A Design of Experiments Approach for Engineering Carbon Metabolism in the Yeast *Saccharomyces cerevisiae*

Submitted by Steven Richard Brown to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Biological Sciences

In September, 2016

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: ……..

Acknowledgements

I dedicate this thesis to Emma without whom this wouldn't have been possible; my greatest hope is that our affection, dedication and drive-to-develop be an inspiration to our children.

Mum, Dad and all of my loving family thank you without bound for your never-ending support.

To Stephen and Tom, I consider myself very lucky to have had such brilliant supervision. I have really enjoyed our time working together and thank you for your suggestion and patience.

Thank you to my colleagues on the Mezzanine particularly Marta, Richard, Dr Lux, Jamie and Alex. Your technical help has been invaluable and sanity preservation measures most appreciated.

Finally, thank you to both my financial sponsor Royal Dutch Shell as well as John, Rob, Dave and Nadim for providing me with the belief and opportunity to continue my development as a scientist.

Abstract

The proven ability to ferment *Saccharomyces cerevisiae* on a large scale presents an attractive target for producing chemicals and fuels from sustainable sources. Efficient and predominant carbon flux through to ethanol is a significant engineering issue in the development of this yeast as a multi-product cell chassis used in biorefineries. In order to evaluate diversion of carbon flux away from ethanol, combinatorial deletions were investigated in genes encoding the six isozymes of alcohol dehydrogenase (ADH), which catalyse the terminal step in ethanol production. The scarless, dominant and counter-selectable *amdSYM* gene deletion method was optimised for generation of a combinatorial ADH knockout library in an industrially relevant strain of *S. cerevisiae*. Current understanding of the individual ADH genes fails to fully evaluate genotype-by-genotype and genotype-by-environment interactions: rather, further research of such a complex biological process requires a multivariate mathematical modelling approach. Application of such an approach using the Design of Experiments (DoE) methodology is appraised here as essential for detailed empirical evaluation of complex systems. DoE provided empirical evidence that in *S. cerevisiae*: i) the *ADH2* gene is not associated with producing ethanol under anaerobic culture conditions in combination with 25 g l^{-1} glucose substrate concentrations; ii) *ADH4* is associated with increased ethanol production when the cell is confronted with a zinc-limited [1 μM] environment; and iii) *ADH5* is linked with the production of ethanol, predominantly at pH 4.5. A successful metabolic engineering strategy is detailed which increases the product portfolio of *S. cerevisiae*, currently used for large-scale production of bioethanol. Heterologous expression of the cytochrome P450 fatty acid peroxygenase from *Jeotgalicoccus* sp., $OleT_{JE}$, fused to the RhFRED reductase from *Rhodococcus* sp. NCIMB 978 converted free fatty acid precursors to C_{13}, C_{15} and C_{17} alkenes (3.81 ng μl^{-1} total alkene concentration).

List of Contents

List of tables

List of figures

List of abbreviations

ADH	Alcohol dehydrogenase
amdSYM	*amdS* (acetamidase) gene yeast marker
BLAST	Basic local alignment search tool
D_{crit}	Critical dilution rate
DoE	Design of experiments
EIC	Extracted ion chromatogram
FACS	Fluorescence activated cell sorter
FBA	Flux balance analysis
GC-FID	Gas chromatography with flame ionisation detector
GC/Q-TOF/MS	Gas chromatography/quadrupole-time of flight/mass spectrometer
GEMs	Genome scale models
GRAS	Generally regarded as safe
HPLC RI	High performance liquid chromatography with refractive index
IEA	International Energy Agency
LB	Lysogeny broth
OFAT	One-factor-at-a-time
ORFs	Open reading frames
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PLS	Partial least squares
PRESS	Predicted residual sum of squares

SD	Synthetic defined
SGD	*Saccharomyces cerevisiae* genome database
SIMPLS	Statistically inspired modification of the PLS method
SM	Synthetic medium
SM Ac	Synthetic medium with acetamide
SM FAc	Synthetic medium with fluoroacetamide
TIC	Total ion chromatogram
VIP	Variable importance for the projection
YEPD	Yeast extract, peptone and dextrose
YNB	Yeast nitrogen base

1 General Introduction

1.1 Replacing petrochemicals

At the beginning of this century global dependence upon oil and the future of energy generation were seen as significant challenges to humanity. The company Royal Dutch Shell plc, a leader of the oil and gas industry, made publicly available “Shell energy scenarios to 2050” (Shell, 2008) which reviewed these challenges and identified “three hard truths”:

i) Step-change in energy use

Industrialisation and increasing use of transportation in developing nations, including the population giants China and India, exacerbate the demand pressures for oil which is predicted to rise from 85 million barrels (1.4×10^7 m^3) per day in 2007 to 104 million barrels (1.7×10^7 m^3) per day by 2030 (IEA, 2008). In late 2015 oil production breached 97 million barrels (1.54×10^7 m^3) per day (IEA, 2016).

ii) Supply will struggle to keep pace

The scenario stated that by 2015, growth in “easily accessible” oil and gas would not match the projected rate of demand, requiring the development of alternative energy sources, such as biofuels, to provide a significant part of the “energy mix”.

iii) Environmental stresses are increasing.

Anthropogenic CO_2 emissions severely threaten human well-being: at current rates it will become increasingly difficult to maintain a CO_2 concentration in the atmosphere within desirable levels. Increasing CO_2 concentrations are associated with increasing global temperatures, and the current globally agreed climate goal is to limit the average global temperature rise to 2 °C, relative to pre-industrial levels. In order to fully implement current pledges, which still fall short of what is required to achieve the globally agreed climate goal, the energy sector alone will require $13.5 trillion investment in low-carbon technologies and efficiencies from 2015 to 2030 (IEA, 2015a).

Though the petroleum-based energy, chemical and fuel industries are economically thriving, concerns such as those detailed by the “three hard truths” and compounded by oil being a finite resource mean these industries also face great socio-political challenges (Lee *et al.*, 2015). Driven by continued policy support, the beginnings of a transformation can be seen, particularly in the energy sector, with the development of renewable technologies such as hydro, wind and solar for power generation. In 2014, renewables accounted for over five times more global electricity generation than oil, and current scenarios anticipate that by 2040 renewables will account for half of additional global power generation, overtaking coal around 2030 to become the largest power source (IEA, 2015b). However, change is proving more difficult in the transport sector where fossil fuels account for about 85 % of the total oil refined (de Jong

et al., 2012), and which is the second biggest source of global greenhouse gas emissions (IEA, 2008). The development of biorefineries for the production of fuels and other chemicals from sustainable sources (such as renewable biomass) therefore is increasing in relevance and interest (de Jong *et al.*, 2012; Lee *et al.*, 2015). Key to this interest in developing biotechnological solutions has been the combined advances in both Systems Biology and Synthetic Biology, permitting design-based engineering of biological systems, analogous to the design-based engineering of electrical circuits (Tyo *et al.*, 2010; Sanchez and Nielsen, 2015).

1.2 Biorefineries

The International Energy Agency (IEA) defines the biorefining concept as "the sustainable processing of biomass into a spectrum of bio-based products (food, feed, chemicals and materials) and bioenergy (biofuels, power and/or heat)". Biorefineries using sugar cane for production of bioethanol, electricity, heat, sugar and fertilizer are in commercial operation worldwide (Jungmeier, 2013). The yeast *Saccharomyces cerevisiae* is the cell factory of choice for this process, achieving in excess of 90 % of the theoretical sugar conversion into ethanol (Basso *et al.*, 2008). One challenge to the development of biorefineries is the limited portfolio of products that are produced at a commercial scale (Tang *et al.*, 2013; Hong and Nielsen, 2012). The total of five products produced by an ethanol biorefinery is considerably less than the hundreds of products derived from the oil refinery process which include: adhesives,

sealants, cosmetics (raw materials), flavourings, packaging, paints, resins and pharmaceutical drugs (PSAC, 2016). The proven ability to ferment *S. cerevisiae* on a large scale presents an attractive target for producing chemicals and fuels from sustainable sources, and Chapter 5 of this thesis details the development of a metabolic engineering strategy using an industrially relevant strain of *S. cerevisiae* to produce hydrocarbons, further increasing the potential product portfolio of this cell factory (van Dijken *et al.*, 2000).

A significant engineering issue in the development of *S. cerevisiae* as a multi-product cell chassis used in biorefineries, is its efficient and predominant carbon flux through to ethanol (Abbott *et al*., 2009; Mouret *et al*., 2006). This includes the propensity of this organism to metabolise high concentrations of carbohydrates to ethanol even in the presence of oxygen, a trait referred to as the Crabtree Effect (Crabtree, 1928). Therefore, in combination with the heterologous expression of genes to produce other chemicals and fuels from sustainable sources, the metabolism of ethanol needs to be decoupled within this cell factory (Ida *et al*. 2013; Runguphan and Keasling, 2014). Chapter 3 of this thesis introduces the six alcohol dehydrogenase (ADH) isozymes that catalyse the terminal reaction step during ethanol metabolism in *S. cerevisiae*, and which together are thought to allow this yeast metabolic flexibility within different environmental niches (Denis *et al*., 1983; de Smidt *et al.*, 2012). The outcomes of previous research activities to divert carbon flux away from ethanol production are discussed in Chapter 3, including a genetic strategy in which metabolism has been evaluated in an *S. cerevisiae* strain with all six ADH isozyme genes deleted (Flikweert *et al.*, 1996; Diderich *et al.*, 2001; Otterstedt

et al., 2004; Jansen *et al.*, 2012; Matsuda *et al.*, 2013). The characteristics of this strain are not suitable for its industrial use within a biorefinery (Ida *et al.*, 2012). The potential of a strategy where a combination of the ADH isozymes are deleted in order to divert carbon flux away from ethanol, however, is shown by the commercial biorefinery process utilising an *S. cerevisiae* strain deleted for a single ADH isozyme; this deletion reduces the amount of ethanol produced, whilst maximising the production of succinic acid (Reverdia (Geleen, Netherlands)). Highlighting the importance of genotypic and environmental interactions, the process was optimised by refinement of culture conditions, particularly pH, which in combination with the engineered strain genotype increased the titre of succinic acid (Jansen *et al.*, 2012). Chapter 3 in this thesis describes the development of a method to generate a full combinatorial *S. cerevisiae* ADH knockout library, applicable for the evaluation of ADH isozyme functional substitutions and genotype-by-environment interactions involved in carbon flux to ethanol (Solis-Escalante *et al.*, 2013). Chapter 4 details development of the bioprocess and multivariate analytical tools required to evaluate the 64 *S. cerevisiae* strains of the ADH knockout library within different culture environments, providing a knowledge base to further engineer carbon metabolism in this yeast.

1.3 Scientific analysis and replicability

Design-based genetic engineering of biological systems is complex. The six ADH isozymes that catalyse the single reaction between acetaldehyde and

ethanol in yeast are an example of this complexity. The isozymes are thought to provide yeast with metabolic flexibility within a natural environment (Denis *et al*., 1983; de Smidt *et al.*, 2012) and therefore metabolic systems must be evaluated under different environmental conditions. Moreover, in order to evaluate complex systems it is not appropriate to employ a scientific method requiring sequential data generation that includes rounds of experiments where all but one condition/factor is maintained (de Smidt *et al.*, 2008). Instead an empirical method that permits multivariate statistical analysis is required, particularly to understand interactions and their impact on multiple responses within the system (Swalley *et al*., 2006; VanderSluis *et al.*, 2014; see also Chapter 4.1). Without an appropriate evaluation, important features of complex system can be missed leading to low reproducibility rates of research. It is estimated that in the United States of America $28,000,000,000 a year is spent on preclinical research that is not reproducible (Freedman *et al.*, 2015). Chapter 4 introduces a multivariate analysis method known as Design of Experiments (DoE), and appraises its application for experimental design, modelling and prediction of a biological system (Lendrem *et al.*, 2016). The availability of system knowledge and advanced molecular biology techniques for the yeast *S. cerevisiae* makes this model organism particularly suitable for the evaluation of genotype and environmental factors in optimising a biological system for biotechnology (Sanchez and Nielsen, 2015).

1.4 Aims

The aim of this thesis is to evaluate the multivariate Design of Experiments methodology for analysing carbon flux in an industrially relevant strain of *S. cerevisiae*. To this end, specific objectives are:

- to develop a molecular biology technique for rapid, multiple and scarless gene deletions in a prototrophic strain of *S. cerevisiae* applicable to industrial use within a biorefinery;
- to generate a combinatorial alcohol dehydrogenase (ADH) isozyme gene knockout library;
- to use multivariate analysis to evaluate carbon flux in the ADH knockout library under different environmental conditions; and
- to develop a metabolic engineering strategy in order to increase the product portfolio of the yeast cell factory.

Subsequent chapters will introduce specific areas in more detail and describe results of experiments used to evaluate the aims of this thesis.

2 Materials and methods

2.1 Materials

2.1.1 Chemicals

Chemicals were purchased from BD Biosciences (Oxford, UK), Fisher Scientific UK Ltd (Loughborough, UK), Formedium (Hunstanton, UK), Sigma-Aldrich (Dorset, UK) and VWR (Lutterworth, UK) unless otherwise stated.

2.1.2 Media

All media were sterilised at 121 °C for 20 min unless otherwise stated. Agar was added to 15 g l^{-1}. Ampicillin was added to 100 µg ml^{-1}, or G418 (Geneticin) to 200 µg ml^{-1} as indicated.

Lysogeny broth (LB) contained 10 g l^{-1} tryptone, 5 g l^{-1} yeast extract and 10 g l^{-1} NaCl (Bertani, 1951).

Yeast extract, peptone and dextrose (YEPD) broth contained 10 g l^{-1} yeast extract, 20 g l^{-1} peptone and 10 g l^{-1} glucose (Green *et al.*, 2012).

Synthetic Medium (SM) was prepared according to Solis-Escalante *et al.* (2013) and was supplemented with acetamide (SM Ac) or fluoroacetamide (SM FAc) as indicated. The non-selective medium (SM) contained 3.0 g l^{-1} KH_2PO_4, 0.5 g l^{-1} $MgSO_4.7H_2O$, 5.0 g l^{-1} $(NH_4)_2SO_4$, 1.0 ml l^{-1} of trace element solution (Verduyn *et al.*, 1992) and 1.0 ml l^{-1} of vitamin solution (Verduyn *et al.*, 1992). Selection for the *amdSYM* marker used SM Ac medium which included 0.6 g l^{-1} acetamide and 6.6 g l^{-1} K_2SO_4 replacing the $(NH_4)_2SO_4$. The counter-selection medium SM FAc had the same composition as the non-selective media with the addition of 2.3 g l^{-1} fluoroacetamide.

S. cerevisiae defined media were prepared according to the manufacturer's instructions (Formedium, Hunstanton, UK). The basal media composition for all defined media was Yeast Nitrogen Base (YNB) without zinc which contained 5.0 g l^{-1} $(NH_4)_2SO_4$, 1.0 g l^{-1} KH_2PO_4, 0.5 g l^{-1} $MgSO_4$, 0.1 g l^{-1} NaCl, 0.1 g l^{-1} $CaCl_2$, 500 µg l^{-1} H_3BO_3, 40 µg l^{-1} $CuSO_4$, 100 µg l^{-1} KI, 200 µg l^{-1} $FeCl_3$, 400 µg l^{-1} $MnCl_2$, 200 µg l^{-1} Na_2MoO_4, 2.0 µg l^{-1} biotin, 400 µg l^{-1} pantothenic acid, 2.0 µg l^{-1} folic acid, 2.0 mg l^{-1} inositol, 400 µg l^{-1} nicotinic acid, 200 µg l^{-1} 4-aminobenzoic acid, 400 µg l^{-1} pyridoxine HCl, 200 µg l^{-1} riboflavin and 400 µg l^{-1} thiamine HCl. Synthetic Defined (SD) / 2.0 % glucose broth additionally contained 20 g l^{-1} glucose and 400 µg l^{-1} $ZnSO_4$.

2.2 Molecular genetic methods

2.2.1 Microbial strains

NEB 5-alpha (New England Biolabs, Ipswich, US) competent *Escherichia coli* strain (genotype: *fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 endA1 thi-1 hsdR17*) was used for microbiological cloning, storage and amplification of plasmid vectors. Single *Escherichia coli* colonies were inoculated into 6.0 ml LB broth with antibiotic selection, as required, and incubated at 37 °C with 220 rpm shaking.

Saccharomyces cerevisiae strain CEN.PK113-7D (*MATa*, *MAL2-8^c SUC2*) was purchased from EUROSCARF (Frankfurt, Germany). Colonies of *S. cerevisiae* were inoculated into 10 ml YEPD broth with antibiotic selection, as required, and incubated at 30 °C with 200 rpm shaking.

Shake flask cultures were set up at 10-20 % (v/v) of the total flask volume. Microbial cultures were grown in a 1.0 cm orbital shaking diameter incubator. For microbial storage, 1.0 ml of stationary phase culture was added to 250 µl of sterile glycerol, thoroughly mixed, snap frozen in liquid nitrogen and stored at −80 °C.

2.2.2 Bacterial transformation

E. coli transformations were performed with 100 ng plasmid DNA following the high efficiency transformation protocol (New England Biolabs, Ipswich, US).

Briefly this comprised using chemically competent *E. coli* with primed cellular membranes which, when stressed using heat shock, will internalise foreign DNA. The *E. coli* strain used is genetically modified in order to increase transformation efficiency, maintain the internalised DNA and prevent its rearrangement.

2.2.3 Plasmid minipreps

Plasmid DNA minipreps were performed from overnight *E. coli* cultures using the GeneJet plasmid miniprep kit (Thermo Scientific, Waltham) according to the manufacturer´s standard protocol. Briefly this comprised alkaline lysis of bacterial cells, clarification of lysate using centrifugation, loading and subsequent binding of DNA on silica columns under high salt conditions, washing away impurities and finally elution of purified DNA using a low salt buffer. DNA was eluted in DNase/RNase free H_2O, quantified using a Qubit™ fluorimeter (Thermo Scientific, Waltham, US) and stored at −20 °C.

2.2.4 Yeast transformation

Transformations of DNA plasmids and linearised deletion cassettes into yeast were performed using lithium acetate and single stranded carrier DNA (Schiestl and Gietz, 1989). The protocol was standardised with the use of the yeast transformation kit (Sigma-Aldrich, Dorset, UK), adapted as follows:

Single yeast colonies were inoculated into 10 ml YEPD broth and incubated overnight at 30 °C with 200 rpm shaking. 100 ml YEPD broth was inoculated with overnight culture at OD_{600nm} of 0.2 and incubated at 30 °C with 200 rpm shaking until an OD_{600nm} of 1.0 was reached (typically after 3-4 hours). Cells were washed in an equal volume of sterile Tris-EDTA buffer solution, pelleted and re-suspended in 1.0 ml transformation buffer (Sigma-Aldrich, Dorset, UK). For each transformation, 100 µl yeast cell suspension was aliquoted and to this 10 µl denatured salmon sperm DNA (100 µg) and 1.0 µg of plasmid/linearised DNA (or an equal volume of sterile Tris-EDTA buffer solution for the negative control) was added. Ethanol was added to 10 % (v/v) and the transformation mix was incubated at room temperature for 5 min. 600 µl PLATE buffer (Sigma-Aldrich, Dorset, UK) was added followed by incubation for 1.0 hour at 30 °C with 200 rpm shaking. The cells were then heat shocked at 42 °C for 20 min, pelleted, re-suspended in 500 µl YEPD broth, incubated for 1.0 hour at 30 °C with 200 rpm shaking, washed in 500 µl of sterile Tris-EDTA buffer, pelleted and re-suspended in 100 µl of sterile Tris-EDTA buffer solution. 90 µl was spread on a selection media plate and 10 µl on a non-selection control plate. Transformation plates were incubated for 2-4 days at 30 °C until discrete colonies were visible.

2.2.5 Restriction digests and ligations

Restriction endonucleases were purchased from New England Biolabs (Ipswich, US) and digests were performed according to the manufacturer's protocol. Linearised DNA was ligated into vectors using rapid DNA ligation kit according

to the manufacturer's protocol (ThermoFisher Scientific, Waltham, US). Ligation mix was transformed into *E. coli* as 2.2.2 and 2.2.3, prior to plasmid verification by diagnostic restriction digests and/or Sanger sequencing using primers in Table 1.

2.2.6 Agarose gel electrophoresis

Unless otherwise stated gel electrophoresis was performed using a 0.8 % w/v (8.0 mg ml^{-1}) agarose gel containing 10 µg ml^{-1} ethidium bromide. 2.0 µl of Hyperladder I (Bioline, London, UK) was loaded as a size reference marker, and gels were run at 110 V for 1.0 hour in TAE buffer. DNA bands were visualised and photographed using a UV transilluminator. If required, DNA was excised and purified from agarose gels using the QIAquick gel extraction kit (Qiagen, Hilden, Germany), according to the manufacturer's instructions.

2.2.7 DNA sequencing

Sanger Sequencing (Sanger *et al.*, 1977) was performed using primers in Table 1 by Source BioScience (Nottingham, UK); sequencing data was evaluated for quality and coverage using 4Peaks software (www.nucleobytes.com).

Table 1. Primers used for sequencing plasmid DNA and genotype verification.

Name	Sequence (5′ → 3′)	Notes
5′SEQfwd	AAGCGGAAGGCGAGAGTAGG	For checking of the 5′ and 3′ homologous sequences in plasmid pADH_Deletion_Cassette (see Figure 2.2) specific for each of the ADH and FAA1 gene deletion cassettes
5′SEQrev	GGGATGTATGGGCTAAATG	
3′SEQfwd	GCGTCAATCGTATGTGAATG	
3′SEQrev	ATAAGGGCGACACGGAAATG	
pEF1aFwd	ATTTGCCCTTTTTGAGTTTGG	For checking pAlkene_3genes cloning variants
EGFPCFwd	CATGGTCCTGCTGGAGTTCGTG	
pCEPFwd	AGAGCTCGTTTAGTGAACCG	
pCMVfwd	GAGCTCGTTTAGTGAACCGTC	
T7Fwd	TAATACGACTCACTATAGGG	
ADH1seqverFwd	CAGCACCAACAGATGTCGTTGTTCC	For sequence verification of *S. cerevisiae* gene deletion genotypes
ADH1seqverRev	CGACCTCATGCTATACCTGAGAAAGCAACC	
ADH2seqverFwd	CGGGAAACCATCCACTTCACGAGAC	
ADH2seqverRev	GAGACGATTCAGAGGAGCAGGACAAAC	
ADH3seqverFwd	CGTTTCTGCGTCCGTACACTGTCC	
ADH3seqverRev	GTTTGGGCGGCTCGATGCTTG	
ADH4seqverFwd	TTGCTGCCTCAAATATCTCACAC	
ADH4seqverRev	GTGCATTATACTGTACGCACAAC	
ADH5seqverFwd	CTGCTATCTGCTTGTAGAAGGGTACGCTAACAGAG	
ADH5seqverRev	CTATTTCAGTTTGTCTTACGCACGCAGTTG	
SFA1seqverFwd	AAAGTACTGACCCTTGAATCG	
SFA1seqverRev	CCAATGTTGGTGATAATAGC	

2.2.8 Polymerase Chain Reaction (PCR)

Thermal cycling reactions were performed using GoTaq® Green master mix (Promega, Madison, US), according to the manufacturer's instructions. A standard reaction mix contained: 12.5 µl of 2 × GoTaq® Green master mix, 3.0 µl $MgCl_2$ [50 mM], 1.0 µl forward primer [10 pmol], 1.0 µl reverse primer [10 pmol] and 5.0 µl template DNA [100 ng or template using single colony]. Total reaction volumes were standardised to 25 µl using nuclease free water. Specific reaction annealing temperatures will be stated as required.

For yeast colony PCR a single colony was picked using a sterile pipette tip and suspended in 50 µl Tris-EDTA buffer pH 8.0. The mixture was heated to 99.9 °C for 5.0 min, cooled to 4.0 °C and placed on ice for at least 5.0 min. The mixture was centrifuged at 13,000 ×*g*. 5.0 µl of the supernatant was used as template DNA for diagnostic PCR for gene deletions of *S. cerevisiae* using primers detailed in Table 2.

Diagnostic colony PCR primers sets for the ADH genes and *FAA1* (Table 2) were designed with similar annealing temperatures to allow concurrent specific analyses. Primers flanked each gene ensuring that the specific primer combination could amplify either the wild type gene, the integration of the *amdSYM* marker, or the scarless removal of both.

Table 2. Primers for colony PCR assessment of gene deletions and marker removal within the genome of *S. cerevisiae* CEN.PK113-7D.

Gene	Name	Sequence (5´ → 3´)	Diagnostic colony PCR fragment size (bp)		
			Wild type gene	*amdSYM* integration	Scarless excision
ADH1	ADH1verFwd	CAGCACCAACAGATGTCGTTGTTCC	1,509	2,963	462
	ADH1verRev	CGACCTCATGCTATACCTGAGAAAGCAACC			
ADH2	ADH2verFwd	CGGGAAACCATCCACTTCACGAGAC	1,675	3,130	628
	ADH2verRev	GAGACGATTCAGAGGAGCAGGACAAAC			
ADH3	ADH3verFwd	CGTTTCTGCGTCCGTACACTGTCC	1,609	2,986	482
	ADH3verRev	GTTTGGGCGGCTCGATGCTTG			
ADH4	ADH4verFwd	TTGCTGCCTCAAATATCTCACAC	1,426	2,781	277
	ADH4verRev	GTGCATTATACTGTACGCACAAC			
ADH5	ADH5verFwd	CTGCTATCTGCTTGTAGAAGGGTACGCTAACAGAG	1,678	3,124	672
	ADH5verRev	CTATTTCAGTTTGTCTTACGCACGCAGTTG			
SFA1	SFA1verFwd	AGTGCCTCAGTCGAATGG	1,997	3,338	836
	SFA1verRev	GGCGATGCAAGTGAAACC			
FAA1	FAA1verFwd	TTCGCGGATATCGTATTATCG	2,501	2,940	400
	FAA1verRev	ACGTGAATGAGGTGATATGAC			

2.2.9 Q5® site-directed mutagenesis

The Q5® site-directed mutagenesis kit was purchased from New England Biolabs (Ipswich, US). Protocols to make insertions, deletion and substitutions to plasmid vectors were developed using NEBaseChanger™ software (www.neb.com), detailing specific PCR primers and thermocycling conditions (see Table 3).

Table 3. Primers used for Q5® site-directed mutagenesis of plasmid DNA.

