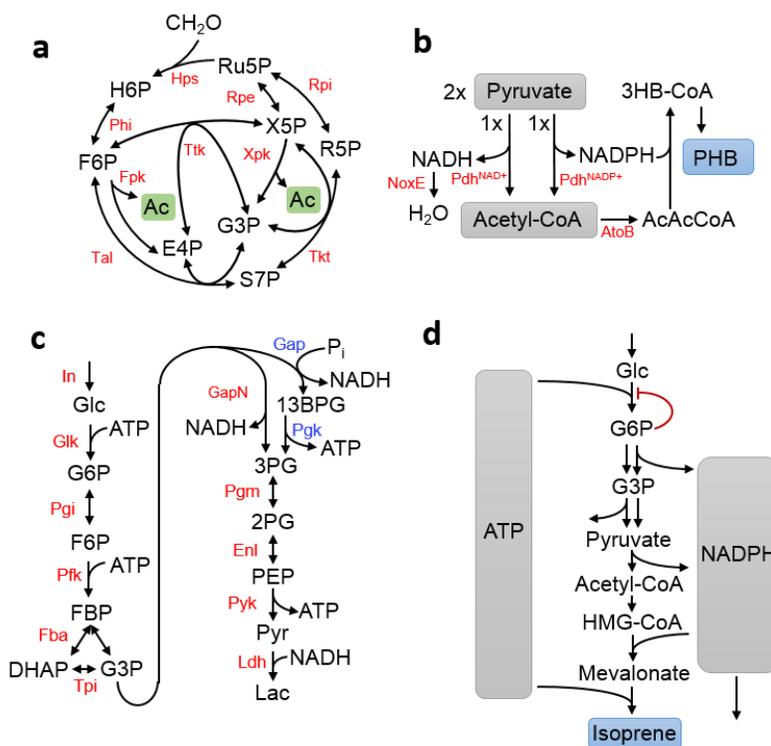
Fig. 3-2) Schematics showing four enzymatic systems which can be investigated by EMRA. **a)** A methanol condensation cycle (MCC) which converts formaldehyde to acetyl-phosphate with 100% carbon efficiency. Acetate can be generated enzymatically. **b)** A molecular purge valve which dissipates reducing power in order to convert pyruvate to polyhydroxybutyrate (PHB) in a redox balanced way. **c)** A chimeric glycolysis system which converts glucose to lactate in a redox- and ATP-balanced route. It uses a non-phosphorylating GapN to maintain ATP balance. The corresponding route through standard Embden-Meyerhof-Parnas (EMP) glycolysis is shown with blue enzyme labels. **d)** Glucose to isoprene pathway which uses NADPH-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and pyruvate dehydrogenase (PDH). An NADPH drain is required to maintain redox balance. This pathway is also ATP-balanced. G6P inhibition is also considered in this system.

3.2.1 Methanol Condensation Cycle (MCC)

Methanol condensation cycle (MCC) (Fig. 3-2a) is a metabolic pathway to convert methanol to higher alcohols with 100% theoretical carbon yield, in contrast to natural pathways like ribulose monophosphate (RuMP) which have a maximum of 67% theoretical carbon yield due to the decarboxylation of pyruvate [50]. The core of the pathway creates a C-C bond between two formaldehyde molecules derived from methanol for the generation of acetyl-phosphate, which can be enzymatically converted to acetate or ethanol.

In this cycle, formaldehyde is incorporated into ribulose-5-phosphate (Ru5P) (Fig. 3-2a) to generate hexulose-6-phosphate (H6P) by hexulose phosphate synthase (Hps). H6P is then isomerized to fructose-6-phosphate (F6P) which can be cleaved by phosphoketolase. Erythrulose-4-phosphate (E4P) and F6P can then recombine via transaldolase, transketolase and isomerases (Tal, Tkt, Rpe, Rpi) to regenerate Ru5P. Alternately, xylulose-5-phosphate (X5P) can be cleaved by phosphoketolase, yielding G3P and acetyl-phosphate. G3P is then shuffled with F6P by transketolase to generate E4P and X5P, which can proceed to regenerate Ru5P via Tal, Tkt, Rpe, and Rpi. The X5P- and F6P-cleaving activities of phosphoketolase are referred to as Xpk and Fpk, respectively, and the pathway is investigated with different combinations of these activities.

3.2.2 Pyruvate to poly-hydroxybutyrate

A molecular purge valve for the production of polyhydroxybutyrate from pyruvate *in vitro* was demonstrated by Opgenorth *et al* (Fig. 3-2b) [15]. This system needs special attention to achieve redox balance, since pyruvate has a more reduced oxidation state than the product. To alleviate this cofactor imbalance, a method for dissipating excess reducing equivalents, termed a molecular purge valve, was designed for the conversion of pyruvate to downstream products like isoprene and poly(hydroxybutyrate) (PHB). Two different pyruvate dehydrogenases (PDH) were used in the system—one with cofactor specificity for NADPH and one with specificity for NADH. The

downstream pathway enzymes use NADPH to reduce metabolites and an NADH oxidase (NoxE) to dissipate the generated NADH. From two acetyl-CoA molecules, two enzymes are required to generate the final product PHB.

3.2.3 A chimeric ATP-balanced glycolysis system

A chimeric glycolysis system was demonstrated by Ye *et al* [14] (Fig. 3-2c). Canonical Embden-Meyerhof-Parnas (EMP) glycolysis generates a net of two ATP per glucose. In the chimeric system, a non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN) was used. This results in a system which is ATP balanced, making it more convenient for *in vitro* assays. Additionally, the system is NADH balanced since the final product was lactate, which has the same redox state as glucose.

3.2.4 Glucose to isoprene system

A system is considered for the conversion of glucose to isoprene (Fig. 3-2d). NADPH-dependent glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and NADPH-dependent pyruvate dehydrogenase are used in the pathway. NADPH is used since the downstream reactions in isoprene synthesis use NADPH. The pathway converts three glucose to two isoprene molecules. Interestingly, this pathway is also ATP-balanced, with the ATP generated by the glycolytic pathway being used stoichiometrically downstream in the isoprene pathway reactions. However, to maintain redox balance, NADPH must be drained from the system, potentially via an oxidase or similar enzyme. This system is investigated both with and without a substrate-level regulation of glucokinase (GK) by G6P, implemented using an irreversible version of modular rate laws [8] proposed by Liebermeister. The kinetic form used is known as competitive inhibition, though many other kinetic forms are plausible. Inhibition of this step by G6P is well-known. For example,

a human enzyme catalyzing this reaction is G6P-inhibited [51]. These equations show the effective

kinetic forms of glucokinase used without and with regulation:

$$No\ Regulation: V_{GK} = \frac{V_{max}}{\frac{K_{m,Glc}}{[Glc]} + \frac{K_{m,ATP}}{[ATP]} + \frac{K_{m,ATP}K_{m,Glc}}{[ATP][Glc]} + 1}$$

$$With\ G6P\ Inhibition\ V_{GK} = \frac{V_{max}}{\frac{K_{m,Glc}}{[Glc]} + \frac{K_{m,ATP}}{[ATP]} + \frac{K_{m,ATP}K_{m,Glc}}{[ATP][Glc]}\left(1 + \frac{[G6P]}{K_{i,eff,ATP}}\right) + 1}$$

3.3 Network intrinsic stability & bifurcational robustness upon perturbation
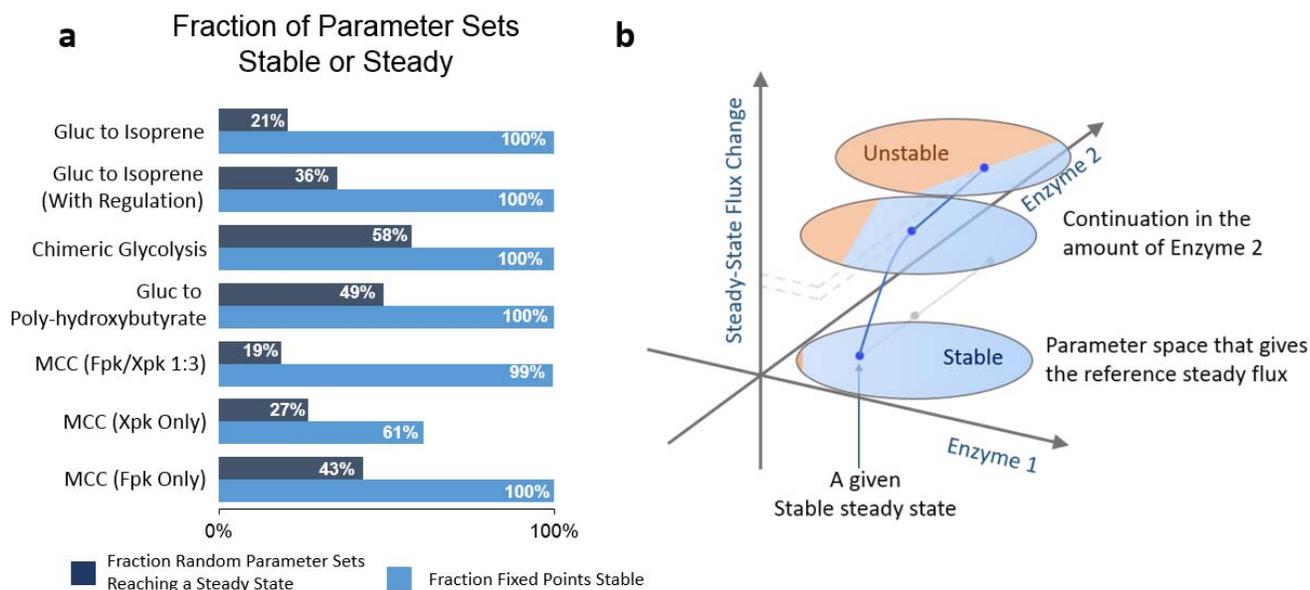
Fig. 3-3) Characterizing intrinsic stability of different pathway systems. **a)** For different pathways, two measures of intrinsic stability are presented. First, in dark blue, is the fraction of unconstrained, random parameter sets which reach a productive steady state. Second, in light blue, is the fraction of EMRA-determined parameter sets constrained to a steady state which are also stable. The intrinsic stability of pathways differs greatly between pathways, and also depending on which measure is used. Thus, a rational method of pathway balancing would be useful. (SD < 2% for all systems, n = 3 x 1000 parameter sets). Since phoshoketolase has two activities, cleaving either F6P (called Fpk) or X5P (called Xpk), we investigated used a ratio of Fpk/Xpk activites, 1:3. **b)** A representation of how steady state is not always stable. After perturbation from a constrained steady state, the fraction of parameter sets which retain stability tends to decrease, and steady state flux may change. Eventually, a parameter set may become unstable after perturbation.

We used the four systems described in Fig. 3-2 to examine the stability problem. In

particular, we investigated how network structure affects the intrinsic possibility of reaching

stability. Previous EMRA work starts from an ensemble of parameter sets that give the same reference steady state, and discards the parameter sets that generate a Jacobian matrix with a real part of an Eigenvalue greater than zero, which indicates instability.

However, experimental systems are not guaranteed to be stable or reach a steady state. To place stability and steady state in a context which is more meaningful to experimental efforts, enzyme parameters were chosen completely at random, and the systems were then integrated in time domain to determine if a productive steady state was reached (Fig. 3-3a, dark blue bars). This method is more representative of experimental efforts which often have either little or indirect control over enzyme amount or activity (*in vivo*), or don't have rational methods for pathway balancing (both *in vivo* and *in vitro*). Interestingly, the results show that pathways have very different likelihoods of resulting in a steady state (Fig. 3-3a, dark blue). The glucose to isoprene system had only 21% of randomly generated parameter sets reaching a non-trivial (non-zero) steady state. This could be because it is a relatively large system in terms of enzyme number and uses two different cofactors (NADPH and ATP). A large system may be less likely to reach a steady state. If each enzyme has an acceptable range of values, then in a large system it is more likely that at least one of these values would be outside the range, resulting in system instability. However, when regulation of glucokinase was introduced via activation by ATP and inhibition by ADP, the likelihood of productivity jumped to 36% (Fig. 3-3a). Overall, these results show that intrinsic pathway structure and kinetic forms (including regulation) have a strong influence on possibility of reaching a productive steady state. The result is a varying, and sometimes low, likelihood of achieving stability and productivity. Thus, finding rational ways to balance pathways is an important goal which can improve and accelerate the pathway development process.

If the enzyme parameters were first constrained to a fixed point by solving for parameter values which give reaction rates equal to the reference flux, then the probability of attaining stability is greatly increased. While this is not practical in experiments, the method proves useful in model construction. We found that (Fig. 3-3a, light blue bars) the fraction of fixed points which were stable varied depending on the network structure. While most systems showed at least 99% of the parameter sets sampled to be at a stable steady state, the MCC (Xpk-only) system showed only 61% of parameter sets to be stable. Although for some systems the fraction of stable steady states is similar—5 systems which all show at least 99% stability by this measure—they have varying tendency to lose stability upon perturbation (Fig. 3-3b). Starting from the reference state, where parameter sets are chosen under the fixed point constraint, the region of instability could grow when enzyme parameter changes (Fig. 3-3b). Depending on the structure of the system, the instability region might grow in a different fashion upon perturbation, and eventually some might lose stability. This shows that stability of fixed points in metabolic systems is not guaranteed and that stability could be a critical factor in metabolic systems.

3.4 Stability of continuous systems can inform results of batch systems
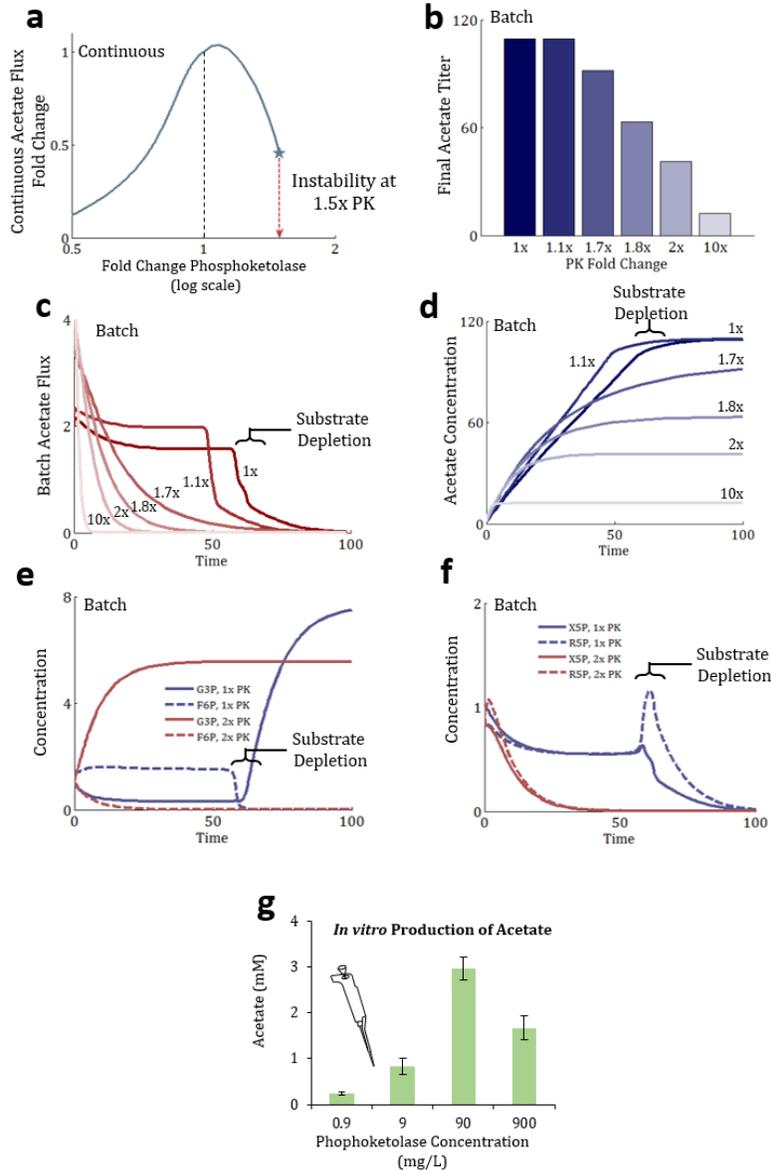
Fig.3-4) Investigating the instability in the MCC pathway using Fpk/Xpk ratio as 1:3. **a)** In a continuous system, an arbitrary parameter set determined by EMRA is perturbed up and down with respect to phosphoketolase, on the X-axis. On the Y-axis, the continuous, steady state, acetate flux is plotted. As phosphoketolase increases, the system bifurcates at ~1.5x increase. **b)** Time domain simulation is performed, at different amounts phosphoketolase (PK). The final titer for each condition is plotted. The production gradient appears gradual, but is the result of a sudden instability. **c)** The production rate for acetate is shown for each phosphoketolase amount over time. In the stable condition (1x), production rate reaches a constant, implying the system enters a "pseudo-steady state", until substrate depletes. In the other conditions, production rate is never steady, but decreases over time. **d)** The amount of acetate is plotted over time. It is observed that as the amount of phosphoketolase increases beyond bifurcation, the production decreases. **e)** At the 1x and 2x conditions, the concentrations of G3P and F6P are plotted. In the 1x condition, F6P is maintained at a nonzero-level throughout production, while in the 2x condition, it is quickly depleted and G3P accumulates. **f)** The R5P & X5P levels are plotted with time in the 1x and 2x conditions. **g)** Data from Bogorad *et al* [50] shows that as phosphoketolase level increased, the amount of acetate produced by the cycle decreased, supporting a link between instability in a continuous system and production in an analogous batch system. An icon shows this data is experimental.

EMRA uses continuous production models to simulate enzymatic systems. However, many experiments, including *in vivo* and *in vitro*, are conducted as batch processes. Thus, it's not clear how a perturbation which causes bifurcation in a continuous system will inform the batch experiment. Where a pseudo-steady state may exist, the pseudo-steady state behavior can be predicted by the continuous model. In these systems, if a parameter set resides in a domain where no stable steady state exists in the continuous mode, then no stable pseudo-steady state exists in the batch mode. This can be justified by locally linearizing the input function to convert a pseudo-steady batch system to a continuous system. However, an experimentalist measuring only the product output at the end point would not detect the lack of stability. In this case, the product yield will gradually decrease even when the system has entered an instability region.

To show how the existence of a continuous bifurcation could manifest itself in a batch system, we simulated a batch system in time domain. First, stable parameter sets were generated via EMRA in a continuous MCC system using Fpk/Xpk ratio as 1:3. Then, the parameter sets were integrated using the continuation method to increase the phosphoketolase level until instability occurs, increasing V_{max} for Fpk & Xpk at the same ratio. A representative parameter set is plotted in Fig. 4a to show the effect of increased phosphoketolase on continuous steady state acetate flux up to the point of instability. As phosphoketolase increases, the flux towards product increases slightly before decreasing and finally becoming unstable.

This parameter set was found to become unstable at a ~1.5-fold increase of phosphoketolase. Different amounts of phosphoketolase perturbation (1x, 1.1x, 1.7x, 1.8x, 2x, 10x – multiplier applied to both Fpk & Xpk V_{max} values) were chosen to show the dynamic response of the system in a batch simulation. All rate equations, parameter values and initial conditions were kept the same as in the continuously model (i.e. all starting metabolite

concentrations were normalized to unity), except that starting formaldehyde concentration was multiplied by 200 and the "in" and "out" reactions used in the continuous mode were eliminated to observe product accumulation in time domain simulation of the batch system. 200-fold increase in initial formaldehyde concentration was chosen arbitrarily to signify a batch reaction, in which the starting substrate was included as a single charge instead of being fed over time. It was found to adequately demonstrate the phenomena we were interested in, though other values could have worked as well. See Tables 1&2 in the method section

Interestingly, the final batch production observed for this system decreases gradually as phosphoketolase (PK) amount (Fig. 3-4b) increases. In the continuous system, the underlying phenomenon is instability, a step change in the nature of the steady state. In the corresponding batch system, pseudo-steady state disappears because of instability. However, the product formation does not stop until key intermediates are depleted. Batch acetate production rate over time is plotted in Fig 4c. For the stable 1x and 1.1x conditions, a pseudo-steady state was achieved in which acetate production rate reached a constant level, only decreasing when the formaldehyde had been consumed. However, for the conditions which are past the instability point (1.7x – 10x), a steady rate of acetate production is never achieved. Instead, the rate decreases monotonically until it reaches zero. The productivity of the 10x condition falls the fastest, eventually resulting in the lowest production. This shows that a decrease in production, even gradually, in a batch system could be associated with an instability issue in an analogous continuous system. In Fig. 3-4d, the acetate concentration over time is plotted to show how the system evolves over time.

For systems that are stable, because the initial concentration of the starting substrate is much higher than the K_m value of the uptake system, the rate of input holds largely constant until the substrate concentration approaches the Km value. During this time, the system is operating

under a pseudo-steady state similar to a stable continuous system. This is seen in Fig, 4 for 1x and 1.1x phosphoketolase concentrations. Thus, the property of continuous system simulation carries over, until substrate concentration approaches K_m. Thus, the system is run almost the same as in a continuous system in the first 50 min time units or so (Fig, 3-4c), when most of the acetate is produced (Fig 3-4d)

For systems that are unstable (Fig. 3-4, 1.7x, 1.8x, 2x, 10x phosphoketolase concentrations), the output flux was not able to reach a steady-state (Fig. 3-4c), and it decreases rapidly from the start and approaches zero despite the presence of the initial substrate. The cumulative product formed (acetate) is the integral of flux over time (Fig. 3-4d), which decreases as the system moves further away from the bifurcation point.

Additionally, we investigated the mechanism by which the bifurcation causes decreased production. In the 1x condition, F6P is maintained at a nonzero-level throughout production, while in the 2x condition, it is quickly depleted (Fig. 3-4e). R5P and X5P are also shown to deplete quickly in the 2x condition (Fig. 3-4f). Thus, it is the depletion of these cycle intermediates which causes cycle failure. A previous experimental effort (Bogorad et al. [6], data reproduced in Fig. 3-4g) showed that in *in vitro* enzymatic experiments, the batch production of acetate with from formaldehyde reached a local maximum with respect to phosphoketolase amount, supporting the EMRA analysis. In sum, EMRA could potentially have useful insights into experimental systems, by identifying enzymes which may be most sensitive to bifurcation, and how they affect the system in question.

