Metabolic Model Design and Elementary Mode Analysis of *Shewanella oneidensis* MR-1 and Derivative Strains

&

Plasmid Construction to Facilitate PHB Production in *Saccharomyces cerevisiae* Using a Single Vector

A THESIS

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL

OF THE UNIVERSITY OF MINNESOTA

BY

Chris M. Flynn

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF SCIENCE

Friedrich Srienc

University of Minnesota

April 2010

Acknowledgements:

I would like to thank my advisors and collaborators on this project, Dr. Jeffrey Gralnick, Dr. Daniel Bond, Dr. Friedrich Srienc, Jeff Flynn, and Kris Hunt. Also, special thanks goes to my colleagues Cong Trinh, Apple Unrean, Alan Gilbert, Pedro Pena, Dan Rouse, and Greg Sitton for their assistance and guidance in developing this research. This research has been supported by the U of M Institute on the Environment/IREE Discovery Grants on Biofuels and the Environment program.

Abstract:

In Chapter 1, a stoichiometric model describing the central metabolism of *Shewanella oneidensis* MR-1 wild-type and derivative strains was developed for use in elementary mode (EM) analysis. An EM model was created and verified by comparing growth phenotypes of single- and double-knockout strains cultivated on differing carbon sources, under aerobic and anaerobic conditions. Furthermore, several single knockout growth phenotypes under specific conditions have been predicted, offering potential for model verification. This wild-type model can be easily adapted to different carbon sources or metabolic products, and allows the prediction of single- and multiple- knockout strains that are expected to operate under defined conditions with increased efficiency when compared to wild type cells. To this end, the *S. oneidensis* wild-type EM model was adapted by addition of both glycerol import reactions and reactions needed to facilitate ethanol production from pyruvate. EM simulations were then performed on this glycerol-ethanol model to predict that the optimal anaerobic, ethanol producing strain cultivated on glycerol would contain the following mutations: *Δpta ΔaldA ΔgcvT*. This strain is expected to require ethanol secretion to accumulate biomass under anaerobic conditions, and as a result to produce ethanol yields between 0.4668 mol ethanol/mol glycerol and 1 mol ethanol/mol glycerol consumed (theoretical yield is 1 mol/mol). Future work should focus on performing these gene deletions and characterizing the resultant mutant strains, as well as determining the role of formate in anaerobic *shewanella* metabolism.

In addition, the steps in generating a single plasmid which expresses all three *R. eutropha* PHA synthesis genes in *Saccharomyces cerevisisae* is described in Chapter 2. Several methods were employed to attempt to generate this plasmid, while plasmid construction was successful using yeast-mediated ligation and inserting the TEF1-*phaC1* promoter-gene into p2DPT-RK(U) to create the plasmid p2DPTT1-RKS1(U). The GAL1-10 promoter is present in the host vector, and it is likely that previous efforts were hampered by plasmid instability resulting from undesired intra-plasmid homologous recombination between these duplicate GAL1-10 promoter regions. The exact sequence of the region downstream of the *phaC1* gene should be determined prior to application PHA production and other experiments. This plasmid should facilitate the direct observation of PHA accumulation *in vivo* using fluorescence microscopy, as well as improve PHA yields in transformed strains when compared to the two-plasmid system.

Contents

Abstract: ii

List of Figures vi

List of Tables vi

Chapter 1: Metabolic Model Design and Elementary Mode Analysis of *Shewanella oneidensis* MR-1 and Derivative Strains 1

1.1: Introduction 1

1.1.1: *Shewanella oneidensis*: introduction and distinguishing characteristics 1

1.1.2: Elementary mode analysis: introduction and central tenets 4

1.2: Materials and methods: elementary mode analysis 6

1.2.1: Elementary mode calculation 6

1.2.2: Biomass term 6

1.2.3: Reaction omission, sorting rationale 8

1.2.4: Condition simulation through reaction omission: 10

1.2.5: Iterative sorting application 13

1.2.6: Elementary mode sorting algorithm 14

1.3: Results 20

1.3.1: Model construction 20

1.3.2: Model verification: 24

1.3.3: Additional knockout target genes for model verification: 25

1.4: Discussion 27

1.5.1 Model implementation: anaerobic conversion of glycerol to ethanol in the presence of an electrode. 31

1.5.2: Elementary mode simulation results 33

1.5.3: Discussion 36

Chapter 2: Plasmid Design & Construction for PHB Production in *Saccharomyces cerevisiae* Using a Single Vector 41

2.1: Introduction & Background 41

2.1.1: Introduction 41

2.1.2: PHB synthesis 42

2.1.3: Single-plasmid expression rationale 43

2.1.4: Single plasmid applications in fluorescence microscopy 47
2.1.5: GAL1-10 divergent promoter 49
2.1.6: TEF1 promoter systems 52
2.2: Materials and Methods 53
2.2.1: Experimental strains 53
2.2.2: Bacterial growth media 53
2.2.3: Yeast growth media 54
2.2.4: Strain preservation 54
2.2.5: Hydrophobic staining with BODIPY (493/503) and microscopic slide preparation 54
2.2.6: PHB determination by propanalysis 55
2.2.7: Plasmid isolation 56
2.2.8: Restriction endonuclease digestion 56
2.2.9: Phenol-chloroform extraction with ethanol precipitation 56
2.2.10: Generating blunt DNA fragments using T4 polymerase 57
2.2.11: Generating blunt DNA fragments with large fragment (Klenow) polymerase I 57
2.2.12: Vector dephosphorylation for use in ligation 58
2.2.13: Ligating blunt fragments 58
2.2.14: DNA electrophoresis and gel purification 58
2.2.15: Preparing chemically competent DH5α cells 59
2.2.16: Transforming chemically competent DH5α cells 59
2.2.17: Transforming chemically competent *S. cerevisiae* 60
2.2.18: Yeast mediated ligation using *in vivo* homologous recombination 61
2.2.19: *E. coli* mediated ligation using *in vivo* homologous recombination 61
2.2.20: Plasmid Sequencing primers, conditions, and analysis 62
2.2.21: Source plasmids 62
2.3: Plasmid Construction Results and Discussion: 63
2.3.1: Plasmid characterization: UraTerm Sequencing 63
2.3.2: Construction of a single vector for PHB expression in yeast: homologous recombination around the aatII site using yeast mediated ligation. 64
2.3.3: Construction of a single vector for PHB expression in yeast: homologous recombination around the aatII site using *E. coli* mediated ligation. 70

2.3.4: Construction of a single vector for PHB expression in yeast: blunt-end ligation at the sstI restriction digest site of p2DPT-RK(U). 73

2.3.5: Construction of a single vector for PHB expression in yeast: ligation at the sstI restriction digest site of p2DPT-RK(U) using yeast mediated ligation..... 78

2.4: Project Summary and Future Work 82

3.1: References..... 86

List of Tables

Table 1.1: Reaction omissions based on culture condition & regulation..... 13
Table 1.2: Model predictions compared with experimental results of single knockout mutants... 25
Table 1.3: All predicted lethal single deletion *S. oneidensis* mutants under conditions tested..... 26
Table 1.4: Potential gene knockouts for model validation..... 27
Table 1.5: The effect of gene deletion on elementary mode number in *S. oneidensis* pGUT2 pPET.....33
Table 2.1: Mitotic stability and segregation rates for high copy number 2µ plasmids..... 44
Table 2.2: Primers used in homologous recombination reactions 66
Table 2.3: Experimental samples for PHB analysis from shake flask 68
Table 2.4: Expected NcoI digestion bands..... 76
Table 2.5: Appropriate test and control transformations for any blunt-end ligation creation and assessment experiment..... 77
Table 2.6: Primers used in homologous recombination reactions II..... 78
Table 2.7: Expected and observed digestion fragments from yeast mediated ligation. 80

List of Figures

Figure 1.1: Preliminary metabolic map describing a “glycerol-utilizing organism". 3
Figure 1.2: Commented pseudocode used for elementary mode sorting and analysis. Part A. 15
Figure 1.3: Commented pseudocode used for elementary mode sorting and analysis. Part B. 16
Figure 1.4: Commented pseudocode used for elementary mode sorting and analysis. Part C. 17
Figure 1.5: Metabolic map describing the *S. oneidensis* MR-1 central metabolism. 24
Figure 1.6: Serine-glycine superpathway and proposed formate assimilation and oxidation reactions. 29
Figure 1.7: Metabolic map of *S. oneidensis* MR-1 pGUT2 pPET. 32
Figure 1.8: Effect of reaction deletions on the number of anaerobic EMs for growth on glycerol. 35
Figure 1.9: Metabolic map showing knockout reactions and conditionally-prohibited reactions expected to result in the optimal glycerol consuming strain which secreted ethanol 39
Figure 2.1: PHB polymerization from acetyl-CoA.. 42
Figure 2.2: D603, D503 p2DPT-RK(U), H1246 fluorescence microscopy using BODIPY staining 49
Figure 2.3: Glucose repression and galactose derepression of GAL1-10 genes. 50
Figure 2.4: The GAL1-10 divergent promoter system 52
Figure 2.5: High copy number yeast PHB production plasmids utilizing the divergent GAL1-10 promoters. 63
Figure 2.6: Sequence result from p2DPT-RK(U) of phaA-uraTerm-ura3 region 64
Figure 2.7: Insertion of the GAL1-10 promoter with *R. eutropha phaC* behind the GAL10 promoter into p2DPT-RK(U) at the AatII digest site.. 65
Figure 2.7: Insertion of the GAL1-10 promoter with *R. eutropha phaC* behind the GAL10 promoter into p2DPT-RK(U) at the AatII digest site.. 65
Figure 2.8: Primer test of GAL1-phaC primers with different homologous recombination target regions. Each set of four lanes is at annealing temperatures between 55-68°C from left to right. 66
Figure 2.9: Comparing isolates of H1246 p2DPT-RKS(U) 67
Figure 2.10: Propanalysis standard curve and production results from shake flask cultivation 68
Figure 2.11: Restriction digestion of PHB plasmids using the AgeI, EcoRI, and HpaI. 69
Figure 2.12: PCR screen results of *E. coli* mediated ligation showing amplification of the GAL1-10-*phaC* region in three of 24 test samples and the positive control. 71
Figure 2.13: Restriction map of p2DPT-RKS(U) isolates generated using *E. coli* mediated homologous recombination. 72
Figure 2.14: Plasmid construction inserting TEF1-*phaC* (*R. eutropha* synthase) and the TEF1-*phaC1* (*P. oleovorans* synthase) cassettes into p2DPT-RK(U) following SstI digest of this vector. 73
Figure 2.15: PHB synthase ligation results 75
Figure 2.16: Restriction map characterization of p2DPT-RKS(U) blunt-end ligation products using NcoI digestion.. 76
Figure 2.17: Restriction map of two p2DPT-RKS1(U) isolates.. 79
Figure 2.18: Promoter-target sequence interaction hypothesis describing observed cloning results. 81

Chapter 1: Metabolic Model Design and Elementary Mode Analysis of *Shewanella oneidensis* MR-1 and Derivative Strains

1.1: Introduction

1.1.1: *Shewanella oneidensis*: introduction and distinguishing characteristics

For decades, species of the genus *Shewanella* have been known to respire a wide variety of extra-cellular compounds under anaerobic conditions during growth, including many soluble and insoluble metal ions, such as iron oxide ($Fe_2O_{3(s)}$), $Mn(IV)_{(s)}$, $Cr(VI)_{(aq)}$, $U(VI)_{(aq)}$, Co(II), among others (1, 2). This reduction capacity is facilitated by the dissimilatory metal reduction complex (DMRC). Briefly, the DMRC is a complex of membrane proteins which oxidizes membrane quinols by transferring electrons to external electron acceptors, including Mn^{2+}, Fe^{3+}, or the cathode contact in an electrical circuit (3).

As a result of its versatile respiratory repertoire, this genus is of great interest in applications including heavy metal remediation, microbial fuel cells, and anaerobic metabolite production in the biotechnology sector. In whole-cell biocatalysis applications, oxygen is often a limiting reactant, due to its low solubility. As a result, it is preferable to operate bioreactors under high density, anaerobic conditions whenever possible. The respiratory flexibility of Shewanellae imparts the potential to generate biomaterials that require aerobic conditions in other strains. This could be achieved through cultivation in the presence of electron acceptors other than oxygen, such as electrodes, and soluble or insoluble electron acceptors.

The creation and verification of accurate metabolic network models is essential to both improving understanding this system, as well as in optimizing whole-cell biocatalysis by guiding the creation of multiple gene-knockout strains with improved metabolite yields. Currently available metabolic networks, however, such as those available in the KEGG and Metacyc databases, are often based largely on gene annotation. In addition, these sources may lack reaction accuracy, manifest in the form of incorrectly annotated genes or incorrect/unspecified reaction directionality, and are generally incapable of generating accurate metabolic pathway predictions because of these limitations. Indeed, reaching such a level of accuracy and numerical soundness is not the goal of such networks, and they should be treated as just one of many tools necessary for the creation of useful metabolic networks. Furthermore, these sources contain thousands of reactions; due to computation constraints, it is vital to establish which pathways are relevant to cellular function and product yield (i.e, pathways that contain significant carbon/energy flux under the desired conditions) while keeping network complexity to a minimum. In this work, we sought to combine all available data describing the *Shewanella oneidensis* MR-1 central metabolism under usual cultivation conditions, and using elementary mode (EM) analysis, generate easily testable viability predictions under several cultivation conditions for knockout mutants. As a result, previous research describing carbon flux characteristics, gene annotation, and experiments regarding consumption, secretion, and viability of gene knockout mutants were combined with central metabolism information obtained from the KEGG and Metacyc databases to generate a wild-type model describing the *S. oneidensis* central metabolism under several culture conditions.

Elementary mode analysis has previously been shown to be effective in designing highly efficient *Escherichia coli* strains that produce ethanol, carotenoids, and other metabolic products (4). Despite having been sequenced and studied extensively, however, no model with predictive capability has been developed to describe the *S. oneidensis* central metabolism to date. A preliminary example was created by Cong Trinh, by modifying previously generated *E. coli* elementary mode model based on information available in public databases and available literature (5-8). This preliminary model, shown in Figure 1.1, constitutes the starting point of the following sections which describe the design of a *S. oneidensis* MR-1 central metabolic model for use in EM analysis.

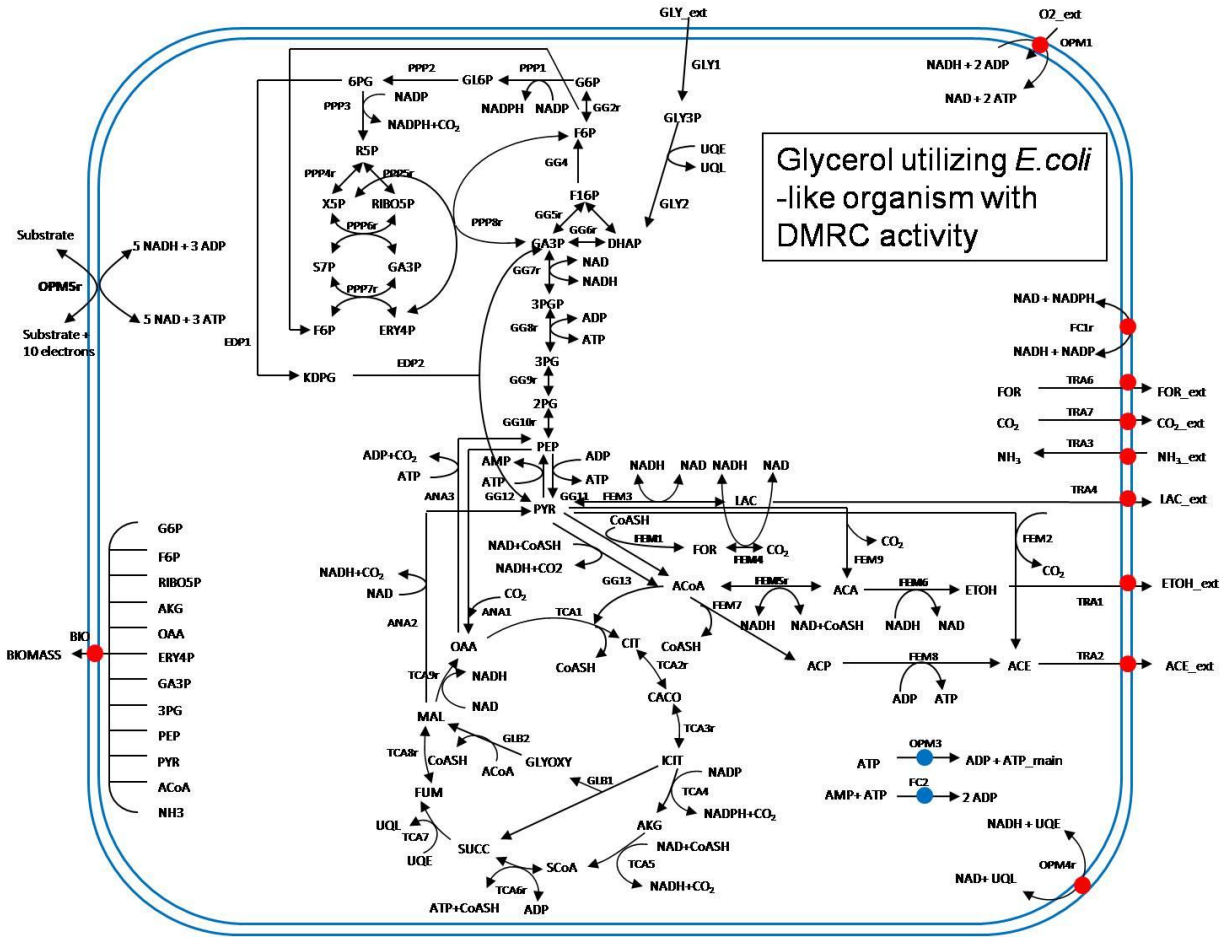


Figure 1.1: Preliminary metabolic map describing a "glycerol-utilizing organism with a central metabolism similar to *E. coli*" with the *S. oneidensis* pathway which allows electron excretion.

1.1.2: Elementary mode analysis: introduction and central tenets

The *S. oneidensis* MR-1 metabolic model was created specifically for application in EM analysis. EM analysis computes all possible (thermodynamically feasible), non-divisible pathways through a given set of system of stoichiometric reactions, assuming pseudo steady state conditions (9, 10). A primary tenet of EM analysis is that the overall flux through a given system can be represented as the weighted sum of all elementary modes, as shown below.

$$M = \sum_i E_i c_i, \text{ and} \qquad \text{(Eq. 1)}$$

$$\sum_i c_i \equiv 1 \qquad \text{(Eq. 2)}$$

Where M is the overall metabolic flux through all described metabolic pathways, E_i is each elementary mode, and c_i being the scalar weighting factor describing the fractional contribution given to each elementary mode (11). Elementary modes with no net flux would be assigned a c_i of zero. In addition, to determine net metabolite flux, the following equations summarize this relationship:

$$R = \alpha M = \alpha \sum_i E_i c_i \qquad \text{(Eq. 3)}$$

Where R is the net metabolic flux through all reactions. If written using matrix notation:

$$R = \alpha E c \qquad \text{(Eq. 4)}$$

If N is the dot product of α and E (the stoichiometric matrix and all elementary modes), N then describes the net accumulation rates of all elementary modes. Because no internal

metabolites accumulate under the pseudo steady state assumption, all non-zero values in N describe external metabolite accumulation (12)[1]. So overall flux accumulation through all elementary modes can be described by the following equation using matrix notation:

$$R = \mathrm{N}c \qquad \text{(Eq. 5)}$$

Finally, if we combine all elementary modes with identical external accumulation rates into mode families, the overall flux can be described by the simplified relationship below, where N' is the matrix containing the external accumulation rates of all mode families, and c' is the weighting factor assigned to each mode family. Each entry in c' (a vector of scalars) must therefore be the sum of all elementary mode weighting factors combined into each family, while the sum of all entries in c' must be unity, as shown previously in Eq. 2.

$$R = N'c' \qquad \text{(Eq. 6)}$$

Using this equation, inferring pathway flux values, (mode weighting factors) no longer requires an exact knowledge of the flux fraction through each individual elementary mode, but instead the total flux through each mode family, substantially reducing the number of unknown values in the system. Indeed, if the number of relationships is equal to or greater than the number of mode families, the weighting factors can be determined, because all relationships in Eq. 6 are linearly independent (11). This is not necessarily true of Eq. 5, where, while all elementary modes are *genetically* independent (non-decomposable), they are not necessarily *linearly* independent (9, 12). As a result,

[1] "External" metabolites need not be actually outside the cell. In this context, external metabolites have been defined as any species that accumulates or depletes. As such, both biomass and glucose are both examples of "external" metabolites in this system (56).

combining elementary modes into mode families provides a valuable tool for determining the composition of fluxes that form a metabolic network by reducing the number of unknown weighting factors and by generating a reduced flux matrix that consist of easily solvable linearly independent vectors. This potentially allows determination of all c' values by measuring only external metabolites.

1.2: Materials and methods: elementary mode analysis

1.2.1: Elementary mode calculation

Elementary mode (EM) analysis was performed on metabolic models using the free publicly available software Metatool v5.1 in the Matlab environment (MathWorks, Inc., Natick, MA) (13). The central metabolism of *S. oneidensis* was described using 65 reactions, 20 of which were reversible, and with 63 metabolites, 14 of which were external. The Metatool software and its application are described in detail by Pfeiffer, Schuster *et al.* (1999).

1.2.2: Biomass term

The biomass term was calculated based on *E. coli* cellular compositions and theory as described by Niedhardt *et al.* (1990), which described biomass composition in terms of 12 precursor metabolites generated in the *E. coli* central metabolism. In addition, 6 energy carriers, elements, or cofactors (ATP, NAD(P)(H), S, NH_4^+, etc) are required for the synthesis of these metabolites. Values were normalized to determine the metabolite requirement to generate 1 g cell dry weight of *E. coli* (5). This method of estimating biomass composition was previously implemented in several experiments involving elementary mode analysis (4, 14-17). Notably, these previous implementations were in the *E. coli* system, and not *S. oneidensis*, presenting the possibility large differences in

biomass composition between species may result in incorrect biomass calculations. However, comparison using C^{13} composition analysis has shown that biomass composition between species is within experimental error in most respects, supporting the conclusion that there is no significant difference in overall biomass composition between *E. coli* and *S. oneidensis* with respect to carbon composition (6). Minor differences between cellular composition is not likely to affect elementary mode outcomes. This is for several reasons. First, biomass flux forms a relatively small fraction of total carbon flux directed toward biomass production (approximately 13.9% under anaerobic, lactate consuming conditions with TMAO reduction (6)). Second, the large number and variety of metabolites contributing to biomass production result in diluting minor deviations in biomass composition, as each of the 13 carbon sources involved in biomass accumulation in this model contributes only a fraction of the final carbon content in the overall cell. Of far greater importance than the absolute quantities of each metabolite leading to biomass production is the accounting for all species required for biomass production (necessitating maintenance of modes through such pathways), and also the large ATP and NADPH requirements of biomass accumulation. Maintaining the correct ratio of metabolite types, such as proteins, lipids, and ATP requirements is likely sufficient to generating accurate, even if somewhat imprecise, pathway predictions.

The composition of the biomass term was modified from containing these 12 carbon sources by separating a fraction of the 3-phosphoglycerate term into serine and glycine components. This modification was essential, due to the inclusion of the serine-glycine superpathway in this system. Carbon and energy requirements for the production

of these amino acids were determined based on equivalents determined by Neidhardt *et al.* (1990).