Name	Sequence (5´ → 3´)	Tm (°C)	Notes
FAA1upQ5fwd	agcatgaaactgctccaagaagaaattatcaatg ccgcgagaagccgCAGCTGAAGCTTCGTACG	64	pADH1_deletion_cassette template DNA, replacing the 5´ homologous sequence with the 5´ homologous sequence specific for the *FAA1* gene
FAA1upQ5rev	cattggcggctttcccaactggaacggtatattg agcaaccatattgtCCCAGTTTCAGCGACTAAG		
FAA1downQ5fwd	catttccatgataggaaagcctcatcatactaaa gcacttttcagttTGTACCCAGTCGTCAATC	62	pFAA1_UP_deletion_cassette template DNA, replacing the 3´ homologous sequence with the 3´ homologous sequence specific for the *FAA1* gene
FAA1downQ5rev	ttgatccaattgttgtctttttttgtcttttgtg tttgttctagttttGCATAGGCCACTAGTGGA		
O2_Q5fwd	GTCGACGATCTCCCTATAG	59	pO1 template DNA, deleting the *Sth* gene
O2_Q5rev	TCTAGACGGTTCACTAAAC		
O5_Q5fwd	ATTTGCCCTTTTTGAGTTTG	59	pO1 template DNA, deleting the *'TesA* gene
O5_Q5rev	GTCGACTGACAAGTTCTTG		
O6_Q5fwd	ATTTGCCCTTTTTGAGTTTG	59	pO2 template DNA, deleting the *'TesA* gene
O6_Q5rev	GTCGACTGACAAGTTCTTG		
O7_Q5fwd	GTCGACGATCTCCCTATAG	62	pO3 template DNA, deleting the *Sth* gene
O7_Q5rev	TCTAGACTTCGAGCGTCC		
O8_Q5fwd	GACGGTTCACTAAACGAG	61	pO5 template DNA, deleting the *RhFRED:OleT* gene
O8_Q5rev	GTCGACTGACAAGTTCTTG		

2.2.10 Plasmid vectors

Plasmid design used the annotated and curated CEN.PK113-7D *S. cerevisiae* genome sequence (*Saccharomyces cerevisiae* Genome Database (SGD) www.yeastgenome.org), visualised and manipulated using a combination of the resources within the SGD website, Artemis (Wellcome Trust Sanger Institute www.sanger.ac.uk), Clone Manager Professional version 8 (Sci-Ed Software, www.scied.com), Gene Designer 2.0 (DNA 2.0 Inc. www.dna20.com) and BLAST (National Center for Biotechnology Information www.ncbi.nlm.nih.gov). A map of the pADH_Deletion_Cassette plasmid backbone used for the deletion of the ADH isozymes in *S. cerevisiae* CEN.PK113-7D is shown in Figure 2.1 Sequence variants for this plasmid are listed in Table 4. A map of the pAlkene_3genes yeast expression vector used in the evaluation of heterologous alkene production in *S. cerevisiae* CEN.PK113-7D is shown in Figure 2.2. Variants for this plasmid are listed in Table 5.

All plasmid expression vectors were synthesised and sequence verified by DNA 2.0 Inc. (California, US). Synthesised vectors were transformed into *E. coli* to maintain and amplify stocks of the DNA.

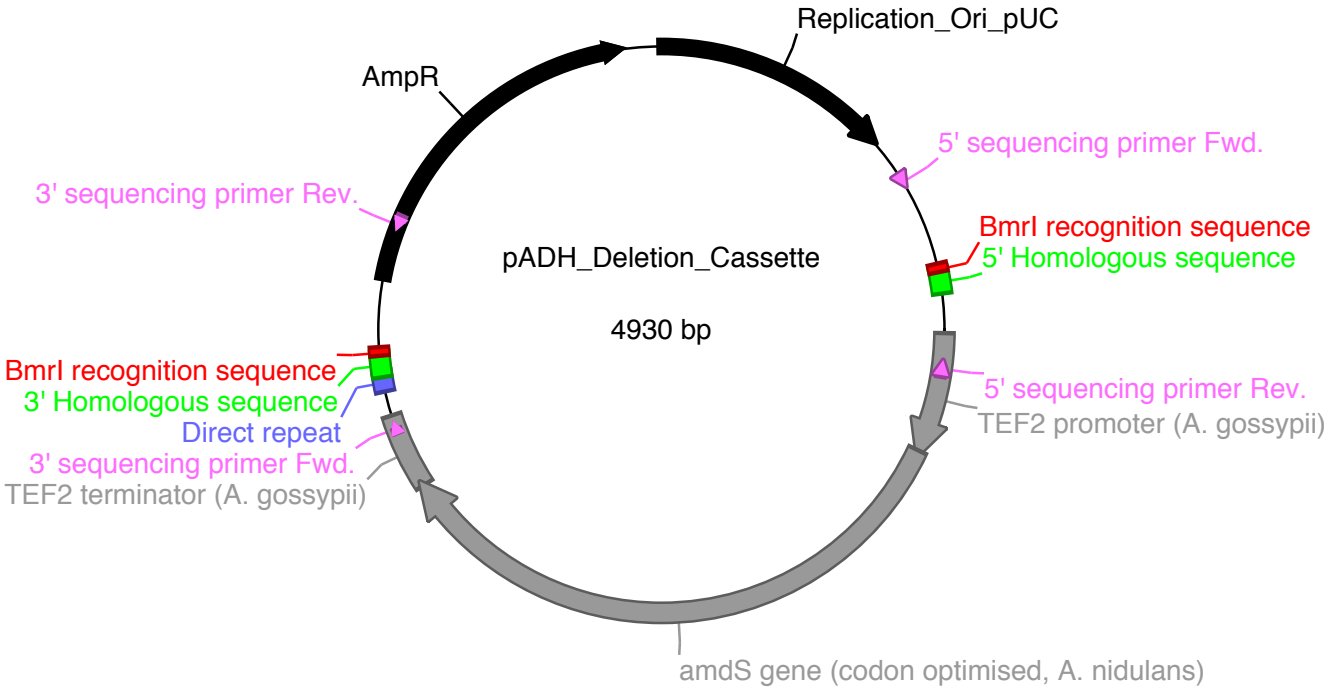


Figure 2.1 Plasmid map of the pADH_Deletion_Cassette plasmid used for the deletion of the alcohol dehydrogenase isozymes in *S. cerevisiae* CEN.PK113-7D. Isozyme-specific homologous recombination and direct repeat marker excision sequences are detailed in Table 4. The bacterial amplification backbone was the pJ154 series of expression vector (DNA2.0 Inc., California, US), allowing the use of the type II restriction enzyme, *Bmr*I, for linearisation of the deletion cassette. The *amdS* gene is the prototrophic selection marker, in *S. cerevisiae.*

Table 4. The pADH_Deletion_Cassette plasmid homologous recombination sequence variants used for the deletion of the alcohol dehydrogenase isozymes in *S. cerevisiae* CEN.PK113-7D and the specific direct repeat sequences for subsequent marker excision. The two 5´ sequences for the *ADH1* deletion cassette were (1) the original sequence and (2) the extended 100 bp homologous sequence used to increase integration efficiency. In addition, the homologous recombination and direct repeat sequences used to delete the *FAA1* gene (a long-chain fatty acyl-CoA synthetase) in *S. cerevisiae* CEN.PK113-7D are shown.

Gene	Name	Sequence (5´ → 3´)
ADH1	5´sequence 1	ATGTCTATCCCAGAAACTCAAAAAGGTGTTATCTTCTACGAATCCCACGGTAAGT
	5´sequence 2	ATGTCTATCCCAGAAACTCAAAAAGGTGTTATCTTCTACGAATCCCACGGTAAGT TGGAATACAAAGATATTCCAGTTCCAAAGCCAAAGGCCAACGAAT
	3´ sequence	TTTCTTATGATTTATGATTTTTATTATTAAATAAGTTATAAAAAAAATAAGTGTA
	Direct repeat	TCAAGCTATACCAAGCATACAATCAACTATCTCATATACA
ADH2	5´sequence	ATGTCTATTCCAGAAACTCAAAAAGCCATTATCTTCTACGAATCCAACGGCAAGT
	3´ sequence	TCTCTTATGTCTTTACGATTTATAGTTTTCATTATCAAGTATGCCTATATTAGTA
	Direct repeat	TACAATCAACTATCAACTATTAACTATATCGTAATACACA
ADH3	5´sequence	ATGTTGAGAACGTCAACATTGTTCACCAGGCGTGTCCAACCAAGCCTATTTTCTA
	3´ sequence	TGTTACGCACCCAAACTTTTTATGAAAGTCTTTGTTTATAATGATGAGGTTTATA
	Direct repeat	GTTAAAACTAGGAATAGTATAGTCATAAGTTAACACCATC
ADH4	5´sequence	ATGTCTTCCGTTACTGGGTTTTACATTCCACCAATCTCTTTCTTTGGTGAAGGTG
	3´ sequence	TCGAACGAACTCATAAACGTCAATTATGCGTGTGCCTTATTTATTTAGTTGTGCG
	Direct repeat	CAAGTTTACATTTGCAACAACTAATAGTCAAATAAGAAAA
ADH5	5´sequence	ATGCCTTCGCAAGTCATTCCTGAAAAACAAAAGGCTATTGTCTTTTATGAGACAG
	3´ sequence	TGTAACGAATTTGATGAATATATTTTTACTTTTTATATAAGCTATTTTGTAGATA
	Direct repeat	AGAAAATTATTTAACTACATATCTACAAAATCAAAGCATC
SFA1	5´sequence	ATGTCCGCCGCTACTGTTGGTAAACCTATTAAGTGCATTGCTGCTGTTGCGTATG
	3´ sequence	CTTAATTAAACTAAGTAAGCATGACTCAAATTTTCTGGAATACTTTGAAAATCAA
	Direct repeat	AATCTCCAAGTAAAGAAGGAATATAAGTAATATAAGTACA
FAA1	5´sequence	ATGGTTGCTCAATATACCGTTCCAGTTGGGAAAGCCGCCAATGAGCATGAAACTG CTCCAAGAAGAAATTATCAATGCCGCGAGAAGCCG
	3´ sequence	TGGATCAACATTTCCATGATAGGAAAGCCTCATCATACTAAAGCACTTTTTCAGT
	Direct repeat	AAAACTAGAACAAACACAAAAGACAAAAAAAGACAACAAT

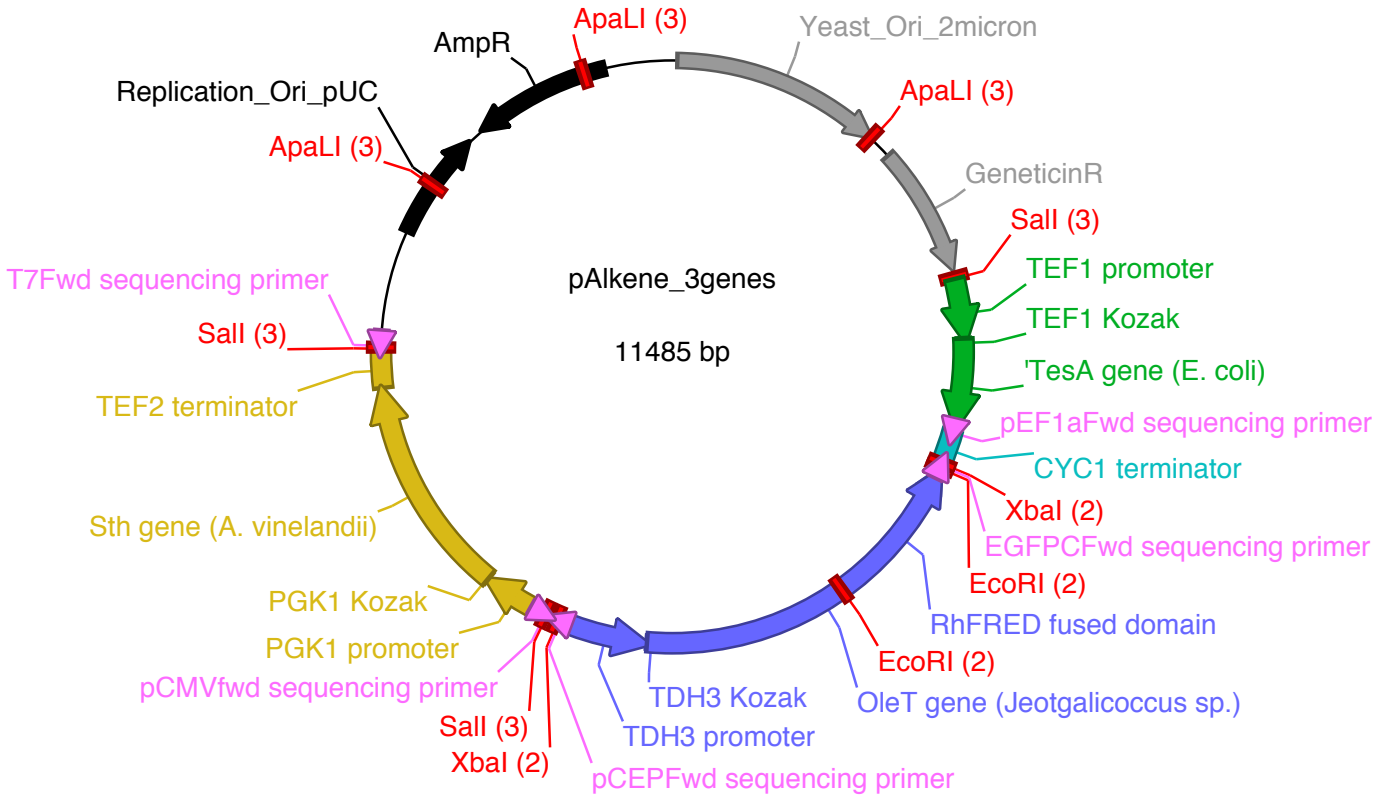


Figure 2.2 Plasmid map of the pAlkene_3genes yeast expression vector used in the evaluation of heterologous alkene production in *S. cerevisiae* strain CEN.PK113-7D. The bacterial amplification backbone was the pJ404 series of expression vector (DNA2.0 Inc., California, US). The Geneticin resistance gene is the selectable marker for yeast. A combination of restriction digest/ligation and Q5 mutagenesis was used to generate the vector variants listed in Table 5.

Table 5. pAlkene_3genes vector variants used to evaluate alkene production in *S. cerevisiae* strain CEN.PK113-7D. Included is the cloning method used to generate the variant.

Plasmid Variant	Size (bp)	Key elements	Cloning method
pO1	11,485	RhFRED[1]:OleT[2] ´TesA[3] Sth[4]	Synthesised by DNA2.0 Inc.
pO2	9,409	RhFRED:OleT ´TesA	Q5 mutagenesis; pO1 template DNA
pO3	8,510	´TesA Sth	*Xba*I digest; pO1 template DNA
pO4	10,489	OleT[2] ´TesA Sth	*EcoR*I digest; pO1 template DNA
pO5	10,517	RhFRED:OleT Sth	Q5 mutagenesis: pO1 template DNA
pO6	8,441	RhFRED:OleT	Q5 mutagenesis; pO2 template DNA
pO7	6,434	´TesA	Q5 mutagenesis; pO3 template DNA
pO8	7,316	Sth	Q5 mutagenesis; pO5 template DNA
pO9	5,240	Empty vector	*Sal*I digest; pO1 template DNA

[1] RhFRED: The NADPH reduction domain from *Rhodococcus sp*. (Liu *et al.*, 2014).

[2] OleT: An alkene-producing gene from *Jeotgalicoccus sp.* (Liu *et al.*, 2014).

[3] ´TesA: A thioesterase with deleted membrane signal peptide sequence from *E. coli* (Runguphan and Keasling, 2014).

[4] Sth: A pyridine nucleotide transhydrogenase from *Azotobacter vinelandii (Nissen et al., 2001).*

2.2.11 Bioinformatic methodology

The phylogenetic analysis of the ADH isozyme protein sequences was performed using Phylogeny.fr bioinformatic pipeline (Dereeper *et al.*, 2008) using default settings. The pipeline incorporates: i) an alignment using MUSCLE, a programme which uses two distance measures (*k*mer and Kimura distances) per pair of protein sequences – distance matrices are clustered using UPGMA providing alignments of multiple protein sequences (Edgar, 2004); ii) curation using Gblocks, a method which uses five threshold values to select for conserved blocks of alignments to perform subsequent phylogenetic analysis (Castresana, 2000); iii) phylogeny using the PhyML algorithm, based upon the maximum-likelihood principle (Guindon *et al.*, 2010; Anisimova and Gascuel, 2006); and iv) tree rendering using TreeDyn (Chevenet *et al.*, 2006).

2.3 *In vivo* methods

2.3.1 *amdSYM* gene deletion

The *amdSYM* method for gene deletion was performed according to Solis-Escalante *et al.* (2013). Chapter 3 discusses method development specific for ADH gene deletions in the prototrophic *S. cerevisiae* strain CEN.PK113-7D.

2.4 Cell culturing methods

Small scale cultures (including overnight inocula) were set up as described in Section 2.2.1.

2.4.1 96 well plates

Sterile 96 well, flat bottom W1515 cell culture plates (Genetix, Berkshire, UK) were used with each well at a 300 µl working volume. The culture medium was YEPD; the outer wells of the plates contained Phosphate Buffer Saline (PBS) solution. Overnight cultures were pelleted, washed in 25 ml of sterile PBS and re-suspended in a calculated volume of PBS to standardise the initial optical density for each well at OD_{600nm} = 0.02. Culture conditions were 30 °C with "normal" shaking using a Sunrise™ microplate reader (TECAN, Männedorf, Switzerland).

Experiments were performed according to the experimental design evaluating 5 level factor interactions using a custom design (JMP Pro v. 12, SAS Institute Inc. USA) in triplicate and in a randomised well position.

2.4.2 Chemostat bioreactor

Chemostat cultures were performed in laboratory fermenters (Labfors 5, INFORS AG, Bottmingen, Switzerland). 10 ml of overnight cultures were used as inocula. The medium for batch and chemostat cultivation was SD / 2.0 % glucose. Chemostats were run at a dilution rate of 0.1 h^{-1} and a working volume of 1.0 l maintained using a peristaltic pump controlled by an electrical level sensor. Culture conditions were maintained at 30 °C, with airflow set to 1.0 vvm and stirrer speed set to 800 rpm. The pH was kept at 5.5 via automatic addition of 5 M NaOH or 5 M HCl. Culture purity was checked by microscopy of daily samples.

2.4.3 Ministat bioreactor

32 parallel ministat bioreactors were constructed according to Miller *et al.* (2013) with adaptations including individual media reservoirs and 50 ml effluent burettes to measure the media flow rate for each of the 32 bioreactors. Dilution rate was controlled using a single peristaltic pump with 32 channels, which were individually fine-tuned. Two heat blocks controlled temperature. Experimentation was performed at an individual bioreactor gas flow rate of 200 ml min^{-1} with either air or nitrogen. Inoculation of the ministat bioreactors used

3.0 ml of overnight culture added to 17 ml of culture medium, grown in batch culture for 24 h prior to operation as a ministat.

2.4.4 Biomass determination

Culture densities were measured at an OD_{600nm} using a spectrophotometer with 1.0 cm path-length.

The BD FACS Aria II flow cytometer (Becton Dickenson, San Jose, US) was used to determine cell count and viability. The blue 488 nm excitation laser was used for analysis of samples (diluted with 1:10 with sterile PBS) and data collection continued until 1,500 events of the CountBright™ absolute counting beads (ThermoFisher Scientific, Waltham, US) had been recorded. Size (forward scatter) and internal complexity (side scatter) detectors were used to identify single cells of *S. cerevisiae.*

The FACS and spectrophotometric assays were performed simultaneously using 16 separate yeast cultures, providing a variety of growth profiles for comparison. The two methods of biomass determination were shown to directly correlate with each other (Figure 2.3).

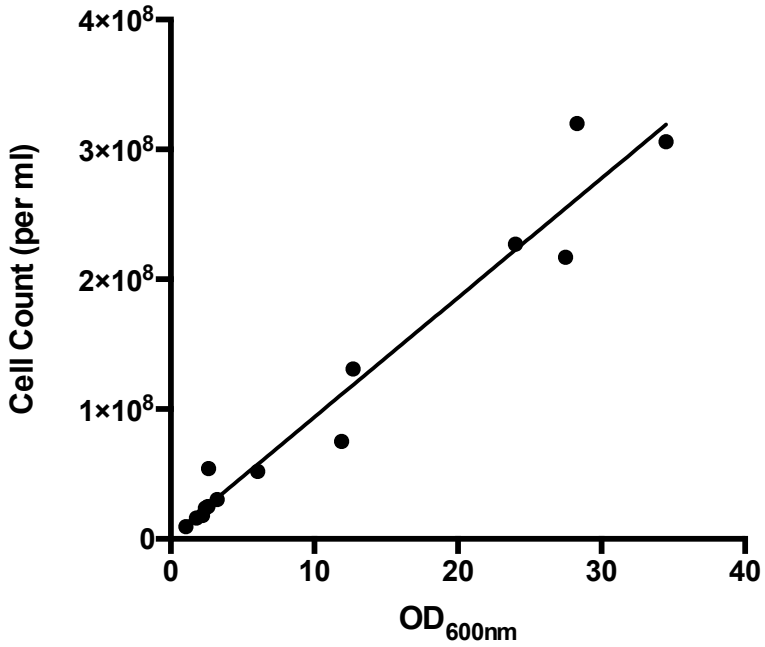


Figure 2.3 Linear regression curve of the correlation between methodologies used for determination of yeast cell density. 1.0 OD_{600nm} = 1.1×10^{7} cells ml^{-1} (R^2 = 0.9556).

Propidium iodide was used to stain yeast suspensions according to Deere *et al.* (1998), in order to determine yeast culture viability. Method validation was performed by mixing viable and non-viable yeast cultures (heated at 95 °C for 10 min) in known proportions prior to the addition of propidium iodide. The measured proportions of cell viabilities were ± 1.15 % of the anticipated result. Cells sorted by FACS on to a YEPD agar plate to assess viability (Figure 2.4) confirmed that propidium iodide is an effective yeast viability indicator.

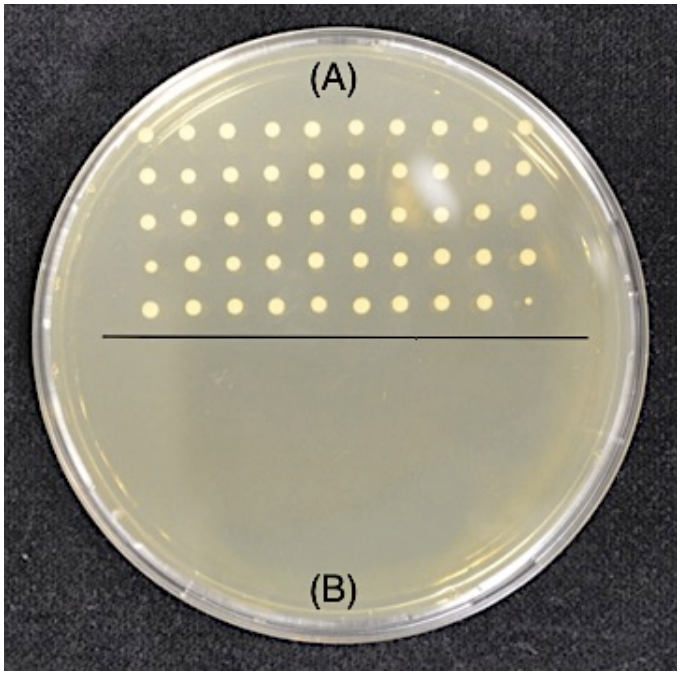


Figure 2.4 Fluorescence activated cell sorter (FACS) separation and plating of propidium-iodide stained viable (A) and non-viable (B) yeast cells.

2.5 Chromatographic methods

2.5.1 High performance liquid chromatography with a refractive index detector (HPLC RI)

200 µl of supernatant from cell samples was combined with 10 µl of cellobiose [85.47 g l^{-1}], used as an internal standard for quantification. Analyses were performed using an Agilent Technologies (Santa Clara, US) 1200 series liquid chromatography system equipped with a 1260 Infinity refractive index detector. 60 µl of each sample and standard were injected on a Rezex™ RHM-Monosaccharide H+ (8%) column, 8.0 µm, 300 × 7.8 mm (Phenomenex, Torrance, US), at 50 °C. The refractive index detector was at 40 °C; chromatographic separation was obtained at a flow rate of 0.6 ml min^{-1} using a 0.005 % (v/v) H_2SO_4 isocratic mobile phase. A 5-point calibration curve and retention time standard including cellobiose, glucose, glycerol and ethanol were performed for each analysis. Data were analysed with OpenLAB CDS ChemStation edition for LC & LC/MS systems (Agilent Technologies, Santa Clara, US).

2.5.2 Gas chromatography (GC)

1.0 ml of yeast cell suspension was disrupted using lysing matrix C tubes and the FastPrep®-24 instrument (MP Biomedicals, California, US), at a speed setting of 6.0 m s^{-1} for 1.0 min and for 8 passes. 100 µl of homogenised sample was combined with 1.5 µl of heptadecanoic acid [1.70 g l^{-1}], used as an internal

standard for quantification of free fatty acids. 300 µl of homogenised sample was combined with 5.0 µl of tetradecene [1.50 g l^{-1}], used as an internal standard for quantification of alkenes.

The free fatty acid analytical method performed on the prepared samples was according to Runguphan and Keasling (2014) with the following adaptations: (1) 100 µl of ethyl acetate (EA)/iodomethane (Mel) was added to the samples and standards; (2) autonomous sample/standard preparation and injection was performed using a Gerstel (Mülheim an der Ruhr, Germany) multipurpose sampler attached to the GC with flame ionisation detector (FID).

Alkenes were extracted from homogenised samples using an equal volume of dichloromethane mixed at 750 rpm, 40 °C for 2.0 h using a Gerstel (Mülheim an der Ruhr, Germany) multipurpose sampler, prior to injection.

2.5.3 Gas chromatography with flame ionisation detector (GC FID) parameters

Analyses were performed using an Agilent Technologies (Santa Clara, US) 7890B GC system equipped with a flame ionisation detector (maintained at 300 °C, hydrogen flow of 30 ml min^{-1}, air flow of 400 ml min^{-1} and a data acquisition rate of 50 Hz) and a Gerstel (Mülheim an der Ruhr, Germany) multipurpose sampler. 2.0 µl of each sample and standard were injected into a non-deactivated, baffled glass liner with a 10:1 split ratio (12 ml min^{-1} split flow)

and the inlet temperature was maintained at 250 °C with a 3.0 ml min^{-1} septum purge flow. A Zebron semi-volatiles (Phenomenex, Torrance, US) column, 30 m × 250 µm × 0.25 µm, coupled to a 10 m guard column was maintained at a constant gas flow rate of 1.2 ml min^{-1}. The temperature gradient of the GC oven was initially held for 2.45 minutes at 40 °C. It was ramped at a rate of 24.52 °C min^{-1} until 310 °C was achieved and held for 4.08 minutes. A 6-point calibration curve and retention time standard including myristic acid ($C_{14:0}$), pentadecanoic acid ($C_{15:0}$), palmitic acid ($C_{16:0}$), heptadecanoic acid ($C_{17:0}$), stearic acid ($C_{18:0}$) and oleic acid ($C_{18:1}$) for free fatty acid analysis or decane ($C_{10:0}$), tridecane ($C_{13:0}$), tridecene ($C_{13:1}$), tetradecene ($C_{14:0}$), tetradecene ($C_{14:1}$), pentadecane ($C_{15:0}$), pentadecene ($C_{15:1}$), heptadecane ($C_{16:0}$), heptadecene ($C_{16:1}$) and octadecane ($C_{18:0}$) for alkene analysis were performed. Data were analysed with OpenLAB CDS ChemStation edition for GC systems (Agilent Technologies, Santa Clara, US).