3.5 Systems Analysis

3.5.1 EMRA predicts behavior of a molecular purge valve
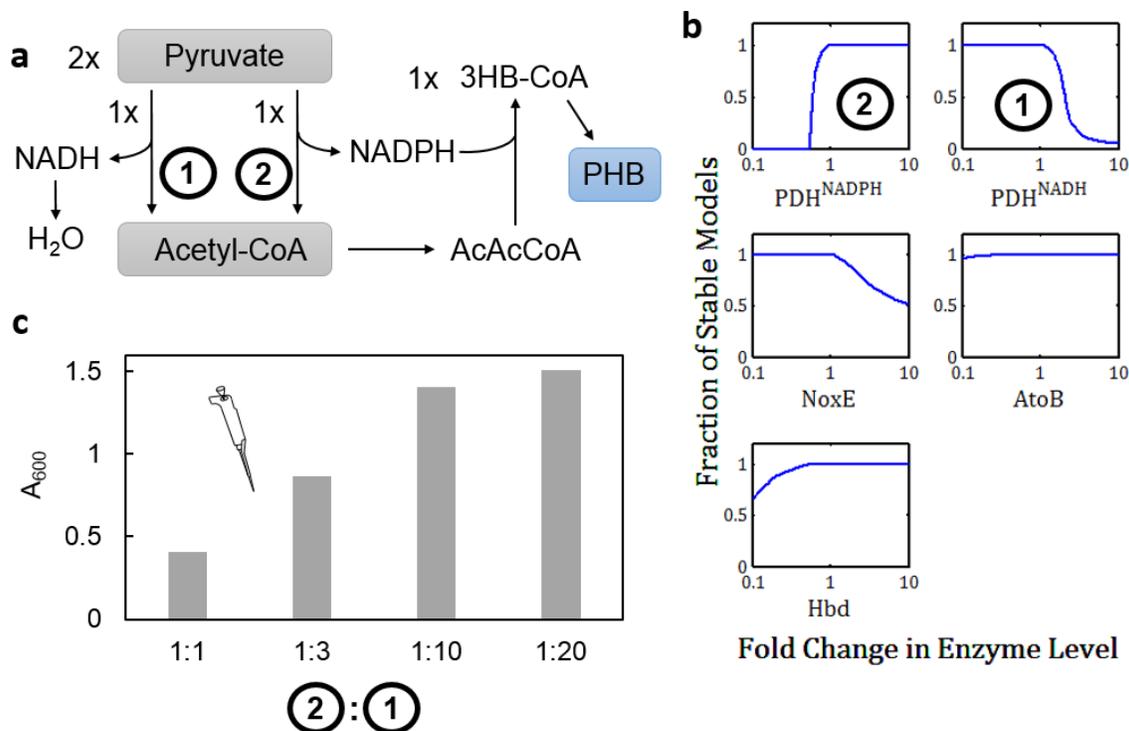
Fig. 3-5) Stability of a biosynthetic purge valve for production of isoprenoids by dissipation of reducing equivalents. **a)** Pathway schematic showing cofactor requirements. **b)** Stability profiles predicted by EMRA (n = 1000) as enzyme amounts vary. It is shown that to maintain stability, high levels of PDH^{NADP} and low levels of PDH^{NAD} are required. **c)** Data from ([14], Figure 4) which shows that a high ratio of PDH^{NADP}: PDH^{NAD} is required for optimal performance of the pathway. Image analysis of line graph figure from reference yielded numerical data to generate the bar graph shown. (1) indicates NADH-dependent PDH and (2) indicates NADPH-dependent PDH. An icon shows this data is experimental.

A purge valve system converting pyruvate to PHB was analyzed (Fig 3-5a). Each enzyme is represented by a canonical Michaelis-Menten kinetic rate law, and the reference flux is fixed since there are no degrees of freedom. EMRA methodology was implemented in this system to show the effects of perturbation of each enzyme. High NADPH-dependent PDH (PDH^{NADPH}) (Fig 3-5b) and low NADH-dependent PDH (PDH^{NADH}) resulted in the most stability for the pathway. PDH^{NADH} must be low to prevent too much pyruvate from taking this route which generates unusable NADH reducing power, while PDH^{NADPH} must be high to ensure that enough NADPH is generated to allow for 100% yield from acetyl-CoA to PHB. The imbalance of these activities may cause system instability, according to EMRA. Indeed, the PHB pathway was experimentally

31

demonstrated to have reduced production with a lower ratio of PDH[NADPH]:PDH[NADH] (Opgenorth,

et al [15])(Fig. 3-5c), matching the results of EMRA.

3.5.2 EMRA predicts chimeric glycolysis system's sensitivity to glucose feed rate
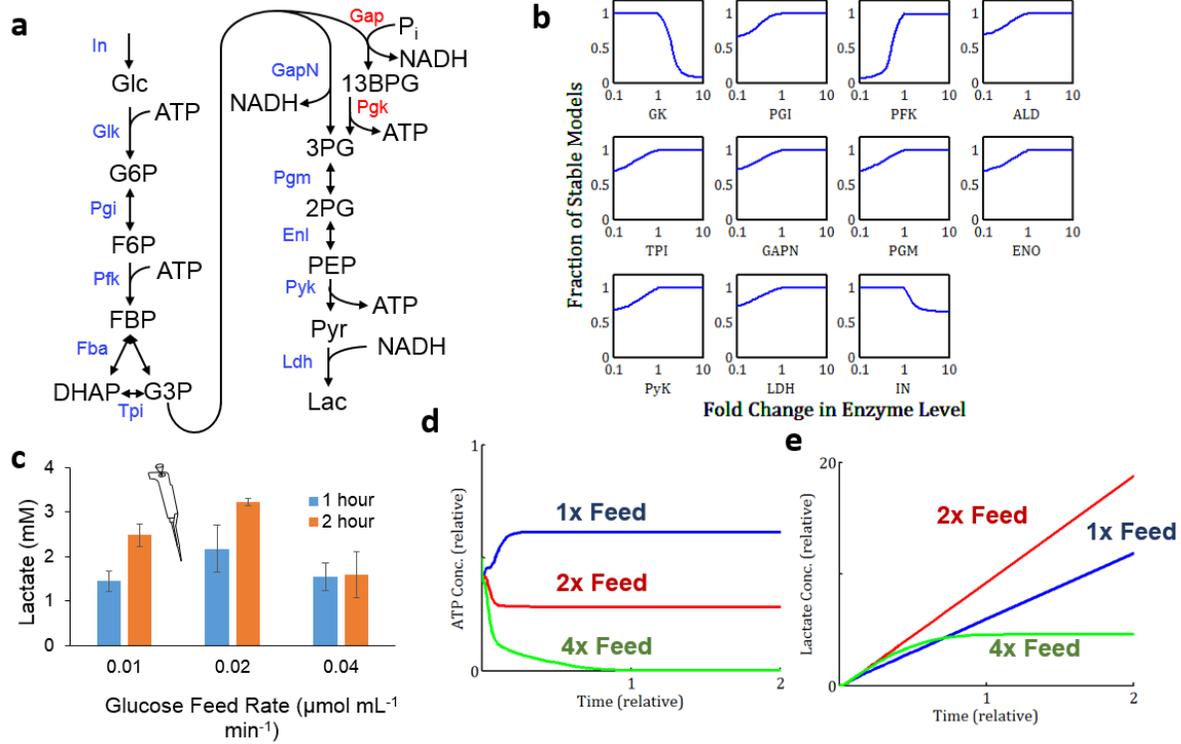
Fig. 3-6) The ATP-balanced synthetic chimeric glycolysis pathway from glucose to lactate (Ye et al[14]). **a)** Pathway schematic contrasting cofactor production between standard Canonical Embden-Meyerhof-Parnas (EMP) glycolysis (Gap & Pgk, red lettering) with the chimeric non-phosphorylating GapN system. **b)** EMRA stability profiles (n = 1000) as enzyme amounts and glucose feed rate (IN) vary. Glucose feed rate is shown to produce moderate instability at higher levels. **c)** Data from (Ye et al, Figure 6A[14]) which shows that increased glucose feed rate can cause lower production. An icon shows this data is experimental. **d-e)** Simulation of fed-batch production of a sample parameter set for the chimeric glycolysis system. Numerical integration of time domain behavior shows instability at higher feed rates caused by ATP depletion and resulting in lower overall lactate production. Priming intermediates are fed in the same proportion as the experimental condition, and feed rates are also demonstrated in the same proportion (1, 2, 4). **d)** ATP concentration over time at the three different glucose feed rates. **e)** Lactate production over time at three different glucose feed rates.

Another example of EMRA application is in a thermotolerant, cell-free glycolysis system

which was demonstrated for the production of lactate from glucose by Ye *et al* (Fig 3-6a) [14].

Canonical Embden-Meyerhof-Parnas (EMP) glycolysis generates two net ATP per glucose, (Fig.

3-6a, Gap & Pgk enzymes). However, in the chimeric system, to prevent cofactor imbalance, a

non-phosphorylating glyceraldehade-3-phosphate dehydrogenase (GapN)[52] was used—resulting

in a net balance of ATP and NADH from glucose to lactate. Again, the system was modeled using EMRA methodology. The results show that increased glucokinase (GK) amount and glucose feed rate (Fig. 3-6b, GK, IN) may cause instability.

This system was experimentally tested for lactate production at different glucose feed rates (Fig. 3-6c, adapted from Ye *et al* [14]). It was found that beyond a certain point, increasing glucose feed rate reduced lactate production, even if the same total amount of glucose had been fed, matching the instability to feed rate (IN) predicted by EMRA. The instability apparently occurs by the depletion of ATP by glucokinase. ATP is required for both glucokinase and phosphofructokinase (PFK). However, if glucose is fed too quickly, ATP may become depleted by glucokinase before it can be regenerated in lower glycolysis. A time domain simulation of this system was carried out using initial conditions similar to the experimental conditions reported [14] (Fig. 3-6d, e) and a parameter set from EMRA which became unstable after increase of feed rate. The time domain simulation showed that at reference feed rate (1x), ATP level is maintained and lactate production continues. However, at 4x feed rate, the ATP is depleted and the lactate production stops.

EMRA was also carried out on canonical EMP glycolysis converting glucose to lactate and similar instabilities were found (Not shown). In both systems, reduction in PFK activity was shown to strongly increase chance of instability. This is because once a metabolite is past PFK, it may be used to regenerate ATP, so it's important to ensure that the flux past PFK is sufficient to supply ATP for all of upper glycolysis. However, it's more paradoxical that an increase in an enzyme would cause productivity and instability issues, particularly glucokinase, or even feed rate.

3.5.3 EMRA predicts unstable enzymes in uncharacterized system from glucose to isoprene
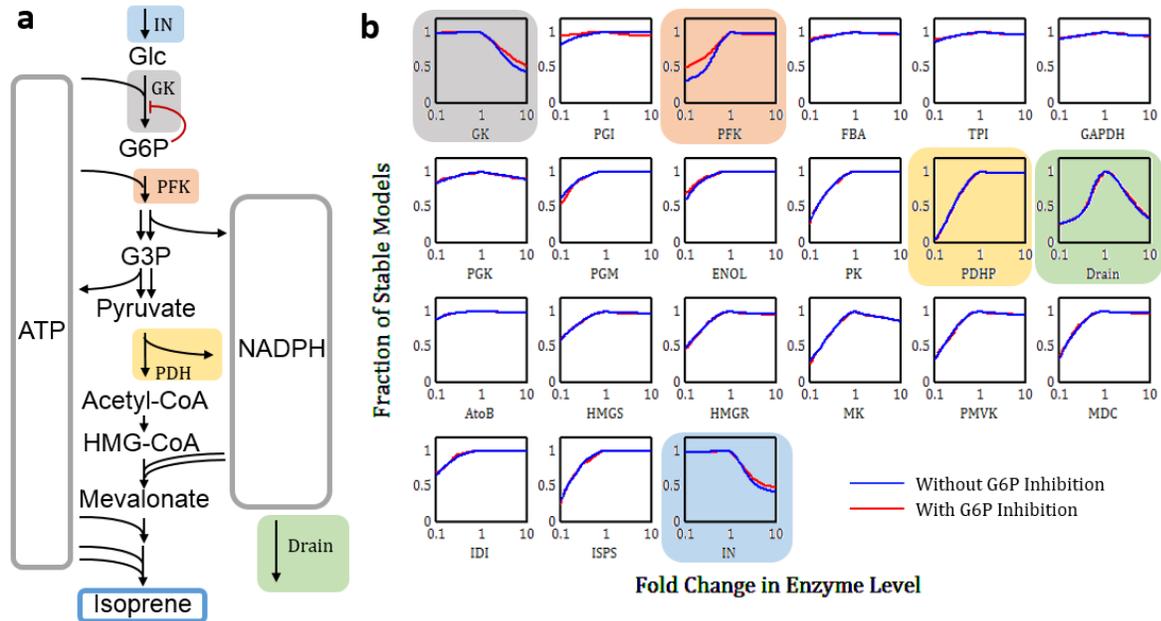
Fig. 3-7) An NADPH-dependent pathway from glucose to isoprene. **a)** A pathway schematic showing an outline of the enzymatic reactions, cofactor flow, and regulation added (glucokinase). **b)** EMRA profiles for all (n = 1000) enzymes in the unregulated isoprene pathway in blue. EMRA profiles for all (n = 1000) enzymes in the GK-regulated isoprene pathway in red. Select enzymes are highlighted to show their position in the pathway. The enzymes dealing with ATP cycling are most changed by the presence of regulation.

To demonstrate the utility of EMRA in identifying potential points of instability, a not yet characterized pathway producing isoprene from glucose was investigated with EMRA. The pathway is ATP-balanced and maintains redox balance using an NADPH drain. EMRA identifies that the NADPH drain must be balanced, not too low or too high. GK & IN must not be too high, while all other enzymes must only not be too low (Fig. 3-7b). By introducing regulation of GK using modular rate laws [8], the fraction of productive steady states increased (Fig 3-3a). The stability to perturbation is also slightly improved for feed rate (IN), GK and PFK (Fig. 3-7b).

Interestingly, the NADPH drain is unstable to both decrease and increase. This could be because if the rate is too low, then $NADP^+$ is not sufficiently available for GAPDH, and lower glycolysis is unable to regenerate ATP needed for earlier in glycolysis and later in the pathway – while if the rate is too high, NADPH will not be available for the biosynthetic steps of the isoprene pathway. This analysis shows that a longer pathway has many complex, interacting factors that

34

can cause instability and that EMRA is able to identify some of these potential issues. Similar to the previously simulated glycolysis system, glucokinase, (GK) and phosphofructokinase (PFK) and feed rate (IN) showed some instability. Other enzymes show less sensitivity upon increase (Fig. 3-7b). Instability caused by decreasing enzyme concentration is common and seen in most if not all pathways.

A rational experimental plan for this pathway would thus focus on having sufficiently high levels of most enzymes (all except feed, glucokinase and drain), for example, by ensuring the total activity of each enzyme is significantly higher than the feed rate. Ensuring these enzymes are at a high level would ensure both stability, according to EMRA results, and the possibility of maximum productivity. For enzymes which become unstable at higher levels, more optimization is required. Levels of glucokinase, NADPH drain, and feed rate should be varied in order to avoid instability and to find the highest productivity condition. This significantly narrows the focus from 21 variables to just 3.

3.6 Discussion & Conclusions

The results show that EMRA has potential to be a valuable tool for investigating the propensity for stability of complex enzymatic pathways without *a priori* knowledge of specific enzyme parameter values. In three cases presented, (MCC, molecular purge valve, chimeric glycolysis) the experimental investigators were able to heuristically identify productivity issues *ad hoc*, but EMRA is able to unify all these results with a theoretical framework based on instability. Importantly, although some of the phenomena were experimentally determined, it was not necessarily known that instability of the system—causing a step change in the nature of the steady state, rather than a smooth change predictable by sensitivity analysis—could be an underlying reason.

The success of the method with these systems presented here shows that it deserves consideration as a design tool in the invention of new pathways. The method has proven versatile enough to successfully predict features in three different pathways investigated in different laboratories and powerful enough to do so without *a priori* knowledge of specific enzyme parameter values. EMRA simulation of a longer and not-yet characterized pathway demonstrates the range of possibilities for potential applications of this technique. While the characterized pathways were optimized based on intuition, it's possible that a longer pathway with more enzymes, such as the glucose-to-isoprene pathway, would be much more difficult to optimize without rational balancing methods like those presented here. The reduction of search space from 22 to 3 variables represents an exponentially more approachable experimental path towards productivity, resulting in 27 (3^3) experiments rather than about 10 million (3^{21}) if three different enzyme amounts are tested.

Another insight provided by EMRA and follow-up analysis is the determination of failure modes for the pathways investigated. Using parameter and enzyme amount values in stable and unstable regions of the parameter continuation, time domain integration allows us to determine the failure modes for these pathways upon instability. In the MCC pathway, it is depletion of pathway intermediates—especially X5P, R5P and F6P—which causes productivity decline and eventual stopping. Although time domain simulations weren't carried out in all systems, the demonstration of failure mechanism in the MCC system may lend credence to the other EMRA examples. In the chimeric glycolysis pathway, depletion of ATP eventually caused that pathway to stop when glucose feed rate was too high. Identifying these failure modes with EMRA is another potentially fruitful area of discovery.

Glycolysis is a fundamental pathway of life and functions successfully in many organisms. However, our simulations and previous experiments (Fig. 3-6, Ye *et al*, [14]) have shown it can be unstable under high glucose feed conditions which apparently deplete ATP and accumulate hexose-monophosphates (Fig. 3-6d, e). Some hexokinase enzymes are product-inhibited by G6P, [53] however, the particular enzyme used in the experimental investigation (from *Thermus thermophilus*) was investigated and no G6P-inhibition was reported.[54] Interestingly, glycolysis has also been shown to be unstable to low levels of inorganic phosphate in yeast, a condition which prevents GAPDH from proceeding [55]. Glycolysis is a nearly universal pathway, but this evidence shows it to be unstable in some cases. This helps to explain the presence of elaborate regulations such as insulin and glucagon [56,57] in animals and the massively sophisticated regulation of phosphofructokinase [58,59] in many organisms. Rather than stability, alternate explanations such as chemical necessity [60] and thermodynamic efficiency [61,62] are more likely reasons for the universality of glycolysis.

In these analyses, EMRA was used to successfully evaluate the stability of complex *cell-free* pathway assays. *In vitro* biocatalysis systems are a powerful alternative and complement to *in vivo* systems [63]. Importantly, however, this does not exclude the possibility of success with simulation of *in vivo* systems. Depending on growth mode (exponential growth, stationary phase, fermentation etc.) *in vivo* systems may have different reference fluxes, so more exploration is required to identify different possibilities.

It is unsurprising that lower amounts of pathway enzymes or feed rate would hinder productivity. The powerful insight provided by these results is that for the pathways identified, increasing levels of certain enzymes or feed rates were shown to cause instability and consequently reduce production. A typical metabolic engineering approach may be to simply maximize the

reaction rate of all pathway enzymes. However, we show here that for many enzymes, this will not always result in an optimal outcome.

Additionally, we have shown that the intrinsic stability of pathways varies significantly depending on structure and kinetic forms. This highlights the importance of stability analysis in understanding metabolic systems. Additionally, it shows that many metabolic systems may be very difficult to balance without sufficient rational methods for analyzing which enzymes are most likely to contribute to pathway instability, and in which amounts. This shows the importance of EMRA and stability analysis in general in understanding pathways theoretically and exploiting them practically.

The lack of requirement for *a priori* knowledge of specific enzyme parameter values could make EMRA particularly approachable for experimental researchers working with new pathways or unknown enzymes. This may be hampered somewhat by the need for sophisticated mathematical operations, though this obstacle could be overcome if an appropriate software suite is made available. We believe EMRA can significantly contribute to pathway development efforts and is an important contribution to the toolbox of metabolic engineering.

3.7 Additional Example: Glycerol-to-alcohol (GtA) pathway

This section did not appear in the manuscript: "Stability Predicts Productivity of Enzymatic Systems", but it serves as an additional example of EMRA predictions proving to be accurate. An application of the previously demonstrated non-oxidative glycolysis (NOG) pathway is the GtA pathway which converts glycerol to fuel alcohols with 100% theoretical carbon yield (Fig 3-8A).[64] GtA uses a unique fructose-6-phosphate aldolase (Fsa) to combine glyceraldehyde-3-phosphate (G3P) with the non-phosphorylated triose dihydroxyacetone (DHA) to form fructose-6-phosphate

(F6P).[65] NOG alone, like MCC, was shown by EMRA analysis to be non-robust with respect to phosphoketolase.[13]

GtA is not redox balanced and requires the input of reducing equivalents to fully utilize the glycerol feedstock. This was overcome by supplementing formate dehydrogenase (Fdh) and formate to the reaction mixture (Fig. 3-8A) However, EMRA shows that in this system, Fdh is the enzyme associated with greatest non-robustness (Fig. 3-8B), whereas non-robustness associated with phosphoketolase (Fpk & Xpk) is minimal, in contrast to the NOG and MCC systems which also include phosphoketolase. This is because at high levels of Fdh, NADH is produced very quickly, preventing the forward operation of glycerol dehydrogenase. Subsequently, the carbon flow to the main portion of the cycle and the reduction to final alcohol product is cut off. This non-robustness of Fdh production was confirmed by *in vitro* experiment (Fig. 3-8C). In the experiment, ethanol production showed a local maximum with respect to Fdh amount [Experiments conducted by Tony Wu], confirming the EMRA prediction.
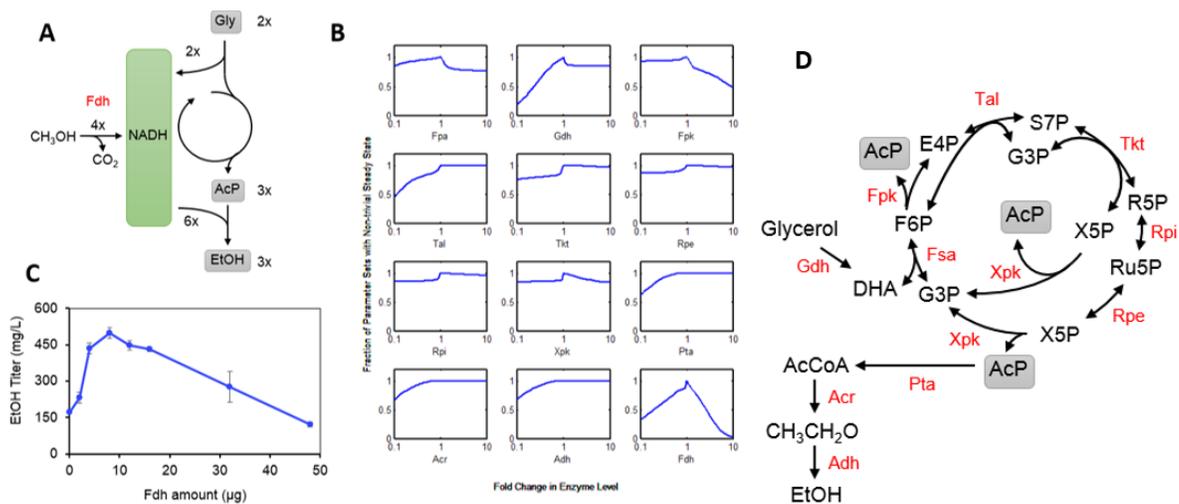
Fig 3-8). The synthetic glycerol-to-alcohol (GtA) pathway. **A)** Pathway schematic showing the requirement for additional reducing equivalents. **B)** EMRA robustness profiles predicted by EMRA (n = 1000) as enzyme amounts vary. Fdh is the shown to have a local maximum for stability. **C)** Data from unpublished *in vitro* experiments [Tony Wu] shows that Fdh has a local maximum for ethanol productivity. **D)** Full GtA pathway.