1.2.3: Reaction omission, sorting rationale

The sorting method employs the property of EM's that, when placed under constrains, such as the presence or absence of specific reactions, the resulting matrix of EM's forms a subset within the original, unconstrained elementary mode space. In analytical form,

$$E' \in \text{E, and} \qquad \text{(Eq. 7)}$$

$$\sum_i^n \{E'_1, E'_2, E'_3, \ldots E'_n\} \leq E \qquad \text{(Eq. 8)}$$

Where E is all possible elementary modes that describe a given metabolic network, and E' is any subset of E, as defined by any number or combination of reaction omissions. Each member of E' is a column vector representing one elementary mode, and is also a member of E. Eq. 8 is simply an implied property of Eq. 7, stating that the combination of all E' (provided any duplicate EM's are combined) is equal to or less than the overall model solution space. From these traits, two forms of analysis are allowed. First, computation time can be saved when iteratively analyzing EM networks with combinations of reaction deletions (applied in section 1.5.1), because the EM output of a particular condition must be calculated only once, instead of requiring re-calculation after each deletion simulation. This allows rapid condition and gene knockout simulation *in silico*, as it is not necessary to recalculate EM's with every reaction omitted from the system. Second, the trait described by Eq. 8 allows separate calculation of major mode sets, and their combination to describe all or nearly all of the possible elementary modes

possible within the original model. Due to combinatorial expansion (i.e., computation constraints), as well as standard culture conditions, EM's are typically calculated by allowing growth using only a single carbon source. In addition, because individual cells often consume only a single carbon source at any given time due to gene regulation, this convention is often relevant even in cases where multiple carbon sources are contained in the culture medium. Possible exceptions to this generalization include some genetically engineered strains, which due to gene deletion/mutation may co-consume several carbon sources at one time[2]. However, to describe an entire metabolic network and its overall complexity, the modes obtained by individual calculation can be combined to represent most relevant elementary modes. The caveat of this trait, however, is that any elementary modes involving co-consumption would not be contained in the combined matrix; this is the source of inequality in Eq 8. Should no co-culture or futile EM's exist, then the combination of all E' would equal E exactly. This is best explained by an example. In a given network the number of elementary modes, when allowed n-acetyl-glucosamine (NAG) as the only carbon source, is 1000. For the same network, instead allowing only lactate as a carbon source, gives 500 modes. Therefore, the number of elementary modes describing that network, if provided with NAG and lactate as carbon sources, would contain at least 1500 elementary modes, with any additional elementary modes being the result of elementary modes containing both lactate and NAG consumption. If no co-consumption modes exist or futile cycles exist, then the total number of elementary

[2] An example of such an engineered co-consumption mutant was created by elementary mode analysis-directed mutation, and is described in Trinh, Unrean, & Srienc (2008) *"Minimal* Escherichia coli *cell for the most efficient production of ethanol from hexoses and pentoses."* Appl Environ Microbiol. 2008 June; 74(12):3634-43

modes would equal 1500 exactly. Finally, this characteristic applies to any combination of reaction omissions, not just carbon source limitation, such as aerobic/anaerobic, etc.

1.2.4: Condition simulation through reaction omission:

Anaerobic conditions were simulated by omitting reactions involving oxygen as an external metabolite, thus preventing aerobic respiration. In conditions requiring the dissimilatory metal reduction complex (DMRC) activity, the reaction describing the reduction of external fumarate to succinate was also omitted, as external fumarate will not be supplied in such conditions. Likewise, in conditions where fumarate is supplied, DMRC reactions will be omitted, and the fumarate reduction reaction included. As previously stated, it is often not possible to calculate all EM's of a particular reaction network that contains multiple carbon sources at one time due to computation constraints. As a result, reaction omission was utilized to simulate carbon source availability; while the full network contains reactions to import NAG, lactate, and acetate, all but one reactant import equation was omitted prior to EM calculation under each growth condition. This method simulates the cultivation of *S. oneidensis* under single-carbon source conditions.

Reactions were also omitted from EM analysis based on known regulation and enzyme deactivation, inferred regulation based on single knockout observations *S. oneidensis* performed by myself and collaborators K. Hunt and J. Gralnick (summarized in Table 1.2), and inferred regulation based on observations made in *E. coli*. Specifically, under aerobic conditions, all EM's containing the reaction catalyzed by pyruvate formate-lyase (Pfl, gene name *pflB*) are omitted from further analysis. This conclusion was made for

two reasons. First, Pfl is irreversibly denatured by oxygen in *E. coli*, resulting in loss of function (18). Second, the inactivity of Pfl under aerobic conditions in *S. oneidensis* is supported by the observed inability of *S. oneidensis* MR-1 *ΔaceE* mutants to grow on both lactate and NAG under aerobic conditions, while maintaining the ability to accumulate biomass anaerobically on those same carbon sources (19). The gene *aceE* encodes an essential component of the pyruvate dehydrogenase complex (pdc), and catalyzes the conversion of pyruvate to acetyl-CoA and carbon dioxide, while also producing NADH. The *ΔaceE* knockout phenotype supports the conclusion that Pfl is inactive under aerobic conditions in *S.* oneidensis MR-1. Similarly, the 2-oxoglutarate dehydrogenase complex (*SucAB*), and isocitrate lyase (*aceA*) containing EM's were omitted from anaerobic simulations based on previously reported activity in *S. oneidensis* (20). Finally, the anaerobic inviability of *ΔpflB* strains on NAG and lactate suggest a lack of anaerobic pdc activity in *S. oneidensis*, similar to *E. coli* in this respect (19). This demonstrates that *aceE* is not active under anaerobic conditions in sufficient quantities to facilitate growth by converting pyruvate to acetyl-CoA.

In addition, isocitrate lyase (*aceA*) and acetyl-CoA synthetase (*Acs*) were assumed to be active only when acetate is used as a sole carbon source. Because of the acetate side chain in NAG, (slow) growth would be expected via NAG degradation, with pyruvate secretion and acetate fixation via Acs and isocitrate lyase activity, unless such activity is repressed strongly enough to prevent all growth through that pathway. The absence of aerobic NAG-consumiing growth in *ΔaceE* strain, as well as anaerobic growth in the *Δpfl* strain, directly indicate that the glyoxylate pathway is not active unless acetate is the sole

carbon source under aerobic conditions (19). This conclusion is further supported by the observed growth phenotype of *Δppc* strains under aerobic conditions, where this strain was not capable of growth on NAG or lactate, unlike wild-type strains. This suggests that the only means of replenishing citric acid cycle intermediates in *S. oneidensis* is through the conversion of PEP to oxaloacetate via Pepc activity when grown on lactate or NAG. Based on simulations, this phenotype was predicted only if there is no glyoxylate flux, which would provide an alternate route to produce oxaloacetate by converting acetyl-coA into malate via isocitrate-lyase.

A complete list of reaction omissions under each respective culture condition presented in Table 1.1. Notably, the anaerobic characterization of knockout mutants was performed with fumarate as an external electron acceptor, so DMRC activity was omitted during viability testing, as no reduction would be expected to occur via this reaction under the studied conditions.[3]

[3] Knockout viability patterns were identical whether anaerobic reduction occurred via DMRC or fumarate reduction, excepting that a reversal of lethality with reduction mechanism deletion (i.e, DMRC deletion was lethal in conditions requiring DMRC activity, but not conditions requiring fumarate reduction and vice versa).

Table 1.1: Reaction omissions based on culture condition & regulation

	lactate consumption		NAG consumption		acetate consumption
gene	**aerobic growth**	**anaerobic growth**	**aerobic growth**	**anaerobic growth**	**aerobic growth**
omitted genes & functions	*pflB* (pyruvate formate-lyase)	oxidative phosphory-lation	*pflB* (pyruvate formate-lyase)	oxidative phosphor-ylation	*pflB* (pyruvate formate-lyase)
	DMRC activity	DMRC activity	DMRC activity	DMRC activity	DMRC activity
	external fumarate reduction	*sucB* (2-oxogluta-rate DH)	external fumarate reduction	*sucB* (2-oxoglutarate DH)	external fumarate reduction
	Glyoxy-late pathway	*aceA* (isocitrate lyase)	*acs* (Acetyl-coA synthe-tase)	*aceA* (isocitrate lyase)	
		aceE (pyruvate decarboxylase)	Glyoxy-late pathway	acs (Acetyl-coA synthetase)acs	
		Glyoxylate pathway		*aceE* (pyruvate decarboxylase)	
				Glyoxylate pathway	

1.2.5: Iterative sorting application

Target genes for deletion were identified using the following semi-iterative method.

Knockout identification was performed using an adaptation of a previously reported semi-automated iterative method (4). Target genes were identified based on the effect of gene (or reaction) deletion on the number of elementary modes remaining, maximum and minimum yield of ethanol, and finally the normalized biomass yield. Deletions were selected based on the following criteria:

1. The most preferable deletion results in the greatest reduction of modes,

2. Any deletion should maintain the potential to produce the highest possible product yield, and

3. The maintenance of non-zero biomass yield under the desired conditions (i.e, any mutation must be non lethal).

This process was then repeated iteratively, until any of the following conditions were met:

1. Additional deletions violate the principles above (i.e, all remaining targets result in cell death or large yield reductions)
2. Additional deletions would be silent, or
3. If remaining knockouts are not expected to result in yield improvements.

This iterative method is a means to achieving two goals in EM reduction: The primary goal is to remove non-producing elementary modes from the system, thereby linking product secretion to biomass accumulation (growth). Secondly, the greatest possible reduction in elementary modes with a minimum number of reaction (gene) deletions is desirable, both to miminize labor costs involved in knockout mutant generation, and because a reduction in mode number, and by extension the number of mode families, may allow determination of flux modes through the cell, as previously described in Section 1.1.2.

1.2.6: Elementary mode sorting algorithm

The following algorithm was developed for sorting EM outputs for all networks being studied. The accuracy of this application was determined by comparing EM results obtained on previously characterized systems by calculating EM's following iterative deletion using both the algorithm below, and network modification and recalculation at each step using Metatool (4). In all cases, identical EM composition and number were

observed (data not shown) using both recalculation of EM matrices and the aforementioned sorting algorithm. The commented pseudocode is reproduced below in Figures 1.2-1.4:

```
function KO_Network(ems,react_name,obj,delrxn, singleDel)
% Notation:                                                              *
%   + obj       : targets whose fluxes are nonzero                       *
%   + delrxn    : deleted reactions                                      *
%   + singleDel : already deleted/omitted reactions
IF nargin < 5
     DISPLAY('Error: Require full supply of inputs\n')
     DOEXIT
END
    FOR i = 1:length(obj)
        i_obj = find(strcmp(react_name,obj(i)));%find obj index, react_name
        obj(i);
        ems=ems(:,find(ems(i_obj,:)~=0));%find EMs with ~=0 fluxes thru obj
        emsA = ems;  %define new EM matrix, reduced to EMS ~=0 flux in obj

       IF ~length(i_obj) %error checking: exit if obj is not in react_name,
            DISPLAY('Error: An rxn name in the obj is not in react_name\n')
            DOEXIT
        END
    END
    DISPLAY(('EMS after obj applied'), length(emsA)) %print ems length
%print length of ems where flux through all 'obj' rxn is not zero
```

Figure 1.2: Commented pseudocode used for elementary mode sorting and analysis. This panel shows the input of relevant data and the algorithms for selection of EM's with non-zero flux through "obj" reactions.

```
    FOR i = 1:length(singleDel)
        i_singleDel = find(strcmp(react_name,singleDel(i))); % find index of
obj in react_name
        singleDel(i);
        emsA = emsA(:,find(emsA(i_singleDel,:)==0)); % find elementary modes
where fluxes through obj reaction is non zero
        emsB = emsA; %define new EM matrix, reduced to include EM's in emsA &
zero flux through deletions in singleDel
        DISPLAY('Number of EMS with singleDel deletion.')
      DISPLAY(size(emsB, 2) %display number of columns in emsB (= # em's)

        IF ~length(i_singleDel) %error checking: if singleDel rxn is not in
react_name, program exits
            DISPLAY('Error: A rxn in singleDel is not in react_name\n')
          DOEXIT
        END
    END

%output bio mode # left after KO's applied
bio_findModes = find(strcmp(react_name,'BIO')); %find index of 'BIO' reaction
in react_name
emsBIO = (emsB(:,find(emsB(bio_findModes,:)~=0))); %create matrix of em's
with non zero BIO term
bio_modes = size(emsBIO, 2)%determine # of biomass producing modes

%product & bio mode production %changed to tra1 = PRODUCT excretion.
PRODUCT_findModes = find(strcmp(react_name,'EXPORT')); %find index of
secreting rxn
emsPRODUCT = (emsB(:,find(emsB(PRODUCT_findModes,:)~=0))); %create matrix of
em's with non zero RXN term
PRODUCT_modes = size(emsPRODUCT,2) %determine # of product producing modes

%find both product secreting & bio producing mode matrices
PRODUCTBio_find = find(strcmp(react_name,'EXPORT')); %find index of secreting
rxn
emsPRODUCTBio = emsBIO(:,find(emsBIO(PRODUCT_findModes,:)~=0)); %create
matrix of em's with both non zero BIO & RXN terms
PRODUCT_BIO_modes = size(emsPRODUCTBio, 2) %determine # of product & biomass
producing modes

%Begin single KO simulation loops
k = 1; %initialize counter for next loop
FOR i = 1:length(delrxn)  %search delrxn
        i_delrxn = find(strcmp(react_name,delrxn(i))); % find index of delrxn
in react_name
        emsC  = emsB(:,find(emsB(i_delrxn,:)==0)); % find elementary modes in
emsB
        n_ems(k) = size(emsC,2); %create vector of number of EM's remaining
```

Figure 1.3: Commented pseudocode used for elementary mode sorting and analysis. This panel shows EM sorting for modes without flux through reactions in "singleDel", the output of elementary mode diagnostics terms for use in yield calculations, and initializing single knockout simulation loops.

```
     p = 1;          %initialize p for nested "for" loop below
     IF n_ems(k) ~= 0;%'For' loop searches all non-zero ems results in the
above 'for' loop for yields of the below metabolites

FOR p = 1:length(n_ems(k))%find yield of each em in each single deletion
          iE = find(strcmp(react_name,'IMPORT'));%find carbon source index
          iP = find(strcmp(react_name,'EXPORT'));%find string for product
          iS = find(strcmp(react_name,'BIO')); % find index for bio accum

           yield_P = emsC(iP,:)./emsC(iE,:); %create yield vector (P/R)
          max_yield(i) = max(yieldP_out(:));%output vtr of max Y per del
          min_yield(i) = min(yieldP_out(:));%output vtr of min Y per del
          yield_S = emsC(iS,:)./emsC(iE,:);%create Bio yield vector (S/R)
          max_yieldS(i) = max(yieldS_out);%output vtr of max Bio Y per del
          p = 1 + p; %counter for nested loop
          END
     END

     IF n_ems(k) == 0%error checking loop; prevents error if no EM's exist
         max_yield(i) = 0;
         max_yieldS(i) = 0;
         max_yieldA(i) = 0;
         min_yield(i) = 0;
     END
k=k+1; %counter for output n X 1 matrices

     IF ~length(i_delrxn)
         DISPLAY('Error: One of the reaction names in the delrxn is not in
react_name\n')
         DOEXIT
     END
   END

mode_frac(1:i) = n_ems(1:i)./size(emsB);%normalize EM remaining after dels
maxBio = max(max_yieldS(1:i)); %find max biomass yield for each deletion
Bio_yield(1:i) = max_yieldS(1:i)./maxBio; %calc normalized biomass yield

XLSOUTPUT(emsA, react_name, delrxn, mode_frac, max_yield, min_yield,
bio_yield, obj', singleDel, size(emsB), bio_modes, PRODUCT_modes,
PRODUCT_bio_modes) %output results to excel

    IF n_ems == 0
      DISPLAY('THERE IS NO ELEMENTARY FOR THE SPECIFIED OBJ and DELRXN\n')
      DOEXIT
    END
END
```

Figure 1.4: Commented pseudocode used for elementary mode sorting and analysis. This panel shows the calculation of max and min yield values for each single deletion in terms of the desired export/import reactions. This calculation is followed by data export to excel for further analysis.

This sorting algorithm functions by selecting rows (reactions) with zero or non-zero flux, and selecting either for or against mode removal based on that criteria. This is demonstrated in the following example:

	EM1	EM2	EM3	EM4	EM5	EM6	EM7	EM8
Rxn1	$\lambda_{(1,1)}$	$\lambda_{(1,2)}$	$\lambda_{(1,3)}$	$\lambda_{(1,4)}$	$\lambda_{(1,5)}$	$\lambda_{(1,6)}$	$\lambda_{(1,7)}$	$\lambda_{(1,8)}$
Rxn2	0	$\lambda_{(2,2)}$	$\lambda_{(2,3)}$	$\lambda_{(2,4)}$	$\lambda_{(2,5)}$	$\lambda_{(2,6)}$	$\lambda_{(2,7)}$	$\lambda_{(2,8)}$
Rxn3	$\lambda_{(3,1)}$	0	0	$\lambda_{(3,4)}$	$\lambda_{(3,5)}$	$\lambda_{(3,6)}$	$\lambda_{(3,7)}$	$\lambda_{(3,8)}$
Rxn 4	$\lambda_{(4,1)}$	0	$\lambda_{(4,3)}$	$\lambda_{(4,4)}$	$\lambda_{(4,5)}$	$\lambda_{(4,6)}$	$\lambda_{(4,7)}$	$\lambda_{(4,8)}$
Rxn 5	$\lambda_{(5,1)}$	$\lambda_{(5,2)}$	0	$\lambda_{(5,4)}$	$\lambda_{(5,5)}$	$\lambda_{(5,6)}$	$\lambda_{(5,7)}$	$\lambda_{(5,8)}$
Rxn 6	$\lambda_{(6,1)}$	$\lambda_{(6,2)}$	$\lambda_{(6,3)}$	0	$\lambda_{(6,5)}$	$\lambda_{(6,6)}$	$\lambda_{(6,7)}$	$\lambda_{(6,8)}$
Rxn 7	$\lambda_{(7,1)}$	$\lambda_{(7,2)}$	$\lambda_{(7,3)}$	$\lambda_{(7,4)}$	0	$\lambda_{(7,6)}$	$\lambda_{(7,7)}$	$\lambda_{(7,8)}$
Rxn 8	**$\lambda_{(8,1)}$**	**$\lambda_{(8,2)}$**	**$\lambda_{(8,3)}$**	**$\lambda_{(8,4)}$**	**$\lambda_{(8,5)}$**	**0**	**0**	**0**
Rxn 9	$\lambda_{(9,1)}$	$\lambda_{(9,2)}$	$\lambda_{(9,3)}$	$\lambda_{(9,4)}$	$\lambda_{(9,5)}$	$\lambda_{(9,6)}$	0	0
Rxn 10	$\lambda_{(10,1)}$	$\lambda_{(10,2)}$	$\lambda_{(10,3)}$	$\lambda_{(10,4)}$	$\lambda_{(10,5)}$	$\lambda_{(10,6)}$	$\lambda_{(10,7)}$	0

Where $\lambda \in \mathbb{R}, \neq 0$

Applying the condition to E, where E' = $E' \in E$, where the element in row 8 $\neq$ 0 removes all EM's (columns) with a zero flux observed in row (reaction) 8. This has the effect of removing any EM's that don't involve required reactions, such as carbon source import. This is often implemented to eliminate futile cycles from further analysis, but may also be used to limit carbon source or oxygen availability.

$$E' = \begin{vmatrix} \lambda_{(1,1)} & \lambda_{(1,2)} & \lambda_{(1,3)} & \lambda_{(1,4)} & \lambda_{(1,5)} \\ 0 & \lambda_{(2,2)} & \lambda_{(2,3)} & \lambda_{(2,4)} & \lambda_{(2,5)} \\ \boldsymbol{\lambda_{(3,1)}} & \mathbf{0} & \mathbf{0} & \boldsymbol{\lambda_{(3,4)}} & \boldsymbol{\lambda_{(3,5)}} \\ \lambda_{(4,1)} & 0 & \lambda_{(4,3)} & \lambda_{(4,4)} & \lambda_{(4,5)} \\ \lambda_{(5,1)} & \lambda_{(5,2)} & 0 & \lambda_{(5,4)} & \lambda_{(5,5)} \\ \lambda_{(6,1)} & \lambda_{(6,2)} & \lambda_{(6,3)} & 0 & \lambda_{(6,5)} \\ \lambda_{(7,1)} & \lambda_{(7,2)} & \lambda_{(7,3)} & \lambda_{(7,4)} & 0 \\ \lambda_{(8,1)} & \lambda_{(8,2)} & \lambda_{(8,3)} & \lambda_{(8,4)} & \lambda_{(8,5)} \\ \lambda_{(9,1)} & \lambda_{(9,2)} & \lambda_{(9,3)} & \lambda_{(9,4)} & \lambda_{(9,5)} \\ \lambda_{(10,1)} & \lambda_{(10,2)} & \lambda_{(10,3)} & \lambda_{(10,4)} & \lambda_{(10,5)} \end{vmatrix}$$

Now, for further matrix sorting, we can simulate reaction deletion (through mutation, regulation, or metabolite availability based on culture conditions) by limiting E'' to contain EM's where the elements in row i are zero (E'' = $E^{'} \in E,$ where element in row $3 = 0$) This eliminates all EM's with flux through the reaction described by row i, thus reducing the solution space (E'') to contain only the mode(s) possible in the absence of that reaction from the network.

$$E'' = \begin{vmatrix} \lambda_{(1,2)} & \lambda_{(1,3)} \\ \lambda_{(2,2)} & \lambda_{(2,3)} \\ 0 & 0 \\ 0 & \lambda_{(4,3)} \\ \boldsymbol{\lambda_{(5,2)}} & \mathbf{0} \\ \lambda_{(6,2)} & \lambda_{(6,3)} \\ \lambda_{(7,2)} & \lambda_{(7,3)} \\ \lambda_{(8,2)} & \lambda_{(8,3)} \\ \lambda_{(9,2)} & \lambda_{(9,3)} \\ \lambda_{(10,2)} & \lambda_{(10,3)} \end{vmatrix} \quad E''' = \begin{vmatrix} \lambda_{(1,2)} \\ \lambda_{(2,2)} \\ 0 \\ 0 \\ \lambda_{(5,2)} \\ \lambda_{(6,2)} \\ \lambda_{(7,2)} \\ \lambda_{(8,2)} \\ \lambda_{(9,2)} \\ \lambda_{(10,2)} \end{vmatrix}$$

E'' could be reduced further by repeating the step above by targeting a different reaction, in this case removing modes with flux through row 4 (reaction 4). This finally results in a single elementary mode that can describe the multiple knock-out system solution space,

following the application of one required reaction (to give E'), and two gene (reaction/row) deletions, finally leaving E'''.

1.3: Results

1.3.1: Model construction

The preliminary model shown in Figure 1.1 constitutes the starting point of the following thesis sections describing *S. oneidensis* MR-1 central metabolism. As a result, specific differences between this model and that of the current *S. oneidensis* model have been focused on in detail.

A reaction was added which reduces fumarate to succinate, regenerating oxidized quinones. This reaction is facilitated by one enzyme, the periplasmic protein Fcc_3, which is encoded by *fccA* (21, 22). This enzyme reduces external fumarate to succinate by oxidizing one quinol (transferring 2 electrons and 2 hydrogen atoms) per fumarate reduced (22).

The current model of this system describes the electron transfer pathway in the following order: cytoplasmic cofactors (NADH, NADPH) transfer electrons to menaquinone located in the inner membrane, which then transfers electrons from menaquinone to external sinks via the DMRC, which is essential for reduction of external electron acceptors by *S. oneidensis* (23, 24). The system was modeled assuming that all quinones are interchangeable throughout the central metabolism, and are equivalent in reduction potential to other quinones, yet lower in reduction potential than NAD(P)H. This necessarily requires that the enzyme encoded by *ndh*, which reduces quinone by oxidizing NADH, must function irreversibly under normal cell conditions. Furthermore,

based on Ndh activity described in *E. coli*, activity of the enzyme encoded by *ndh* was assumed to generate no electrochemical gradient (25-27).