2.5.4 Quadrupole time of flight gas chromatography mass spectrometer (GC/Q-TOF/MS) parameters

Analyses were performed using an Agilent Technologies (Santa Clara, US) 7200 series GC(7890A)/Q-TOF/MS system. The emission current (35 µA), emission voltage (70 eV), acquisition rate (5.0 spectra s^{-1}) and acquisition time (200 ms $spectra^{-1}$) were maintained throughout analysis. 0.8 µl of each sample and standard were injected into a non-deactivated, baffled glass liner with a 10:1 split ratio (12 ml min^{-1} split flow) and the inlet temperature was maintained at 250 °C with a 3.0 ml min^{-1} septum purge flow. A Zebron semi-volatiles

(Phenomenex, Torrance, US) column, 30 m × 250 µm × 0.25 µm, coupled to a 10 m guard column was maintained at a constant helium flow rate of 1.5 ml min^{-1}. The temperature gradient of the GC oven was initially held for 4.0 minutes at 70 °C. It was ramped at a rate of 15 °C min^{-1} until 310 °C was achieved and held for 6.0 minutes. A 6-point calibration curve and retention time standard including decane, tridecane, tridecene, tetradecane, tetradecene, pentadecane, pentadecene, heptadecane, heptadecene and octadecane for alkene analysis were performed. Data were analysed using Agilent Technologies (Santa Clara, US) MassHunter qualitative, TOF quantitative, unknowns analysis and library editor software packages.

2.6 Experimental design and statistical analysis

The Design of Experiments (DoE) methodology including optimal design of experimentation, data modelling and visualisation utilised the commercially available DoE modelling software package (JMP Pro v. 12, SAS Institute Inc. USA). All defaults for the software were maintained unless otherwise stated.

3 Engineering an industrially relevant strain of *S. cerevisiae* with combinatorial deletions of alcohol dehydrogenase isozymes

3.1 Introduction

3.1.1 The metabolism of ethanol in *S. cerevisiae*

Saccharomyces cerevisiae is physiologically classified as a facultatively fermentative yeast and is one of the few yeasts able to grow anaerobically (Visser *et al.*, 1990). It is also categorised as a Crabtree-positive yeast (Crabtree, 1928; De Deken, 1966) due to its propensity to perform alcoholic fermentation in the presence of oxygen. There are two types of Crabtree effect described in yeast. In **the long-term Crabtree effect,** ethanol is produced at high growth rates regardless of the fermentation mode. This is due to an insufficient respiratory capacity because of the repression of respiratory genes (Postma *et al.*, 1989). In contrast, **the short-term Crabtree effect** arises due to saturation of respiratory metabolism (Flikweert *et al.*, 1996). This occurs on transition from glucose limitation to glucose excess.

In addition to oxygen concentration, the availability of fermentable sugars in yeast cultures is a key factor in the regulation of enzyme activities involved in respiration and alcoholic fermentation (van Dijken *et al.*, 1993). In aerobic culture conditions *S. cerevisiae* controls the balance between fermentation and respiration primarily in response to the environmental sugar concentration,

switching to a mixed respiro-fermentative metabolism as soon as the external glucose concentration exceeds 0.8 mM (Verduyn *et al.*, 1984). Carbon metabolism in *S. cerevisiae*, and indeed aerobic ethanol production, is dependent upon environmental conditions which determine cellular metabolic fluxes, governed by substrate pools and enzyme activities (van Dijken *et al.*, 1993). High glucose levels result in the glycolytic rate exceeding that of the pyruvate dehydrogenase enzyme complex, which is the branch-point reaction between glycolysis and oxidative phosphorylation (Figure 3.1), the resulting overflow of carbon is metabolised using ethanolic fermentation. At low external glucose levels and in the presence of oxygen, *S. cerevisiae* does not produce ethanol. It has been postulated that the Crabtree effect permits rapid production of ATP within yeast, circumventing the slower metabolic rates involved in energy production via oxidative phosphorylation (Otterstedt *et al.*, 2004; Kappeli, 1986).

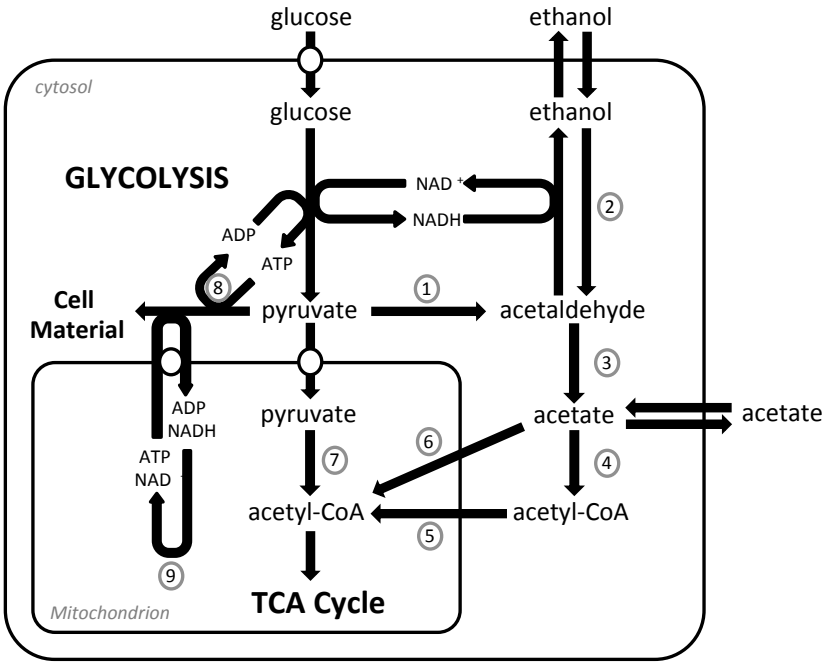


Figure 3.1. Glucose metabolism in *Saccharomyces cerevisiae* (adapted from van Dijken *et al.*, 1993). (1) Pyruvate decarboxylase; (2) alcohol dehydrogenase(s); (3) acetaldehyde dehydrogenase(s); (4) acetyl-CoA synthetase (cytoplasmic); (5) transport of acetyl-CoA into mitochondria via the carnitine shuttle; (6) transport of acetate into mitochondria and formation of acetyl-CoA via mitochondrial acetyl-CoA synthetase; (7) pyruvate dehydrogenase complex; (8) pyruvate carboxylase; (9) formation of ATP via oxidative phosphorylation.

The production of ethanol via fermentation oxidises NADH. Regeneration of NAD^+ is required for the oxidation of glyceraldehyde-3-phosphate during glycolysis and is therefore essential for maintaining the redox balance in the cytoplasm of fermenting yeast cells (Drewke *et al.*, 1990). The energetic net reaction equation of the ethanolic fermentation of glucose by *S. cerevisiae* is shown below:

$$C_6H_{12}O_6 + 2\ ADP + 2\ P_i \quad = \quad 2\ C_2H_6O + 2\ ATP + 2\ CO_2$$

Increasing glycolytic flux rather than maximising ATP yield is the apparent goal of *S. cerevisiae* metabolism. The production of ethanol (through the Crabtree effect) combined with a metabolic diauxic shift (assimilating the produced ethanol, in the presence of oxygen), highlights *S. cerevisiae's* evolutionary development to 'make-accumulate-consume'. This yeast kills its competitors by producing ethanol, to which it is resistant, and can then consume the generated ethanol later (Piškur *et al.*, 2006).

3.1.2 Perturbations of ethanol biosynthesis

Metabolism in *S. cerevisiae* has been the focus of much research not only due to the importance of this yeast as a model eukaryotic organism but also because of its application as an industrial host strain in the manufacture of a plethora of products besides ethanol. The propensity of *S. cerevisiae* to ferment is advantageous for industrial fermentative processes such as ethanolic biofuel production. Carbon routed toward ethanol production is however disadvantageous for manufacturing biomass-linked products such as heterologous proteins (Mouret *et al.*, 2006), and production of ethanol competes for cofactors (Abbott *et al.*, 2009). Additionally increasing ethanol concentrations during fermentation reduces the growth rate and cell viability of yeast cultures (Brown *et al.*, 1981). The following examples detail environmental and genetic strategies employed during research activities to divert carbon flux away from ethanol metabolism.

The culture conditions of yeast fermentations can be maintained to ensure that a bioprocess is carried out in an aerobic fed-batch mode where glucose feeding is finely controlled to keep a very low residual concentration (<1.0 g l^{-1}) and to maintain the specific growth rate of *S. cerevisiae* below a critical value around 0.25 to 0.3 h^{-1} (strain dependent); this is known as the critical dilution rate (Sonnleitner and Kappeli, 1986). These process controls provide environmental conditions that do not permit the Crabtree effect. However, at the very large industrial scales required for biofuel production, controlling culture conditions becomes expensive and is a difficult strategy to employ due to poor mixing and non-homogeneity within the system (Flikweert *et al.*, 1997).

Large scale difficulties in mixing, as with the above process strategy, can be prevented with the use of non-fermentable carbon sources. These prevent the transition towards respiro-fermentative metabolism at high substrate concentrations (Ochoa-Estopier *et al.*, 2011). Of the known oxidative substrates assimilated by *S. cerevisiae,* both acetic acid and trehalose are considered expensive substrates and not suitable for the production of a high-volume, low-value product such as transport fuel, and ethanol usage as a carbon source would lead to direct competition with biofuel production. Glycerol is a conceivable feedstock substrate due to the growth of the bioethanol and biodiesel industries. Glycerol is a byproduct of these processes accounting for 5 to 10 % (w/w) of the final process yield (da Silva *et al.*, 2009). The key disadvantage, however, of all these non-fermentable carbon sources is that they only permit very slow growth of *S. cerevisiae*, with a maximum specific growth rate ranging from 0.05 to 0.07 h^{-1}. By comparison with glucose as the

substrate the rate is 0.44 h^{-1} (Ochoa-Estopier *et al.*, 2011; de Smidt *et al.*, 2012). Moreover, metabolism of glycerol by *S. cerevisiae* is being developed to increase ethanol yields (Yu *et al.*, 2010) potentially reducing the availability of this byproduct feedstock and raising its price.

Prevention of the Crabtree effect has also been achieved by removal of all but a single modified hexose transporter. This strain has restricted glucose uptake, switching to fermentation only when oxygen is removed. The engineered strain produces negligible amounts of ethanol compared to the wild type, however the glucose consumption rate is also severely diminished compared to the wild type strain (3.5 mmol g^{-1} h^{-1} compared to 16 mmol g^{-1} h^{-1}) leading to extended batch times of growth which negatively impact on the industrial applicability of the strain (Otterstedt *et al.*, 2004).

Addition of the metabolic inhibitor pyrazole to the growth medium has been shown to reduce the production of ethanol in a dose-dependent manner within shake flask cultures. This strategy has several disadvantages, including off-target impacts of the inhibitor leading to various effects on *S. cerevisiae* metabolism (Matsuda *et al.*, 2013).

Alleviation of glucose repression of the respiratory metabolism of *S. cerevisiae* has been achieved by deletion of the gene encoding hexokinase II, an isozyme involved in glycolysis, yielding a strain with a Crabtree-diminished phenotype. The strain initiates respiro-fermentative metabolism at a higher glucose concentration compared to the wild type. However, only a partial redirection of metabolism toward respiration is achieved and pyruvate accumulates intracellularly (Diderich *et al.*, 2001; Otterstedt *et al.*, 2004).

Diversion of carbon flux from ethanol production has been demonstrated in a strain in which pyruvate decarboxylase activity is removed through deletion of three genes. This prevents conversion of pyruvate into acetaldehyde (Figure 3.1, step 1). Experimentation however highlighted the requirement of this metabolic step to provide cytosolic acetyl-CoA as a precursor essential for lipid synthesis; a pyruvate decarboxylase-negative mutant of *S. cerevisiae* requires an exogenous source of acetyl-CoA to remain viable when metabolising glucose and is hypersensitive to high glucose concentrations (Flikweert *et al.*, 1996; Flikweert *et al.*, 1997).

Finally, deletion of the genes encoding the alcohol dehydrogenase (ADH) isozymes of *S. cerevisiae*, which catalyse the ultimate step in ethanol production (Figure 3.1, step 2) has been attempted (Drewke *et al.*, 1990; Ida *et al.*, 2012). Unlike the pyruvate decarboxylase knockout, stable disruption of this metabolic step permits production of cytosolic acetyl-CoA and synthesis of lipids. The strains however accumulate large amounts of glycerol (highlighting a redox cofactor imbalance) ultimately resulting in an overall fermentative metabolism which is ATP neutral and insufficient to drive the process of cellular maintenance (Abbott *et al.*, 2009). The strains also grow poorly due to toxic accumulation of acetaldehyde (Drewke *et al.*, 1990). It has been proposed that under aerobic culture conditions wild-type cells can rely on respiration to reoxidise excess NADH formed in biosynthesis (Flikweert *et al.*, 1997).

An industrial example of this strategy and its commercialisation is the fermentative production of succinic acid from renewable sources (Jansen *et al.*, 2012) by the company Reverdia (Geleen, Netherlands), a joint venture involving

DSM (Heerlen, Netherlands) and Roquette (Lestrem, France). The *S. cerevisiae* metabolic engineering strategy included the deletion of *ADH1* (René Verwaal (DSM Biotechnology Center, Delft, Netherlands), personal communication) in combination with process optimisation, yielding culture conditions where the yeast strain produces succinic acid at a high titre and at a low pH. This industrial bioprocess includes the seemingly wasteful diversion of carbon flux toward the production of ethanol (although at reduced amounts) and concomitant production of CO_2 in the strain. This beneficially influences succinic acid production due to the impact of CO_2 on phosphoenol pyruvate carboxykinase converting higher amounts of oxaloacetate from pyruvate (Jansen *et al.*, 2012).

In another study, Ida *et al.* (2012) evaluated a *Δadh1* genotype (similar to that used for commercial production of succinic acid) in continuous culture and suggested that the associated decrease in ethanol production was partially compensated for by the up-regulation of other ADH isozyme genes. This hypothesis was supported by increased expression of *ADH2* and *ADH4* with a concomitant increase in ethanol production during long-term culture experiments. Reversion of ethanol production would negatively impact on the industrial suitability of the strain. Cell recycling requirements for certain bioprocesses (as with current large scale cellulosic ethanol production (Basso *et al.*, 2008)), would likely lead to the adaptation of the *Δadh1* yeast strain's metabolic carbon distribution, reverting to production of ethanol and altering the economics of the process.

The advantageous production of lower amounts of ethanol in the commercial production of succinic acid, and demonstration of ethanol reversion in a single ADH isozyme deletion strain of *S. cerevisiae*, highlights the requirement to optimise both genotypic and environmental factors, the applicability of which will be dependent on the specific product and associated manufacturing process. Elucidating the impact of environmental conditions on the metabolism of ethanol in combination with various perturbations of the ADH genes, and subsequently applying this developed understanding to the industrial manufacture of products derived from *S. cerevisiae,* will be the focus of chapters 4 and 5.

3.1.3 Alcohol dehydrogenases (ADHs)

The single step enzymatic reduction of acetaldehyde to ethanol and concomitant regeneration of NAD^+ within *S. cerevisiae* is catalysed by multiple alcohol dehydrogenase (ADH) isozymes. These have been studied intensively for over half a century with data indicating that different enzymes may be expressed depending on the environmental constraints and/or specific yeast strain (Denis *et al.*, 1983). Because this reaction is reversible, ethanol also serves as a carbon substrate during aerobic respiration, when a fermentable carbon source is depleted. *S. cerevisiae* assimilates previously excreted ethanol, through the action of specific ADH isozymes. The ADH isozymes therefore play a key role in both fermentative and respiratory (oxidative) carbon metabolism, allowing the optimal use of the carbon substrates (de Smidt *et al.*, 2012).

Proteins of the enzymatic family classed as alcohol dehydrogenases (E.C. 1.1.1) act on the CH-OH group of electron donors with either NAD^+ or $NADP^+$ used as an acceptor. They are oxidoreductases that catalyse the reversible reaction of alcohols to aldehydes or ketones (having a broad range of specificity), with the concomitant reduction of either NAD^+ or $NADP^+$ (Kanehisa *et al.*, 2014). The canonical ADH monomers comprise two domains: i) a coenzyme binding domain, that contains one reactive sulfhydryl group, which upon binding one atom of zinc and one molecule of NAD^+ or $NADP^+$ forms the active site and ii) the catalytic domain of the active dimeric or tetrameric enzyme. The proposed mechanism is electrophilic catalysis mediated by the active site zinc atom. A second zinc atom can bind and provides conformational stability of the tertiary structure (Leskovac *et al.*, 2002).

3.1.4 The ADH isozymes of *Saccharomyces cerevisiae*

Research of yeast ADH genetics in haploid *adh* mutants, led to the identification of four unlinked genes, namely *ADH1*, *ADH2*, *ADH3* and *ADH4* (de Smidt *et al.*, 2008). Elucidation of yeast ADH amino acid sequences allowed structural comparisons to be made with horse liver alcohol dehydrogenase, which had been studied in detail, and a comparative evaluation of the yeast ADH isozymes was performed (Wills and Jörnvall, 1979). The complete sequence of the *S. cerevisiae* genome (Goffeau *et al.*, 1996) revealed the existence of c. 6,000 open reading frames (ORFs); Table 6 details the subsequent identification and functions of the ORFs encoding proteins classified as or related to the ADH class of enzymes.

Table 6. The alcohol dehydrogenase genes of *Saccharomyces cerevisiae* (adapted from Dickinson, 2004).

Gene	ORF	Functions
ADH1	YOL086c	Main cytosolic ADH, forms ethanol in glycolysis
ADH2	YMR303c	Cytosolic, glucose repressed, used for growth on ethanol
ADH3	YMR083w	Mitochondrial, role as redox shuttle
ADH4	YGL256w	Induced in response to zinc deficiency
ADH5	YBR145w	Paralog of *ADH1*, arose from whole genome duplication
SFA1	YDL168w	Part of a bifunctional formaldehyde dehydrogenase
BDH1	YAL060w	2,3-butanediol dehydrogenase
BDH2	YAL061w	Putative ADH, 51 % identical to *BDH1*
SOR1	YJR159w	Sorbitol dehydrogenase
SOR2	YDL246c	Unknown function, 99 % identical to *SOR1*
XDH1	YLR070c	Xylitol dehydrogenase
ADH6	YMR318c	NADPH-dependent medium chain cinnamyl ADH
ADH7	YCR105w	NADPH-dependent medium chain cinnamyl ADH

Experimentation on a quadruple (Δ*adh1/2/3/4*) non-revertant strain of *S. cerevisiae* and evaluation of the transcriptional profile of the *ADH5* isozyme under standard growth conditions, revealed the existence of a novel acetaldehyde-reducing activity (Drewke *et al.*, 1990). The strain still produced ethanol in the absence of the five evaluated ADH isozyme genes, in addition Ida *et al.* (2012) demonstrated that the knockout of six ADH isozymes (Δ*adh1/2/3/4/5 and* Δ*sfa1*) is required to stably disrupt ethanol production under standard growth conditions. These ADH isozymes specifically involved in ethanol metabolism in *S. cerevisiae* are the focus of this research.

The relatedness of the protein sequences of the six ADH isozymes (Figure 3.2) has been summarised, along with their individual regulation, expression and activity, in Table 7. The presence of multiple isozymes catalysing a single reaction step is thought to allow *S. cerevisiae* greater flexibility within different environmental niches, and the ADH isozymes involved with ethanol metabolism have been shown to functionally substitute for one another (de Smidt *et al.*, 2012).

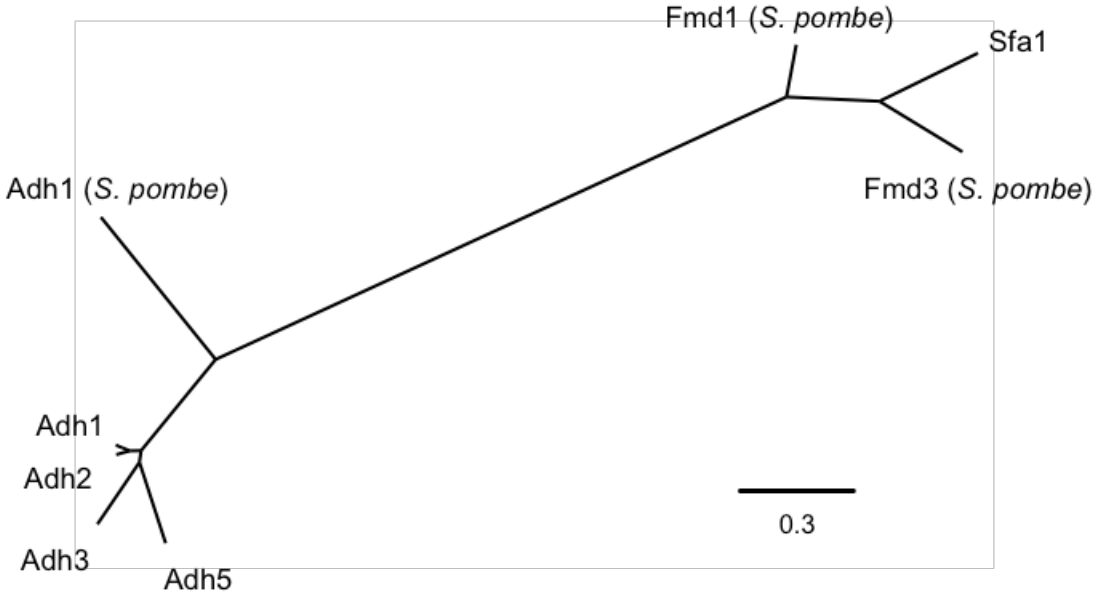


Figure 3.2. Phylogenetic tree of the amino acid sequences of ADH isozymes of *S. cerevisiae* and the fission yeast *Schizosaccharomyces pombe* specifically involved in ethanol metabolism. Adh4 proteins have little sequence similarity to the other ADH isozymes and are not included. Branch lengths are proportional to the number of substitutions per site, as indicated by the marker bar (phylogenetic analysis of ADH isozyme protein sequence performed according to method as Dereeper *et al.*, 2008).

The evolutionary history leading to divergence of the ADH isozymes has yet to be fully established. Comparative genomic and biochemical evidence points toward an ancestral progenitor gene which did not provide yeast the means to

accumulate ethanol for later consumption but rather encoded a recycle function for NADH generated in the glycolytic pathway (Thomson *et al.*, 2005). This ancestral gene was duplicated (including part of the non-coding region), and subsequently part of the gene encoding the N-terminal 28 amino acids (associating the protein with the mitochondrial membrane) was lost from one copy. Further mutations gave rise to two genes: the highly expressed *ADH1*, which encoded an isozyme specialised for fermentation, and the *ADH2* gene, encoding an isozyme that was more adapted to ethanol oxidation during gluconeogenesis. This meant that *S. cerevisiae* could now re-assimilate ethanol. The diverged roles of Adh1 and Adh2 allowed the isozyme with the intact amino terminal sequence, Adh3, to diverge with a specialised role in respiratory metabolism within the mitochondria (Young and Pilgrim, 1985).

Table 7. Comparative features of the six ADH isozymes in *S. cerevisiae* specifically involved in ethanol metabolism.

ADH Isozyme	Length (amino acid residues)	Size (kDa)	Sequence identity to Adh1 (%)	Quaternary structure [a]	Cellular localisation [e]	K_m for NAD^+ (μM) [g]	K_m for ethanol (μM) [g]	K_m for acetaldehyde (μM) [g]
Adh1	348	36.85	n/a	Homotetrameric [b]	Cytoplasm	170	17,000	11,000
Adh2	348	36.74	93	Homotetrameric [b]	Cytoplasm	110	810	90
Adh3	375	40.37	80	Tetrameric [c]	Mitochondrial	240	12,000	440
Adh4	382	41.15	8	Dimeric [d]	Cytoplasm / Mitochondrial	59	16,700	2,830
Adh5	351	37.65	77	Tetrameric	Cytoplasm? [f]			
Sfa1	386	41.05	23	Tetrameric	Cytoplasm / Mitochondrial	90		

[a] All isozymes have associated zinc ion(s); [b] Mortlock (2013); [c] Zimmermann and Entian (1997); [d] Bird *et al.* (2006); [e] Huh *et al.* (2003); [f] Drewke *et al.* (1990); [g] Adh1 and Adh2 data from Ganzhorn *et al.* (1987), Adh3 and Adh4 data from Drewke and Ciriacy (1988), Sfa1 data from Wehner *et al.* (1993).

Reduction of acetaldehyde to produce ethanol (essential for ATP production and regeneration of NAD^+ from NADH formed during glycolysis) is catalysed primarily by the Adh1 isozyme (Drewke *et al.*, 1990; Thomson *et al.*, 2005). The abundant activity of the isozyme meant that originally it was proposed to be constitutively expressed. However studies have since shown *ADH1* to be completely repressed by growth under conditions of extreme aerobiosis or on non-fermentable substrates (de Smidt *et al.*, 2012). Reduced protein expression when cells are grown on ethanol however indicates a lack of glucose induction rather than ethanol repression (Bennetzen and Hall, 1982; van den Berg *et al.*, 1998). Indeed expression can be induced 5- to 10-fold when grown on glucose (Zimmermann and Entian, 1997). At the onset of the ethanol assimilation phase of diauxic growth, *ADH1* mRNA transcription levels decrease 33-fold before gradually increasing as growth on ethanol continues (de Smidt *et al.*, 2012). Denis *et al.* (1983) performed radiolabelling degradation studies on the Adh1 protein and found that the protein does not rapidly turn over and therefore can give the appearance of constitutive expression in some experiments. The Adh1 protein has a high K_m for ethanol (17,000 μM), ensuring ethanol can reach high concentrations before the reaction equilibrium shifts (Wills, 1976a; Mortlock, 2013).

