UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Application of 2-Scale 13C Metabolic Flux Analysis to Growth Phenotypes in S. cerevisiae

Permalink

https://escholarship.org/uc/item/99j5w5q9

Author

Shymansky, Christopher Michael

Publication Date

2015

Peer reviewed|Thesis/dissertation

Application of 2-Scale ^{13}C Metabolic Flux Analysis to Growth Phenotypes in
S. cerevisiae

by

Christopher Michael Shymansky

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Jay D. Keasling, Chair
Professor Emeritus Harvey Blanch
Professor David Schaffer
Professor John Dueber

Fall 2015

Application of 2-Scale ^{13}C Metabolic Flux Analysis to Growth Phenotypes in *S. cerevisiae*

Abstract

Application of 2-Scale ^{13}C Metabolic Flux Analysis to Growth Phenotypes in *S. cerevisiae*

by

Christopher Michael Shymansky

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Jay D. Keasling, Chair

Fluxes are useful because they are the most phenotypically relevant output we can infer. The dominant methods of obtaining these profiles, FBA and ^{13}C MFA, have advantages and disadvantages. 2S-^{13}C MFA combines the advantages of both to obtain more reliable genome-scale flux profiles that are self-consistent with both the carbon transition model and data used to infer them. This makes them a better basis for troubleshooting phenotypic changes and for making predictions.

We do not fully understand the role of Sip1 in glucose repression systems. A better understanding of it and the general phenomenon of carbon catabolite-repression could result in better ways to engineer cells and utilize various feedstocks. Previous work in our lab resulted in an unreported growth phenotype upon knockout of *SIP1* in mixed glucose/galactose medium. To better understand the relative roles of galactose and knockout of *SIP1*, we constructed base and *sip1Δ* mutant strains in a CEN.PK113-7D *ura3Δ gal1Δ* background and compared their 2S-^{13}C MFA-derived flux profiles in both glucose-only and mixed glucose/galactose media. Our original hypothesis, that deletion of *SIP1* was necessary to see an effect from the presence of galactose, was incorrect. The presence of galactose was necessary to see a phenotypic difference in growth rate upon knockout of *SIP1*. Both the base and mutant strains exhibited, upon addition of galactose, increases in specific growth rate, decrease of PPP pathway activity to 1/100 of its initial value, shift from use of NAD- to NADP-dependent malic enzyme, and redistribution of flux towards branched-chain amino acid biosynthesis. Despite the expected lack of change in growth rate upon knockout of *SIP1* in glucose repressing conditions, extracellular ethanol flux decreased, mitochondrial flow completely shut down, and flow was directed toward a part of the network involving arginine and threonine biosynthesis. Regardless of the noted differences, all strain/condition pairs were accompanied by certain normal glucose-repressing phenotypes (i.e. ethanol production and repression of TCA/glyoxylate cycle activity). Additionally, a cycle from cytosolic pyruvate through cytosolic malate and the mitochondria back to cytosolic pyruvate through NAD-dependent malic enzyme occurred. Clearly, glucose repression and the role of Sip1 is more complicated than we realized. Recent work makes it plausible that galactose enters the cell at the galactose/glucose ratio in this study. The more-reliable and self-consistent flux

profiles generated by 2S-^{13}C MFA were instrumental in studying this problem at unprecedented detail. However, more work is necessary to fully elucidate the role of Sip1 in carbon catabolite-repressing conditions.

Carbon lost to non-target metabolites (e.g. ethanol) represents a missed opportunity. It might be possible to redirect this lost flux towards biomass and/or target-molecule production. Additionally, the question of whether we can use an NADH oxidase to correct cofactor imbalances is an interesting one. We attempted to rescue the growth of an *ADH1* null mutant via low-copy plasmid expression of 15 different promoter/gene pairs. These promoter/gene pairs consisted of all combinations of 5 mutated *TEF1* promoters and 3 species-specific NADH oxidases from *L. lactis*, *S. pneumoniae*, and *A. capsulatus*. The growth of these 15 variants were characterized, along with that of base and *adh1*Δ mutant strains, in low-replicate shake flask, 96-well, 24-well, and high-replicate shake flask experiments. Expression of the NADH-oxidase appeared to result in a growth rescue in the initial low-replicate flask and the 96-well plate experiments. There was no clear difference in 24-well plate experiments and the high-replicate shake flask experiment definitively showed that heterologous expression of one particular variant resulted in no growth rescue. More experiments are necessary to determine if growth rescue by balancing cofactor utilization is a viable strategy. Higher replicate experiments with the same variant, high-copy expression, chromosomal integration, and/or codon-optimization could all be tried as extensions of this work. It's interesting how there appeared to be a difference in the 96-well experiment, admittedly with about half the replicates of the final one. Differing rates of aeration at these two scales may be responsible for this observation, since all NADH oxidases used in this study converted oxygen to water.

Flux profiles are the most phenotypically relevant output we can infer from experimental data. Any advance in the speed or accuracy of flux profile inference from ^{13}C labeling data could increase the range of its applications. The optimization inherent to ^{13}C MFA is a high-dimensional, nonconvex least squares minimization subject to third-order polynomial constraints. Nonconvex systems tend to be harder to solve due to local minima. We were interested to see if we could find bounds or solver starting points and/or even recover the same global minimum solution using a convex semidefinite program relaxation with forms of rank-sparcity encouragement. For a six-reaction toy model, we found that the SDP-relaxation resulted in a lower bound on the objective and the rank-sparcity encouragement methods had their advantages and disadvantages. The weighted-trace objective method resulted in an optimal solution and argument closer to that found in the paper, albeit requiring more user supervision and resulting in a less-feasible point. The eigenvalue inflation method was more automatable, faster, and possessed better mass balance constraint satisfaction. Though these results are encouraging, more work remains to be done to determine whether convex relaxations could improve overall ^{13}C MFA flux profile inference. The space of solutions possesses a semi-linear nature so, it is plausible that convex relaxations could be useful. The SDP relaxation attempted here and the rank-sparcity encouragement methods, among others, should be attempted on larger networks.

Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude to my research advisor Prof. Jay Keasling for his guidance, patience, compassion, and support. He and the environment he has created have enriched me as a person and given me the opportunity to pursue my various intellectual interests. My work draws from so many disciplines; I doubt I could have managed without his ability to connect me with the people I needed to move my projects forward.

I am deeply grateful to Hector Garcia Martin for his vision and advice in general. Thank you for encouraging me to learn how to code and guiding me through languages (e.g. Matlab, Python, Unix, Git) I expect will be the core tools I use throughout my career. Finally thank you for working with me so closely on finishing my thesis.

I'd also like to thank Laurent El Ghaoui for teaching me convex optimization and for guiding me though its application to my work. The knowledge I've gained from that particular project and your courses have opened up entire branches of mathematics and science to me.

Thanks to the rest of my thesis committee for reading and commenting on my thesis and for advice in general.

So many people at JBEI have helped shape me academically and personally. I greatly appreciate George and Edward's help with my sample processing and for advice in general. Jennifer and Victor were essential to my learning how to perform my strain characterization experiments. Thanks to Bilge, Zane, Jennifer, Sarah, Leo, Mario, Charles, Ricky, and, especially, Amanda for teaching me how to genetically engineer yeast and for providing me with various materials and yeast-related advice. I've also appreciated discussions with CJ and Jim Kirby on various topics. Many thanks to members of the Quantitative Metabolic Modeling group, especially Dan Weaver, Liz Brunk, and Amit Ghosh for help with computational/mathematical aspects of my work. Thanks to Ee-Been for lab-related advice. I'm very grateful to Sarah Richardson for setting up and teaching me how to use my LaTeX environment. I doubt I could have finished my thesis in time without it. Many thanks to Olga, Karen, Beverly, Linda, and especially Susan for guiding me through my graduate school career. Thanks to the rest of JBEI, especially Daniel, Richard, and Jeff for making my experience more enjoyable.

I owe my loving thanks to Chuck for his support and so much more.

Chapter 1

Introduction

1.0.1 The role of the fluxome in systems biology

Cells are dynamic. Thus understanding their behavior requires understanding how cellular components change with respect to each other. It would be easier to understand cellular behavior if rates of expression determined rates of translation and these determined metabolic reaction rates. Unfortunately, this is greatly complicated by transcriptional, translational, and post-translational regulation[1,2]. Genetic sequences and regulator proteins modify the rates of transcription. Secondary structure of mRNA and modifications of ribosomal recruitment rate modify rates of translation. Protein folding and metabolite inhibition modify protein function, including reaction rates. Other mechanisms such as chromosomal location of genes, histone acetylation, and chromosomal structure further complicate our understanding.

While a full mathematical model of all cellular component interactions has yet to be realized, our understanding of metabolic interconversion rates, or the fluxome, has probably been our most successful. Fluxes represent the flow of mass through a metabolic network and are effectively a proxy for the number of reactions per hour per cell. Our ability to calculate them and the resolution at which we're able to do so have greatly increased as our experimental techniques and computational power have. Thankfully, flux profiles are the network output most relevant to a large number of cellular behaviors we are most interested in. Growth rate is essentially a flux to biomass, cellular composition results from relative individual component fluxes to biomass, rates of excretion/consumption of extracellular metabolites are fluxes, etc. Various techniques have been developed to infer these intracellular reaction rates, as discussed below.

1.0.2 Flux balance analysis

The common assumptions of fluxomics, in general, are conservation of mass and the completeness of the metabolic model used to derive it. Stoichiometric intracellular metabolite balances are the manifestation of these assumptions in flux calculations. They are constructed by recognizing that the rate of accumulation of each intracellular metabolite's mass

is equal to the sum of the individual rates from each reaction it is involved in. Oftentimes, these individual rates of accumulation are set to zero for exponential phase situations where excretion/consumption fluxes appear to be constant. This results in a set of algebraic instead of differential equations.

Extracellular metabolite and cellular concentration data are often used in fluxomic methods. Extracellular fluxes are calculated from these metabolite concentrations. Often referred to as "exchange" fluxes, these excretion/consumption rates are used to place bounds on metabolic network inputs and outputs. Another core assumption of many fluxomic methods is that of cellular metabolite composition percentages, encoded in a biomass rate equation. Specific growth rates are, sometimes, used to place bounds on these biomass rate equations.

One approach to obtaining flux profiles is Flux Balance Analysis (FBA). FBA uses these stoichiometric balances, the biomass equation linked to specific growth rate, and flux sign constraints. Modern FBA uses highly-curated genome scale models representing the most complete models of an organism's metabolic network available. These genome-scale stoichiometric equations coupled with the constraints imposed by the measured data result in an underdetermined system of linear equations with an infinite solution set, thus requiring a biological objective (e.g. maximization of growth rate or ATP production) to obtain a single flux profile. Flux balance analysis usually takes the form of a linear program (LP), similar to that displayed in Equation 1.2.

The main advantage of this approach is the completeness of the genome-scale model. Also, FBA can be used to, in some cases, to accurately predict growth and excreted metabolites. Additionally, certain methods (e.g. MOMA) require genome-scale flux profiles to make predictions. However, the biological objective isn't always true and the technique has difficulty resolving fluxes in bidirectional reactions and cycles.

1.0.3 ^{13}C metabolic flux analysis

An alternative method of obtaining flux profiles is ^{13}C metabolic flux analysis (^{13}C MFA). The basic idea is to introduce extra information, other than the normally measured extracellular flux and specific growth rates of FBA, by performing a tracer experiment where an organism is cultured in medium containing ^{13}C-labeled substrate(s). The labeled carbons accumulate in cellular biomass that is processed and analyzed using nuclear magnetic resonance (NMR) or mass spectrometry to obtain ^{13}C labeling probability distributions for cellular components (e.g. intracellular metabolites and/or proteinogenic amino acids). Labeled species mass balances (usually steady-state) are used to model the flow of labeled carbons and are derived from reaction carbon transitions. These, coupled with their corresponding stoichiometric balances and sign constraints, are used to generate simulated species labeling distributions. An iterative fitting is conducted until a modified Euclidean distance between the simulated and measured labeling data is minimized.

^{13}C MFA improves upon FBA by replacing the biological assumption objective with one that ties flux profiles to measured data and by being able to resolve bidirectional reaction rates and those in cycles through the use of a more detailed model. FBA often uses extracellular flux- and biomass equation-constraining extracellular metabolite and cellular

concentration data. ^{13}C MFA uses both this data and the measured metabolite labeling probability distributions resulting from the tracer experiment. Also, the labeled data allows a measure of data and model consistency by replacing the biological assumption objective of FBA with a Euclidean distance between measured and simulated labeling data. Additionally, confidence intervals of fluxes obtained via ^{13}C MFA are tighter and the finer-scale model allows determination of reaction fluxes involved in cycles and forward/reverse values in reversible reactions. Despite these advantages, ^{13}C MFA uses a non-comprehensive stoichiometric model, including a simplified biomass equation that doesn't include all components necessary to produce cells, and the limited size of the network makes it less than ideal for predictive methods. Despite the constraining nature of the labeled species balances resulting in a finer flux profile resolution, they are the source of nonconvexity that results in multiple local minima and greatly increased computational complexity. Finally, ^{13}C MFA assumes the reactions it doesn't include would have no effect on the simulated labeling output.

1.0.4   Elementary metabolite unit decomposition

We, in particular, use an elementary metabolite unit (EMU) decomposition algorithm[3] to derive steady-state labeling balances that retain the information of previous methods while reducing the number of equations and unknowns by about 1/10 relative to previous methods. Labeled species mass balance equations are obtained from a given network of reactions and their corresponding carbon transitions, which describe where each carbon atom in the reactants ends up in the products of a given reaction.

An elementary metabolite unit (EMU) is any distinct subset of carbon atoms in a metabolite for which the fractional labeling is being considered. That is, an EMU is a piece of a molecule. Each piece has a corresponding labeling distribution known as its mass distribution vector (MDV), designated f_e (Equation 1.1), for which each entry, $f_{e,m}$ is the probability that any m number of carbons is labeled (^{13}C instead of ^{12}C). As such, the first entry contains the probability that EMU, e, has no labeled carbons (m=0), the second entry contains the probability that any one of the carbons is labeled (m=1), etc.

$$f_e = \left\{ f_{e,m} \ : \sum_{m=0}^{m_e} f_{e,m} = 1, f_{e,m} \geq 0 \right\} \in \mathbb{R}^{m_e+1} \tag{1.1}$$

The EMU algorithm uses the carbon transitions to decompose the network into multiple small networks consisting of EMUs. This is accomplished by starting from the molecules with measured labeling patterns, defining what pieces of other molecules were used to create them and through what reactions, then defining these newly encountered EMUs in terms of others as well, and continuing this process until the labeling of all EMUs is completely determined from those encountered and that of the feed molecule's. For instance, for the six-reaction toy model obtained from the founding EMU paper and presented in Table 1.1 the EMU reaction network resulting from the decomposition is displayed in Figure 1.1.

These EMU decompositions are used to derive steady-state mixing equations for each EMU labeling state that serve as the ^{13}C MFA labeled species balance constraints in Equa-

Table 1.1: Set of reactions and corresponding carbon transitions for a six-reaction toy model

Reaction number	Reaction stoichiometry	Atom transitions
1	A → B	abc → abc
2 and 3	B ↔ D	abc ↔ abc
4	B → C + E	abc → bc + a
5	B + C → D + E + E	abc + de → bcd + a + e
6	D → F	abc → abc

Figure 1.1: EMU decomposition of a six-reaction toy model[3]. Metabolite (nodes) subscripts denote carbon atoms for which ^{13}C labeling probabilities are being kept track of.

tion 1.4e. The labeling variables of these equations are MDVs mentioned earlier. The mathematical structure of the models resulting from our models containing 130 reactions about 6000 dimensional (i.e. MDV entries and fluxes) 3rd-order polynomials (i.e. sums of products of 1 or 2 unknown labeling values and 1 unknown flux), depending on the size of the model.

1.0.5 2-scale ^{13}C metabolic flux analysis

2-scale ^{13}C metabolic flux analysis (2S-^{13}C MFA)[4], developed at the Joint BioEnergy Institute (JBEI), is an attempt to combine the advantages of both FBA and ^{13}C MFA. The name reflects the fact that the metabolic model has two scales of resolution. A comprehen-

sive genome scale model is used, yet with a "core" subset of reactions containing finer-scale carbon transition information.

2S-^{13}C MFA uses a cycle of computations that are iteratively performed until the possible ranges of simulated labeling distribution data is minimized. This cycle of computations starts with a "limit-flux-to-core" step, then a ^{13}C MFA, and then an External Labeling Variability Analysis (ELVA).

The first step involves identifying reactions contributing flux to the core network from the non-core network. A procedure is separately performed for each of these non-core-to-core reactions to determine the minimum flux through it that is consistent with the network stoichiometry of the full genome-scale model. This is achieved by iterating through certain percentages of the incoming glucose flux for that single reaction until a genome-scale flux balance analysis is successful. The mathematical form of the Flux Balance Analysis computation is displayed in Equation 1.2 and descriptions of the sets and variables involved are in Equation 1.3.

$$\max_{v} \quad v_{obj} \tag{1.2a}$$

$$\text{s.t.} \quad \sum_{j \in J} S_{ij} v_j = 0\,, \ \forall i \in I^N \tag{1.2b}$$

$$lb_j \leq v_j \leq ub_j\,, \ \forall j \in J \tag{1.2c}$$

$$
\begin{aligned}
I^N \subset I : \ & \text{Set of non-exchange metabolites} \\
J = j : \ & \text{Set of fluxes} \\
S_{ij} : \ & \text{Stoichiometric coefficient of metabolite i in reaction j} \\
ub_j, lb_j : \ & \text{Upper/lower bounds for reaction j} \\
v_j : \ & \text{Flux value of reaction j, in mmol/gDcW/h} \\
v_{obj} : \ & \text{Objective flux to maximize (e.g. biomass or atp)}
\end{aligned} \tag{1.3}
$$

Once minimum values for all non-core-to-core reactions are found their values are set for these fluxes for a modified ^{13}C MFA. The mathematical form of this problem is displayed in Equation 1.4 and the corresponding set and variable descriptions are in Equation 1.5.

$$\min_{v,V,f} \quad \sqrt{\sum_{e \in E_{meas}} \left(\sum_{m \in M_e} \left(\frac{f_{em}^{exp} - f_{em}}{\Delta_{em}} \right)^2 / |M_e| \right) / |E_{meas}|} \tag{1.4a}$$

$$\text{s.t.} \quad \sum_{j \in J} S_{ij} v_j = 0 \,, \ \forall i \in I^N \tag{1.4b}$$

$$lb_j \leq v_j \leq ub_j \,, \ \forall j \in J \tag{1.4c}$$

$$\sum_{m \in M_e} f_{em} = 1 \,, \ \forall e \in E_{co} \tag{1.4d}$$

$$\left(\sum_{l | S_{il}^* < 0} S_{il}^* V_l \right) f_{em} + \sum_{e' \in E_{co}} \left(\left(\sum_{l | EMM_{e' -> e}^l > 0} EMM_{e' -> e}^l V_l \right) f_{e'm} \right) = 0 \,, \tag{1.4e}$$

$$\forall m \in M_e, \forall e \in E_i, \forall i \in I_{co}^N \tag{1.4f}$$

$$f_{em} = \sum_{w \in W_{em}} \prod_{n=1}^{|E_e|} f_{e_n m_n} \,, \ \forall m \in M_e, e \in E_{co}^c \tag{1.4g}$$

$$v_j = \sum_{l \in J_{co}^B} (map_{jl}) V_l, \forall j \in J \tag{1.4h}$$

$$
\begin{aligned}
I = \{i\} : &\quad \text{Set of all metabolites} \\
I_{co} \subset I : &\quad \text{Set of core metabolites} \\
I^N \subset I : &\quad \text{Set of non-exchange metabolites} \\
I_{co}^N \subset I_{co} : &\quad \text{Set of non-exchange core metabolites} \\
J : &\quad \text{Set of fluxes} \\
J_{co} \subset J : &\quad \text{Set of core fluxes} \\
J^B : &\quad \text{Set of fluxes with backward and forward fluxes differentiated} \\
J_{co}^B \subset J^B : &\quad \text{Set of core fluxes with reverse/forward fluxes differentiated} \\
E = \{e\} : &\quad \text{Elementary Metabolite Units (EMUs) or carbon groups} \\
E^c \subset E : &\quad \text{Combined EMUs} \\
E_i \subset E : &\quad \text{EMUs from metabolite } i \in I \\
E_{co}^c \subset E^c : &\quad \text{Core combined EMUs} \\
E_e \subset E : &\quad \text{EMUs that produce combined EMU e} \\
E_{co} \subset E : &\quad \text{EMUs corresponding to core metabolites} \\
E_{meas} \subset E : &\quad \text{EMUs corresponding to measured EMUs} \\
W_{em} : &\quad \text{Set of every possible mass isotopomer multiplet} \\
&\quad \text{of } E_e \text{ that produce the mass isotopomer m of e} \\
M_e : &\quad \text{m values ofr MDV of emu e: 0, 1, } \ldots, \# \text{ of carbons in e} \\
EMM_{e' ->e}^l : &\quad \frac{1}{k} \text{ if } e' \text{ produces e through reaction } l \in J_{co}^B, \text{ 0 otherwise} \\
S_{ij} : &\quad \text{Stoichiometric coefficient of metabolite i in reaction j} \\
S_{ij}^* : &\quad \text{Stoichiometric coefficient of metabolite i in reaction j in network} \\
&\quad \text{corresponding to that with backward/forward fluxes differentiated} \\
ub_j, lb_j : &\quad \text{Upper/lower bounds for reaction j} \\
f_{em}^{exp} \in [0,1] : &\quad \text{Experimentally measured MDV for emu e from metabolite m} \\
\Delta_{em} : &\quad \text{Measurement error for } f_{em}^{exp} \\
map_{jl} : &\quad V_{glucupt} \text{ if l corresponds to forward flux of j and} \\
&\quad \text{-}V_{glucupt} \text{ if l corresponds to backward flux of j} \\
v_j : &\quad \text{Flux value of reaction } j \in J, \text{ in mmol/gDcW/h} \\
V_l : &\quad \text{Flux value of reaction } l \in J_{co}^B, \text{ normalized to glucose input rate} \\
f_{em} \in [0,1] : &\quad \text{Mass isotopomer fraction (MDV) for emu e from metabolite m}
\end{aligned}
\tag{1.5}
$$

There are a number of differences between a normal ^{13}C MFA and this version. The differences consist of there being two types of flux variables, the use of a set of genome-scale stoichiometric equations utilizing only one of these flux variable types, a mapping between the two flux variable sets, and the upper and lower flux bounds are limited by the previous

limit-flux-to-core step. The two sets of flux variables used are: 1) absolute net fluxes where some of them can be negative or positive and 2) positive relative fluxes that are normalized to the absolute incoming glucose flux. The genome-scale stoichiometric balances in Equation 1.4b correspond to all fluxes of the absolute flux type, whose values are positive if flow is in the designated "forward" direction and negative if flow is the designated "reverse" direction. The EMU balances are in terms of the second set of relative fluxes that correspond only to the core. For clarity, core reactions have both types of flux variables corresponding to them. The relationship between these two flux variables is expressed in the mapping equation between them in Equation 1.4h. This mapping is simply that the absolute value of a particular reaction, j, is calculated by its relative value times the positive or negative glucose consumption flux, depending on the direction. This is encoded in the map_{jl} variable, whose value is the glucose consumption flux if net flow of reaction l is in the forward direction or the negative of the glucose consumption flux if net flow is in the reverse direction. All equations involving flux variables, other than the EMU balances and flux mapping equations, are in terms of the absolute fluxes, v_j.

Note how the objective is a modified Euclidean distance between simulated and measured fractional labelings. The modification manifests in three different normalizations designed to give more weight to better data and to counteract overrepresentation of distributions with more values (i.e. more carbons) in the objective and to make it a little more comparable among different data sets. The division of each difference between simulated and measured mass isotopomer distribution values by its measured error, Δ_{em}, causes those with higher error to have a smaller term in the sum. The division of the sum of squares corresponding to each EMU's MDV by its number of entries, $|M_e|$, is meant to decrease the value of that term to decrease the weight of vectors with higher lengths. The final division by the number of EMUs is an attempt to counteract the higher objectives that result from having more data so that data sets of different sizes have some measure of comparison.

Once a ^{13}C MFA flux profile is obtained, we characterize the effect of the hypothetical labeling contributed to the core network by the non-core section by finding minimum and maximum simulated labeling probabilities corresponding to the measured data using an ELVA. We can not track the non-core contribution because we do not have carbon transition information for the reactions in that part of the network. Small simulated ranges give us confidence that the simulated value is actually close to the corresponding measured value from the ^{13}C MFA. This is accomplished by first identifying metabolites on the edge of the core network that have reactions generating them from the non-core network. These reactions are replaced by a single "extended flux" with a value given by their sum. The labeling contributed by this reaction is an "extended" metabolite with the same number of carbons. The labeling of all extended metabolites represents the labeling non-core-to-core network flow contributes to the core network. These extended labeling probability distributions are varied to minimize and maximize the simulated data corresponding to that measured subject to the ^{13}C MFA constraints and the flux profiles found in the ^{13}C MFA step. The mathematical form of ELVA is represented in Equation 1.6 and the corresponding set/variable definitions are in Equation 1.7.

$$\min/\max \quad f_{em}\,, \quad \forall m \in M_e, e \in E_{meas} \tag{1.6a}$$

$$\text{s.t.} \quad V_j = \overline{V}_j\,,\ j \in J^B_{coext} \tag{1.6b}$$

$$\sum_{j \in J^B_{coext}} S^*_{ij} V_j = 0\,,\ \forall i \in I^N_{coext} \tag{1.6c}$$

$$LB_j \leq V_j \leq UB_j\,,\ \forall j \in J^B_{coext} \tag{1.6d}$$

$$\sum_{m \in M_e} f_{em} = 1\,,\ \forall e \in E \tag{1.6e}$$

$$\left(\sum_{j|S^*_{ij}<0} S^*_{ij} V_j \right) f_{em} + \sum_{e' \in E} \left(\left(\sum_{j|EMM^j_{e'->e}>0} EMM^j_{e'->e} V_j \right) f_{e'm} \right) = 0\,, \tag{1.6f}$$

$$\forall m \in M_e, \forall e \in E_i, \forall i \in I^N \tag{1.6g}$$

$$\sum_{w \in W_{em}} \prod_{n=1}^{|E_e|} f_{e_n m_n}\,,\ \forall m \in M_e, e \in E^c_{coext} \tag{1.6h}$$

$$
\begin{aligned}
I^B_{coext} :\quad & \text{Set of extended metabolites} \\
J^B_{coext} :\quad & \text{Set of extended fluxes with backward and forward fluxes differentiated} \\
\overline{V}_j :\quad & \text{Solutions to the } {}^{13}\text{C MFA problem given by Equation 1.4}
\end{aligned} \tag{1.7}
$$

The cycle of the limit-flux-to-core, ^{13}C MFA, and ELVA steps is repeated until the simulated data ranges are tightened. This usually involves identifying the highest valued non-core-to-core flux. An example of this process is displayed in Figure 1.2. It took the addition of carbon transitions to a new reaction at the end of 3 separate cycles to tighten simulated data ranges to a satisfactory degree. Each cycle takes about 25-45 minutes to complete. The overall process is highly variable, taking anywhere from 1 hour to 4 days to complete in our hands. The reactions added to the core and their order of addition varies for each data set. Even minor changes to data inputs require that the process be completed over again.