4. *In vivo* EMRA applications

4.1 Using Kinetically Accessible Yield (KAY) and EMRA for *in vivo* applications

Although constructing models of in vitro systems is much simpler, construction of in vivo models using EMRA is nonetheless feasible and desirable. Because of the complexity of *in vivo* systems, it is often necessary to make choices about the scope of the model—i.e. how much of the overall metabolism to include. These choices are often related to the overall purpose for the model, usually the production of some compound of interest.

Additionally, it is necessary to determine a reference flux which is appropriate for the organism and condition to be modeled. For instance, the reference flux for a bacteria in growth phase would be different for that in stationary phase, etc. One method to provide information about the reference flux, used extensively in this chapter, is to use information about the extracellular metabolites produced under the condition of interest, and use that to constrain the flux. This can be accomplished in combination with diverting metabolites to a reaction signifying the biomass accumulation, important in growth phase simulations. In such cases, it is necessary to define a realistic biomass objective function.

This chapter demonstrates three instances in which EMRA is applied to in vivo systems, bringing together modelling, data from literature reports, metabolomics data from collaborators, all three of which examples were conducted in collaboration with other institutions. These examples demonstrate the flexibility and power of the EMRA method, and provide actionable insights for experimental researchers.

4.1.1 Conceptual idea of kinetically accessible yield

Space of chemical transformations
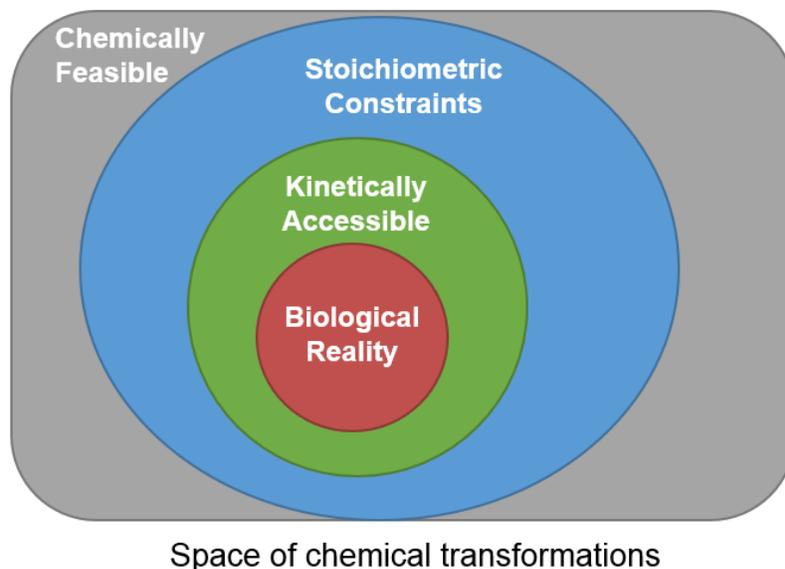
Fig. 4-1) Classifying the chemical transformation space. Chemical transformations are bounded by several factors. In the most permissive formulation, any chemical transformation is allowed—bounded only by an atom balance. This defines the "chemically feasible range". In a given biochemical system, only certain transformations, governed by the stoichiometry of the system, are allowed. This defines the stoichiometric range. Another constraint, within the stoichiometric constraints, is the kinetically accessible region. This kinetically accessible region governs the stability of the system. Imposition of growth requirements will further limit the allowed set of chemical transformations, since some amount of flux must be directed to biomass.

When considering the conversion yield of a raw material to a product, several factors come into play. First, the balance of atoms, which gives the largest feasible yield range (Fig. 4-1). The atomic balance does not involve energetics or kinetics, thus this yield involves the minimum constraint and represents the highest bound. When enzymatic pathways are chosen for the conversion, the stoichiometry of the pathway imposes an additional constraint, within the atomic balance, and gives the maximum theoretical yield allowable by the pathway. At this level, energetics plays an implicit role through the choice of the pathway, but kinetic effects are not considered.

4.1.2 Generation of Toy Model

A simplified model of microbial metabolism with butanol synthesis[66] was generated to illustrate the concept of KAY. This pathway for 1-butanol synthesis passes through the keto acid synthesis pathway,[67] and differs from the common 1-butanol pathway that proceeds by coenzyme-

A (CoA) dependent reverse beta-oxidation.[67] The model consists of 11 reactions (Fig. 4-2A), and is heavily abridged, where pathways of multiple reactions, such as glycolysis, have been modeled as a single reaction. Michaelis-Menten style rate laws with saturation features were chosen according to reversibility and number of substrates, and products if reversible. Parameters are single valued positive numbers. The model outputs up to six exometabolites and has no biomass generation. The model was assigned arbitrary parameters so that five outputs, other than butanol, had roughly equal orders of magnitude for their output flux.

4.1.3 Determination of kinetically accessible yield using flux integration

Here we examine the kinetic stability of the system when a flux is directed to generate a product. As a starting point for developing the concept, we consider a host system with complete kinetic parameters.

$$\frac{dX}{dt} = F(X, k) = Sv(X, k) = 0 \tag{4-1}$$

Flux v is a function of metabolite concentrations X, kinetic parameters k, and is multiplied by a stoichiometric matrix S, to provide an equation, F, describing a system of differential equations. Suppose we wish to incorporate information for a new pathway for which kinetics are unknown.

Although a new pathway may introduce new metabolites that may affect the kinetics of the host, for parsimony, we assume that the metabolites in the new pathway do not interact with the host system and only consider how the new pathway flux might affect the kinetics of the existing host metabolic system. Under this assumption, we can represent the new model as:

$$\frac{dX}{dt} = F(X, k, \varphi) = Sv(X, k) + S_{new}\varphi \tag{4-2}$$

where S_{new} is a vector representing the stoichiometry of the new pathway and φ is a scalar parameter which represents the amount of flux going through the pathway. S_{new} would include the metabolites drawn out of native metabolism into the heterologous pathway. Since, we are not considering the effect new metabolites may have on the host system, S_{new} does not include the stoichiometry of metabolites not present in the host system. By its definition, parameter φ is 0 at the reference state (host only).
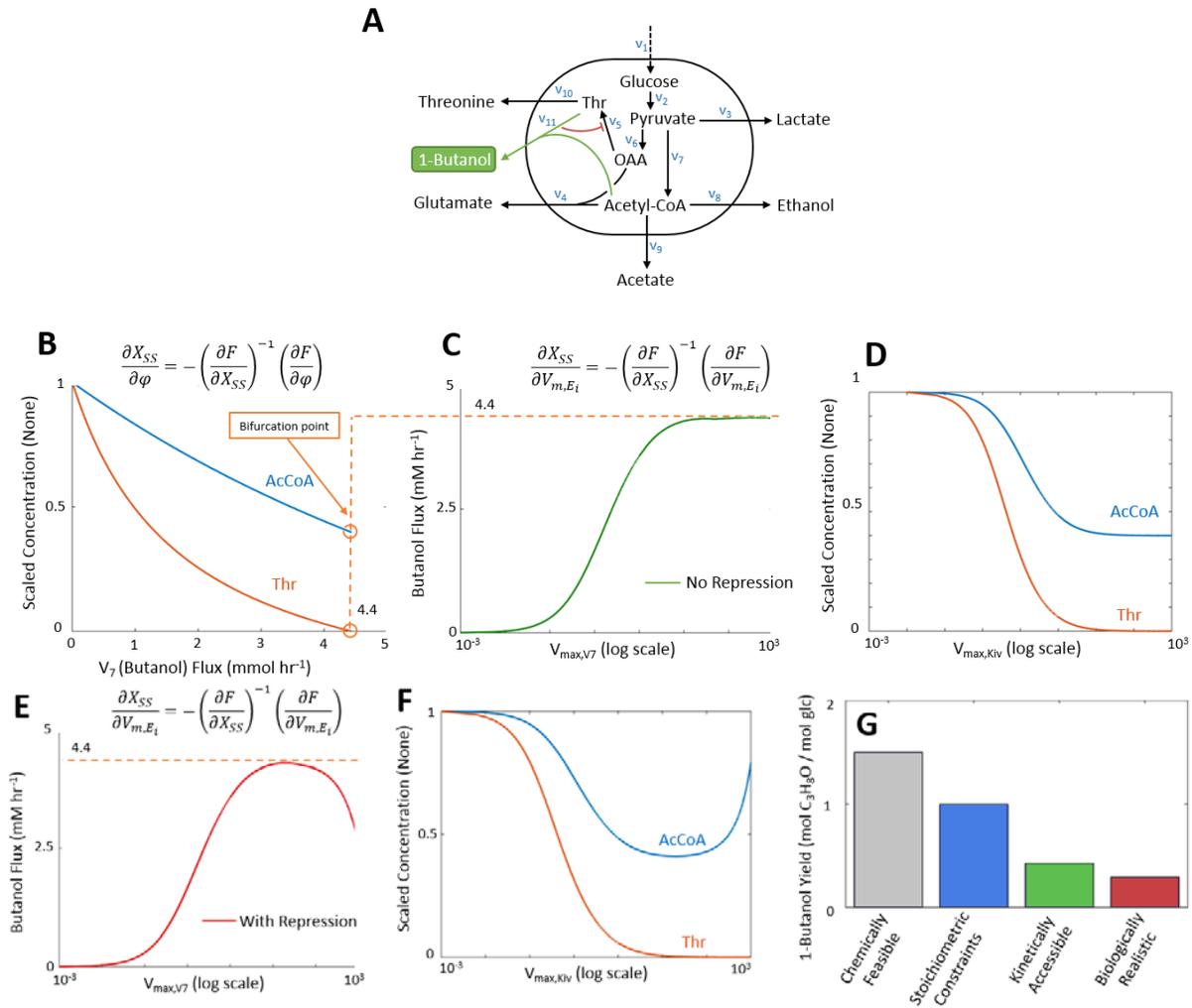
Fig. 4-2) **Overexpression of the butanol pathway in the toy model. A)** Diagram of toy model with butanol production reaction circled. **B)** The overexpression of the butanol pathway accomplished by integration with respect to reaction flux using Eqn (5). Acetyl-CoA and threonine concentrations are plotted—scaled to the maximum of each metabolite—with respect to pathway flux. **C)** Plot of butanol flux with respect to Vmax integration of the butanol pathway. The system does not bifurcate as in B), but the flux does reach a plateau as Vmax increases, corresponding to the same flux value observed in B). **D)** A plot of scaled threonine and acetyl-CoA concentrations with respect to V_{max} for the butanol output reaction. For C) and D), all other parameter values remained fixed (no repression). Threonine depletes, while acetyl-CoA, doesn't fully deplete. **E)** The overexpression of the butanol pathway accomplished by integration with respect to reaction V_{max}. This integration included a repression effect where the threonine synthesis Vmax was decreased as the butanol reaction was overexpressed. **F)** A plot of scaled threonine and acetyl-CoA

concentrations, for repression the condition. **G)** A plot showing how different measures of yield compare. Chemically feasible is limited by carbon atom balance, while stoichiometry is limited by the pathways present in the toy model. KAY value depicts the yield corresponding to maximum flux determined in B) & C), while biologically likely lies between the theoretical (stoichiometric) yield and the biologically realistic value, placing a lower ceiling on calculations of yield in a specific genetic background.

In the toy model (Fig 4-2A), one can investigate different methods of simulation, and their effects on predicted yield. For example one can use integration with flux as a parameter, as demonstrated in Eqn (5) using the continuation method shown in Eqn (3). In that case, the flux parameter (ϕ), was used to represent the 1-butanol production reaction, instead of the kinetic rate law [V11, Thr + AcCoA -> (butanol out)]. The concentrations of Thr and AcCoA are plotted vs. increasing 1-butanol flux (ϕ) in Fig 4-2B. Thr decreases to zero, while AcCoA decreases, but not to zero. As threonine reaches zero, the integration fails as the system bifurcates. Note the maximum flux achievable is 4.4 mmol hr^{-1}. The yield (product flux/substrate consumption flux) corresponding to this flux is defined as the Kinetically Accessible Yield (KAY). In this integration, we did not consider the kinetic rate law of the butanol reaction (V11), but integrated with respect to the butanol flux as described in Eqn (5).

4.1.4 Determine kinetically accessible yield using kinetic parameter integration

One can also incorporate kinetics of the perturbed pathway into the model:

$$\frac{dX}{dt} = F(X, k, \varphi) = Sv(X, k) + S_{new}V_{new}(X, k_{new}) \tag{4-3}$$

The kinetic form used to represent this reaction is shown in Table 1, Reaction V11. We can determine the steady state metabolite concentrations using the continuation as a function of V_{max} (K_{14} in Reaction V_{11}). Interestingly, the system is fully robust against increases in V_{max} (K_{14} in Reaction V_{11}) for the butanol reaction. The system does not encounter a bifurcation point, but the butanol flux reaches a plateau at the value corresponding to the KAY in the previous integration (Fig 4-2C). With the overexpression of the butanol reaction itself, V_{max} may be increased to an arbitrarily high value (Fig 4-2C), and the system remains stable, although flux through the pathway

asymptotically approaches a maximum. During this integration, it was assumed that other parameters would not be changed (i.e., the other V_{max}'s etc.), (Fig. 4-2C). The fact that the flux value approached asymptotically by the V_{max} integration (Fig 4-2C) corresponds to KAY determined using flux integration (Fig 4-2B) lends support to the idea that there is an intrinsic kinetic limitation to the amount of flux that can be directed towards a pathway by overexpression of that pathway.

In the V_{max} integration, As V_{max} increases, steady state flux plateaus, but threonine decreases, approaching zero (Fig 4-2D). While acetyl-CoA is consumed in the same reaction, acetyl-CoA concentration never approaches zero. From this, we can infer that threonine is the limiting metabolite for the production of 1-butanol, and even though there is sufficient carbon to produce more 1-butanol, the kinetics of the system prevent the acetyl-CoA flux from redistributing towards threonine. This is one reason the KAY is less than the maximum theoretical yield (Fig 4-2B). It is possible that with other genetic manipulations, such as overexpression of other upstream enzymes or knocking out competing enzymes, a higher yield would result.

4.1.5 Kinetic parameter integration when the new pathway negatively affects the host

We can also include a system which may be more biologically realistic (Fig 4-2E). As one enzyme or pathway becomes highly overexpressed, the new pathway may negatively affect the host metabolism through various mechanisms. It is possible that the expression and activity of other enzymes could decrease due to competition for energy, amino acids, or amino acyl-tRNAs. To investigate these effects, in the toy system, the butanol pathway overexpression was tied to repression of threonine synthesis. Again, the steady-state metabolite concentrations were determined by continuation and the butanol flux was calculated as a function of V_{max} for V_{11}, k_{14}. In this integration, the resulting butanol flux went through a maximum and decreased as V_{max}

continued to increase (Fig 4-2E). Thus, the final butanol production yield was less than for the unrepressed condition (Fig 4-2C). Additionally, for the repressed condition, threonine was decreased to zero as before (Fig 4-2D), but the acetyl-CoA did not.

4.1.6 Chemical & Biological Reasoning

To fully cover the chemical transformation space, one must consider other hypothetical reaction systems. For example, if only chemical constraints are considered, the synthesis of butanol is limited by the number of carbons, for example. The maximum butanol synthesis is given by the following reaction:

$$1.5 \ H_2O + C_6H_{12}O_6 \rightarrow 1.5 \ C_4H_{10}O + 3 \ O_2$$
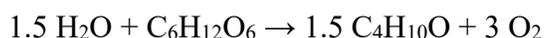
Resulting in a total yield of 1.5 butanol/glc. In our simplified cellular model (Fig 4-1), stoichiometry limits butanol synthesis to 1 butanol/glc, since one pyruvate and one acetyl-CoA are required for each butanol. The kinetically accessible flux and biologically realistic flux (Figs 4-2C, 4E) are determined from the integration carried out above, as the V_{max} for butanol synthesis approaches infinity for the no repression and with repression conditions, respectively. A comparison of all these yield determinations is presented in Fig 4-2G.

Biological realism in this system was modeled as repression of the reaction for threonine synthesis as overexpression of the propanol pathway increases. The exact implementation of 'biological realism' is arbitrary, but the analysis is meant to show the limitations of the KAY method, and that it may not correspond to observed yields due to mismatches between model and biological reality. Although in this case, the interaction causes the yield to be less than the KAY value, it cannot be ruled out that actual yield may be higher than predicted KAY, through interactions not included in a kinetic model. In other words, positive regulation of the host by the new pathway would cause the maximum yield to increase beyond the KAY value calculated above.

In general, these effects would involve inhibition of competing pathways relative to the pathway of interest. This could be caused by various signal transduction pathways, such as sigma factors influencing [68] expression level, redox state [69], or substrate-level regulation not included in the model. Any mathematical model is imperfect, but the KAY concept allows us to use kinetic models in a powerful way and push them to their limits of predictive power and usefulness. The KAY concept is used in the upcoming sections to predict the effects of genetic manipulations on exometabolite yields.

4.2 EMRA for predictions in *E. coli*, *Yarrowia lipolytica* and *Clostridium thermocellum*

Y. lipolytica and *C. therm* are non-model organisms which have promise to address rising metabolic engineering challenges. *Y. lipolytica* is an oleaginous yeast which produces lipids at high titer (~30 g/L).[70] Canonical understanding of yeast biomass combined with more recent knowledge of the lipid composition can give insights about the reference steady state of this strain.[71] *C. therm* is a cellulolytic thermophile, addressing the need to address more recalcitrant substrates for bioprocessing.[72,73] Models of *C. therm*, *Y. lipolytica* as well as the model organism *E. coli* will investigate ways to improve production of desired compounds and demonstrate the viability of using EMRA and the KAY concept on *in vivo* systems.

4.3 Constructing a model of n-butanol production in *E. coli*

The n-butanol production model was based on a previously generated model of *E. coli* generating isobutanol during growth phase based on the data published in [74]. The original model used pathway information from the Ecocyc database [75]. The original model was constructed by lab mate Jimmy Rivera and was featured in other published work.

Several changes were made to the model for the purpose of using the model to analyze n-butanol production data provided by Osaka University. First, the biomass term in the model was

eliminated, since the experiments data provided by Osaka University were conducted at high cell density and in stationary phase, so biomass accumulation is likely to be small. Second, the n-butanol synthesis pathway was added. This pathway encompasses six reactions, starting from two acetyl-CoA molecules and requiring four reducing equivalents per butanol [67]. Third, export reactions were added for eight metabolites: pyruvate, succinate, lactate, formate, acetate, butyrate, ethanol and butanol. The final model contained 60 reactions and 47 metabolites.
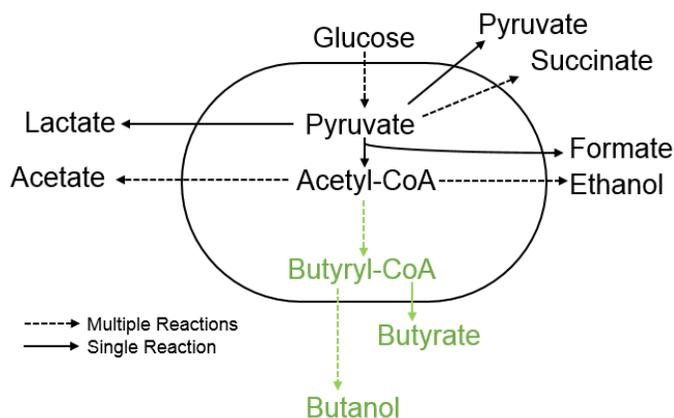
Fig. 4-3) Overview of the model used for generating insights about n-butanol production in E. coli. Overall, the model contained 60 reactions and 47 metabolites. 8 exo-metabolites export reactions were included in the model to make use of data provided by the Osaka University lab.

4.3.1 Using data from Osaka University to Generate a Reference Steady State

Osaka University provided data for the extracellular concentrations of the eight metabolites mentioned. There were four different strains, the first of which JCL16, was used to build the reference steady state discussed here. The overall configuration of the model is shown in Fig 4-3. These metabolite concentrations were used to represent the fluxes of the export reactions in the model. The export fluxes were constrained to the OU data and linear programming was used to find a solution minimizing the respiration reaction flux, since the culture was carried out anaerobically. Data was also provided for glucose consumption. However, when glucose consumption was constrained to the observed value, no solution was found. The system was

overdetermined, such that no solution simultaneously satisfied all of the observations. Possible explanation for this may be error in experimental measurement, or the imperfections of the stoichiometric model.