The metabolic importance of the Adh1 protein in combination with the Adh2 protein is highlighted by their abundance, together constituting 1.0 % of the total soluble protein from yeast cell preparations (Wills, 1976b). The primary function of the Adh2 protein is to oxidise ethanol (typically previously excreted) to acetaldehyde, which can then be metabolised via the tricarboxylic acid cycle or

as an intermediate in gluconeogenesis (Young and Pilgrim, 1985; de Smidt *et al.*, 2012). The Adh2 protein has a remarkably conserved sequence homology with the Adh1 protein (93% identity) and indeed in their putative binding pockets only a methionine to leucine change at position 294. This is however not the cause for the difference in substrate binding (Ganzhorn *et al.*, 1987). The K_m for ethanol of Adh2 is one-tenth that of Adh1 and therefore low concentrations of ethanol are metabolised efficiently by Adh2 compared to Adh1 (Mortlock, 2013).

The molecular signals that regulate expression of *ADH2* (as with many glucose-sensitive genes) remain to be fully elucidated. The gene is oxygen inducible and is also thought to be regulated by catabolite repression, with transcription being markedly but not completely repressed by growth on glucose (Ciriacy, 1979; van den Berg *et al.*, 1998; de Smidt *et al.*, 2008). Additional complexity in ADH2 regulation was highlighted by van den Berg *et al.* (1998) who showed that *Adr1*, a transcription factor regulating *ADH2* transcription, is phosphorylated (and thereby inactivated) during growth on both glucose and ethanol. Repression of *ADH2* in the presence of the enzymes preferred substrate, ethanol, highlights the remarkable and complex regulation of the ADH isozymes within yeast.

It has been proposed that Adh2 arose from a gene duplication event (Wills and Jörnvall, 1979; Kellis *et al.*, 2004) and that it became responsible for the dominant activity replacing the function of Adh3*,* which has now acquired a more specialised role within the mitochondria. An N-terminal extension of 28 residues which associates Adh3 with the mitochondria is cleaved during the translocation process (Zimmermann and Entian, 1997). Adh3 functions as part

of the ethanol acetaldehyde shuttle (Figure 3.3), which under anaerobic conditions is required to reoxidise mitochondrial NADH (Bakker *et al.*, 2000). The pyridine nucleotides NAD^+ and NADH cannot cross the mitochondrial inner membrane (Heux *et al.*, 2006) whereas ethanol and acetaldehyde can and, providing there is a cytosolic alcohol dehydrogenase (de Smidt *et al.*, 2008), the essential reoxidation of mitochondrial NADH during anaerobic growth can occur.

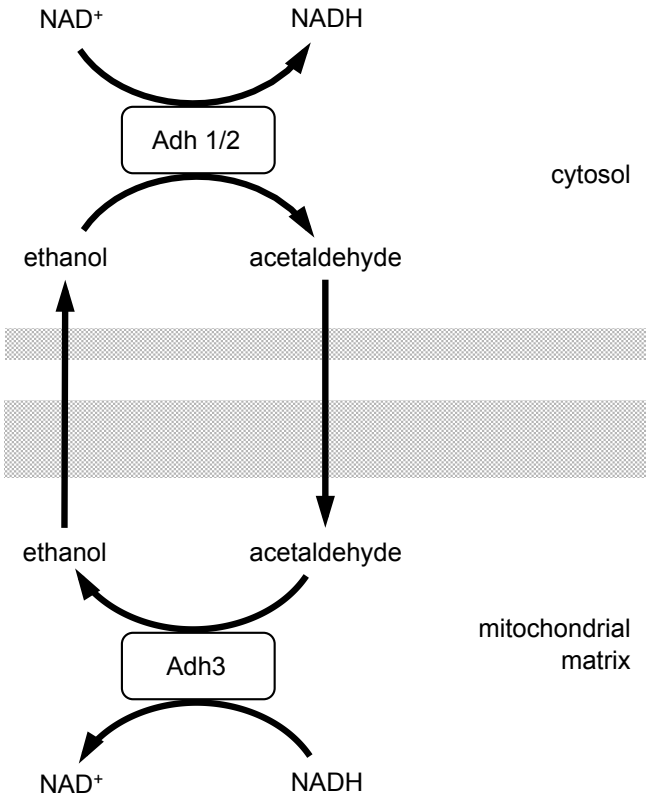


Figure 3.3. The ethanol-acetaldehyde shuttle of *S. cerevisiae* (adapted from Bakker *et al*., 2000).

Transcription of *ADH3* is constitutive during growth on both glucose and ethanol (de Smidt *et al.*, 2012), and is repressed (as is the transcription of *ADH1*) in zinc-limited conditions (Bird *et al.*, 2006). The expression of *ADH4* is, however,

upregulated under low zinc conditions (de Smidt *et al.*, 2012) and it has been postulated that its expression is advantageous during zinc starvation either due to the protein binding only one atom of zinc per subunit (it is a dimer as opposed to the other tetrameric ADH isozymes) or it has a more efficient catalytic activity than Adh1 under conditions of zinc limitation (Bird *et al.*, 2006).

The expression and importance of *ADH4* in *S. cerevisiae* is yet to be fully elucidated, with reports of a lack of expression in some laboratory strains of *S. cerevisiae* (Zimmermann and Entian, 1997; van den Berg *et al.*, 1998; de Smidt *et al.*, 2012) and expression in others (Drewke and Ciriacy, 1988). Additionally, comparison of mRNA expression levels in the *S. cerevisiae* brewing strain NCYC1245 suggests Adh4 to be the main cytosolic ADH (Mizuno *et al.*, 2006).

Adh4 shares very little sequence similarity with the other ADH isozymes; it is more closely related to the iron-activated ADH from *Zymomonas mobilis* and other bacteria (Branden, 1975; Williamson and Paquin, 1987) but requires zinc for activation (Drewke and Ciriacy, 1988; Yuan, 2000). The *in vitro* substrate specificity profile of Adh4 also supports its separation from the other ADH isozymes as it cannot utilise 1-butanol as a substrate (Williamson and Paquin, 1987). It is thought that *ADH4* may represent a gene that has undergone significant evolutionary changes. This may be a consequence of not being essential for yeast growth (Drewke and Ciriacy, 1988), and/or that the strong sequence homology to the iron-activated ADH from bacteria suggests *ADH4* may have entered the yeast genome as a result of horizontal gene transfer, an event which, although rare, has been found to occur in *S. cerevisiae* (Hall *et al.*, 2005).

Adh5 was discovered from sequence homology to Adh1 (77 % amino acid identity) during sequencing of the *S. cerevisiae* genome (Feldmann *et al.*, 1994). The distinct function and transcriptional regulation of *ADH5* remain unclear. Chemostat glucose-pulse experiments concluded that *ADH5* was expressed at a constant level during experimentation, and in a study deleting *ADH1/2/3/4*, the strain still produced ethanol and this was proposed to be due to the function of Adh5 (Drewke *et al.*, 1990; van den Berg *et al.*, 1998).

A stable disruption of ethanol production including knockout of the *SFA1* gene (Ida *et al.*, 2012) highlighted the involvement of Sfa1 in ethanol metabolism. Gene deletion studies of *SFA1* (sensitive to formaldehyde) demonstrated the gene to be non-essential (Wehner *et al.*, 1993). Sfa1 is bi-functional, with both glutathione-dependent formaldehyde and NADH-dependent alcohol dehydrogenase activities (Dickinson, 2004) and, as with the other ADH isozymes, it is involved in fusel alcohol production via the Ehrlich pathway (Hazelwood *et al.*, 2008). The similarity of *Sfa1* to mammalian and yeast ADH isozymes (Dickinson *et al.*, 2003) again highlights its potential function in ethanol metabolism despite reports that the isolated protein showed no *in vitro* catalytic activity with either ethanol or acetaldehyde (Wehner *et al.*, 1993).

3.1.5 ADH gene deletions in *S. cerevisiae*

S. cerevisiae strains with a *Δadh1* genotype have a significantly reduced rate of growth and production of ethanol with concomitant accumulation of glycerol (Zimmermann and Entian, 1997; Ida *et al.*, 2012). Drewke *et al.* (1990) constructed a quadruple deletion (Δ*adh1/2/3/4*) non-revertant strain of *S. cerevisiae* and concluded that deletion of the ADH isozymes increased production of glycerol, attributed to the operation of Neuberg's second form of fermentation, where the reduction equivalents are transferred to dihydroxyacetone phosphate to form glycerol-3-phosphate that is then converted to glycerol. The deletion strain produced significant quantities of ethanol, indicating the existence of other ADH isozymes. Subsequent analysis was performed by de Smidt *et al.* (2012), where quadruple knockout combinations of the ADH genes *ADH1/2/3/4/5* were generated in order to evaluate individual ADH isozyme functions (5 separate genotypes). These experiments demonstrate that the ADH isozymes of *S. cerevisiae* are capable of functionally substituting for each other. Evaluation of different ADH genotypes using either glucose or ethanol substrates show that *ADH1* is capable of utilising ethanol, a function previously attributed to *ADH2*. Additionally, growth on glucose with concomitant ethanol production is observed in a genotype with only *ADH3* present. Moreover the same genotype can assimilate ethanol, indicating that Adh3 can perform both the functions of Adh1 and Adh2. However, differences in strains, methods of gene knockout (including potential functional reversion and the use of auxotrophic markers)

and environmental conditions used in experimentation are all potential sources of disagreements within the literature.

Generating strains to include all permutations of the ADH genotypes would provide a useful toolset for understanding how these enzymes functionally substitute for each other *in vivo*. Additionally, the ADH genotype toolset would be useful in elucidating the interactions of genotype and environmental conditions applicable for minimisation of ethanol by-product formation, maintenance of rapid carbon flux and retention of other beneficial biotechnological characteristics (Jouhten *et al.*, 2008).

3.1.6 Aims

The aims of experiments described in this chapter are to optimise a dominant method of gene deletion and marker recycling in an industrially relevant strain of *S. cerevisiae* and generate a complete set of *S. cerevisiae* genotypes deficient in all permutations of the ADH genes. This will provide a library for subsequent evaluation of functional substitutions and genotype by environment interactions related to perturbations of the ADH genes.

3.2 Results and discussion

3.2.1 *S. cerevisiae* strain selection

Auxotrophic laboratory strains of *S. cerevisiae* are a convenient tool for research activities (Pronk, 2002). However, the genetic modification related to the auxotrophy and cultivation requiring the addition of amino acids or other supplements complicates physiological studies (van Dijken *et al.*, 2000). It is therefore desirable to use a prototrophic strain of *S. cerevisiae*.

The prototrophic, haploid *S. cerevisiae* strain CEN.PK113-7D was selected for experimentation as it offers a good compromise between genetic accessibility and physiological properties (industrially relevant advantages) based on its growth characteristics in shake-flask cultures, range of specific growth rates that support respiratory growth, and transformation efficiency (van Dijken *et al.*, 2000). The strain has an available genome sequence, assembled, annotated and analysed (Nijkamp *et al.*, 2012). Even after prolonged cultivation in chemostat cultures CEN.PK113-7D shows no mating-type reversal. It has been successfully grown in high-cell-density fermentations (130 g dry biomass l^{-1}), and additionally the strain is an excellent host for heterologous protein production (specific productivities ≤ 2 mg g^{-1} biomass h^{-1}) (van Dijken *et al.*, 2000). The strain has characteristics which make it less applicable for certain studies. For example, signal transduction studies of this strain revealed a lack of characteristic transient accumulation of cAMP upon growth on excess glucose, however this had no apparent effect on growth or extracellular product formation (van Dijken *et al.*, 2000).

Within the yeast research community, in attempts to quantifiably standardise various physiological and genetic studies, CEN.PK113-7D is becoming a popular choice as the reference strain for *S. cerevisiae*. The beneficial characteristics of the CEN.PK113-7D strain mean it is also being developed for use within current industrial processes (Westfall *et al.*, 2012; Ito *et al.*, 2014; Ahn *et al.*, 2016).

3.2.2 Multiple ADH gene deletion strategy and design of deletion cassettes

The 64 genotypes required to evaluate all possible permutations of the ADH isozyme deletions are shown in Figure 3.4. Manipulation of this organism's genome in order to achieve these 64 combinations is greatly enhanced by its ability to effectively perform homologous recombination, allowing integration of desirable DNA sequences at targeted genomic sites using endogenous cellular DNA repair mechanisms, and deletion of target genes (Fishman-Lobell *et al.*, 1992; Brachmann *et al.*, 1998).

Genotype

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
ADH1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ADH2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
ADH3	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−
ADH4	+	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−
ADH5	+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−
SFA1	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−

Genotype

	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64
ADH1	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
ADH2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−
ADH3	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−	+	+	+	+	+	+	+	+	−	−	−	−	−	−	−	−
ADH4	+	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−	+	+	+	+	−	−	−	−
ADH5	+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−	+	+	−	−
SFA1	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−	+	−

Figure 3.4 An orthogonal array highlighting the 64 genotypes required for evaluation of all permutations of the ADH isozymes of the yeast *S. cerevisiae*. (+) represent intact ADH isozyme genes; (−) represent gene deletions.

Using multiple selectable markers to achieve sequential gene knockouts is likely to exacerbate the issues associated with auxotrophic strains (van Dijken *et al.*, 2000) and limit subsequent genetic engineering of the developed strain. Additionally, selection of strains which use selectable markers that confer resistance to antibiotics or other toxic compounds are cost prohibitive when used in large-scale processes (Pronk, 2002).

Marker recovery strategies can be used to circumvent these issues; a widely utilised example is that of the Cre-*loxP* site-specific recombination system of bacteriophage P1 (Sauer, 1987). The strategy involves having *loxP* palindromic repeat sequences flanking the selectable marker. After deletion of a target gene the marker module is excised by the inducible Cre recombinase that specifically targets the *loxP* sites, replacing the marker with a single *loxP* sequence. However, *loxP* sites are retained within the genome as “scars” and repetitive use of the Cre-*loxP* system for gene deletion and marker recycling has been shown to contribute towards large genome alterations influencing the physiology and response of *S. cerevisiae,* specifically chromosomal translocations, truncations and gene loss (Solis-Escalante *et al.*, 2015). These are all undesirable side effects for generation of an industrial strain of *S. cerevisiae*.

The *amdSYM* method of gene deletion (Solis-Escalante *et al.*, 2013) has the potential to eliminate undesirable genome instability side effects and limitations on strain choice. This method utilises a codon-optimised acetamidase gene from *Aspergillus nidulans*, which can be heterologously expressed in *S. cerevisiae* and will allow breakdown of acetamide into acetate and ammonia,

providing a nitrogen source to support growth of the culture. This provides selection linked to nitrogen availability which is both dominant and requires no specific genetic background. Effective *amdSYM* marker removal is performed by incorporating into the marker module a 40 base pair sequence of the region directly upstream of the "left hand" homologous sequence used for integration of the deletion cassette (Figure 3.5A – orange sequences). Homologous recombination between this sequence and the corresponding direct repeat sequence in the host genome leads to scarless excision of the *amdSYM* marker. This can be selected by growing the culture in the presence of a defined metabolisable nitrogen source in the presence of fluoroacetamide; if the selection marker (the acetamidase gene) is still present fluoroacetamide is metabolised into the toxic product fluoroacetate, disrupting cell growth. Once the genotype is verified the strain can be used for subsequent gene deletions, allowing the generation of a stable ADH deletion library for assessment of the ADH genes in combination with environmental impacts on the fermentation capacity of *S. cerevisiae*.

The dominant counter-selectable *amdSYM* method of genetic manipulation permits use of prototrophic strains of *S. cerevisiae*, including industrial strains, which is a requirement of the aims of this research. Subsequent assessments of phenotypes are not influenced by remnants of the gene knockout method which ultimately provides an industrially relevant assessment of the system.

Figure 3.5(A) shows that the use of the homologous sequences for gene deletion described by Solis-Escalante *et al.* (2013) deletes a portion of flanking sequence upstream of the open reading frame (ORF), and the stop codon of

the gene remains. The *amdSYM* method described was therefore adapted to generate genotypes where only the ORF of targeted genes was deleted. Design of the deletion cassettes therefore included the start codon and subsequent 50-55 bp of ORF for the “left hand” homologous sequence, and removal of the stop codon from the “right hand” homologous sequence (Figure 3.5B). These adaptations permit an exact deletion of the genomic ORF with no loss of upstream flanking sequences (Figure 3.5B).

In order to increase throughput of the methodology, the production of linearised deletion cassettes specific to each of the ADH isozymes was optimised. The deletion cassettes were synthesised within an *Escherichia coli* expression vector, which provided increased amounts of DNA compared to PCR amplification. The deletion cassettes were flanked by *Bmr*I (a type II nuclease) sites within the expression vector such that plasmid linearisation using *Bmr*I left no remnants of the restriction site or the expression vector backbone within the specific deletion cassette used for yeast transformations.

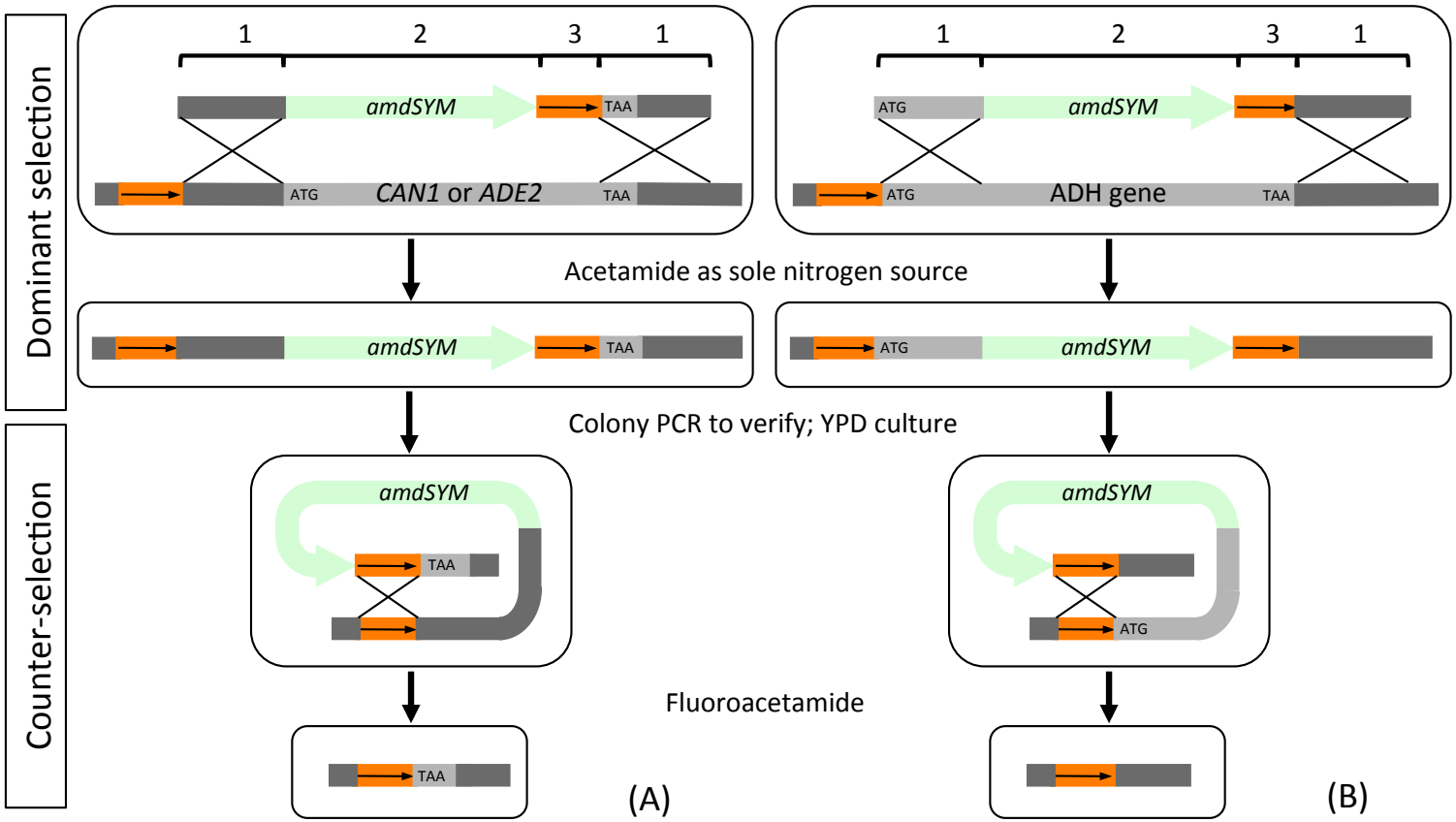


Figure 3.5 The *amdSYM* method of gene deletion in *S. cerevisiae*. The deletion cassette includes 1) 50-55 bp gene-specific sequences for homologous recombination; 2) *amdSYM* marker module; and 3) 40 bp direct repeat sequence for scarless marker recovery. (A) As described by Solis-Escalante *et al.* (2013); note that this leads to deletion of upstream flanking sequences of the target gene (dark grey) during the counter-selection step. (B) Adaptation of homologous recombination sequences, which after marker excision permit exact ORF deletion with no loss of upstream flanking sequences.

3.2.3 Workflow for combinatorial ADH isozyme gene deletion(s)

The workflow detailed in Figure 3.6 was developed for generation of the 64 ADH isozyme genotypes. The workflow included decision review points and validation of the desired genotypes using diagnostic colony PCR and replica plating on to selection media. The successful implementation of the workflow ensured rapid and accurate production of the genotypes required to evaluate all combinations of the ADH isozymes.

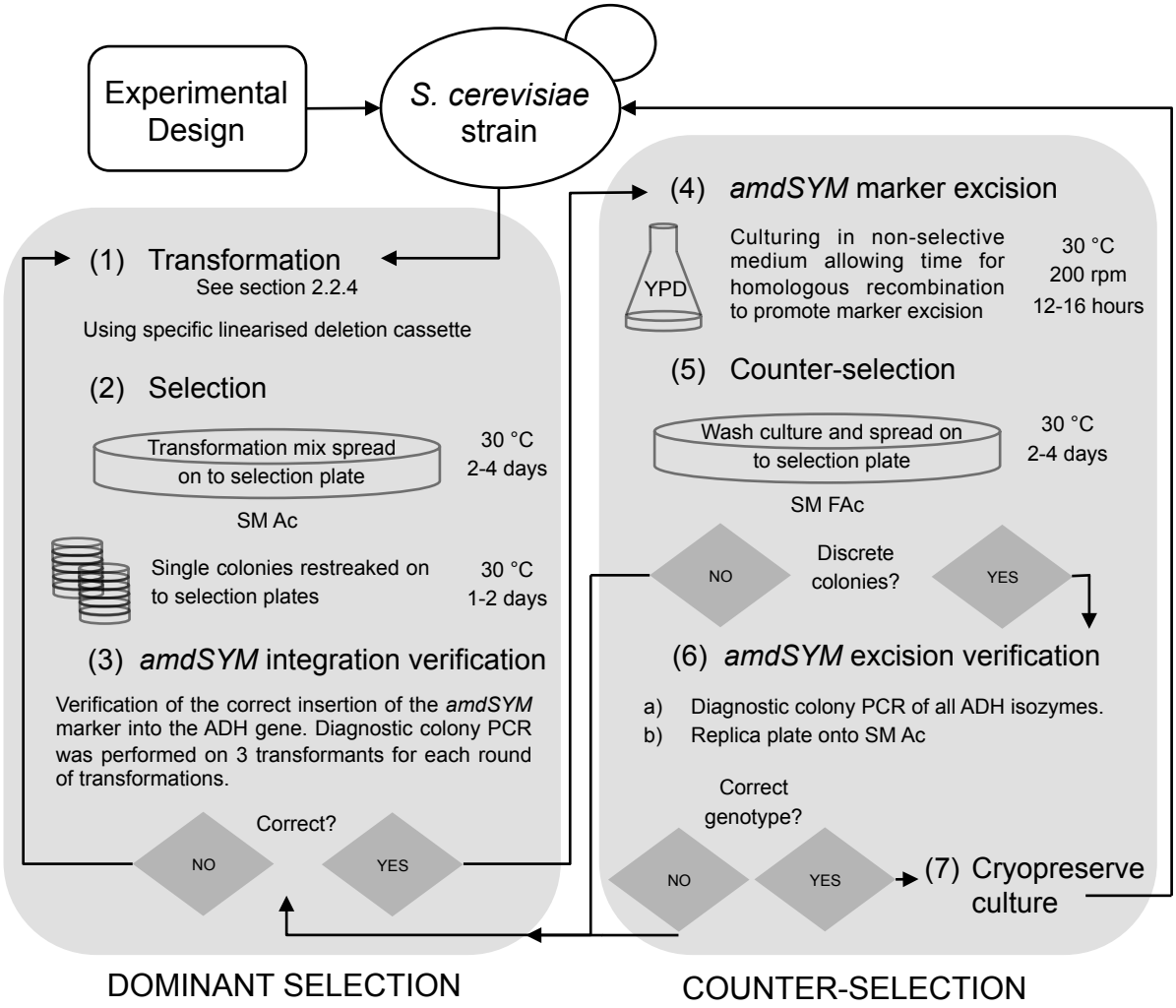


Figure 3.6 Workflow diagram for generating *S. cerevisiae* combinatorial ADH isozyme gene deletions. Ac = acetamide; FAc = fluoroacetamide.

3.2.4 Genotype verification using diagnostic colony PCR

Standardisation of diagnostic primers and PCR conditions was desirable to reduce experimental complexity and increase the throughput of experimentation. Primer recognition sites were designed to flank the coding sequence of each of the ADH genes, and were therefore not changed due to the adapted *amdSYM* gene deletion method (Figure 3.5B). A single set of primers was used for each of the ADH genes, providing a positive diagnostic for the relevant intact gene, the integration of the selection marker, or the excision of both.

At least a single G or C nucleotide was included as the terminal 3´ nucleotide of each primer sequence in order to provide maximum binding stability. Primer lengths of between 22-25 nucleotides were used in order to maintain specificity and provide comparable annealing temperatures for specific primer combinations as well as between distinct primer sets; this was standardised to 63.3 °C after gradient PCR evaluation (data not shown).