Once simulated data ranges are tightened to a point where confidence in their values is established, a separate ^{13}C Flux Variability Analysis (^{13}C FVA) is performed to find the absolute minimum and maximum values for all fluxes in the network compatible with the labeling. This, essentially, involves minimizing and maximizing each flux subject to ^{13}C MFA constraints from earlier and new bounds on the simulated data variables in Equation 1.8i. The mathematical form of this computation can be found in Equation 1.8. The label bounds are constructed such that the worst error encountered, δ_{em}, is set to either the corresponding measured standard deviation or 110% of corresponding difference between measured and simulated values found from solution of Equation 1.4. More detail can be found in 2S-^{13}C
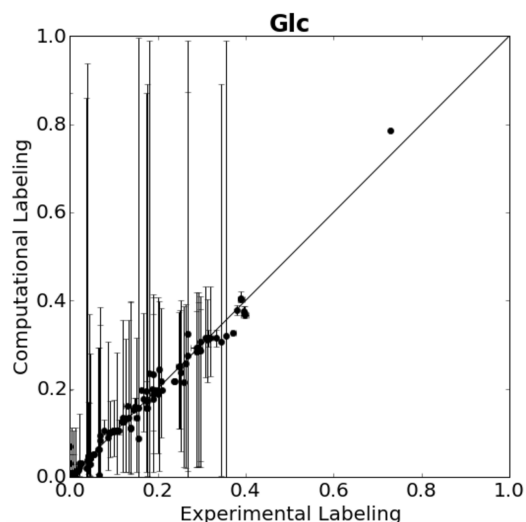
(a) Starting point ELVA plot
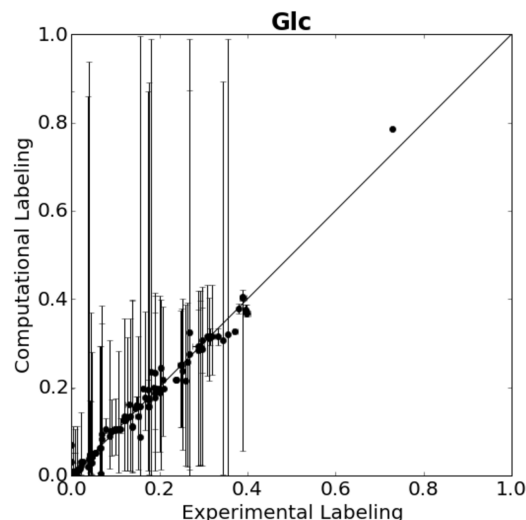
(b) ELVA plot after addition of a carbon transition for reaction 2HBt2
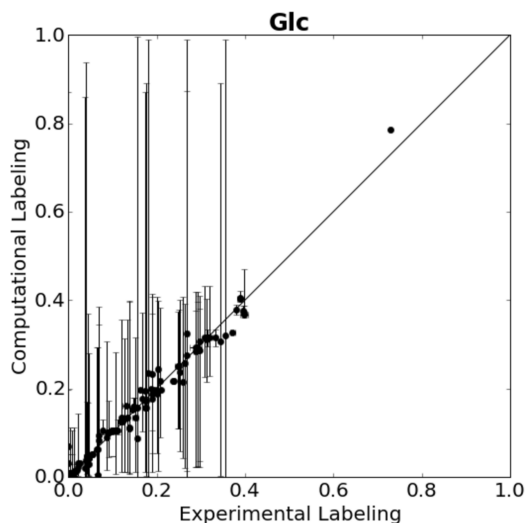
(c) ELVA plot after addition of a carbon transition for reaction ALAt2r
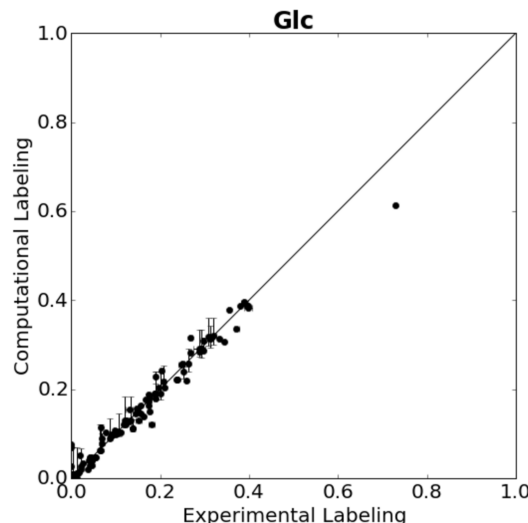
(d) ELVA plot after addition of a carbon transition for reaction FRDcm

Figure 1.2: External Labeling Variability Analysis (ELVA) plots illustrating the process of expanding the core network to improve simulated data confidence intervals. Each data point is an individual simulated versus measured labeling distribution mass isotopomer percentage. Horizontal error bars represent one standard deviation from the mean of the measured data. Vertical error bars represent the absolute minimum and maximum simulated labeling values from the ELVA. Each cycle of computations was performed ending with the ELVA plot. Carbon transitions were added for the reactions with the highest non-core-to-core flux and the cycle was repeated until the simulated ranges tightened.

MFA paper[4].

$$\min/\max \quad v_j, \quad \forall j \in J \tag{1.8a}$$

$$\text{s.t.} \qquad \sum_{j \in J} S_{ij} v_j = 0 \,, \ \forall i \in I^N \tag{1.8b}$$

$$lb_j \leq v_j \leq ub_j \,, \ \forall j \in J \tag{1.8c}$$

$$\sum_{m \in M_e} f_{em} = 1 \,, \ \forall e \in E_{co} \tag{1.8d}$$

$$\left(\sum_{l | S_{il}^* < 0} S_{il}^* V_l \right) f_{em} + \sum_{e' \in E_{co}} \left(\left(\sum_{l | EMM_{e' \to e}^l > 0} EMM_{e' \to e}^l V_l \right) f_{e'm} \right) = 0 \,, \tag{1.8e}$$

$$\forall m \in M_e, \forall e \in E_i, \forall i \in I_{co}^N \tag{1.8f}$$

$$f_{em} = \sum_{w \in W_{em}} \prod_{n=1}^{|E_e|} f_{e_n m_n} \,, \ \forall m \in M_e, e \in E_{coext}^c \tag{1.8g}$$

$$v_j = \sum_{l \in J_{co}^B} map_{jl} V_l \,, \ \forall j \in J \tag{1.8h}$$

$$(f_{em} - f_{em}^{exp})^2 \leq \delta_{em}^2 \,, \ \forall e \in E_{meas}, m \in M_e \tag{1.8i}$$

$$\tag{1.8j}$$

1.1 Dissertation Scope

The main theme of this dissertation is the use of 2S-^{13}C MFA to investigate different growth phenotypes resulting from genetic and environmental perturbations. Additionally, it includes our investigation of approximations to the mathematical structure of ^{13}C MFA. Chapter 2 will cover our application of 2S-^{13}C MFA to investigate an unusual growth phenotype involved in the interaction between the knockout of *SIP1*, a gene involved in glucose-repression, and the presence of galactose. Chapter 3 covers our attempt to perturb redox metabolism and rebalance it using expression of heterologous NADH oxidases. Chapter 4 describes our attempt to improve our ^{13}C MFA calculations using a novel series of convex approximations and rank-sparcity encouragement corrections.

Chapter 2

Investigation of an unusual growth phenotype upon knockout of *SIP1*

2.1   Introduction

Saccharomyces cerevisiae prefers fermentative growth in the presence of glucose. This phenomenon, known as glucose repression, involves the repression of genes and pathways involved in respiration (TCA cycle, etc), the use of alternative fermentable (e.g. sucrose and galactose) and non-fermentable (e.g. ethanol and acetate) carbon sources, and gluconeogenesis[5,6]. Consisting of a catalytic α-subunit Snf1, a regulatory γ-subunit Snf4, and one of three β-subunits Gal83, Sip1, or Sip2, the Snf1 kinase complex plays a central role in glucose repression. Upon depletion of glucose, Snf1 is phosphorylated by one of its upstream kinases (Sak1, Elm1, and Tos3) and the complex forms. That associated with Sip1 is sequestered in the vacuole, that with Sip2 remains in the cytosol, and that containing Gal83 localizes to the nucleus and activates genes involved in catabolism of other carbon sources[5] (Figure 2.1). This heterotrimer is completely dissociated and all of its components remain in the cytosol in the presence of glucose. Little is known about the role of Sip1 under these conditions due to a lack of phenotypic difference between wildtype and *sip1*Δ mutants[5,7,8].

While attempting to increase heterologously expressed biofuel production under a *GAL1* promoter in a previous study[9], we noticed an unreported increase in specific growth rate upon deletion of *SIP1* in a S288c *ura3*Δ *gal1*Δ background in medium containing both glucose and galactose. Sip1 has been shown to be a negative regulator of genes involved in galactose catabolism[10] and has been found to play a role in adherent growth[11]. A depiction of the relationship of Sip1 with the *GAL* gene system is shown in Figure 2.2. Galactose enters the cell through its transporter, Gal2, and is known to indirectly activate the main activator of the system, Gal4, via activation of Gal3, the repressor of Gal4's repressor, Gal80. We originally thought it possible that increased growth rate could have resulted from galactose entering the cell, upon the reported 2-to-3-fold increase in galactose transporter gene transcription upon knockout of *SIP1*, thereby indirectly activating Gal4, which is known to increase expression of genes involved in protein metabolism and transcription machinery[7,12].
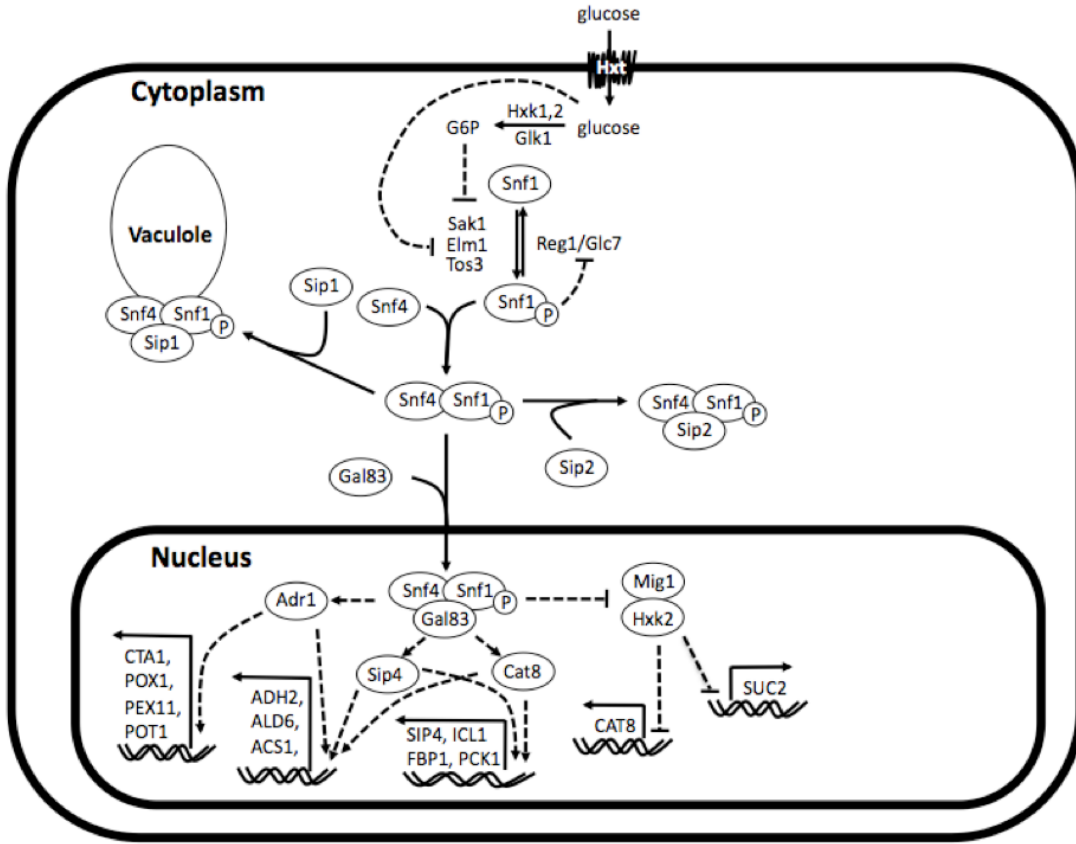
Figure 2.1: Simplified depiction of the role of Snf1 in glucose repression, its upstream promoters and repressors, and the role of its β-subunits in localization of the Snf1 kinase complex during high-glucose conditions. Reconstructed and modified from a 2008 paper by Zaman et. al.[5]

In this study, we investigate the relative roles of galactose and knockout of *SIP1* from a fluxomic perspective. We construct base and *sip1*Δ mutants in an industrially more-relevant CEN.PK113-7D *ura3*Δ *gal1*Δ background, similar to that from our previous work[9], and characterize their growth and flux profiles in minimal medium containing either glucose-only or glucose and galactose using 2-scale ^{13}C metabolic flux analysis.
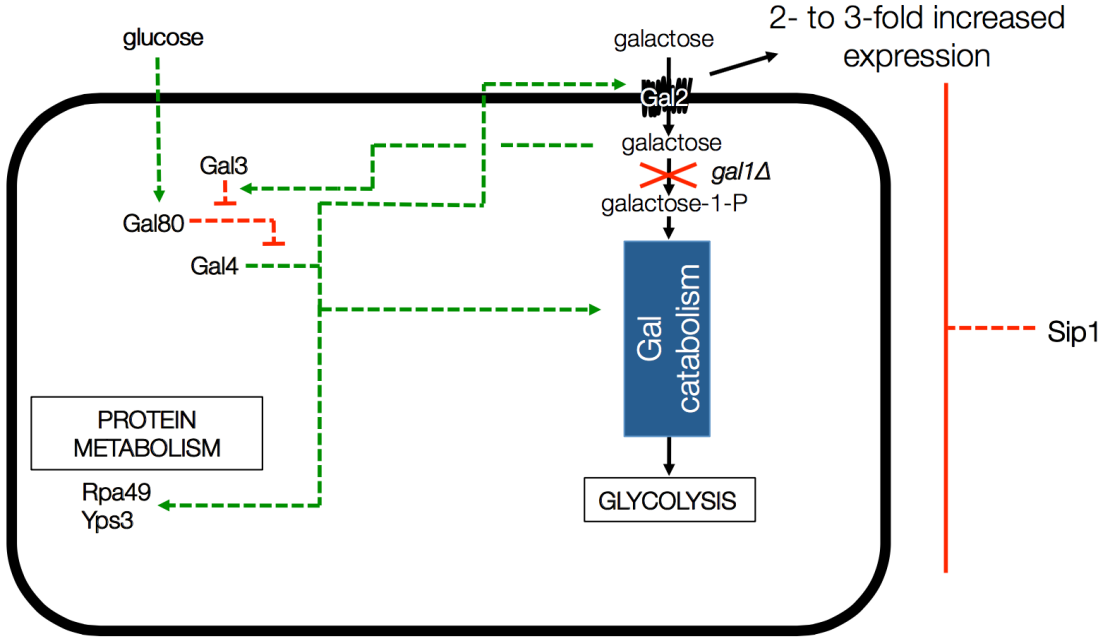
Figure 2.2: Depiction of *GAL* gene interactions with Sip1 in a *GAL1* knockout background. Green arrows indicate activation and red blunt arrows indicated repression

2.2 Materials & Methods

2.2.1 Media & culturing conditions

Media used in this study, along with their component concentrations are listed in Table 2.1. For both genetic manipulations and growth and tracer experiments, all strains were grown in non-baffled shake flasks at 30°C at 200 rpm in either minimal glucose medium (Min), minimal glucose medium with galactose (Min+Gal), YPD, or Sc-Ura. All stable strain intermediates were kept in 20% glycerol stocks at -80°C. Labeled media used 80% 1-^{13}C glucose and 20% U-^{13}C glucose at the same total concentration of 2% glucose. Exponential phase cells were obtained, in general, by streaking from -80°C glycerol stocks on YPD plates, incubating 5 mL YPD cultures overnight, inoculating into 40 mL of unlabeled media of the final desired composition, and grown until exponential phase (usually 0.6-0.9 OD).

2.2.2 Strain construction

Prototrophic base (U) and mutant (S) *S. cerevisiae* strains were constructed in a haploid CEN.PK113-7D[13] (*Mat***a** *MAL2-8^c SUC2*) background containing a *URA3* knockout provided by Bilge Ozaydin of JBEI. Intermediate and final base and *sip1*Δ mutant strains are listed in Table 2.2 with their strain designations, parent strain, genotype descriptions,

Table 2.1: Media and component concentrations

Media	Component	Concentration [g/100 mL]
Min	Glucose	2
	Yeast nitrogen base w/o amino acids	0.67
Min+Gal	Glucose	2
	Galactose	0.2
	Yeast nitrogen base w/o amino acids	0.67
Min+Ura+5-FOA	Glucose	2
	Yeast nitrogen base w/o amino acids	0.67
	Uracil (Ura)	0.002
	5-Fluoroorotic acid (5-FOA)	0.1
YPD	Glucose	2
	Bacto-yeast extract	1
	Bacto-peptone	2
YPD+G418	Glucose	2
	Bacto-yeast extract	1
	Bacto-peptone	2
	Geneticin (G418)	0.02
Sc-Ura	Glucose	2
	Yeast nitrogen base w/o amino acids	0.67
	CSM-Ura	0.077
pSH47	Glucose	0.2
	Galactose	1.8
	Yeast nitrogen base w/o amino acids	0.67
	CSM-Ura	0.077

Note: All plates contained 2 g/100 mL bacto-agar

and references. All knockouts were constructed via a near-markerless loxP/Cre recombinase strategy[14] and PCR verified. Briefly, each knockout cassette was amplified from a loxP-kanMX-loxP plasmid, pUG6[14], using the primers listed in Table 2.3, transformed into yeast using a heat shock method[15], selected on YPD+G418 (geneticin) plates, and PCR verified using primers listed in Table 2.4. In order to loop out the kanMX marker, a Cre recombinase plasmid was transformed in the resulting kanMX cassette integrants and plated on selective medium. The selective plate varied depending on the knockout. For knockout of *SIP1*, the cre recombincase promoter was Gal1p (pSH47[14]) and selection occurred on pSH47 plates. A different plasmid was necessary for knockout of *GAL1*, since the strain couldn't grow on galactose. We opted for expression of the cre recombinase under a constitutive *TEF1* promoter. This new plasmid, pCMS1, was constructed via yeast cloning using SacI and XbaI digested pSH47, to excise Gal1p, and Tef1p amplified with regions homologus to the cut ends and subsequent selection on Sc-Ura plates. All loop-outs were PCR verified using the same verification primers in Table 2.4 and pCMS1 was sequence verified. Cre recombinase plasmids were cured by streaking on YPD plates, growing overnight in liquid YPD medium, streaking to single colonies on YPD plates, simultaneously streaking on YPD and Sc-Ura plates, and glycerol storing YPD plate colonies whose corresponding Sc-Ura colonies didn't grow. Plasmids pUG6 and pSH47 were graciously provided by Dr. Ozaydin.

Prototrophic final base and mutant strains were completed via transformation of a *URA3* plasmid, pRS416[16], provided by Dr. Amanada Reider Apel of JBEI.

Table 2.2: List of strains, their parents, genotypes, and references

Strain name	Parent strain	Description	Reference
CPU	CEN.PK113-7D	CEN.PK113-7D *ura3Δ*	From Bilge Ozaydin (JBEI)
CMSY3	CPU	CPU *SIP1::loxP-kanMX-loxP*	This work
CMSY4	CMSY3	CPU *sip1Δ*	This work
CMSY5	CPU	CPU *GAL1::loxP-kanMX-loxP*	This work
CMSY7	CMSY5	CPU *gal1Δ*	This work
CMSY6	CMSY4	CPU *sip1Δ GAL1::loxP-kanMX-loxP*	This work
CMSY8	CMSY6	CPU *sip1Δ gal1Δ*	This work
U	CMSY7	CPU *gal1Δ* [pRS416]	This work
S	CMSY8	CPU *sip1Δ gal1Δ* [pRS416]	This work

Table 2.3: Deleted genes and corresponding templates and forward/reverse primers used to construct knockout cassettes. Uppercase indicates homologous flanking regions and lowercase designates regions binding to pUG6 plasmid

Knocked out gene	Template	F-primer	R-primer
GAL1	pUG6	AAAAATTGTTAATATACCTCTATACTTTAACGTCAAGGAGAAAAAACTATAagctgaagcttcgtacgc	GTAGAAAAAAATGAGAAGTTGTTCTGAACAAAGTAAAAAAAAGAAGTATACcataggccactagtggatctg
SIP1	pUG6	CTGACATCTTGGAAAGTTGAACTGTCATATTATATAGTTGTTGCAGCCGCCagctgaagcttcgtacgc	AGAAAAAAATTGAATTAATAGAGTTCGTGAGAATCATTGCGAATTGAGATTcataggccactagtggatctg

Table 2.4: Primers used to PCR verify specific gene deletions

Knocked out gene	F-primer	R-primer
GAL1	TTATTTCTGGGGTAATTAATCAGCGAAG	TCCCTGTGTTTCAAAGTTTGTGG
SIP1	GCACTTCTTTTTTTGCGTGTGG	CGTTCTAGGAGCCATAGGAATC

2.2.3 Growth characterization and tracer experiments

Cell and extracellular metabolite concentrations were monitored during exponential phase in strain characterization batch experiments. This data was necessary to calculate extracel-

lular fluxes and specific growth rates used to mathematically constrain flux profile inference. Exponentially growing cells, obtained as described in Section 2.2.1, were used to inoculate, in quadruplicate, the final 40 mL shake flask cultures to achieve exponential growth the following morning. Optical density was monitored at 600 nm via UV-VIS and 200 μL samples were spin-filtered and kept at -20°C for subsequent HPLC analysis.

Labeled mid-log biomass samples were taken during similar labeled batch experiments and immediately quenched for subsequent processing and analysis. The measured labeling obtained from analyzing these samples was used to judge the quality of the inferred flux profiles. Flux profiles that resulted in the best fit between measured and simulated labeling distributions were kept for further analysis. Exponentially growing cells (obtained as described above) were used to inoculate 40 mL labeled shake flask cultures in quadruplicate and monitored via UV-VIS. 1 ml samples were taken during mid-log (0.75 OD) and quenched, To prevent changes in intracellular metabolite labeling patterns, 1 mL mid-log samples were taken, spun down (1 min, max speed, 4°C), immediately quenched with 300 μL ice-cold methanol, and kept at -80°C.

2.2.4 Labeled biomass sample processing

Labeling distributions were obtained from processed labeled biomass samples for intracellular 3-phosho-D-glycerate (3pg), alanine (Ala), arginine (Arg), asparagine (Asp), glutamine (Gln), glutamate (Glu), isoleucine (Ile), leucine (Leu), lysine (Lys), phenylalanine (Phe), threonine (Thr), tyrosine (Tyr), valine (Val), citrate (cit_m), fructose 1,6-bisphosphate (fdp), and succinate (succ_m). We assumed that the succinate and citrate were mitochondrial. The closeness of fit of this data with corresponding simulated values provided a measure of the quality of inferred flux distributions. Labeled biomass samples were mixed with 300 μL ice-cold chloroform and 150 μL ice-cold water, spun down, bead-beated with 500 μL acid-washed beads (10 times, 10 s, 1 min ice breaks) in 1.7 mL screw cap tubes, the bottom of the tube was punctured with a needle, and the beads were separated from the solution by spinning (1 min, 1000g, 4°C) into a 2 mL collection tube. The aqueous layer was filtered (3K MW cut-off (Amicon), 1.5h, 13000g, 4°C), mixed with 1 mL ice-cold H$_2$O, and snap frozen in liquid nitrogen. Three holes were punched in the tube cap and the samples were lyophilized for 24 hours. Lyophilized samples were resuspended in 40 μL 50/50 MeOH/H$_2$O, and stored at -80°C. Samples were analyzed by George Wang to obtain intracellular amino acid and non-amino acid labeling data via LCMS as previously described[17,18].