Instead, the system used linear programming to minimize the difference between the observed glucose consumption and the glucose intake flux value in the model, rather than setting glucose flux as a hard constraint. The result showed that the model was able to replicate the observed product yields within less than 10% (Fig 4-4). This discrepancy could be from random experimental error, unexpected sources of carbon in the culture, or missing reactions in the model which might allow for more carbon-efficient metabolism. In any case, error within 10% was considered proceed with the model to make further predictions.
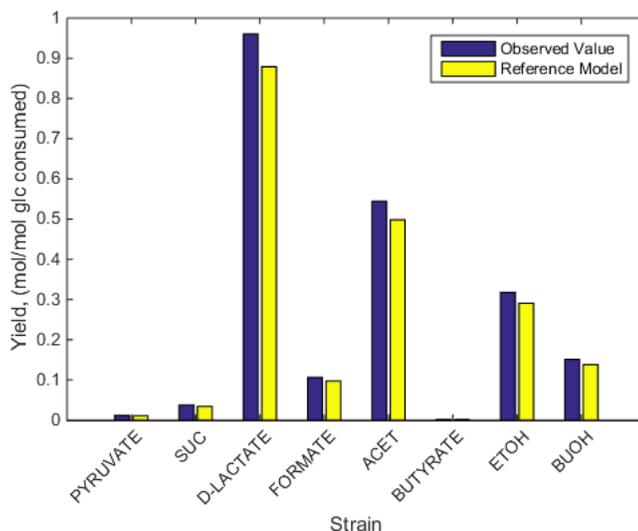
Fig. 4-4) Linear programming was used to minimize the difference in measured yield (dark blue bars, exometabolite measurement/glucose consumption measurement) and the result obtained when fluxes were calculated subject to steady state and the model stoichiometry (yellow bars). The difference in the two was about 5%, within possible measurement or other experimental error, and the flux distribution thus calculated was used as the reference state for further calculations.

4.3.2 Testing the model on different genotypes

After setting the reference steady state of the model (Fig 4-4), work can continue on the application of the EMRA to the model. In the EMRA framework, realistic rate laws are assigned to each reaction using the network stoichiometry and reversibility of each reaction. After the rate

laws are determined, the reference steady state is used, a suitable number of parameter sets (n = 500) were generated, comprising an ensemble. The parameter sets were constrained to the reference steady state fluxes, and were determined to be dynamically stable, or discarded. Bacterial strain JCL16 was used as the reference steady state.

Table 4-1) Genotypes for the four strains involved with the butanol production project.

Strain Name	Genotype (*E. coli* strains)
JCL16	Wild Type(BW25113/F'[traD36 proAB+ lacIqZ M15 (Tetr)]) + Butanol Plasmid
JCL16F	JCL16 + fdh
JCL166F	JCL16F ΔldhA ΔadhE ΔfrdBC
JCL299F	JCL166F Δpta

The JCL16 ensemble was then used to 'predict' the extracellular metabolite fluxes on the basis of genetic manipulations. These prediction values were then compared to the experimental observations of the Osaka University group.

The predictions were generated using parameter continuation methods that have been described previously[13]. In short, the system is constrained to steady state and perturbed by changing the amount of some enzyme. This is accomplished mathematically by parameter-domain integration, rather than time-domain integration, which would have a higher computational burden. The amount of enzyme is changed by this method until the system becomes unstable, or until a pre-set fold-change is reached (in this instance, 10-fold). In this way, overexpression and knockdown can be suitably represented.

In the case of JCL16F, the overexpression target is Fdh (formate dehydrogenase). For this genotype, the Fdh reaction flux was integrated from zero to bifurcation for all members of the ensemble (the method termed kinetically accessible flux). For JCL166F & JCL299F the appropriate knockouts were made first, by integrating the relevant V_{max} values to 10% of their original values, then applying the Fdh overexpression to bifurcation.

The compound of interest in this investigation is n-butanol. Thus, an important feature of the model is its ability to reproduce the correct n-butanol yields from glucose, in comparison with the experimental observations.
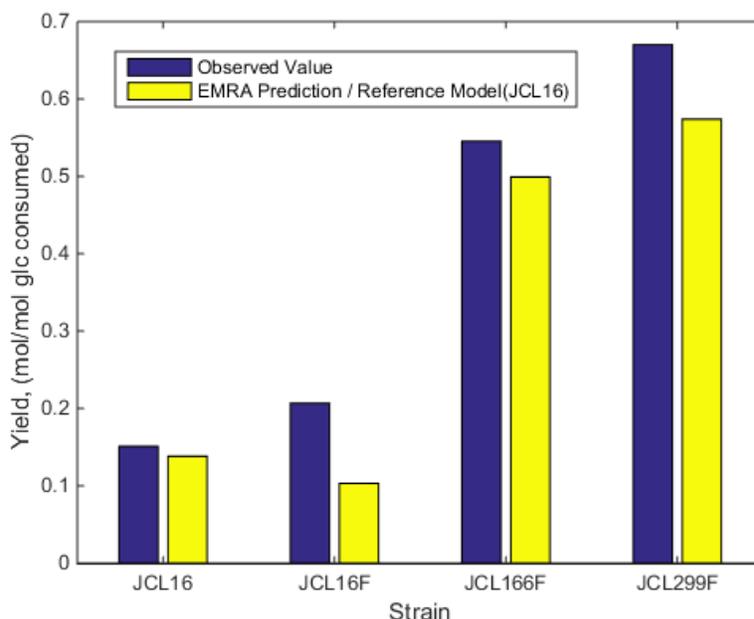
Fig. 4-5) Comparison of observed n-butanol yields for each genotype (dark blue bars) with the yield values obtained by ensemble modeling and robustness analysis (yellow bars). The predictions match well overall for the three simulated genotypes (JCL16F, JCL166F, JCL299F) despite the fact that the observed value changes dramatically, roughly four-fold, from the reference, JCL16 genotype.

The results show that the model performs well overall, particularly with the JCL166F and JCL299F conditions (Fig 4-5), which have the biggest differences from the original reference state. These values, the model predicts well, within 15% of actual value, even when, in the case of JCL299F, the observed value is four-fold higher than the original reference state. This shows that the model is capable of correctly capturing large changes of behavior in model, a feat which may be difficult to achieve with linear or less sophisticated kinetic models.

Even though we are most interested in n-butanol as a product in this model, we can still compare the predictions of the model for other compounds as well. As a byproduct of the parameter continuation calculations, the simulations also generate expected values for reaction

fluxes of the eight export fluxes. These can readily compared to i) the original reference values (JCL16) and ii) the measured experimental values.
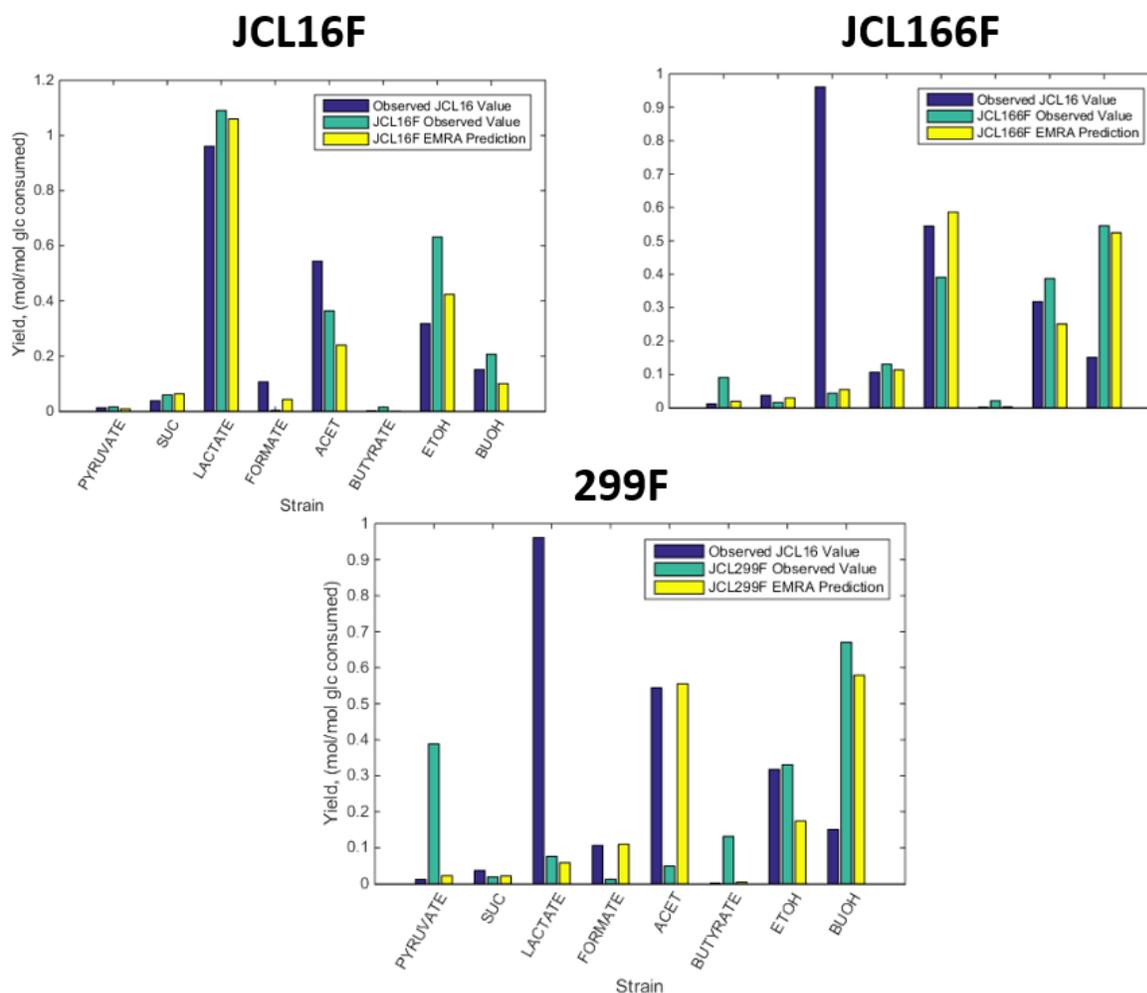
Fig. 4-6) Comparison of observed exometabolite yields for each genotype (teal bars) with the reference (JCL16) values (dark blue) and the simulated yield values obtained by ensemble modeling and robustness analysis (yellow bars). The predictions match well overall for the three simulated genotypes (JCL16F, JCL166F, JCL299F) despite the fact that the observed values sometimes change dramatically from the reference, JCL16 genotype. Notable exceptions are the pyruvate and acetate predictions for JCL299F.

We find that in general, the model performs well for most compounds in JCL16F and JCL166F (Fig 4-6). However, the model fails drastically in predicting the amounts of pyruvate and acetate for the JCL299F strain (Fig 4-6, 299F, lower panel). Overall, however, and in particular with regard to the performance with butanol, the model performs well, and demonstrates

the feasibility of using EMRA to integrate experimental data with kinetic metabolic models to make realistic predictions about the effect of genetic manipulations.

4.3.3 Finding new genetic targets

Top 5 OE Targets

Enzyme Name	Predicted Butanol Yield (mol/mol glc)
Butyryl-CoA dehydrogenase (PduP)	0.58
Thiolase (AtoB)	0.56
Pyruvate Dehydrogenase (Pdh)	0.54
Butanol Dehydrogenase (Bdh)	0.54
Hbd	0.53

Top 5 KO Targets

Enzyme Name	Predicted Isobutanol Yield (mol/mol glc)
Butyryl-CoA Hydrolase (butyrate-forming)	0.63
Acetyl-CoA Dehydrogenase (acylating)	0.59
Ethanol Export	0.58
Alcohol Dehydrogenase (ethanol)	0.58
Pyruvate Export	0.57

JCL299F Reference Flux Baseline Yield = 0.52

Fig. 4-7) A model of the JCL299F as a reference steady state was generated from the JCL299F observed yields. Each enzyme was then perturbed up and down and n-butanol yield for each perturbation was tabulated. The top 5 knockout and overexpression targets are shown here.

After using the JCL16 reference strain as a means of demonstrating the model, we can identify which targets would be most effective in further increasing the butanol yield of the best-performing 299F strain. To do this, a reference state based on the JCL299F data was constructed. Then all of the enzymes in the model were subjected to perturbation, overexpression and knockdown of 10-fold. The genetic changes that result in the highest butanol yields from glucose were ranked and the top 5 knockdowns and overexpressions are presented here (Fig 4-7).

These genetic targets form the basis of a set of recommendations that are actionable by experimental researchers. Such actionable insights are often touted as justifications for metabolic simulations, but rarely provided. Here I have provided them, based on a validated kinetic model capable of reproducing highly non-linear phenotypes.

4.3.4 Changes to the Model That Increase Model Performance
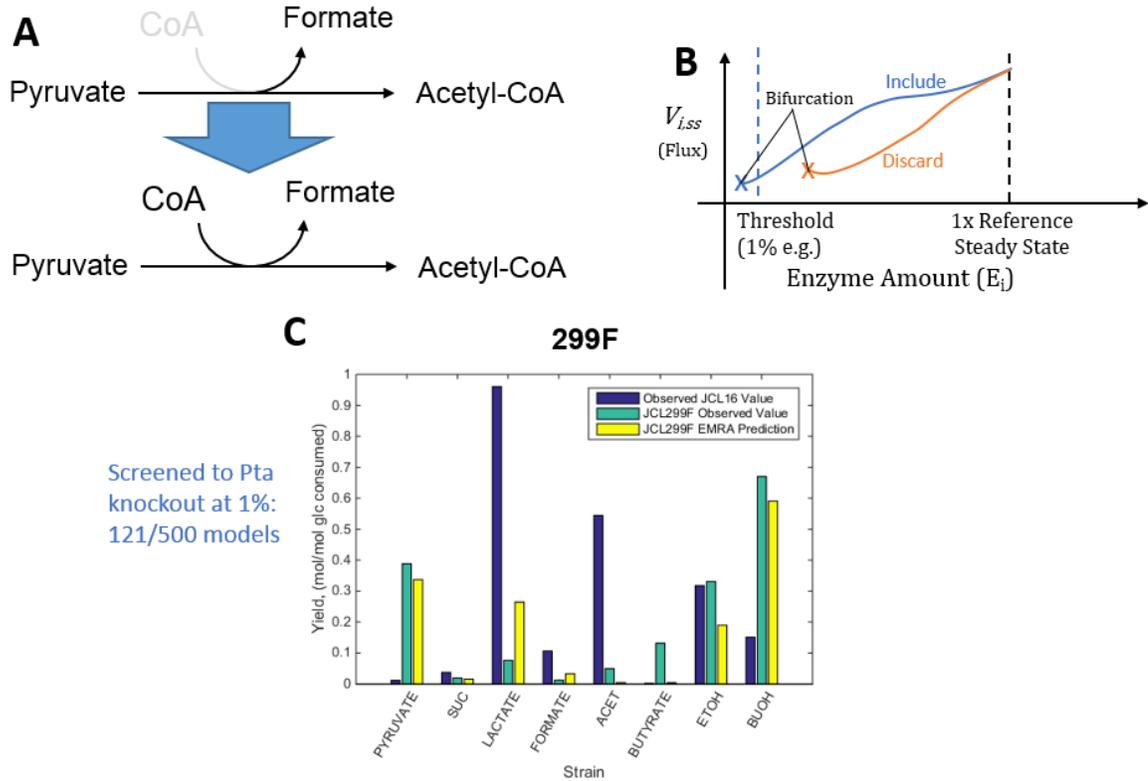
Fig. 4-8) Comparison of observed exometabolite yields for each genotype (teal bars) with the reference (JCL16) values (dark blue) and the simulated yield values obtained by ensemble modeling and robustness analysis (yellow bars). The predictions match well overall for the three simulated genotypes (JCL16F, JCL166F, JCL299F) despite the fact that the observed values sometimes change dramatically from the reference, JCL16 genotype. Notable exceptions are the pyruvate and acetate predictions for JCL299F.

While overall, the model performs relatively well in predicting exometabolite yield in the various genotypes, there are a notable exceptions with the pyruvate and acetate predictions for the 299F genotype. In consultation with the Osaka University Group, we hypothesized that CoA limitation was an important factor in determining the behavior of the cells and the 299F. In the original model, free CoA was not included as a metabolite in the model. Thus, effects of the depletion of CoA on the model were not manifest in this model.

To correct this, and other defects in the model, changes were made. First, free CoA was added as a metabolite to the model. Thus, CoA would appear as a reactant or product in the relevant reactions in the model (Fig 4-8A).

Second, the pta knockout in 299F was used as a screen for models. Previously, to model knockouts, all parameter sets in the ensemble were perturbed by changing V_{max} down to 1% of the original reference value, and the parameter set was included in the final prediction regardless of whether bifurcation occurred before the 1% threshold. In other words, models which bifurcated at 20% of reference V_{max} value would still be included, at the values calculated just before bifurcation. As a change, I included only models which could be successfully perturbed to 1% of pta V_{max} in the JCL299F predictions (Fig 4-8B). Combined, these two changes to the model were found to greatly improve the predictions of pyruvate and acetate production for the JCL299F genotype (Fig 4-8C).

Using successful knockout to 1% as a screen for other knockout perturbations was performed, but it was found not to make a significant difference in predictions for these genotypes [data not shown].

4.3.5 Lessons Learned from Modeling *E. coli* Production of n-Butanol

The performance of this modelling strategy with this production system shows that ensemble modelling can go beyond just the *in vitro* systems described previously. This application shows that ensemble modelling and robustness analysis (EMRA) provide methods to go from a low information description of a system to effective kinetic models in an automated way.

While the EMRA method is or can be automated, it is also clear from these results that attention to model construction is also required. It is not clear *a priori* whether inclusion of CoA as a free metabolite would result in a more accurate model or not. For instance, a cell under CoA limitation could respond by synthesizing more. However, we now have evidence that the cell does not, mostly, respond in such a fashion. This is an interesting point both from a modelling

perspective to improve model performance, and from a descriptive biological perspective to better understand the nature of *E. coli* metabolism.

4.4 Ensemble Modelling Using *Clostridium thermocellum*

Using non-edible plant material as a feedstock for bioprocessing is an important goal [76]. *Clostridium thermocellum* has shown the ability to convert cellulose, a recalcitrant store of carbohydrates, into a useful product like isobutanol[77]. A model of *C. therm* metabolism was created by a team at Penn State University including Satyakam Dash and Ali Khodayari under Professor Costas Maranas. The model included a stoichiometric matrix of relevant reactions, as determined by discussion with the Lee Lynd group at Cornell. In collaboration with Penn State, I developed the model for analysis with EMRA.

4.4.1 Development of *C. thermocellum* Ensemble Model

Stability analysis, termed ensemble modeling robustness analysis (EMRA)[13] was carried out. The stoichiometric matrix of the underlying model was used to create a kinetic model of the system. Reversibilities were assigned according to reported reversibilities or thermodynamics. Stoichiometry and reversibilities were used to assign realistic reaction modular rate laws as described by Liebermeister and used previously [8,13].

H^+, H_2 and H_2S metabolites were removed. It is likely that the maintenance of these metabolites' steady states is accomplished by mechanisms not reflected in the model, so steady maintenance of steady state for these metabolites is an unnecessary kinetic constraint which may distort the model.

Secreted metabolite (exometabolite) concentration data was obtained from a previously published study of *C. therm* [77]. In the previous exploration, ten exometabolites (isobutanol, ethanol, valine, citrate, malate, succinate, lactate, formate, acetate, fumarate) (Fig 4-9A) were

measured after a 75hr high density fermentation for both a wild type (WT) and isobutanol overproducing strain (CT24). CT24 overexpresses the isobutanol production pathway (Fig 4-9B).

Reference fluxes were assigned in the wild type by constraining export reactions for the ten measured exometabolites to the measured value. Other export reactions were removed. Theoretical glucose consumption was estimated by determining the molar amount of glucose required to produce each metabolite (0.5 mol glc for ethanol, succinate, lactate, acetate and fumarate—1 mol glc for isobutanol, valine, citrate, malate—0 for formate). Through discussion with the Penn State group, biomass was left in the model, with 3% glucose flux directed to biomass. Low biomass accumulation is expected since it is a high density fermentation. The biomass equation from the underlying model was used to determine the flux towards each biomass component and separate kinetic equations for each component were used.

Reasonable estimates were made for other flux determinations. The ratio of POR:FDH was set at 1:1. PEPCKr to PPDK was set at 1:1. PGM:PGCD was set at 8:1. PGCD catalyses the conversion of 3PG to 3P-hydroxypyruvate, which can be converted to serine and recycled in a futile cycle through hydroxypyruvate and glycerate back to 3PG. The overall flux determination was carried out using linear programming to minimize the ATP lost from a futile glycogen cycling pathway.

Ensemble modeling was carried out by selecting realistic kinetic parameter values constrained to the reference fluxes used and discarding models which were not dynamically stable. The models (n = 100) were then perturbed by increasing the V_{max} of the isobutanol production pathway enzymes and export reaction by 100x using parameter continuation integration. If a bifurcation was detected by finding a singular Jacobian during any step of the integration, it was halted. About 10% of the models remained stable to 100x overexpression. The flux values at

bifurcation or 100x overexpression (whichever came first) were used to generate an ensemble average of predicted fluxes. To determine enzymes which have the best impact on further improving isobutanol yield from glucose in the CT24 strain, all enzymes were overexpressed 10x and knocked down 10x using parameter continuation to determine steady state isobutanol production.