In addition, this model was designed considering the possibility that this electron transfer from quinones to external electron sinks (metals or electrodes) may result in the translocation of hydrogen ions across the cell membrane. However, this translocation has been treated as one that does not result in a proton gradient that drives ATP synthesis. As a result, the most conservative estimates of energy conservation and yield have been utilized in developing the current model, where the regeneration of oxidized cofactors in anaerobic fermentation does not result in ATP production via oxidative phosphorylation in anaerobic cells (i.e., ATP fixation occurs by substrate-level phosphorylation only). This assumption is supported by the observed inability of *S. oneidensis* MR-1 to grow anaerobically on lactate when *ackA*[4] or *pta* have been deleted (28).The gene *ackA* encodes acetate kinase, which converts acetyl phosphate to acetate, producing one ATP from substrate-level phosphorylation, while Pta phosphorylates acetyl-coA to acetylphosphate (29). Because cells of these genotypes are incapable of normal anaerobic growth on lactate in the presence of external fumarate as an electron acceptor, it is implied that this reaction is the only source of ATP generation available to the cell during anaerobic lactate metabolism, meaning that there can be no oxidative phosphorylation occurring during anaerobic cofactor regeneration. As a result, all anaerobic reactions which may result in H^+ gradient production have been assumed to result in no ATP

[4] It has also been suggested that the *Δpta* strain suffers from the toxic buildup of acetyl-phosphate, resulting in the observed lethal phenotype when cultivated anaerobically on lactate.

production, and an H^+ balance has not been attempted in this model. The primary consequences of this assumption are that the NADH dehydrogenase enzymes *nuo* and *ndh* are treated as interchangeable, and formate oxidation is treated as resulting in no net ATP generation or other energy synthesis (i.e, all NADH dehydrogenases have been treated as Type II dehydrogenases) (26). Type II dehydrogenases are defined as dehydrogenases whose activity does not produce a membrane potential, usually in the form of an H^+ gradient, whereas type I dehydrogenases produce a membrane potential in this way (26). The potential consequences of these simplifications will be discussed in section 1.4.

In addition, because no annotated gene has been identified to catalyze the reaction converting pyruvate to acetate (encoded by *poxB* in *E. coli*), and the observed elimination of acetate secretion in anaerobic *S. oneidensis* MR-1 *ΔackA* when fed lactate, this reaction was not included in the updated model describing *S. oneidensis* MR-1 central metabolism. Furthermore, the uptake pathways for N-acetylglucosamine, acetate, and lactate were added to this model. Lactate secretion, which has not been observed previously in *S. oneidensis* MR-1, has been removed by making lactate import irreversible. The conversion of lactate to pyruvate is also treated as unidirectional in *S. oneidensis* MR-1, and catalyzed by the enzymes and transport proteins *lldEGF; dld-II* , in contrast to *E. coli*, where it has been treated as reversible, with lactate being a common secretion product under anaerobic conditions (31, 14).

Finally, the serine-glycine pathway was inserted based on the carbon flux information published by Tang (2007), where 2-9% of carbon when cultivated on lactate under

various conditions was diverted through the serine-glycine pathway. This level of carbon flux makes the serine-glycine pathway the auxiliary pathway with the highest flux, where ¾ of the flux from phosphoenolpyruvate (PEP) was directed into the serine-glycine biosynthesis pathway(6). Due to this level of flux, far in excess of other included auxiliary/biomass producing pathways, such as the Entner-Doudoroff and PPP pathways, it was determined that this pathway must be included to accurately describe *S. oneidensis* metabolism. In addition, due to the reported ability of S. *oneidensis* to fix formate via the combination of glycine with tetrahydrofolate (THF) to create serine, the inclusion of this pathway is required to accurately describe the central metabolism of this species as it is currently understood (6, 32). The reaction stoichiometry for the serine-glycine pathway, as well as the net stoichiometry of THF-formate-serine interconversion were both obtained from the publicly available database Metacyc (29). Due to the insertion of this pathway, the biomass term, which describes the carbon and energy requirement to accumulate 1 gram of dry cell mass, was recalculated to include terms for serine and glycine based on the published cellular composition requirements for *E. coli*, described in detail in section 1.2.2. The final result of the implementation of these changes was incorporated into the creation of the metabolic map described in Figure 1.5. This model has been limited to describe the utilization of glycerol, NAG, lactate, and acetate as carbon sources, allowing cofactor regeneration by aerobic growth, anaerobic fumarate reduction, and via anaerobic DMRC activity. Furthermore, secretion of pyruvate, succinate, acetate, and carbon dioxide have been accommodated. This wild-type model contains 65 reactions, 20 of which are reversible, with 49 internal metabolites and 14 external metabolites. Simple changes, such as omission of carbon import pathways,

simulate growth on additional/alternative/single carbon sources, and allow rapid model adaptation to specific growth conditions as needed.

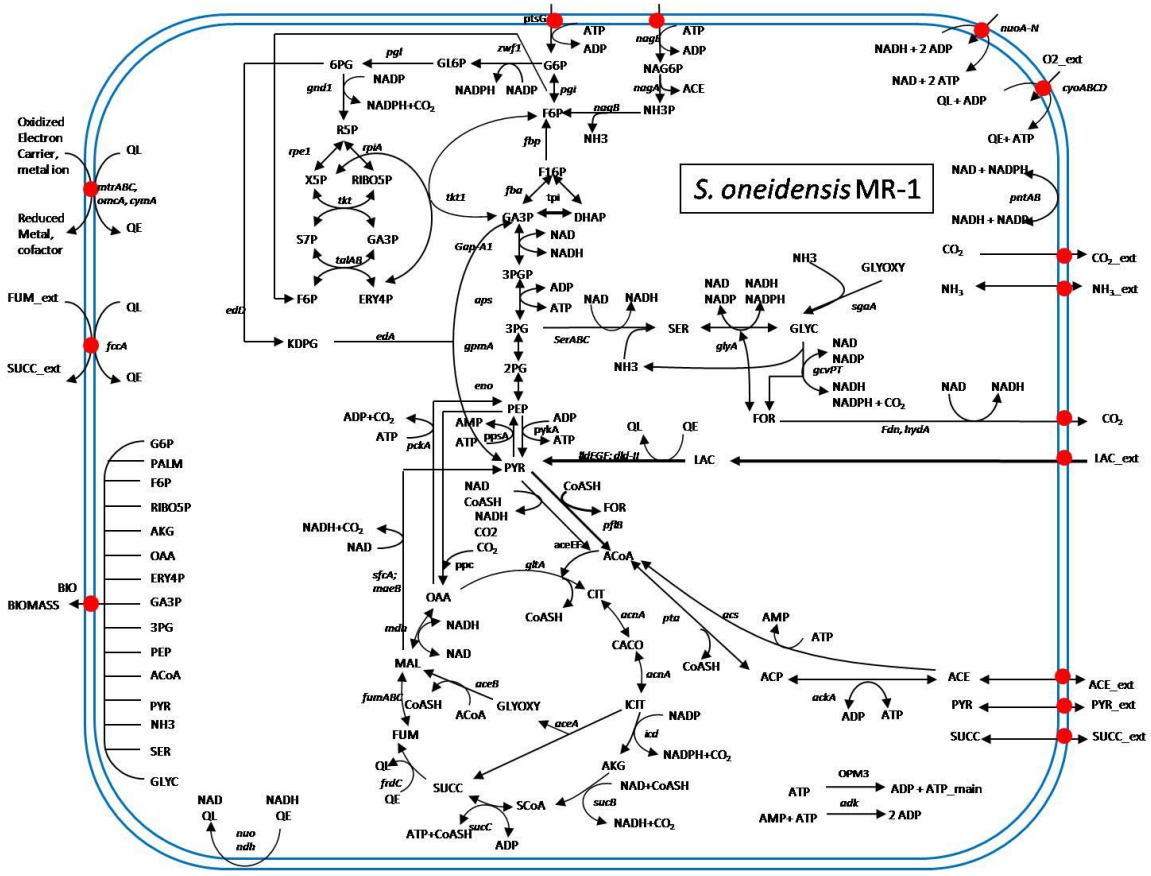


Figure 1.5: Metabolic map describing the *S. oneidensis* MR-1 central metabolism and auxiliary pathways. Reactions/reactants that differ from those in the preliminary model have been highlighted in blue, while reactions with external metabolites are signified by red dots (●).

1.3.2: Model verification:

This model was verified by comparing model predictions with viability phenotypes observed in single knockout strains under various growth conditions, the simulation of which is summarized in Table 1.1. For this analysis, samples were cultivated under anaerobic conditions in the presence of exogenous fumarate as an electron acceptor. As shown in Table 1.2, all available knockout viability phenotypes were successfully simulated using this model, when applying regulation and nutrient

conditions described in Section 1.2.4. Mutant strain creation and characterization was performed by K. Hunt and J. Gralnick (19).

Table 1.2: Model predictions compared with experimental results of single knockout mutants

	Carbon source							
	lactate				N-acetyl Gluccosamine			
	anaerobic		aerobic		anaerobic		aerobic	
Gene knockout	Experi-ment	Predic-tion	Experi-ment	Predic-tion	Experi-ment	Predic-tion	Exper iment	Pred-iction
Δgnd1	+	+	+	+	+	+	+	+
ΔackA	-	-	+	+	+	+	+	+
ΔldhA	+	+	+	+	+	+	+	+
ΔaceE	+	+	-	-	+	+	-	-
Δpta	-	-	+	+	+	+	+	+
ΔpckA	+	+	+	+	+	+	+	+
ΔsfcA	+	+	+	+	+	+	+	+
ΔaceA	+	+	+	+	+	+	+	+
ΔpykA	+	+	+	+	+	+	+	+
ΔpflB	-	-	+	+	+	+	+	+
ΔgcvT	+	+	+	+	+	+	+	+
Δacs	+	+	+	+	+	+	+	+
ΔackA; Δpta	-	-	+	+	+	+	+	+
Δppc	nd	-	-	-	nd	-	-	-

1.3.3: Additional knockout target genes for model verification:

The method of elementary mode analysis described in section 1.2.3 and applied using the algorithm presented in section 1.2.6 allows fast and easy identification of additional conditionally-lethal gene deletions that may be tested experimentally through the creation of single reaction deletion strains[5]. As a result, a list of gene deletions that

[5] Single *reaction* deletions often requires single *gene* deletion. However, due to both metabolic redundancy and protein complex formation, effective *reaction* deletion may require multiple *gene* deletions. An

would test these predictions is presented in their entirety in Table 1.3 for all simulated conditions.

Table 1.3: All predicted lethal single deletion *S. oneidensis* mutants under conditions tested

lactate		NAG		acetate
aerobic	anaerobic	aerobic	anaerobic	acetate aerobic
ΔaceE	*Δacs*	*ΔnagE*	*ΔacnA*	*ΔaceA*
ΔacnA	*Δadk*	*ΔnagA*	*ΔfccA*	*ΔaceB*
Δacs	*Δfba*	*ΔnagB*	*ΔfdnG; Δ hydA*	*ΔacnA*
Δadk	*Δfbp*	*Δzwf1*	*ΔgltA*	*Δacs*
Δfba	*ΔfccA*	*Δpgi*	*ΔglyA*	*Δadk*
Δfbp	*ΔfdnG; ΔhydA*	*ΔgltA*	*Δicd*	*Δfba*
ΔfdnG; Δ hydA	*Δgap-A1*	*Δicd*	*ΔnagA*	*Δfbp*
Δgap-A1	*ΔgltA*	*ΔaceE*	*ΔnagB*	*ΔfrdC*
ΔgltA	*ΔglyA*	*Δpgi*	*ΔnagE*	*ΔfumABC*
ΔglyA	*ΔgpmA*	*ΔrpiA*	*Δndh*	*Δgap-A1*
ΔgpmA	*Δicd*	*ΔacnA*	*ΔpflB*	*ΔgltA*
Δicd	*Δlld, Δdld*		*Δpgi*	*ΔgpmA*
Δlld, Δ dld	*Δndh*		*Δpgi*	*Δicd*
Δpgi	*Δpfl*		*Δppc*	*Δmdh*
Δpgk	*Δpgi*		*ΔrpiA*	*Δnuo*
Δppc	*Δpgk*		*ΔserA*	*Δpgi*
ΔpykA	*Δppc*		*Δzwf1*	*Δpgk*
ΔrpiA	*ΔpykA*			*ΔpykA*
ΔserA	*ΔrpiA*			*ΔrpiA*
Δtpi	*ΔserA*			*Δtpi*
	Δtpi			

To further facilitate verification, this list has been reduced into Table 1.4 to include only predicted lethal single gene deletions expected to be capable of growth under at least one growth condition. This list has been limited to conditionally- lethal gene deletions, as

example of this is *ΔlldEGF*, *Δdld-II*, which are needed for import of L and D stereoisomers, respectively. Both deletions are expected to be necessary for effective elimination of lactate import.

opposed to all expected lethal mutations, because at least one "growth" phenotype is needed to verify the existence of any mutant strain, versus the possibility of simply an unsuccessful transformation during knockout generation.

Table 1.4: Potential gene knockouts for model validation

	lactate consumption		NAG consumption		Acetate consumption
mutation	aerobic growth	anaerobic growth	aerobic growth	anaerobic growth	aerobic growth
ΔaceE	-	+	-	+	-
ΔlldEGF Δdld-II	-	-	+	+	+
ΔfdnG, ΔhydA	+	-	+	-	+
Δndh; Δnuo	+	-	+	-	+
ΔpflB	+	+	+	+	+
ΔglpX	-	-	+	+	-
ΔaceA	+	+	+	+	-
Δacs	+	+	+	+	-
Δzwf1	+	+	-	-	+
Δppc	-	-	-	-	+
ΔserA	+	-	+	-	+
ΔglyA	+	-	+	-	+

1.4: Discussion

Elementary mode analysis has been a useful tool in both characterizing previous metabolic networks, as well as directing gene mutations for rational organism design. The elementary mode model presented here represents a starting point for the design of derivative strains, and can quickly be adapted to represent growth on a wide variety of substrates and the production of a wide variety of products. One example of this is presented in section 1.5.1, where this wild-type network was adapted to model glycerol

consumption and ethanol secretion by adding two metabolic pathways to the wild-type network described above.

The results described in Table 1.3 highlight one caveat of this, as well as any model: predictions that apply to a system may not apply to a derivative, or changed, system. This is exemplified in the predicted lethality of the *ΔfdnG, ΔhydA* mutations. This model (shown in Figure 1.5) lacks formate secretion mechanisms, because formate secretion in the wild-type organism has not been observed in any previous experiments. It is may be the case, however, that this organism is capable of producing biomass in the absence of formate oxidation mechanisms, or that formate secretion in the absence of formate oxidation mechanisms will occur. It is very likely that highly active formate oxidation enzymes simply oxidize formate before it can be secreted in measureable quantities in wild-type strains, and that in the absence of the necessary enzymes, formate secretion will occur.

Future work on this topic should focus on three primary subjects, illustrated in Figure 1.6. First, model predictions should be tested by creating the single reaction knockout mutants described in Table 4, most notably the Δ*fdh*, Δ*glyA*, and Δ*serA* mutants. The results of these knockouts would not only indicate network accuracy, but would also be useful resources in testing previously proposed mechanisms of formate & serine assimilation via the serine-glycine pathway. Nealson (1994) claimed to observe anaerobic biomass accumulation with formate as a sole carbon source, while proposing a mechanism of assimilation via reverse flux through the serine synthetic pathway, producing 3-phosphoglycerate (20). No modes could be found which assimilated formate

into biomass via this proposed route under anaerobic conditions using elementary mode analysis. However, if such an anaerobic formate assimilation pathway exists in *S. oneidensis*, it is also possible that this pathway proceeds from formate to serine as stated before, but then directly catabolize serine to pyruvate and ammonia, entering the Embden-Myerhoff pathway via serine deaminase activity (Sda), to produce pyruvate and ammonia (*sdaA*, geneID 1169979). As a result, the growth characteristics of a Δ*serA* mutant would strengthen confidence in the proposed pathways of formate fixation and serine degradation reaction in *S. oneidensis* by eliminating one possible route from serine to the glycolytic pathway.

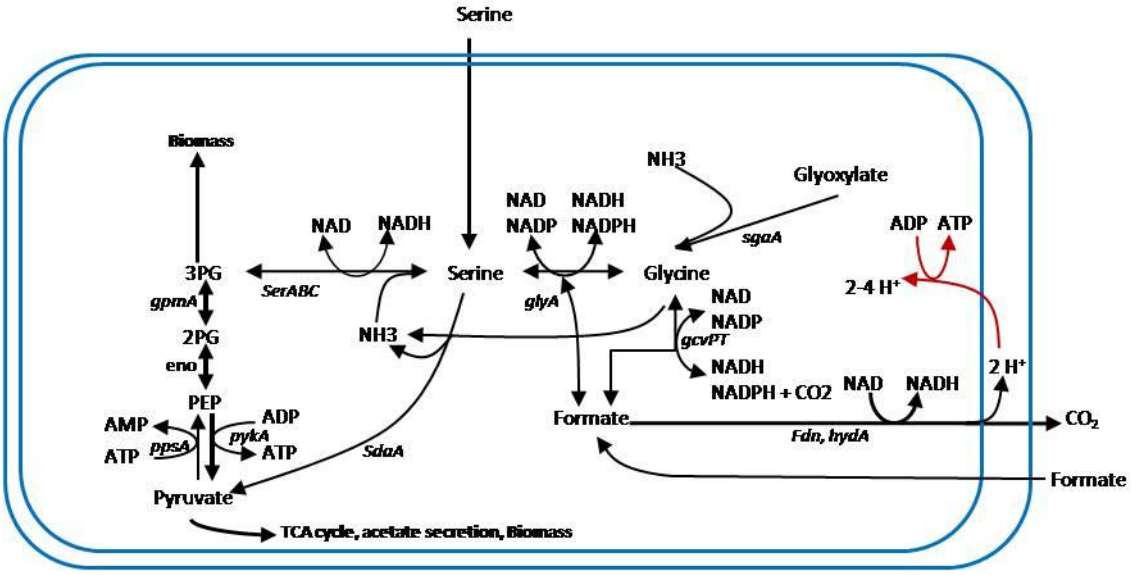


Figure 1.6: Serine-glycine superpathway and proposed formate assimilation and oxidation reactions. ATP generation from membrane potential has been highlighted in red, as the presence of this reaction under anaerobic conditions has not been confirmed. Serine/formate assimilation have been proposed to occur via reverse flux through reactions catalyzed by *serABC* related enzymes, while a competing reaction for serine catabolism exists via serine deaminase, which converts serine directly to pyruvate, eliminating both NAD and ATP generated from *serABC* reverse flux.

Second, this model also assumed that no ATP accumulation occurs as a result of formate oxidation under anaerobic conditions, as described in section 1.3.1. By this assumption, we would expect little or no reduction in biomass accumulation in *ΔfdnG, ΔhydA* mutants when grown on NAG or lactate. In addition, such strains would be expected to secrete formate if the assumption that formate does not generate ATP via proton gradient generation is indeed correct. If true, this would raise the question of why *S. oneidensis* MR-1 would maintain high formate dehydrogenase activity, if no net energy or carbon benefit occurs from its oxidation (20). Because it is suspect that the cell would expend energy to oxidize formate without benefit, and the necessity of Pta activity to grow on anaerobically on lactate, it follows that formate oxidation may produce a proton motive force, but this force may not or cannot be used to directly generate ATP anaerobically. Nevertheless, incomplete understanding of reaction pathways, as well as ambiguous data regarding reaction stoichiometry, have required conservative energy production assumptions that ignore any effects of proton gradient generation in formate metabolism. It appears that, if *S. oneidensis* can in fact grow solely on formate anaerobically, then it must be capable of obtaining usable energy from formate oxidation, and the only known means of such energy generation occur through proton gradient-generating formate dehydrogenase activity. This issue will be addressed as new information becomes available.

Third, because the proposed mechanism for formate (or formaldehyde) assimilation in *S. oneidensis* requires the reversible conversion of glycine to serine to incorporate C1-units, a Δ*glyA* strain would directly test this hypothesis, as well as the existence of serine-

glyoxylate aminotransferase (*sgaA)*, which has been proposed to convert glyoxylate to glycine through serine deamination via serine deaminase activity (8, 32). A Δ*glyA* mutant would be expected to be incapable of biomass accumulation under anaerobic conditions, if the formate incorporation pathway via the serine-glycine pathway is correct.

Finally, this model required the assumption that quinones (menaquinone, ubiquinone) be both equal in reduction potential and interchangeable within reaction and cell states. However, studies regarding anaerobic DMRC activity have shown that this reaction occurs with menaquinone, and no other quinone reaction has been reported (33). It is unclear at the current time, however, if this is significant. For example, if ubiquitous enzymes exist to transfer electrons between quinones and menaquinone, or if all quinone reactions under anaerobic conditions involve menaquinone, a distinction between quinones would not affect the network outcome. Currently, no evidence could be found to indicate that this assumption affects EM composition detrimentally.

1.5.1 Model implementation: anaerobic conversion of glycerol to ethanol in the presence of an electrode.

The *S. oneidensis* wild-type model has been further specialized to describe a derivative strain that contains two plasmids, pGUT2 and pPET, developed by J. Flynn and J. Gralnick (34). The former plasmid allows *S. oneidensis* to import glycerol by inserting the genes *glpF*, *glpK*, *glpD*, and *tpiA*, while the latter facilitates conversion of pyruvate to ethanol conversion in *S. oneidensis* through the insertion of the genes *pdc* and *adhE*. A detailed description of this system is described in correlated works (34). The development of this strain, and its efficient production of ethanol from glycerol, has great

potential to both the development of microbial fuel cells (this conversion liberates 2 electrons per glycerol, creating an electric current), as well as the biofuels industry, by converting waste glycerol to the biofuel ethanol, a conversion not possible in other anaerobic systems, being electronically unbalanced, and typically requiring mixed-acid fermentation.

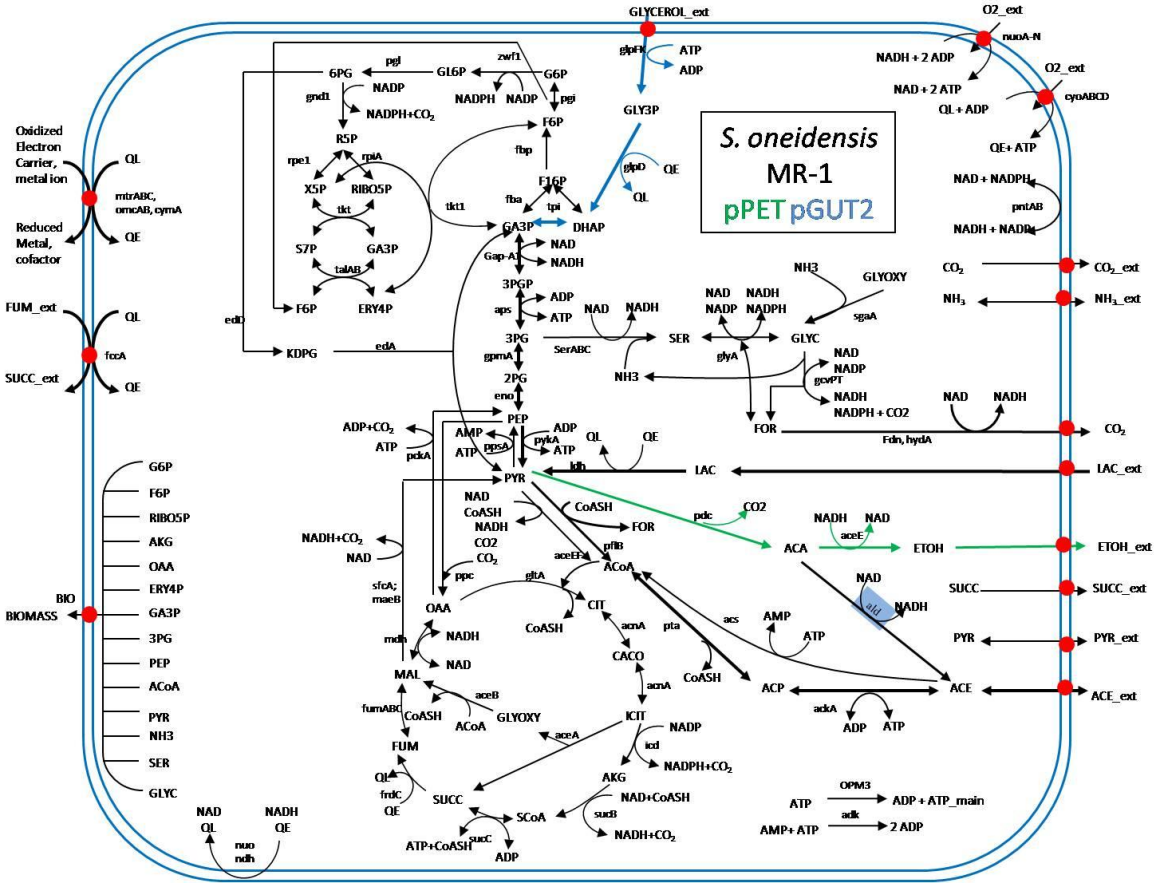


Figure 1.7: Metabolic map of *S. oneidensis* MR-1 pGUT2 pPET. This figure shows the central metabolism from glycerol import through ethanol secretion, including DMRC remain uninhibited by the suggested deletion set, while branch pathways and the secretion of undesired fermentation acids have been eliminated. Note the addition of the reaction catalyzed by Ald, which converts acetaldehyde (ACA) to acetate (ACE). External reactions are highlighted with red dots (●), while reactions different from the wild-type model have been highlighted in blue, and inserted reactions are shown with blue (pGUT2) or green (pPET) reaction arrows, as determined by the plasmid conferring each respective reaction capacity.