2.2.5 Extracellular concentration determination

Extracellular concentrations for glucose, galactose, ethanol, glycerol, succinate, lactate, acetate, and formate were measured via HPLC. These concentrations, along with corresponding culture specific growth rates, were necessary to calculate extracellular fluxes. A 1200 Series HPLC (Agilent Technologies, CA) outfitted with UV and refraction index detectors and an Organic Acid Analysis Column (Aminex HPX-87H Ion Exclusion Column, 300 mm 7.8 mm,

Cat# 125-0140 Bio-Rad, CA) was used along with standards to identify metabolite retention times and sample concentrations.

2.2.6 Extracellular flux and intracellular labeling input calculations

Extracellular fluxes and specific growth rates were derived from extracellular concentration and optical density time curves. Their means and standard deviations were used to constrain exchange fluxes for consumed and excreted metabolites and biomass fluxes during flux profile inference. Flask-specific maximum specific growth rates were determined from the slope of lnOD versus time data in manually determined linear ranges. The same time points were used with corresponding concentration data to calculate extracellular fluxes using Equation 2.1. The extracellular flux of metabolite p is ν_p, M_p is its corresponding molecular weight, α is the conversion factor between OD and cell mass concentration in gDcW/L, and $d\overline{C}p/dOD$ is the slope of the concentration of metabolite p versus OD. The value of α was taken to be 0.7742 based on multiple in-house experiments (data not shown). The average plus and minus the corresponding standard deviation was used to constrain all extracellular fluxes and specific growth rates.

$$\nu_p = 1000 \frac{\mu}{M_p \alpha} \frac{d\overline{C}p}{dOD} \tag{2.1}$$

2.2.7 Flux profile inference via 2S-^{13}C MFA

Flux profiles were inferred from growth and tracer experiment data using 2-scale-^{13}C metabolic flux analysis (2S-^{13}C MFA)[4]. They, along with specific growth rates and extracellular fluxes were used to characterize the relative effects of the presence of galactose and/or knockout of *SIP1*. 2S-^{13}C MFA was chosen over ^{13}C MFA for its ability to infer cofactor balances and quantify the internal consistency of resulting flux profiles with the carbon transition model and feasible growth. The means and standard deviations for strain/condition-specific extracellular fluxes, intracellular metabolite LCMS fractional labeling distributions, specific growth rates, and feed glucose labeling were used as inputs to in-house code. We used the iMM904 *S. cerevisiae* genome-scale model[19] where applicable. The higher-resolution carbon transition model differed for each strain/condition pair as demanded by the ELVA requirements. Starting from a base core reaction network, carbon transition information was added to reactions (i.e. the reaction was added to the core network) with the largest non-core-to-core flux. This was performed iteratively until ELVA plot simulated data confidence intervals were minimized. A final ^{13}C Flux Variability Analysis (^{13}C FVA) was used to determine individual flux ranges (See Introduction for more details). Our code uses the CONOPT Solver in a GAMS framework to perform the ^{13}C MFA step using 30 initial flux starting points. That resulting in the closest simulated-to-measured labeling distributions is kept for the next step.

2.3 Results

2.3.1 Growth rates

An unexpected increase in growth rate occurred for the base strain in medium containing both glucose and galactose relative to that containing glucose only. This is, to our knowledge, the first reported incidence of increased growth rate upon exposure to galactose in the presence of glucose. The average maximum specific growth rates for the 4 strain/condition pairs are presented in Figure 2.3. An 8.5% increase was observed when the base strain was grown in medium with supplemented galactose instead of glucose-only medium (U versus UG). Similarly, the *sip1∆* mutant grew 18.2% faster in galactose-supplemented medium relative to that without (S versus SG).

Also, the presence of galactose appeared to be necessary for knockout of *SIP1* to increase growth rate. Consistent with the literature[7,8], no change in maximum specific growth rate was observed upon knockout of *SIP1* in glucose-only medium (U versus S). It wasn't until galactose was present that knockout of *SIP1* resulted in an 8.2% higher specific growth rate (UG versus SG).

2.3.2 Extracellular fluxes

Lactate, formate, and succinate were not detected in any strain/pair condition, hence their extracellular fluxes were set to zero. All absolute fluxes were the same for all strain/condition pairs within error except for the ethanol flux of the *sip1∆* mutant. When comparing the base and mutant strains, there appeared to be a 89% decrease in glucose-only medium (U versus S). The addition of galactose to the medium of the *SIP1* null mutant appeared to restore the ethanol flux to its value before the gene knockout (S versus SG), within error. The extracellular flux input ranges, as the mean plus or minus one standard deviation, for the code for strain condition pairs U, S, UG, and SG are presented in the Tables 2.5, 2.6, 2.7, and 2.8, respectively.

2.3.3 Fits and ELVA plots

The External Labeling Variability Analysis (ELVA) plots for all strain/condition pairs are presented in Figure 2.4. Strain/condition pair SG exhibited more variability in both its measured and simulated data errors. Also, SG had a somewhat worse fit.

Detailed fits between simulated and measured LCMS data for metabolites 3-phosho-D-glycerate (3pg), alanine (Ala), arginine (Arg), asparagine (Asp), glutamine (Gln), glutamate (Glu), isoleucine (Ile), leucine (Leu), lysine (Lys), phenylalanine (Phe), threonine (Thr), tyrosine (Tyr), valine (Val), citrate (cit_m), fructose 1,6-bisphosphate (fdp), and succinate (succ_m) are displayed in Figures 2.5, 2.6, 2.7, and 2.8. We decided to exclude the data for citrate from the fitting to see how well the other data predicted its labeling. Overall, the predicted citrate labeling looks good. That for strain/condition SG is a little less tight.
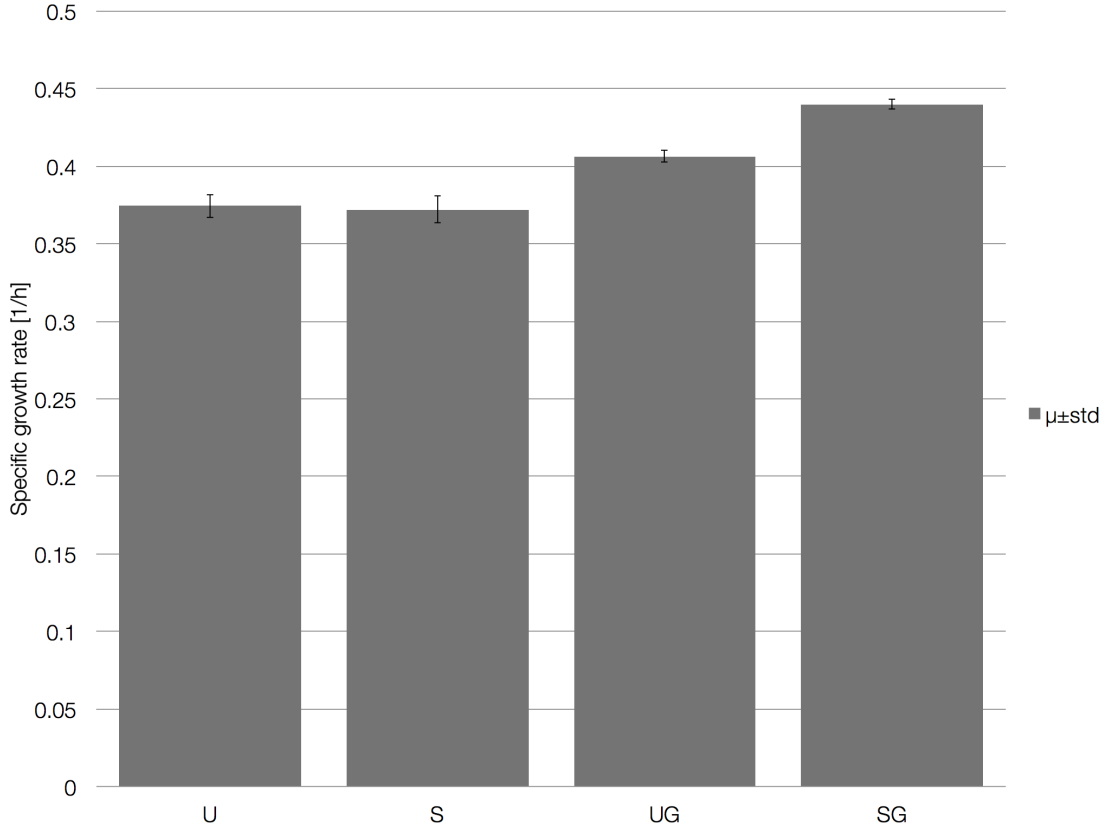
Figure 2.3: Average maximum specific growth rates for the strain/condition pairs U, S, UG, and SG. U refers to the base strain, S refers to the *SIP1* knockout mutant, and the presence of a "G" in the label indicates 2% galactose minimal medium (Min+Gal) (Ex "UG" is the base strain in Min+Gal) instead of glucose-only medium (Min). Error bars represent one standard deviation from the mean

The whole flux profiles corresponding to these strain/condition pair ELVA plots are displayed in the appendix in Figures 2.12, 2.13, 2.14, and 2.15 for U, S, UG, and SG, respectively. All values are normalized to the absolute glucose uptake rate. As indicated in the legend in the lower-right of the figure, differently colored small arrows indicate the use of particular cofactors. Cofactors displayed are NADPH, NADH, ATP, GLN-L, AKG-L, NADP, NAD, ADP, GLU-L, ACCOA (acetyl-CoA), FOR (formate), CO_2, AMP, and CoASH.

2.3.4 Pentose phosphate pathway activity

The presence of galactose appeared to greatly reduce pentose phosphate pathway (PPP) activity. PPP fluxes for strain/condition pairs U, S, UG, and SG are displayed in Figure 2.9. When switching from Min to Min+Gal the base strain's PPP activity reduced by 98% (U

(a) U
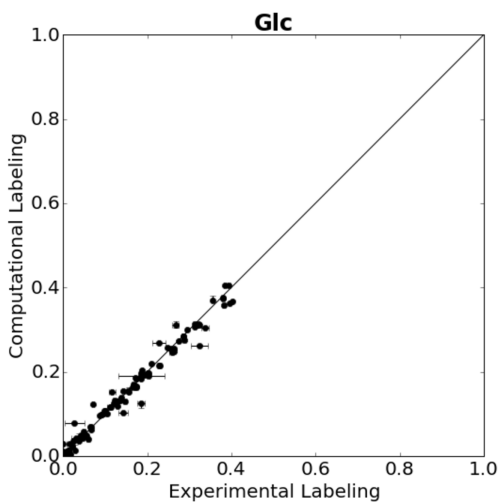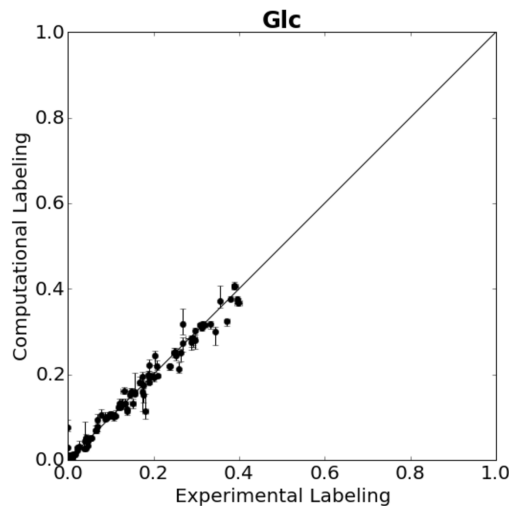
(b) S
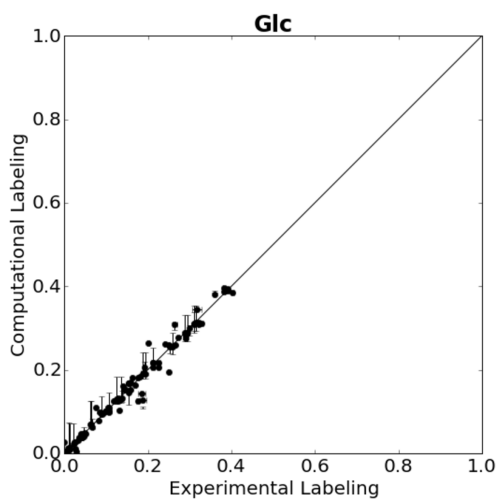
(c) UG
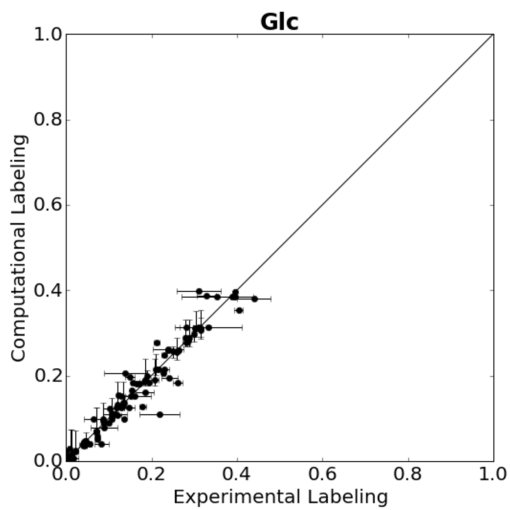
(d) SG

Figure 2.4: External Labeling Variability Analysis (ELVA) plots for strain/condition pairs U, S, UG, and SG. A plot of the simulated versus measured mass isotopomer probabilities for all measured intracellular metabolites. The center line is that corresponding to equal simulated/measured values. Horizontal error bars represent one standard deviation from the mean of measured values. Vertical error bars represent the maximum/minimum possible values of each mass isotopomer labeling probability based on the ELVA.

versus UG). That for the *SIP1* null mutant decreased about 93% when additional galactose was present in the medium (S versus SG).

2.3.5   Inactive glyoxylate and TCA cycles

The glyoxylate cycle, as expected from glucose repression, appeared to be completely repressed across all strain/condition pairs. This was virtually true for the TCA cycle as well. This is mostly consistent with the ^{13}C MFA literature, though these studies tend to indicate a small amount of activity (usually about 1-2% of total glucose consumption flux) in glucose-repressing conditions for CEN.PK113-7D[20–23]. A comparison of fluxes surrounding mitochondrial import and export is displayed for strain/condition pairs U, S, UG, and SG in Figure 2.10. Only the mitochondrial conversion of malate to oxaloacetate and then to citrate appears to have a small amount of flux (~2% of total glucose flux).

2.3.6   Mitochondrial import/export

Our analysis appears to indicate, for the base strain in glucose-repressing conditions, most mitochondrial flux flows from cytosolic pyruvate to oxaloacetate via pyruvate carboxylase, then to cytosolic malate through malate dehydrogenase (MDH), which is then imported into the mitochondria where most of it is converted to mitochondrial pyruvate through NAD-dependent malic enzyme, and is finally exported to the cytosol.

Interestingly, the addition of galactose to the medium at 0.2% appears to increase flow through MDH, subsequent import of malate into the mitochondria, reroutes flow through NADP-dependent malic enzyme, and switches to import of pyruvate to the mitochondria for redirection towards export of 3-methyl-2-oxobutanote to the cytosol and branched-chain amino acid biosynthesis. Import of malate into the mitochondria increased by 490% for the base strain (U versus UG). That for the mutant (S versus SG) went from virtually no import to import.

Despite no interesting growth phenotype, in glucose-only medium, deletion of *SIP1* appears to reduce mitochondrial activity almost to inactivation (U versus S). There was also a less pronounced decrease in the rate of mitochondrial import of malate by 60% in medium containing both carbon sources (UG versus SG).

2.3.7   Asparagine/threonine biosynthesis

In addition to the correlated decrease in mitochondrial activity upon deletion of *SIP1*, in glucose-only medium flux was rerouted from the mitochondria to the part of the metabolic network responsible for asparagine and threonine biosynthesis (U versus S). This section of the network is displayed in Figure 2.11 for all strain/condition pairs. The same trend occurred in glucose/galactose medium with a 165% increase in flux towards this part of the network. Finally, the addition of galactose to the *sip1Δ* mutant appears to have reversed this trend with a 77% reduction in flow (S versus SG).

2.4 Discussion

We encountered a number of unexpected phenotypic differences in this study. The addition of galactose to normally glucose-repressing conditions resulted in unreported differences in growth and flux profile phenotypes in both base and mutant strains. Knockout of *SIP1* in glucose-only conditions resulted in no change in specific growth rate as expected, yet flux profile redistributed and changes in ethanol flux resulted. Despite changes noted in this study, glucose repression still appeared to result largely in ethanol fermentation and repression of TCA and glyoxylate cycle activity in all strain/condition pairs.

SIP1 appears to be an obligatory footnote in yeast glucose repression literature. If mentioned at all, its mostly described as the β-subunit of the Snf1 kinase complex. Sometimes details about its localization or role in sequestering the complex in the vacuole are mentioned but it appears to be largely ignored. We suspect this is due to the lack of growth phenotype accompanying knockout of *SIP1*. To our knowledge this and our previous work[9] are the only studies to even attempt to monitor excreted metabolites and infer fluxes in a *SIP1* null mutant. Hence, we were surprised when deletion of *SIP1* in glucose-only medium appeared to effectively shut down mitochondrial fluxes, reroute flow towards asparagine and threonine biosynthesis, and decrease absolute ethanol production. The same effects occurred in glucose and galactose medium but to a lesser extent.

Even more surprising, the addition of galactose to the medium containing the base strain resulted in increased growth rate, in what would normally be considered glucose-repressing conditions. More so, upon the addition of galactose, pentose phosphate activity dropped to about 1/100 of its original value, rate of import of malate into the mitochondria increased, a resulting switch from NAD- to NADPH-dependent malic enzyme occurred, and flux was directed toward mitochondrial 3-methyl-2-oxobutanote export and subsequent flow towards branched-chain amino acid biosynthesis. This occurred for both strains.

We suspect this apparent violation of glucose repression was plausible based on a paper from earlier this year. Escalante-Chong et. al.[24] found that the space of galactose-to-glucose ratios were sparsely sampled in the literature and that a particular galactose/glucose concentration ratio resulted in YFP expression under a *GAL1* promoter. The beginning of this expression activation happened to occur at the same 1/10 ratio of galactose-to-glucose used in this study. It's possible that the results of that study might not apply since they monitored growth in microwell plates and their strain background was S288c while this study took place in shake flasks using a CEN.PK113-7D background. CEN.PK113-7D is known to exhibit different phenotypic differences relative to S288c[25], though the Escalante study at least makes it plausible that a similar effect might be occurring. Perhaps the phenotypic changes noted in this study represent some sort of sensing of and preparation for degradation of galactose. Further studies using strains with an intact *GAL1* gene that can catabolize galactose might reveal more information on this.

Our original hypothesis, that the deletion of *SIP1* allowed galactose to enter the cell and affect growth by increasing expression of its transporter, appears to be incorrect. This assumes that the *SIP1* knockout was necessary to see an effect from the presence of galactose.

In fact, the opposite was observed; the presence of galactose was required to see a growth phenotype from knockout of *SIP1*. Sip1 is known to be a negative regulator of *GAL* gene activity[10]. Perhaps the absence of Sip1 allowed the effect of galactose to be more pronounced.

The cytosolic/mitochondrial cycle of pyruvate for the base strain in glucose-only medium through the mitochondria is itself interesting. This route diverts carbon flux away from ethanol production through two NADH-producing pathways. It's possible the cell is routing flux through this cycle to produce more NADH for ethanol fermentation, since both MDH and NAD-dependent malic enzyme produce this cofactor. It should be noted that most ^{13}C metabolic flux analysis studies do not include these mitochondrial transport reactions. In fact, the initial carbon transition model we used didn't include them. It was only through the process of adding them to tighten the simulated data ranges in the ELVA plot and then to the map visualizations did it become apparent that this cycle was occurring.

This is the first study to investigate the relative effects of the presence of galactose and knockout of *SIP1* in normally carbon repressing conditions from a fluxomic perspective. It's also one of the first to apply the 2-scale version of ^{13}C metabolic flux analysis to yeast, allowing an unprecedented confidence in the self-consistency of the mass isotopomer model and data. We also encountered unreported increases in growth rate upon the addition of galactose to normally glucose-repressing medium.