4.4.2 Results of *C. thermocellum* Modeling
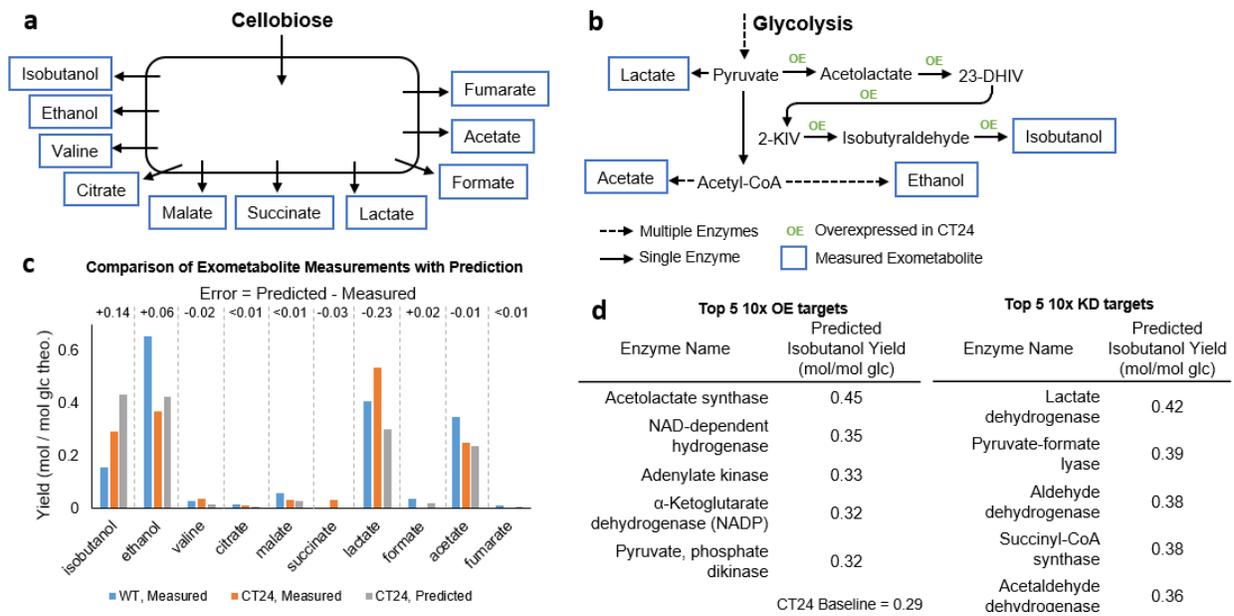
Fig. 4-9) A) Diagram showing input and output from experimental *C. therm* data. Exometabolite data was used to constrain metabolic fluxes in the reference steady state. The reference steady state is used to constrain kinetic parameters to realistic values. **B)** Detailed view of major pyruvate- and acetyl-CoA-derived exometabolites. The isobutanol pathway from pyruvate was overexpressed in the CT24 isobutanol overproducing strain. **C)** Comparison of measured exometabolite yields for WT, CT24 and CT24 simulations. Simulation was accomplished by increasing Vmax for isobutanol pathway enzymes using parameter continuation until bifurcation or 100x increase. Values are presented as molar yield per mol glucose. Theoretical glucose was calculated by determining total amount of glucose required to generate all observed products for all 3 conditions. **D)** To further improve the CT24 strain, the top overexpression and knockdown targets for isobutanol production were identified.

For eight out of ten exometabolites (all except valine and isobutanol), the sign of the measured change of the molar yield from glucose was predicted correctly by the simulation (Fig 4-9C). Interestingly, the acetyl-CoA derived products were predicted almost correctly. Acetate had error of only 0.01 out of a total change of about 0.1 while ethanol was about 0.06 out of a measured change of 0.26. However, larger errors were found in the pyruvate-derived

exometabolites. Isobutanol showed error of +0.14 while lactate was -0.23. Overall, it appears that the simulation correctly predicted the shift from acetyl-CoA-derived to pyruvate-derived exometabolites, but did not correctly determine the split between isobutanol/lactate. This could indicate there were unanticipated effects on the activity of lactate dehydrogenase (e.g. the activity of LDH was upmodulated in CT24 by some unknown mechanism) which were not reflected in the parameter continuation integration used. Another possibility is that the fermentation took place in distinct stages with flux distributions varying as time increased to 75 hours. A steady state simulation would not necessarily capture all dynamics at play.

The top 5 knockdown and overexpression targets were also identified to improve the isobutanol yield in the CT24 genotype (Fig 4-9D). Only one overexpression and one knockdown target were predicted to raise isobutanol yield from 0.29 mol/mol glc equiv to above 0.4. The knockdown target was lactate dehydrogenase while the overexpression target was acetolactate synthase. These enzymes compete for pyruvate. Thus, it makes sense that these enzymes controlling pyruvate-derived exometabolites would be important in isobutanol production in a CT24 strain. Acetolactate synthase is already overexpressed in the CT24, but it could be that further overexpression of this enzyme may further improve isobutanol yield. Other enzymes in the isobutanol pathway did not appear in the top 5 overexpression targets.

Other identified targets also have some intuitive rationale. α-Ketoglutarate dehydrogenase, was also an overexpression target, possibly because it provides NADPH for the isobutanol pathway. Pyruvate, phosphate dikinase is predicted to operate in the pyruvate direction in this model and was another oeverexpression target. Other knockouts included reactions which produced acetaldehyde and acetate (Acetaldehyde and aldehyde dehydrogenase) which could further shift the overall exometabolome from acetyl-CoA-derived to pyruvate-derived products.

Overall, the results indicate the feasibility and usefulness of kinetic models using ensemble modeling for *C. therm* in particular, and organism-wide metabolic simulations in general. At the same time, this endeavor shows the shortcomings of assuming that overexpressing a protein in a cellular system will increase the activity of that protein while leaving others unchanged. Potential follow-up work could include integration of this approach with proteomics to include system-wide changes.

4.5 Ensemble modeling of acetate conversion in *Y. lipolytica*

Previous work undertaken by the UCLA and Chalmers University Teams has resulted in a plausible model of *Yarrowia lipolytica* metabolism[78] under conditions of glucose conversion to fatty acids. This model construction consisted of finding a suitable set of reactions for the purpose of representing metabolism and lipid synthesis. This encompassed glycolysis, pentose phosphate pathway for the generation of NADPH, pyruvate transport to the mitochondria and tracylglycerol synthesis in the lipid body. The model found contained a total of 120 reactions and 106 metabolites and was able to fit fluxes from both the wild type and the lipid overproducing strain.

4.5.1 Acetic Acid Metabolism in *Yarrowia*

Acetic acid metabolism by yeast is a common trait. *Yarrowia* and other yeast strains including the model strain *Saccharomyces cerivisiae* have been observed to be capable of growth on acetic acid. A quantitative mRNA study of *S. cerevisiae* was undertaken which pinpointed many of the genetic changes associated with a switch of carbon source from glucose to acetic acid [79]. By this study, it was found that the major route for acetic acid metabolism is by the reaction acetyl-CoA synthetase, which converts acetic acid into acetyl-CoA powered by the conversion of ATP to AMP. We used this as the route for acetic acid metabolism in *Yarrowia* for the model. Additionally, we achieved acetyl-CoA transport into the mitochondria via the acyl-

carnitine/carnitine translocase system. The glyoxylate cycle is presumed to be active in the mitochondria under acetate conditions, which allows for a net conversion of two acetyl-CoA molecules into a 4-carbon dicarboxylic acid (malate). Malate is then decarboxylated in the mitochondria by malic enzyme which has been noted to be localized there[80]. A pyruvate carrier [81] then moves pyruvate to the cytosol where it can be used for gluconeogenesis etc.

In contrast to glucose conditions, this represents a reversal of the usual pyruvate and acetyl-CoA flow from mitochondria to cytosol. In glucose conditions, pyruvate is transported into the mitochondria by the pyruvate carrier, where it undergoes decarboxylation by pyruvate dehydrogenase. Acetyl-CoA is then effectively exported to the cytosol by the citrate-malate antiporter. This system is not reversible since cytosolic ATP-citrate lyase and mitochondrial citrate synthase are irreversible.

4.5.2 Using Previously Reported Acetic Acid Conversion Experiments to Find Reference Fluxes

One of the most comprehensive studies of acetic acid consumption and conversion by *Yarrowia* to date was conducted by Fontanille *et al* [82]. The authors reported that in bioreactor experiments using *Yarrowia* in growth phase, the yield of biomass from acetic acid was 0.50 (g biomass / g acetic acid consumed) and the lipid yield was 0.15 (g lipid / g acetic acid consumed).

To use the conditions reported by the author, the mass/mass yields of lipid and biomass were converted into molar fluxes for the model. In the original *Yarrowia* model, a biomass term was included which was based on work done by Gombert *et al* [83]. On this basis, a biomass function was derived which used the correct ratio of each amount of each metabolite for yeast in rapid growth conditions. The average molar mass of each unit of biomass was calculated, and from that and the biomass mass yield a biomass flux was determined. The lipid composition of the acetic-

acid grown *Yarrowia* was reported in [82] and the average molar mass of each fatty acid was calculated, allowing the molar flux of lipid production to be set. After fixing biomass and lipid production, linear programming was used to bring acetate uptake as close as possible to the observed value, and achieved a result within 4% of the reported value. Acetyl-CoA was removed from the biomass term, since that is represented by the lipid synthesis reactions.
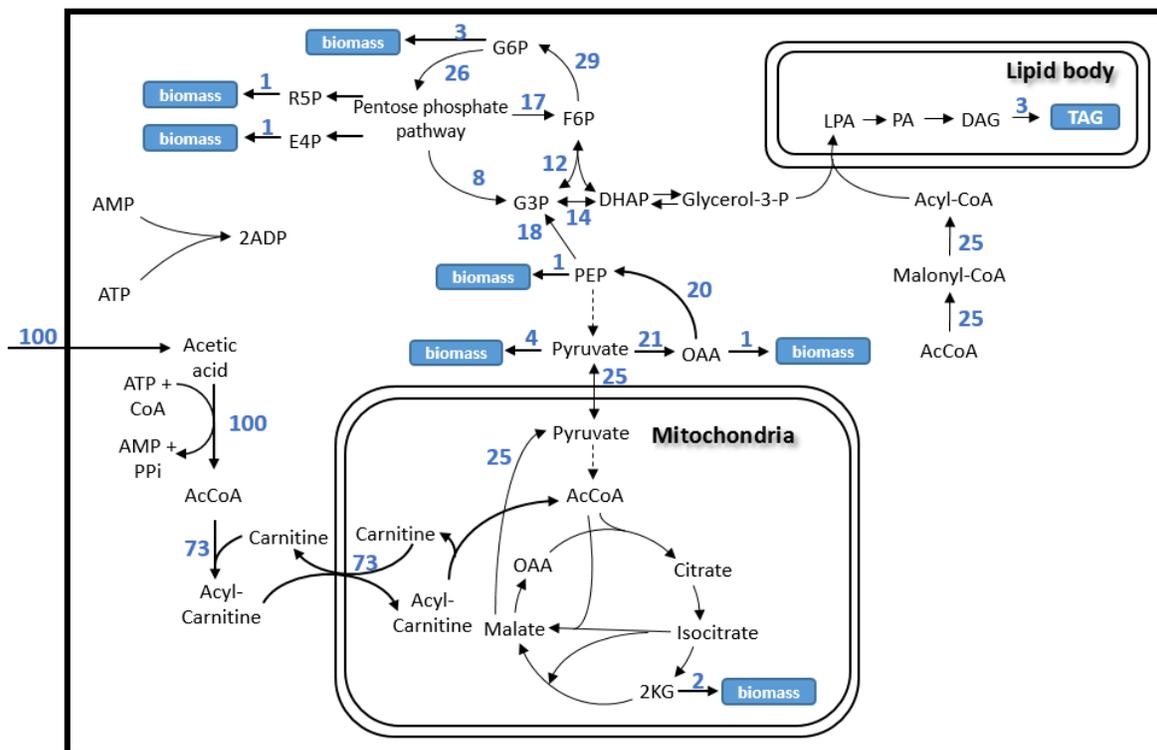
Fig. 4-10) Diagram showing the flux distribution determined for *Y. lipolytica* growth under acetate conditions. This distribution is approximated by using biomass objective function and acetate metabolism information from the model yeast species A. cerivisiae. Compartmentalization is maintained by including separate metabolites for each compartment.

4.5.3 Determination of genetic targets

This model of *Yarrowia* metabolism (Fig 4-10) under acetic acid feed conditions is a reasonable representation of the system and can be used for further simulations. Key assumptions about compartmentalization and metabolite flow have been identified, which will allow this model to be adapted to many other conditions, including high density production schemes with low biomass accumulation. Additionally, the network and fluxes identified here allow for analysis of

genetic changes to this organism under these conditions to identify which changes would increase lipid yield using the ensemble modelling framework.

The model was perturbed by changing each enzyme amount 10-fold up or down for a suitable number of parameter sets (n=100). The resulting fluxes were investigated to determine which enzymes could be perturbed to cause the greatest increase in lipid yield from acetate. The results are presented here, for top 6 knockdown and overexpression targets.
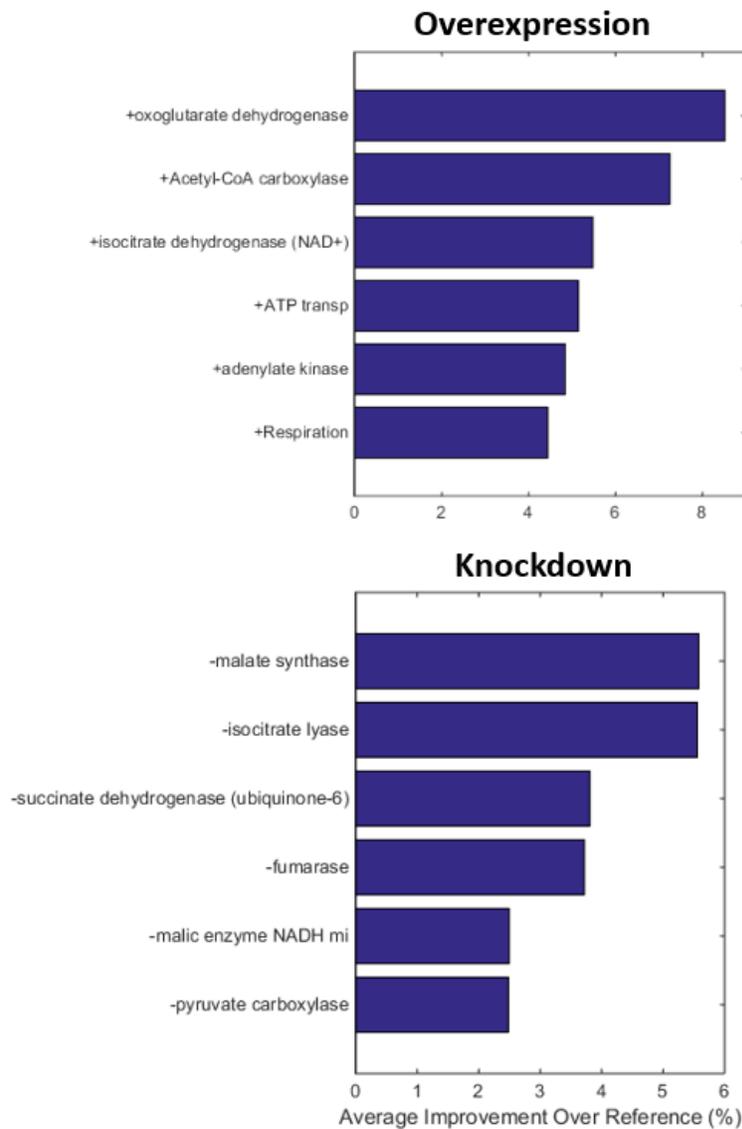

Fig. 4-11) Diagram showing the flux distribution determined for *Y. lipolytica* growth under acetate conditions. This distribution is approximated by using biomass objective function and acetate metabolism information from the model yeast species A. cerivisiae. Compartmentalization is maintained by including separate metabolites for each compartment.

These genetic targets provide useful insights (Fig 4-11) for experimentalists looking to improve the performance of *Y. lipolytica* with acetic acid. The conversion/valorization of acetic acid with *Y. lipolytica* is an area of lively and greatly expanding interest, and this work will allow for rational targeting of new research directions.

4.6 Summary & Conclusion

The performance and success of EMRA with these in vivo systems paves the way for its use in yet more systems. This variety of applications show that the EMRA method is both highly flexible and surprisingly powerful. It is able to take a low information network skeleton and convert it into a queryable kinetic model. These models have demonstrated a high degree of fidelity to observed exometabolite yield, especially in the *E. coli n*-butanol model. The 3 genotype, 8 exometabolite data set is a high-dimensional standard to match, yet the model performs well without any modifications. With small tweaks, the model matches every yield almost perfectly, illuminating biological significance in the process.

5. Network Structural Features Affect Stability of Calvin Bassham Benson Cycle in Plants

The stability of the Calvin Bassham Benson (CBB) cycle remains an area of active computational research. Our understanding of biology and the prospect for bioengineered plants with higher productivity may both be impacted by a greater understanding of this area. Here we use the ensemble modelling robustness analysis (EMRA) framework to show that the action of the phosphate/G3P antiporter is much more significant for maintenance of stability than a recently proposed G6P shunt. Additionally, we interpret recent results suggesting that overexpression of RuBiSCO does not improve growth rate of plants but overexpression of sedohuptulose-bisphosphate phosphatase (SBPase) does. Our simulations reproduce this result, but only in models which do not include the G6P shunt. Taken together, these results may suggest a situational role for the G6P shunt, possibly in dynamic situations under starvation or other stress conditions.

5.1 Introduction

The Calvin Bassham Benson cycle (CBB) is responsible for CO_2 fixation by plants, including the C3 & C4 variants, of which the C4 is an adaptation which allows for plants in high temperature or low water environments [84–87]. Plants have advanced regulatory systems which allow them to successfully grow and thrive in an unpredictable and changing world [88–92]. For example, sugars generated from CO_2 during daylight are stored as starch in photosynthetic and non-photosynthetic chloroplasts. Nighttime consumption of starch is tuned to leave only a small amount remaining by morning—and this consumption rate is dynamically tuned to adjust for changing day length[93–95].

Among canonical metabolic pathways, the CBB pathway is highly branched and complex, much more so than simple linear pathways like glycolysis or simple loops like the TCA cycle. This is in some ways the result of the chemical difficulty of aerobic CO_2 fixation [96,97] which seems

to require a carbon reshuffling step to regenerate a suitable starting substrate like ribulose-1,5-bisphosphate [98]. The complexity of the pathway, in which there is not a linear pathway from substrate to product, results in instability if intermediates are depleted. For example, if the sugar phosphates in a chloroplast become depleted, the pathway is not able to continue since some starting substrate (RuBP) is required to continue CO_2 fixation [99].

There are two main mechanisms of transport of sugars across the chloroplast membrane. First, there is the G3P/phosphate antiporter [100]. This transporter moves a G3P from the CBB pathway in the chloroplast to the cytosol, where it is used for various cellular functions. In return a phosphate molecule is transported into the chloroplast, effectively keeping the total number of phosphates (including phosphate attached to sugars) in the chloroplast constant. Second, there are glucose (putative) [101] and maltose transporters [102], of which the maltose transporter is known to be essential for starch breakdown. G3P is a CBB intermediate and is directly interconvertible with other sugar phosphates, so a depletion of G3P would be problematic for CBB. However, glucose and maltose are more removed from the CBB pathway itself and are possibly only produced as starch breakdown products [95].

The direct regulation of plastidic enzymes involved in photosynthesis is accomplished by redox-mediated proteins called thioredoxins [103]. In light conditions, the NADPH/NADP$^+$ ratio is higher because the photosystems which generate NADPH from light are active. As a result, the disulfide bonds in thioredoxins and other regulated proteins are broken, mediating enzyme activity. In *Arabidopsis thaliana* several enzymes are known to be redox regulated in this manner [104].

Some enzymes of the CBB cycle are activated in a reducing (light) environment by the breaking of their disulfide bonds. In the dark, these enzymes are attenuated in the oxidizing environment. Of the 12 enzymes of the canonical CBB cycle, 4 are known to be redox regulated

in the ferredoxin/thioredoxin system [104]. First, GAPDH converts 1,3-bisphosphoglycerate to glyceraldehyde-3-phosphate using reducing power from NADPH. GAPDH is reversible, although in dark conditions scarce NADPH is required for other critical cellular functions.

In addition to GAPDH, enzymes which catalyze the cleavage of high-energy phosphate bonds are also thioredoxin-regulated, presumably to reduce thermodynamic losses in dark conditions. Phosphoribulokinase (Prk) catalyzes the cleavage of ATP to ADP coupled with the conversion of ribulose-5-phosphate to ribulose-1,5-bisphosphate. Prk and GAPDH are inactivated in the non-enzymatic oligomerization with chloroplast protein CP12 in oxidizing conditions, which is reversed by NADPH [105]. Sedoheptulose-1,7-bisphosphatase catalyzes the irreversible loss of phosphate from the sedoheptulos-1,7-bisphosphate to result in sedoheptulose-7-phosphate. Fructose-1,6-bisphosphatase catalyzes an analogous reaction and loss of phosphate to result in F6P. These enzymes are all regulated to lose function in dark conditions when NADPH is low and CO2 fixation cannot continue [104].

The Calvin cycle has many enzymes in common with the pentose phosphate pathway, except that it functions in the reverse direction, leading to the distinction between the traditional or oxidative pentose phosphate pathway (oPPP) and the CBB-synonymous reductive pentose phosphate pathway (rPPP) [106]. Distribution of oPPP and rPPP enzymes within plant cellular compartments (plastid vs. cytosol) is an area of research [107], but in *Arabidopsis*, it is recognized that the first three steps of the oPPP (glucose-6-phosphate dehydrogenase, gluconolactonase and 6-phosphogluconate dehydrogenase) are localized to both the plastids and the cytosol [108]. In addition to the CBB enzymes above, the plastidic enzymes of the oPPP, particularly G6PDH, are subject to redox-based regulation [109].

G6PDH is most active in oxidizing conditions which prevail in night darkness. The oPPP provides NADPH for critical cell functions when light is unavailable. Activity is highly attenuated by the presence of light. This is rationalized to be for the prevention of thermodynamic losses due to a futile cycle [110]. However, interestingly, the attenuation of G6PDH in reducing conditions is far from complete and varies widely by species. In the investigation of three different plastidic G6PDHs, activity is attenuated to anywhere from 10-30% of maximum in reducing conditions [109,111,112]. It has recently been suggested that flux through G6PDH and the next two oPPP enzymes (generating Ru5P) may stabilize the CBB pathway [113]. This opens the door for investigation into possible competitive benefits of a futile cycle which in terms of thermodynamics, is a clear loss.

Some previous efforts have attempted to address stability in the CBB pathway, but these have had shortcomings such as not considering phosphate [114] which our work suggests has a critical role in stability, or considering only a single set of parameter values [115], which doesn't reflect the range of stochastic and environmental variability encountered in biological reality. Other works focused on the well-documented oscillations of the CBB pathway [116], without considering general propensity towards stability (nonsingular Jacobian), or instability (singular Jacobian). In this work, we consider the present evidence that multiple structural features of CBB in plants and *Arabidopsis thaliana* in particular stabilize the pathway, independent of their effect on oscillatory behavior. In particular, we investigate the role of the G3P/phosphate translocator, the oxidative pentose phosphate pathway, as well as covalent modification of triose phosphate isomerase [117]. We use ensemble modeling robustness analysis, a method which investigates the stability of metabolic pathways using network information such as reference flux, network stoichiometry, reaction reversibility and substrate-level regulations [13]. We also consider the potential applications toward biotechnological work attempting to increase the productivity and growth rates of plants.