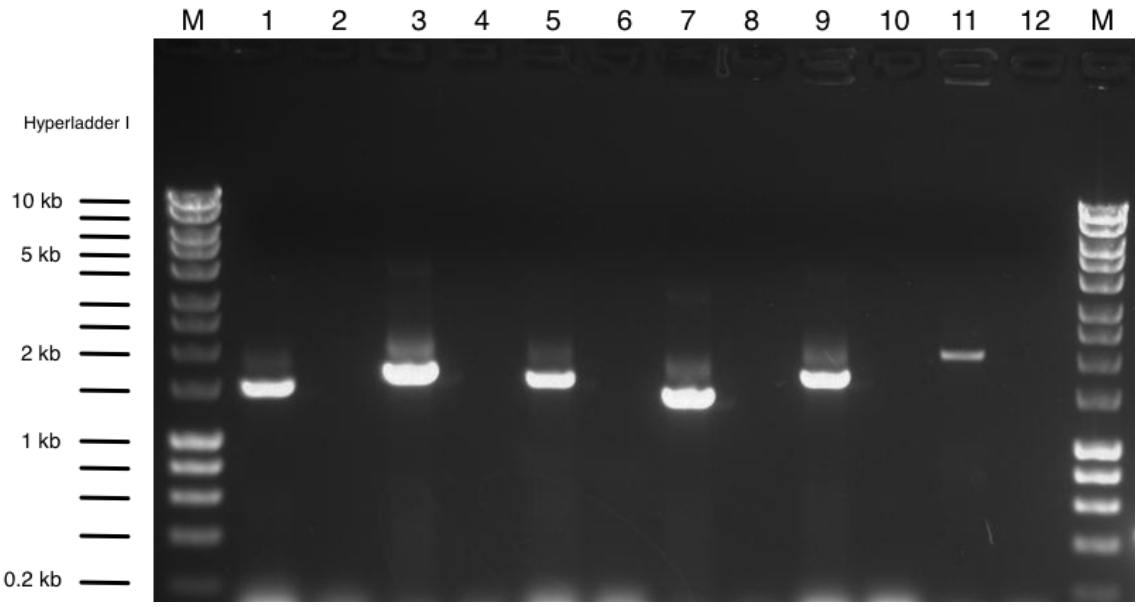


Figure 3.7 Diagnostic colony PCR verification of ADH genes of the wild type *S. cerevisiae* CEN.PK113-7D strain. All PCRs were performed at the standardised annealing temperature of 63.3 °C. Genes and expected fragment sizes are: Lane 1, *ADH1* (1,509 bp); Lane 3, *ADH2* (1,675 bp); Lane 5, *ADH3* (1,609 bp); Lane 7, *ADH4* (1,426 bp); Lane 9, *ADH5* (1,678 bp); Lane 11, *SFA1* (1,997 bp). Lanes 2, 4, 6, 8, 10 and 12 were respective negative controls.

Figure 3.7 shows the colony PCR screening for all six of the ADH genes in the wild type CEN.PK113-7D *S. cerevisiae* strain. The band positions indicating amplified DNA fragment sizes for all the ADH genes are of the expected sizes for the intact wild-type gene. The no template negative controls show the results are not from contamination or non-specific DNA amplification.

The full evaluation of the verification primers was performed once gene deletion and marker excision templates had been generated using the *amdSYM* method.

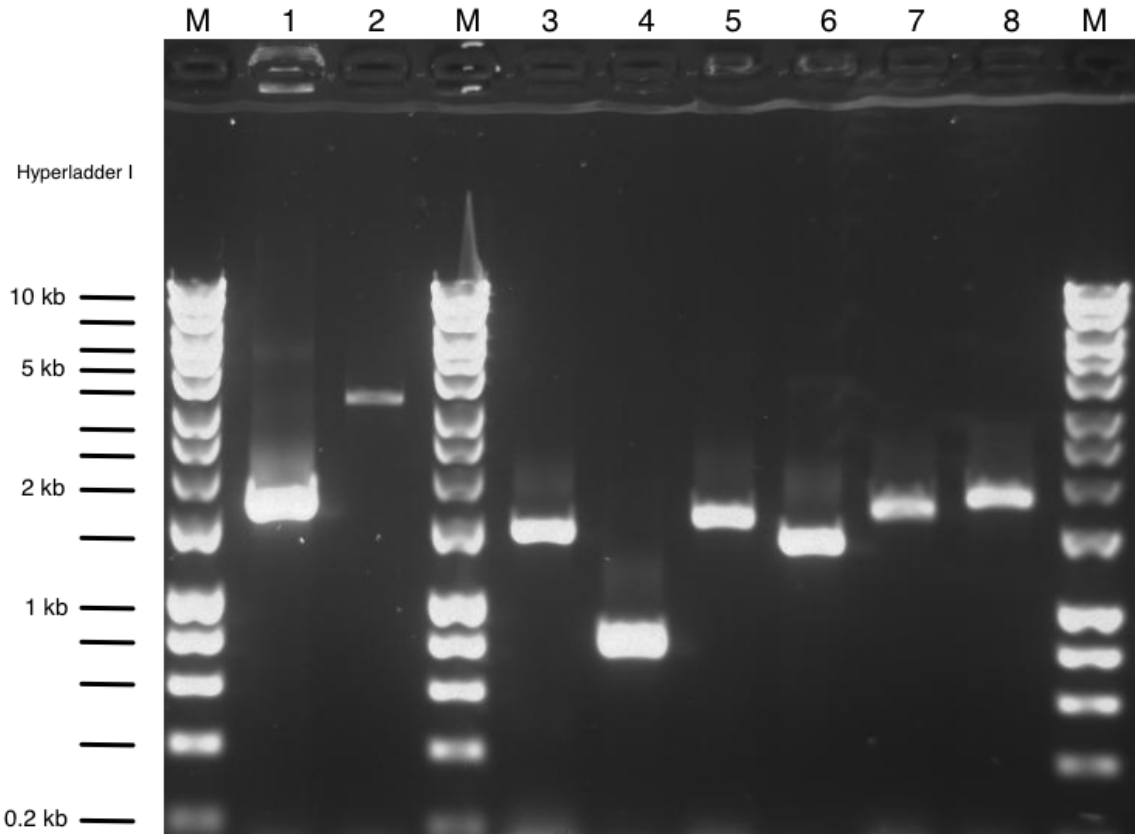


Figure 3.8 Diagnostic colony PCR verification of a CEN.PK113-7D *Δadh2*::*amdSYM* genotype and subsequent excision of the *amdSYM* selectable marker. All PCRs were performed at the standardised annealing temperature of 63.3 °C. Genes, genotype variants and expected fragment sizes are: Lane 1, *ADH2* (1,675 bp); Lane 2, *Δadh2*::*amdSYM* (3,130 bp); Lane 3, *ADH1* (1,509 bp); Lane 4, *Δadh2* (628 bp); Lane 5, *ADH3* (1,609 bp); Lane 6, *ADH4* (1,426 bp); Lane 7, *ADH5* (1,678 bp); Lane 8, *SFA1* (1,997 bp).

An example diagnostic PCR validation is presented in Figure 3.8 for a correct deletion of *ADH2* from the wild type *S. cerevisiae* CEN.PK113-7D. The single primer combination evaluated the intact *ADH2* (using the wild type reference strain; Lane 1), the successful integration of the *amdSYM* selection marker in place of *ADH2* (*Δadh2*::*amdSYM*; Lane 2), and the subsequent excision of the selection marker (*Δadh2*; Lane 4).

The strain genotype can be determined by the banding patterns of amplified DNA specific to each of the ADH genes (Lanes 3-8). The DNA fragment size in lane 4 is that expected of *Δadh2* with complete removal of the *amdSYM* selectable marker. All other ADH primer combinations verify corresponding wild type genes. Therefore this genotype is *Δadh2*, which is genotype number 17 of the ADH isozyme combinations shown in Figure 3.4.

The standardised conditions and primer combinations for verification of the ADH deletions were found to be applicable to all of the ADH genes and intermediates of the *amdSYM* gene deletion method (data not shown). This provided an accurate and high throughput methodology for the screening, selection and validation necessary for generation of the 64 different ADH isozyme genotypes required.

3.2.5 Transformation and *amdSYM* selectable marker integration efficiency

Solis-Escalante *et al.* (2013) reported the average transformation efficiency from *amdSYM*-mediated deletion of *CAN1* in *S. cerevisiae* strain CEN.PK113-7D to be 14 transformants μg^{-1} of DNA, with 100 % of the colonies having the correct integration. The transformation efficiency was reduced compared to studies using other dominant selectable markers; this was thought not to be due to the nature of the selection marker but rather the localisation of the gene deletion. Nucleosome density has been shown to be a critical factor determining transformation efficiency (Aslankoohi *et al.*, 2012), and therefore the different

ADH genes may exhibit different efficiencies which may also require method optimisation.

Transformations (as detailed in section 2.2.4) of each of the linearised ADH gene deletion cassettes were performed using the parental *S. cerevisiae* CEN.PK113-7D strain, and the efficiencies recorded (Table 8).

Table 8. Transformation efficiencies for each of the ADH gene deletion cassettes.

Deletion cassette	Transformants μg^{-1} DNA
ADH1::*amdSYM*	2
ADH2::*amdSYM*	29
ADH3::*amdSYM*	30
ADH4::*amdSYM*	21
ADH5::*amdSYM*	48
SFA1::*amdSYM*	72

The transformation efficiencies were deemed acceptable for all deletion cassettes apart from the *ADH1*-specific cassette, which resulted in only two transformants per µg of DNA. This would likely reduce the throughput for generation of the multiple ADH isozyme genotypes required. There are many potential causes for this low transformation efficiency including: (i) reduced fitness of an *Δadh1* genotype (maximum specific growth rate of 0.44 h^{-1} of the wild type *S. cerevisiae* strain is reduced to 0.20 h^{-1} with a *Δadh1* genotype (de Smidt *et al.*, 2012)), (ii) impact of nucleosomes (Aslankoohi *et al.*, 2012) and (iii) poor specificity of the homologous sequences for integration of the *amdSYM* selection marker (the 5´ homologous sequence within the *ADH1* gene including the start codon shares 89 % identity to the closely related *ADH2* gene).

The length of the 5´ homologous sequence for the *ADH1* deletion cassette was increased from 55 bp to 100 bp (sequence variation as section 2.2.10, Table 4) using the Q5® site-directed mutagenesis protocol. The transformation efficiency of the longer *ADH1* deletion cassette (sequence verified) was increased to 37 transformants per µg of DNA and this cassette was used for all subsequent relevant transformations.

To determine correct integration (the correct location of genome integration and the integration of the complete deletion cassette) efficiencies of *amdSYM* at each ADH locus, diagnostic PCR screening was carried out on 12 discrete colonies for each transformation, after acetamide selection.

Table 9. *amdSYM* marker integration expected PCR fragment sizes and observed efficiencies (12 discrete colonies) for each of the ADH genes.

Deletion cassette integration	Expected PCR fragment size (bp)	Integration efficiency (%)
Δadh1::*amdSYM*	2,963	83
Δadh2::*amdSYM*	3,130	100
Δadh3::*amdSYM*	2,986	83
Δadh4::*amdSYM*	2,781	92
Δadh5::*amdSYM*	3,124	83
Δsfa1::*amdSYM*	3,338	83

The *amdSYM* integration efficiencies were high for all deletion cassettes (Table 9). As a result the workflow was adapted to evaluate a reduced number of transformants: three discrete colonies in each case.

3.2.6 Method optimisation for *amdSYM* deletion cassette excision

The excision of the *amdSYM* deletion cassette after counter-selection with fluoroacetamide was apparently successful for five out of the six ADH genes, with 10-60 % of discrete colonies tested giving the short PCR fragment expected when the ADH gene has been deleted and the deletion cassette has been excised (Figure 3.9).

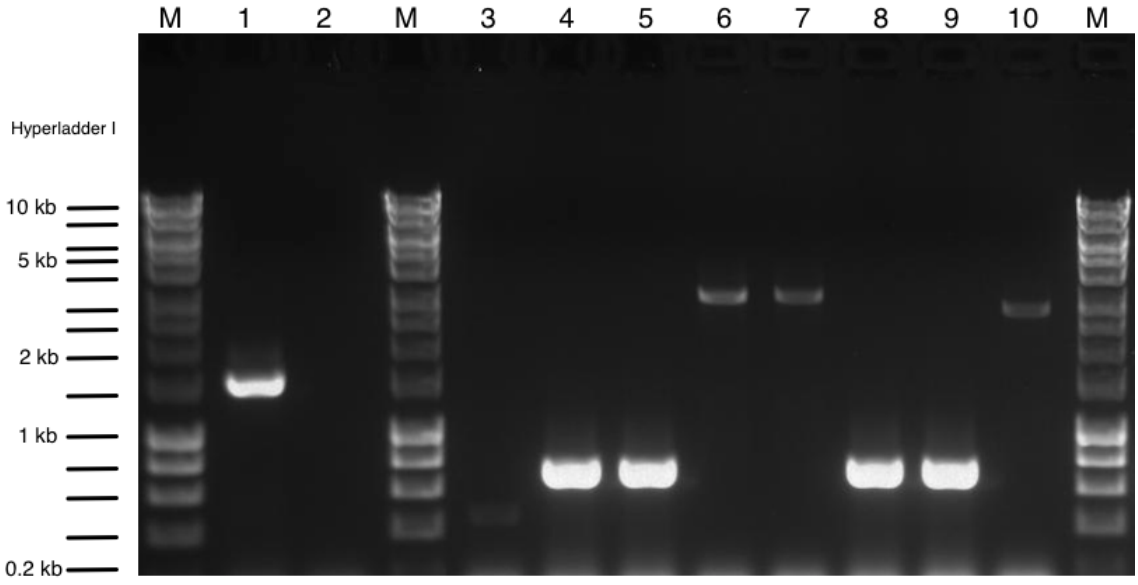


Figure 3.9 Diagnostic colony PCR verification of discrete colonies screened for a CEN.PK113-7D *Δadh5* genotype including excision of the *amdSYM* selectable marker. All PCRs were performed at the standardised annealing temperature of 63.3 °C. Expected fragment sizes are: Lane 1, *ADH5* (1,678 bp); Lane 2, Negative control; Lane 3-10, *Δadh5* (672 bp). The band positions in Lanes 6,7 and 10 indicate the intact *ADH5 amdSYM* deletion cassette (3,124 bp).

Unlike the other ADH gene deletion cassettes, the excision of the correctly integrated *amdSYM* marker used for the deletion of the *ADH2* gene was unsuccessful. Diagnostic colony PCR assessment (not shown) of 20 discrete

colonies cultured on SM agar containing fluoroacetamide (SM FAc agar), to promote the excision of the *amdSYM* deletion cassette, showed the *amdSYM* marker had not excised; the diagnostic *Δadh2*::*amdSYM* PCR band of 3,130 bp was visible for all colonies. The growth of *Δadh2*::*amdSYM* transformants in the presence of the counter-selection, fluoroacetamide, remains unexplained. It was proposed that the intact *ADH2* promoter sequence may be interfering with heterologous expression of acetamidase under the counter-selection conditions; *ADH2* expression is affected by trans-acting regulatory genes such as the transcription factor *ADR1* (van den Berg *et al.*, 1998) which appear to have no effect on *ADH1* expression (Denis *et al.*, 1983). The medium used for the counter-selection of the *ADH2*-specific *amdSYM* cassette was adapted (Table 10) to reduce possible catabolite repression associated with the promoter of *ADH2* and low expression of *amdSYM*.

Table 10. Adaptations of the counter-selection medium described by Solis-Escalante *et al.* (2013) required to identify *ADH2* deletion and associated *amdSYM* marker excised strains.

Media component	Original counter-selection medium	Adapted counter-selection medium
Glucose	20 g l^{-1}	2.0 g l^{-1}
Ethanol	Not present	1.0 %
Fluoroacetamide	2.3 g l^{-1}	4.6 g l^{-1}

The original and adapted *amdSYM* counter-selection media were comparatively assessed.

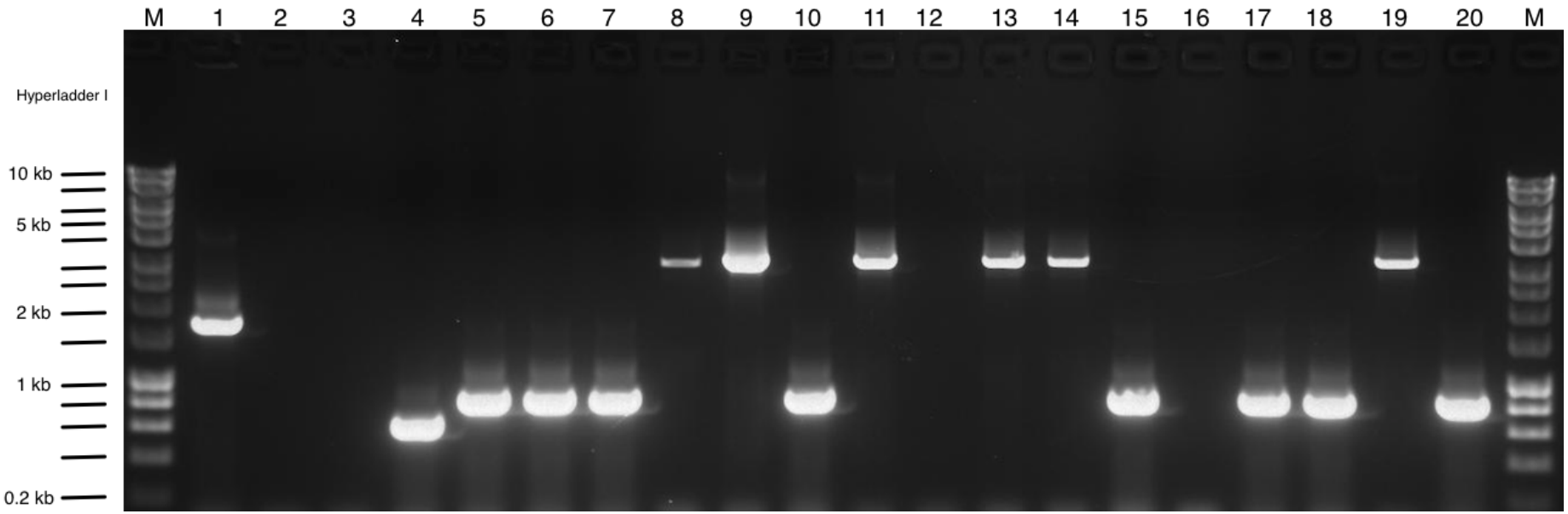


Figure 3.10 Diagnostic colony PCR evaluation of a CEN.PK113-7D *Δadh2* genotype and subsequent marker excision using the adapted *amdSYM* counter-selection medium. All PCRs were performed at the standardised annealing temperature of 63.3 °C. Genes, genotype variants and expected fragment sizes are: Lane 1, *ADH2* (1,675 bp); Lane 2, Negative control; Lanes 3-20, 18 independent colonies assessed after fluoroacetamide counter-selection (*Δadh2*::*amdSYM* (3,130 bp); *Δadh2* (628 bp)).

None of the 20 discrete colonies screened for *ADH2* marker excision, using the original counter-selection medium, were as desired (data not shown). The adapted *amdSYM* counter selection medium (Table 10) provided colonies that had excised the *amdSYM* marker (Figure 3.10). However, the banding pattern of the excised colonies unexpectedly showed two sizes for the excision of the selectable marker, the correct size of 628 bp (lane 4) and a higher frequency of incorrect excision sizes of approximately 750 bp (lanes 5, 6, 7, 10, 15, 17, 18 and 20). This size discrepancy was also seen for other ADH gene deletions. Frequencies of the correct excision sizes were low (approximately one in seven) and therefore the number of colonies that needed to be screened for correct marker removal was increased. The size of the incorrect excisions remained similar for all of the ADH gene deletions, an increase of approximately 150 bp. The incorrectly excised ADH deletion genotypes were sequence analysed using specific sequencing primers (section 2.2.7, Table 1) in an attempt to understand the cause. Figure 3.11 shows an example of a sequence obtained, with highlight colours as for Figure 3.5. From this sequence it can be seen that both direct repeats (orange) are still present, therefore homologous recombination has not occurred between these two sequences. Instead, a 161 bp region remains including 55 bp of the ADH gene (including the start codon) and 106 bp of *amdSYM* deletion cassette (Figure 3.11 – light grey and green highlighted sequences). Examination of the sequence showed that excision had occurred at a *loxP* site within the ADH gene deletion cassette (Figure 3.11 - red sequence).

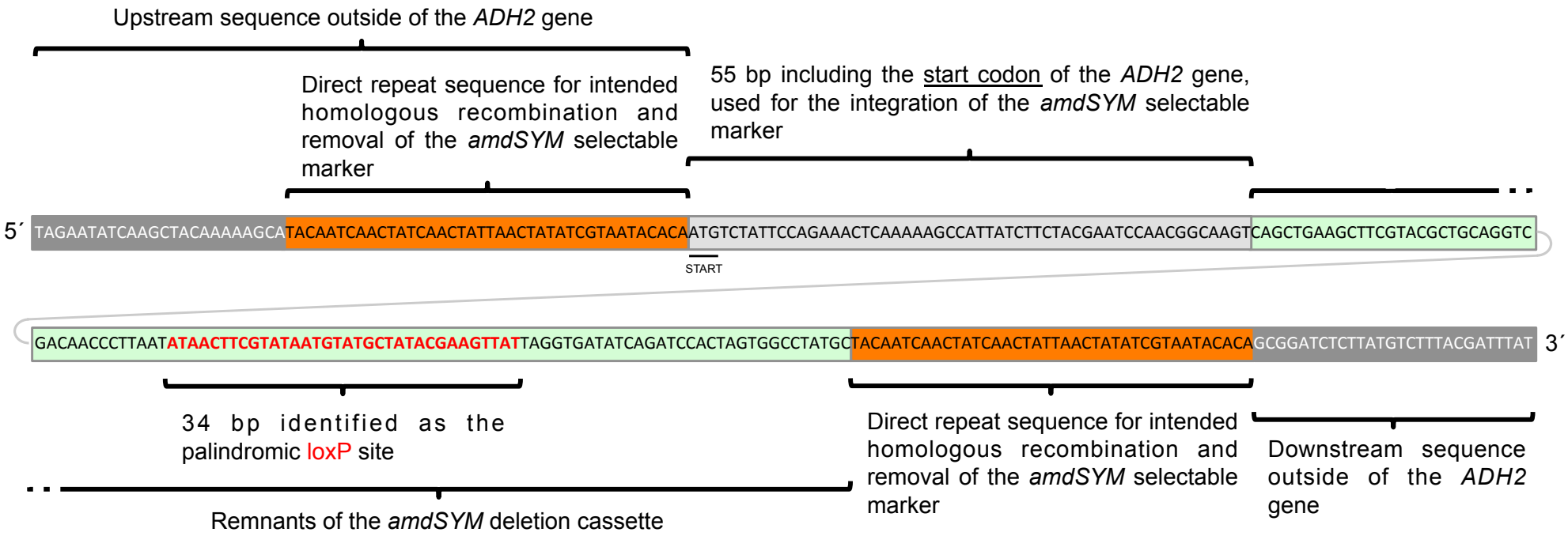


Figure 3.11 Annotated sequence of an incorrect excision of the *Δadh2*::*amdSYM* selectable marker. The features described were present in all incorrect *Δadh2*::*amdSYM* excisions, and for gene deletions of the other five ADH genes. Highlight colours match those in Figure 3.5.

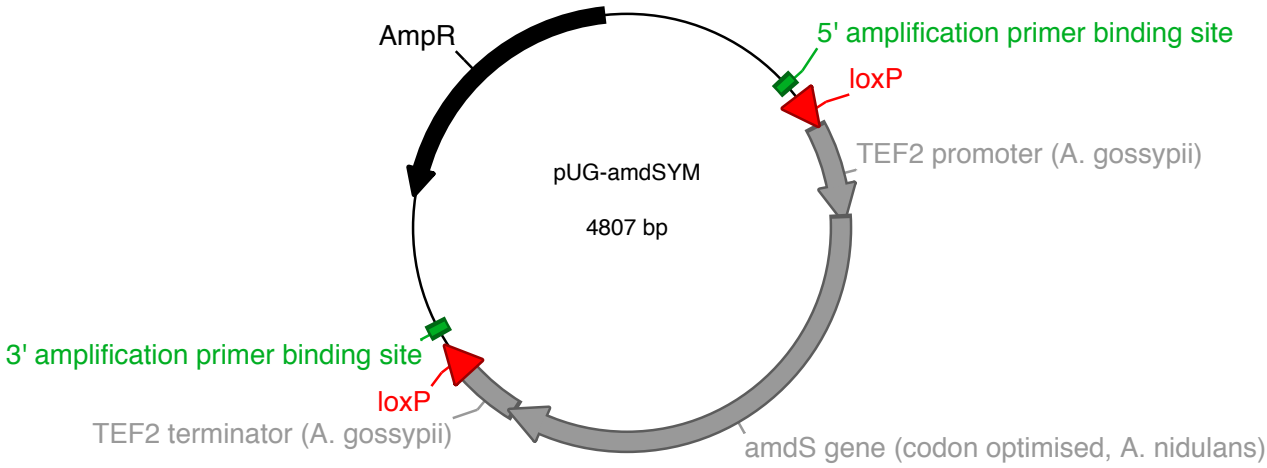


Figure 3.12 pUG-*amdSYM* plasmid map. The primer binding sites used to amplify the linear *amdSYM* deletion cassette and attach the specific homologous and direct repeat sequences for gene deletion are shown (adapted from Solis-Escalante *et al.*, 2013).

The backbone sequence of the *amdSYM* deletion cassette is obtained by PCR amplification of the pUG-*amdSYM* template (Solis-Escalante *et al.*, 2013; Figure 3.8). Initially it was thought that the presence of the two *loxP* sequences (Figure 3.12, red) would not be problematic for the purpose of marker recycling, due to their short size (34 bp) compared with the 40 bp of direct repeat sequence intended for homologous recombination and marker excision. However, comparison of the two *loxP* regions within the pUG-*amdSYM* plasmid revealed the 34 bp *loxP* site and a previously unknown historical cloning site adjacent to it that increase the length of sequence homology to 54 bp. These comparatively long homologous sequences (compared with the intended direct repeat sequences) explain the low frequency of correct excisions observed in Figure

3.10. Figure 3.13 shows the mechanistic differences leading to either a correctly excised selectable marker (A) or an excision scar including a single *loxP* site and remnants of the specific ADH deletion cassette (B).