The central metabolism of *S. oneidensis* pGUT2 pPET was modeled implementing elementary mode analysis, using the publicly available free software Metatool v5.1 in the Matlab environment (MathWorks, Inc., Natick, MA) (13). The central metabolism of *S. oneidensis* was described using 64 reactions, 20 of which were reversible, and with 60 metabolites, 11 of which were external. Anaerobic DMRC active conditions were simulated by omitting reactions involving oxygen as an external metabolite. Glycerol consuming conditions were simulated by omitting lactate, acetate, and NAG import reactions. Finally, the reaction describing the reduction of external fumarate to succinate was also removed, as external fumarate will not be supplied in anaerobic conditions requiring DMRC activity. Target genes for deletion were identified using the semi-iterative method described in section 1.2.5.

Because no elementary modes involving the secretion of acetate achieve maximum yield of ethanol, the acetate producing pathway (corresponding to the Δ*pta* mutation) was deleted prior to the implementation of the aforementioned iterative deletion method.

1.5.2: Elementary mode simulation results

When cultivated under anaerobic conditions on glycerol, with an electrode as the external electron acceptor, *S. oneidensis* Δ*pta* Δ*ald* Δ*gcvT* pGUT2 pPET describes the optimal knockout strain for this application. The progression of gene deletions, as well as maximum and minimum ethanol yield at each point, is summarized in Table 1.4.

Table 1.4: The Effect of Gene Deletion on Elementary Mode Number in *S. oneidensis*

	total modes	biomass produ-cing modes	ETOH secretin g modes	ETOH secreting & Biomass producing modes	maximum ETOH yield (# modes)	minimu m ETOH yield (# modes)
wild type, glycerol growth	85412	82670	8070	7886	1 (76)	0 (77342)
anaerobic, DMRC growth	378	260	84	52	1 (28)	0 (294)
fcc3 inactive (fumarate -)	189	130	42	26	1 (14)	0 (147)
Δpta	77	52	38	26	1 (10)	0 (39)
Δpta, Δald	37	26	36	26	1 (8)	0 (1)
Δpta, Δald, ΔgcvPT	28	18	28	18	1 (8)	0.4668 (2)
Δpta, Δald, ΔgcvPT, Δzwf1	8	2	8	2	1 (1)	0.4668 (2)
Δpta, Δald, ΔgcvPT, Δgnd1	11	3	11	3	1 (1)	0.4668 (2)

The central metabolism of this optimized strain can be described by only 28 elementary modes, of which all produce ethanol and 18 produce biomass and ethanol, in contrast to the wild type strain, which can be described by 85412 total modes, of which only 8070 (9.4%) produce ethanol. Finally, yield of ethanol from this multiple-knockout strain was predicted to be between 0.46 moles ethanol/moles glycerol (0.23 g ETOH/g glycerol) and 1 mole ethanol/mole glycerol (0.5 g ETOH/g glycerol) under the described conditions, while releasing between two and eight electrons per glycerol to the exterior of the cell,

with two electrons being exported in those modes that produce ethanol at theoretical yield, as expected. This is also presented graphically in Figure 1.7.

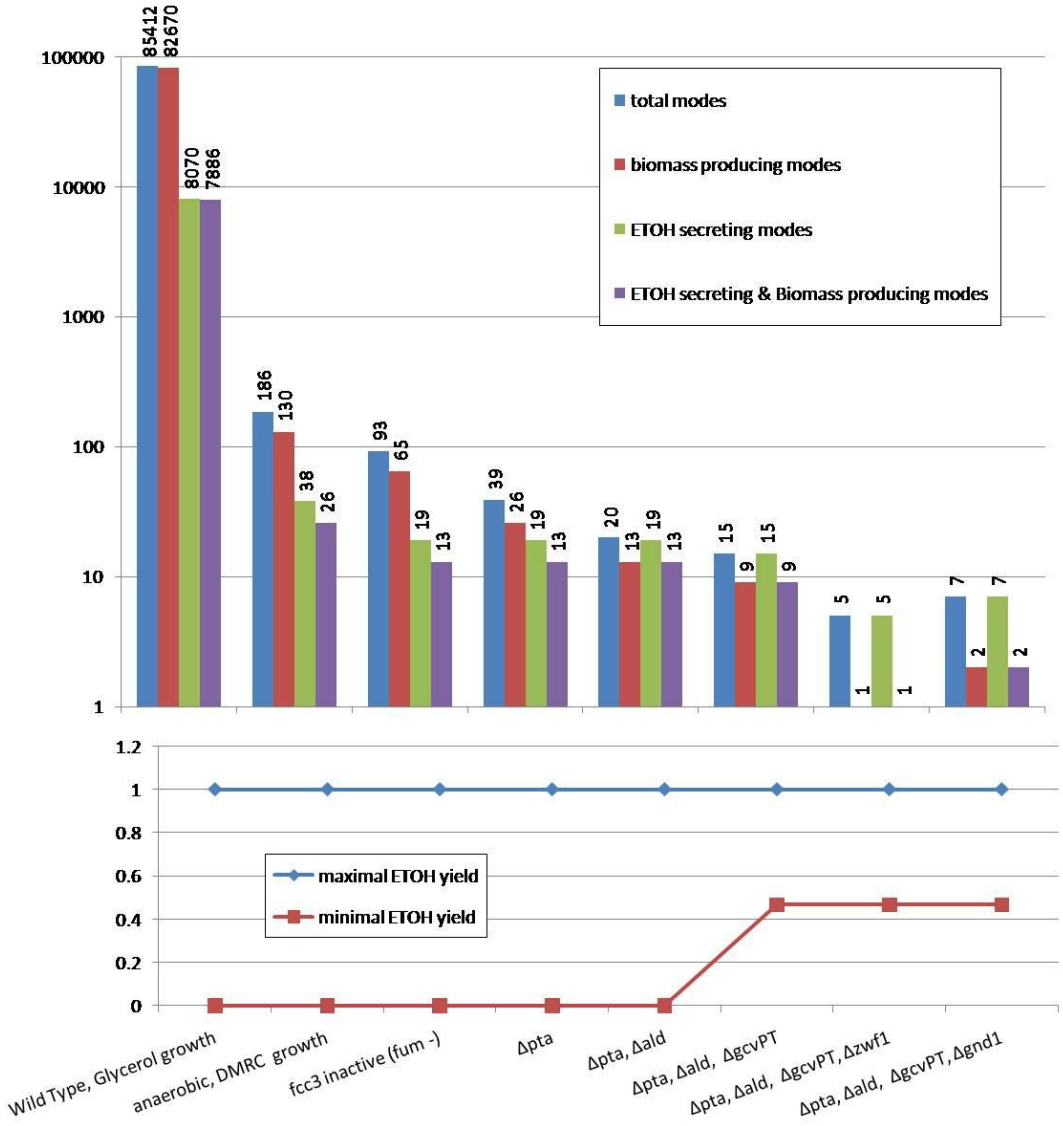


Figure 1.8: Effect of reaction deletions on the number of anaerobic EMs for growth on glycerol. In each group of bars, the number of (i) total EMs, (ii) biomass producing modes, (iii) ethanol secreting modes, and (iv) ethanol and biomass producing modes. Note the increase in minimum ethanol yield with *gcvPT* deletion, which together with *ald* and *pta* deletions, forces all carbon flux through ethanol secreting modes.

1.5.3: Discussion

Surprisingly, the optimal ethanol producing strain does not require the removal of any portion of the pentose phosphate pathway (PPP). This is in contrast to the results predicted earlier when using the fully iterative approach to selecting knockouts in this and other related systems(4). When the PPP is uninhibited, the range of possible ethanol yields is unaffected, as the active PPP results in a large number of modes with intermediate yields. In the living cell, however, a cell with an entirely inactivated forward PPP (Δ*zwf1*) or partially inactivated PPP (Δ*gnd1*), which allows growth on NAG with flux the Entner-Doudoroff pathway[6], may exhibit increased ethanol yields when compared with the PPP active, quadruple knockout described above, if carbon flux through the PPP pathway is significant in the wild-type cell. This apparent contradiction is due to the principle of elementary modes that, under any given condition, the metabolism of the organism being studied can be described by the weighted combination of all possible elementary modes (Eq. 4, Section 1.1.2). While the calculation of the weights assigned to any individual mode is nearly impossible, it is possible to infer the weight on mode families in this case based on measured flux through a particular pathway. Previously performed flux analysis, based on growth on lactate, indicates that the PPP is only active in biomass accumulation, and as a result exhibits a minor metabolic flux (~0.5% of total carbon), while the Entner-Doudoroff pathway exhibited slightly higher flux (0.7-1.1% of total carbon) (6). However, these values have the potential to change when cultivated on glycerol or NAG. In an experiment performed by

[6] Because *S. oneidensis* lacks 6-phosphofructokinase II function (required for conversion of β-D-glucose-6-phosphate to β-D-fructose-6-phosphate), NAG consumption requires carbon flux to proceed through PPP and/or the Entdner-Doudoroff pathways in *S. oneidensis*.

Kris Hunt, *S. oneidensis Δgnd1* pPET was cultivated anaerobically on NAG, and which showed that ethanol yield was observed to not be significantly different than the wild type strain (19). This indicates that the flux through the PPP is indeed minor under several conditions, as expected based on independently performed carbon flux measurements, and likely does not contribute significantly to overall metabolite flux (6). A detailed analysis has yet to be performed on *S. oneidensis Δgnd1* pGUT2 to directly determine differences in yield resulting from forward flux through the PPP when consuming glycerol anaerobically. However, there is no reason to expect any increase in PPP flux when cultivated on glycerol vs. lactate, as both conditions would require reverse flux from glyceraldehyde-3-phosphate and pyruvate to pass through the PPP, a path unlikely to be taken except in biomass accumulation, as observed in lactate-based growth conditions (6). This further reinforces the conclusion that *Δgnd* or Δ*zwf1* knockouts are highly unlikely to affect the ethanol yield of the resulting strain.

In addition, aldehyde dehydrogenase deletion (Ald, genotype *ΔaldA*) is expected to result in yield improvements, as the removal of the reaction it catalyzes eliminates the conversion of acetaldehyde to acetate. Because the inserted ethanol production pathway involves acetaldehyde production, and its subsequent conversion to ethanol, Ald competes directly for substrate with the ethanol secretion pathway. The presence of this reaction results in many elementary modes without any ethanol production, and is expected to result in significant yield reduction if present in the cell.

Finally, the removal of the glycine cleavage complex (Gcv) is expected to improve ethanol yield in *S. oneidensis* MR-1. The glycine cleavage complex is encoded by three

genes, *gcvT* and *lpdA*. *lpdA* has been reported to be active in other enzymatic reactions, so the removal of either or both *gcvPT* genes is expected to successfully inhibit glycine cleavage without affecting other known reactions *in vivo* (29, 30). This reaction has been targeted for deletion based on two factors: carbon flux through this reaction, and expected yield range based on elementary mode analysis. In contrast to the earlier discussed PP pathway, a significant carbon flux has been observed through the serine-glycine pathway, ranging from 2% to 9.1% of total lactate flux, with flux through GCV ranging from 1.1% to 5.6% depending on nutrient conditions (6). Secondly, elementary mode analysis predicts several (11 of 38, or 28% at point of deletion) elementary modes proceeding through the Gcv, with all resulting in eliminated or reduced ethanol yield below 0.4668 moles ethanol per mole glycerol. By elementary mode analysis, the minimum ethanol yield of the metabolic network is raised with the removal of the reaction catalyzed by Gcv, from 0 to 0.4668, as a mole fraction. As a result of these two factors, both a significant carbon flux, and improvement in minimum predicted ethanol yield, it is likely that the *ΔgcvT* genotype strain will result in an improvement in ethanol yield.

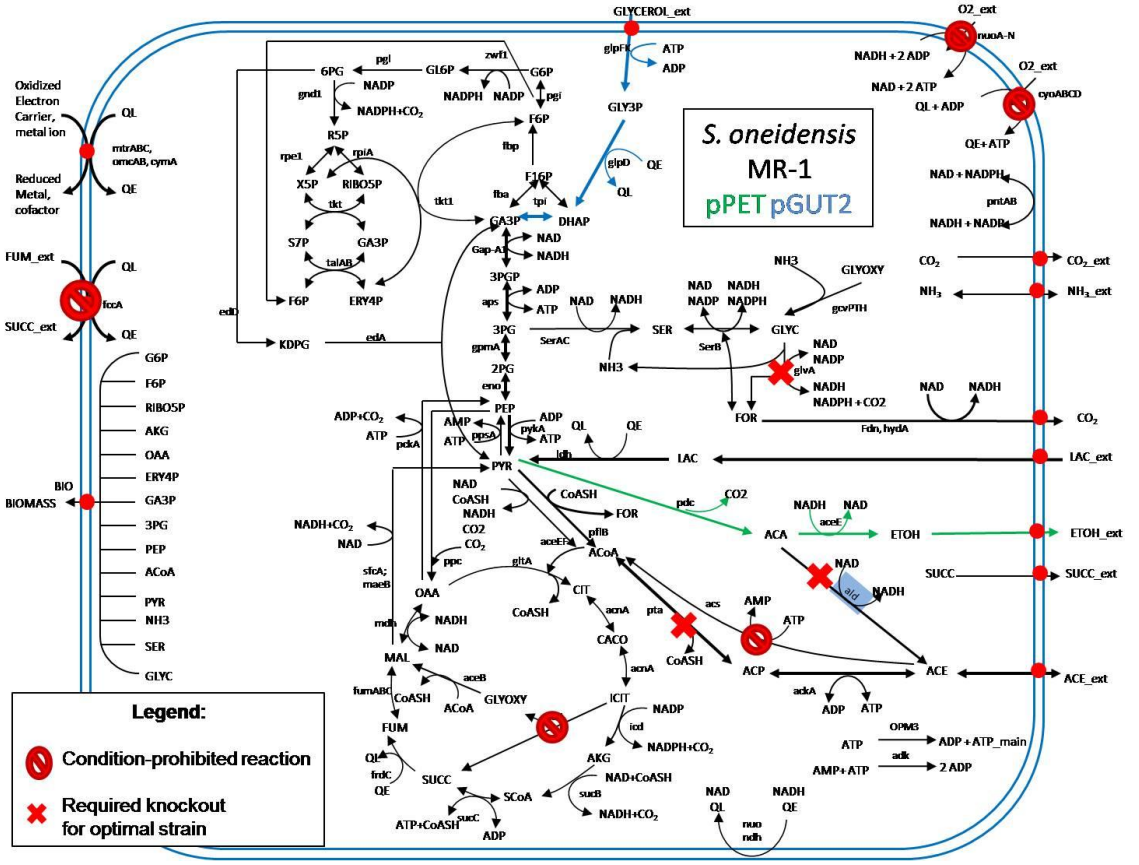


Figure 1.9: Metabolic map showing knockout reactions and conditionally-prohibited reactions expected to result in the optimal glycerol consuming strain which secreted ethanol. Major pathways remaining after deletion have been emboldened. This figure shows the central metabolism from glycerol import through ethanol secretion and the DMRC, remain uninhibited by the suggested deletion set, while branch pathways and the secretion of undesired fermentation acids have been eliminated.

The optimal ethanol producing *S. oneidensis* pGUT2 pPET strain would contain the following minimum set of gene deletions: *Δpta ΔaldA ΔgcvPT,* the combination of which prevents the secretion of acetate and the degradation of glycine. As a result, major carbon flux is predicted to be directed toward the production of ethanol, as shown in Figure 1.8. Finally, due to the prevention of acetate secretion, the cell is expected to be capable of accumulating biomass anaerobically only when transformed with pPET, which facilitates the production of ethanol, as observed in similar multiple knockout *E. coli* strains optimized to produce ethanol by coupling ethanol secretion to biomass formation (4).

Future work should focus primarily on producing the prescribed knockout strains, as well as generating a single plasmid containing pPET and pGUT2 genes, to aid in strain stability and reduce antibiotic requirements. Once performed, single knockout strains can be transformed/cultivated under required conditions to test the yield predictions made by this model.

Chapter 2: Plasmid Design & Construction for PHB Production in *Saccharomyces cerevisiae* Using a Single Vector

2.1: Introduction & Background

2.1.1: Introduction

Saccharomyces cerevisiae is a well known eukaryote, cultivated widely as a model organism in academia and utilized in many industrial and commercial applications, including baking, brewing, and recently, biofuel production. Because of this wide usage, facilities and expertise exist for large volume cultivation of *S. cerevisiae* industrially, with little need for the development of new infrastructure. Finally, as a generally recognized as safe (GRAS) eukaryote, products made by this strain are likely to be readily applicable in pharmaceutical and medical device applications, in contrast to the common poly-hydroxy alkanoate producing, gram negative soil prokaryotes *Pseudomonas oleovorans* and *Ralstonia eutropha*, (35, 36). As a result of these factors, *S. cerevisiae* is a major target organism for use in biotechnology applications, most specifically biopolymer production based on heterologous expression of prokaryotic genes.

Despite the benefits of biopolymer production in *S. cerevisiae* relative to natively producing strains outlined above, a number of hurdles remain before industrial production of biopolymers may be feasible. Previous experiments have demonstrated successful expression of *R. eutropha* PHA genes under control of the GAL1-10 promoter

in *S. cerevisiae* strain D603, with poly[(R)-3-hydroxybutyrate] (PHB) accumulation at up to 9% cell dry weight (CDW) using a two plasmid system (37). However, this accumulation was far below that observed in natively producing *R. eutropha*, which has been observed to consistently produce PHA at 80% cell dry weight (38). In addition, significant culture heterogeneity and production instability have hampered both reproducibility and total PHB yields in this strain (37).

In this work, we sought to reduce culture heterogeneity, and thereby improve PHB yield in *S. cerevisiae* D603, by expressing all three *R. eutropha* PHA genes in one plasmid.

2.1.2: PHB synthesis

PHB accumulation in *R. eutropha* is facilitated by three enzymes, which convert Acetyl-coA into 3-hydroxybutyrate, then subsequently polymerize these monomers to form PHB granules within the cytoplasm. Step one of this reaction (Figure 2.1) is catalyzed by β-ketothiolase, the transcription product of *phaB*, the condensation of two acetyl-coA molecules to form acetoacetyl-coA with the release a free CoASH. PHA reductase (transcription product of *phaA*) catalyzes the NAD(P)H linked reduction of acetoacetyl-coA to β-hydroxybutyryl-CoA. Finally, these β-hydroxybutyryl-CoA monomers are then polymerized by PHB polymer synthase (*phaC*), coupled with CoASH removal.

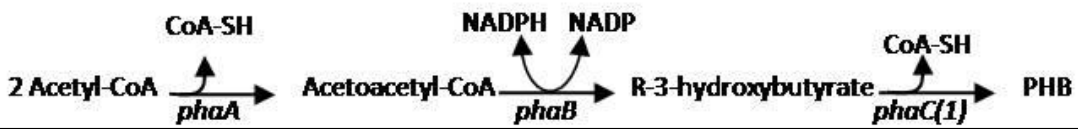


Figure 2.1: PHB polymerization from acetyl-CoA. Acetyl-CoA is first dimerized by PHA ketothiolase (*phaA*), then reduced by PHA reductase (*phaB*), and finally the 3-hydroxybutyrate monomer polymerized by PHA polymer synthase (*phaC, phaC1*).

PHA synthesis may also be catalyzed by *P. oleovorans* PHA polymer synthase (encoded by *phaC1)*, which has also been observed to incorporate medium-chain monomers into PHA polymers. The ability of the phaC1 synthase to incorporate monomers with 4-12 carbons can be used to generate block-copolymers with improved properties versus PHB, including reduced brittleness, crystallinity, and increased tensile strength (39).

2.1.3: Single-plasmid expression rationale

As noted in previous works, gene expression using multiple plasmids would be expected to result in sub-optimal bioproduct formation for several reasons. First, it is highly unlikely that plasmids would be maintained in either equal or optimal proportions in most cells to maximize PHB pathway flux, resulting in a reduction in overall product yield. Second, additional stress placed on cells to both produce additional DNA and to express multiple selection markers may also hamper optimal function (40). Finally, due to natural plasmid segregation during mitosis, plasmid addition is more likely to result in an increase in cell fraction lacking any or all plasmids, thus increasing the population fraction unable to produce the desired product due to lack of a complete production pathway. This phenomenon occurs regardless of which selection marker is used, as a certain fraction of cells in any population are expected to entirely lose plasmid presence through normal growth processes, even in high copy number plasmids (41).

Mitotic stability and segregation rates

Vector[a]	Marker scored	Mean values for five transformants after twelve doublings				Single transformants after 24 doublings			
		Percent (P) plasmid-carrying cells				Percent (P) plasmid-carrying cells			
		After 10 doublings selective[b]	After g doublings nonselective[c]		Segregation rate[d]	After 10 doublings selective	After g doublings nonselective[e]		Segregat rate
		P_1	g	P_2	m	P_1	g	P_3	m
pRS423	*HIS3*	73 ± 5	12.5 ± 0.4	42 ± 7	4.3 ± 1.3	80	24.7	21	5.3
pRS424	*TRP1*	82 ± 8	12.6 ± 0.2	68 ± 8	1.5 ± 1.7	80	23.9	69	0.6
pRS425	*LEU2*	73 ± 6	11.9 ± 0.3	42 ± 4	4.6 ± 0.6	73	23.5	29	3.8
pRS426	*URA3*	78 ± 5	12.8 ± 0.3	45 ± 11	4.4 ± 1.4	82	24.6	37	3.2
YEp24	*URA3*	88 ± 3	12.7 ± 0.3	69 ± 4	2.0 ± 0.7	88	24.0	55	2.0
pRS313	*HIS3*	76 ± 5	12.3 ± 0.5	51 ± 8	3.2 ± 1.4	77	24.4	36	3.0

Table 2.1: Mitotic stability and segregation rates for high copy number 2µ plasmids (pRS420 series), yeast integrative plasmid (Yep24 and pRS313). Obtained from Christianson *et al.*

As shown in Table 2.1, the percent of plasmid carrying cells with the 2µ origin, following 10 generations, varies between approximately 70-90% of a given population. Plasmid segregation kinetics, in terms of number of cell divisions, are described below. Beginning with Eq. 1 to determine the number of doublings between inoculation and accession to stationary phase,

$$g = \frac{\ln(\frac{N2}{N1})}{\ln 2} \qquad \text{Eq. 1}$$

Where g is the number of doublings, N_1 is the viable cell concentration at 1:1000 dilution, and N_2 is the viable cell concentration at stationary phase. g can then be used to determine segregation rate (m) for a single plasmid using Eq. 2 which is the number of plasmid-free segregants appearing after each doubling,

$$m = 1 - e^{\frac{\ln(\frac{P2}{P1})}{g}} \qquad \text{Eq. 2}$$

Where P_2 and P_1 are percent plasmid carrying cells following non-selective growth and selective growth, respectively (41). Using this equation, the segregation rate for 2µ plasmids with a uracil selection marker was determined to be 4.4%, meaning that after each doubling, 4.4% of mitosis products would be expected to be without any plasmid. To determine segregation rate for multiple plasmid systems, Eq. 2 can easily be modified to give Eq. 3, where m represents the number of cells lacking at least one distinct plasmid type following a single doubling.

$$m = 1 - \prod_i^n e^{\ln(\frac{P2}{P1})/g} \qquad \text{Eq. 3}$$

given i is an integer representing the number of different plasmids (ideally with different selection markers) are maintained within a population. This equation assumes no interaction between plasmids; that insertion of additional plasmids to a system does not increase overall instability through homologous recombination of vectors, a stipulation known to be overly simplistic based on observations made in other yeast 2µ systems (42). Because of this assumption, the plasmid stability obtained from Eq. 3 represents a theoretical maximum plasmid stability, and the reality may be between 20% and 50% lower than this result, due to homologous recombination between 2µ regions on each plasmid (42).