5.2 Building a model of chloroplastic metabolism

First, a consensus model of chloroplast metabolism flow in light conditions was developed (Fig. 5-1). Steps of the CBB cycle, starch synthesis and starch degradation were included. Additionally, G3P transport from chloroplast to cytosol is also included. NADPH generation by the light reactions and ATP generation through respiration were included as single reactions in the model.

Fluxes were set by linear programming to determine a reference steady state. Carbon was assumed to be split 50:50 between G3P and starch synthesis. Starch degradation was assumed to be at 2/3 the rate of starch synthesis. Starch degradation is represented as non-negligible in the model, since starch degradation rate was found to be almost unchanged by light in spinach leaves [118]. G3P export and import were modeled as parallel reactions in dynamic equilibrium, additionally at a 2:3 ratio for parsimony. Beyond these specifications, the system has no degrees of freedom so flux rate was completely determined.

Reactions were modeled kinetically using realistic rate laws which take into account number of substrates and products, and the reversibility of the reaction using modular reaction rates according to the method of Liebermeister [8]. Substrate-level regulations were added to the model as described in the Methods section. Parameter values were sampled constrained to reference fluxes and stability at the reference steady state was insured. A suitable number of parameter value sets (n = 300) were generated and tested. Enzyme levels were perturbed by using the parameter continuation method, where the system is perturbed, constrained to a fixed point, until the Jacobian becomes singular, or a metabolite concentration becomes negative. The fraction of parameter sets, or 'models' which become unstable at each level of integration, is plotted. Further details about the model are available in the methods section.
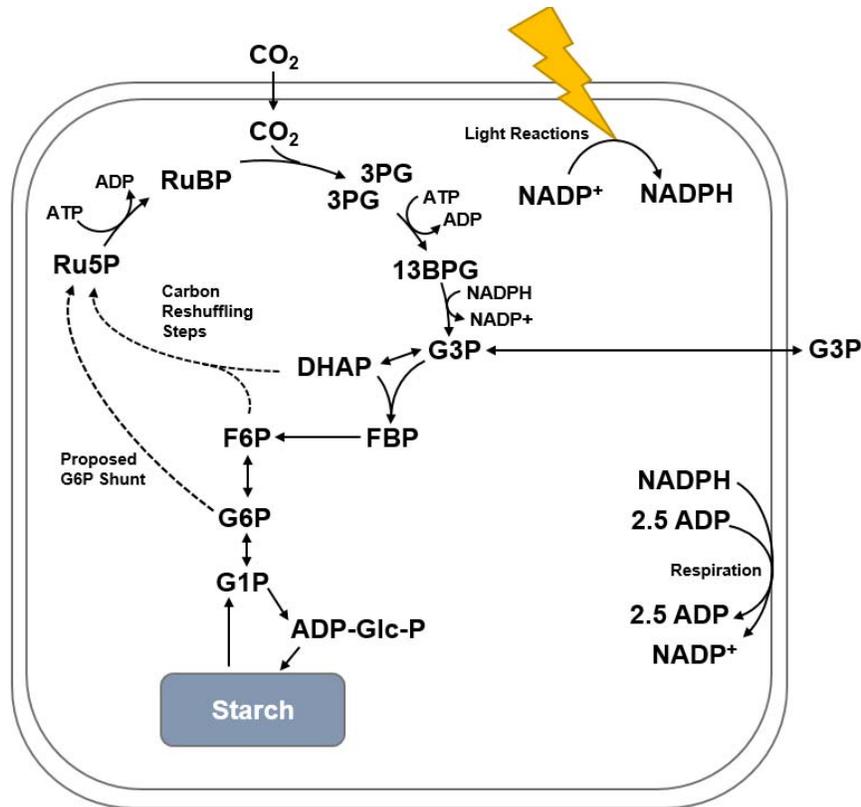
Fig. 5-1 The overall model of chloroplast flux used in this paper. Reactions in the carbon shuffling steps and G6P shunt are modelled individually but are shown in a simplified format. Light reactions and respiration were modeled as single reactions for simplicity. NADPH/NADP+ & ATP/ADP cofactors included in all simulations, free phosphate held constant in some simulations as noted. CO_2 was held constant in all simulations.

5.3 G3P/phosphate translocator almost completely stabilizes CBB

The G3P/phosphate translocator and phosphate in general is known to have an important role in the action of the CBB pathway (Fig. 5-2A) [119]. However, to date, this role has not been thoroughly tested by simulation efforts. Here we test the idea of the phosphate antiporter as a CBB pathway stabilizer by doing EMRA simulations of the CBB enzymes with and without holding phosphate constant. Allowing plastidic phosphate to vary freely as a metabolite in the simulation is a proxy for the effects of the antiporter, since if phosphate was transported independently from the cytosol, there would be effectively no steady state requirement for phosphate—any deviation would simply be made up by transport to or from the cytosol. The inclusion of phosphate fixes the steady state requirement to the one-to-one antiport of G3P and inorganic phosphate.

70

Without P_i, the model was found to have noticeable instability in essentially all of the CBB enzymes, and, seven of the 12 CBB enzymes were noted to have instability upon increase. Inclusion of phosphate as a metabolite was shown to almost completely eliminate instability with one notable exception. With phosphate, triose phosphate isomerase was noticeably unstable to decrease. Without phosphate, that enzyme was unstable to increase—the inclusion of phosphate reversed the tendency towards instability (Fig. 5-2B, red lines).
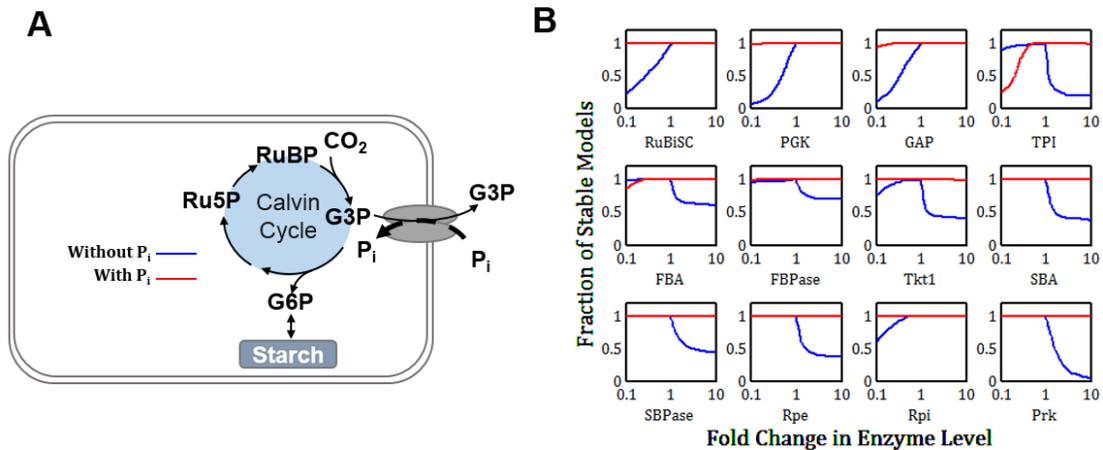
Fig. 5-2 Comparison of stability with and without holding phosphate constant. **A)** Schematic showing the flow of phosphate through the phosphate/G3P antiporter in relation to the CBB pathway. **B)** EMRA stability profile for the enzymes of the CBB pathway upon perturbation of 10x and 0.1x. Both Tkt reactions were perturbed simultaneously (n = 300). Including the effects of the G3P/phosphate antiporter (red line) significantly stabilizes the pathway.

5.4 Glutathionylation of triose phosphate isomerase improves stability (with phosphate)

The one enzyme of the CBB pathway which was unstable after the inclusion of phosphate was triose-phosphate isomerase. There seems to be further experimental confirmation of the importance of sufficient triose phosphate isomerase (TPI) activity. A plastidic TPI mutant with reduced activity was installed in *Arabidopsis* and the resulting plants were found to grow at a highly stunted rate [120]. Interestingly, if grown in the dark with nutrients provided (heterotrophic growth), there was no growth deficiency, indicating that the plastidic TPI is important for autotrophic (light) metabolism, but not critical for heterotrophic (dark) metabolism.

There are multiple possible methods for accommodating this loss of stability. First, triose-phosphate isomerase is a highly active, reversible enzyme with no stability penalty indicated (Fig. 2B) for high activity, so it's possible TPI is operating mostly or exclusively in the high activity domain, where stability is not an issue. Another possibility is that TPI instability is partly rescued by the effect of glutathionylation. A recent analysis showed the first evidence of glutathionylation of plant enzymes. The authors found that a cytosolic TPI from *Arabidopsis thaliana* was inactivated in the presence of oxidized glutathione (GSSG) but reactivated in the presence of reduced glutathione [117]. Since GSH is regenerated by the reducing power of NADPH, this regulatory network can be represented as NADPH activation of TPI combined with NADP$^+$ inactivation (Fig. 5-3A).

Interestingly, the stability of the TPI with NADPH/NADP+ regulation improves noticeably (Fig. 5-3B). Although there is no direct evidence if glutathionylation of plastidic TPI (pdTPI) in Arabidopsis, the protein sequences show 62% sequence identity and have similar numbers of methionine residues (2 & 3), (UniProt entries Q9SKP6 & P48491 [121], aligned by BLASTP 2.3.0+ [122,123]) . Regulation of plastidic TPI may be an interesting area of future research.
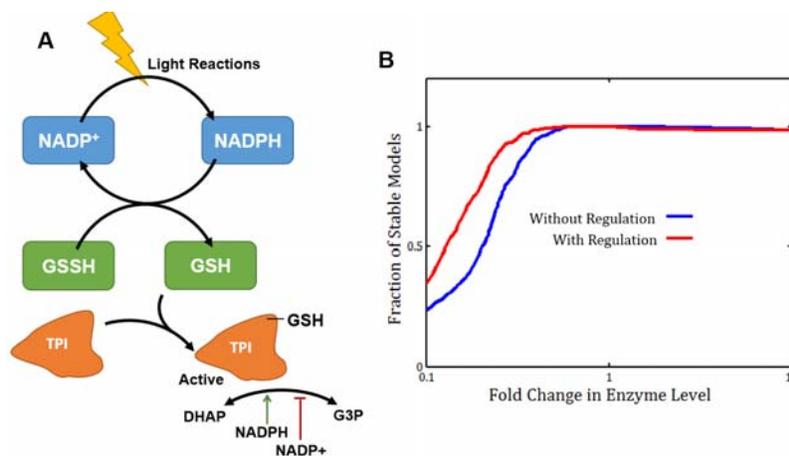
Fig. 5-3 Possible regulatory mechanism for plastidic TPI. **A)** Possible schematic for TPI activation by glutathionylation. This is represented in the model by NADPH activation and NADP$^+$ repression. **B)** Stability profile for the TPI enzyme in the 'with phosphate' model showing the effect of NAPDH regulation on TPI.

5.5 Glucose-6 phosphate shunt affects stability of no phosphate condition

A perhaps paradoxical aspect of the plastidic glucose 6 phosphate dehydrogenase enzyme is that it retains some activity after deactivation, which seems to be thermodynamically unfavorable, since carbon decarboxylated by the oxidative pentose phosphate pathway has to be re-fixed by RuBiSCO, including the use of 3 ATP per carbon fixed. It has recently been proposed that this is a feature of chloroplastic metabolism which may stabilize the CBB pathway itself [113]. This was discussed in great detail but so far has seen no mathematical justification. Looking at the CBB with stabilization by the phosphate translocator, there is little stability improvement to be made. In stress conditions, however, such as phosphate limitation, plant metabolism is known to change radically [124–126], including changing expression of plastidic transporters [127]. This could potentially alter the stabilizing, protective effects of the phosphate/G3P antiporter, which can be modelled (as before) by the removal of phosphate as a metabolite. In such cases, other structural features would be required to provide stability.

The so-called glucose-6 phosphate shunt (Fig. 5-4A) has been proposed to provide stability to the CBB. To test the effects of the proposed glucose-6 phosphate shunt, simulation of the no-phosphate condition with various levels (0% of RubisCO, 10%, 30%) of flux through the first three enzymes of the oxidative pentose phosphate pathway (G6PDH, GLNase & GLNDH) was undertaken via EMRA. For nearly all enzymes, the 10% & 30% conditions showed stability improvements over the 0% condition for increases in enzyme activity from the reference steady state (Fig. 4B, red & green lines). However, interestingly, several enzymes showed slightly higher instability in the 10% and 30% conditions upon *decrease* in enzyme amount, though higher stability upon increase. One possible explanation is that transketolase, and the aldolases are highly active, reversible enzymes, and thus more likely to operate in the high activity regime than the low

activity. Another possible explanation is that the operation of the G6P shunt is situational, and that it is meant to operate in dynamic scenarios to replenish cycle intermediates, rather than to operate continuously to maintain steady state.
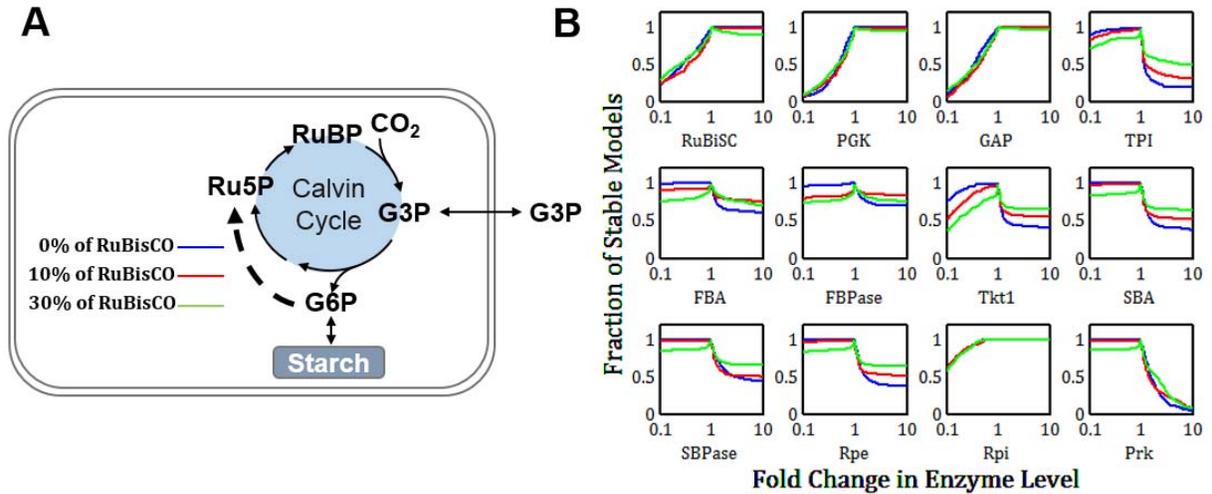
Fig. 5-4 Comparison of stability of various fluxes through the proposed G6P shunt. **A)** Schematic showing the flow of metabolites through the G6P shunt relative to the CBB pathway. **B)** EMRA stability profile for the enzymes of the CBB pathway upon perturbation of 10x and 0.1x. Both Tkt reactions were perturbed simultaneously (n = 300). Including the effects of the G6P shunt (red line & green lines) improves stability of the pathway upon increase of many enzymes, hurts stability upon decrease of many enzymes (Tkt, aldolases, phosphatases particularly).

5.6 Assessing methods for improving plant productivity, SBPase and RuBiSCO overexpression

Use of stability analysis to provide biological insight into the mechanisms of stability in the CBB is one powerful demonstration of its capabilities. However, it doesn't provide insight into engineering and biotechnological efforts which are aimed at increasing the productivity of plants, particularly relating to growth rate and the CO2-fixing rate of the CBB pathway. Thus, in addition to assessing the effect of genetic changes on stability, we can additionally look at the predicted impact on net carbon fixation rate.

Many efforts to increase growth rate and carbon fixation rate of plants have understandably focused on RuBiSCO. Some projects have focused on methods to modify the amino acid sequence of RuBiSCO [128,129]. Others have attempted to overexpress RuBiSCO or, more recently, replace

native RuBiSCO with a heterologous enzyme which has higher specific activity [130]. These efforts have increased the content and activity of RuBiSCO, but they have not convincingly increased plant productivity [131]. However, looking at the CBB pathway as a network problem rather than a problem with a single enzyme opens up many different possibilities. Interestingly, one group reported that overexpression of SBPase increased carbon fixation rate by 6-12% [132].

To investigate consistency of these results with simulation, the average model-predicted net CO2-fixation rate for different genetic changes and flux configurations can be compared. Interestingly, results show that for the 0% G6PDH condition with phosphate, overexpression of SBPase slightly increased CO2-fixation rate, while RuBiSCO overexpression was, counterintuitively, found to decrease RuBiSCO flux. Other targets in the CBB pathway which were also investigated, with Prk showng the largest projected increase on carbon fixation rate (Fig. 5-5A). For other conditions (no phosphate, with G6PDH flux) (Fig. 5-5B & C), no improvement was observed for either, except a small improvement for SBPase in the no phosphate model. This suggests that perhaps network effects are more determinative of the response of the CBB pathway than performance of individual enzymes. Additionally, it seems to suggest that in laboratory conditions, the models not including G6P shunt flux are more reflective of biological reality, and thus that the role of the G6P shunt may be situational.
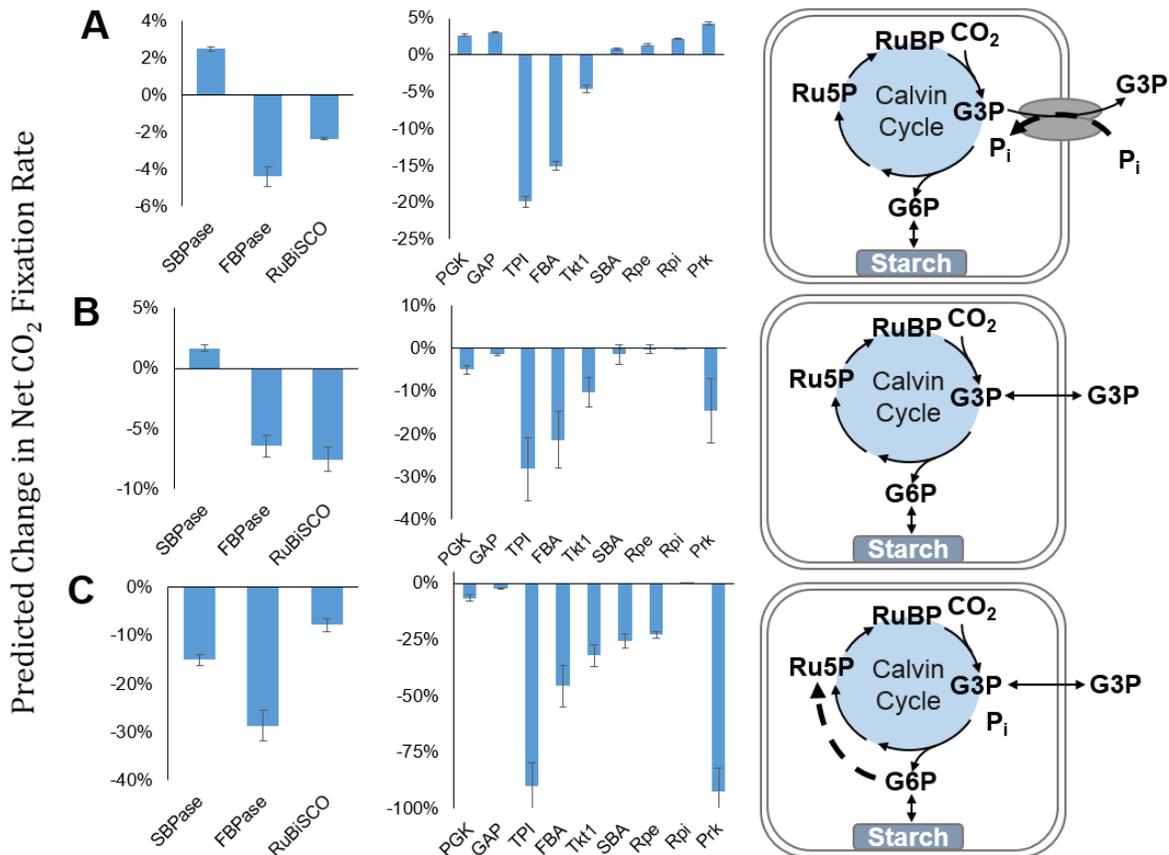
Fig. 5-5 Figure showing predicted effect of 10x increase of various CBB enzymes for three different flux models. SEM shown (n = 300). **A)** Allowing plastidic phosphate to vary freely. **B)** Holding phosphate constant. **C)** With G6P shunt at 30% of RuBiSCO flux.

5.7 Discussion

This analysis reveals the importance of structural features for the stability of the CBB pathway in plants. Stability is an important characteristic of metabolic pathways, since they are subject to stochastic variability in protein expression as well as different environmental conditions which can perturb the system. While oscillations in the CBB are a point of previous research [116], we here present an analysis of the stability of the underlying fixed points involved. So far, stability, and in particular the ensemble modeling robustness analysis framework has been applied to explore the performance relatively simple *in vitro* pathways, but this paper shows how it can also uncover and illuminate biologically significant features and phenomena.

Additionally, this manuscript sheds additional light on some specific details of these mechanisms. For instance, these results indicate that the G3P/phosphate antiporter is more significant for the stability of the CBB than the G6P shunt under normal steady state. However, if the one-to-one link between phosphate- and G3P-transport is broken (as in the no phosphate simulations), the action of the glucose-6-phosphate shunt does change the stability profile of CBB enzymes noticeably. However, the true purpose of the G6P shunt may be to restore steady state in dynamic situations. This sheds light on the apparent paradox of thermodynamic losses in this 'futile' cycle. The thermodynamic involved in one turnover of the oPPP would be involve loss of one ATP in the Prk step and two ATP at the Pgk step.