2.5   Acknowledgements
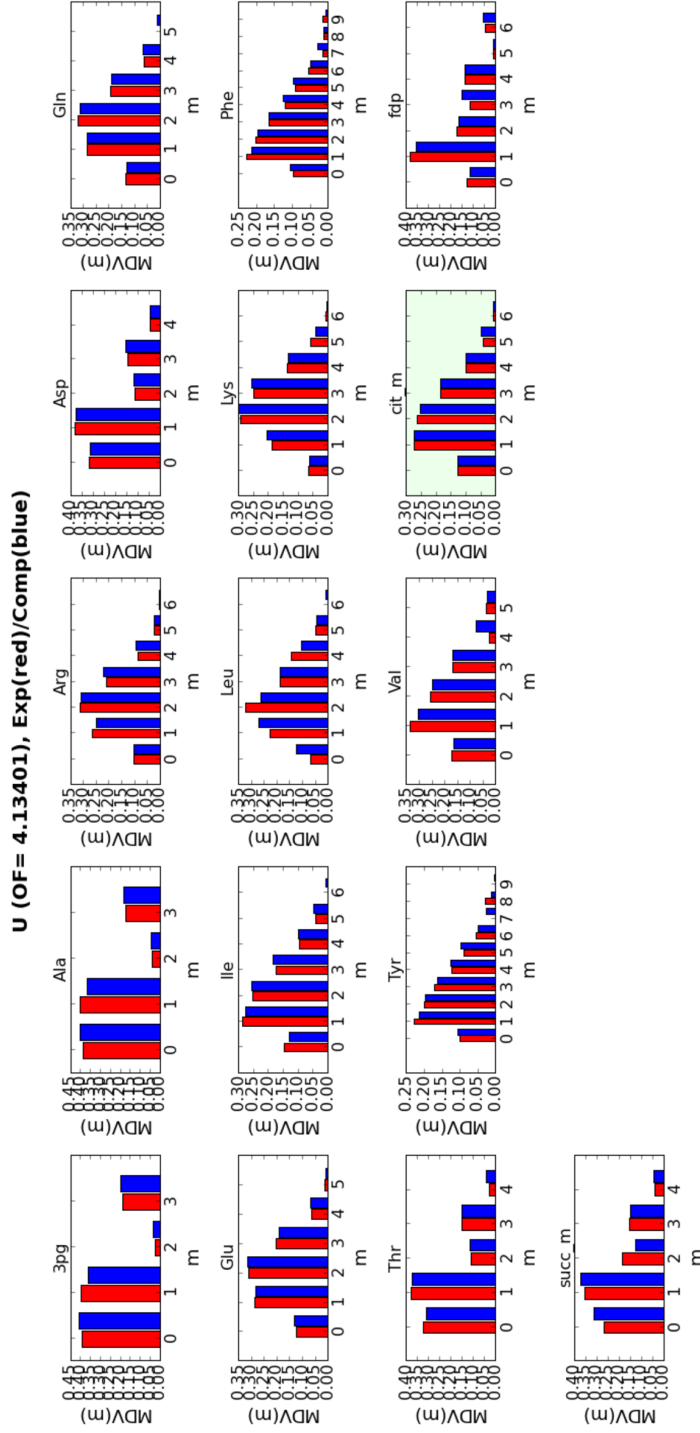
Figure 2.5: Detailed fits between simulated (blue bars) and measured (red bars) intracellular metabolite labeling distributions for strain/condition pair U. The green box corresponds to a metabolite whose measured data was excluded from computations and the predicted labeling was compared to that measured
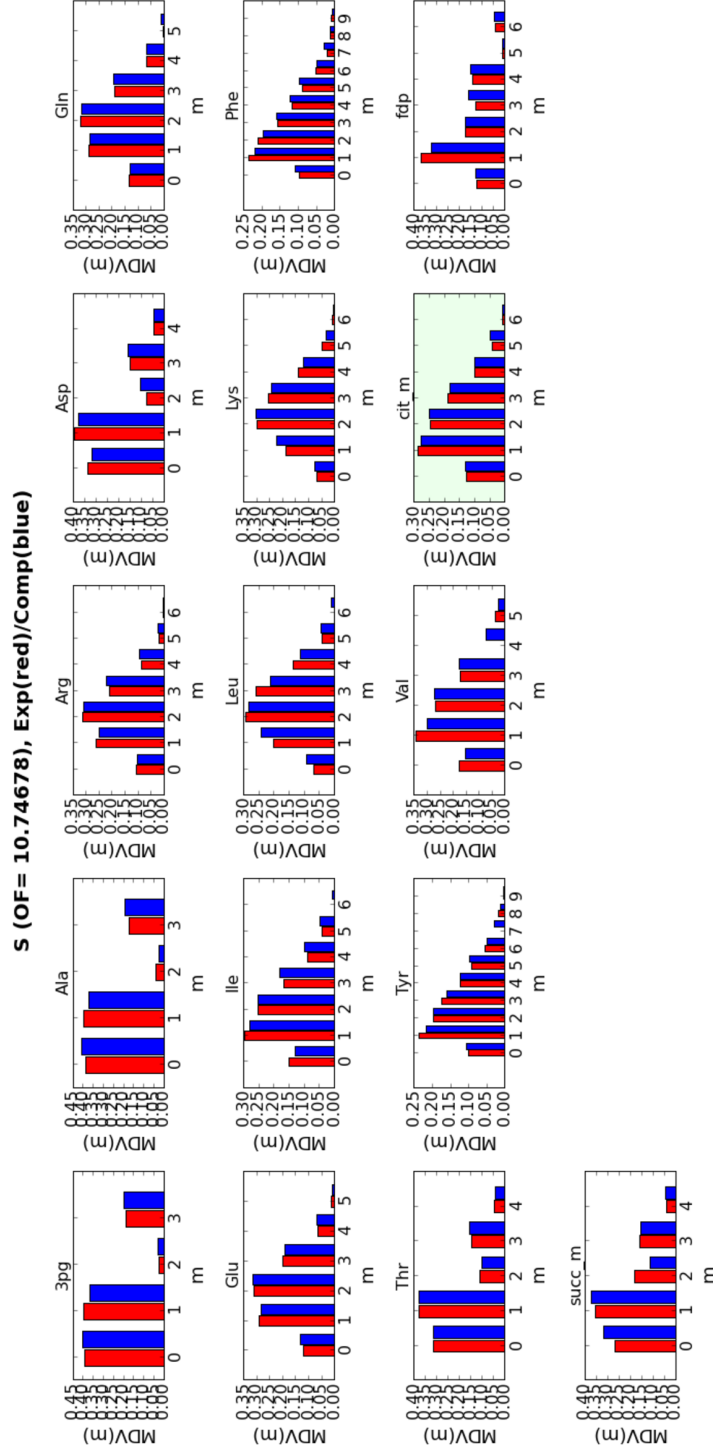
Figure 2.6: Detailed fits between simulated and measured intracellular metabolite labeling distributions for strain/condition pair S. The green box corresponds to a metabolite whose measured data was excluded from computations and the predicted labeling was compared to that measured
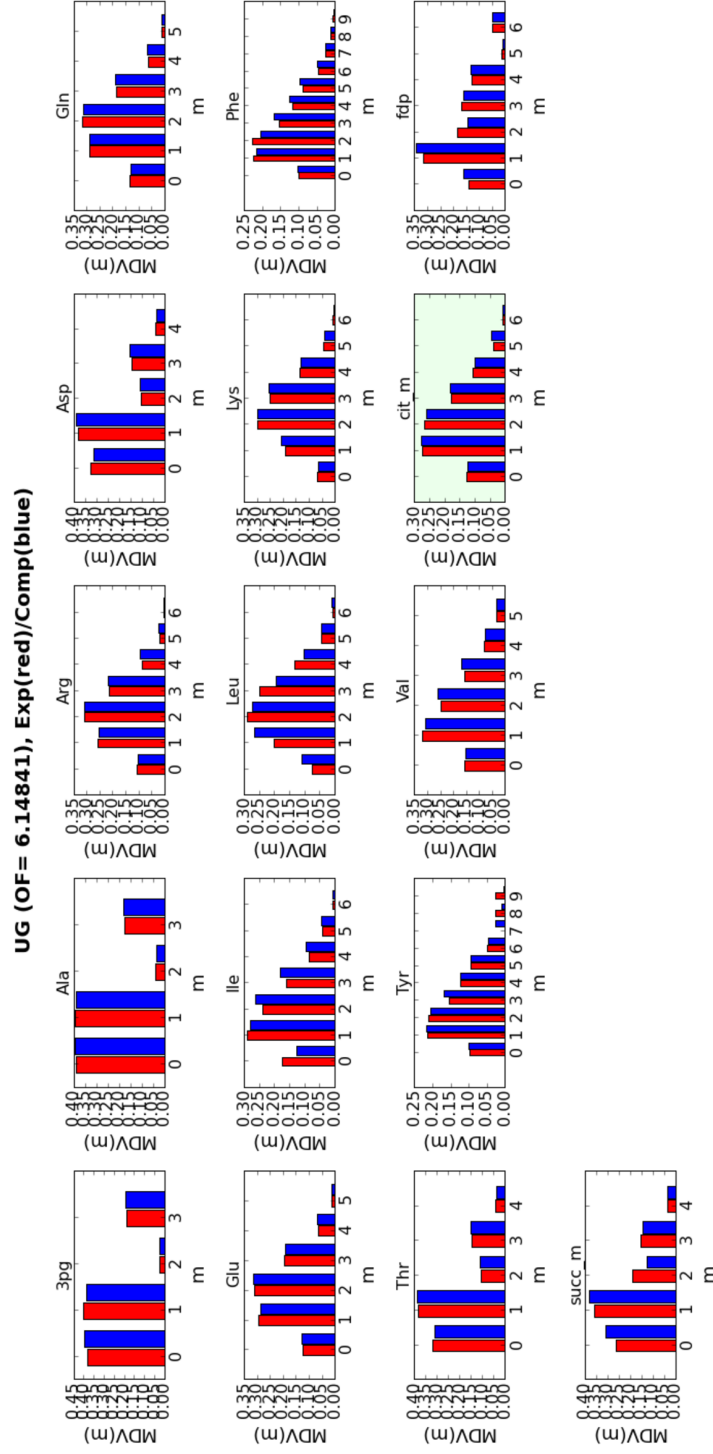
Figure 2.7: Detailed fits between simulated and measured intracellular metabolite labeling distributions for strain/condition pair UG. The green box corresponds to a metabolite whose measured data was excluded from computations and the predicted labeling was compared to that measured
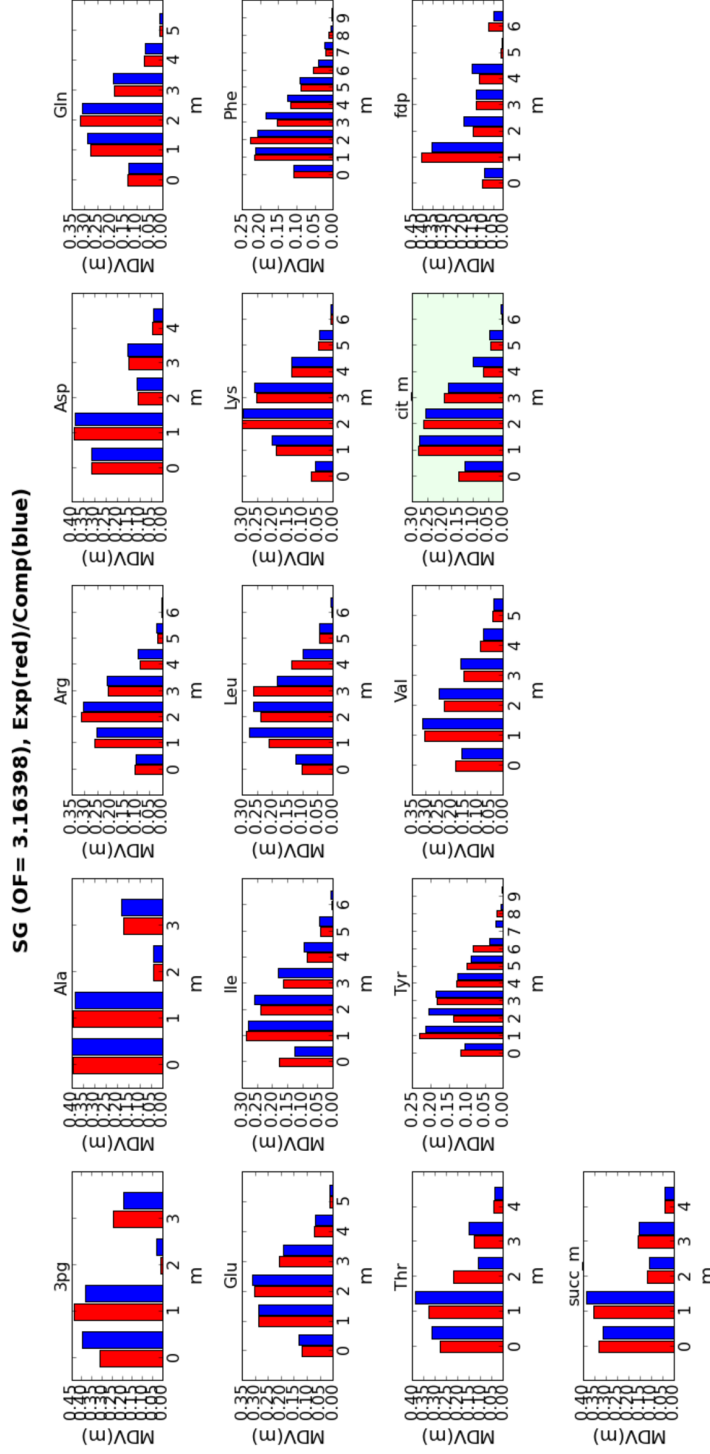
Figure 2.8: Detailed fits between simulated and measured intracellular metabolite labeling distributions for strain/condition pair SG. The green box corresponds to a metabolite whose measured data was excluded from computations and the predicted labeling was compared to that measured

(a) U

(b) S

(c) UG

(d) SG
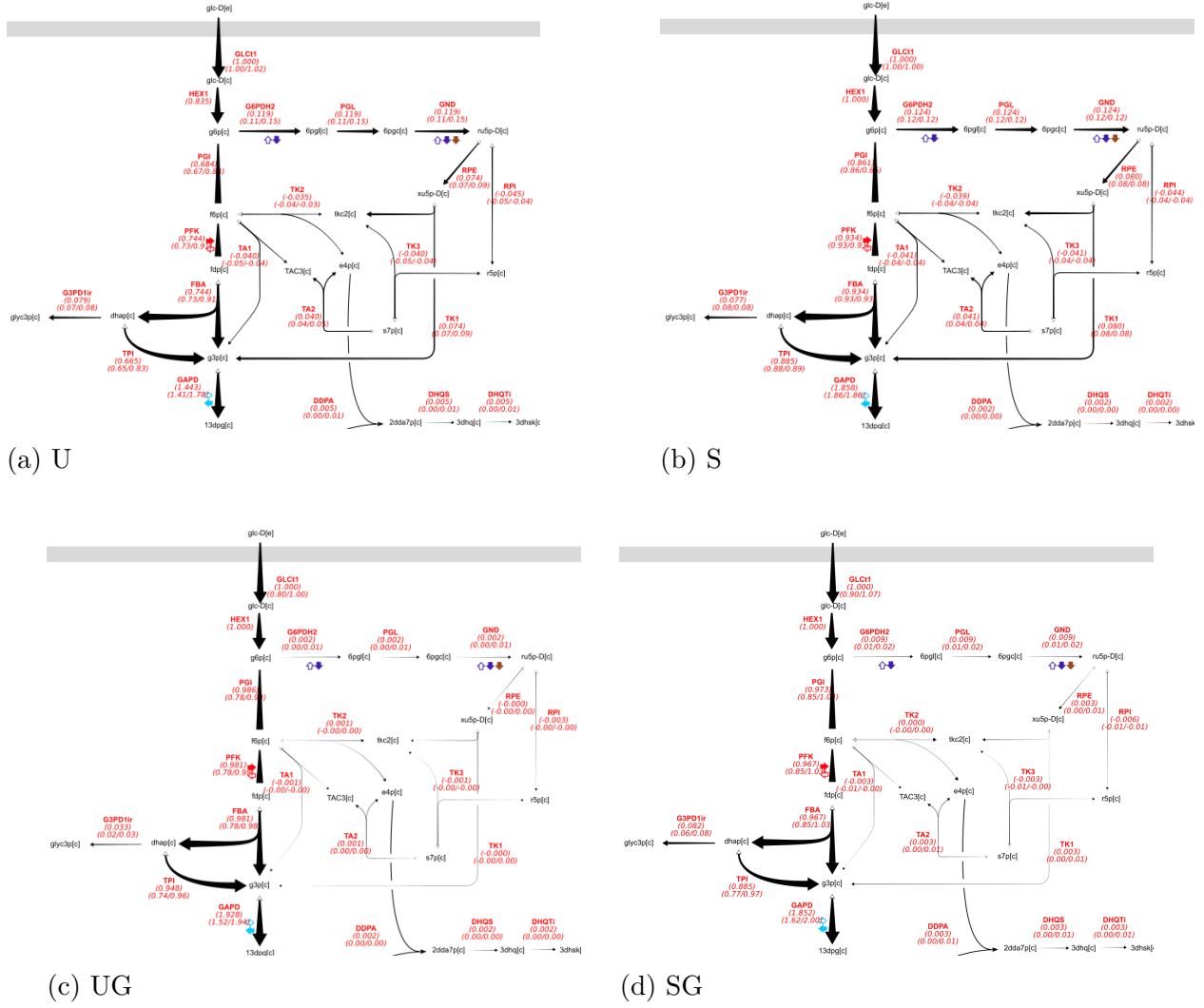
Figure 2.9: PPP fluxes for strain/condition pairs U, S, UG, and SG

(a) U
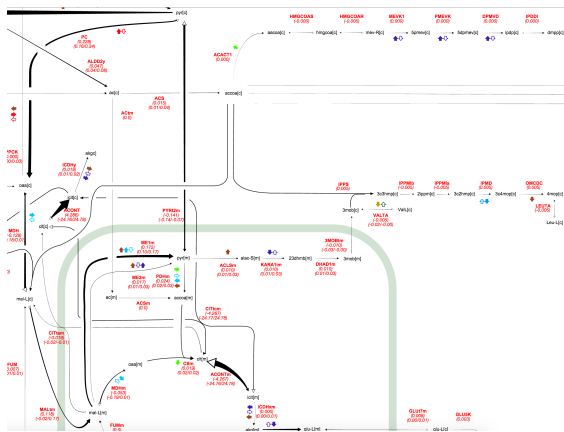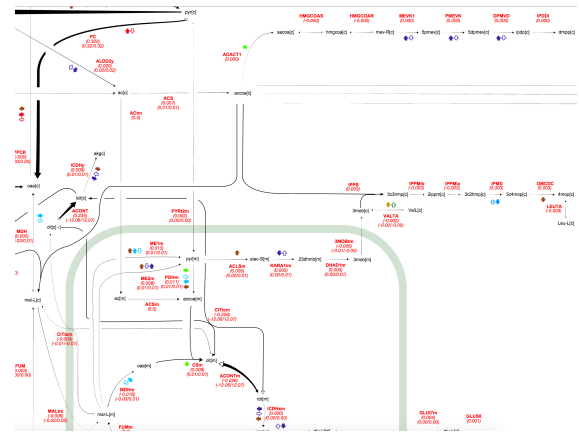
(b) S
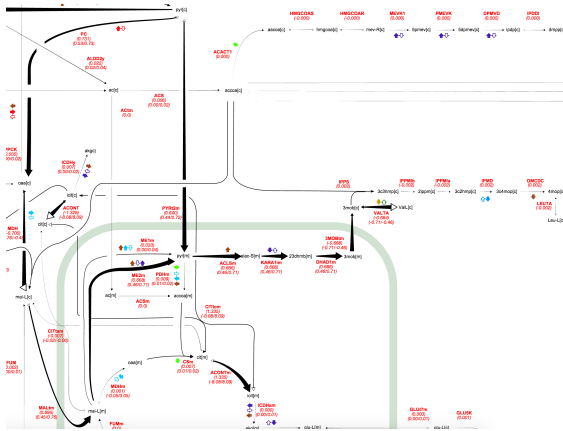
(c) UG
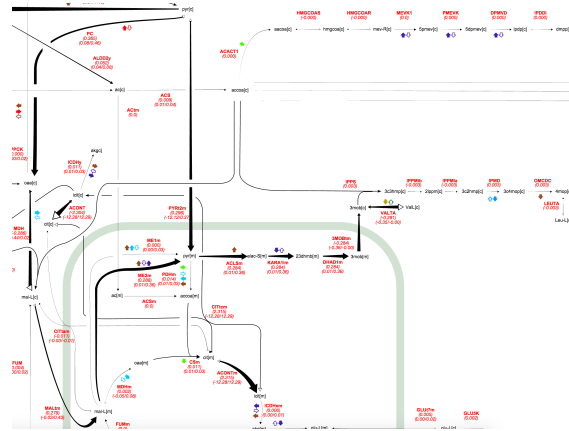
(d) SG

Figure 2.10: Mitochondrial import/export fluxes for strain/condition pairs U, S, UG, and SG

(a) U

(b) S

(c) UG

(d) SG
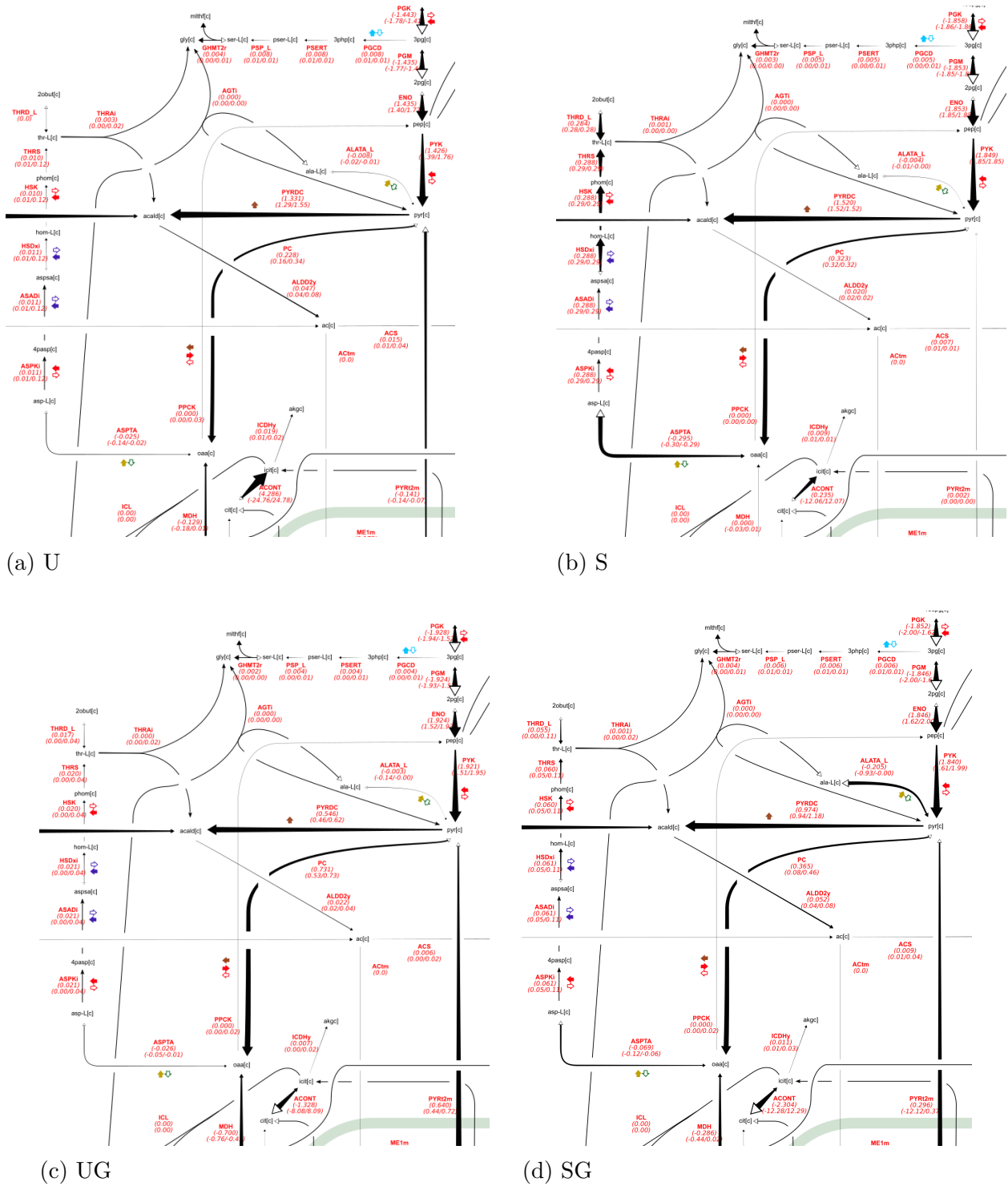
Figure 2.11: Fluxes near asparagine and threonine biosynthesis for strain/condition pairs U, S, UG, and SG

2.6 Appendix

Table 2.5: Extracellular metabolite flux bounds as the mean plus or minus one standard deviation for strain/condition pair U

```
biomass_SC5_notrace:     0.368076597958 [==] 0.380967956696
GLCt1:  20.1775027573 [==] 32.7249994091
EX_glc(e):      -32.7249994091 [==] -20.1775027573
EX_ac(e):       0.649171302952 [==] 0.807467832378
EX_etoh(e):     18.7087083907 [==] 22.4121111429
EX_for(e):      -0.042937921881 [==] 0.0115051814927
EX_glyc(e):     1.39040825811 [==] 1.58632290283
EX_succ(e):     0.0 [==] 0.0
EX_lac_L(e):    0.0 [==] 0.0
EX_fum(e):      0.0 [==] 0.0
EX_mal_L_e_:    0.0 [==] 0.0
ACt2r:  0.0 [==] 0.0
```

Table 2.6: Extracellular metabolite flux bounds as the mean plus or minus one standard deviation for strain/condition pair S

```
biomass_SC5_notrace:     0.364770870897 [==] 0.379551153717
GLCt1:  5.54594695637 [==] 41.4228038048
EX_glc(e):      -41.4228038048 [==] -5.54594695637
EX_ac(e):       0.545019159258 [==] 0.7502366638
EX_etoh(e):     -0.0805353085846 [==] 4.74144877392
EX_for(e):      0 [==] 0
EX_glyc(e):     -0.978472290387 [==] 2.01628121859
EX_succ(e):     0.0 [==] 0.0
EX_lac_L(e):    0.0 [==] 0.0
EX_fum(e):      0.0 [==] 0.0
EX_mal_L_e_:    0.0 [==] 0.0
ACt2r:  0.0 [==] 0.0
```

Table 2.7: Extracellular metabolite flux bounds as the mean plus or minus one standard deviation for strain/condition pair UG

```
biomass_SC5_notrace:    0.403289617257 [==] 0.409715634064
GLCt1:  12.7988591665 [==] 61.8351699967
EX_glc(e):      -61.8351699967 [==] -12.7988591665
EX_ac(e):       1.02930219015 [==] 1.21244774938
EX_etoh(e):     12.1449617938 [==] 21.3745300987
EX_for(e):      0.0 [==] 0.0
EX_glyc(e):     1.38970011823 [==] 2.03596203336
EX_succ(e):     0.0 [==] 0.0
EX_lac_L(e):    0.0 [==] 0.0
EX_fum(e):      0.0 [==] 0.0
EX_mal_L_e_:    0.0 [==] 0.0
ACt2r:  0.0 [==] 0.0
```

Table 2.8: Extracellular metabolite flux bounds as the mean plus or minus one standard deviation for strain/condition pair SG

```
biomass_SC5_notrace:    0.437253740884 [==] 0.442552150354
GLCt1:  36.7422490021 [==] 43.3764465986
EX_glc(e):      -43.3764465986 [==] -36.7422490021
EX_ac(e):       1.42911300659 [==] 1.78851951154
EX_etoh(e):     17.4413404772 [==] 20.9769286945
EX_for(e):      0.0 [==] 0.0
EX_glyc(e):     2.32527136848 [==] 3.31789211791
EX_succ(e):     0.0 [==] 0.0
EX_lac_L(e):    0.0 [==] 0.0
EX_fum(e):      0.0 [==] 0.0
EX_mal_L_e_:    0.0 [==] 0.0
ACt2r:  0.0 [==] 0.0
```

Table 2.9: Intracellular metabolite labeling distribution input for strain/condition U

	Amino acid	Mass distribution									
		m0	m1	m2	m3	m4	m5	m6	m7	m8	m9
3pg	M-0	0.3936	0.3963	0.0238	0.1863						
Ala	M-0	0.3854	0.4034	0.0398	0.1713						
Arg	M-0	0.1045	0.2643	0.3117	0.2089	0.0864	0.0217	0.0026			
Asp	M-0	0.3176	0.3810	0.1122	0.1429	0.0464					
Gln	M-0	0.1352	0.2860	0.3219	0.1935	0.0634	0.0000				
Glu	M-0	0.1219	0.2855	0.3123	0.2021	0.0655	0.0127				
Ile	M-0	0.1466	0.2877	0.2536	0.1739	0.0954	0.0427	0.0000			
Leu	M-0	0.0705	0.2265	0.3238	0.1862	0.1434	0.0496	0.0000			
Lys	M-0	0.0662	0.1883	0.2944	0.2476	0.1359	0.0597	0.0080			
Phe	M-0	0.0979	0.2272	0.2021	0.1656	0.1207	0.0898	0.0533	0.0150	0.0127	0.0158
Thr	M-0	0.3247	0.3787	0.1133	0.1531	0.0302					
Tyr	M-0	0.1012	0.2292	0.1996	0.1724	0.1220	0.0900	0.0563	0.0000	0.0293	0.0000
Val	M-0	0.1727	0.3367	0.2583	0.1684	0.0267	0.0371				
fdp	M-0	0.1288	0.3830	0.1744	0.1164	0.1365	0.0119	0.0489			
cit_m	M-0	0.1279	0.2737	0.2643	0.1841	0.0993	0.0422	0.0085			
succ_m	M-0	0.2665	0.3554	0.1848	0.1554	0.0379					

Table 2.10: Intracellular metabolite labeling distribution standard deviation input for strain/condition U

		m0	m1	m2	m3	m4	m5	m6	m7	m8	m9
3pg	M-0	0.0059	0.0053	0.0020	0.0047						
Ala	M-0	0.0040	0.0025	0.0013	0.0005						
Arg	M-0	0.0023	0.0088	0.0046	0.0025	0.0007	0.0003	0.0000			
Asp	M-0	0.0082	0.0063	0.0130	0.0054	0.0041					
Gln	M-0	0.0017	0.0069	0.0025	0.0009	0.0032	0.0000				
Glu	M-0	0.0077	0.0091	0.0083	0.0053	0.0019	0.0010				
Ile	M-0	0.0076	0.0066	0.0045	0.0057	0.0032	0.0015	0.0000			
Leu	M-0	0.0047	0.0323	0.0396	0.1083	0.0219	0.0117	0.0000			
Lys	M-0	0.0005	0.0025	0.0030	0.0040	0.0011	0.0094	0.0007			
Phe	M-0	0.0034	0.0092	0.0047	0.0046	0.0033	0.0020	0.0029	0.0150	0.0009	0.0071
Thr	M-0	0.0070	0.0083	0.0030	0.0021	0.0198					
Tyr	M-0	0.0018	0.0085	0.0017	0.0027	0.0026	0.0014	0.0019	0.0033	0.0033	0.0033
Val	M-0	0.0090	0.0171	0.0108	0.0097	0.0463	0.0027				
fdp	M-0	0.0032	0.0090	0.0023	0.0156	0.0049	0.0002	0.0058			
cit_m	M-0	0.0026	0.0037	0.0004	0.0005	0.0027	0.0007	0.0004			
succ_m	M-0	0.0159	0.0053	0.0204	0.0015	0.0008					