If Eq. 3 is used to describe the D603 p2DPT-RK(U) p2DP-S(H) system for PHB production, which contains two plasmids with 2µ origins of replication, as well as a histidine or a uracil selection marker on each plasmid, the following segregation rate is

expected after 10 doublings, indicating the population fraction lacking either or both plasmids, using data in Table 2.1.

$$m = 1 - \prod_{i}^{n} e^{\ln(\frac{P2}{P1})/g} = 1 - \left(e^{\ln\frac{\frac{42}{73}}{12.5}} * e^{\ln\frac{\frac{45}{78}}{12.8}} \right) = 0.08354 \cong \mathbf{8.4}\%$$

As a result of Eq. 3, segregation rate increases with increasing plasmid type, resulting in significant increase in the rate of combined plasmid loss, and therefore reaction network loss from the population. Further analysis of these results show that in the simplest possible case, where we assume no interactions between plasmids, only 56.9% of cells possess both plasmids after 10 generations using the data in Table 2.1. Clearly, plasmid instability is a significant source of culture heterogeneity, and must result in significant deviation from optimal PHB yields, when compared to both integrative and single plasmid systems. Product yields could therefore be significantly improved through the use of integrative or single plasmid systems for PHB gene expression.

Single plasmid systems were chosen for PHB expression over integrative systems for several reasons. First the portability and relative ease of modification of plasmids versus integrative systems in yeast are highly desirable for research applications, where transfer between several strains was desired. Second, previous experiments have shown highest yields, as well as fewest protein/growth defects, using PHB expression behind GAL1-10 promoters using high copy number plasmids. This is relative to PHB and green fluorescent protein (GFP) expression behind integrated the constitutive TEF1 promoter in integrated expression systems (43).

2.1.4: Single plasmid applications in fluorescence microscopy

An additional benefit to single-plasmid PHB gene expression in yeast is that this allows selection based on a single selection marker. This is specifically applicable to the lipid-storage knockout strain H1246 (W303 MATα yor245::KanMX4 pdat::TRP1 are1::HIS3 are2:LEU2 ADE2 ura3), which contains 4 knockouts that nearly eliminate lipid storage. However, unlike D603, H1246 is not auxotrophic in histidine, preventing selection for the PHA synthase bearing plasmid p2DP-S(H)(44). As a result, any PHA synthase-bearing must either be modified to select based on adenine complementation, or the PHA synthase gene and its promoter transferred to p2DPT-RK(U) to express all genes on one plasmid, with uracil-based selection. For reasons discussed above, the latter option was preferred.

The ability to express all three PHB producing genes in strain H1246 was essential to obtaining further understanding of the physiology of PHB accumulation in yeast cells under this expression system. As previously discussed, significant culture heterogeneity has been observed regarding PHB expression, where a small fraction of a given population accumulated PHB as a high fraction of intracellular volume, while many cells showed little accumulation (37). When attempting to analyze PHB accumulation using the hydrophobic fluorescent dye BODIPY (493/403), however, visualizing PHB granule formation was hampered by co-staining of naturally occurring intracellular lipids produced by the D603 strain, as shown in Figure 2.2. Such interference could be removed through the transformation of H1246 with PHB producing plasmid(s), allowing direct, unambiguous observation of intracellular PHB accumulation

in yeast. One potential drawback of this strain is that, unlike D603, H1246 has a fully regulated galactose metabolism (D603 contains the Δ*reg1* mutation), which may prevent full expression of genes behind GAL1-10 in the presence of low glucose concentrations (45).

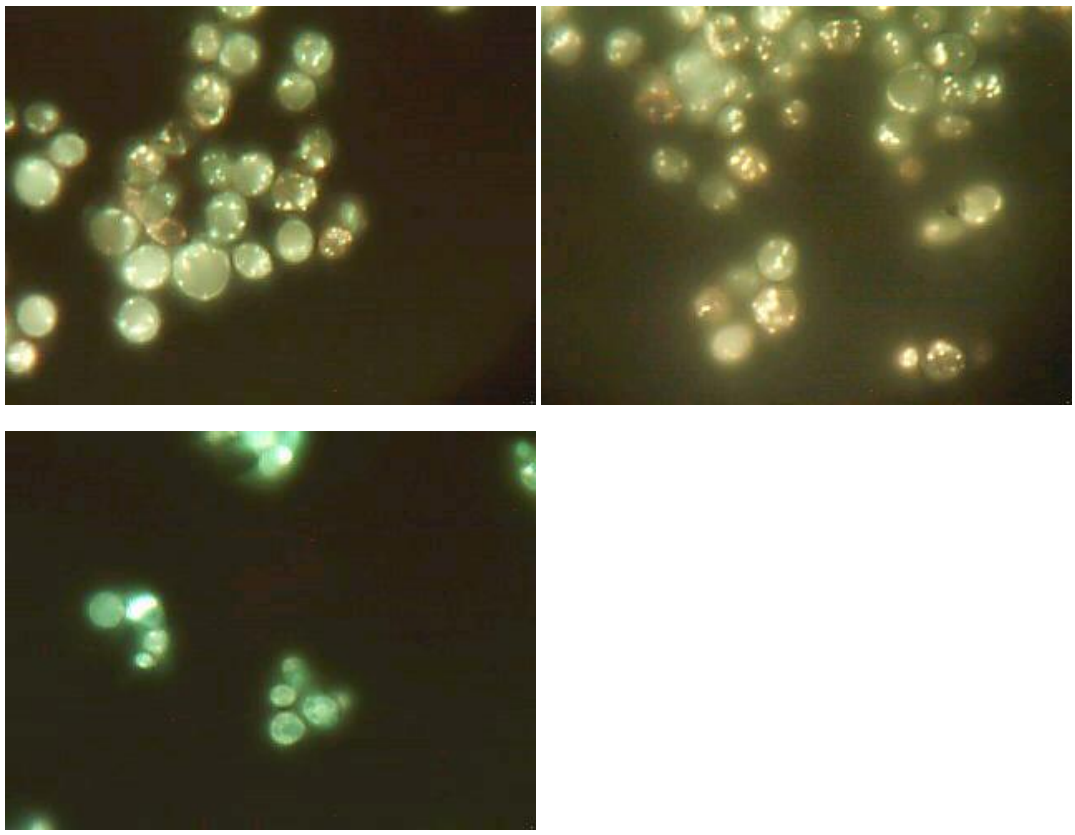


Figure 2.2: Panel A: D603 p2DPT-RK(U), p2DP-S(H) grown on 2% glucose SD medium for 72 hrs. Panel B: D603 p2DPT-RK(U), p2DP-S(H) grown on 1% galactose/1% glucose SD medium for 72 hrs. Panel C: lipid-storage knockout strain H1246 following 72 hrs growth on 1% galactose/1% glucose SD medium. PHB accumulation of ~1% CDW was measured by propanalysis in the cells in panel B, while non-expressing cells (panel A) and wild-type H1246 (panel C) did not produce any PHB.

The intracellular bright spots within all cells are either lipids or PHB; as demonstrated when comparing panels A (PHB neg) and B (PHB positive), identifying PHB granules in the presence of storage lipids is not trivial. If the storage-lipid negative strain H1246 is transformed to produce PHB, such determination may be possible, given the lack of interfering lipids within this strain.

2.1.5: GAL1-10 divergent promoter

The GAL1-10 divergent promoter is a widely studied inducible promoter system used in *S. cerevisiae* to express recombinant proteins under conditions where the primary carbon source is galactose. This system is strongly repressed in wild-type *S. cerevisiae* by

the presence of glucose, even in very low concentrations. As a result, a mutant strain, D603, containing the *reg1-501* mutation, partially prevents glucose inhibition of GAL1-10 expression, allowing GAL1-10 expression in low-glucose concentrations (45). This is particularly useful because, as the preferred carbon source for *S. cerevisiae*, the growth rate of *S. cerevisiae* when cultivated on glucose is higher than that of the same strain when cultivated on galactose alone. This mutation facilitates relatively high expression of genes behind GAL promoters when cultivated on mixed-sugar media before complete exhaustion of extracellular glucose in *reg1-501* strains (45).

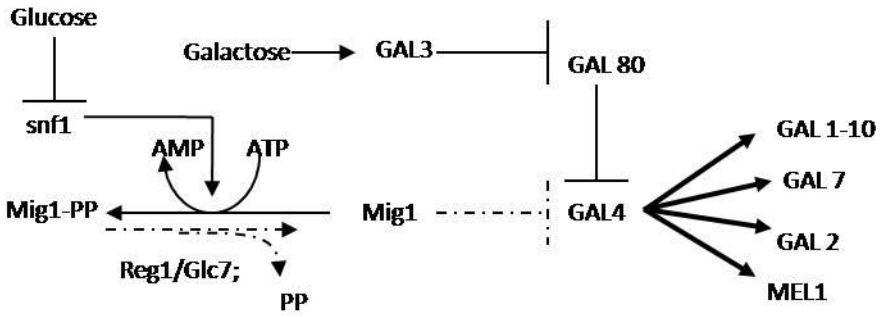


Figure 2.3: Glucose repression and galactose derepression of GAL1-10 genes. Glucose presence inhibits Snf1 activity, which prevents Snf1 from dephosphorylating Mig1. Dephosphorylated Mig1 represses GAL4 expression. In addition, galactose presence activates GAL3, which binds to and deactivates GAL80, resulting in derepression of GAL4, and results in GAL and MEL enzyme expression in the absence of glucose.

While the exact interaction between *reg1* and GAL expression *in vivo* are not known, their interaction has been proposed to follow the signal cascade diagrammed in Figure 2.3. While the interaction between GAL80-GAL4 and subsequent GAL expression has been well established, the relationship between *reg1* and other proteins are not completely understood, and proposed interactions have been represented as dashed lines. GAL regulation is primarily regulated by GAL4 expression, which is in turn

repressed by GAL80 transcription. Regarding glucose repression, at high glucose concentrations the Snf1 complex is deactivated, preventing Mig1 phosphorylation. At low glucose concentrations, Snf1 triggers Mig1 phosphorylation, preventing its repression of GAL4 transcription. In addition, it is known that *reg1* encodes a dephosphorylase that binds to Glc7 in the presence of glucose, forming a complex that has been proposed to dephosphorylate the Mig1 protein, reactivating it and strengthening glucose repression by reversing Snf1 complex activity (46)(47). In turn, the Mig1 protein has been proposed to repress galactose-related and other carbon utilization genes by repressing GAL4 expression unless phosphorylated (47). Correspondingly, galactose presence inhibits GAL80 expression by activating GAL3 expression, which binds to GAL80, therefore derepressing GAL4 and allowing transcription of the GAL1, 2, 7, 10 and MEL1 genes(47, 48). GAL4 binds DNA in the absence of galactose, and has been observed in multiple phosphorylation states, so it may require (de)phosphorylation to be activate transcription (47). Taken together, the wild type system requires both the absence of glucose, and the presence of galactose to produce active GAL4, and allow transcription to occur by removing repression of GAL4. In the case of *reg1* mutants, the glucose repression pathway is weakened by preventing Mig1 rephosphorylation, effectively making the Snf1 mediated activation of this enzyme irreversible reducing glucose repression.

In addition, the GAL1-10 divergent promoter describes a 681 bp region containing two genes promoters. The sequence of the entire region, and the sequences required for expression from each promoter were determined by Johnston and Davis

(1984). Using the start of the GAL10 gene as a reference point, the GAL10 promoter requires bases at least 324 bases upstream to achieve full expression. Likewise, at least 384 bases upstream of GAL1 are required for full expression (49). Because these promoters are on opposite DNA strands, there is an overlap of approximately 50 bp. The divergent promoter cassette used, with restriction digest sites added by cloning is presented in Figure 2.4. The cloning of this cassette is described by Carlson (2006).

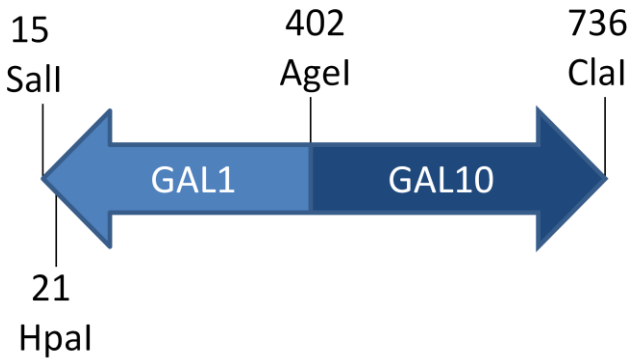


Figure 2.4: The GAL1-10 divergent promoter system. Although not shown in this diagram, these promoters overlap for approximately 25 base pairs, 50 bp downstream of the AgeI restriction site.

2.1.6: TEF1 promoter systems

The Translational Elongation Factor 1 promoter (TEF1) constitutive promoter was utilized in plasmid construction as a promoter for *R. eutropha phaC* and *P. oleovorans phaC1* PHB polymer synthase gene expression. The TEF1 promoter sequence is 407 bp in length, the creation of the source plasmids used in this study, pNTG1-S(500) and pNTG1-S(700) is described in Zhang (2005). This promoter is bounded by HindIII and ClaI restriction digestion sites.

Further information regarding this promoter sequence is described by Planta and Raué (1988), Tornow *et al.* (1993), and Uemura *et al.* (1996), and Bi and Broach (1999). In

addition, expression of heterologous genes in *S. cerevisiae* is described by Kacmar, *et al.* (2004), and Carlson (2003).

2.2: Materials and Methods

2.2.1: Experimental strains

All yeast studies were performed using *S. cerevisiae* strain D603 (MATa/MATα *ura3-52 lys2-801 met his3 ade2-101 reg1-501 [cir+])*(50).The reg1-501 mutation at least partially releases glucose repression of galactose catabolism (45)(51). This inhibited repression facilitates effective expression of genes under control of the GAL1-10 promoter, allowing induction in the presence of glucose with galactose concentration as low as 0.2 g/l (40). *S. cerevisiae* strain H1246 (W303 MATα yor245::KanMX4 pdat::TRP1 are1::HIS3 are2:LEU2 ADE2 ura3), a quadruple knockout strain with greatly reduced intracellular lipid accumulation capacity, was also considered for application in microscopic analysis of PHB expression (44).

All common plasmid manipulation and construction was performed in *Escherichia coli* strain DH5α (Invitrogen, Carlsbad, CA).

2.2.2: Bacterial growth media

DH5α was cultivated in standard Luria-Bertani (LB) media (5 g/L Bacto yeast extract, 10 g/L Bacto Tryptone, (Difco, Detriot, MI) 10 g/L NaCl, and 15 g/L agar as necessary). When necessary for plasmid maintenance, plates and media were supplemented with 100 mg/L ampicillin. Strains were cultivated at 37ºC. Liquid cultures

were grown in either 14 ml shake tubes (5 ml liquid volume) or 250 mL shake flasks (50 mL working volume) with shaking at 200 rpm.

2.2.3: Yeast growth media

Wild-type yeast strains were cultivated in YPD (10 g/L yeast extract, 20 g/L peptone (Difco), 20 g/L glucose, (15 g/L agar when necessary)). Recombinant cultures were cultivated in SD dropout medium, (20 g/L glucose, 1.7 g/L yeast nitrogen base w/o AA's (Sigma, St. Louis, MO), 5 g/L ammonium sulfate, 120 mg/L leucine, 120 mg/L adenine, 200 mg/L methionine, 300 mg/L lysine, 160 mg/L histidine, 120 mg/L uracil). Liquid SD media was filter sterilized using 0.2 µm pore size nylon membranes (Millipore, Billerica, MA). All yeast vectors utilized either URA3 or HIS2 selection markers, and SD dropout media was modified accordingly to select for plasmid maintenance.

S. cerevisiae liquid cultures were grown in either 14 mL shake tubes (5 ml liquid volume) or 250 mL shake flasks (50 mL working volume) at ~200 rpm and 30ºC.

2.2.4: Strain preservation

All strains were cryopreserved at -80ºC in 15% glycerol solution (52).

2.2.5: Hydrophobic staining with BODIPY (493/503) and microscopic slide preparation

Cells were diluted to an optical density of 3 and were fixed in 30% ethanol for 10 minutes, then washed twice in PBS. After resuspension in 1 mL PBS, 10 µl BODIPY solution (100 µg/ml BODIPY in DMSO) was added and incubated for 5 additional minutes. Cells were immobilized by adherence to glass cover slips using 0.1%

Polyethyleneimine (PEI). Slides were initially washed with filtered 100% ethanol, air dried, followed by the application of PEI solution to one side of each cover slip and incubated for an additional 10 minutes. Slides were gently rinsed with 1 mL filtered dH_2O and allowed to air dry. Stained cell solutions were then applied to the PEI coated side of each cover slip, incubated for 10-15 minutes, and gently rinsed with 1 mL filtered dH_2O to remove excess non-adhered cells. Cover slips were then inverted and sealed to glass microscope slides using clear nail polish. All incubation steps were at room temperature.

2.2.6: PHB determination by propanalysis

PHB content was determined by the following procedure. A known volume of culture was spun at 600 rpm using a Beckman J2-21 centrifuge (Brea, CA) for 15 minutes in glass, screw-top tubes lined with PTFE tape. Cell pellets were then washed 2X with cold dH_2O, given a final wash in 1 mL methanol, centrifuged, the methanol decanted, and the resulting cell pellet dried overnight at 50°C. This dried cell pellet was then suspended in 1.5 mL propanalysis solution (50% dichloroethane, 10% HCL, 40% 1-propanol). An internal standard of 20 µl benzoic acid was added to all samples and vortexed. Samples were then incubated, with appropriate standards, for 2 hours in a boiling water bath. Propanalysis reaction was then stopped by the addition of 1-2 mL dH_2O, and the organic layer inserted into GC flasks for analysis using a GC-17A gas chromatograph (Shimadzu, Columbia, MD). PHB content was calculated by dividing the area of the PHB peak by the area of the benzoic acid peak of known PHB standards to create a standard curve.

2.2.7: Plasmid isolation

Plasmids were purified from overnight DH5α cultures using the Wizard Plus miniprep DNA purification system (Promega, Madison, WI). Five mL cultures were pelletted by centrifugation for 10 minutes at 6000 rpm in a Beckman J2-21 centrifuge (Beckman, Brea, CA), and the supernatant decanted. Cells were resuspended in 250 µL "cell resuspension buffer", then lysed using 250 µl "cell lysis buffer" and 10 µl "alkanine protease" for 5 minutes. Lysis was stopped by addition of 350 µL "neutralization buffer", and spun for 10 minutes at maximum speed (13.2 rpm) on a tabletop centrifuge. The supernatant was then placed in a DNA binding column, washed twice in a buffered ethanol solution, and eluted using 100 µL purified water. Purified plasmid solutions were frozen at -20^0 C for further analysis.

2.2.8: Restriction endonuclease digestion

Restriction digestion was commonly performed using 10X New England Biolab buffers (NEB), at 37^0 C for 1 hour or overnight, as empirically determined to be necessary for complete digestion. 100 ng to 5 µg DNA was used in each digestion, depending on application (lower for analysis, higher for plasmid construction). Enzyme concentration was modified accordingly to achieve at least a 1:1 ratio of µg DNA to units of enzyme (1 unit enzyme = amount required for complete digestion of 1 µg DNA in 1 hour at 37^0 C).

2.2.9: Phenol-chloroform extraction with ethanol precipitation

A volume of phenol:chloroform:isopropanol (25:24:1) solution equal to that of the DNA solution was added in a 1.5 mL centrifuge tube and vortexed for 15 sec, followed

by 1 minute centrifugation on a tabletop centrifuge at maximum speed. The aqueous layer was then removed by pipette, and a 1/10 volume of 3M NaOAc (pH 5.2) was then added to the aqueous solution and quickly vortexed. A 2.5X volume of cold, filtered 100% ethanol was then added to this solution, and incubated at -80ºC for 10 minutes. This was then centrifuged at room temperature for 5 minutes, the supernatant removed, and washed in 500 µL cold, filtered, pure ethanol. Following another 5 minute centrifugation and supernatant removal, the pellet was then vacuum dried at room temperature until visibly dry (approximately 5 minutes). DNA was then dissolved by addition of 20 µL sterile TE (10 mM Tris, 1mM EDTA, pH 7.5) (52).

2.2.10: Generating blunt DNA fragments using T4 polymerase

Following phenol:chloroform extraction, relevant DNA samples (those with 3'-5' overhangs) were treated with T4 polymerase in 50 µL total volume reactions, containing 10 µL 5X T4 Buffer, 0.5 µL Bovine Serum Albumin (BSA) to a final concentration of 100 µg/ml, and 5 units of T4 enzyme, with the remaining volume consisting of sterile filtered dH_2O and DNA solution. The reaction occurred at 11ºC for 20 minutes, and was halted by gel purification. This procedure is described in the "T4 DNA polymerase" directions (Invitrogen, catalog number 18005-017).

2.2.11: Generating blunt DNA fragments with large fragment (Klenow) polymerase I

Following phenol:chloroform extraction, relevant DNA samples (those with 5'-3' overhangs) were treated with klenow fragment polymerase in 50 µL total volume reactions, containing 5 µL 10X NEB Buffer 2, 5 units of Klenow fragment polymerase,

and with the remaining volume consisting of sterile filtered dH_2O and DNA solution. The reaction occurred at 11^0 C for 20 minutes, and was halted by gel purification (52).

2.2.12: Vector dephosphorylation for use in ligation

Gel purified, blunted vector was mixed with 10X rAPid Alkaline Phosphatase Buffer and 2 Units rAPid alkaline phosphatase (Roche, Basel, Switzerland) and allowed to react at 37ºC for 30 minutes, before heat inactivation using a 75ºC water bath. This solution was then used, without purification, in subsequent ligation steps.

2.2.13: Ligating blunt fragments

Dephosphorylated, blunted vector and blunted insert fragments were combined in a 1:5 mass ratio (~1:20 molar ratio)(33.89 ng vector, 177 ng insert). Ligation solution was prepared by appropriately diluting 5X ligation buffer in DNA solution, and using 2 units of T4 DNA ligase enzyme. The ligation reaction was allowed to proceed at room temperature for 1 hour, then overnight at 15ºC to compare efficacy of different reaction conditions. For blunt ligations, overnight ligation and transformation using Subcloning Efficiency DH5α chemically competent cells (Invitrogen) were necessary. "Sticky-end" ligations were successful using 1 hour ligations as well as both Invitrogen and prepared competent cells.

2.2.14: DNA electrophoresis and gel purification

Electrophoresis was used to separate linear DNA fragments, typically following restriction digestion or PCR amplification. Samples were run on 0.7% (wt/v) GenPure LE agarose (ISC BioExpress, Kaysville, UT), prepared in standard Tris/borate/EDTA buffer

(TBE) (53 g/L Tris base, 27.5 g/L boric acid, 10 mM EDTA). Samples were loaded into wells using Bluejuice loading buffer (Invitrogen) and typically exposed to 130 V for 30-60 minutes until band separation was visible. Gels were imaged using a FOTO/Eclipse Benchtop Darkroom system and accompanying FOTO/analyst PC Image V10.21 software (Fotodyne Inc, Hartland, WI), using an ethidium bromide filter. Bands were excise from 0.7% agar gels and purified using a QIAquick gel extraction kit (Qiagen, Valencia, CA).