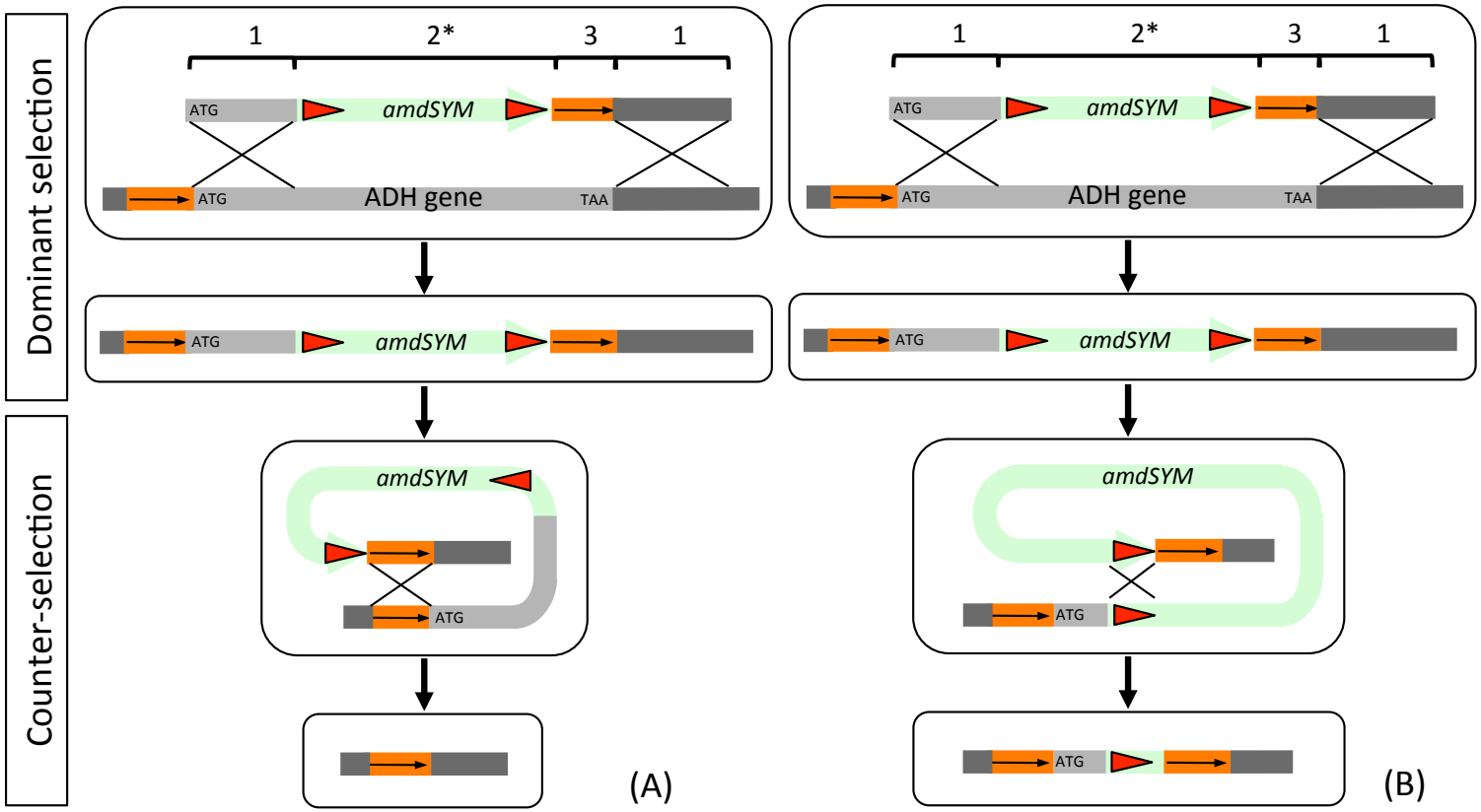


Figure 3.13 Comparison of *amdSYM* methods of gene deletion and marker recycling leading to: (A) the correct scarless excision of the selectable marker, (B) an excision scar including a single loxP site and remnants of the specific ADH deletion cassette. The deletion cassette includes 1) 50-55 bp gene-specific sequences for homologous recombination; 2*) *amdSYM* marker module including extended *loxP* sites (red triangles, 54 bp); and 3) 40 bp direct repeat sequence intended for scarless marker recovery.

3.2.7 Generation of ADH gene deletion library

The *amdSYM* method workflow was modified to screen a larger number of counter-selected colonies for correct, scarless excisions. The *amdSYM* method of gene deletion was used to successfully generate all 64 genotypes required for the evaluation of all permutations of the ADH genes in the industrially relevant prototrophic *S. cerevisiae* strain CEN.PK113-7D (examples in Figure 3.14). In all cases ADH ORF deletions were precise and scarless, with the exception of a subset of deletion strains in which either a single *ADH1* or *SFA1* gene contained an incorrect *loxP* excision; this would not result in genomic instability, as only one *loxP* site is present in the genome. All excision scar sequences were evaluated using the protein family database, Pfam (EMBL-EBI, pfam.xfam.org): the very short predicted peptide sequences from the remnant start codon to the next stop codon (33-48 amino acid residues) were analysed and found to encode no functional domains.

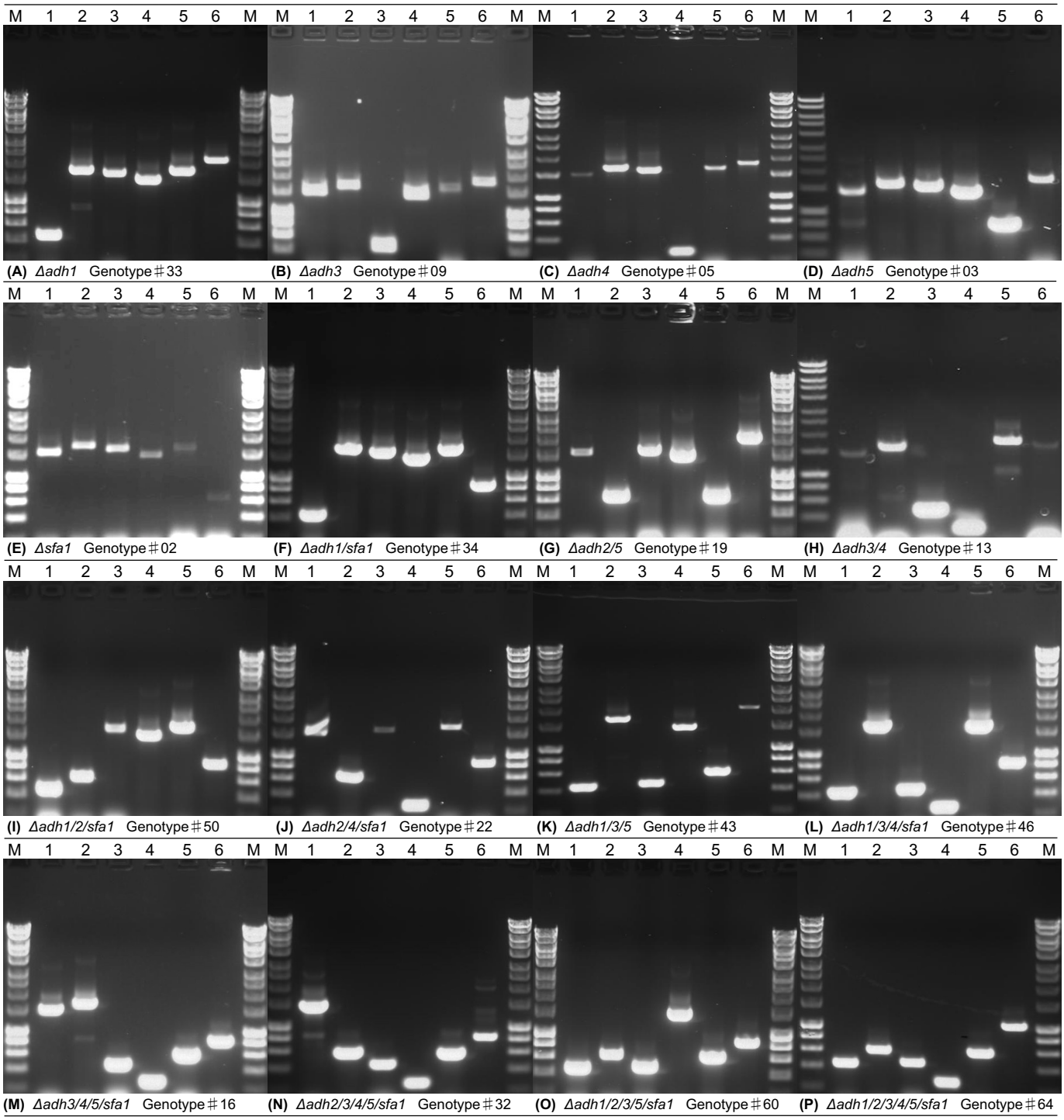


Figure 3.14 Diagnostic colony PCR verification examples of ADH knockout genotypes. All PCRs were performed at the standardised annealing temperature of 63.3 °C. Diagnostic fragment sizes are: Lane 1, *ADH1* (1,509 bp), *Δadh1* (462 bp); Lane 2, *ADH2* (1,675 bp), *Δadh2* (628 bp); Lane 3, *ADH3* (1,609 bp), *Δadh3* (482 bp); Lane 4, *ADH4* (1,426 bp), *Δadh4* (277 bp); Lane 5, *ADH5* (1,678 bp), *Δadh5* (672 bp); Lane 6, *SFA1* (1,997 bp), *Δsfa1* (836 bp). The genotype number corresponds to the orthogonal array list of required genotypes, Figure 3.4.

The successful generation of the ADH gene deletion library (including all 64 permutations of possible ADH gene combinations) can be used to evaluate the effect of genotypes and environmental conditions on the physiology of an industrially relevant strain of *S. cerevisiae*.

3.3 Conclusion

The prototrophic CEN.PK113-7D strain of *S. cerevisiae* was selected for development of the ADH gene deletion toolbox as it offers a good compromise between genetic accessibility and physiological properties (van Dijken *et al.*, 2000; Pronk, 2002).

The dominant counter-selectable *amdSYM* method of gene deletion (Solis-Escalante *et al.*, 2013) is a useful metabolic engineering tool for multiple gene deletions within a prototrophic *S. cerevisiae* strain. The method workflow was optimised, including standardisation of genotype screening, providing all 64 deletion permutations of ADH gene combinations. Further optimisation to increase the efficiency of screening correct excisions could be achieved by using the Q5® site-directed mutagenesis protocol to delete extended *loxP* sequences in the existing pADH_Deletion_Cassette plasmids or, indeed, re-synthesising these plasmids to remove the *loxP* sequences.

The ADH gene deletion library can be used to comparatively study functional substitutions, genotype-by-environment interactions, and adaptations in metabolism (carbon flux) due to perturbations of the ADH genes in *S. cerevisiae,* with the aims to further understanding of the ADH genes (either alone or in combination) and the use of *S. cerevisiae* as an industrial chassis for the production of chemicals.

4 Evaluation of the ADH knockout library

4.1 Introduction

A comprehensive review of the ADH isozymes in *S. cerevisiae* by de Smidt *et al.* (2008) concluded that further investigations (assessing the *in vivo* roles of the ADH enzymes) should be conducted not only under standard growth conditions or conditions of carbon repression, but also under other nutrient conditions such as zinc limitation (shown to regulate the expression of the *ADH4* gene (de Smidt *et al.*, 2012; Bird *et al.*, 2006)). Current understanding of the individual ADH enzymes, as described in Section 3.1, details several environmental factors that differentially impact ADH gene expression and enzyme activity. The full range of these genotype-by-genotype and genotype-by-environment interactions however has yet to be fully explored. The prototrophic *S. cerevisiae* ADH gene knockout library built in chapter 3 was developed to include all permutations of ADH gene knockouts and permits a complete phenotypic evaluation of associated ethanol metabolism. This chapter describes and discusses the statistical methods employed and the analyses of results in the study of yeast ethanol metabolism using the *S. cerevisiae* ADH gene knockout library.

4.1.1 Empirical evaluation of the ADH genes in *S. cerevisiae*

Experimental studies to evaluate the *in vivo* roles of the ADH genes in *S. cerevisiae* have used different *S. cerevisiae* strains, making iterative and

comparative assessments difficult (Drewke *et al.*, 1990; Bakker *et al.*, 2000; de Smidt *et al.*, 2012; Ida *et al.*, 2012). There is an appreciation of the need to standardise laboratory evaluations in an industrially tractable *S. cerevisiae* strain. The strain CEN.PK113-7D (used in this research) is becoming a popular choice in both the academic (van Dijken *et al.*, 2000; Nijkamp *et al.*, 2012) and industrial (Ahn *et al.*, 2016) research communities. Additionally, current understanding of the individual ADH genes (and how they can functionally substitute for one another) fails to fully evaluate the interactions of the ADH isozymes under different environmental conditions. This is a requirement for understanding how environmental niches and industrial process parameters affect the dynamics of the ADH isozymes within this industrially relevant model biological system. This can be attributed in part due to the number of combinations of factors involved and associated effort required to experimentally explore the system: for example, the evaluation of all combinations of ADH genes (see orthogonal array, Figure 3.4) using standardised conditions requires 2^6 (or 64) experiments. The number of experiments and requirement for resources then increases significantly when assessing the environmental impacts on the genotypic combinations. For example, four different environmental parameters, at two different levels of each results in over one thousand experiments:

$$\underset{\text{(genotype)}}{2^6} \times \underset{\text{(environment)}}{2^4} = 1024 \text{ experiments}$$

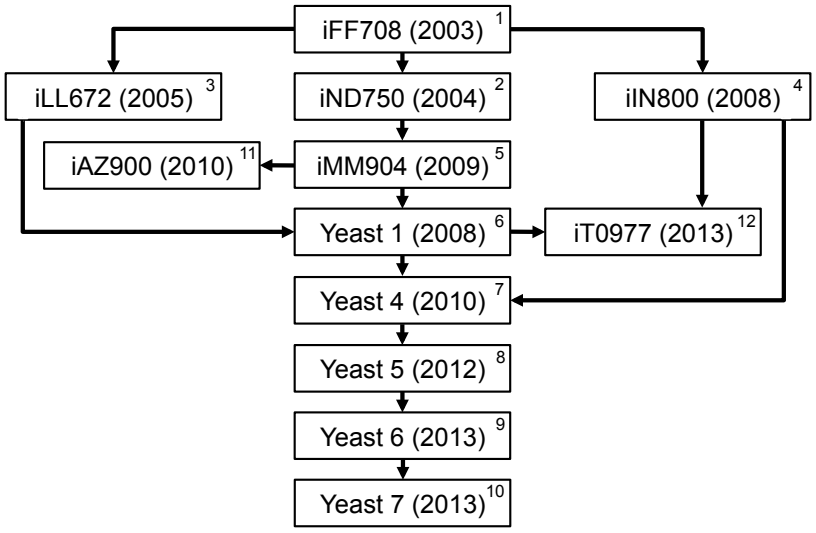
The impact on resources is exacerbated when additional interacting genes, other environmental factors, experimental replication (validation), and potential engineering strategies (production of heterologous chemicals for example) are included. The limitations of current experimentation and requirement to evaluate

the large number of combinations mean a mathematical modelling approach becomes a necessity.

4.1.2 Genome scale models of *S. cerevisiae*

The most widely studied model eukaryotic unicellular organism, the yeast *S. cerevisiae*, is understood to a level of detail that is fundamental to its application in systems biology and important in attempts to develop genome scale models (GEMs) of metabolism (Sanchez and Nielsen, 2015). There are two broad classifications of mathematical models applied in systems biology: “top-down” models which are inductively generated using large data sets, and “bottom-up” models that are deductively generated and built on detailed and localised mechanistic knowledge. The development of global scale modelling will involve building models with the mechanistic quality of the “bottom-up” models at the scale of “top-down” models (Soh *et al.*, 2012). A schematic history of several *S. cerevisiae* GEMs (Figure 4.1A) shows the iterations of models that have identified and addressed important challenges in GEM development including protocols for producing high-quality genome scale models (Thiele and Palsson, 2010), improvement of bench mark evaluation (used to increase accuracy and measure comparison quality for the different GEMs (Sanchez and Nielsen, 2015)), addressing annotation problems and dissimilar nomenclature (Kumar *et al.*, 2012), standardisation and curation of datasets, and detailed referencing of modelling constraints. Failure to implement these standards can result in different predictions arising from the same studies in different laboratories (Soh *et al.*, 2012).

(A)



1 (Forster *et al.*, 2003)
2 (Duarte *et al.*, 2004)
3 (Kuepfer *et al.*, 2005)
4 (Nookaew *et al.*, 2008)
5 (Mo *et al.*, 2009)
6 (Herrgard *et al.*, 2008)
7 (Dobson *et al.*, 2010)
8 (Heavner *et al.*, 2012)
9 (Heavner *et al.*, 2013)
10 (Aung *et al.*, 2013)
11 (Zomorrodi and Maranas, 2010)
12 (Osterlund *et al.*, 2013)

(B)

Genome Scale Model	Genes (reaction modifiers)	"Dubious" ORFs	Cellular compart-ments	Metabolites	Total reactions	Gene-associated reactions	Blocked reactions
iFF708	619	17	3	1,420	1,387	944 (68 %)	325 (23 %)
iND750	750	17	8	1,061	1,266	810 (64 %)	522 (41 %)
iIN800	707	13	4	1,693	1,324	836 (63 %)	316 (24 %)
iMM904	905	18	8	1,228	1,577	1,043 (66 %)	553 (35 %)
Yeast 4	924	4	11	1,319	1,865	1,243 (67 %)	520 (28 %)
iAZ900	901	15	8	1,234	1,597	1,049 (66 %)	393 (25 %)
Yeast 5	918	0	8	1,655	2,110	1,217 (58 %)	800 (38 %)
iTO977	961	20	4	1,218	1,560	1,046 (67 %)	505 (32 %)
Yeast 6	900	0	10	1,458	1,888	1,180 (63 %)	738 (39 %)
Yeast 7	910	0	14	2,220	3,498	2,310 (66 %)	725 (21 %)

Figure 4.1 The *S. cerevisiae* genome scale models: (A) a schematic history (adpated from Sanchez and Nielsen, 2015); (B) summary statistics (adapted from Heavner and Price, 2015) including "dubious" ORFs curated as unlikely to code for a functional protein, and the number of blocked reactions (in grey) for model biomass definition (reactions which cannot carry flux due to network structural constraints). All of the ADH isozymes are included as reaction modifying genes in all models apart from the omission of *ADH1* in models iND750, iMM904, iAZ900 and iTO977.

The yeast GEMs (Figure 4.1) differ in scope, content and terminologies relating to the main levels of information (genes, reactions and metabolites) and therefore an on-going collaborative curation effort is being performed in an attempt to decide on which to base further models. No GEM is suited for all purposes and the current suggestion is that larger models be used for studying particular pathways not present in other models, whereas smaller models may be better for predicting growth phenotypes of knockout strains (Sanchez and Nielsen, 2015). Identification of the best model for a specific performance is a paradigm shift in focus toward realistic representation and evidence-based selection of reactions, rather than creating a complete reconstruction with simulation in mind (Dobson *et al.*, 2010).

There is a limited number of examples where GEMs have been used for developing yeast metabolic engineering strategies, with the research community generally focused on developing new and larger GEMs rather than evaluating their application (Soh *et al.*, 2012). However, application examples of yeast GEMs have achieved satisfactory results; the model iFF708 was used to decrease glycerol production in *S. cerevisiae* (with a strain producing 40 % less glycerol, by engineering redox metabolism (Bro *et al.*, 2006)). Moreover, the constraint-based global scale model iMM904 was used in combination with economic trade-off evaluations of six target products (acetate, ethanol, hydrogen, formate, succinate and D-lactate), highlighting engineering strategies based on environmental and genotypic perturbations in *S. cerevisiae* (Byrne *et al.*, 2012). Validation of engineering strategies was performed using pre-existing experimental data to rank designs by how likely they are to be accurately

reproduced experimentally. The prediction accuracy was poor and limited by the accuracy of the underlying constraint-based models (Byrne *et al.*, 2012). Constraint-based modelling is a widely used approach to study the metabolic capacity of genome scale biochemical networks and Flux Balance Analysis (FBA) is the most popular constraint-based approach to computationally predict the phenotypes under environmental and genetic perturbations in steady state (VanderSluis *et al.*, 2014). The biggest limitation of the approach is the inability to predict responses to changes in enzyme activities. FBA cannot extrapolate a dynamic response, instead providing a snapshot (Soh *et al.*, 2012).

A trade-off in GEMs exists between completeness and prediction accuracy, with models having to be simplified and constrained in order to provide better predictive power. For example: Yeast 5 generalises the many possible triacylglycerol compounds which may be synthesized from fatty acyl moieties of various lengths into a single model species, called “triglyceride” (Heavner *et al.*, 2012). Generalisation is required as the current understanding of metabolism is still developing due to the complexity of biological systems (Sanchez and Nielsen, 2015): in order to develop global scale models of yeast there is a need to generate more understanding using experimentation and integrate data sets into models. There are examples of datasets being developed using prototrophic yeast strains with genetic screens performed in specifically defined media, helping to refine the understanding of yeast metabolism (Heavner and Price, 2015). VanderSluis *et al.* (2014) quantified growth analysis for mutants in all 4,772 non-essential genes from a prototrophic deletion collection across a total of 28 different well-defined metabolic environments. The analysis was

performed using multivariate statistical methods revealing numerous interactions between carbon and nitrogen sources on growth rate, highlighting the requirement to perform growth experiments in a combinatorial fashion. The data set was used to evaluate the Yeast 5 (Heavner *et al.*, 2012) and iMM904 (Mo *et al.*, 2009) GEMs and both demonstrated a modest ability to predict the experimental dataset.

The development of global scale models by the yeast research community has addressed key challenges relevant to both modelling and modern scientific practice including standardisation, experimental reproducibility and data curation. The iterations of current GEMs have provided useful insights and engineering strategies associated with *S. cerevisiae* (Bro *et al.*, 2006; Byrne *et al.*, 2012) but examples of their application are limited due to their nascent stage of development, the requirement to provide accessible tools for visualisation (King *et al.*, 2016), and the necessity for highly skilled scientists to interpret current models and adapt them for a specific purpose (Soh *et al.*, 2012). Additionally, the current models are evaluated using empirical data subject to the same limitations highlighted in section 4.1.1 – for example, models have been validated with laboratory results from auxotrophic strains originally designed to facilitate genetic investigation. It is likely that the reconstruction of portions of the yeast metabolic network (such as nitrogen and sulphur metabolism) is incomplete (Heavner and Price, 2015) and therefore the prediction accuracy of these models can be poor (Byrne *et al.*, 2012).

The current global scale models of *S. cerevisiae* are not suitable for the detailed evaluation of the ADH gene deletion library in combination with environmental

perturbations because of potential differences in evaluation including: i) differences in strain selection and ii) a lack of accuracy of GEMs in predicting dynamic responses to environmental stresses. An empirical approach in combination with data modelling is required in order to evaluate this multifactorial research space.

4.1.3 Design of Experiments

Classically, experimentation has been performed one-factor-at-a-time (OFAT) where all but one factor is fixed. The factor under investigation is varied using scientific intuition or preconceived notions in order to provide understanding of the system being researched. Figure 4.2A shows the limited exploration of the research space using the OFAT experimental approach, which is not an appropriate research method for studying multidimensional spaces. The multifactorial approach (Figure 4.2B) is fundamental to Design of Experiments (DoE), an empirical statistical methodology applicable to both the design and the analysis of experiments. This parallel multifactorial method explores the entirety of the research space without bias, and this is achieved efficiently with fewer experimental runs and resource requirements.

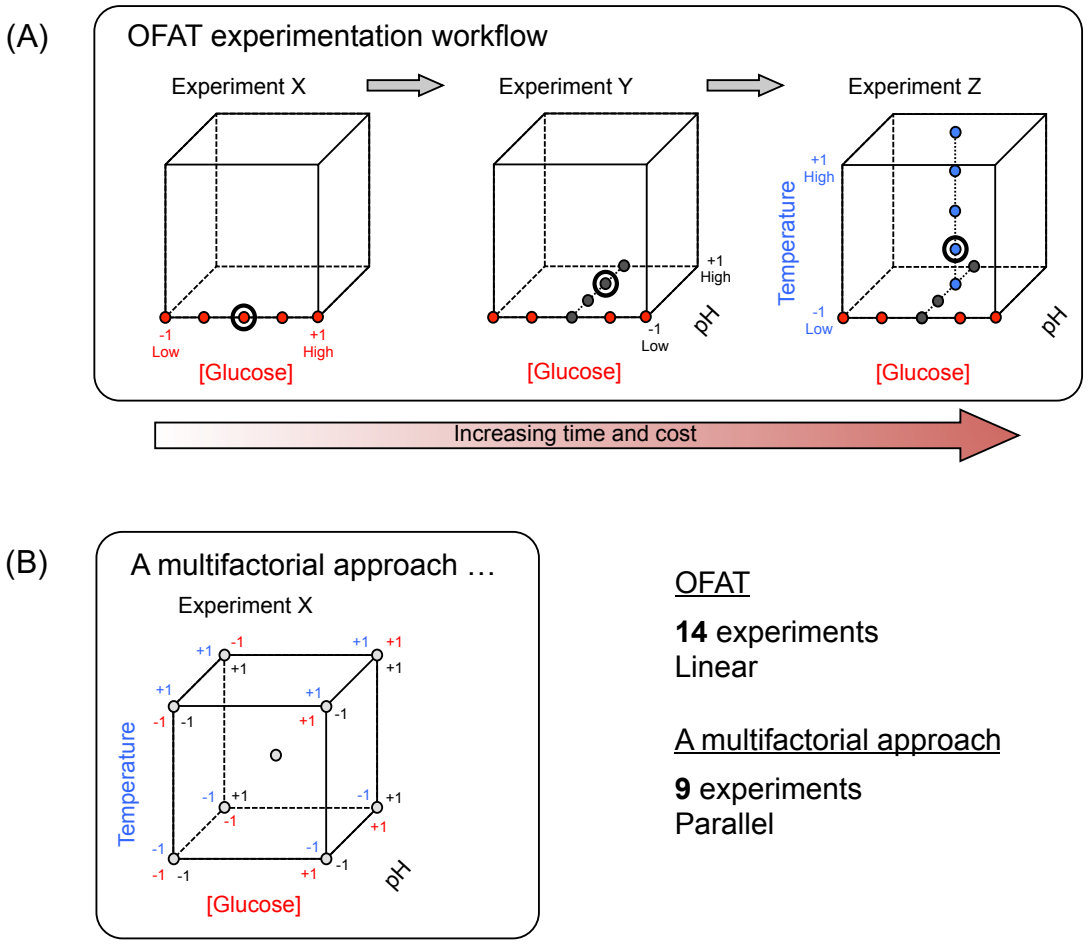


Figure 4.2 Comparing the exploration of multidimensional spaces using the one-factor-at-a-time (OFAT) and a multifactorial approach.

(A) The classical OFAT approach workflow fixes all but one factor in order to explore the research space linearly. Experiment X evaluates different values of [glucose] ranging arbitrarily from a low value (-1) to a high value (+1). The circled data point is the chosen value (potentially an optimum or preconceived choice) for [glucose], maintained for Experiment Y and evaluation of the next factor, pH. The values for pH are then varied (-1 to +1); the circled value is the level of pH maintained for the final evaluation of temperature. Experiment Z has the fixed, experimentally verified, values for [glucose] and pH and the effect of temperature on this system is then evaluated.