Table 2.11: Intracellular metabolite labeling distribution input for strain/condition S

	Amino acid	Mass distribution									
		m0	m1	m2	m3	m4	m5	m6	m7	m8	m9
3pg	M-0	0.3895	0.3977	0.0258	0.1870						
Ala	M-0	0.3879	0.3976	0.0405	0.1740						
Arg	M-0	0.1063	0.2631	0.3121	0.2083	0.0866	0.0210	0.0026			
Asp	M-0	0.3328	0.3948	0.0780	0.1485	0.0460					
Gln	M-0	0.1341	0.2882	0.3199	0.1895	0.0655	0.0028				
Glu	M-0	0.1209	0.2922	0.3090	0.2011	0.0646	0.0122				
Ile	M-0	0.1524	0.2956	0.2512	0.1680	0.0902	0.0425	0.0000			
Leu	M-0	0.0684	0.2018	0.2916	0.2578	0.1370	0.0435	0.0000			
Lys	M-0	0.0680	0.1893	0.2972	0.2526	0.1391	0.0466	0.0071			
Phe	M-0	0.0972	0.2358	0.2087	0.1565	0.1181	0.0883	0.0518	0.0213	0.0126	0.0098
Thr	M-0	0.3160	0.3789	0.1106	0.1490	0.0456					
Tyr	M-0	0.1019	0.2377	0.1967	0.1728	0.1223	0.0920	0.0556	0.0000	0.0171	0.0039
Val	M-0	0.1759	0.3450	0.2679	0.1728	0.0000	0.0384				
fdp	M-0	0.1239	0.3716	0.1749	0.1313	0.1445	0.0128	0.0410			
cit_m	M-0	0.1269	0.2876	0.2476	0.1894	0.0992	0.0420	0.0072			
succ_m	M-0	0.2687	0.3553	0.1809	0.1562	0.0388					

Table 2.12: Intracellular metabolite labeling distribution standard deviation input for strain/condition S

	Amino acid	Mass distribution									
		m0	m1	m2	m3	m4	m5	m6	m7	m8	m9
3pg	M-0	0.0140	0.0164	0.0025	0.0046						
Ala	M-0	0.0021	0.0035	0.0002	0.0031						
Arg	M-0	0.0035	0.0063	0.0071	0.0087	0.0009	0.0006	0.0001			
Asp	M-0	0.0054	0.0025	0.0006	0.0047	0.0046					
Gln	M-0	0.0054	0.0284	0.0129	0.0073	0.0029	0.0000				
Glu	M-0	0.0061	0.0058	0.0019	0.0005	0.0022	0.0002				
Ile	M-0	0.0134	0.0035	0.0046	0.0049	0.0018	0.0003	0.0000			
Leu	M-0	0.0032	0.0061	0.0099	0.0020	0.0016	0.0031	0.0000			
Lys	M-0	0.0009	0.0024	0.0012	0.0018	0.0017	0.0015	0.0003			
Phe	M-0	0.0027	0.0037	0.0026	0.0008	0.0022	0.0010	0.0123	0.0007	0.0002	0.0010
Thr	M-0	0.0057	0.0029	0.0039	0.0010	0.0009					
Tyr	M-0	0.0024	0.0037	0.0066	0.0018	0.0013	0.0007	0.0010	0.0000	0.0083	0.0039
Val	M-0	0.0050	0.0015	0.0019	0.0030	0.0000	0.0007				
fdp	M-0	0.0030	0.0127	0.0037	0.0156	0.0024	0.0001	0.0008			
cit_m	M-0	0.0096	0.0103	0.0125	0.0074	0.0044	0.0033	0.0005			
succ_m	M-0	0.0078	0.0049	0.0128	0.0006	0.0005					

Table 2.13: Intracellular metabolite labeling distribution input for strain/condition UG

	Amino acid	Mass distribution									
		m0	m1	m2	m3	m4	m5	m6	m7	m8	m9
3pg	M-0	0.3833	0.4029	0.0239	0.1898						
Ala	M-0	0.3905	0.3920	0.0407	0.1768						
Arg	M-0	0.1058	0.2587	0.3098	0.2134	0.0883	0.0213	0.0026			
Asp	M-0	0.3278	0.3820	0.1051	0.1435	0.0415					
Gln	M-0	0.1345	0.2884	0.3160	0.1879	0.0609	0.0122				
Glu	M-0	0.1248	0.2953	0.3100	0.1944	0.0644	0.0110				
Ile	M-0	0.1753	0.2909	0.2398	0.1615	0.0842	0.0406	0.0077			
Leu	M-0	0.0756	0.2009	0.2896	0.2496	0.1318	0.0440	0.0084			
Lys	M-0	0.0674	0.1910	0.3000	0.2515	0.1372	0.0456	0.0073			
Phe	M-0	0.0991	0.2250	0.2263	0.1539	0.1169	0.0894	0.0456	0.0266	0.0113	0.0059
Thr	M-0	0.3213	0.3835	0.1072	0.1459	0.0421					
Tyr	M-0	0.0973	0.2133	0.2118	0.1540	0.1237	0.0939	0.0515	0.0000	0.0264	0.0281
Val	M-0	0.1582	0.3213	0.2497	0.1577	0.0807	0.0324				
fdp	M-0	0.1411	0.3165	0.1850	0.1696	0.1283	0.0119	0.0476			
cit_m	M-0	0.1279	0.2726	0.2662	0.1801	0.1058	0.0403	0.0073			
succ_m	M-0	0.2636	0.3593	0.1876	0.1531	0.0364					

Table 2.14: Intracellular metabolite labeling distribution standard deviation input for strain/condition UG

	Amino acid	Mass distribution									
		m0	m1	m2	m3	m4	m5	m6	m7	m8	m9
3pg	M-0	0.0078	0.0083	0.0011	0.0127						
Ala	M-0	0.0026	0.0016	0.0019	0.0011						
Arg	M-0	0.0013	0.0013	0.0013	0.0013	0.0012	0.0003	0.0000			
Asp	M-0	0.0112	0.0054	0.0036	0.0030	0.0017					
Gln	M-0	0.0012	0.0028	0.0005	0.0021	0.0004	0.0007				
Glu	M-0	0.0009	0.0013	0.0003	0.0012	0.0004	0.0008				
Ile	M-0	0.0067	0.0038	0.0024	0.0020	0.0008	0.0010	0.0002			
Leu	M-0	0.0019	0.0042	0.0029	0.0025	0.0009	0.0010	0.0009			
Lys	M-0	0.0008	0.0014	0.0012	0.0007	0.0021	0.0010	0.0003			
Phe	M-0	0.0058	0.0082	0.0046	0.0110	0.0042	0.0040	0.0020	0.0003	0.0004	0.0002
Thr	M-0	0.0017	0.0021	0.0023	0.0022	0.0019					
Tyr	M-0	0.0015	0.0105	0.0076	0.0021	0.0017	0.0025	0.0010	0.0000	0.0028	0.0022
Val	M-0	0.0012	0.0021	0.0018	0.0007	0.0031	0.0003				
fdp	M-0	0.0102	0.0208	0.0196	0.0036	0.0015	0.0003	0.0025			
cit_m	M-0	0.0015	0.0051	0.0045	0.0026	0.0019	0.0012	0.0003			
succ_m	M-0	0.0108	0.0046	0.0146	0.0015	0.0007					

Table 2.15: Intracellular metabolite labeling distribution input for strain/condition SG

	Amino acid	Mass distribution								
		m0	m1	m2	m3	m4	m5	m6	m7	m8	m9
3pg	M-0	0.3098	0.4402	0.0082	0.2419						
Ala	M-0	0.3950	0.3952	0.0386	0.1712						
Arg	M-0	0.1056	0.2602	0.3153	0.2094	0.0859	0.0211	0.0025			
Asp	M-0	0.3115	0.3878	0.1058	0.1506	0.0444					
Gln	M-0	0.1331	0.2794	0.3163	0.1864	0.0712	0.0135				
Glu	M-0	0.1201	0.2866	0.3043	0.2070	0.0707	0.0113				
Ile	M-0	0.1795	0.2859	0.2384	0.1637	0.0869	0.0405	0.0050			
Leu	M-0	0.1024	0.2124	0.2385	0.2606	0.1348	0.0464	0.0050			
Lys	M-0	0.0717	0.1888	0.2991	0.2506	0.1352	0.0473	0.0072			
Phe	M-0	0.1084	0.2161	0.2267	0.1552	0.1168	0.0882	0.0547	0.0209	0.0112	0.0018
Thr	M-0	0.2808	0.3283	0.2193	0.1294	0.0423					
Tyr	M-0	0.1186	0.2301	0.1383	0.1833	0.1288	0.1009	0.0830	0.0000	0.0170	0.0000
Val	M-0	0.1842	0.3021	0.2294	0.1528	0.0894	0.0421				
fdp	M-0	0.1068	0.4036	0.1500	0.1331	0.1220	0.0122	0.0723			
cit_m	M-0	0.1483	0.2825	0.2626	0.1938	0.0647	0.0431	0.0051			
succ_m	M-0	0.3319	0.3541	0.1146	0.1602	0.0392					

Table 2.16: Intracellular metabolite labeling distribution standard deviation input for strain/condition SG

|	Amino acid	Mass distribution								
		m0	m1	m2	m3	m4	m5	m6	m7	m8	m9
3pg	M-0	0.1021	0.0786	0.0115	0.0377						
Ala	M-0	0.0041	0.0034	0.0031	0.0033						
Arg	M-0	0.0006	0.0011	0.0016	0.0027	0.0002	0.0003	0.0000			
Asp	M-0	0.0058	0.0051	0.0040	0.0031	0.0007					
Gln	M-0	0.0024	0.0082	0.0063	0.0041	0.0119	0.0015				
Glu	M-0	0.0029	0.0088	0.0117	0.0180	0.0053	0.0011				
Ile	M-0	0.0146	0.0043	0.0057	0.0035	0.0013	0.0008	0.0035			
Leu	M-0	0.0319	0.0145	0.0725	0.0214	0.0052	0.0035	0.0037			
Lys	M-0	0.0047	0.0015	0.0024	0.0024	0.0024	0.0012	0.0004			
Phe	M-0	0.0117	0.0029	0.0087	0.0069	0.0044	0.0044	0.0071	0.0149	0.0004	0.0025
Thr	M-0	0.0311	0.0452	0.0939	0.0161	0.0047					
Tyr	M-0	0.0208	0.0225	0.0981	0.0209	0.0133	0.0088	0.0316	0.0000	0.0240	0.0000
Val	M-0	0.0445	0.0059	0.0112	0.0017	0.0634	0.0118				
fdp	M-0	0.0074	0.0198	0.0230	0.0059	0.0085	0.0006	0.0053			
cit_m	M-0	0.0252	0.0165	0.0122	0.0006	0.0458	0.0020	0.0023			
succ_m	M-0	0.1574	0.1669	0.0561	0.0755	0.0185					
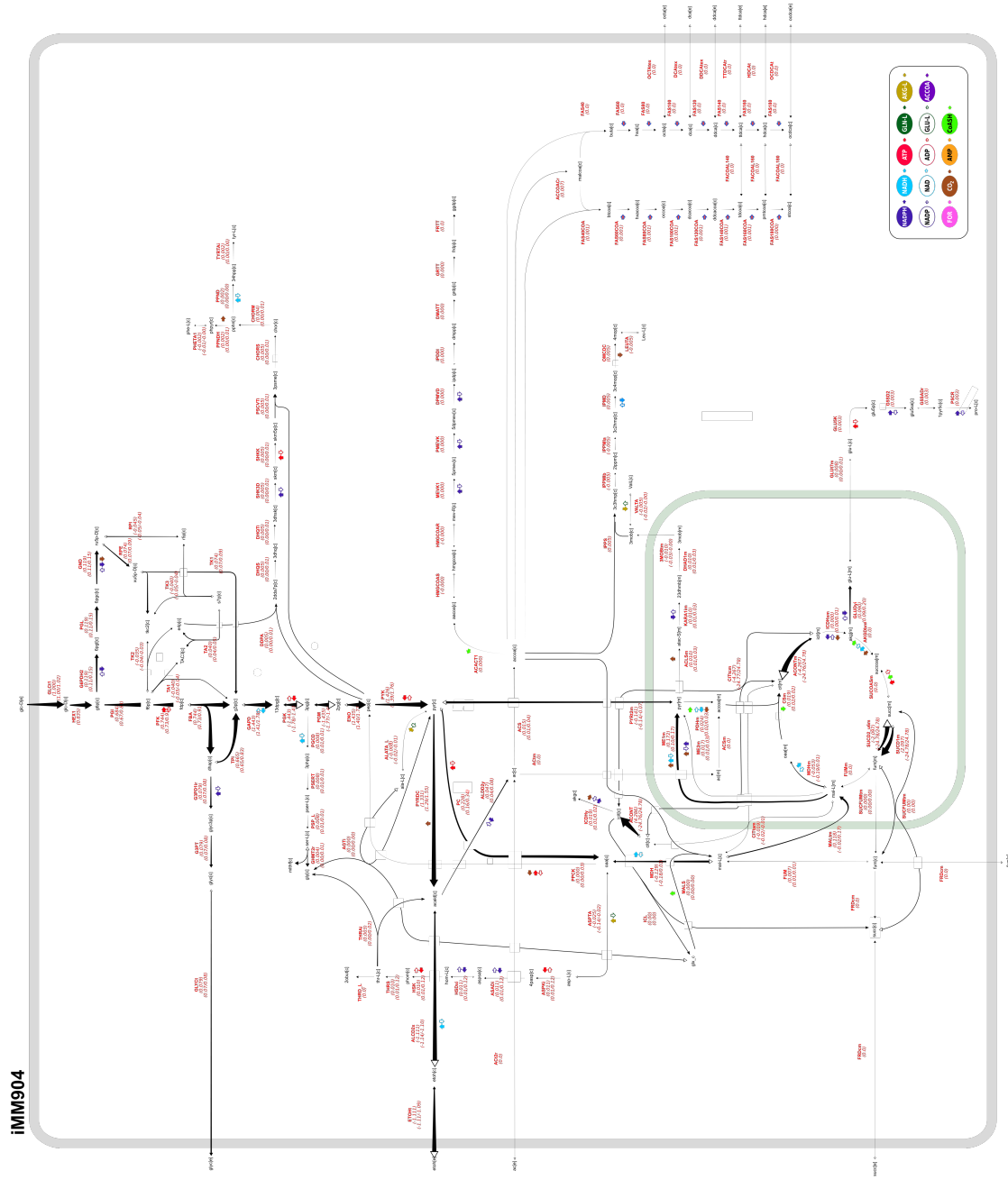
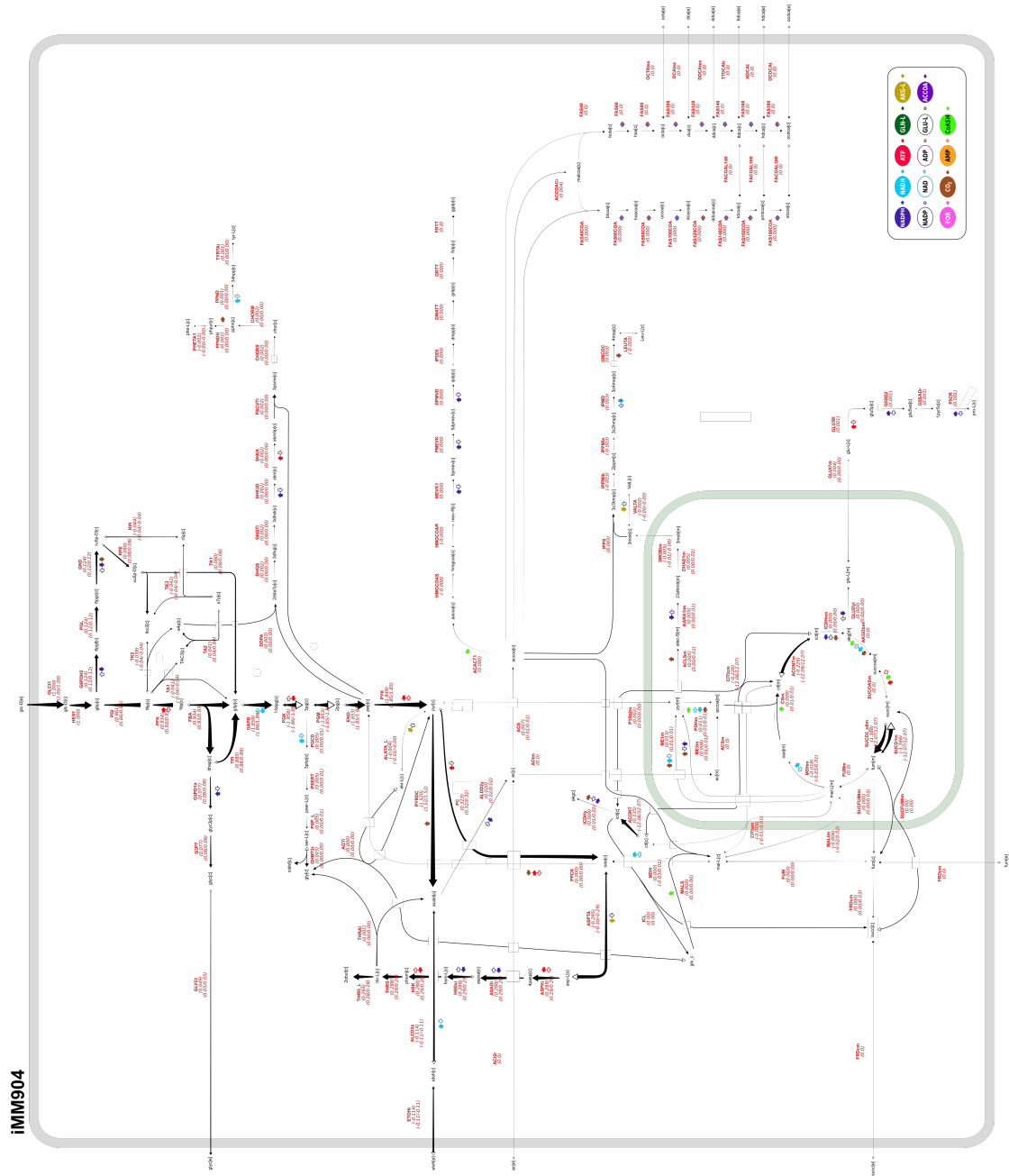Figure 2.12: Whole flux profile for strain/condition U

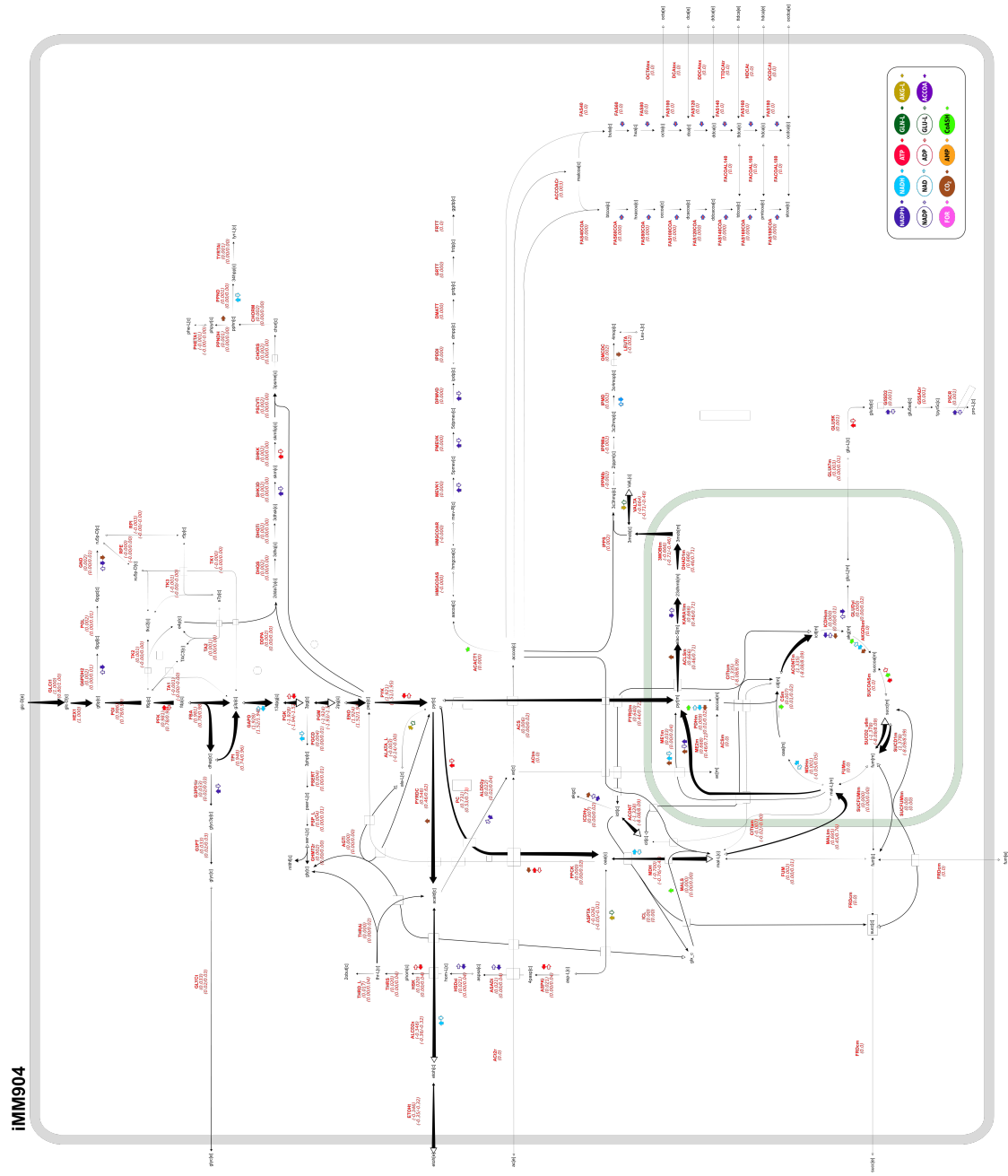Figure 2.13: Whole flux profile for strain/condition S

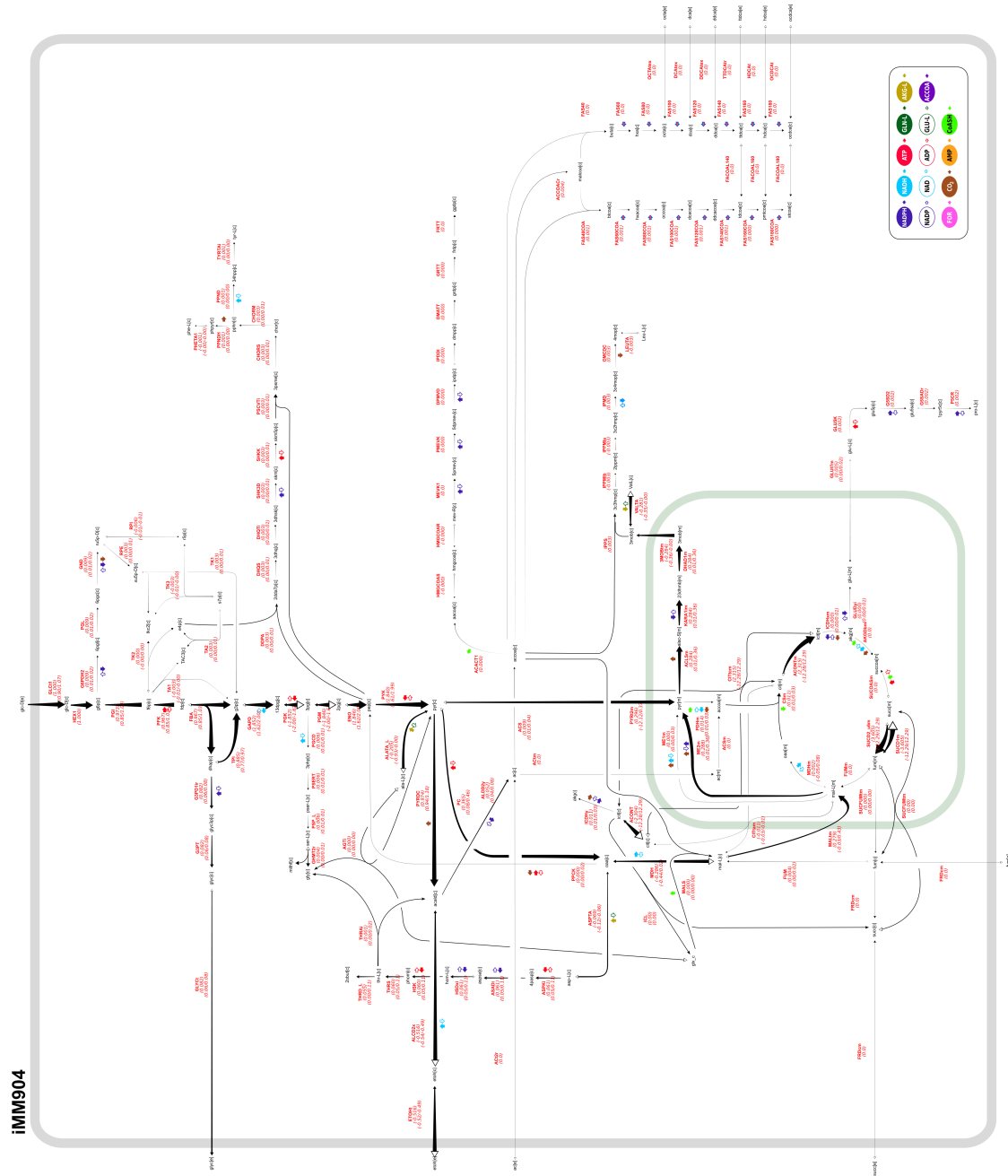Figure 2.14: Whole flux profile for strain/condition UG

Figure 2.15: Whole flux profile for strain/condition SG

Chapter 3

Attempted growth rescue of adh1Δ via heterologous NADH oxidase expression

3.1 Introduction

Carbon flow to cellular biomass and/or some designated target metabolite is preferred over ethanol production, yet deletion of the gene responsible for the majority of ethanol production, *ADH1*, roughly halves growth rate[26]. Since this is partially due to reduction of *S. cerevisiae's* ability to oxidize NADH, it may be possible to correct this growth defect via expression of an enzyme that performs this oxidation. Previous studies have used heterologous expression of NADH oxidases that also convert oxygen to water for cofactor engineering[27–30] though we intended for this to be the first study to do so from the perspective of balancing carbon capture and growth from a fluxomic perspective.