2.2.15: Preparing chemically competent DH5α cells

Overnight DH5α cells were diluted 100X in 100 mL LB and allowed to grow to an OD of approximately 0.4-0.6. Solutions were then placed on ice for 10-20 minutes and centrifuged at 600 rpm for 10 minutes. Cells were then resuspended in 10 mL cold 100 mM $CaCl_2$ and placed on ice for 20 minutes. Samples were centrifuged and resuspended in 5 mL 100 mM $CaCl_2$, 15% glycerol (all sterile), and aliquotted into microtubes in 100-300 µL volumes. These were then stored at -80ºC until use.

2.2.16: Transforming chemically competent DH5α cells

Aliquots were taken from -80ºC and allowed to thaw slowly on ice. Volume was adjusted as necessary (100 µL for self-made competent cells, 50 µL for Invitrogen cells), and 2 ul ligation mixture (Invitrogen) or 20 µL ligation mixture (lab-treated DH5α) was added to cells and incubated on ice for 20 minutes. Cells were then heat-shocked at 42ºC for 1 minute and placed on ice for an additional 2 minutes. Two hundred fifty µL of S.O.C. liquid medium (Invitrogen) was then added to each sample, and incubated at 37ºC for 1 hour. Samples were then centrifuged for 1 minute at maximum velocity on a

tabletop centrifuge, excess supernatant removed, and the pellet resuspended in 100 µL total volume, and spread on selective LB plates containing 100 µg/mL ampicillin and allowed to grow overnight. This procedure was adapted from Short Protocols in Molecular Biology (1995). It is especially important that plates be fresh (stored for <1 month at 4ºC), and DH5α growth at 37ºC to be less than one day, because the rapid degradation of ampicillin results in the appearance of satellite colonies lacking antibiotic resistance, and thus intact plasmids when cultivated for longer periods.

2.2.17: Transforming chemically competent *S. cerevisiae*

The yeast transformation procedure was modified from "Rapid and simplified procedure" described by Soni, Rajeev *et al* (1993). 1.5 mL of *S. cerevisiae* cultures exponential phase (5mL YPD total volume) were centrifuged and the supernatant removed. In minimal volume, cells were resuspended in10 ul Salmon sperm DNA (10 mg/ml) and 1-20 ul DNA solution, depending on DNA concentration. Typically, total DNA ranged between 50 ng -1 ug, depending on availability and reaction type. This solution was then resuspended by pipetting in 500 ul LiAc solution (40% poly-ethylene glycol 3350, 100 mM LiAc, 10 mM Tris pH7.5, 10 mM EDTA, and 10% v/v DMSO). This was then incubated with gentle shaking on a vortex mixer at room temperature for 20 minutes. Samples were then incubated for 20 minutes in a 42ºC water bath, with occasional mixing by inversion. Following heat shock, samples were placed on ice for 2 minutes, and washed 2X in sterile TE buffer, and resuspended in 100 uL TE buffer. Entire samples were then spread on dropout SD medium.

2.2.18: Yeast mediated ligation using *in vivo* homologous recombination

Yeast mediated ligation was performed using methods adapted from Shanks *et al.* (2006). Target regions were cloned with 60 bp primers for insertion into single-cut vectors. Typical primers consisted of ~40 bp of homology to the target region in the vector, with the remaining ~20 bp serving as forward or reverse primers, typical of PCR cloning. Target regions were less than 1000 bp upstream/downstream of the single digestion site in the vector. Best results were obtained when minimizing vector concentration, while maximizing insert concentration. As a result, insert: vector molar ratios of 20:1 are recommended, with total DNA mass of 175 ng :35 ng. Transformation was performed using the LiAc method described in section 2.2.9. *S. cerevisiae* strain D603or H1246 was used in all homologous recombination experiments. Strains were grown up to 5 days and colonies selected for preservation at -80°C or streaked onto 0.1%/1% glc/gal SD dropout plates for selection of high-producing strains by microscopic analysis using BODIPY staining to assess PHB production.

2.2.19: *E. coli* mediated ligation using *in vivo* homologous recombination

E. coli mediated ligation was performed using homologous recombination as described by Bubeck *et al.* (1993) and Oliner *et al.* (1993). Target regions were cloned using PCR using 60 bp primers for insertion into vectors typically digested with a single restriction endonuclease, and a single cut site. Typical primers consisted of ~40 bp of homology to the target region in the vector, with the remaining ~20 bp serving as forward or reverse primers, typical of PCR cloning. Target regions were less than 1000 bp upstream/downstream of the single digestion site in the vector. *E. coli* strain DH5α was

used in all transformations. Transformants were prepared using chemically competent cells, prepared and stored at -80°C as described in sections 2.2.7 and 2.2.8. 1.5 mL *E. coli* exponential phase (OD 0.4-0.6) cells were transformed with 40 ng single-cut vector and 175-260 ng insert (1:17-25 mole:mole) and overnight on selective plates. Eight colonies were arbitrarily selected for shake tube cultivation.

2.2.20: Plasmid Sequencing primers, conditions, and analysis

Primers for sequencing were selected to target regions approximately 50-100 bp upstream of the region to be sequenced, to transcribe known regions for reference. Anealing takes place at 55°C, so primers with Tm above 58°C were found to be successful. Successful sequencing required primers to be prepared at exactly 0.267 pmol/µl, with final plasmid concentration of 27ng/µl, in a total volume of 12 µl sterile filtered water. If DNA secondary structure is expected to interfere with sequencing, 5% vol/vol DMSO can also be added. Sequences were analyzed using ClustalW (53). Fresh primers and vector DNA were deemed to be necessary, and care was taken to ensure that sequencing primers were never frozen, instead being stored at 4°C and used with days.

2.2.21: Source plasmids

Previous experiments performed by Ross Carlson demonstrated successful PHB accumulation (up to 10% CDW) in *S. cerevisiae* strain D603 using two plasmids, p2DPT-RK(U) and p2DP-S(H) (37). Plasmid p2DPT-RK(U) contains a 2 micron origin of replication (signified by the "2" in its name), the GAL1-10 divergent promoter (DP), the 236 bp uraTerminator region (T), and the *R. eutropha* PHB reductase (*phaB*) gene in the gal1 promoter position, and the *R. eutropha* ketothiolase (*phaA*) gene behind the GAL10

promoter region. This plasmid contains the URA3 selection marker. The plasmid p2DP-S(H) contains the same general structure, except lacking a gene behind the GAL1 promoter, no uraTerm region, and the *R. eutropha* ketothiolase (*phaC*) gene is expressed behind the GAL10 promoter. A detailed description of the construction of these plasmids can be found in Carlson *et al* (2006), and these plasmids are presented in Figure 2.4.

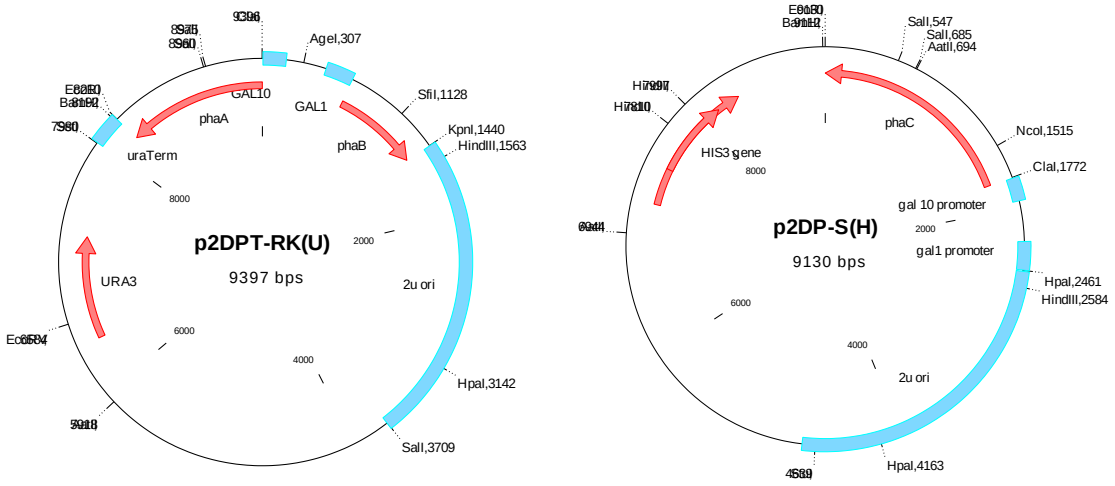


Figure 2.5: High copy number yeast PHB production plasmids utilizing the divergent GAL1-10 promoters, with URA3 or HIS2 selection markers. p2DPT-RK(U) contains the *phaA* gene (PHA ketothiolase) in the GAL10 promoter slot, with the *phaB* gene (PHA reductase) in the GAL1 promoter position. p2DPT-S(H) carries the *phaC* gene (PHA synthase) in the GAL10 promoter position.

2.3.1: Plasmid characterization: UraTerm Sequencing

The exact sequence of the uraTerm region, located at the terminal end of the phaA gene on p2DPT-RK(U), could not be determined from source literature. As a result, this region was sequenced using the following primer: GAAGAAGGGCCTGGCCTC. This primer has a recorded Tm of 58.4ºC in 50 mM NaCl, and corresponds to the 1128-1130 bp region of the *R. eutropha* H16 *phaA* gene (geneID 4249783), bp upstream of the stop codon of *phaA*. Raw sequencing results are below, with sequence identities determined

by alignment using clustalW to known restriction sites, *phaA*, and the reverse complement of the vector from the *ura3* gene to the sstI restriction site. The final uraTerm sequence was determined to be the region bound by ecoRI-sstI digestion sites. The uraTerm sequence was verified by one additional sequencing reaction using the same primer with DMSO addition.

1 TGGGTGGGCGCTGAGCGTGGCGCTGGCAGTCGAGCGCAAATAA [GAATTC]
51 CTGCAGCCCGGGGGATCCACTAGGGTTTTTGTTCTGTGCAGTTGGGTTAA
101 GAATACTGGGCAATTTCATGTTTCTTCAACACTACATATGCGTATATATA
151 CCAATCTAAGTCTGTGCTCCTTCCTTCGTTCTTCCTTCTGTTCGGAGATT
201 ACCGAATCAAAAAAATTTCAAGGAAACCGAAATCAAAAAAAAGAATAAAA
251 AAAAAATGATGAATTGAAAAGGTG [GAGCTC] *CAATTCGCCCTATAGTGAGT*
301 CGTATTACAATTCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAAC
351 CCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAG
401 CTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGC
451 GCAGCCTGAATG

Figure 2.6: Sequence result from p2DPT-RK(U) of phaA-uraTerm-ura3 region (reverse complement direction). Underlined sequence: *phA.* [Brackets]: restriction sites ecoRI, sstI, known to straddle the uraTerm region. **Bold:** uraTerm region. Italics: perfectly aligned region of p2DPT-RK(U) reverse complement region downstream of the *ura3* gene.

2.3.2: Construction of a single vector for PHB expression in yeast: homologous recombination around the aatII site using yeast mediated ligation.

As a first attempt to generate a recombinant plasmid containing all three *R. eutropha* PHB synthesis genes, yeast—mediated homologous recombination was attempted using primers with targets on both sides of the *aatII* restriction digestion site on p2DPT-RK(U). The GAL1-10-*phaC* (PHA synthase) cassette was cloned from the plasmid pDP306-syn, a previously prepared plasmid containing the GAL10-*phaC* cassette (37).

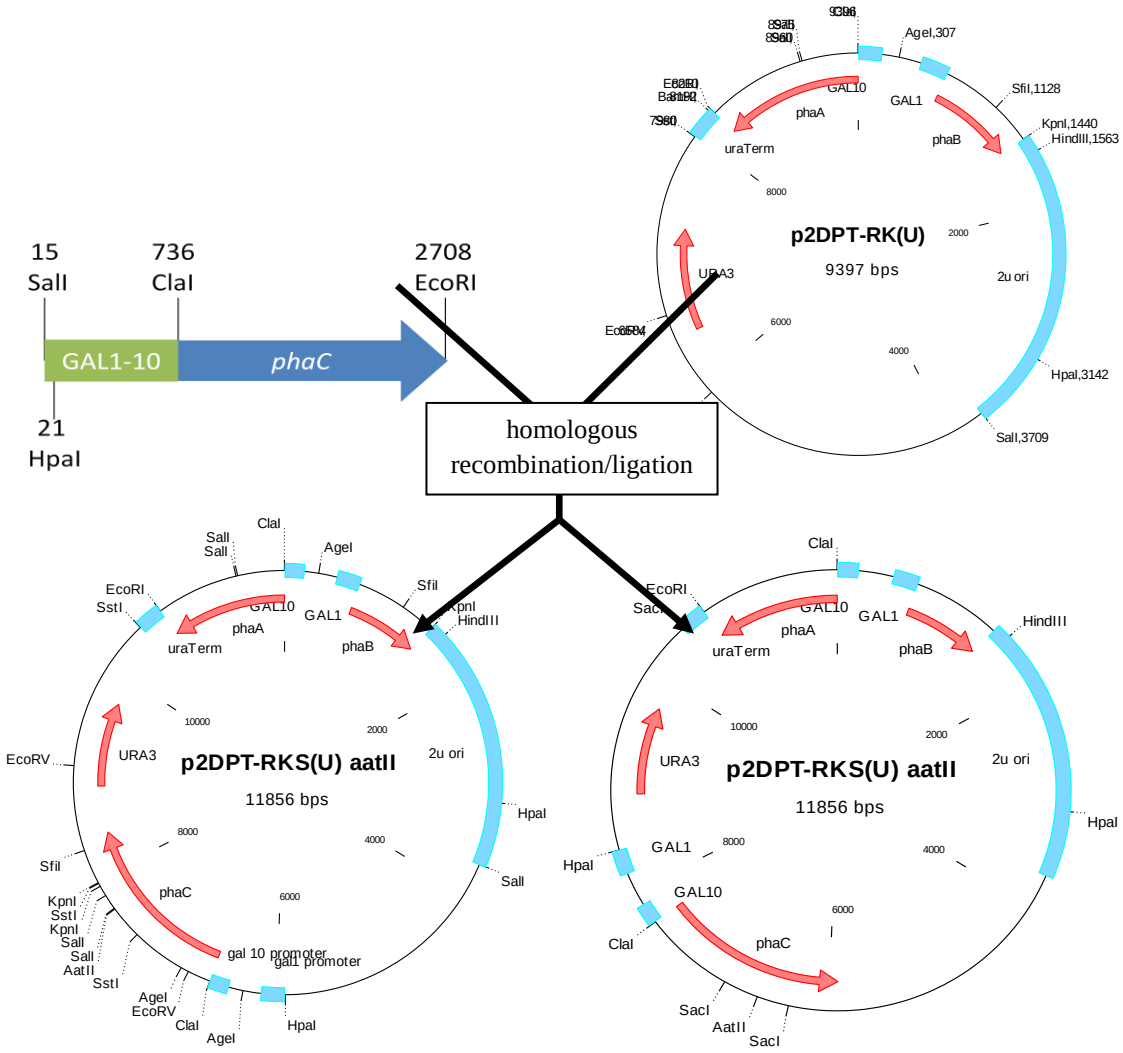


Figure 2.7: Insertion of the GAL1-10 promoter with *R. eutropha phaC* behind the GAL10 promoter into p2DPT-RK(U) at the AatII digest site. Using blunt-end ligation, this may give two plasmids, with *phaC* facing in opposite directions. Yeast mediated-ligation, if successful, would yield only the plasmid shown on the lower left, with *phaC* on the same strand as the *ura3* gene. It was hypothesized that this orientation would reduce homologous recombination between GAL promoter regions. The AatII restriction site on p2DPT-RK(U) is highlighted with a blue arrow.

Table 2.2 Primers used in homologous recombination reactions

Primer ID	Description	length (bp)	Target bp on p2DPT-RK(U), relative to aatII digestion site	sequence [cloning target in brackets]
GAL-Tm	High Tm forward primer	60	-121 to -78	AATATTATTGAAGCATTTAT CAGGGTTATTGTCTCATGAG CGG[CTCTCCTTGA CGTTAAA]
GAL1	Forward primer, GAL1	60	-121 to -84	AATATTATTGAAGCATTTAT CAGGGTTATTGTCTCAT[CTC TCCTTGACGTTAAAG TATAG]
phaCrevA	reverse primer, phaC	52	233 to 266	AATTGAATTGAAAAGCGTG GTGCACTCTCA [TCATGCCTTGGCT TTGACGTAT]
phaCrevB	reverse primer, phaC	59	233 to 274	AATTGAATTGAAAAGCGTG GTGCACTCTCAGTACAAT[T CATGCCTTGGCTT TGACGTAT]
phaC-303	reverse primer, phaC	60	289 to 327	AATCAAAAAAATTTCAAAG AAACCGAAATCAAAAAAAA [TCATGCCTTGG CTTTGACGTAT]

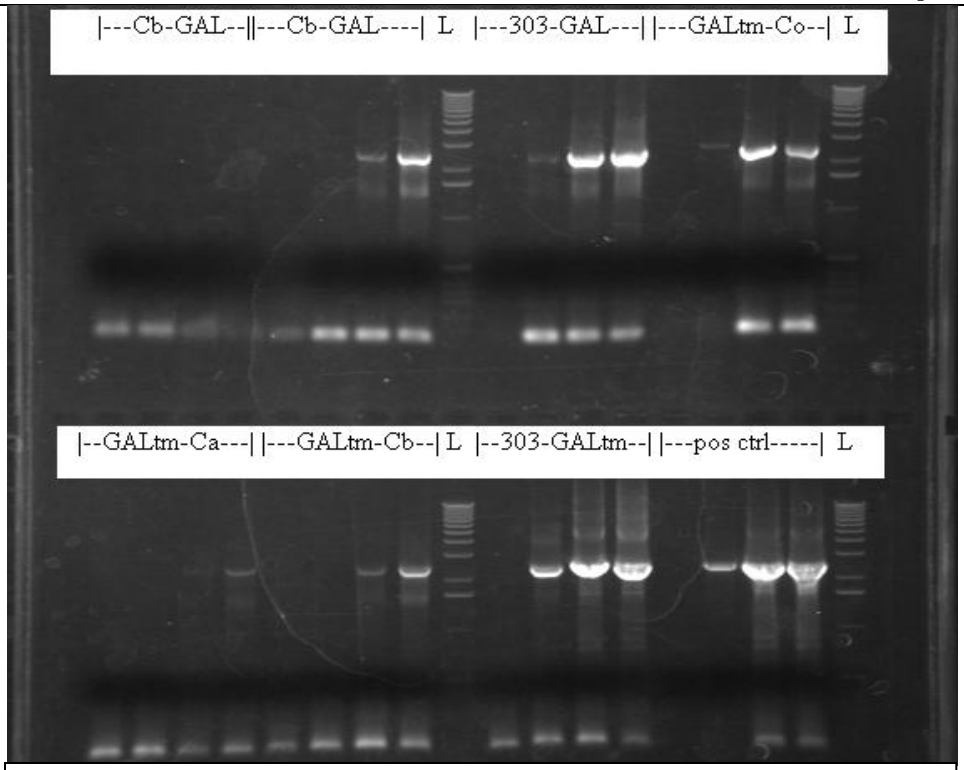


Figure 2.9: Primer test of GAL1-phaC primers with different homologous recombination target regions. Each set of four lanes is at annealing temperatures between 55-68°C from left to right. Cloned products at highest concentration were extracted by gel purification for plasmid construction.

Yeast D603 and H1246 strains were transformed as described in section 2.2.18 using the primers described above. Transformants were then streaked onto 0.1%/1% glc/gal SD dropout plates for PHB production assay by BODIPY staining. Plates were streaked with 8 colonies/plate. The densest portion of each colony streak was then swabbed using 1 mL pipette tips, suspended in PBS, and stained with BODIPY for PHB accumulation analysis using fluorescence microscopy, with an excitation wavelength of approximately 430 nm. The isolate showing highest granularity, following six days' analysis, were then selected for further analysis, as this was expected to produce the highest PHB yield. Microscope images comparing granularity, are presented in Figure 2.9. Sample p2DPT-RKS(U) 303.4 (gal forward primer, 303 reverse primer, sample 4) appeared to contain large quantities of PHB.

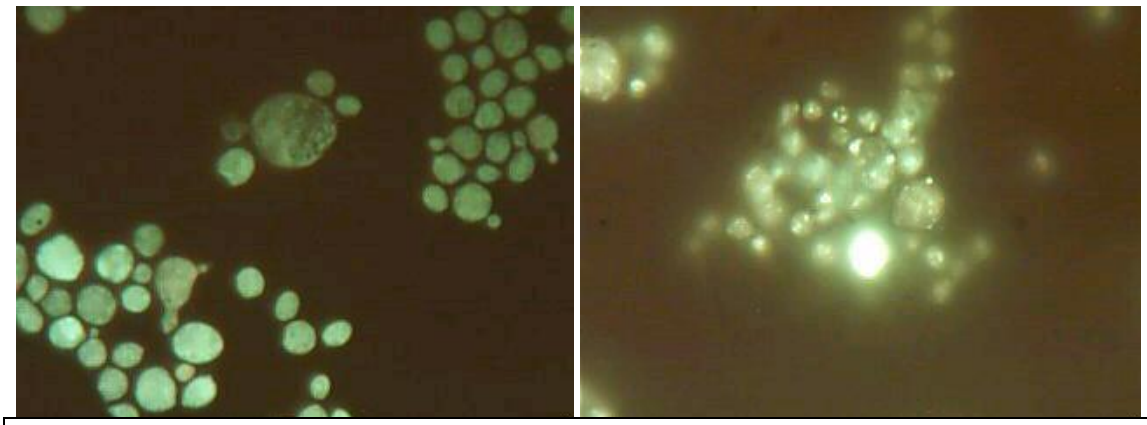


Figure 2.10: Comparing different isolates of H1246 p2DPT-RKS(U) after 6 days' cultivation on 0.1%/1% glc/gal SD dropout plates. Presumably, the isolate in panel B appears to be expressing PHB, while that in panel A appears like wild-type H1246.

Following this promising result, this strain was cultivated in 250 mL shake flasks on 1%/1% glc/gal SD dropout medium, using the previously described D603 p2DPT-

RK(U) p2DP-S(H) as a positive PHB producing control (37). While negative controls were supplied by D603 and H1246 wild-type cultures in medium supplemented accordingly to facilitate growth on minimal medium. The experiment setup is outlined in Table 6 below. Samples -303.4a and -303.4b are duplicates. Relative PHB accumulation as determined by cell dry weight (CDW).