Among heterotrophic organisms using the CBB cycle, there is a remarkable amount of diversity in the arrangement and function of metabolism [133–137]. Thus, it is likely that depending on environmental constraints and chance occurrences in evolutionary history, the stabilizing mechanisms used by different species are a combination of those presented here and those yet to be discovered. Thus, this manuscript is not a comprehensive or conclusive look at the mechanisms of stability in the CBB pathway but is an initial, provisional investigation into some possible explanations for the success of the CBB pathway despite its apparently unstable underlying structure. The model presented here is advances on some previously described models [114–116] in important ways. This work provides answers and more questions to pave the way for yet more complete and sophisticated simulation of CBB.

So far, attempts to increase the productivity of plants have mostly focused on individual enzymes, rather than investigating the CBB pathway as a network. Here, we give plausible explanation to results that show SBPase overexpression increases plant growth rate while RuBiSCO overexpression has so far not shown any increase in plan performance. While the

methods employed here are not conclusive, they provide new insights which lay out potential targets of future exploration in the biotechnological engineering of plants.

5.8 Methods

The model of chloroplast metabolism, including the CBB, the G3P/phosphate translocator and the was constructed by inspecting the latest literature about plastid metabolism[95]. The full stoichiometric matrix, reversibilities and reference flux are shown in Supplementary Table 1. Adjustments were made as necessary (removal of phosphate, adjustment of fluxes to include G6P shunt etc.). Based on stoichiometry and reversibility, realistic Michaelis-Menten style rate laws were assigned. Regulation of PGM, G6PDH were included and TPI was regulated in some simulations. Parameters were obtained by randomly sampling normalized affinity parameters from a uniform distribution (0.1,10) as described previously. V_{max} was then solved for, constraining the rate law to the reference steady state. Simulations of steady state perturbations were carried out using the parameter continuation method described previously [13]. Calculations were done in MATLAB and full code is at: https://github.com/theis188/CBB-theisen.

6 Future work

6.1 User-friendly web based EMRA simulation tool

The development of a user-friendly EMRA simulation tool is an important next step for the expanding its impact as a simulation tool. Although EMRA simulation requires some fairly sophisticated mathematical operations (matrix multiplication, linear programming, numerical integration, differentiation, calculating eigenvalues, etc.), the required input is minimal, and may be further minimized by the generation of curated reaction sets. In many cases, picking the relevant enzymes from a pre-constructed list would be enough. Further, the javascript library NumericJS, (created by Sebastian Loisel) allows all these operations to be carried out by a web browser. Work will proceed using SEASNet-provided hosting, with other options being pursued if more sophisticated hosting needs arise.

6.2 Evolution of E. coli using mutD5 to consume methanol/induce methylotrophy

Formaldehyde tolerance, methanol consumption and methylotrophy have the potential to be significantly enhanced by the use of laboratory enhanced evolution techniques. One strategy that has great promise is the use of mutator strains such as mutD5-containing strains. MutD5 is a mutated copy of dnaQ which is a subunit of DNA Polymerase with proofreading function and dominant mutagenesis activity[138–140]. MutD5 strains have been shown to have their highest mutation rates in early stages when other mechanisms for mutation correction are saturated [141,142].

If the correct selection pressure is provided, a mutD5 strain of *E. coli* may undergo evolution at an accelerated rate and adapt to a nutrient-poor methanol-containing environment. If limited other sources of carbon are provided, (e.g. dilute LB or glucose), then the cells would benefit if they are able to utilize the carbon and reducing power locked up in methanol. Thus,

mutations that facilitate this ability would be enriched in the culture and would eventually dominate.

6.2.1 Approaches for enhanced laboratory evolution

It may be the case that growth-promoting mutations do not arise in the first round attempts. Thus, it will be necessary to increase the probability of finding beneficial mutations. To facilitate faster mutational iteration, I have proposed a system of mutation accumulation and enrichment termed GOGOGO (**G**oal-**O**riented, **G**enetic **O**ptimization for **G**rowth on **O**ne-carbon compounds). This system is a multi-reactor system in which selection pressure for methylotrophy and methanol consumption are applied at different levels. At low selection pressure (i.e. relatively LB-rich and lower methanol), growth will be very fast and mutations will accumulate rapidly. However, they will not necessarily be enriched very quickly since the selection pressure would be relatively low. To accelerate the process of selection, cells from the low selection pressure reactors can be inoculated into higher selection pressure (dilute LB, higher methanol) reactors. Thus, the beneficial mutations which had accumulated in the fast-growth reactor would be enriched. This would likely not be accomplished as quickly using only high selection pressure reactors since growth of strains is slow and mutation accumulation, especially using mutD5 will be low. This is because at low growth rate, the other proofreading mechanisms of *E. coli* will be active, reducing the overall mutation rate.

GoGoGo

- Less Selective
- Faster Growth
- Accumulation of Mutations

- More Selective
- Slower Growth
- Enrichment of Beneficial Mutations
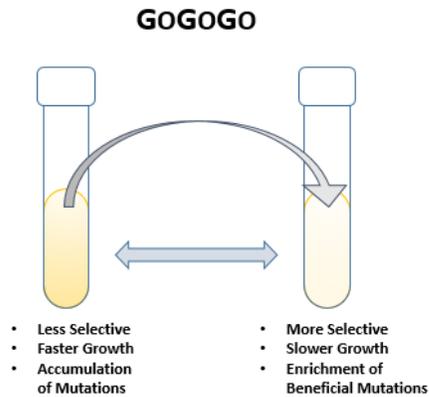
Fig. 6-1) GOGOGO for the generation of a methylotrophic E. coli strain.

6.2.2 Inspiration from natural methylotrophs

Natural methylotrophs utilize methanol by a variety of mechanisms. For example, the serine and RuMP pathways are used by different classes of methylotrophs. The RuMP pathway itself has multiple variations that appear in different organisms. For example, obligate methylotrophs use a version of the pathway that is similar to MCC, with the difference being the presence of Fba and the absence of Fpk. Facultative methylotrophs use a version of the RuMP pathway which is ATP-dependent [5].

Implementing methylotrophy in *E. coli* can take inspiration from natural pathways. RuMP is a natural candidate because of its overlap with both the MCC pathway and the pentose phosphate pathway in *E. coli*. Implementation of RuMP in *E. coli* requires enzyme expression to allow for flux throuhg the RuMP pathway. In natural methylotroph *Bacillus methanolicus* MGA3, enzyme expression is heavily influenced by the presence of methanol. In the presence of methanol, enzymes of the RuMP pathway of MGA3 are heavily overexpressed, with transcripts becoming 6-40x more abundant in the presence of methanol.[16]

Additionally, the enzyme substrate specificity and allosteric regulation between *E. coli* and *Bacillus methanolicus* enzymes may be different. For example, GlpX is an enzyme which catalyzes the removal of phosphate from sugar bispohosphates. MGA3 has two versions of the

enzyme, one of which has activity on seven-carbon sedoheptulose-1,7-bisphosphate (SBP) and is overexpressed in response to the presence of methanol. The second one which has activity only on fructose-1,6-bisphosphate (FBP).[143] This indicates that substrate specificity may play an important role in the RuMP pathway of MGA3. Another enzyme of possible concern is phosphofructokinase. Phosphofructokinase in *E. coli* is regulated allosterically by AMP, which may be detrimental to the performance of the RuMP pathway.[144]

6.2.3 Strategies for Implementation of methylotrophy and MCC in *E. coli*

Implementation of methanol assimilation and carbon shuffling will in theory allow for *E. coli* growth on methanol. However, it is unlikely that an efficient methylotrophic *E. coli* strain can be designed a priori. Combinatorial expression of enzymes has been used as an effective tool in many biosynthetic endeavors[145–148], however, it requires labor-intensive library generation and screening, and when selection is available as an option it is preferred. Therefore, it will be useful to construct an MCC production strain in multiple steps, optimizing each one separately, and the entire pathway as a whole. For example, implementation of assimilation and carbon shuffling in *E. coli* can be optimized first using selection for methanol growth. After that, combinatorial expression of key enzymes in the full n-butanol production pathway may provide a means of maximizing production.

Taking inspiration from previous combinatorial construction and screening efforts, MCC production can be optimized. For example, the production of isoprenoids such as lycopene has been boosted by combinatorial expression of enzymes.[149] Additionally, using different promoter strengths increased the production of taxol, an important anti-cancer precursor, several-thousand fold.[146]

6.3 Use wild type *E. coli* fluxes to generate *in silico* ensembles and subject them to installation of methylotrophy & identify targets for knockout and overexpression to support methylotrophy

Wild type models of *E. coli* will be a useful starting point for the generation of targets for methylotrophy. In stoichiometric models, it is typical to provide a reaction for the production of biomass, which may be maximized to approximate wild type steady state fluxes.[150] From this steady state, an ensemble of kinetic parameters may be generated without prior knowledge of kinetic properties.[10]
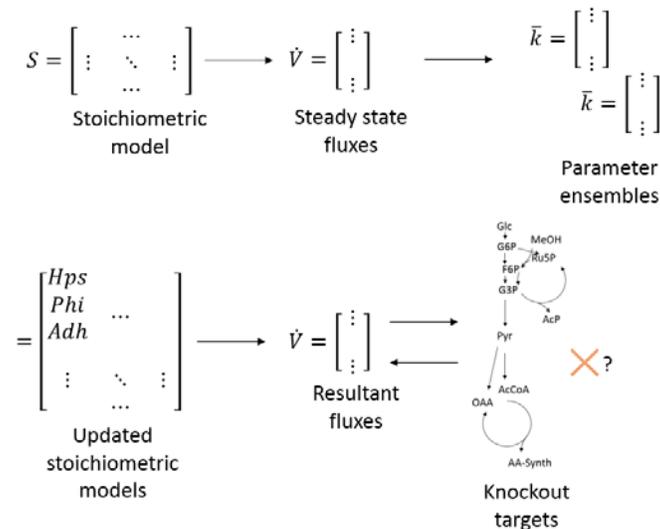
Fig. 6-2) The process by which ensemble models will be used to generate knockout (or overexpression) targets to generate methylotrophic *E. coli*.

After the generation of this ensemble, addition of methylotrophy reactions can be provided to identify how their addition affects intracellular flow of metabolites, especially the production of biomass. Finally, knockouts of many genes can be simulated to identify what knockouts may enhance the methylotrophic performance of the 'methylotrophic' *E. coli* strain. Further insights might be gained by examination of metabolism of methylotrophs, which mostly cannot grow on multi-carbon compounds.[151] Until recently, computational methods for the identification of knockout and overexpression targets have used mostly stoichiometric, rather than kinetic, models

of cellular metabolism.[19,152] Recently, kinetic parameters are starting to be used in these models.[153] Additionally, ensemble modeling has also been used to engineer strains for production by analyzing productivity rather than robustness.[10,154]

6.4 Understanding and inventing metabolic cycles

Finding new metabolic cycles to improve over existing pathways is an area of increasing research interest. In addition to MCC, many other efforts to design better synthetic routes are under development. A recent novel cycle, non-oxidative glyxolysis (NOG) allows for conversion of one glucose to three acetic acids[155], compared to the typical two through normal metabolism[156]. Also, another synthetic pathway for carbon dioxide fixation, reverse glyoxylate shunt (rGS) has been proposed and its pieces have been functionally demonstrated in *E. coli*[157]. Identifying previously unknown cycles in existing organisms has also recently been an area of importance, with new carbon fixation and metabolic cycles being identified in the last decade or two[98,158–161]. Understanding what features unify all metabolic cycles may lead towards yet more discoveries.

6.5 Robustness as a ranking characteristic for metabolic cycles

Another goal is to develop an algorithm for identifying and ranking novel cycles. A previous approach had the same goal and used KEGG, a database of many known enzymes, to search for alternative carbon fixation pathways[162,163]. Pathways were ranked in terms of productivity per enzyme mass and thermodynamic feasibility. Additionally, another criteria 'Topological Compatibility' was used which investigated the pathways in a flux balance analysis (FBA) framework, where a model of cellular metabolism was compared before and after institution of the novel pathway. Differences in flux distribution were quantified and smaller differences confered a higher rank.

However, the analysis is limited in multiple ways. First, only CO_2 fixation pathways were considered, while other metabolic cycles that may be of interest were not investigated. Secondly, FBA analysis is not based on a kinetic understanding of metabolism, but only stoichiometric. Installation of the enzymes of a pathway, even if they are functionally expressed, does not guarantee that flux will go in the directions hoped. For metabolic cycles, the problem is more pronounced, since in many cases, depletion of cycle intermediates will halt cycle function[21]. Thus, ranking of cycles will be refined by consideration of kinetic limitations. Addition of this consideration may be accomplished by the use of ensemble model robustness analysis (EMRA)[13]. EMRA uses knowledge of functional forms of enzyme flux equations to determine bifurcational stability of a metabolic system to perturbation of parameters.

Appendix. Enzyme and compound names

Enzyme names: Mdh = methanol dehydrogenase; Hps = 3-hexuolse-6-phospate synthase; Phi = phosphohexulose isomerase; Fpk = phosphoketolase (F6P activity); Xpk = phosphoketolase (X5P activity); Tal = transaldolase; Tkt = transketolase; Rpe = D-ribulose-5-phosphate 3-epimerase; Rpi = ribose-5-phosphate isomerase; PduP = acylating aldehyde dehydrogenase; Adh = alcohol dehydrogenase; Glk = glucokinase ; Zwf = glucose-6-phosphate dehydrogenase; Pgi = glucose-6-phosphate isomerase. Compound names: CH2O = formaldehyde; H6P = 3-hexulose-6-phosphate; F6P = fructose-6-phosphate; E4P = erythrose-4-phosphate; S7P = sedoheptulose-7-phosphate; X5P = xylulose-5-phosphate; R5P = ribose-5-phosphate; Ru5P = ribulose-5-phosphate; AcP = acetyl phosphate; EtOH = ethanol.

References

1. US Department of Energy. Funding Opportunity Announcement - Advanced Research Projects Agency – Energy (ARPA-E) - Reducing Emissions Using Methanotrophic Organisms For Transportation Energy (REMOTE). (2013).

2. Maverick Synfuels Introduces Maverick Oasis^{TM} – Affordable Small-Scale Methane Gas-

to-Liquid Modular Methanol Plants. *Maverick Synfuels* (2014).

3. Kozlowski, J. T. & Davis, R. J. Heterogeneous Catalysts for the Guerbet Coupling of Alcohols. *ACS Catal.* **3,** 1588–1600 (2013).

4. Fox, J. R. Formation of Higher Alcohols from Methanol Metal Acetylides in the Presence of Although catalytic processes for the con- version of synthesis gas and / or methanol to useful hydrocarbons and oxygenates are well known , the mechanistic details of many of t. *J. Catal.* **138,** 127–138 (1984).

5. Anthony, C. *The biochemistry of methylotrophs*. *Comp. Biochem. Physiol. Part A Physiol.* **75,** (Academic Press, 1983).

6. Bogorad, I. W. *et al.* Building carbon-carbon bonds using a biocatalytic methanol condensation cycle. *Proc. Natl. Acad. Sci.* **111,** 15928–15933 (2014).

7. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28,** 245–248 (2010).

8. Liebermeister, W., Uhlendorf, J. & Klipp, E. Modular rate laws for enzymatic reactions: Thermodynamics, elasticities and implementation. *Bioinformatics* **26,** 1528–1534 (2010).

9. Holmberg, A. On the practical identifiability of microbial growth models incorporating Michaelis-Menten type nonlinearities. *Math. Biosci.* **62,** 23–43 (1982).

10. Tran, L. M., Rizk, M. L. & Liao, J. C. Ensemble modeling of metabolic networks. *Biophys. J.* **95,** 5606–17 (2008).

11. Tan, Y. & Liao, J. C. Metabolic ensemble modeling for strain engineers. *Biotechnol. J.* **7,** 343–53 (2012).

12. Müller, J. E. N. *et al.* Engineering Escherichia coli for methanol conversion. **28,** 190–201 (2015).

13. Lee, Y., Lafontaine Rivera, J. G. & Liao, J. C. Ensemble Modeling for Robustness Analysis in engineering non-native metabolic pathways. *Metab. Eng.* 1–9 (2014). doi:10.1016/j.ymben.2014.06.006

14. Ye, X. *et al.* Synthetic metabolic engineering-a novel, simple technology for designing a chimeric metabolic pathway. *Microb. Cell Fact.* **11,** 120 (2012).

15. Opgenorth, P. H., Korman, T. P. & Bowie, J. U. A synthetic biochemistry molecular purge valve module that maintains redox balance. *Nat. Commun.* **5,** 4113 (2014).

16. Jakobsen, Ø. M. *et al.* Upregulated Transcription of Plasmid and Chromosomal Ribulose Monophosphate Pathway Genes Is Critical for Methanol Assimilation Rate and Methanol Tolerance in the Methylotrophic Bacterium Bacillus methanolicus. *J. Bacteriol.* (2006). doi:10.1128/JB.188.8.3063

17. Egli, T., Bosshard, C. & Hamer, G. Simultaneous utilization of methanol-glucose mixtures by Hansenula polymorpha in chemostat: Influence of dilution rate and mixture composition on utilization pattern. *Biotechnol. Bioeng.* **28,** 1735–1741 (1986).

18. Lynd, L. H. & Zeikus, J. G. Metabolism of H2-CO2, methanol, and glucose by

Butyribacterium methylotrophicum. *J. Bacteriol.* **153,** 1415–1423 (1983).

19. Burgard, A. P., Pharkya, P. & Maranas, C. D. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **84,** 647–57 (2003).

20. Yevenes, A. & Frey, P. a. Cloning, expression, purification, cofactor requirements, and steady state kinetics of phosphoketolase-2 from Lactobacillus plantarum. *Bioorg. Chem.* **36,** 121–7 (2008).

21. Owen, O. E., Kalhan, S. C. & Hanson, R. W. The key role of anaplerosis and cataplerosis for citric acid cycle function. *J. Biol. Chem.* **277,** 30409–12 (2002).

22. Jitrapakdee, S. *et al.* Structure, mechanism and regulation of pyruvate carboxylase. *Biochem. J.* **413,** 369–87 (2008).

23. Arrowsmith, D. in *Dyn. Syst.* (Chapman & Hall, 1992).

24. Kien, C., Chang, D., Murray, R., Ailabouni, A. & Kepner, J. Measurement of Stable Isotopic Enrichment of Underivatized Acetate by Gas Chromatography/Mass Spectrometry: Application to in vivo Estimation of Acetate Production. *Biomed. Environ. Mass Spectrom.* **19,** 554–558 (1990).

25. NIST/EPA/NIH Mass Spectral Library. Entry Number 227635.

26. Frankel-Conrat, H. & Olcott, H. S. The Reaction of Formaldehyde with Proteins. *J. Am. Chem. Soc.* **70,** 2673–2684 (1948).

27. Quievryn, G. & Zhitkovich, A. Loss of DNA – protein crosslinks from formaldehyde-exposed cells occurs through spontaneous hydrolysis and an active repair process linked to proteosome function. *Carcinogenesis* **21,** 1573–1580 (2000).

28. Heck, H. D., Casanova, M. & Starr, T. B. Formaldehyde toxicity--new understanding. *Crit. Rev. Toxicol.* **20,** 397–426 (1990).

29. Kaulfers, P. & Brandt, D. Isolation of a conjugative plasmid in Escherichia coli determining formaldehyde resistance. *FEMS Microbiol. Lett.* **43,** 161–163 (1987).

30. Kümmerle, N., Feucht, H.-H. & Kaulfers, P.-M. Plasmid-mediated formaldehyde resistance in Escherichia coli: characterization of resistance gene. *Antimicrob. Agents Chemother.* **40,** 2276–2279 (1996).

31. Orita, I., Sakamoto, N., Kato, N., Yurimoto, H. & Sakai, Y. Bifunctional enzyme fusion of 3-hexulose-6-phosphate synthase and 6-phospho-3-hexuloisomerase. *Appl. Microbiol. Biotechnol.* **76,** 439–445 (2007).

32. Patel, R. N., Hou, C. T. & Derelanko, P. Microbial Oxidation of Methanol : Purification and Properties of Formaldehyde Dehydrogenase from a Pichia sp . NRRL-Y-11328 purified to homogeneity from soluble extracts of methanol-grown yeast. *Achives Biochem. Biophys.* **221,** 135–142 (1983).

33. Ellis, S. B. *et al.* Isolation of Alcohol Oxidase and Two Other Methanol Regulatable Genes from the Yeast Pichia pastoris. *Mol. Cell. Biol.* **5,** 1111 (1985).

34.   Atsumi, S. *et al.* Evolution, genomic analysis, and reconstruction of isobutanol tolerance in Escherichia coli. *Mol. Syst. Biol.* **6,** 449 (2010).

35.   Keseler, I. M. *et al.* EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* **41,** D605–12 (2013).

36.   Theisen, M. K., Lafontaine Rivera, J. G. & Liao, J. C. Stability of Ensemble Models Predicts Productivity of Enzymatic Systems. *PLOS Comput. Biol.* **12,** e1004800 (2016).

37.   Rapp, P. in *Chaos* 179–203 (1986).

38.   Hess, B., Boiteux, a & Krüger, J. Cooperation of glycolytic enzymes. *Adv. Enzyme Regul.* **7,** 149–167 (1969).

39.   Mees, a I. & Rapp, P. E. Periodic metabolic systems: oscillations in multiple-loop negative feedback biochemical control networks. *J. Math. Biol.* **5,** 99–114 (1978).

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

41.   Steuer, R. *et al.* Structural kinetic modeling of metabolic networks. *Proc. Natl. Acad. Sci.* **103,** 11868–11873 (2006).

42.   Bass, J. & Takahashi, J. S. Circadian integration of metabolism and energetics. *Science* **330,** 1349–1354 (2010).

43.   Lafontaine Rivera, J. G., Lee, Y. & Liao, J. C. An entropy-like index of bifurcational robustness for metabolic systems. *Integr. Biol.* (2015). doi:10.1039/C4IB00257A

44.   Voit, E. *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*. (Cambridge University Press, 2000).

45.   Voit, E. *Canonical Nonlinear Modeling*. (Chapman and Hall, 1991).

46.   Savageau, M. Biochemical systems analysis. I. Some mathematical properties of the rate law for the component enzymatic reactions. *J. Theor. Biol.* **25,** 365–369 (1969).

47.   Savageau, M. A. Development of fractal kinetic theory for enzyme-catalysed reactions and implications for the design of biochemical pathways. *BioSystems* **47,** 9–36 (1998).

48.   Savageau, M. Biochemical Systems Analysis. 379 (1976).

49.   Fell, D. a. Metabolic Control Analysis : experimental development of its theoretical and. *Biochem. J.* **330,** 313–330 (1992).