In this work, we attempted to rescue the growth of an *ADH1* null mutant via low-copy plasmid expression of the NADH oxidase (LnoxE) from *L. lactis* under the control of 5 mutated *TEF1* promoters[26,31,32]. Initial shake flask experiments led to expansion of our combinatorial assembly to include the NADH oxidases from bacterial *S. pneumoniae* and fungal *A. capsulatus*[28]. Growth characterization experiments in 96-well and 24-well plate reader experiments led to higher replicate batch experiments. Our hope was to rescue growth and explain the resulting phenotype at a finer level of detail using 2-scale ^{13}C metabolic flux analysis.

3.2 Materials & Methods

3.2.1 Strain & plasmid construction

Strains were constructed using a combination of a loxP/Cre-recombinase method, as described in the *Sip1* chapter, and yeast cloning in a CEN.PK113-7D *ura3*Δ, provided by Bilge

Ozaydin of JBEI. A list of strains generated in this study is presented in Table 3.1. The *ADH1* knockout cassette was PCR amplified using forward primer M-F and reverse primer M-R. Integration of the knockout cassette was verified using forward and reverse primers M-Ver-F and M-Ver-R, respectively. The sequences for these primers can be found in Table 3.3. Strain CMSY2, an *ADH1* null mutant, resulted from looping out the loxP-kanMX-loxP cassette from CMSY1 using Cre-recombinase and subsequent curing of the plasmid. Strains CPU and CMSY2 were transformed with pRS416 to produce the prototrophic base strain U and the *adh1Δ* strain M, used as references in this study.

Table 3.1: List of strains, their parents, genotypes, and references

Strain name	Parent strain	Description	Reference
CPU	CEN.PK113-7D	CEN.PK113-7D *ura3Δ*	From Bilge Ozaydin (JBEI)
CMSY1	CPU	CPU *ADH1::loxP-kanMX-loxP*	This work
CMSY2	CMSY1	CPU *adh1Δ*	This work
U	CPU	CPU [pRS416]	This work
M	CMSY2	CPU *adh1Δ* [pRS416]	This work
L6	CMSY2	CPU *adh1Δ* [pRS416-Tef1p6-LnoxE-Tef1t]	This work
L10	CMSY2	CPU *adh1Δ* [pRS416-Tef1p10-LnoxE-Tef1t]	This work
L11	CMSY2	CPU *adh1Δ* [pRS416-Tef1p11-LnoxE-Tef1t]	This work
L3	CMSY2	CPU *adh1Δ* [pRS416-Tef1p3-LnoxE-Tef1t]	This work
L7	CMSY2	CPU *adh1Δ* [pRS416-Tef1p7-LnoxE-Tef1t]	This work
S6	CMSY2	CPU *adh1Δ* [pRS416-Tef1p6-SnoxE-Tef1t]	This work
S10	CMSY2	CPU *adh1Δ* [pRS416-Tef1p10-SnoxE-Tef1t]	This work
S11	CMSY2	CPU *adh1Δ* [pRS416-Tef1p11-SnoxE-Tef1t]	This work
S3	CMSY2	CPU *adh1Δ* [pRS416-Tef1p3-SnoxE-Tef1t]	This work
S7	CMSY2	CPU *adh1Δ* [pRS416-Tef1p7-SnoxE-Tef1t]	This work
A6	CMSY2	CPU *adh1Δ* [pRS416-Tef1p6-AnoxE-Tef1t]	This work
A10	CMSY2	CPU *adh1Δ* [pRS416-Tef1p10-AnoxE-Tef1t]	This work
A11	CMSY2	CPU *adh1Δ* [pRS416-Tef1p11-AnoxE-Tef1t]	This work
A3	CMSY2	CPU *adh1Δ* [pRS416-Tef1p3-AnoxE-Tef1t]	This work
A7	CMSY2	CPU *adh1Δ* [pRS416-Tef1p7-AnoxE-Tef1t]	This work

15 plasmids were combinatorially assembled in the CMSY2 background via yeast cloning from all promoter-gene combinations of 5 mutated *TEF1* promoters and 3 NADH oxidases and the *TEF1* terminator in a pRS416 plasmid backbone containing a *URA* marker. This combinatorial assembly is depicted in Figure 3.1. Plasmid names take the form pRS416-Tef1p**X**-**Y**noxE-Teft1, where X is the promoter variant number (same as that in the Nevoight paper it originated from[32]) and Y refers to the species the NADH oxidase originated from. The promoter variant numbers or X are, in order of increasing expected expression according to the results of the Nevoight paper, 6, 10, 11, 3, and 7. The NADH oxidase species abbreviations or Y are L, S, and A for *L. lactis*, *S. pneumoniae*, and *A. capsulatus*, respectively. The strain names for those containing these 15 plasmid variants are designated YX, where X and Y still have the same meaning. For example, strain S3 refers to the prototrophic *ADH1* strain containing the plasmid with the *S. pneumoniae* NADH oxidase under *TEF1* promoter variant 3, pRS416-Tef1p3-SnoxE-Tef1t.
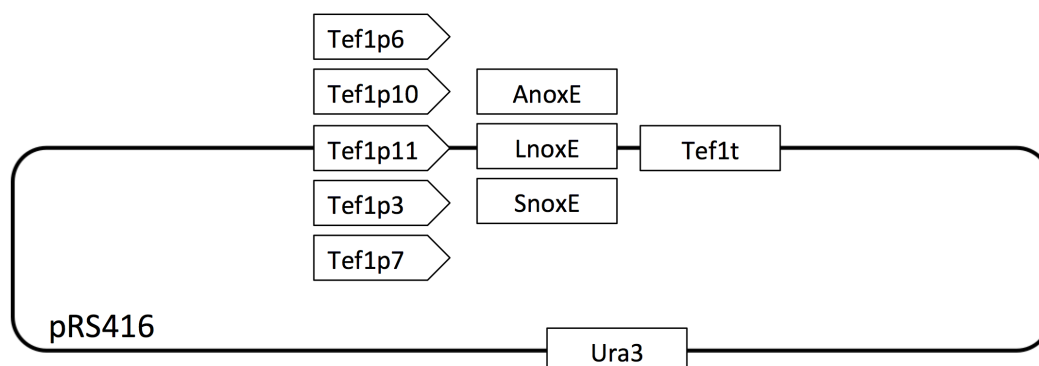
Figure 3.1: Depiction of combinatorial assembly of five mutated *TEF1* promoters, three NADH oxidases, and one *TEF1* terminator in a pRS416 Ura-plasmid backbone

Each plasmid was constructed by transforming CMSY2, using a heat shock method[15], with BamHI-digested pRS416, and overlapping PCR amplicons containing the corresponding *TEF1* promoter, the corresponding NADH oxidase, and Tef1t. All of the *TEF1* promoter amplification's upstream sections were homologous with the upstream side of the cut-site. The upstream and downstream sections of the NADH oxidase piece overlapped with the 3-prime end of the promoter and the 5-prime end of the terminator, respectively. The terminator's 3-prime end overlapped with the downstream side of the cut-site. The forward and reverse primers and templates for each of the 5 potential Tef1p and 3 NADH oxidase amplicons is displayed in Table 3.2. The corresponding sequences of those primers are provided in Table 3.3. Successful transformations were selected on Sc-Ura plates and sequence verified. All stable strain intermediates were stored in 20% glycerol stocks at -80ºC. The plasmids containing the *TEF1* sequences were provided by Sarah Rodriguez of JBEI and the plasmid containing the *L. lactis* gene and S288c gDNA were provided by Mario Ouellet of JBEI. The plasmids containing the remaining NADH oxidase genes were ordered from IDT (Integrated DNA Technologies, Coralville, IA).

3.2.2 Media

The same media compositions as those described in the *Sip1* chapter were used throughout this study.

3.2.3 Plate reader and shake flask experiments

Maximum specific growth rates were compared in 96- and 24-well plate growth experiments. Cells were streaked from -80ºC glycerol stocks onto Min plates, inoculated into 1 mL Sc-Ura deep-96-well plates in 5-6 biological replicates, incubated at 30ºC at 200 rpm until

Table 3.2: PCR amplicons used to construct pRS416-Tef1pX-YnoxE-Teft1 plasmids and their PCR templates and forward/reverse primers

Plasmid piece	Template	Template reference	F-primer	R-primer
Tef1p6	pTef-M6	Sarah Rodriguez	P-6	P-R
Tef1p10	pTef-M10	Sarah Rodriguez	P-10	P-R
Tef1p11	pTef-M11	Sarah Rodriguez	P-11	P-R
Tef1p3	pTef-M3	Sarah Rodriguez	P-3	P-R
Tef1p7	pTef-M7	Sarah Rodriguez	P-7	P-R
LnoxE	pLL	Mario Ouellet	L-F	L-R
SnoxE	pSP	This work	S-F	S-R
AnoxE	pAC	This work	A-F	A-R
Tef1t	S288c gDNA	Mario Ouellet	T-F	T-R

Table 3.3: Sequences of primers used in this study

Primer id	Sequence
M-F	TGCACAATATTTCAAGCTATACCAAGCATACAATCAACTATCTCATATACAagctgaagcttcgtacgc
M-R	TTTTTTTATAACTTATTTAATAATAAAAATCATAAATCATAAGAAATTCGCcataggccactagtggatctg
M-Ver-F	GTTCTCGTTCCCTTTCTTCCTTG
M-Ver-R	ggagtacggataaaatgcttgatgg
P-6	tgatatcgaattcctgcagcccgggatagcttcaaaatgtctctactcctttttttac
P-10	tgatatcgaattcctgcagcccgggatagcttcaaaatgtttctactcctttgttac
P-11	tgatatcgaattcctgcagcccgggatagcttcaaaatgtttctactcctttgttac
P-3	tgatatcgaattcctgcagcccgggatagcctcaaaatgtttctactccttc
P-7	tgatatcgaattcctgcagcccgggatagcttcaaaatgtctctactcctttttttac
P-R	GTTCGTACCGATAACTACGATTTTCATttttctagaaaacttagattagattgctatgctttc
L-F	ATGAAAATCGTAGTTATCGGTACGAAC
L-R	AACTAGAAAAGTCTTATCAATCTCCTTATTTGGCATTCAAAGCTGCAAC
S-F	aatctaatctaagttttctagaaaaatgagtaaaatcgttgtagtcggtg
S-R	AACTAGAAAAGTCTTATCAATCTCCttattttttcagccgtaagggcag
A-F	aatctaatctaagttttctagaaaaatgtatccgacctcaggatgc
A-R	AACTAGAAAAGTCTTATCAATCTCCtcatatcacctcatcccgttcc
T-F	GGAGATTGATAAGACTTTTCTAGTTGC
T-R	cggtggcggccgctctagaactagtACTTTAAAATTGACGATTCCAATACTTCAATTG

exponential phase, diluted 1/100 into a 96-well plate with 100 μL Min per well, and monitored at 30ºC at 600 nm every 15 min using a plate reader (BioTekSynergy 4, USA). Experiments were repeated for a subset of the strains in four different 24-well plates with 750 μL per well using 3-4 replicates per strain. Triplicate shake flask minimal medium experiments were conducted for a smaller subset of strains the same way as described in the *SIP1* chapter.

3.2.4 Maximum specific growth rate determination

Specific growth rates were determined from slopes of manually determined linear regions of the lnOD versus time curves.

3.3 Results

3.3.1 *L. lactis* batch culture growth

We first decided to construct the *L. lactis* NADH oxidase variants and characterize their growth in 40 mL shake flasks relative to strains U and M. The maximum specific growth rates for these 5 variants and the base and *adh1Δ* are displayed in Figure 3.2.
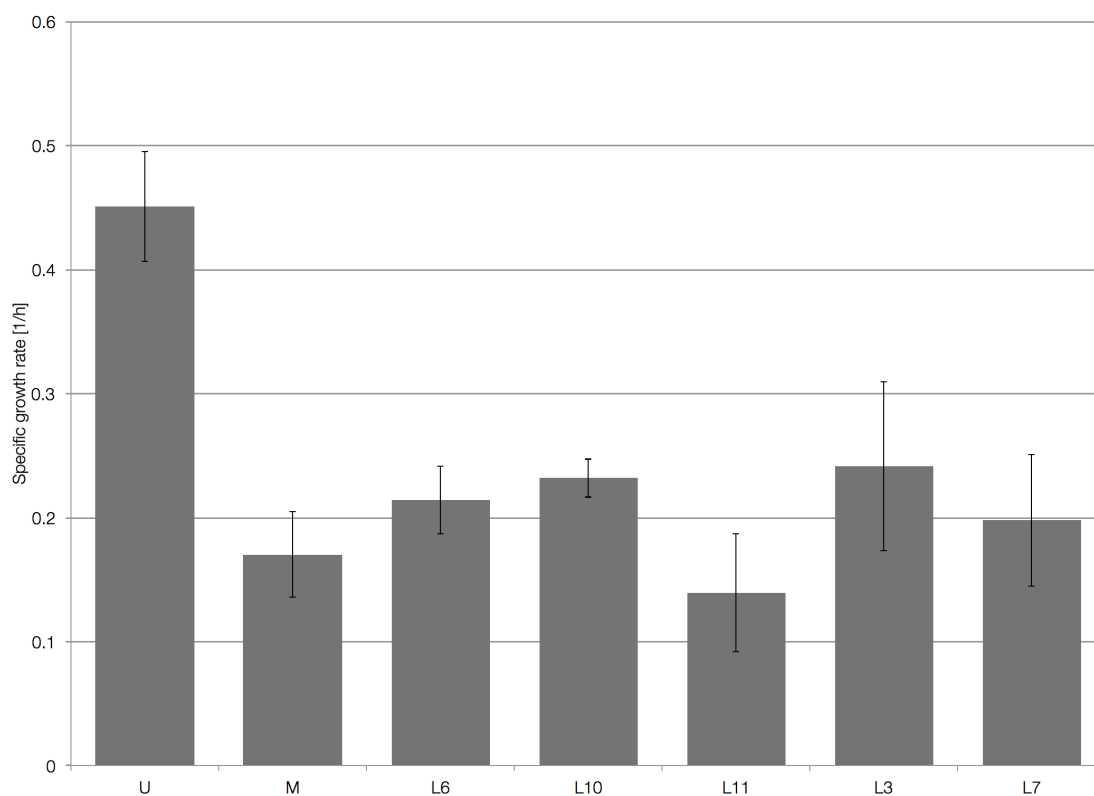
Figure 3.2: Averaged maximum specific growth rates for 40 mL shake flask experiments in triplicate for strains U, M, and 5 Tef1pX-LnoxE strains. Error bars represent one standard deviation from the mean

Deletion of *ADH1* resulted in a 62% decrease in specific growth rate (U versus M) to 38% of the base strain's. Unfortunately, it was difficult to tell if there was a real improvement upon expression of the NADH oxidases under the 5 different TEF1 promoters. It appeared

there may have been an increase of 36% (M versus L10) upon expression of the Tef1p10-LnoxE promoter/gene pair restoring growth to 51% of the base strain.

3.3.2 96-well growth

In an attempt to obtain a stronger effect and more definitive result, we decided to construct an additional 10 Tef1pX-YnoxE combination plasmids using the same 5 mutated *TEF1* promoters and bacterial *S. pneumoniae* and fungal *A. capsulatus* NADH oxidases. The growth of strains U, M, and all 15 NADH oxidase strain variants were compared in a 96-well plate in the same medium as the initial batch cultures. The maximum specific growth rates resulting from this effort, at 5-6 replicates per strain, are displayed in Figure 3.3. Deletion
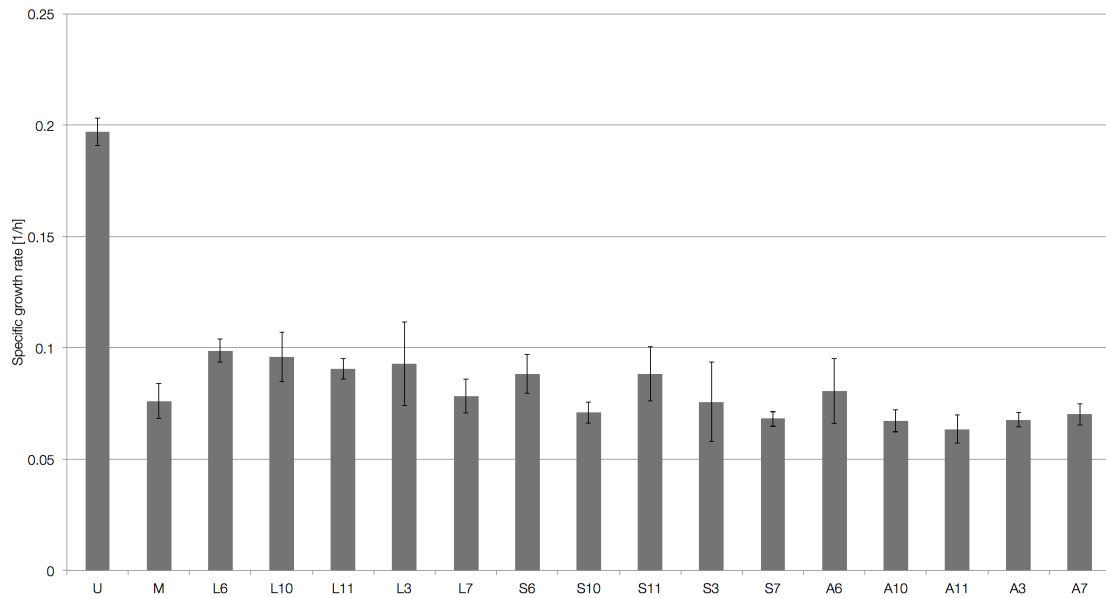

Figure 3.3: Averaged maximum specific growth rates for 96-well plate experiment for strains U, M, and the 15 different Tef1pX-YnoxE strains. Error bars represent one standard deviation from the mean

of *ADH1* resulted in a drop in maximum specific growth rate of 61% to ~37% of the base strain's value. Most of the constructs performed the same with each other and the *adh1Δ* mutant within error. The only real exceptions were from the original *L. lactis* group. The strain containing promoter/gene pair Tef1p6-LnoxE performed the best with about a 30% increase in maximum specific growth rate to 50% of the base strain's value.

3.3.3 24-well growth

We also characterized the growth of all 17 strains of this study in 750 μL 24-well experiments. The results of these experiments, with 3-4 replicates per strain, are displayed in Figure 3.4.

Deletion of *ADH1* resulted in a drop in growth rate of 61% to ~39% of the base strain's
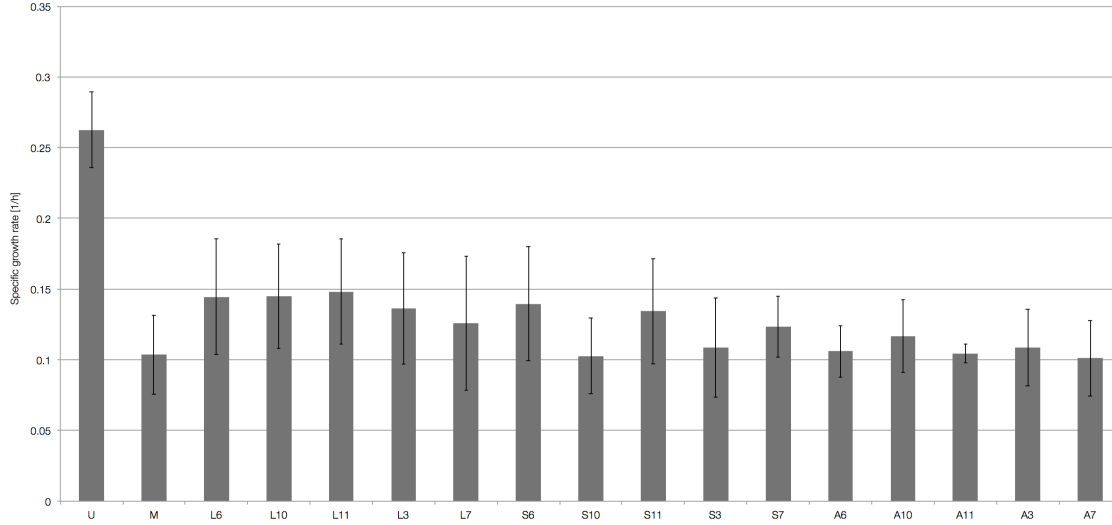

Figure 3.4: Averaged maximum specific growth rates for 24-well plate experiment for strains U, M, and the 15 different Tef1pX-YnoxE strains. Error bars represent one standard deviation from the mean

value. The values for the 15 Tef1p-noxE variants were indistinguishable from each other and the *adh1Δ* mutant's.

3.3.4 Flask growth

Due to the difficulty in resolving the 24-well plate experiment data and the similarity of the batch and 96-well experiments for the *L. lactis* strains, we decided to repeat the 40 mL shake flask experiments in quadruplicate for strain U and using 10 replicates for strains M and L6, to better distinguish their growth rates from each other and to use the data for 2S-^{13}C MFA if we decided to continue. The resulting specific growth rates are displayed in Figure 3.5. Deletion of *ADH1* resulted in a drop in maximum specific growth rate of about 53% to 45% of the base strain's value. There appeared to be no difference between the *ADH1* knockout mutant and that expressing the Tef1p6-LnoxE plasmid.

3.4 Discussion

We could not definitively show that expression of an NADH oxidase under a mutated *TEF1* promoter could rescue growth of an *ADH1* knockout mutant. It's possible that there was some rescue for two *L. lactis* variants when grown in a 96-well plate. Unfortunately, the additional *S. pneumoniae* and *A. capsulatus* constructs appeared to be indistinguishable from the *adh1Δ* mutant in any of the experiments they were involved in. The higher replicate
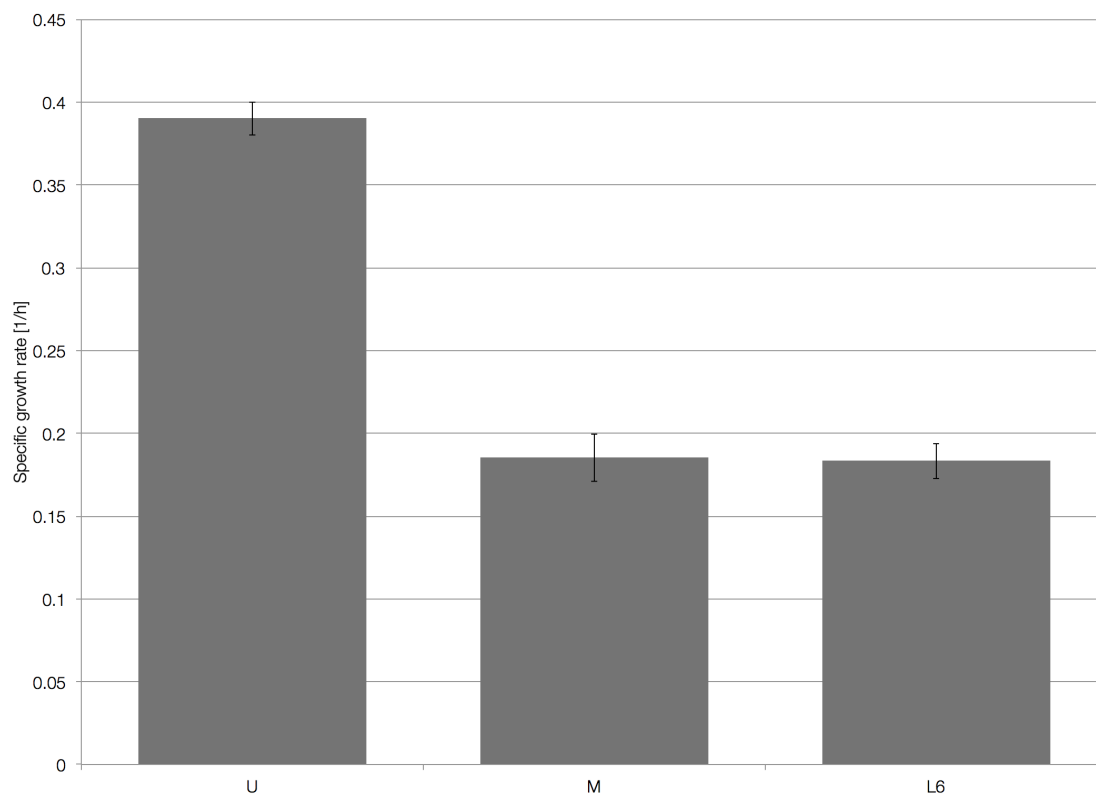
Figure 3.5: Averaged maximum specific growth rates for 2nd round of batch experiments with 4 replicates for strain U and 10 replicates for strains M and L6. Error bars represent one standard deviation from the mean

flask experiments with the best-performing promoter/gene pair had better resolution yet indicated no improvement in specific growth rate. This apparent difference between the 96-well plate and high-replicate flask experiments might have something to do with aeration, since the NADH oxidase converts oxygen to water while oxidizing NADH.

This study highlights the importance of experimental replication and how results can vary at different scales. While the initial three-replicate 40 mL shake flask experiment and that in the 96-well plate appeared to indicate a possible growth rescue, the 10 replicate shake flask experiment for the presumed best promoter/gene pair definitively showed no improvement in conditions most relevant to obtaining flux profiles. We decided against pursuing flux experiments due to these results.