Table 2.3: Experimental samples for PHB analysis from shake flask

SD 1% glc, 1% gal
H1246
D603
D603 p2DPT-RK(U) p2DP-S(H)
H1246 p2DPT-RKS(U)303.4a
H1246 p2DPT-RKS(U) 303.4b

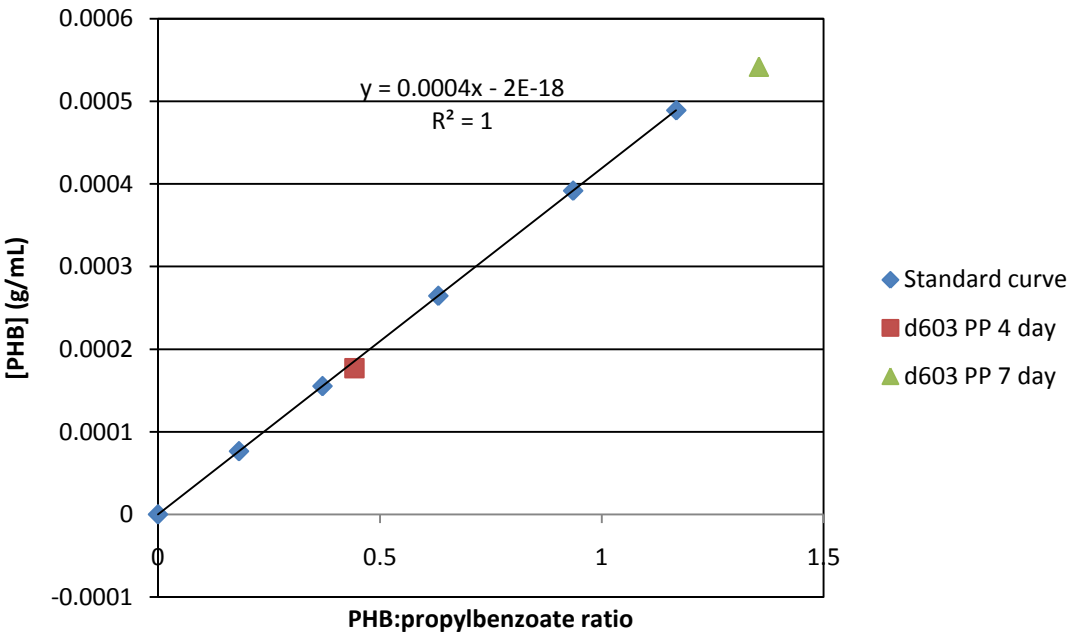
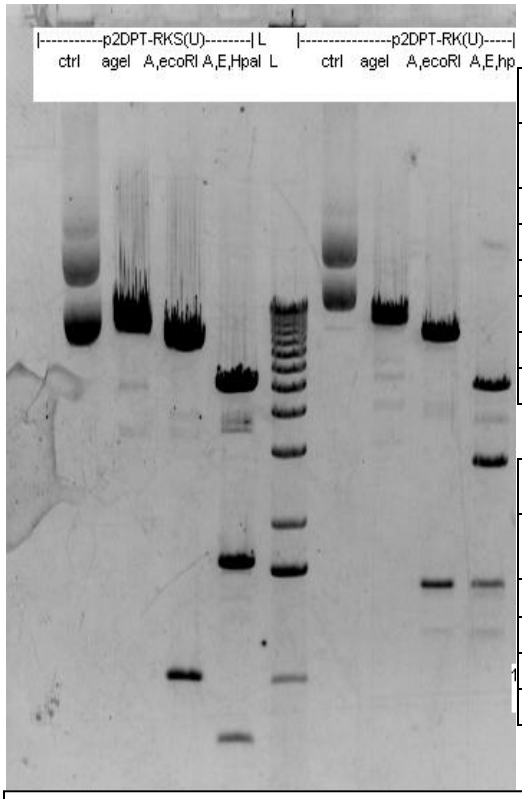


Figure 2.11: Propanalysis standard curve and production results from shake flask comparison of H1246 p2DPT-RKS(U) 303.4 strain, and positive and negative controls. Only samples with measurable PHB content have been included in this graph. The only strain with measured PHB accumulation was the positive control, D603 p2DPT-RK(U) p2DP-S(H), which accumulated PHB at 0.9% CDW (t = 4 days) and 1.1% CDW (t = 7 days) incubation.

Unfortunately, no PHB was observed in either H1246 p2DPT-RKS(U) 303.4 sample, suggesting either poor expression of PHB genes, or incorrect plasmid construction. To determine between these possibilities, *E. coli* Subcloning Efficiency DH5α samples (Invitrogen) were transformed with yeast-harvested plasmid using the procedure described in section 2.2.16. DNA was then extracted from these cells by miniprep, and analyzed by restriction digestion using AgeI, EcoRI, and HpaI enzymes.



Expected fragments (KB)						
fragment number	p2DPT- RKS(U)			p2DPT- RK(U)		
1	AgeI	A,ecoRI	A,E,hpaI	ageI	A,ecoRI	A,E,hpaI
2	591	591	591	9221	1494	1494
3	4865	1494	1494		7727	2845
4	5908	3371	1700			4882
5		5908	3371			
6			4208			

observed fragments (KB)						
fragment number	p2DPT- RKS(U)			p2DPT- RK(U)		
1	AgeI	A,ecoRI	A,E,hpaI	AgeI	A,ecoRI	A,E,hpaI
2	9	1	0.8	9	1.5	1.5
3		8	2		8	2.8
4			5			5

Figure 2.12: Restriction digestion of PHB plasmids using the AgeI, EcoRI, and HpaI restriction enzymes. The lack of ageI digestion site is especially relevant, in that successful Gal1-10-phaC incorporation would add two cut sites, while experimental results show only a single cut by this enzyme in both plasmids. However, the significant difference between each plasmid was observed, indicating some degradation occurred, not simply self-ligation of the host vector.

The results shown in Figure 2.11 demonstrate that the plasmid p2DPT-RKS(U) 303.4 does not match the expected sequence. The observed sequence does not match the

host vector either, reducing the likelihood that this is simply the result of a self-annealing vector during the initial transformation of linear components. Rather, this result suggests undesired intra-plasmid recombination, likely the result of instability brought on by the either the large plasmid size or the presence of ~600 bp of identical sequence in GAL promoters on opposite sides of the plasmid. While either of these factors may be involved in reducing plasmid stability, it is likely that this is the result of successfully transformed strains expressing PHB under galactose induced conditions, and that significant selection pressure was placed on these cells by PHB production, leading to the propagation of population members lacking complete PHB pathway expression when cultures were selected for propagation. Future work involving any system defined by non-wild-type gene expression should involve duplicate plating on both expressing and non-expressing media, to prevent selection against gene expression.

2.3.3: Construction of a single vector for PHB expression in yeast: homologous recombination around the AatII site using *E. coli* mediated ligation.

E. coli mediated ligation has been shown previously to be effective, with success rates of approximately 49% under reaction conditions similar to those described in section 2.3.11 (54). This method could potentially allow plasmid construction more quickly than is possible for yeast-mediated ligation, due to the growth kinetics of *E. coli*. Eight samples from each ligation reaction were analyzed using the homologous cloning primers listed in Table 2.2 to select for samples with the GAL10-phaC cassette in the miniprep DNA solution. Primers used in these reactions are listed in Table 2.2. The results are presented below in Figure 2.12.

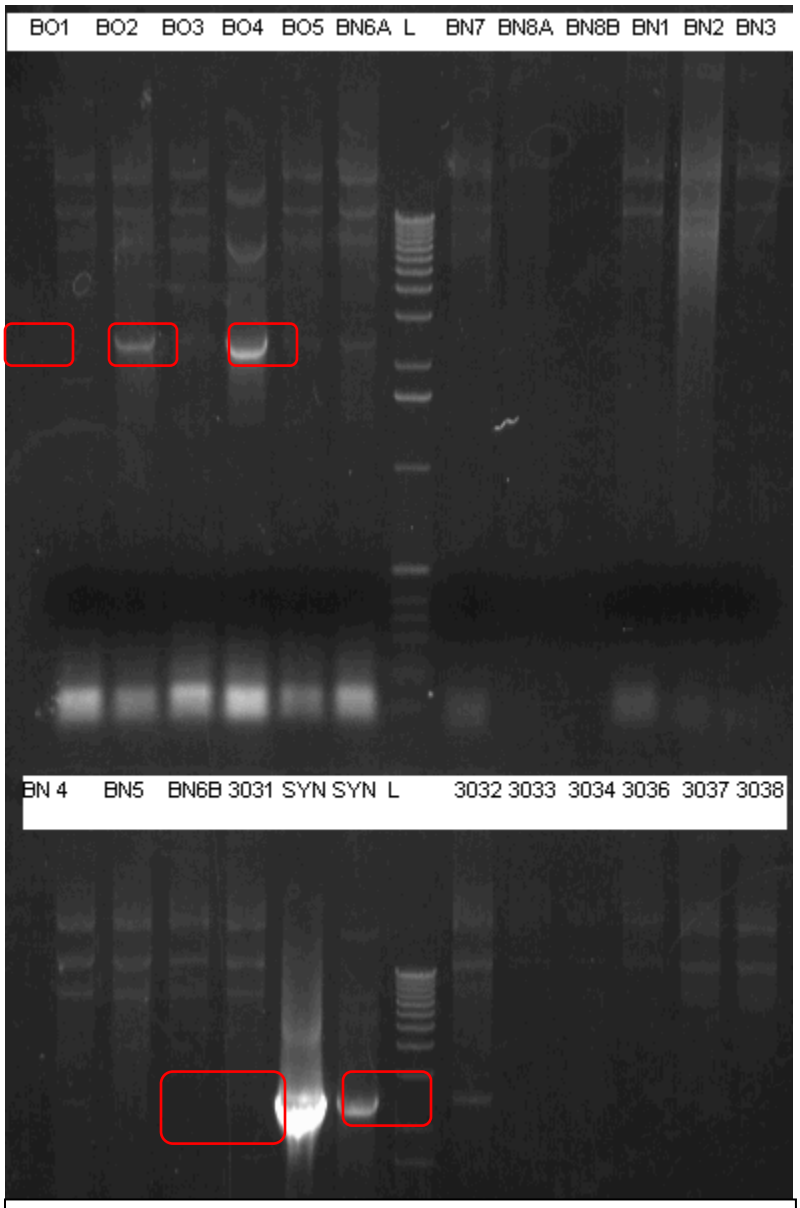
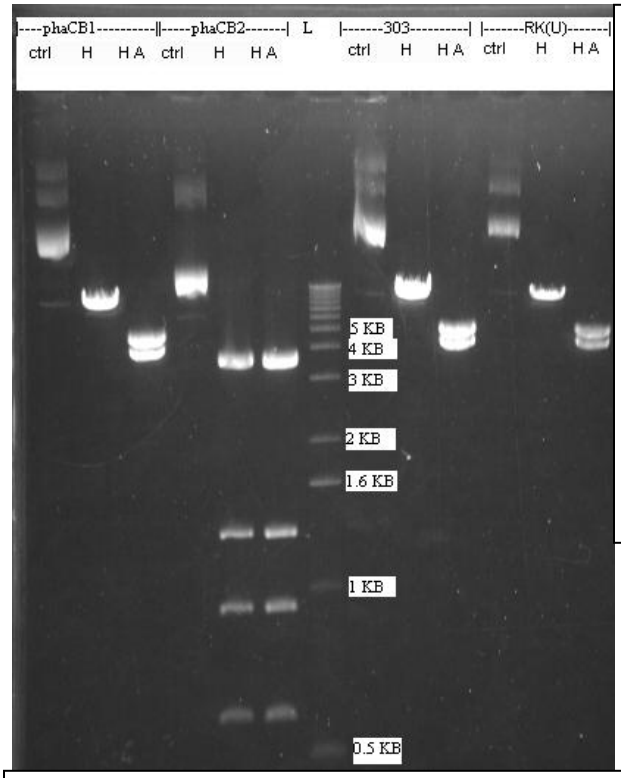


Figure 2.13: PCR screen results of *E. coli* mediated ligation showing amplification of the GAL1-10-phac region in three of 24 test samples and the positive control. A 2.3 KB plasmid was expected; samples showing this band have been highlighted with red boxes.

These results indicated the presence of the desired insert, GAL1-10-*phaC*, in three of the twenty-four samples above. However, because of the sensitivity of PCR, these

result may lead to false positives. The best choice of primers to verify plasmid identity would have been a combination of inside: outside primers. However, plasmid identification can also be performed using restriction digestion, and this route was chosen over PCR-based techniques for these three samples. The result of hindIII and aatII digestion reactions are presented in Figure 2.13, showing patterns not consistent with GAL1-10-phaC cassette insertion in any sample.



p2DPT-RK(U) HindIII, AatII digest			
expected pattern	**observed pattern**		
4350	4300		
5038	5000		
p2DPT-RKS(U)			
expected pattern	**observed pattern**		
	phaCB#1	**phaCB#2**	**303**
5300	4300	4300	3500
6000	5000	5000	1300
			900
			600

Figure 2.14: Restriction map of p2DPT-RKS(U) isolates generated using *E. coli* mediated homologous recombination. The control plasmid matched expected results, while no test plasmid exhibited the expected banding pattern, supporting the conclusion of unsuccessful plasmid construction.

The restriction map shown in Figure 2.13 shows that none of the previously identified samples match the desired restriction map for p2DPT-RKS(U) isolates. This negative result, as well as the large variation in plasmid structure, lead directly to the

conclusion that homologous recombination mediated by *E. coli* strain DH5α is not reliable, and was not attempted further.

2.3.4: Construction of a single vector for PHB expression in yeast: blunt-end ligation at the sstI restriction digest site of p2DPT-RK(U).

To remove possible interference caused by homologous recombination, the TEF1 constitutive promoter was chosen for expression of the *phaC* gene in the final three gene PHB plasmid. A construct containing a TEF1-*phaC*, as well as a TEF1-*phaC1* (*P. oleovorans* PHA synthase) cassette were constructed previously, contained in the plasmids pNTG1-S500(H), pNTG1-S700(H) (55). These cassettes were used in all subsequent plasmid construction, which is diagrammed in Figure 2.14.

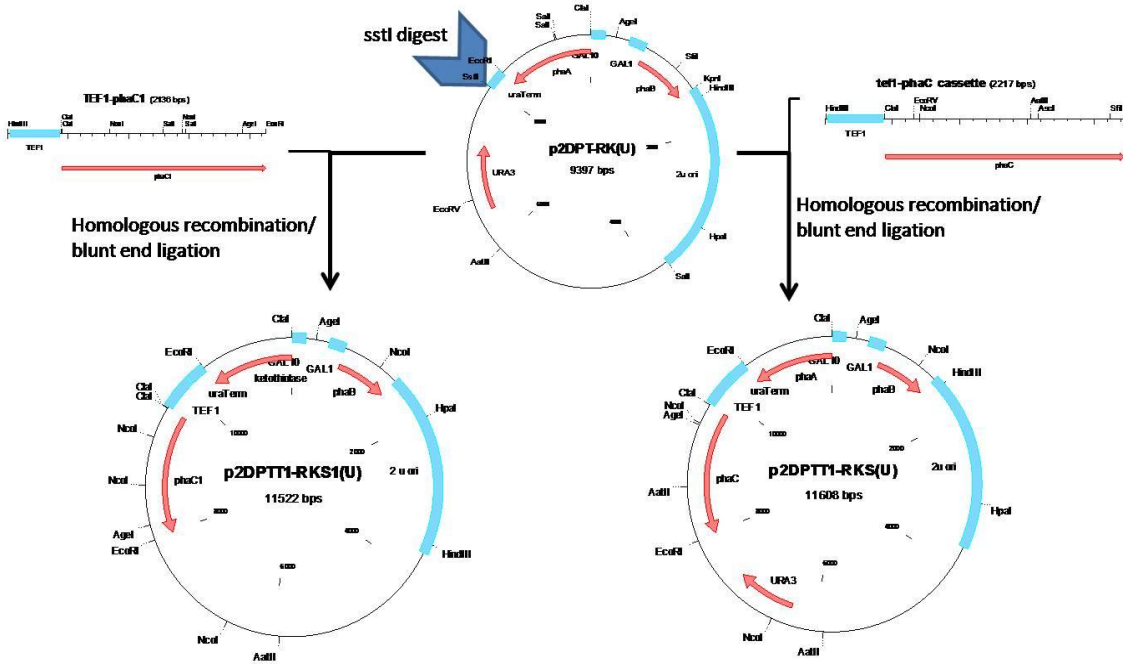


Figure 2.15: Plasmid construction inserting TEF1-*phaC* (*R. eutropha* synthase) and the TEF1-*phaC1* (*P. oleovorans* synthase) cassettes into p2DPT-RK(U) following SstI digest of this vector. The resulting plasmid would contain the two *R. eutropha* genes which convert acetyl-coA to 3-(R)-hydroxybutyrate, and either the *P. oleovorans* or *R. eutropha* polymer synthase genes expressed by the constitutive TEF1 promoter.

Blunt end ligation was attempted to generate these plasmids. Miniprep-purified vector samples were digested overnight with sstI, and complete digestion was verified by gel electrophoresis of a sample (data not shown). Similarly, insert samples were digested using HindIII and EcoRI in the appropriate buffer overnight to excise the TEF1-*phaC(1)* cassettes. All samples were then purified by phenol:chloroform extraction and ethanol precipitation. DNA digested with SstI was then treated with T4 DNA polymerase to blunt the 3'→5' overhangs produced by SstI, while inserts generated with EcoRI and HindIII were similarly treated with Klenow polymerase in excess dNTP, to repair the 5'→3' overhangs produced by these enzymes. The vector was then treated with Calf Intestine Alkaline Phosphatase (CIAP) to prevent self ligation. These prepared fragments were then run on a 0.7% agarose gel, the appropriate fragments excised and extracted by gel purification. These fragments were then combined in a 1:2 vector: insert molar ratio (176 ng insert, 400 ng vector). In addition, the re-ligation of pNTG1-S500(H) from EcoRI, HindIII digests served as a positive control, showing success of insert ligation with "sticky" ends. Additional control reactions were performed, testing the self ligation of p2DPT-RK(U) without dephosphorylation. The transformation results of these ligations are presented in Figure 2.15 below, which shows the expected pattern of transformation efficiencies, and colony formation in overnight blunt-end ligation to create p2DPTT1-RKS(U).

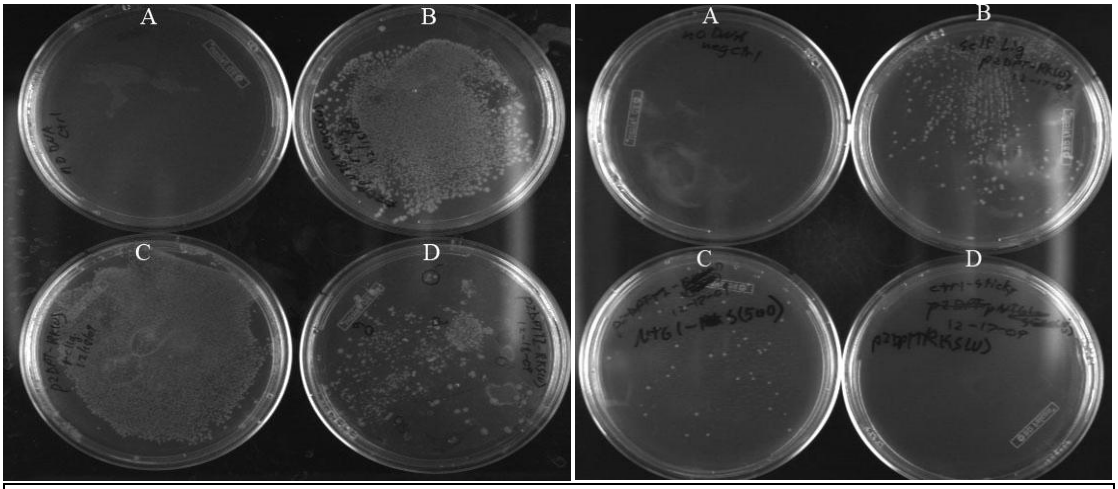


Figure 2.16: Panel 1: Overnight ligation, Invitrogen cells. A: no DNA control, B: pNTG1-S500(H) re-ligation "sticky end" control, C: self-ligation, non-dephosphorylated, D: Blunt end ligation, p2DPTT1-RKS(U). Panel 2: Invitrogen cells, 1 hour ligation. A: no DNA control, B: pNTG1-S500(H) re-ligation "sticky end" control, C: self-ligation, non-dephosphorylated, D: Blunt end ligation, p2DPTT1-RKS(U). Panel 2: Invitrogen cells, 1 hour ligation

Taken together, 1 hr and overnight ligation methods are successful in stick-end ligations (plates B and C), while Invitrogen cells and overnight ligations appear to be required for blunt end ligation. The large difference in colony formation between plates C and D suggest effective dephosphorylation, although further analysis (discussed below) show that the blunt-end ligation plate actually contains only self-ligated vector, meaning incomplete dephosphorylation must be

Six colonies were then picked arbitrarily from the p2DPTT1-RKS(U) blunt ligation plate with plasmid DNA obtained from overnight cultures by miniprep purification. To assess ligation success, isolates were digested with NcoI, which was expected to give the following band sizes for test and control plasmids (Table 2.4).

Table 2.4: Expected NcoI digestion bands	
p2DPT-RK(U) ctrl	**p2DPT-RKS(U)**
5.4	**5.4**
4	**2.9**
	3.3
Observed pattern-all samples	
5.4	
4	

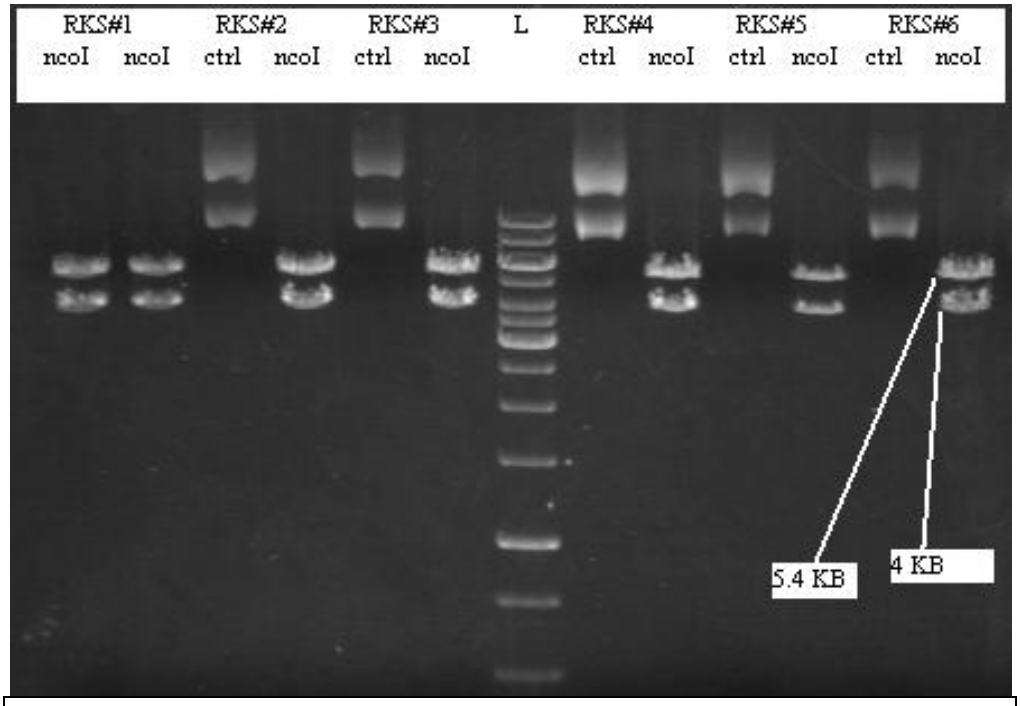


Figure 2.17: Restriction map characterization of p2DPT-RKS(U) blunt-end ligation products using NcoI digestion. This figure shows two things: all ligation products are identical, and all ligation products are digested in the pattern expected for p2DPT-RK(U) (two bands at 5.4 and 4 KB). As a result, all blunt-end ligation products have been determined to be the self-ligated vector.

Based on the restriction digestion result shown in Figure 2.16, it was concluded that blunt-end ligation was not successful, and that all transformed isolates in this sample were both identical and the result of self-ligation. It is apparent that vector dephosphorylation occurred, but was incomplete. Dephosphorylation appears to have

occurred sufficiently to eliminate self-ligation in 1 hour ligation reactions, but not to prevent all self-ligation in overnight ligations.

For this experiment to be repeated successfully, new samples of CIAP are expected to be necessary. In addition, controls should be used to assess success at every step in the ligation. Furthermore, efficiency can be increased by minimizing the need for solvent purification by performing compatible reactions in one reaction buffer, with heat inactivation between enzyme steps. For example, if digestion is performed in NEB buffer 2 (New England Biolabs,Ipswich, MA), digestion, blunting with T4 polymerase or Klenow (large fragment DNA polymerase I) polymerase, and vector dephosphorylation can all be performed in the same buffer, with 70-75°C heat inactivation between each step (and EDTA addition following blunting). Following these digestions, samples should then be gel purified, excised, and combined in 1:3 to 1:5 vector: insert molar ratios for overnight ligation at 15°C, followed by transformation of chemically competent DH5α stocks using the heat-shock method described in Section 2.3.8. Controls should include, at minimum, the transformation of DH5α strains described in Table 2.5.