50.   Bogorad, I. W. *et al.* Building carbon-carbon bonds using a biocatalytic methanol condensation cycle. *Proc. Natl. Acad. Sci. U. S. A.* (2014). doi:10.1073/pnas.1413470111

51.   Liu, X., Kim, C. S., Kurbanov, F. T., Honzatko, R. B. & Fromm, H. J. Dual mechanisms for glucose 6-phosphate inhibition of human brain hexokinase. *J. Biol. Chem.* **274,** 31155–31159 (1999).

52.   Ettema, T. J. G., Ahmed, H., Geerling, A. C. M., van der Oost, J. & Siebers, B. The non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase (GAPN) of Sulfolobus solfataricus: a key-enzyme of the semi-phosphorylative branch of the Entner-Doudoroff pathway. *Extremophiles* **12,** 75–88 (2008).

53.   Renz, A. & Stitt, M. Substrate specificity and product inhibition of different forms of fructokinases and hexokinases in developing potato tubers. *Planta* **190,** 166–175 (1993).

54.   Nakamura, T., Kashima, Y., Mine, S., Oku, T. & Uegaki, K. Characterization and crystal structure of the thermophilic ROK hexokinase from Thermus thermophilus. *J. Biosci. Bioeng.* **114,** 150–154 (2012).

55.   van Heerden, J. H. *et al.* Lost in Transition: Start-Up of Glycolysis Yields Subpopulations of Nongrowing Cells. *Science (80-.).* **343,** 1245114–1245114 (2014).

56.   Sugano, M. & Ishiwaki, N. soya-bean protein on serum lipids , apolipoproteins , insulin. (1982).

57.   Paolisso, G., Scheen, a J., Albert, a & Lefebvre, P. J. Effects of pulsatile delivery of insulin and glucagon in humans. *Am. J. Physiol.* **257,** E686–E696 (1989).

58.   Hofmann, E. The significance of phosphofructokinase to the regulation of carbohydrate metabolism. *Rev. Physiol. Biochem. Pharmacol.* **75,** 1–68 (1976).

59.   Reeves, R. E. & Sols, A. Regulation of Escherichia coli Phoshpfructokinase in situ. *Biochem. Biophys. Res. Commun.* **50,** 459–466 (1973).

60.   Bar-Even, A., Flamholz, A., Noor, E. & Milo, R. Rethinking glycolysis: on the biochemical logic of metabolic pathways. *Nat. Chem. Biol.* **8,** 509–517 (2012).

61.   Meléndez-Hevia, E., Waddell, T. G., Heinrich, R. & Montero, F. Theoretical approaches to the evolutionary optimization of glycolysis--chemical analysis. *Eur. J. Biochem.* **244,** 527–543 (1997).

62.   Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W. & Milo, R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 10039–44 (2013).

63.   Krutsakorn, B. *et al.* In vitro production of n-butanol from glucose. *Metab. Eng.* **20,** 84–91 (2013).

64.   Liao Lab. Unpublished experiments.

65.   Schurmann, M. & Sprenger, G. a. Fructose-6-phosphate aldolase is a novel class I aldolase from Escherichia coli and is related to a novel group of bacterial transaldolases. *J. Biol. Chem.* **276,** 11055–61 (2001).

66.   Shen, C. R. & Liao, J. C. Metabolic engineering of Escherichia coli for 1-butanol and 1-propanol production via the keto-acid pathways. *Metab. Eng.* **10,** 312–20 (2008).

67.   Shen, C. R. *et al.* Driving forces enable high-titer anaerobic 1-butanol synthesis in Escherichia coli. *Appl. Environ. Microbiol.* **77,** 2905–15 (2011).

68.   Gribskov, M. & Burgess, R. R. Sigma factors from E. coli, B. subtilis, phage SP01, and phage T4 are homologous proteins. *Nucleic Acids Res.* **14,** 67456763 (1986).

69.   Vemuri, G. N., Altman, E., Sangurdekar, D. P., Khodursky, a B. & Eiteman, M. a. Overflow Metabolism in Escherichia coli during Steady-State Growth : Transcriptional Regulation and Effect of the Redox Ratio Overflow Metabolism in Escherichia coli during

Steady-State Growth : Transcriptional Regulation and Effect of the Redox Ratio †. *Appl. Environ. Microbiol.* **72,** 3653–3661 (2006).

70.   Tai, M. & Stephanopoulos, G. Engineering the push and pull of lipid biosynthesis in oleaginous yeast Yarrowia lipolytica for biofuel production. *Metab. Eng.* **15,** 1–9 (2013).

71.   Gombert, A. K. & Moreira, M. Network Identification and Flux Quantification in the Central Metabolism of *Saccharomyces cerevisiae* under Different Conditions of Glucose Repression Network Identification and Flux Quantification in the Central Metabolism of Saccharomyces cerev. *J. Bacteriol.* **183,** 1441–1451 (2001).

72.   Lamed, R., Setter, E. & Bayer, E. a. Characterization of a cellulose-binding, cellulase-containing complex in clostridium thermocellum. *J. Bacteriol.* **156,** 828–836 (1983).

73.   Krauss, J., Zverlov, V. V. & Schwarz, W. H. In vitro reconstitution of the complete Clostridium thermocellum cellulosome and synergistic activity on crystalline cellulose. *Appl. Environ. Microbiol.* **78,** 4301–4307 (2012).

74.   Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451,** 86–9 (2008).

75.   Weaver, D. *et al.* The EcoCyc Database. *EcoSal Plus* **1,** (2014).

76.   Lynd, L., Wyman, C. & Gerngross, T. Biocommodity Engineering. *Biotechnol. Prog.* **15,** 777–793 (1999).

77.   Lin, P. P. *et al.* Consolidated bioprocessing of cellulose to isobutanol using Clostridium thermocellum. *Metab. Eng.* **31,** 44–52 (2015).

78.   Kerkhoven, E. J., Pomraning, K. R., Baker, S. E. & Nielsen, J. Regulation of amino-acid metabolism controls flux to lipid accumulation in Yarrowia lipolytica. *npj Syst. Biol. Appl.* **in press,** 1–7 (2016).

79.   Paiva, S., Devaux, F., Barbosa, S., Jacq, C. & Casal, M. Ady2p is essential for the acetate permease activity in the yeast Saccharomyces cerevisiae. *Yeast* **21,** 201–210 (2004).

80.   Wasylenko, T. M., Ahn, W. S. & Stephanopoulos, G. The oxidative pentose phosphate pathway is the primary source of NADPH for lipid overproduction from glucose in Yarrowia lipolytica. *Metab. Eng.* **30,** 27–39 (2015).

81.   Bricker, D. K. *et al.* A Mitochondrial Pyruvate Carrier Required for Pyruvate Uptake in Yeast, Drosophila, and Humans. *Science (80-.).* **337,** 96–100 (2012).

82.   Fontanille, P., Kumar, V., Christophe, G., Nouaille, R. & Larroche, C. Bioconversion of volatile fatty acids into lipids by the oleaginous yeast Yarrowia lipolytica. *Bioresour. Technol.* **114,** 443–9 (2012).

83.   Gombert, A. K., Moreira dos Santos, M., Christensen, B. & Nielsen, J. Network identification and flux quantification in the central metabolism of Saccharomyces cerevisiae under different conditions of glucose repression. *J. Bacteriol.* **183,** 1441–51 (2001).

84.   Raines, C. a. The Calvin cycle revisited. *Photosynth. Res.* **75,** 1–10 (2003).

85.   Yamanaka, T. in *Chemolithoautotrophic Bact.* 103–117 (2008). doi:10.1007/978-4-431-78541-5

86.   Bassham, J. & Calvin, M. in *Biog. Nat. Subst.* 72 (1960).

87.   Hatch, M. D. C(4) photosynthesis: discovery and resolution. *Photosynth. Res.* **73,** 251–6 (2002).

88.   Smirnoff, N. *Environment and plant metabolism: flexibility and acclimation.* (Bios Scientific Publishers, 1995).

89.   Hasegawa, P. M., Bressan, R. A., Zhu, J.-K. & Bohnert, H. J. PLANT CELLULAR ANDMOLECULAR RESPONSES TOHIGH SALINITY. *Annu Rev Plant Physiol Plant Mol Biol* **51,** 463–99 (2000).

90.   Slesak, I., Libik, M., Karpinska, B., Karpinski, S. & Miszalski, Z. The role of hydrogen peroxide in regulation of plant metabolism and cellular signalling in response to environmental stresses. *Acta Biochim. Pol.* **54,** 39–50 (2007).

91.   Plaxton, W. C. The Organization and Regulation of Plant Glycolysis. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47,** 185–214 (1996).

92.   Kauss, H. SOME ASPECTS OF CALCIUM-DEPENDENT REGULATION IN PLANT METABOLISM H. *Ann Rev Plant Physiol* **38,** 47–72 (1987).

93.   GIBON, Y. *et al.* Adjustment of growth, starch turnover, protein content and central metabolism to a decrease of the carbon supply when Arabidopsis is grown in very short photoperiods. *Plant. Cell Environ.* **32,** 859–874 (2009).

94.   Gibon, Y. *et al.* Adjustment of diurnal starch turnover to short days: depletion of sugar during the night leads to a temporary inhibition of carbohydrate utilization, accumulation of sugars and post-translational activation of ADP-glucose pyrophosphorylase in the followin. *Plant J.* **39,** 847–862 (2004).

95.   Stitt, M. & Zeeman, S. C. Starch turnover: pathways, regulation and role in growth. *Curr. Opin. Plant Biol.* **15,** 282–292 (2012).

96.   Tcherkez, G. G. B., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 7246–7251 (2006).

97.   Galmés, J. *et al.* Rubisco catalytic properties optimized for present and future climatic conditions. *Plant Sci.* **226,** 61–70 (2014).

98.   Bar-Even, A., Noor, E. & Milo, R. A survey of carbon fixation pathways through a quantitative lens. *J. Exp. Bot.* **63,** 2325–42 (2012).

99.   Bowsher, C., Steer, M. & Tobin, A. in *Plant Biochem.* 500 (2008).

100.   Flügge, U.-I. Phosphate Translocators in Plastids. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **50,** 27–45 (1999).

101.   Weber, A. P. M., Schwacke, R. & Flügge, U.-I. Solute Transporters of the Plastid Envelope Membrane. *Annu. Rev. Plant Biol.* **56,** 133–164 (2005).

102. Trevisan, M., Chen, J., Smith, A. M. & Zeeman, S. C. A Previously Unknown Maltose Transporter Essential for Starch Degradation in Leavs. *Science (80-.).* **303,** 87–90 (2004).

103. Balmer, Y. *et al.* Proteomics gives insight into the regulatory function of chloroplast thioredoxins. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 370–375 (2003).

104. Michelet, L. *et al.* Redox regulation of the Calvin–Benson cycle: something old, something new. *Front. Plant Sci.* **4,** 1–21 (2013).

105. Wedel, N. & Soll, J. Evolutionary conserved light regulation of Calvin cycle activity by NADPH-mediated reversible phosphoribulokinase/CP12/ glyceraldehyde-3-phosphate dehydrogenase complex dissociation. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 9699–9704 (1998).

106. Winkenbach, F. & Wolk, C. P. Activities of enzymes of the oxidative and the reductive pentose phosphate pathways in heterocysts of a blue-green alga. *Plant Physiol.* **52,** 480–483 (1973).

107. Averill, R. H., Bailey-Serres, J. & Kruger, N. J. Co-operation between cytosolic and plastidic oxidative pentose phosphate pathways revealed by genotypes of maize. *Plant J.* **14,** 449–457 (1998).

108. Kruger, N. J. & von Schaewen, A. The oxidative pentose phosphate pathway: structure and organisation. *Curr. Opin. Plant Biol.* **6,** 236–246 (2003).

109. Wenderoth, I., Scheibe, R. & Von Schaewen, A. Identification of the cysteine residues involved in redox modification of plant plastidic glucose-6-phosphate dehydrogenase. *J. Biol. Chem.* **272,** 26985–26990 (1997).

110. Hauschild, R. & von Schaewen, A. Differential regulation of glucose-6-phosphate dehydrogenase isoenzyme activities in potato. *Plant Physiol.* **133,** 47–62 (2003).

111. Cardi, M. *et al.* Overexpression, purification and enzymatic characterization of a recombinant plastidial glucose-6-phosphate dehydrogenase from barley (Hordeum vulgare cv. Nure) roots. *Plant Physiol. Biochem.* **73,** 266–73 (2013).

112. Née, G., Zaffagnini, M., Trost, P. & Issakidis-Bourguet, E. Redox regulation of chloroplastic glucose-6-phosphate dehydrogenase: A new role for f-type thioredoxin. *FEBS Lett.* **583,** 2827–2832 (2009).

113. Sharkey, T. D. & Weise, S. E. The glucose 6-phosphate shunt around the Calvin – Benson cycle. (2015). doi:10.1093/jxb/erv484

114. Grimbs, S. *et al.* Spatiotemporal dynamics of the Calvin cycle: Multistationarity and symmetry breaking instabilities. *BioSystems* **103,** 212–223 (2011).

115. HAHN, B. D. A Mathematical Model of the Calvin Cycle: Analysis of the Steady State. *Ann. Bot.* **57,** 639–653 (1986).

116. Laisk, A. & Walker, D. Control of phosphate turnover as a rate-limiting factor and possible cause of oscillations in photosynthesis: a mathematical model. *Proc R Soc L. B* **227,** 281–302 (1986).

117. Ito, H., Iwabuchi, M. & Ogawa, K. The sugar-metabolic enzymes aldolase and triose-

phosphate isomerase are targets of glutathionylation in Arabidopsis thaliana: Detection using biotinylated glutathione. *Plant Cell Physiol.* **44,** 655–660 (2003).

118. Stitt, M. & Heldt, H. W. Simultaneous synthesis and degradation of starch in spinach chloroplasts in the light. *Biochim. Biophys. Acta (BBA)-Bioenergetics* **638,** 13–16 (1981).

119. Rao, I. in *Handb. Photosynth.* (Pessarakli, M.) 173–194 (Marcel Dekker, 1997).

120. Chen, M. & Thelen, J. J. The Plastid Isoform of Triose Phosphate Isomerase Is Required for the Postgerminative Transition from Heterotrophic to Autotrophic Growth in Arabidopsis. *Plant Cell* **22,** 77–90 (2010).

121. The Uniprot Consortium. Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* **40,** D71–D75 (2012).

122. Altschul, S. F. *et al.* Protein database searches using compositionally adjusted substitution matrices. *FEBS J.* **272,** 5101–9 (2005).

123. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).

124. Rouached, H., Arpat, a. B. & Poirier, Y. Regulation of phosphate starvation responses in plants: Signaling players and cross-talks. *Mol. Plant* **3,** 288–299 (2010).

125. Secco, D. *et al.* Spatio-Temporal Transcript Profiling of Rice Roots and Shoots in Response to Phosphate Starvation and Recovery. *Plant Cell* **25,** 4285–4304 (2013).

126. Yang, X. J. & Finnegan, P. M. Regulation of phosphate starvation responses in higher plants. *Ann. Bot.* **105,** 513–526 (2010).

127. Wu, P. *et al.* Phosphate starvation triggers distinct alterations of genome expression in Arabidopsis roots and leaves. *Plant Physiol.* **132,** 1260–71 (2003).

128. Cai, Z., Liu, G., Zhang, J. & Li, Y. Development of an activity-directed selection system enabled significant improvement of the carboxylation efficiency of Rubisco. *Protein Cell* **5,** 552–562 (2014).

129. Parry, M. A. J. Manipulation of Rubisco: the amount, activity, function and regulation. *J. Exp. Bot.* **54,** 1321–1333 (2003).

130. Lin, M. T., Occhialini, A., Andralojc, P. J., Parry, M. A. J. & Hanson, M. R. A faster Rubisco with potential to increase photosynthesis in crops. *Nature* **513,** 547–550 (2014).

131. SUZUKI, Y., MIYAMOTO, T., YOSHIZAWA, R., MAE, T. & MAKINO, A. Rubisco content and photosynthesis of leaves at different positions in transgenic rice with an overexpression of *RBCS*. *Plant. Cell Environ.* **32,** 417–427 (2009).

132. Lefebvre, S. *et al.* Increased sedoheptulose-1, 7-bisphosphatase activity in transgenic tobacco plants stimulates photosynthesis and growth from an early stage in development. *Plant Physiol.* **138,** 451–460 (2005).

133. Hartmann, T. Diversity and variability of plant secondary metabolism: a mechanistic view. *Entomol. Exp. Appl.* **80,** 177–188 (1996).

134. Rogers, M. & Keeling, P. J. Lateral transfer and recompartmentalization of calvin cycle

enzymes of plants and algae. *J. Mol. Evol.* **58,** 367–375 (2004).

135. Berman-Frank, I., Lundgren, P. & Falkowski, P. Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. *Res. Microbiol.* **154,** 157–164 (2003).

136. Michels, A. K., Wedel, N. & Kroth, P. G. Diatom plastids possess a phosphoribulokinase with an altered regulation and no oxidative pentose phosphate pathway. *Plant Physiol.* **137,** 911–920 (2005).

137. Doyle, J. J., Luckow, M. A., Hortorium, L. H. B., Biology, P. & Library, M. Update on Phylogeny The Rest of the Iceberg . Legume Diversity and Evolution in a Phylogenetic Context 1. **131,** 900–910 (2003).

138. Fowler, R. G., Degnen, G. E. & Cox, E. C. Mutational Specificity of a Conditional Escherichia coli Mutator , mutD5. *Mol. genomics Genet.* **133,** 179–191 (1974).

139. Takano, K., Nakabeppu, Y., Maki, H., Horiuchi, T. & Sekiguchi, M. Structure and function of dnaQ and mutD mutators of Escherichia coil. *Molec* **205,** 9–13 (1986).

140. Maruyama, M., Takashi, H., Maki, H. & Sekiguchi, M. A Dominant (mutD5) and a Recessive (dnaQ49) Mutator of Escherichia coli. *J. Mol. Biol.* **167,** 757–771 (1983).

141. Damagnez, V., Doutriaux, M. & Radman, M. Saturation of Mismatch Repair in the mutD5 Mutator Strain of Escherichia coli. *J. Bacteriol.* **171,** 4494 (1989).

142. Schaaper, R. M. & Radman, M. The extreme mutator effect of Escherichia coli mutD5 results from saturation of mismatch repair by excessive DNA replication errors. *EMBO J.* **8,** 3511–6 (1989).

143. Stolzenberger, J., Lindner, S. N., Persicke, M., Brautaset, T. & Wendisch, V. F. Characterization of fructose 1,6-bisphosphatase and sedoheptulose 1,7-bisphosphatase from the facultative ribulose monophosphate cycle methylotroph Bacillus methanolicus. *J. Bacteriol.* **195,** 5112–22 (2013).

144. Atkinson, D. E. & Walton, G. M. Kinetics of Regulatory Enzymes. Escherichia coli Phosphofructokinase. *J. Biol. Chem.* **240,** 757–763 (1965).

145. Chang, M. C. Y., Eachus, R. a, Trieu, W., Ro, D.-K. & Keasling, J. D. Engineering Escherichia coli for production of functionalized terpenoids using plant P450s. *Nat. Chem. Biol.* **3,** 274–7 (2007).

146. Ajikumar, P. K. *et al.* Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science* **330,** 70–4 (2010).

147. Lütke-Eversloh, T. & Stephanopoulos, G. Combinatorial pathway analysis for improved L-tyrosine production in Escherichia coli: identification of enzymatic bottlenecks by systematic gene overexpression. *Metab. Eng.* **10,** 69–77 (2008).

148. Zelcbuch, L. *et al.* Spanning high-dimensional expression space using ribosome-binding site combinatorics. *Nucleic Acids Res.* **41,** e98 (2013).

149. Alper, H., Miyaoku, K. & Stephanopoulos, G. Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. *Nat. Biotechnol.* **23,** 612–6 (2005).

150. Reed, J. L., Vo, T. D., Schilling, C. H. & Palsson, B. O. An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biol.* **4,** R54 (2003).

151. Chistoserdova, L., Kalyuzhnaya, M. G. & Lidstrom, M. E. The expanding world of methylotrophic metabolism. *Annu. Rev. Microbiol.* **63,** 477–99 (2009).

152. Pharkya, P., Burgard, A. P. & Maranas, C. D. OptStrain: a computational framework for redesign of microbial production systems. *Genome Res.* **14,** 2367–76 (2004).

153. Chowdhury, A., Zomorrodi, A. R. & Maranas, C. D. k-OptForce: integrating kinetics with flux balance analysis for strain design. *PLoS Comput. Biol.* **10,** e1003487 (2014).

154. Contador, C. a, Rizk, M. L., Asenjo, J. a & Liao, J. C. Ensemble modeling for strain development of L-lysine-producing Escherichia coli. *Metab. Eng.* **11,** 221–33 (2009).

155. Bogorad, I. W., Lin, T.-S. & Liao, J. C. Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature* **502,** 1–6 (2013).

156. Clark, D. P. The fermentation pathways of Escherichi coli. *FEMS Microbiol. Rev.* **63,** 223–234 (1989).

157. Mainguet, S. E., Gronenberg, L. S., Wong, S. S. & Liao, J. C. A reverse glyoxylate shunt to build a non-native route from C4 to C2 in Escherichia coli. *Metab. Eng.* **19,** 116–27 (2013).

158. Huber, H. *et al.* A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic Archaeum Ignicoccus hospitalis. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 7851–6 (2008).

159. Fischer, E. & Sauer, U. A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry Escherichia coli. *J. Biol. Chem.* **278,** 46446–51 (2003).

160. Buchanan, B. B. & Arnon, D. I. A reverse KREBS cycle in photosynthesis: consensus at last. *Photosynth. Res.* **24,** 47–53 (1990).

161. Herter, S., Fuchs, G., Bacher, A. & Eisenreich, W. A bicyclic autotrophic CO2 fixation pathway in Chloroflexus aurantiacus. *J. Biol. Chem.* **277,** 20277–83 (2002).

162. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).

163. Bar-Even, a., Noor, E., Lewis, N. E. & Milo, R. Design and analysis of synthetic carbon fixation pathways. *Proc. Natl. Acad. Sci.* **107,** 8889–8894 (2010).