It's possible that a stronger expression of an NADH oxidase could still rescue growth. The pRS416 plasmid backbone is a low-copy plasmid and none of the NADH oxidases were codon-optimized. Perhaps expression of a codon-optimized version on low-copy and/or high-copy plasmid or integrated into a highly expressed region of the chromosome could allow for a growth rescue to occur.

3.5   Acknowledgements

Chapter 4

SDP-relaxation of ^{13}C metabolic flux analysis with rank sparcity

4.1 Introduction

During the course of our fluxomics work we decided to explore some convex approximations of the mathematical structure of ^{13}C metabolic flux analysis to see if we could achieve the same solution as the original nonconvex form, determine lower objective function bounds, and, possibly, speed up our code. In this work, the computational side of ^{13}C MFA was reformulated as a rank-relaxed semidefinite program (SDP). Python code was written to obtain all relevant parameters from a user-provided list of metabolic reactions and corresponding carbon transitions, designated input metabolites and those for which ^{13}C labeling probabilities have been measured, and values for all input and output fluxes. These parameters were then used by the code to solve a six-reaction toy model from the literature using a convex optimization solver, CVX, in Matlab[3,33] (Antoniewicz 2007, Grant 2011). The global optimum was verified by gridding over the one remaining degree of freedom obtained using the algorithm proposed in the Antoniewicz paper and the solution was used, along with constraint violation and numerical decision matrix rank to ascertain the effectiveness of the method. Rank-sparsity was encouraged using two different terms added to the objective function: 1) a scaled trace of the decision matrix and 2) a trace inner product between the decision matrix and an iteratively updated matrix formulated from the eigenvalue decomposition of the previous decision matrix solution. All methods were compared with the solution from the paper and with each other.

4.2 Materials & Methods

4.2.1 Equivalent quadratic reformulation

EMUs or elementary metabolite units are subsets of carbon atoms for which ^{13}C labeling probabilities are being kept track of. For instance EMU D_{23} corresponds to the second and

third carbon of metabolite D. Each EMU has an associated mass distribution vector (MDV), in which the first element describes the percentage of the given EMU that is labeled in zero carbons, the second element describes the percentage that is labeled in any one carbon, etc. The founding EMU paper focused on fitting simulated EMU MDV values with their experimentally determined counterparts by exploiting the structure of EMU steady-state mass balance and stoichiometric equations so that a subset of reaction rates determined the values of all other variables[3]. Hence, these reaction rates were the decision variables while the EMU MDVs served as intermediates. The generalized problem, as discussed in the 2-scale ^{13}C MFA section in the Introduction, is a least-squares minimization subject to third-order polynomial mass isotopomer balance and other constraints. Introduction of slack variables results in second-order constraints as shown in Equation 4.1.

$$
\min_{\overline{x}} \left\| \overline{C}\overline{x} - \overline{y} \right\|_2^2 : \overline{S}\overline{x} = 0
$$

$$
\begin{aligned}
& A_e \overline{x} = 1, \forall e \in E_i \cup E^C, I \in I^N \\
& \overline{x} \geq 0 \\
& D\overline{x} \leq 1 \\
& \begin{bmatrix} \overline{x} \\ 1 \end{bmatrix}^T Q_{em} \begin{bmatrix} \overline{x} \\ 1 \end{bmatrix} = 0, \forall m \in \{1...n_e\}, \forall e \in E_i \cup E^C, I \in I^N
\end{aligned}
$$

$$(4.1)$$

The decision vector, $\overline{x}$, consists of intracellular reaction rates and MDVs of all EMUs encountered by the EMU decomposition method (described in the introduction) and those generated when two EMUs combine[3]. The entries of these so-called combined EMU MDVs are the slack variables mentioned above and are defined by the elements of the Cauchy products of the EMUs that combine to generate them. A Euclidean norm serves as the objective function and is minimized as a measure of how close the experimentally determined labeling distribution values, $\overline{y}$, are to the corresponding simulated values in the decision vector selected by matrix $\overline{C}$. The bottom constraint represents the quadratic EMU balances and slack variable definitions. The subscript of the quadratic matrix denotes the mth MDV entry of EMU, e, for all metabolite carbon groups in the set of balanced EMUs, E_i, for all non-input metabolites in the set I^N and for all combined EMUs in the set E^C. The mass indices, m, range from 0 to the number of carbons, n_e, present in EMU, e. The affine inequality above this set of quadratic constraints describes the fact that all labeling probability values are less than one. Hence, D, is a selection matrix that picks out all labeling values. The inequality above that asserts that all labeling and flux values are non-negative. The affine equality above that represents the fact that the labeling probability distributions for each regular and combined EMU sums to one. The first equation represents flux conservation at each non-input metabolite. That is, the sum of the fluxes that generate a given metabolite is equal to the sum of the fluxes that consume it. If s_{ij} is equal to the stoichiometric coefficient, which is positive if metabolite i is created by the flux in the jth position of the decision vector and negative if consumed by it and whose absolute value is equal to the number of molecules involved in a single reaction, the stoichiometric matrix, $\overline{S}$, is given by the rules laid out in

Equation 4.2.

$$\overline{S}_{ij} = \begin{cases} 0, & \text{reaction j isn't connect to metabolite i} \\ s_{ij}, & \text{reaction j is connected to metabolite i} \end{cases} \tag{4.2}$$

4.2.2 Equivalent non-convex trace reformulation

An equivalent problem can be written by exchanging the quadratic constraints with trace equalities involving a rank-one decision matrix that is the outer product of the decision variable, as shown in Equation 4.3.

$$\min_{\overline{x},X} \left\| \overline{C}\overline{x} - \overline{y} \right\|_2^2 : \overline{S}\overline{x} = 0$$

$$A_e \overline{x} = 1, \forall e \in E_i \cup E^C, I \in I^N$$
$$\overline{x} \geq 0$$
$$D\overline{x} \leq 1$$
$$\begin{bmatrix} X & \overline{x} \\ \overline{x}^T & 1 \end{bmatrix} \succeq 0 \tag{4.3}$$
$$Tr\left(Q_{em} \begin{bmatrix} X & \overline{x} \\ \overline{x}^T & 1 \end{bmatrix} \right) = 0, \forall m \in M_e, \forall e \in E_i \cup E^C, i \in I^N$$
$$rank(X) = 1$$

4.2.3 Convex positive semidefinite reformulation

Since the trace constraint matrices are positive semidefinite (PSD), the objective and all but the rank constraint are convex. Dropping the non-convex rank requirement results in the convex SDP given in Equation 4.4.

$$\min_{\overline{x},X} \left\| \overline{C}\overline{x} - \overline{y} \right\|_2^2 : \overline{S}\overline{x} = 0$$

$$A_e \overline{x} = 1, \forall e \in E_i \cup E^C, I \in I^N$$
$$\overline{x} \geq 0$$
$$D\overline{x} \leq 1 \tag{4.4}$$
$$\begin{bmatrix} X & \overline{x} \\ \overline{x}^T & 1 \end{bmatrix} \succeq 0$$
$$Tr\left(Q_{em} \begin{bmatrix} X & \overline{x} \\ \overline{x}^T & 1 \end{bmatrix} \right) = 0, \forall m \in M_e, \forall e \in E_i \cup E^C, i \in I^N$$

It was noticed during the course of this work that there were negative entries in the decision matrix when the relaxed SDP was solved. This should not be the case since the decision matrix should be the outer product of the non-negative decision vector. The addition of an element-wise non-negative sign constraint on the decision matrix results in a problem of the form given in Equation 4.5 that was used for the SDP results.

$$\min_{\overline{x},X} \left\| \overline{C}\overline{x} - \overline{y} \right\|_2^2 : \overline{S}\overline{x} = 0$$

$$A_e\overline{x} = 1, \forall e \in E_i \cup E^C, I \in I^N$$
$$\overline{x} \geq 0$$
$$X \geq 0$$
$$D\overline{x} \leq 1 \qquad (4.5)$$
$$\begin{bmatrix} X & \overline{x} \\ \overline{x}^T & 1 \end{bmatrix} \succeq 0$$
$$Tr\left(Q_{em} \begin{bmatrix} X & \overline{x} \\ \overline{x}^T & 1 \end{bmatrix} \right) = 0, \forall m \in M_e, \forall e \in E_i \cup E^C, i \in I^N$$

4.2.4   Rank sparsity encouragement via weighted trace term

The first method used to encourage rank sparsity in the decision matrix was to add its trace weighted by a constant, λ, to the objective function resulting in the problem in Equation 4.6. The trace of a matrix is equal to the sum of its eigenvalues. Since these eigenvalues are positive (since X is PSD) the solver is encouraged to decrease this term by encouraging as many of the eigenvalues to be zero as possible. Since this term is linear, with regard to the decision matrix, it is a convex way to encourage rank sparsity.

$$\min_{\overline{x},X} \left\| \overline{C}\overline{x} - \overline{y} \right\|_2^2 + \lambda Tr\left(X\right) : \overline{S}\overline{x} = 0$$

$$A_e\overline{x} = 1, \forall e \in E_i \cup E^C, I \in I^N$$
$$\overline{x} \geq 0$$
$$X \geq 0$$
$$D\overline{x} \leq 1 \qquad (4.6)$$
$$\begin{bmatrix} X & \overline{x} \\ \overline{x}^T & 1 \end{bmatrix} \succeq 0$$
$$Tr\left(Q_{em} \begin{bmatrix} X & \overline{x} \\ \overline{x}^T & 1 \end{bmatrix} \right) = 0, \forall m \in M_e, \forall e \in E_i \cup E^C, i \in I^N$$

4.2.5   Rank sparsity encouragement via eigenvalue inflation

The second method used to encourage rank sparsity was to artificially decrease the trace term added to the objective function by multiplying the decision matrix by a matrix that directly divides its eigenvalues by those of the previous solution plus a small constant. This causes the solver to underestimate the rank of X thereby encouraging rank sparsity. The iterative scheme is given in Equation 4.7. The weighted trace term is replaced with the matrix inner product between the decision matrix and the current iterate of the inflation matrix, Y_i. Note that, for compactness, V is the set of constraints present in Equation 4.6.

Starting with the identity matrix, the problem is solved and the diagonal matrix from the eigenvalue decomposition of the decision matrix is used to construct the next Y_{i+1} matrix iterate by inverting the matrix created from adding the identity matrix scaled by a small constant, ϵ, to the eigenvalue decomposition diagonal matrix and multiplying this by the orthogonal matrix and its transpose on the left and right, respectively. This process is repeated until the problem's decision matrix solution converges.

$$Y_0 = I$$
$$X_i = arg \min_{(\overline{x}, X) \in V} \left\| \overline{C}\overline{x} - \overline{y} \right\|_2^2 + Tr(Y_i X)$$
$$U_i \Lambda U_i^T = eig(X_i)$$
$$Y_{i+1} \longleftarrow U_i \left(\Lambda_i + \epsilon I \right)^{-1} U_i^T$$

(4.7)

4.2.6 Python/Matlab code

Python code was written to derive all parameters in Equation 4.4 from a user-input similar to that in Table 4.1 below and solve the resulting convex problem using the Matlab version of a convex optimization solver named CVX[33].

4.2.7 Toy model and its EMU decomposition

The six-reaction toy model from the introduction, obtained from the founding EMU paper, is presented in Table 4.1 for convenience. It consists of four irreversible reactions and one reversible pair. Carbon transitions are provided to generate the EMU reaction network dis-

Table 4.1: Set of reactions and corresponding carbon transitions for a six-reaction toy model

Reaction number	Reaction stoichiometry	Atom transitions
1	A → B	abc → abc
2 and 3	B ↔ D	abc ↔ abc
4	B → C + E	abc → bc + a
5	B + C → D + E + E	abc + de → bcd + a + e
6	D → F	abc → abc

played in Figure 4.1.

4.2.8 Toy Model Input

The toy model carbon transitions, three-carbon input metabolite labeling of 100% in the second carbon, and flux values of 100 and 80 for reactions 1 and 6 were used as inputs to the code.
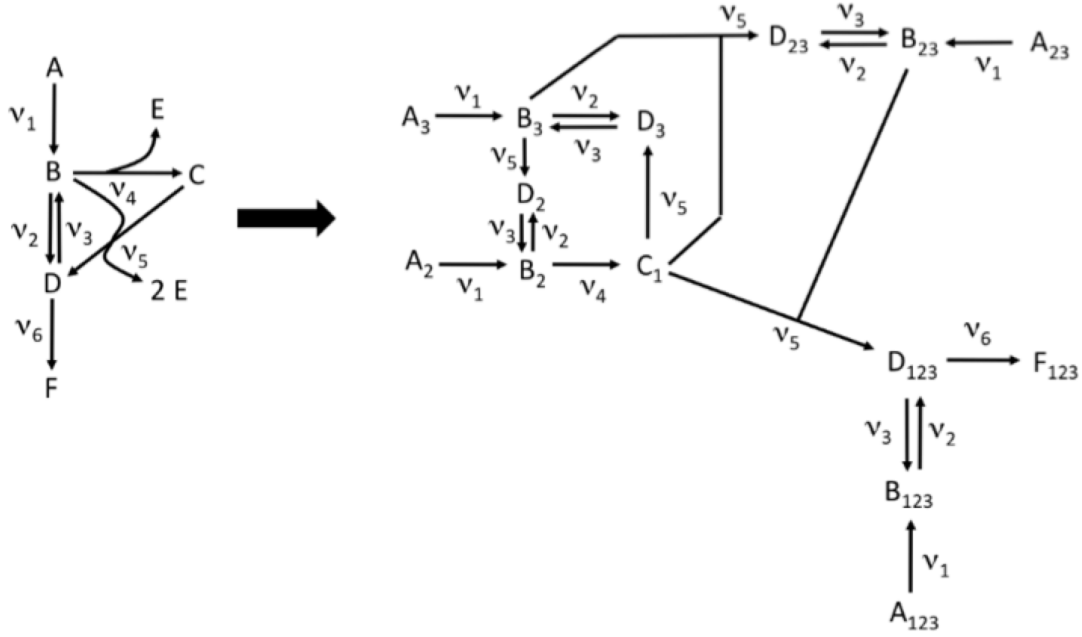
Figure 4.1: EMU decomposition of a six-reaction toy model[3]. Metabolite (nodes) subscripts denote carbon atoms for which ^{13}C labeling probabilities are being kept track of.

4.3   Results

4.3.1   Toy model global minimum

The stoichiometric matrix corresponding to the six-reaction toy model is presented in Equation 4.8.

$$\begin{bmatrix} -1 & 1 & -1 & -1 \\ 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \nu_2 \\ \nu_{2-rev} \\ \nu_4 \\ \nu_5 \end{bmatrix} = 0 \tag{4.8}$$

The reduced row echelon form of the stoichiometric equation revealed that all other fluxes are uniquely determined by the value of ν_{2-rev}. Note ν_{2-rev} is the same as ν_3 above yet this name is used here since it is automatically generated by the Python code to cut down on user input/error. The EMU balances, arranged as discussed in the Antoniewicz paper[3], are displayed in Equation 4.9.

$$\begin{bmatrix} -\nu_4 I_2 & \nu_4 I_2 & 0 & 0 & 0 \\ 0 & -(\nu_1+\nu_3)I_2 & \nu_3 I_2 & 0 & 0 \\ 0 & \nu_2 & -(\nu_2+\nu_5)I_2 & \nu_5 I_2 & 0 \\ 0 & 0 & 0 & -(\nu_1+\nu_3)I_2 & \nu_3 I_2 \\ 0 & 0 & 0 & \nu_2 I_2 & -(\nu_2+\nu_5)I_2 \end{bmatrix} \begin{bmatrix} C_1 \\ B_2 \\ D_2 \\ B_3 \\ D_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -\nu_1 I_2 & 0 \\ 0 & 0 \\ 0 & -\nu_1 I_2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A_2 \\ A_3 \end{bmatrix}$$

$$\begin{bmatrix} -(\nu_2+\nu_5)I_3 & \nu_2 I_3 \\ \nu_3 I_3 & -(\nu_1+\nu_3)I_3 \end{bmatrix} \begin{bmatrix} D_{23} \\ B_{23} \end{bmatrix} = \begin{bmatrix} -\nu_5 I_3 & 0 \\ 0 & -\nu_1 I_3 \end{bmatrix} \begin{bmatrix} (B_3 \times C_1) \\ A_{23} \end{bmatrix} \quad (4.9)$$

$$\begin{bmatrix} \nu_6 I_4 & -\nu_6 I_4 & 0 \\ 0 & -(\nu_2+\nu_5)I_4 & \nu_2 I_4 \\ 0 & \nu_3 I_4 & -(\nu_1+\nu_3)I_4 \end{bmatrix} \begin{bmatrix} F_{123} \\ D_{123} \\ B_{123} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -\nu_5 I_4 & 0 \\ 0 & -\nu_1 I_4 \end{bmatrix} \begin{bmatrix} (B_{23} \times C_1) \\ A_{123} \end{bmatrix}$$

The capital letters represent EMU MDVs. For instance, B_{23} is the probability distribution for the carbon group consisting of the 2nd and 3rd carbons of B. The 'x' on the right-hand-side of the two- and three-carbon EMU balances represents the Cauchy product between the labeling distributions of the two vectors. Note how everything on the right-hand-side of the system at the top is known allowing for the solution of all one-carbon EMU vectors. Also, notice how once the one-carbon balances are solved, everything on the right-hand-side of the two-carbon EMU balances is known. The system can be solved in a cascaded manner to find the labeling of all unknown quantities if a flux profile is specified. Hence, the labeling generation code plots the objective function value as a function of v_{2-rev}, since it determines all reaction rates and labeling values. This graph is displayed in Figure 4.2. The global optimum is found at a v_{2-rev} value of 50.0 with an objective function value of 5.361e-5. This is the same flux value that was specified in the founding EMU paper[3]. The labeling distribution from this solution is used to evaluate the solutions obtained below.

4.3.2 Solution of the SDP relaxation

Solving the relaxed SDP of Equation 4.5 results in the comparison given in Figure 4.3. The corresponding fluxes and those from the paper are displayed in Table 4.2.

Table 4.2: Reaction rates from the founding EMU paper and that obtained via solution with CVX

Reaction flux	Paper value [rxn/hr/gdw]	CVX solution [rxn/hr/gdw]
v_2	110.000	60.448
v_{2-rev}	50.000	0.448
v_4	20.000	20.000
v_5	20.000	20.000

The objective function was 5.025e-5, lower than the true global optimum of 5.361e-5 as expected. The net fluxes are the same as the solution from the Antoniewicz paper, though the individual reversible reactions differ. The comparison above indicates that the CVX solution and that from the paper differ greatly. The constraints were all rearranged so that they equal zero and evaluated using the decision vector to judge their violation. Error versus the constraint index is plotted in Figure 4.4. Constraints 0 through 2 are the stoichiometric constraints. Those with indices 3 until 12 represent how the labeling distributions of each EMU should add up to one. 13 to 40 are the EMU balances and 40 to 50 are the slack variable definitions. Violations arise mostly from the EMU balances. A plot of the first nine singular values of the decision matrix is displayed in Figure 4.5 and indicates that it is not rank one. This should not be the case since the decision matrix should be the outer product of the decision vector. Unfortunately, this corresponds to the rank one requirement that was relaxed making us unable to communicate a vital piece of information to the solver.
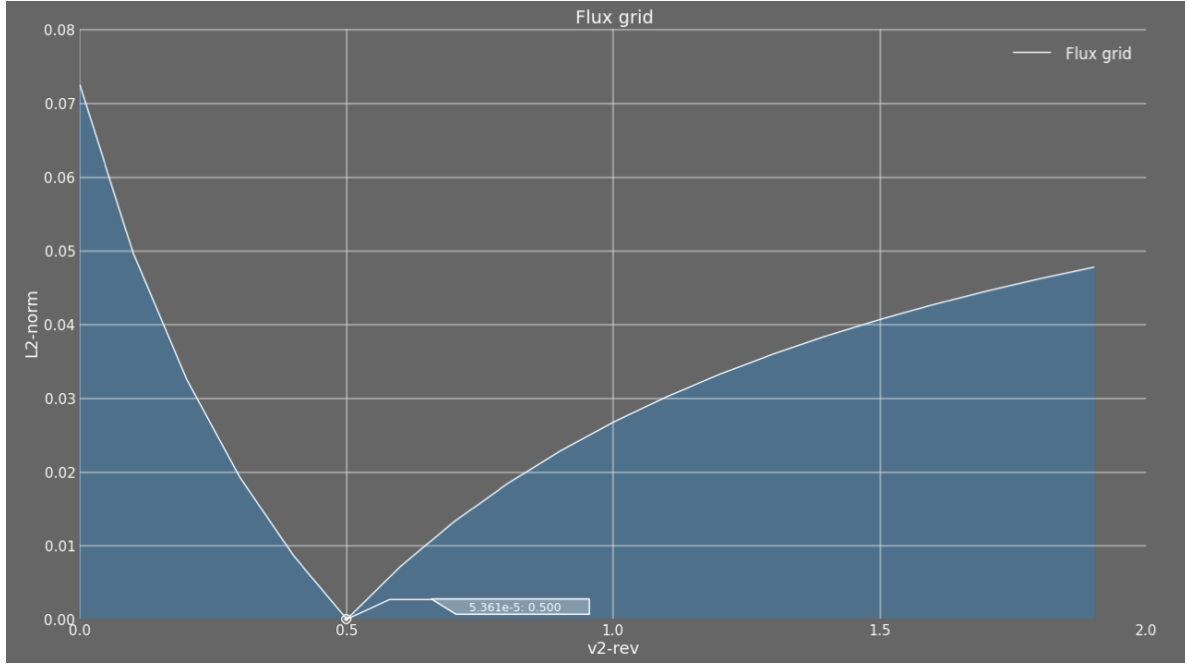
Figure 4.2: Toy model objective function gridded over v_{2-rev}, the single degree-of-freedom for the system

4.3.3 Rank sparsity encouragement via weighted trace term

The weighted trace problem was solved over a range of λ values and the resulting l2-norm objective and the ratio of the second and first singular values, a measure of the rank, were plotted as functions of λ and are displayed in Figure 4.6. The optimal λ-value was determined to be 0.070145, where the two curves crossed. The solution found using this framework in CVX is displayed in Figure 4.7 and corresponds to an objective function of 1.439e-4, which is above that for the paper solution.

The labeling values have become much closer to those from the paper solution. The fluxes corresponding to the solution are presented in Table 4.3. The net fluxes are the same as in the paper and the value of v_2 and v_{2-rev} have integer values.

Table 4.3: Reaction rates from the founding EMU paper and that obtained via solution with CVX with trace objective

Reaction flux	Paper value [rxn/hr/gdw]	CVX w/ trace [rxn/hr/gdw]
v_2	110.000	60.000
v_{2-rev}	50.000	0.000
v_4	20.000	20.000
v_5	20.000	20.000

The constraint violation plot for this λ-value is shown in Figure 4.8. The scale of the error is much lower in this instance since the previous errors were as high as 70 and these reach no higher than a magnitude of 0.2. Also, most of the constraint violation appears to be concentrated in the slack variable definitions, the least important constraints. The first nine singular values of this decision matrix are displayed in Figure 4.9 and indicate that the matrix is nearly rank one at a σ-ratio of 1.438e-4.
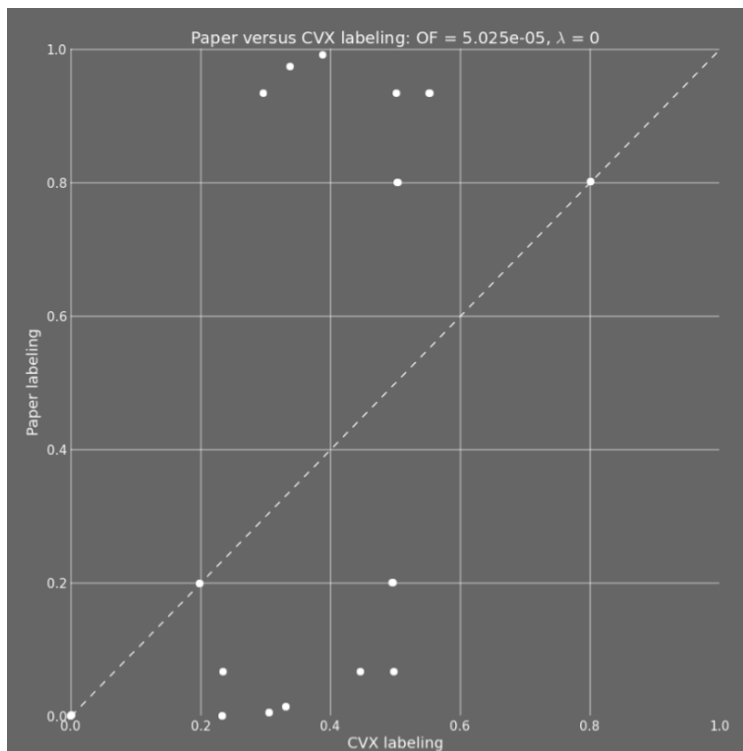
Figure 4.3: Paper solution versus that obtained using CVX

4.3.4 Rank sparsity encouragement via eigenvalue inflation

The rank sparsity encouragement algorithm in Equation 4.7 was implemented in Matlab and run for 3 iterations. The paper solution versus that obtained using the iterative scheme above is displayed in Figure 4.10. The solution is a little less close than that obtained using the scaled decision matrix trace in the previous section. The objective function for this method was also a couple orders of magnitude higher at 0.0716 and, therefore, farther from the global optimum. However, the constraint violations are roughly two orders in magnitude lower as shown in Figure 4.11. Also, the rank σ-ratio is two orders of magnitude lower as well, indicating that the decision matrix is closer to being rank one. Despite this method resulting in a solution further from the global optimum it satisfies the constraints better, is more automatable, and is less time-consuming since an optimal constant doesn't need to be chosen.