(B) The multifactorial approach varies [glucose], pH and temperature in parallel providing coverage of all areas of the research space. More expansive coverage of the research space is achieved with fewer experiments. Experimentation including the centre point, the midpoint of all the factors, allows curvature to be evaluated using data visualisation tools that are used in combination with the multifactorial approach to experimental design.

The requirement to use multivariate statistical methodologies to evaluate complex and often noisy biological and chemical processes is essential (VanderSluis *et al.*, 2014). The application of these methodologies, such as DoE, are most successful when coupled with the prior knowledge of the scientist, capturing the dimensionality of the problem i.e. potential factors involved and the range of their values (Lendrem *et al.*, 2016). The existing literature evaluating the yeast ADH genes (Chapter 3) provides the dimensionality required for investigation of genotype-by-environment interactions – the focus of this thesis.

Figure 4.3 provides an example showing that, without the use of multivariate methods, the OFAT approach in combination with the scientist's preconceived notions regarding the dimensionality of the research space can lead to geographically sensitive solutions and/or a self-fulfilling prophecy of conclusions (Lendrem *et al.*, 2015). In the example given in Figure 4.3, empirical coverage of the research space and the result of highest yield is dependent upon the initial level of factor 2 chosen to remain static during the OFAT experimentation (geographical sensitivity). The two different groups of scientists (**a** and **b**, Figure 4.3B) can provide empirical evidence for differing conclusions, and the initial presupposition regarding the level of factor 2 for both scientific groups is apparently vindicated (self-fulfilling prophecy). The example highlights a potential cause for scientific research being irreproducible or variable in different laboratories (i.e. yield conclusions for groups **a** and **b**). The varying conclusions due to limited evaluation of the research space ultimately corrupt our scientific understanding of the biological or chemical process.

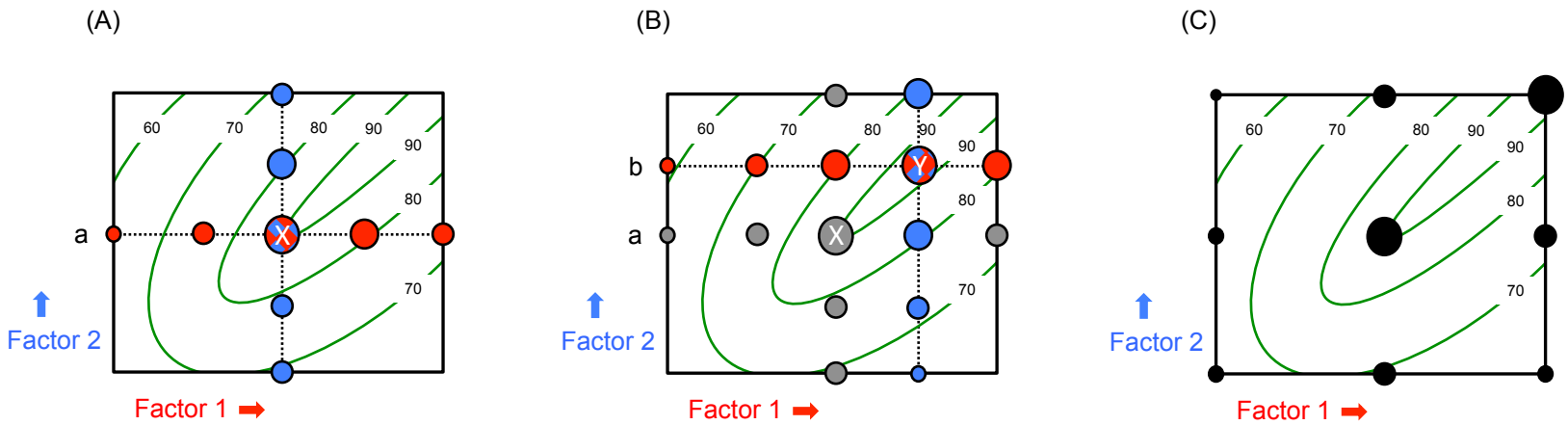


Figure 4.3 Comparing an OFAT evaluation of yield as a function of two factors performed by two different research groups (adapted from Lendrem *et al.*, 2016). As the yield increases the size of the experimental data point (bubble) increases. In (A) research group **a** fixes the value of factor 2 according to a presupposition, and varies factor 1 (●). The level of factor 1 representing the highest experimentally verified yield is then maintained while factor 2 (●) is varied, and group **a** reports the corresponding values of each of the factors providing the highest yield **X**. In (B) research group **b** fixes the value of factor 2 according to a different presupposition, performs the same OFAT evaluation, and reports different values of each of the factors giving the highest yield **Y**. The application of the DoE methodology evaluating the same research space (C), provides a more expansive research of combinations of the two factors and their interactive impact on yield.

The OFAT approach fails to fully evaluate the entirety of the research space (Figures 4.2 and 4.3), and preconceived notions can lead to research bias. The statistical DoE approach for designing experiments can help remove bias from experimentation and provide more scientific insight by screening the entire research space, achieved by varying several factors in parallel 4.2). More expansive design of experimentation is essential when evaluating interactions of factors; at a low glucose concentration, for example, temperature has no effect on a measured response, whereas at high glucose concentrations varying the temperature has a significant impact on the response. The nine experimental design points (Figure 4.3c) determined using the DoE methodology permit the mapping of the underlying contour plot (yield) for the entire research space and would allow factor interactions to be elucidated.

Figure 4.2 shows three factors ([glucose], pH and temperature) involved in a multidimensional research area, represented as a cube. The complexity and multidimensionality of biological and chemical processes makes it difficult to evaluate the impacts of factors and their potential interactions. Research spaces requiring evaluation of more factors and/or combinations of measured responses (for example comparative evaluation of growth rate, productivity and yield) need improved data visualisation tools. The major advantages of the statistical data analysis associated with the DoE methodology is that: i) it allows visualisation of interactions for multiple experimental variables, and ii) it provides predictions of the biological system or process with the application of data modelling (Swalley *et al.*, 2006).

The DoE methodology (a combination of expansive, statistical experimental design, modelling and data visualisation) provides a solution to the empirical evaluation of the ADH genes in combination with environmental perturbations in the *S. cerevisiae* strain CEN.PK113-7D.

4.1.4 Aims

The aims of experiments described in this chapter are to investigate the flux of carbon toward ethanol production in the prototrophic *S. cerevisiae* ADH gene deletion library. A multivariate analysis will investigate genotype and environmental interactions. The application of the statistical Design of Experiments (DoE) methodology for experimental design, modelling and prediction will be appraised.

4.2 Results and discussion

Optimal design of experimentation evaluating the 64 variants of the *S. cerevisiae* ADH gene deletion library and associated data modelling and visualisation utilised a commercially available DoE modelling software package (JMP Pro v. 12, SAS Institute Inc. USA).

4.2.1 Initial screening of the ADH gene deletion library

Evaluation of applying the DoE methodology to assess the 64 variants of the *S. cerevisiae* ADH gene deletion library had the initial aim of elucidating the impact of the ADH genotypes on ethanol production under the same environmental conditions. Standardisation of laboratory media and growth conditions for *S. cerevisiae*, in combination with testing in 96 well plates, provided a high throughput evaluation that included online biomass measurement and offline measurements of glucose, glycerol and ethanol.

The full factorial custom experimental design provided by the DoE methodology defined 68 experiments providing an estimate for all the main effects (each of the ADH genes), as well as 2, 3, 4 and 5 level interactions of the ADH genes' effects on ethanol production in *S. cerevisiae*. The 68 experimental runs were performed in triplicate and were assigned to a randomised position in 96 well plates; the average value for a measured response was used for modelling (see Appendix A for experiment dataset). One experimental constraint was that variants with an intact *ADH1* gene have a quicker specific growth rate

compared to variants with a Δ*adh1* genotype (Zimmermann and Entian, 1997), so these were evaluated in separate plates.

The PLS (Partial Least Squares) modelling platform was used for initial screening as it is a flexible, effective method for modelling experimentation where the experimental variables are correlated (multicollinearity) and noisy (Cox and Gaudard, 2013). The modelling validation was performed using KFold (10) cross validation and the SIMPLS (Statistically Inspired Modification of the PLS method) algorithm applied to the data set. The coefficient estimates of X and Y for models were comparatively evaluated using the root mean PRESS (Predicted REsidual Sum of Squares) statistic. Root mean PRESS provides an estimate of the squared prediction error between an observed validation set value and the model predicted value; a lower root mean PRESS value indicates a lower variance of the model prediction for a given response compared to the validation dataset.

A PLS model fitted using 3 factors to describe the relationship between X and Y matrices had a minimum root mean PRESS of 0.1998 and explained 14.3 % of the cumulative variation in the X score and 99.1 % of the cumulative variation in the Y score. The model's actual by predicted diagnostic plots of the measured responses (Figure 4.4) shows that the data split into the same two distinct groups as required by the constrained experiment run order: *S. cerevisiae* ADH variants with an intact *ADH1* gene and those with a Δ*adh1* genotype. The ADH isozymes have been reported to functionally substitute for one another (de Smidt *et al.*, 2012), however, contrary to this, the data suggest that an intact copy of the *ADH1* isozyme is essential for ethanol production in *S. cerevisiae*

(Figure 4.4). This discrepancy was thought to be a result of limitations associated with experimental design and the format of equipment used to evaluate the ADH gene deletion library. The critical dilution rate (D_{crit}) value (which in a chemostat culture is equal to the culture's specific growth rate i.e. dilution rate (D) = growth rate (μ)) indicates the threshold at which the respiratory capacity of *S. cerevisiae* is exceeded and fermentation occurs: the onset of the Crabtree effect. The fermentative capacity of the CEN.PK113-7D strain of *S. cerevisiae* (used in this study) has previously been evaluated in aerobic chemostat culture and the onset of fermentation reported to be at a D_{crit} = 0.26 - 0.29 (van Hoek *et al.*, 2000; Vemuri *et al.*, 2007). The results of the aerobic 96 well plate format evaluation of the ADH gene variants with a Δ*adh1* genotype (Figure 4.4), showed they had maximum specific growth rates ($0.08 \leq \mu \leq 0.12\ h^{-1}$) lower than the reported corresponding D_{crit} threshold value for the onset of alcoholic fermentation; these strains produced no ethanol as they were metabolising sugar via respiration only. Anaerobic growth conditions are required to assess the Δ*adh1* genotype variants as under this condition only fermentative metabolism can occur. Moreover the difference in the growth rates of the ADH variants make it difficult to utilise the 96 well plate format to standardise sampling times and comparatively analyse cultures. Additionally, *S. cerevisiae* is capable of assimilating produced ethanol and evaporative loss of ethanol from a culture can lead to variability in the measurement of ethanol production.

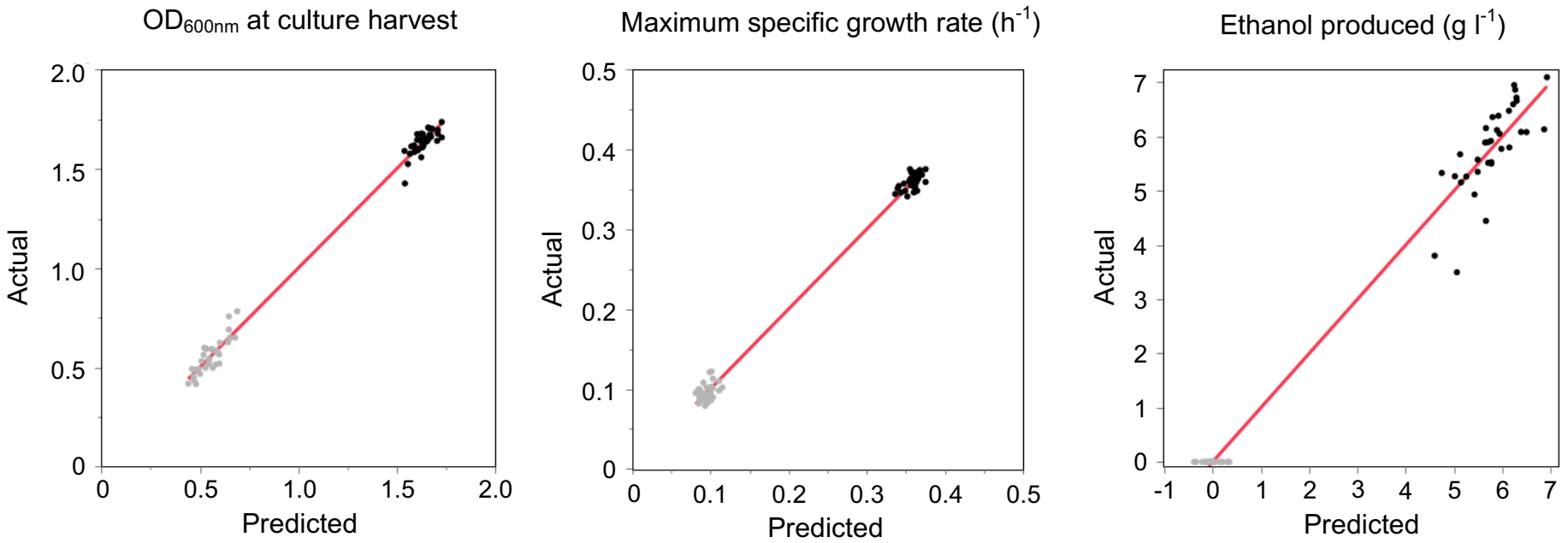


Figure 4.4 The actual by predicted diagnostic plots for the PLS model evaluating the ADH gene deletion library at standardised laboratory growth conditions using the high-throughput 96 well plate equipment format. The measured responses include OD_{600nm} at the time of harvest, maximum specific growth rate and ethanol produced. Data points representing experimental runs evaluating *S. cerevisiae* ADH variants that include an intact *ADH1* gene are shown in black. The intact *ADH1* genotype cultures have a maximum specific growth rate that exceeds the rate required for the yeast to exhibit the Crabtree effect, see the text for details.

The high-throughput 96 well plate equipment format used to evaluate ethanol production in the ADH gene deletion library strain variants is therefore not appropriate; the requirement to provide comparative steady state measurements for the cultures means that culture evaluation using a chemostat is seemingly more appropriate. Chemostat cultures are the method of choice to demonstrate the Crabtree effect in yeast (Postma *et al.*, 1989). Operation of chemostat equipment permits the fixing of the dilution rate of a culture, which under substrate limiting conditions and after equilibration (3 - 5 reactor volume changes) will represent the steady state metabolism of the culture within the environmental conditions provided and at a specific growth rate. Compared to the 96 well plate format, the increased capacity to control and measure bioprocess factors using a chemostat is attractive. However, the associated increased cost of equipment and set up time make the application of chemostats less feasible. To overcome this, a small-scale adaptation of the chemostat, the ministat, was evaluated; this is described in the next section.

4.2.2 Development and testing of the ministat bioprocess equipment

In order to address the limitations of the 96 well plate format to assess the *S. cerevisiae* ADH gene deletion library, the ministat (Miller *et al.*, 2013) was evaluated. A bioprocess system was built that included 32 parallel ministat bioreactors (Figure 4.5). The smaller volumes required increase the throughput of experimentation reducing both the time to achieve steady state measurements and the associated consumable costs as compared to a standard chemostat. The equipment was adapted to include multiple feed vessels: 32 Drechsel bottles contained 250 ml of sterile media and permitted media-associated factors to be changed for each ministat bioreactor. The dilution rate for each reactor was measured in-line using 32 burettes connected to the effluent line of the ministat bioreactors. The gas addition manifold was adapted enabling robust gas flow of either air or nitrogen to each ministat at a flow rate of 200 ml min^{-1} ensuring adequate mixing of the culture broth.

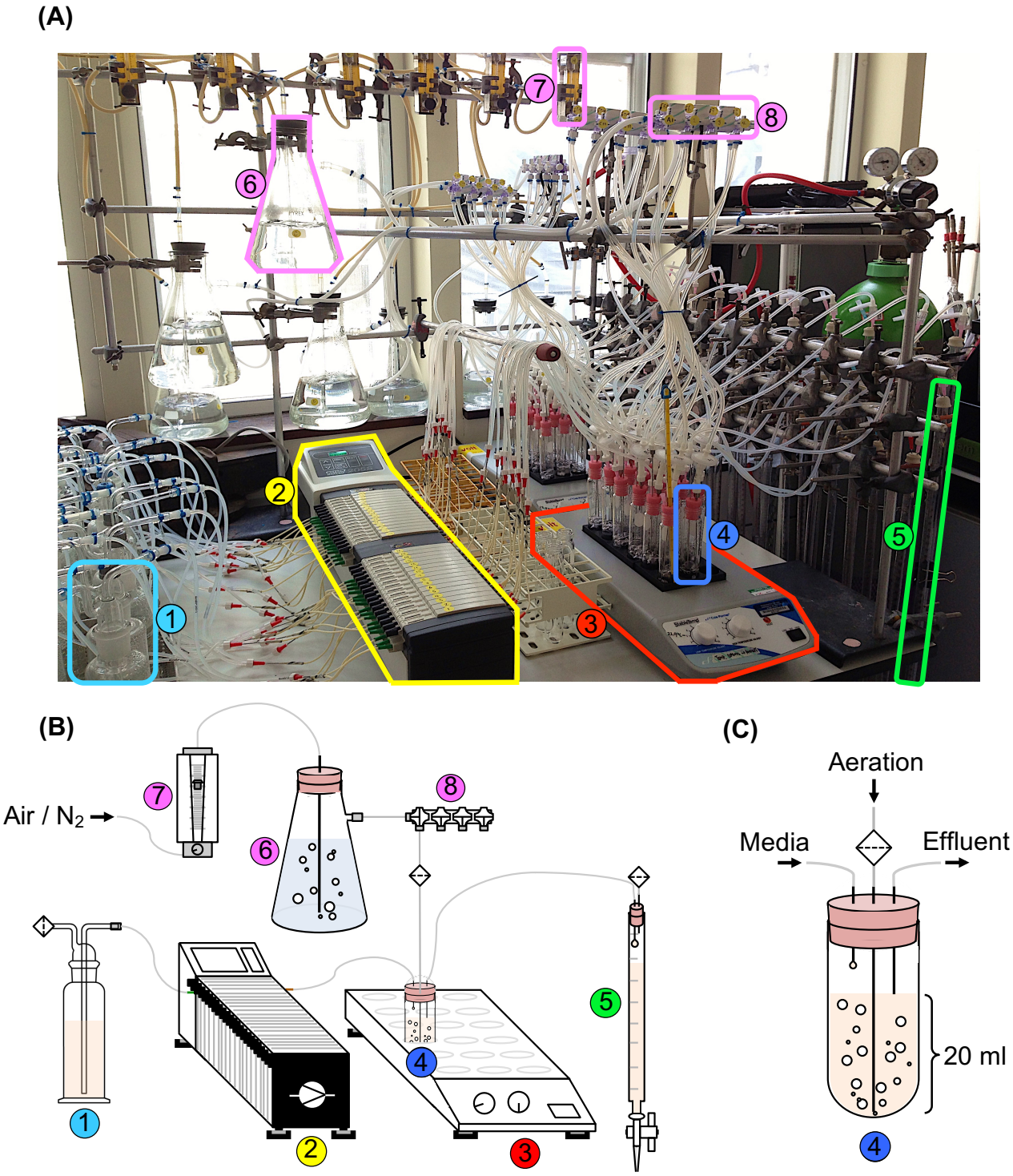


Figure 4.5 The ministat equipment including: **(A)** in situ photograph; **(B)** a detailed schematic of the units of operation; and **(C)** a 20 ml ministat bioreactor. (1) Drechsel bottle for sterile media (×32); (2) a 32 channel peristaltic pump; (3) heatblock (×2); (4) 20 ml working volume ministat bioreactors (×32); (5) burette (×32); (6) gas washing bottle (×8); (7) gas rotameter (×8); (8) gas four-port manifold (×8).

Standard laboratory media were again used for all experiment runs. The lowest maximum specific growth rate (μ) of the previous 96 well plate format experiment was 0.08 h^{-1} and therefore the minstats were operated at the set dilution rate (D) of 1.2 ml h^{-1} = (μ) 0.06 h^{-1} to prevent wash out of cultures, which occurs when the dilution rate exceeds the maximum specific growth of a culture. In addition to the assessment of the variants of the *S. cerevisiae* ADH gene deletion library, the effect of the environmental factor, aeration, on ethanol production was evaluated. When cultured anaerobically, *S. cerevisiae* is unable to synthesise sterols and unsaturated fatty acids (Andreasen and Stier, 1954; Ishtar Snoek and Yde Steensma, 2007), therefore media were supplemented with Tween [420 mg l^{-1}] and ergosterol [10 mg l^{-1}] for anaerobic cultures. The ministat bioreactors were operated as chemostats for 83.3 hours, which permitted 5 reactor volume changes to occur prior to harvesting. The dilution rate of each culture was measured, adjusted and maintained throughout the fermentations.

The custom experimental design using the DoE methodology defined 36 experiment runs providing an estimate for all the main effects (each of the ADH genes and the effect of aeration), as well as two-level interaction effects of the ADH genes and aeration on ethanol production in *S. cerevisiae*. The design was constrained to 32 experiment runs allowing a single batch run of the ministat bioprocess equipment to be performed. This constraint is reflected in the power analysis (the probability of detecting a difference in an effect at a significance level of 0.05) of the design, which is reduced by 10 - 15 % for the main effects if an anticipated coefficient of 1.0 is assumed for each factor. The

measured responses for the experiment included concentrations of glucose, glycerol and ethanol using HPLC as well as cell count using flow cytometry (see Appendix B for experiment dataset).

Table 11 details the experiment runs removed (5 out of 32) from the data set and subsequent analysis, as there was no growth. This included the complete ADH gene deletion strain of *S. cerevisiae* in the presence of nitrogen (experiment run 24).

Table 11. The ADH genotype and aeration of experiment runs removed from the data set and subsequent analysis, as there was no growth. (+) represents intact ADH isozyme genes; (−) represents gene deletions.

Exp. run	ADH genotype						Aeration
	1	2	3	4	5	SFA1	
5	−	−	−	+	+	−	Nitrogen
16	−	+	+	+	−	+	Nitrogen
21	−	+	−	−	+	+	Nitrogen
24	−	−	−	−	−	−	Nitrogen
31	−	−	+	−	+	+	Nitrogen

The PLS (Partial Least Squares) modelling platform including KFold (10) cross validation and the SIMPLS algorithm were applied to the dataset. A PLS model fitted using 4 factors to describe the relationship between X and Y matrices was shown to have a root mean PRESS of 0.8581 and explained 43.2 % of the cumulative variation in the X score and 81.3 % of the cumulative variation in the Y score. Appropriate to the evaluation of the ADH gene deletion library, the model's actual by predicted diagnostic plots of measured responses (Figure 4.6) show that the steady state evaluation using the ministat bioprocess

equipment provides validated models that cover a range for each of the measured responses; the effects of the ADH isozymes and their interactions on ethanol metabolism can now be evaluated.

Additional confidence regarding the application of the ministat bioprocess equipment in evaluating the *S. cerevisiae* ADH gene deletion library can be derived from how the data and corresponding explanatory models are congruent with what has been reported in the literature. The raw data analysis in Figure 4.7 shows *S. cerevisiae* produces more biomass (represented by cell count) cultured under aerobic conditions as compared to anaerobic conditions; this is reported to be due to the low ATP yield from alcoholic fermentation (Vieira *et al.*, 2013).

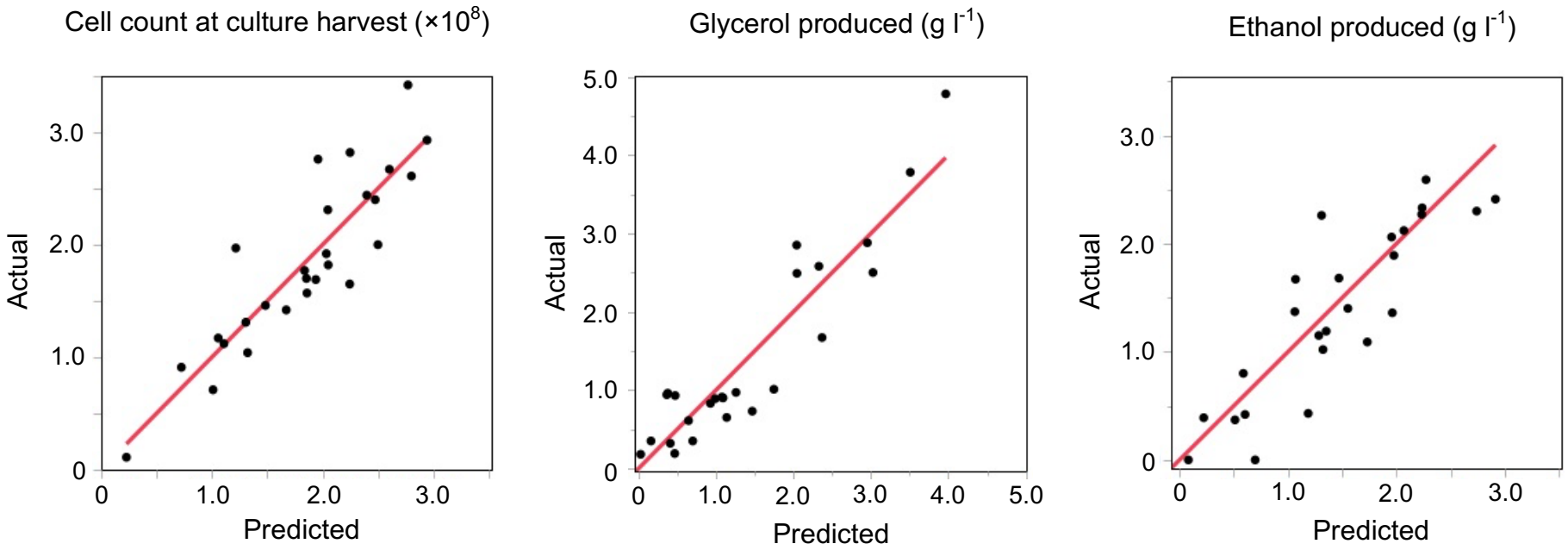


Figure 4.6 The actual by predicted diagnostic plots for the PLS model evaluating the ADH gene deletion library using the ministat bioprocess equipment, additionally the effect of the environmental factor, aeration, on ethanol production was evaluated. The measured responses include cell count at the time of culture harvest, glycerol and ethanol produced. As an enhancement to the evaluation performed using the 96 well plate equipment format, the ministat system permits steady state measurements of cultures, providing applicable models that evaluate the interacting effects of the ADH isozymes.