Table 2.5: Appropriate test and control transformations for any blunt-end ligation creation and assessment experiment

Reaction Number	Description	Purpose
C1	No DNA	Negative control to verify selection plates
C2	Dephosphorylated vector	Negative control to verify dephosphorylation reaction
C3	Non-dephosphorylated vector	Positive control to assess dephosphorylation
C4	Re-ligation of doubly-digested vector-insert	Positive control to compare sticky and blunt ligation efficiency
T1	Blunt ligation-3:1 insert:vector ratio	Test reaction
T2	Blunt ligation-5:1 insert:vector ratio	Test reaction

2.3.5: Construction of a single vector for PHB expression in yeast: ligation at the sstI restriction digest site of p2DPT-RK(U) using yeast mediated ligation.

Yeast mediated vector construction was utilized to insert the TEF1-*phaC* and TEF1-*phaC1* promoter-synthase cassettes into the host vector p2DPT-RK(U) in order to yield the PHB producing p2DPT-RKS(1)(U), as diagrammed in Figure 2.14. These primers, shown in Table 2.6, were used to clone the TEF1-*phaC* cassettes from the vectors pNTG-S700(H) and pNTG1-S500(H), although only pNTG1-S700(H) should contain *the phaC* gene itself.

Table 2.6: Primers used in homologous recombination reactions

Primer ID	Description	Length (bp)	Target bp on p2DPT-RK(U), relative to sstII digestion site	Sequence [cloning target in brackets]
TEF1-homoUraTerm_for	forward primer, from hindIII of TEF1, homo to uraterm region	60	+1 to +36	AAAGAATAAAAAAAAAATGATGAATTGAAAAGGTG [GAGCTTCTTAAAGCTTCTAGCTGAG]
PhaC-homoSstI_rev	reverse primer, phaC end		-1 to -39	CCAGTGAATTGTAATACGACTCACTATAGGGCGAATTG [CATGCCTTGGCTTTGACGTATC]

Successful amplification of a 2300 bp region was only achieved using these primers with pNTG1-S500(H), which contains only the *P. oleovorans phaC1* gene. Further analysis of plasmids created by homologous recombination support two conclusions. First, homologous recombination was successful in generating a very large plasmid, at 11.6 KP that was stable in both *S. cerevisiae* and *E. coli*. This plasmid showed no noticeable deviation from the expected sequence after 5 days selective growth on SD plates, and overnight culture in liquid SD dropout medium, miniprep isolation and

E. coli transformation, and finally two overnight *E. coli* cultivations on solid and liquid LB + ampicillin medium. As a result, we can conclude that such a plasmid can be created and maintained for time spans needed for research purposes. In combination with the plasmid instability suggested in Section 2.3.2 when generating a large plasmid with two GAL1-10 promoter sequences, this result suggests that the use of the TEF1 promoter in place of the GAL1-10 sequence for *phaC(1)* expression has resulted in significant improvements in plasmid stability.

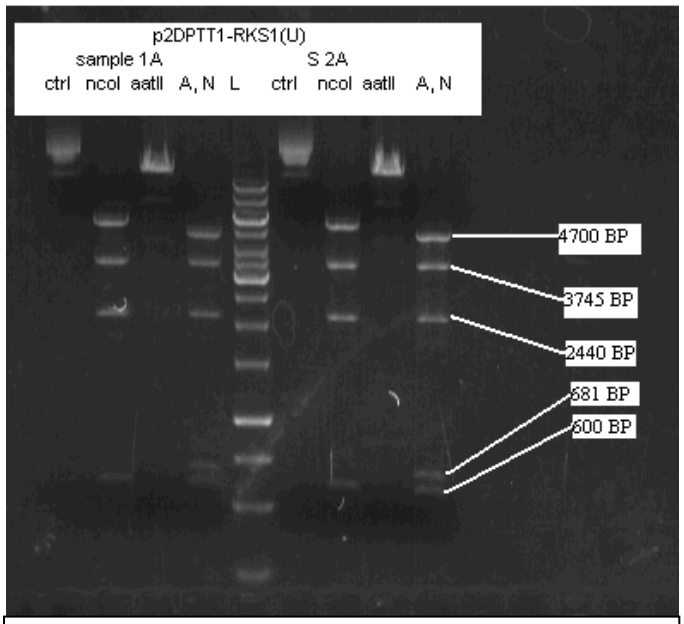


Figure 2.18: Restriction map of two p2DPT-RKS1(U) isolates. All samples show identical banding patterns consistent with *phaC1* amplification using *phaC* primers.

Table 2.7: Expected and observed digestion fragments from yeast mediated ligation.

p2DPT-RK(U) ncoI, aatII digest					
expected pattern			observed pattern		
AatI	**NcoI**	**AatII, NcoI**	**AatI**	**NcoI**	**AatII, NcoI**
9397	5420	4735			
	3977	3977			
		685			
p2DPT-RKS1(U)					
expected pattern			observed pattern		
NcoI	**AatII**	**AatII, NcoI**	**NcoI**	**AatII**	**AatII, NcoI**
5420	11522	4735	5400		4700
3438		3438	3745	Larger than	3745
2067		2067	2440	ladder.	2440
597		685	600		681
		597			600
p2DPT-RKS(U)					
expected pattern					
NcoI	**AatII**	**AatII, NcoI**			
5420	5917	4735			
2892	2935	2114			
2114	2756	2071			
1182		1182			
		821			
		685			

Second, the restriction map presented in Figure 2.17 and Table 2.6 show an unexpected result, that the TEF1- *R. eutropha phaC* primers used in cloning the same region from the plasmids pNTG1-S500(H) and pNTG1-S700(H) resulted in no amplification from the expected vector, pNTG1-S700(H), while successful cloning of the TEF1-*phaC1* gene from pNTG1-S500(H) was observed. In addition, the restriction pattern presented in Figure 2.17 confirms that the TEF1-*phaC1* cassette was inserted into the linearized p2DPT-RK(U) plasmid, creating the p2DPTT1-RKS1(U) plasmid shown in Figure 2.10. The question of why TEF1-*phaC1* cloning was successful, while the targetted cassette TEF1-*phaC* was not, must be addressed.

The most likely explanation for this observed amplification pattern is that DNA synthesis was initiated not by binding of the reverse primer to the final 22 bases of the

phaC gene as expected (Figure 2.18), but instead by the homologous recombination sequence 54- 92 bp downstream from the end of the *phaC* gene. If this is correct, the resulting cloned region should then contain bases + 92 to + 54 of the original vector, while the sequence from bases +53 to + 31 would match the final 22 base pairs of *phaC*. Most importantly, the *phaC1* gene observed in the final construct p2DPT-RKS1(U) is expected to be intact. This hypothesis must be confirmed with sequencing or a repeated cloning of the TEF1-*phaC(1)* cassette using only the 38 bp homologous recombination sequence shown in the Table 2.6.

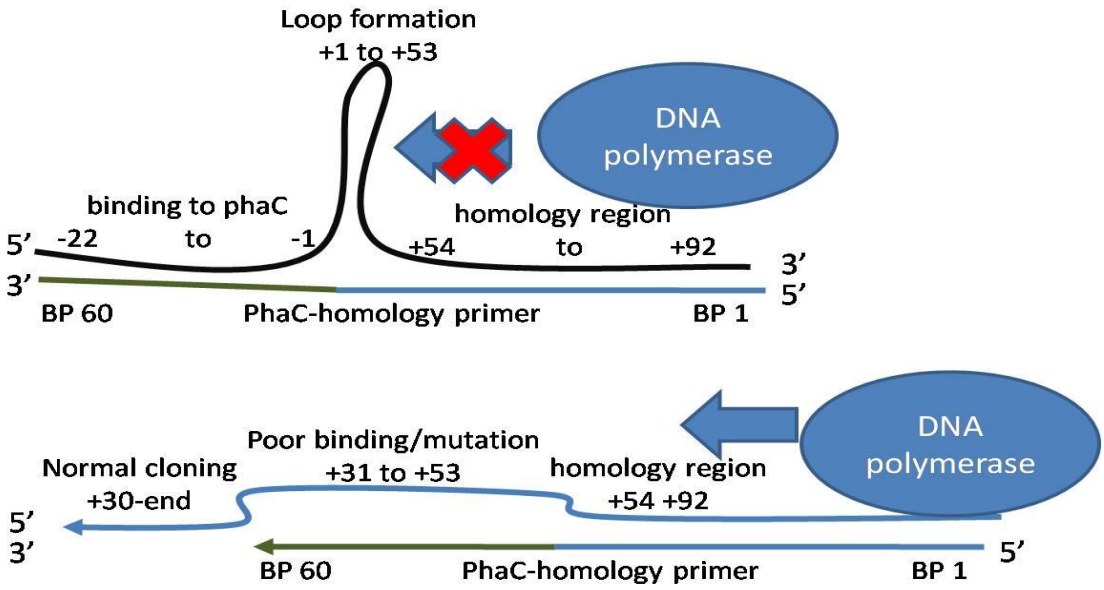


Figure 2.19: Promoter-target sequence interaction hypothesis describing observed cloning results. Cloning of the TEF1-*phaC* region is inhibited by the perfect binding of both *phaC* target regions *and* homologous recombination sequences, while *phaC1* cloning is successful, with the final 22 bp of *phaC* being substituted for the +31 to +54 bp sequence downstream of the *phaC1* stop codon. This mutation is not expected to affect protein expression or function, as it would occur downstream of the coding sequence.

The homologous recombination sequence targets the region 54-92 bp downstream of the end of the *phaC* gene. Because all plasmids are derivatives of the pRS300 series of

plasmids generated by Sikorsky & Heiter (1989), they all share this region of homology, between the sstI restriction site and the plasmid's selection marker, whether the *his2* or the *ura3* gene. If cloning begins from the homologous recombination region instead of the *phaC* region, no difference in gene function would be expected, as the final construct would simply contain an additional 92 bp at the trailing end of the *phaC* gene. What was surprising, however, was that successful construction was achieved with *phaC1* amplification, and not *phaC* as well, as both plasmids contain this homology sequence. One possible reason for this counter-intuitive result could be that, due to the perfect homology between the +92-54 region, followed by the 22 bp homology between the -1 to -22 region of the *phaC* gene, primer binding resulted in DNA loop formation, preventing DNA polymerase from initiating effectively, as diagrammed in Figure 2.18. Contrarily, because of the lack of a competing, perfect homology region in plasmids containing *phaC1*, perfect binding to the +92 to +54 bp region is accompanied by imperfect binding in the +53 to +31 bp region, followed by normal sequence cloning from +30 to the end of the cloned region at the start of the TEF1 promoter, resulting in the successful cloning of the TEF1-*phaC1*cassette, and its insertion into the host vector, creating the p2DPTT1-RKS1(U) plasmid, which contains the *R. eutropha* PHB ketothiolase and reductase encoding genes, while also containing the *P. oleovorans* PHA synthase enzyme.

2.4: Project Summary and Future Work

The results presented in the previous chapter have been included to illustrate several methods for plasmid construction, and the difficulties experienced from implementing each method. In addition, She successful plasmid construction presented in Section 2.3.5 is of by far greatest importance for several reasons. First, the results

presented demonstrate that stable plasmids of this size can be created and maintained, and suggest that the removal of homologous regions within the plasmid results in increased stability, as expected. Furthermore, this plasmid, and others created using the same techniques, may be directly applied to several experiments, with great potential for both improving our understanding of PHA accumulation in yeast through fluorescence microscopic analysis, PHA accumulation comparison using different PHA synthase genes. In addition, due to plasmid segregation kinetics (Section 2.1.2), strains expressing the PHA production pathway using this plasmid are expected to exist in a more homogenous population, with PHA accumulation higher than that previously observed in two plasmid systems due an increase in both plasmid bearing cells, and cells containing a complete and proportionally expressed PHA anabolic pathway.

Future work should concentrate first on sequencing the region downstream of *phaC1* in the plasmid p2DPTT1-RKS1(U), to verify the hypothesis presented in Figure 2.12 using a primer containing only the 38 bp of homology +92 to +54 downstream of the *phaC1* gene. Furthermore, this same primer can be used, regardless of the sequencing results above, to clone bothTEF1- *phaC(1)* regions using identical primers for plasmid construction using yeast-mediated ligation quickly and easily. Once created, these plasmids contain different PHA synthase genes, but would otherwise be identical, while containing a complete PHA synthesis pathway. Further analysis is necessary to establish the effect of expressing *phaC(1)* constitutively, while expressing *phaA* and *phaB* using inducible galactose expression. It is highly likely that constitutive *phaC(1)* expression

will be significantly higher than that of the other pathway enzymes, potentially resulting in inclusion bodies or other detrimental side effects within target cells.

If stable PHB production can be achieved, the transformation of the lipid accumulation-negative *S. cerevisiae* strain H1246 with this single plasmid would facilitate unencumbered observation of PHA accumulation within yeast cells, shedding light on important physiological processes involved in cytoplasmic expression of this and other biotechnology related metabolic pathways.

References

1. Hau HH, Gralnick JA. Ecology and biotechnology of the genus *Shewanella* . Annu Rev Microbiol. 2007;61:237-58.

2. Hau HH, Gilbert A, Coursolle D, Gralnick JA. Mechanism and consequences of anaerobic respiration of cobalt by *Shewanella oneidensis* strain MR-1. Appl Environ Microbiol. 2008 Nov;74(22):6880-6.

3. Hartshorne RS, Jepson BN, Clarke TA, Field SJ, Fredrickson J, Zachara J, et al. Characterization of *Shewanella oneidensis* MtrC: A cell-surface decaheme cytochrome involved in respiratory electron transport to extracellular electron acceptors. J Biol Inorg Chem. 2007 September;12(7):1083-94.

4. Trinh CT, Unrean P, Srienc F. Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. Appl Environ Microbiol. 2008 June; 74(12):3634-43.

5. Neidhardt FC, Ingraham JL, Schaechter M. Biosynthesis and fueling. In Physiology of the Bacterial Cell; 1990. p. 133-73.

6. Tang YJ, Hwang JS, Wemmer DE, Keasling JD. *Shewanella oneidensis* MR-1 fluxome under various oxygen conditions. Appl Environ Microbiol. 2007 Feb; 73(3):718-29.

7. Tang YJ, Meadows AL, Keasling JD. A kinetic model describing *shewanella oneidensis* MR-1 growth, substrate consumption, and product secretion. Biotechnol Bioeng. 2007 January 1; 96(1):125-33.

8. Tang YJ, Meadows AL, Kirby J, Keasling JD. Anaerobic central metabolic pathways in *Shewanella oneidensis* MR-1 reinterpreted in the light of isotopic metabolite labeling. J Bacteriol. 2007 February; 189(3):894-901.

9. Schuster S, Hilgetag C, Woods JH, Fell DA. Reaction routes in biochemical reaction systems: Algebraic properties, validated calculation procedure and example from nucleotide metabolism. J Math Biol. 2002 August; 45(2):153-81.

10. Klamt S, Stelling J. Two approaches for metabolic pathway analysis? Trends Biotechnol. 2003 February; 21(2):64-9.

11. Wlaschin AP, Trinh CT, Carlson R, Srienc F. The fractional contributions of elementary modes to the metabolism of *Escherichia coli* and their estimation from reaction entropies. Metab Eng. 2006 July; 8(4):338-52.

12. Llaneras F, Pico J. Stoichiometric modelling of cell metabolism. J Biosci Bioeng. 2008 January; 105(1):1-11.

13. Pfeiffer T, Sanchez-Valdenebro I, Nuno JC, Montero F, Schuster S. METATOOL: For studying metabolic networks. Bioinformatics. 1999 March; 15(3):251-7.

14. Trinh CT, Carlson R, Wlaschin A, Srienc F. Design, construction and performance of the most efficient biomass producing *E. coli* bacterium. Metab Eng. 2006 Nov; 8(6):628-38.

15. Trinh CT, Unrean P, Srienc F. Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. Appl Environ Microbiol. 2008 June; 74(12):3634-43.

16. Carlson R, Srienc F. Fundamental *Escherichia coli* biochemical pathways for biomass and energy production: Creation of overall flux states. Biotechnol Bioeng. 2004 April 20; 86(2):149-62.

17. Unrean P, Trinh CT, Srienc F. Rational design and construction of an efficient *E. coli* for production of diapolycopendioic acid. Metab Eng. 2010 March; 12(2): 112-122.

18. Knappe J, Sawers G. A radical-chemical route to acetyl-CoA: The anaerobically induced pyruvate formate-lyase system of *Escherichia coli*. FEMS Microbiol Rev. 1990 August; 6(4):383-98.

19. Hunt K, Gralnick J. Personal communication.

20. Scott JH, Nealson KH. A biochemical study of the intermediary carbon metabolism of *Shewanella putrefaciens*. J Bacteriol. 1994 June;176(11):3408-11.

21. Leys D, Tsapin AS, Nealson KH, Meyer TE, Cusanovich MA, Van Beeumen JJ. Structure and mechanism of the flavocytochrome C fumarate reductase of *Shewanella putrefaciens* MR-1. Nat Struct Biol. 1999 December; 6(12):1113-7.

22. Maier TM, Myers JM, Myers CR. Identification of the gene encoding the sole physiological fumarate reductase in *Shewanella oneidensis* MR-1. J Basic Microbiol. 2003; 43(4):312-27.

23. Hartshorne RS, Jepson BN, Clarke TA, Field SJ, Fredrickson J, Zachara J, et al. Characterization of *Shewanella oneidensis* MtrC: A cell-surface decaheme cytochrome involved in respiratory electron transport to extracellular electron acceptors. J Biol Inorg Chem. 2007 September; 12(7):1083-94.

24. Shi L, Squier TC, Zachara JM, Fredrickson JK. Respiration of metal (hydr)oxides by *Shewanella* and *Geobacter*: A key role for multihaem c-type cytochromes. Mol Microbiol. 2007 July; 65(1):12-20.

25. Hayashi M, Miyoshi T, Takashina S, Unemoto T. Purification of NADH-ferricyanide dehydrogenase and NADH-quinone reductase from *Escherichia coli* membranes and their roles in the respiratory chain. Biochim Biophys Acta. 1989 October 26; 977(1):62-9.

26. Matsushita K, Ohnishi T, Kaback HR. NADH-ubiquinone oxidoreductases of the *Escherichia coli* aerobic respiratory chain. Biochemistry. 1987 December 1; 26(24):7732-7.

27. Calhoun MW, Gennis RB. Demonstration of separate genetic loci encoding distinct membrane-bound respiratory NADH dehydrogenases in *Escherichia coli*. J Bacteriol. 1993 May; 175(10):3013-9.

28. Hunt K, Naranjo B, Flynn JM, Gralnick J. Substrate level phosphorylation is the major source of energy during anaerobic growth on *Shewanella oneidensis* MR-1. In press 2010.

29. Caspi R, Foerster H, Fulcher CA, Hopkinson R, Ingraham J, Kaipa P, et al. MetaCyc: A multiorganism database of metabolic pathways and enzymes. Nucleic Acids Res. 2006 January 1; 34(Database issue):D511-6.

30. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. Nucleic Acids Res. 2004 January 1; 32(Database issue):D277-80.

31. Pinchuk GE, Rodionov DA, Yang C, Li X, Osterman AL, Dervyn E, et al. Genomic reconstruction of *Shewanella oneidensis* MR-1 metabolism reveals a previously uncharacterized machinery for lactate utilization. Proc Natl Acad Sci U S A. 2009 February 24; 106(8):2874-9.

32. Serres MH, Riley M. Genomic analysis of carbon source metabolism of *Shewanella oneidensis* MR-1: Predictions versus experiments. J Bacteriol. 2006 July; 188(13):4601-9.

33. DiChristina TJ, Bates DJ, Burns JL, Dale JR, Payne AN. *Shewanella*: Novel strategies for anaerobic respiration. In: Neretin LN, editor. Past and Present Water Column Anoxia. 1st ed. Netherlands: Springer; 2006. p. 443-55.

34. Flynn J.M. Engineering with electricigens: Hardwiring *Shewanella* [dissertation]; 2009.

35. Rehm BH. Polyester synthases: Natural catalysts for plastics. Biochem J. 2003 November 15; 376(Pt 1):15-33.

36. Porro D, Sauer M, Branduardi P, Mattanovich D. Recombinant protein production in yeasts. Mol Biotechnol. 2005 November; 31(3):245-59.

37. Carlson R, Srienc F. Effects of recombinant precursor pathway variations on poly[(R)-3-hydroxybutyrate] synthesis in *Saccharomyces cerevisiae*. J Biotechnol. 2006 July 25; 124(3):561-73.

38. Rehm BH, Antonio RV, Spiekermann P, Amara AA, Steinbuchel A. Molecular characterization of the poly(3-hydroxybutyrate) (PHB) synthase from *Ralstonia eutropha*: In vitro evolution, site-specific mutagenesis and development of a PHB synthase protein model. Biochim Biophys Acta. 2002 January 31; 1594(1):178-90.

39. McChalicher CW, Srienc F. Investigating the structure-property relationship of bacterial PHA block copolymers. J Biotechnol. 2007 November 1; 132(3):296-302.

40. Carlson R. Metabolic pathway analysis of prokaryotic and eukaryotic organisms [dissertation]. 2003.

41. Christianson TW, Sikorski RS, Dante M, Shero JH, Hieter P. Multifunctional yeast high-copy-number shuttle vectors. Gene. 1992 January 2; 110(1):119-22.

42. Futcher AB, Cox BS. Copy number and the stability of 2-micron circle-based artificial plasmids of *Saccharomyces cerevisiae*. J Bacteriol. 1984 January; 157(1):283-90.

43. Kacmar J, Zamamiri A, Carlson R, Abu-Absi NR, Srienc F. Single-cell variability in growing *Saccharomyces cerevisiae* cell populations measured with automated flow cytometry. J Biotechnol. 2004 April 29; 109(3):239-54.

44. Sandager L, Gustavsson MH, Stahl U, Dahlqvist A, Wiberg E, Banas A, et al. Storage lipid synthesis is non-essential in yeast. J Biol Chem. 2002 February 22; 277(8):6478-82.

45. Da Silva, N.A., Bailey, J.E. Effects of inducer concentration on GAL regulated cloned gene expression in recombinant *Saccharomyces cerevisiae*. J Bacteriol. 1989(10):253-66.

46. Tu J, Carlson M. REG1 binds to protein phosphatase type 1 and regulates glucose repression in *Saccharomyces cerevisiae*. EMBO J. 1995 December 1; 14(23):5939-46.

47. Gancedo JM. Yeast carbon catabolite repression. Microbiol Mol Biol Rev. 1998 Jun;62(2):334-61.

48. Johnston M. A model fungal gene regulatory mechanism: The GAL genes of *Saccharomyces cerevisiae*. Microbiol Rev. 1987 December; 51(4):458-76.

49. Johnston M, Davis RW. Sequences that regulate the divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. Mol Cell Biol. 1984 August; 4(8):1440-8.

50. Srienc F, Campbell J, Bailey J. Flow cytometry analysis of recombinant *Saccharomyces cerevisiae* populations. Cytometry. 1986; 7:132-41.

51. Matsumoto K, Yoshimatsu T, Oshima Y. Recessive mutations conferring resistance to carbon catabolite repression of galactokinase synthesis in *Saccharomyces cerevisiae.*. J Bacteriol. 1983 March; 153(3):1405-14.

52. Ausubel F, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, et al, editors. Short protocols in molecular biology. 3rd ed. John Wiley & Sons; 1995.

53. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and clustal X version 2.0. Bioinformatics. 2007 November 1; 23(21):2947-8.

54. Oliner JD, Kinzler KW, Vogelstein B. In vivo cloning of PCR products in *E. coli*. Nucleic Acids Res. 1993 November 11; 21(22):5192-7.

55. Zhang B. Engineering yeast for production of polyhydroxyalkanoates. 2005 [dissertation].

56. Carlson R, Fell D, Srienc F. Metabolic pathway analysis of a recombinant yeast for rational strain development. Biotechnol Bioeng. 2002 July 20; 79(2):121-34.