4.4 Discussion

This is the first reported instance of a quadratic reformulation of ^{13}C-metabolic flux analysis. It's also the first to report a SDP relaxation and to attempt to encourage rank sparsity using either a weighted trace term or an iterative eigenvalue inflation scheme. As expected, the SDP relaxation resulted in a lower objective than the actual solution. The parameter optimization method resulted in an optimal solution and argument closer to that found in the paper. However, the iterative matrix method proved more automatable, faster, and possessed better mass balance constraint satisfaction. It should be noted that the space of solutions of the six-reaction toy model is limited and further work needs to be completed before it can become clear whether these kinds of convex relaxations can improve performance. Extensions of this work would include using larger networks, comparing both the solutions and objectives to the convex versions of these larger networks, and exploring the use of these solutions as initial points for randomized methods. Other extensions include exploring further approximations such as Reformulation-Linearization Technique (RLT) and/or successive linear restriction of the EMU balance equation quadratic matrices.
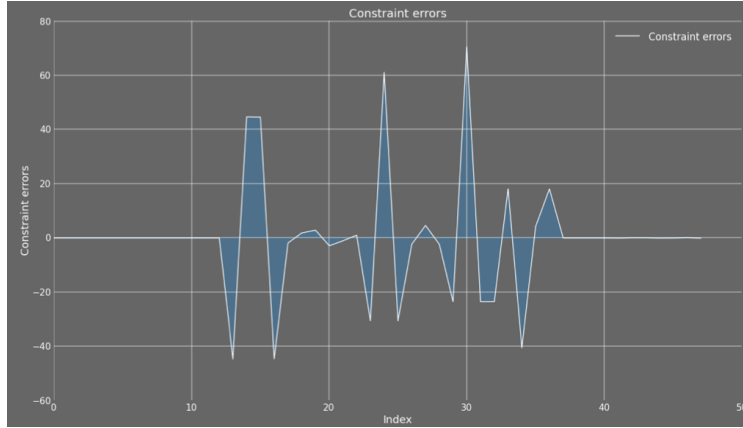
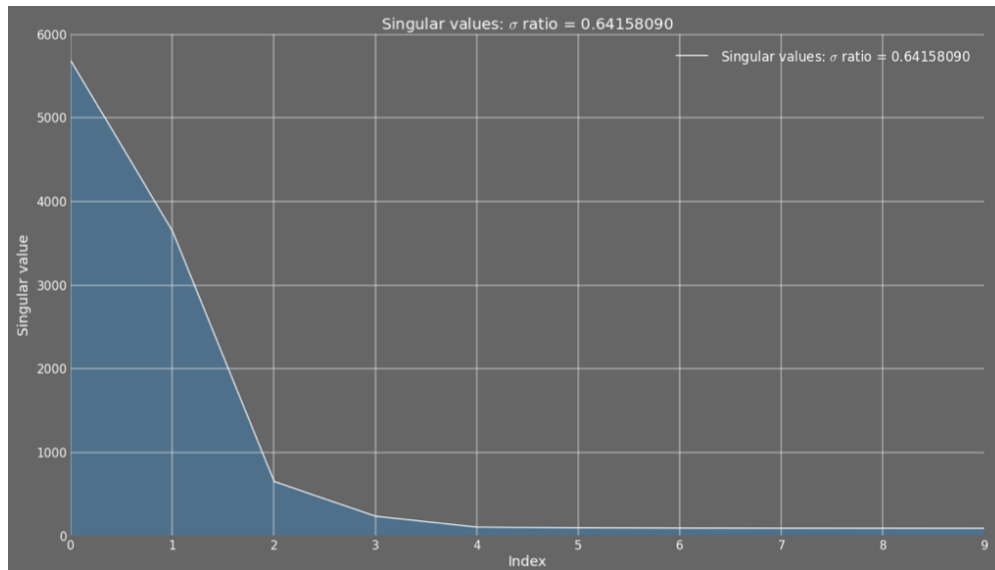Figure 4.4: Constraint violation error plotted as function of constraint index

Figure 4.5: Plot of the first 9 singular values

4.5 Acknowledgements

We would like to thank Professor Laurent El Ghaoui of UC Berkeley for guiding us through this project, teaching us about convex optimization, and his general advice.
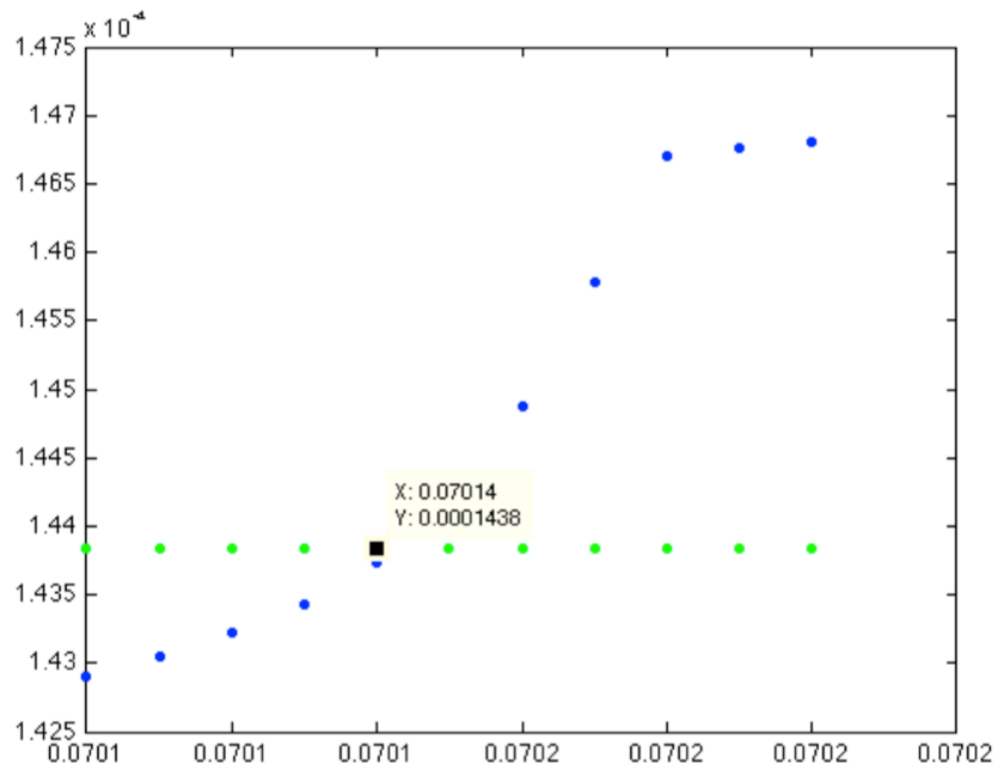
Figure 4.6: Objective function (blue) and ratio of 1st and 2nd singular values (green) plotted as a function of λ
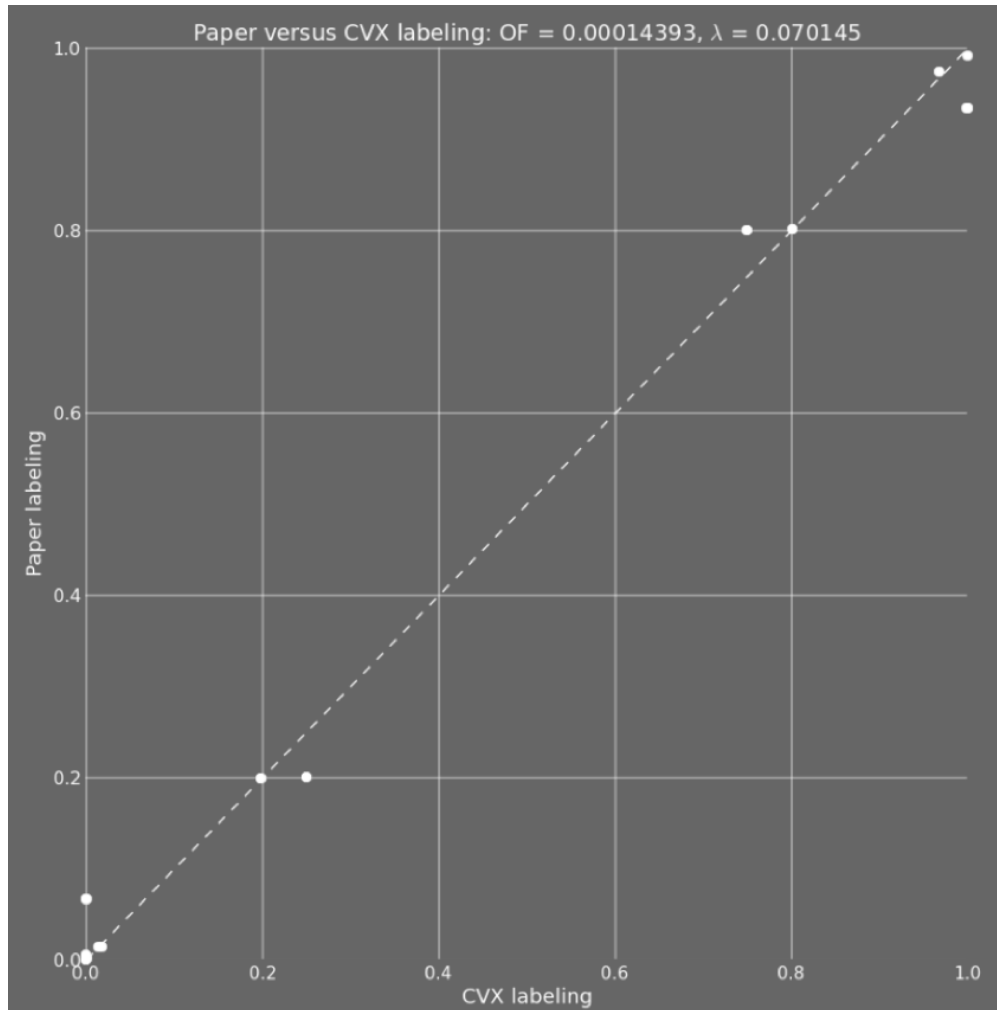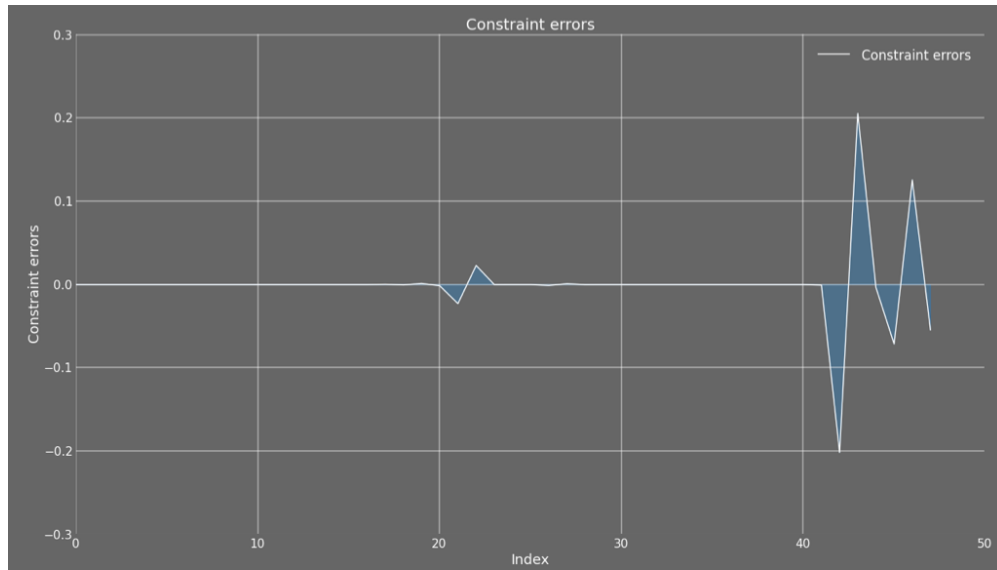
Figure 4.7: Paper solution versus weighted trace solution

Figure 4.8: Constraint violation plot after the addition of the weighted decision matrix trace term to the objective function with an optimal λ of 0.070145
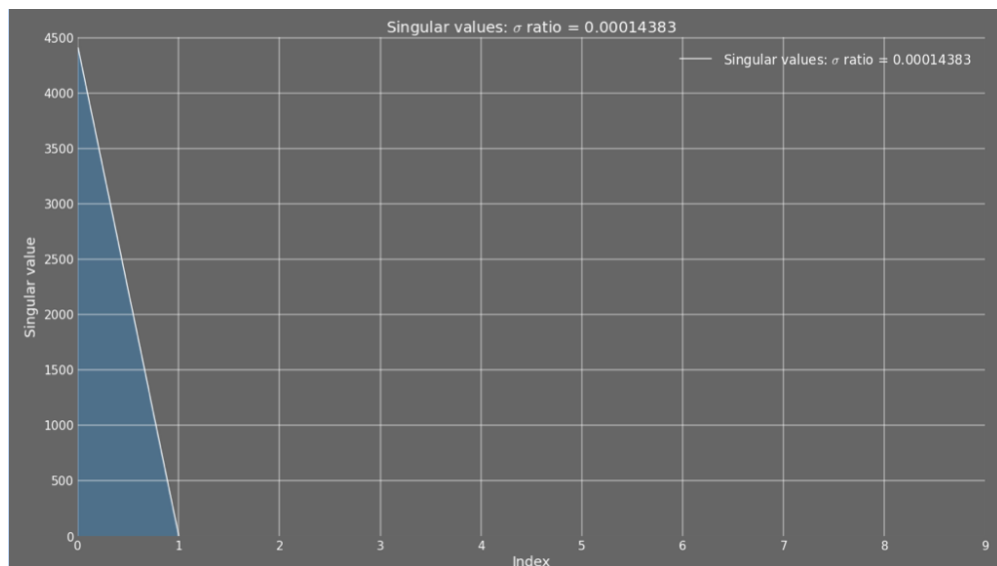

Figure 4.9: Plot of first nine singular values after the addition of the weighted decision matrix trace term to the objective function with an optimal λ of 0.070145
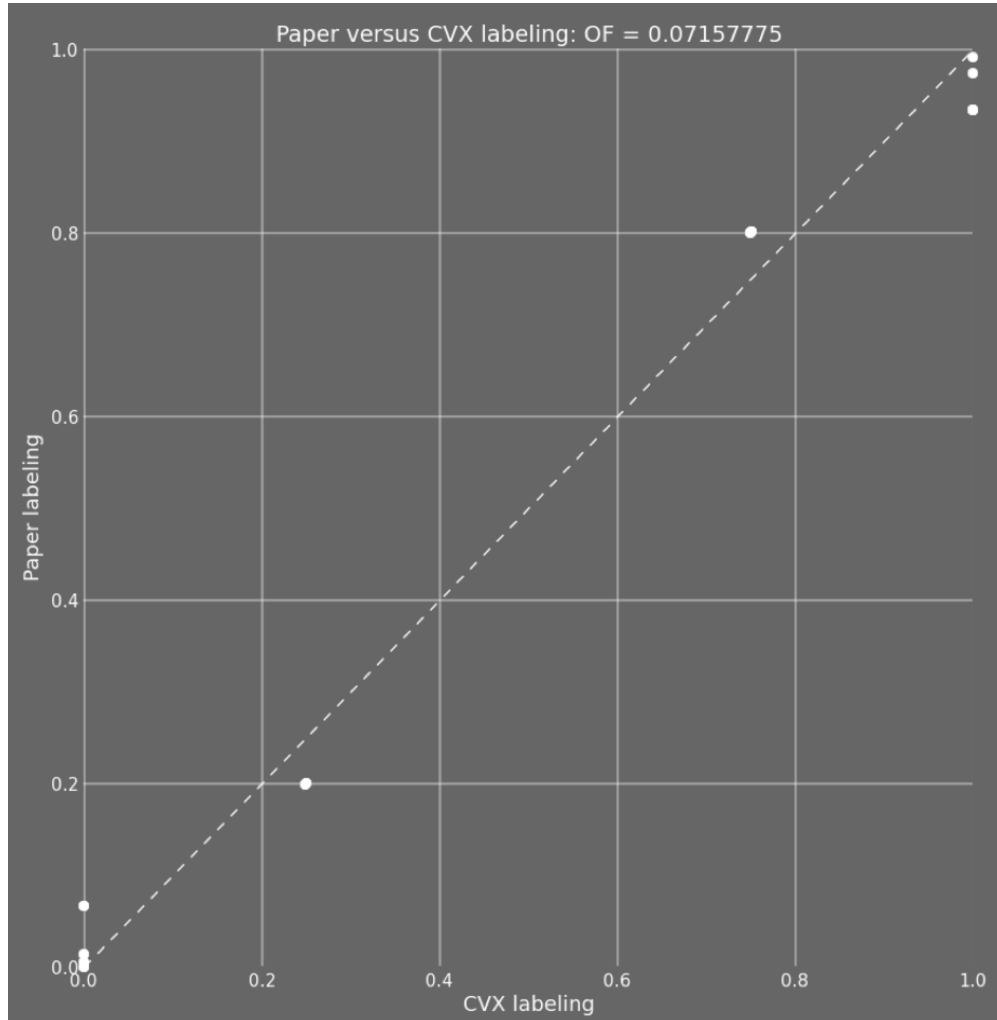
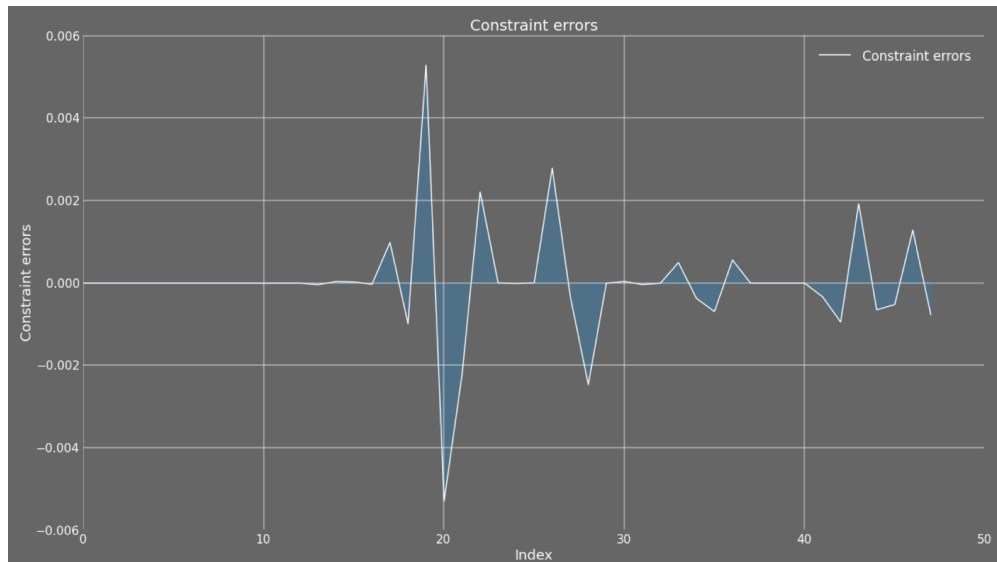Figure 4.10: Paper solution versus that obtained via the eigenvalue inflation scheme above

Figure 4.11: Constraint violation plot for the eigenvalue inflation scheme

References

1. Heinemann, M. & Sauer, U. Systems biology of microbial metabolism. *Current Opinion in Microbiology* **13,** 337–343. ISSN: 1369-5274 (2010).

2. Kitano, H. Computational systems biology. **420** (2002).

3. Antoniewicz, M. R. & Kelleher, G. S. Joanne K. Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions. *Metabolic Engineering* **9,** 68–86 (2007).

4. Garcia Martin, H. *et al.* A Method to Constrain Genome-Scale Models with ^{13}C Labeling Data. *PLoS Comput Biol,* 1–34 (2015).

5. Zaman, S., Lippman, S. I., Zhao, X. & Broach, J. R. How Saccharomyces Responds to Nutrients. *Annual Review of Genetics* **42,** 27–81 (2008).

6. Kayikci, Omur and Nielsen, J. Glucose repression in *Saccharomyces cerevisiae. FEMS Yeast Research* **15,** 1–8 (2015).

7. Zhang, J., Olsson, L. & Nielsen, J. The β-subunits of the Snf1 kinase in Saccharomyces cerevisiae, Gal83 and Sip2, but not Sip1, are redundant in glucose derepression and regulation of sterol biosynthesis. **77,** 371–383 (2010).

8. Breslow, D. K. *et al.* A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nature methods* **5,** 711–718. ISSN: 1548-7091 (2008).

9. Shymansky, C. M. ^{13}C *Metabolic flux-analysis-aided exploration of the high-glucose role of the Sip1 β-subunit of the Snf1 kinase complex in Saccharomyces Cerevisiae.* PhD thesis (University of California, Berkeley, 2011).

10. Mylin, L. M. *et al. SIP1* Is a Catabolite Repression-Specific Negative Regulator of *GAL* Gene Expression. *Molecular Biology* **137,** 689–700 (1994).

11. Vyas, V. K., Kuchin, S., Berkey, C. D. & Carlson, M. Snf1 Kinases with Different β-Subunit Isoforms Play Distinct Roles in Regulating Haploid Invasive Growth. *Molecular and Cellular Biology* **23,** 1341–1348 (2003).

12. Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science (New York, N.Y.)* **292,** 929–934. ISSN: 00368075 (2001).

13. Entian, K.-D. & Kötter, P. 25 yeast genetic strain and plasmid collections (2007).

14. Güldener, U., Heck, S., Fiedler, T., Beinhauer, J. & Hegemann, J. H. A New Efficient Gene Disruption Cassette for Repeated Use in Budding Yeast. *Nucleic Acids Research* **24,** 2519–2524 (1996).

15. Agatep, R., Kirkpatrick, R. D., Parchaliuk, D. L., Robin, A. & Gietz, R. D. Transformation of Saccharomyces cerevisiae by the lithium acetate / single-stranded carrier DNA / polyethylene glycol (LiAc/ss-DNA/PEG) protocol. **3,** 133–137 (1998).

16. Sikorski, R. S. & Hieter, P. A System of Shuttle Vectors and Yeast Host Strains Designed for Efficient Manipulation of DNA in Saccharomyces Cerevisiae. *Genetics* **122,** 19–27 (1989).

17. Bokinsky, G. *et al.* HipA-Triggered Growth Arrest and β-Lactam Tolerance in Escherichia coli Are Mediated by RelA-Dependent ppGpp Synthesis. **195,** 3173–3182 (2013).

18. Weaver, L. J. *et al.* A Kinetic-Based Approach to Understanding Heterologous Mevalonate Pathway Function in *E. coli.* **112,** 111–119 (2015).

19. Mo, M. L., Palsson, B. Ø. & Herrgård, M. J. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Systems Biology* **3,** 1–17 (2009).

20. Blank, L. M., Kuepfer, L. & Sauer, U. Large-scale ^{13}C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. **6,** 1–16 (2005).

21. Blank, L. M. & Sauer, U. TCA cycle activity in *Saccharomyces cerevisiae* is a function of the environmentally determined specific growth and glucose uptake rates. **150,** 1085–1093 (2004).

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

23. Maaheimo, H. *et al.* Central carbon metabolism of *Saccharomyces cerevisiae* explored by biosynthetic fractional ^{13}C labeling of common amino acids. **268,** 2464–2479 (2001).

24. Escalante-Chong, R. *et al.* Galactose metabolic genes in yeast respond to a ratio of galactose and glucose. **112.** doi:`10.1073/pnas.1418058112` (2015).

25. Nijkamp, J. F., Broek, M. V. D., Datema, E., Kok, S. D. & Bosman, L. De novo sequencing , assembly and analysis of the genome of the laboratory strain Saccharomyces cerevisiae CEN . PK113-7D , a model for modern industrial biotechnology De novo sequencing , assembly and analysis of the genome of the laboratory strain Saccharomyces cerevisiae CEN . PK113-7D , a model for modern industrial biotechnology. *Microbial Cell Factories* **11,** 36. ISSN: 1475-2859 (2012).

26. Cordiera, H., Mendes, F., Vasconcelos, I. & Francois, J. A metabolic and genomic study of engineered *Saccharomyces cerevisiae* strains for high glycerol production. **9,** 364–378 (2007).

27. Zhang, G.-c., Liu, J.-j. & Ding, W.-t. Decreased Xylitol Formation during Xylose Fermentation in *Saccharomyces cerevisiae* Due to Overexpression of Water-Forming NADH Oxidase. **81,** 1081–1086 (2012).

28. Vemuri, G., Eiteman, M., McEwen, J., Olsson, L. & Nielsen, J. Increasing NADH oxidation reduces overflow metabolism in *Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America* **104,** 2402–2407. ISSN: 0027-8424 (2007).

29. Hou, J. *et al.* Fine-tuning of NADH oxidase decreases byproduct accumulation in respiration deficient xylose metabolic *Saccharomyces cerevisiae. BMC Biotechnology* **14,** 1–10. ISSN: BMC Biotechnology (2014).

30. Kim, J.-w., Seo, S.-o., Zhang, G.-c., Jin, Y.-s. & Seo, J.-h. Expression of Lactococcus lactis NADH oxidase increases 2,3-butanediol production in Pdc-deficient Saccharomyces cerevisiae. *Bioresource Technology* **191,** 512–519. ISSN: 0960-8524 (2015).

31. Stephanie Heux, Remy Cachon, S. D. Cofactor engineering in *Saccharomyces cerevisiae*: Expression of a H_2O-forming NADH oxidase and impact on redox metabolism. **8,** 303–314 (2006).

32. Nevoigt, E. *et al.* Engineering of Promoter Replacement Cassettes for Fine-Tuning of Gene Expression in *Saccharomyces cerevisiae. Applied and Environmental Microbiology* **72,** 5266–5273. ISSN: 00992240 (2006).

33. Grant, M. & Boyd, S. *CVX: Matlab software for disciplined convex programming, version 1.21.* <http://cvxr.com/cvx/> (2011).