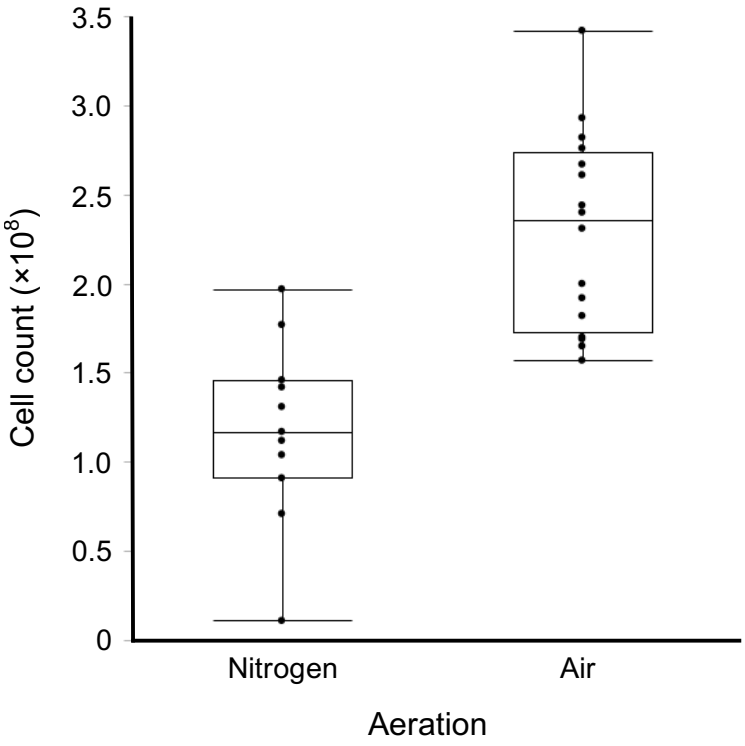


Figure 4.7 The impact on cell count of controlled aeration culture conditions using the ministat bioprocess equipment. The boxes show the interquartile range and median values for the data and the whiskers show the highest and lowest data values. The actual data values are also shown.

The modelled responses include the summary statistic VIP (variable importance for the projection), which is a weighted sum of squares score calculated from the amount of variance attributed to each predictor (factor) defining the model projection (Wold *et al.*, 2001); moreover the model summarises the coefficient for each predictor, which indicates whether the predictor is having an overall positive or negative effect on the measured response (in this instance metabolite production). Table 2 lists the predictors with the greatest VIP scores for the modelled responses of ethanol and glycerol production. The important predictors reflect what has been shown in the literature previously, for example the interaction of deleting both *ADH1* and *ADH2* in *S. cerevisiae* has the most impact on the model projection (VIP = 1.98). The importance of the *ADH1* and *ADH2* isozymes regarding metabolism is inferred by their abundance, together constituting 1.0 % of the total soluble protein from yeast cell preparations (Wills, 1976). The isozymes have been shown to be able to functionally substitute for one another (de Smidt *et al.*, 2012) and their combined deletion reduces ethanol production. Additionally, Table 12 shows the important predictors elicit opposing effects on the production of ethanol and glycerol. This relates to the function of ethanol and glycerol in the maintenance of the cellular redox balance, shifting between the first and second forms of Neuberg's fermentation (Cao *et al.*, 2007; Drewke *et al.*, 1990).

Table 12. The predictors with the greatest VIP (variable importance for the projection) scores for the modelled responses of ethanol and glycerol production in *S. cerevisiae*, derived from the ministat bioprocess equipment evaluation in which all environmental factors were kept constant with the exception of culture aeration (air or nitrogen). The model coefficients for each predictor are not shown. * indicates an interaction of the factors.

Modelled response	Predictors with a **negative** impact on production		Predictors with a **positive** impact on production	
	Factor	VIP score	Factor	VIP score
Ethanol production	Δ*adh1* * Δ*adh2*	1.98	Aeration [Nitrogen]	1.83
	Aeration [Air]	1.83	*ADH1*	1.80
	Δ*adh1*	1.80	*ADH1* * Aeration [Nitrogen]	1.53
Glycerol production	*ADH1*	1.80	Δ*adh1*	1.80
	ADH1 * Aeration [Air]	1.71	Δ*adh1* * Aeration [Air]	1.68
	ADH2 * Aeration [Air]	1.19	Δ*adh1* * *ADH3*	1.47

The ministat bioprocess equipment permits steady state measurements of the ADH gene deletion library, which has been shown to be applicable to its evaluation. The empirical DoE derived models of the measured responses are useful explanatory models that detail factor interactions, which can be difficult to evaluate. Moreover, the model details the interaction between genotypes and an environmental factor (aeration type); the exploration of other environmental interactions with the ADH isozymes and associated impact on ethanol metabolism will now be evaluated.

4.2.3 The ADH gene deletion library: assessment of genotype-by-genotype and genotype-by-environment interactions

In addition to the ADH isozyme variant genotypes and aeration factors evaluated previously using the ministat bioprocess equipment, Table 13 shows the associated levels of other environmental parameters (or factors) that were identified as potentially influencing the production of ethanol in the prototrophic *S. cerevisiae* ADH gene deletion library. The upper level for the dilution rate was set at the critical dilution rate threshold required to elicit the Crabtree effect in aerobic chemostat culture, D_{crit} = 0.29 (van Hoek *et al.*, 2000; Vemuri *et al.*, 2007). The carbon substrate levels were chosen to provide glucose concentrations shown to be either limiting or previously used to evaluate the Crabtree effect in *S. cerevisiae* within chemostat culture (van Hoek *et al.*, 2000). The temperatures represented the laboratory standard growth temperature for yeast culture and a higher value balancing ethanol yield and culture viability (Torija *et al.*, 2003). Evaluation of the interaction of zinc limitation on the ADH isozymes included assessment in zinc-limiting and zinc-replete media (North *et al.*, 2012). The initial pH of the media included the level used for standard growth under laboratory conditions as well as a more acidic condition, corresponding to the pH of industrial consolidated bioprocessing of lignocellulose to bioethanol (Hasunuma and Kondo, 2012).

Table 13. List of the experimental factors, and associated levels of each, for empirical evaluation of the impact on ethanol metabolism of the ADH isozymes and environmental factors.

Factor	Role	Factor level	
ADH1	Categorical	Δ*adh1*	ON (Wild type)
ADH2	Categorical	Δ*adh2*	ON (Wild type)
ADH3	Categorical	Δ*adh3*	ON (Wild type)
ADH4	Categorical	Δ*adh4*	ON (Wild type)
ADH5	Categorical	Δ*adh5*	ON (Wild type)
SFA1	Categorical	Δ*sfa1*	ON (Wild type)
Aeration (200 ml min^{-1})	Categorical	Nitrogen	Air
Dilution rate (D)	Continuous	0.06	0.29
Glucose (g l^{-1})	Continuous	7.5	25
Temperature (°C)	Continuous	30	33
Zinc (μM)	Continuous	1.0	38
Media pH	Continuous	4.5	5.5

Evaluating all permutations of the levels of factors in Table 13 would require 4096 experiments. The custom experiment design using the DoE methodology defined 88 experiments providing an estimate for all the main effects of the factors listed in Table 13, as well as two-level interaction effects of the genotype and environment factors on ethanol production in *S. cerevisiae* (see Appendix C for experiment dataset). The run order of the design was constrained to four whole plots (experiment batches) and eight subplots; this was because of equipment constraints including the single pump used to control dilution rate and two heat blocks for temperature control of the ministats. Potential effects of the reduction in run randomisation was assessed by modelling the effect of the

whole and subplots factors (in addition to those detailed in Table 13); the experiment run order had no impact on the model projection of ethanol production (data not shown).

Table 14 details the experiment runs removed (12 out of 88) from the data set and subsequent analysis, as there was no growth. This included the complete ADH gene deletion strain of *S. cerevisiae* in the presence of nitrogen; the prevalent factors were a Δ*adh1* genotype in combination with nitrogen aeration condition and the upper limit set point of the dilution rate, D = 0.29.

Table 14. The ADH genotype, aeration and dilution rate set points of experiment runs removed from the data set and subsequent analysis, as there was no growth. (+) represents intact ADH isozyme genes; (−) represents gene deletions.

Exp. run	ADH genotype						Aeration	Dilution Rate (D)
	1	2	3	4	5	SFA1		
5	−	−	−	−	+	−	Nitrogen	0.29
7	−	−	+	−	+	+	Nitrogen	0.29
9	−	+	−	+	+	+	Nitrogen	0.29
11	−	+	−	−	+	+	Nitrogen	0.29
18	−	+	+	−	+	−	Nitrogen	0.29
21	−	+	+	−	+	+	Nitrogen	0.29
29	−	−	−	+	−	−	Nitrogen	0.29
49	−	−	−	+	−	+	Nitrogen	0.29
56	−	−	−	−	−	−	Nitrogen	0.29
64	−	+	+	+	−	+	Nitrogen	0.29
71	−	−	+	−	−	+	Nitrogen	0.06
80	−	−	−	+	+	−	Nitrogen	0.06

The PLS modelling platform including KFold (10) cross validation and the SIMPLS algorithm were applied to the dataset. A PLS model fitted using 3 factors to describe the relationship between X and Y matrices had a root mean PRESS of 0.6772 and explained 13.9 % of the cumulative variation in the X score and 95.0 % of the cumulative variation in the Y score. The actual by predicted diagnostic plot of the model for ethanol production shows data were predicted well by the validated model (Figure 4.8); 15 of the 19 experiment runs shown in the blue box were aerobic cultures operated at a dilution rate of 0.06 (which corresponds to a culture growth rate below the D_{crit} threshold value for mixed respiro-fermentative growth). The zero values for ethanol production for these are due to the cultures performing cellular respiration without exhibiting the Crabtree effect.

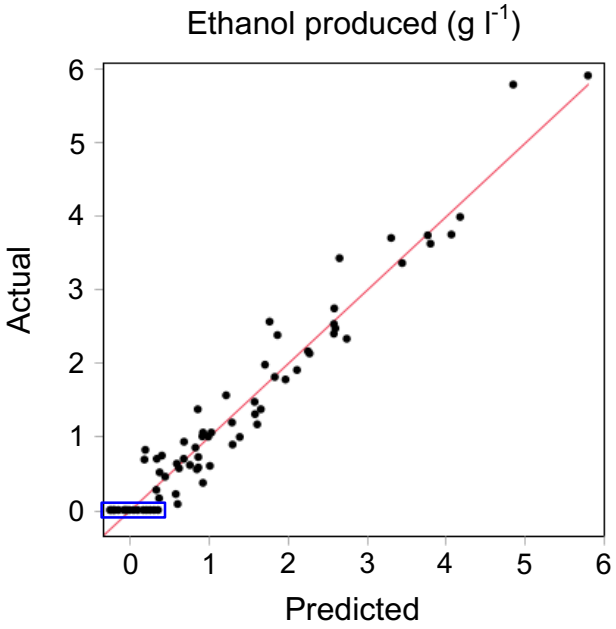
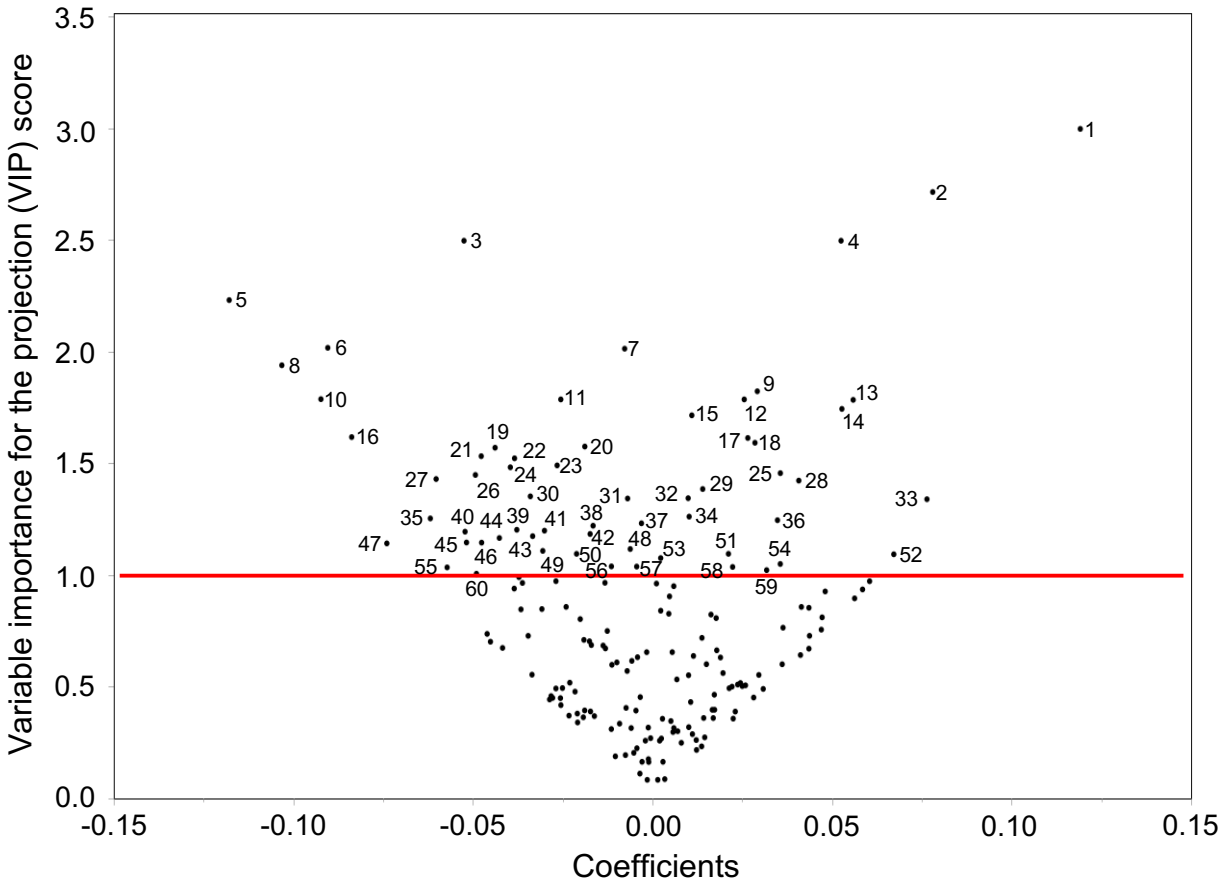


Figure 4.8 The actual by predicted diagnostic plot for the modeled ethanol production response evaluating the ADH gene deletion library in combination with different environmental factors. The blue box highlights 19 experiment runs that produced zero g l^{-1} ethanol.

Figure 4.9 shows the 60 different factor and factor interactions (VIP score > 1.0) as important for the model projection of ethanol production associated with the *S. cerevisiae* ADH gene deletion library in combination with different environmental conditions. This highlights the complexity of regulation and adaptability of the natural system under investigation and shows the necessity to use structured experimental design that permits evaluation using appropriate modelling methodologies.

The interaction of *ADH3* and nitrogen was the most important predictor for the model (Figure 4.9 (**1**); VIP = 2.99) followed by the interaction of *ADH1* and nitrogen (Figure 4.9 (**2**); VIP = 2.71); this was somewhat unexpected as Adh1 has been shown to primarily catalyse the reduction of acetaldehyde to produce ethanol (Drewke *et al.*, 1990; Thomson *et al.*, 2005). However the explanatory model is indeed congruent with what is widely accepted in the literature. Adh3 functions as part of the ethanol acetaldehyde shuttle, which under anaerobic conditions is required to reoxidise mitochondrial NADH (Figure 3; Bakker *et al.*, 2000); the ethanol produced prediction profiler (Figure 4.10) shows this, as well as the removal of this interaction effect when aeration is applied to the system.

Figure 4.9 Model analysis for ethanol production in the ADH gene deletion library using the ministat bioprocess equipment. The variable importance plot of VIP (variable importance for the projection) scores for experimental predictors against the centred and scaled data coefficients is shown. Standardised coefficients indicate if a predictor is having a positive or negative impact, as well as the magnitude of that impact on the measured response. Predictors above the VIP value of 1.0 (red line) are important to the explanatory model for ethanol production (Cox and Gaudard, 2013). * indicates an interaction of the factors.



1	*ADH3* * Aeration [N_2]	21	Δ*adh1* * Δ*adh4*	41	*ADH4* * Aeration [Air]
2	*ADH1* * Aeration [N_2]	22	Glucose	42	Δ*adh1* * Δ*adh5*
3	Aeration [Air]	23	Δ*adh4* * Aeration [Air]	43	Δ*adh1* * *SFA1*
4	Aeration [N_2]	24	Δ*adh2* * Aeration [Air]	44	*SFA1* * Glucose
5	*ADH1* * Glucose	25	Δ*adh5* * Aeration [N_2]	45	Δ*adh5* * Media pH
6	*ADH3* * Glucose	26	Δ*sfa1* * Aeration [Air]	46	*ADH5* *Glucose
7	Δ*adh1* * Aeration [Air]	27	Δ*adh2* * Glucose	47	Zinc * Glucose
8	Aeration [N_2] * Glucose	28	*ADH1* * Δ*adh4*	48	Δ*adh5* * Glucose
9	Δ*adh4* * Aeration [N_2]	29	*ADH2* * Aeration [N_2]	49	Dilution rate
10	*ADH1* * Dilution rate	30	Δ*adh4* * Glucose	50	Δ*adh3*
11	Δ*adh1*	31	*SFA1* * Aeration [Air]	51	*ADH3*
12	*ADH1*	32	*SFA1* * Aeration [N_2]	52	Δ*adh4* * Zinc
13	Δ*sfa1* * Aeration [N_2]	33	Δ*adh1* * Glucose	53	Δ*adh1* * Δ*sfa1*
14	Δ*adh2* * Aeration [N_2]	34	*ADH1* * Δ*adh5*	54	Aeration [Air] * Glucose
15	*ADH1* * *ADH3*	35	Δ*adh5* * Dilution rate	55	Δ*adh3* * Aeration [N_2]
16	*ADH3* * Aeration [Air]	36	*ADH4* * Aeration [N_2]	56	Δ*sfa1* * Glucose
17	Δ*adh3* * Aeration [Air]	37	Δ*adh1* * Δ*adh2*	57	*ADH1* * Δ*sfa1*
18	*ADH5* * Aeration [N_2]	38	*ADH2* * Aeration [Air]	58	*ADH1* * *ADH2*
19	Δ*adh1* * Δ*adh3*	39	Δ*adh5* * Aeration [Air]	59	*ADH1* * *SFA1*
20	*ADH5* * Aeration [Air]	40	Δ*adh4* * Dilution rate	60	*ADH1* * Aeration [Air]

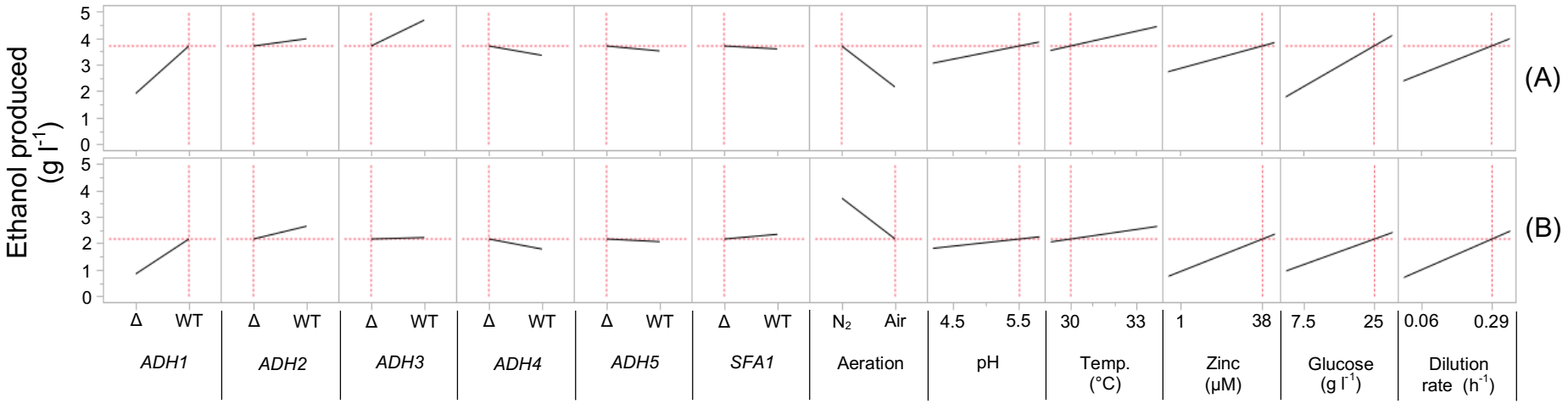


Figure 4.10 The prediction profiler for the model projection of ethanol production in the ADH gene deletion library. The red dashed vertical lines highlight the selected set point for each of the experimental factors; their selection affects the horizontal red dashed line that indicates ethanol production dependent upon the validated model prediction formula. The set points for each of the factors have been maintained apart from aeration set to nitrogen (in A) and air (in B). The data show a marked positive effect of *ADH3* on ethanol production in the presence of nitrogen, but not in the presence of air. This supports the function of Adh3 as part of the ethanol acetaldehyde shuttle, which under anaerobic conditions is required to reoxidise mitochondrial NADH (Bakker *et al.*, 2000).

Moreover, the profiler (Figure 4.10) indicates *ADH1* is the main ADH isozyme involved in ethanol production as it has the biggest gradient / effect of isozyme deletion on ethanol production.

The molecular signals regulating *ADH2* expression have been discussed previously including oxygen induction and the proposed regulation of the gene by catabolite repression (section 3.1.4). Adh2 has been shown to catalyse the reaction producing ethanol from acetaldehyde (de Smidt *et al.*, 2012). The profiler provides empirically derived evidence that this activity does not occur under anaerobic conditions in combination with the higher glucose concentration of the experiment (Figure 4.11 (A)), whereas the presence of the gene increases ethanol production when oxygen is present and at a glucose concentration shown to be limiting in chemostat culture (Figure 4.11 (B)).

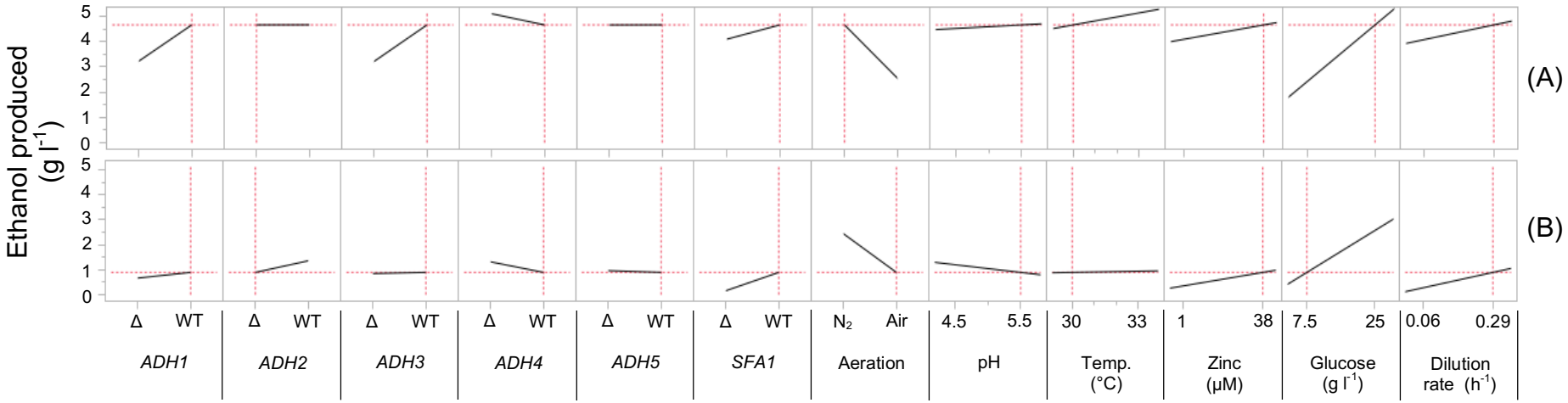


Figure 4.11 The prediction profiler for the model projection of ethanol production in the ADH gene deletion library. The red dashed vertical lines highlight the selected set point for each of the experimental factors; their selection affects the horizontal red dashed line that indicates ethanol production dependent upon the validated model prediction formula. The set points for each of the factors have been maintained apart from aeration and glucose set to nitrogen and 25 g l^{-1} (in A) and air and 7.5 g l^{-1} (in B). The data support the function of Adh2 corresponding to the regulation of the *ADH2* gene, which is oxygen inducible and thought to be catabolite repressed.

The DoE methodology includes the advantage of being able to assess the interactions of factors. Data can be visualised using the interaction profiler providing clear understanding of the combination of factors on a measured response, a capability that is difficult or indeed impossible using a classical OFAT approach. The ADH isozymes exhibit zinc-dependent ADH activity (Larroy *et al.*, 2002; Yuan, 2000), however, only the expression of the Adh4 isozyme has been shown to be upregulated under zinc-limited conditions, alluding to a putative function as an isozyme which produces ethanol during zinc starvation (de Smidt *et al.*, 2012; Bird *et al.*, 2006). Figure 4.12 shows the interaction of the ADH isozymes with zinc concentrations in the media and the corresponding impact on ethanol production, providing empirical evidence that the Adh4 isozyme does indeed provide this function whereas the other ADH isozymes produce more ethanol as media zinc concentrations are increased.

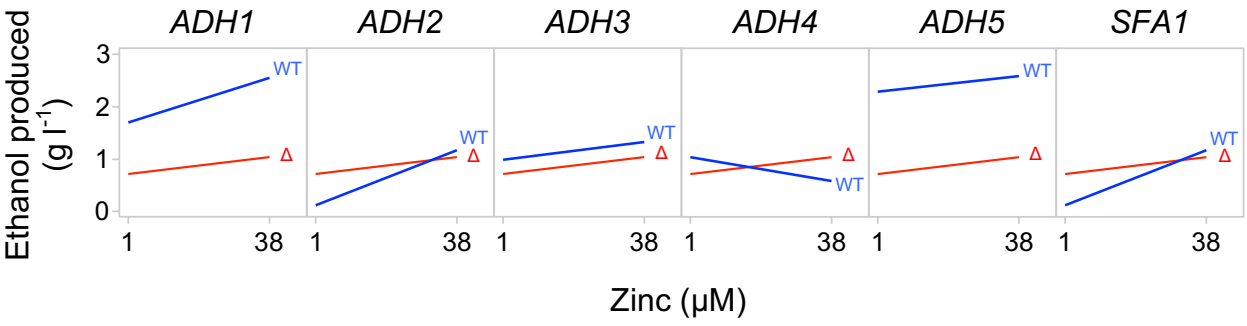


Figure 4.12 Interaction profiler showing the interaction of each of the ADH isozymes at both zinc-limiting and zinc-replete concentrations in the media and the corresponding impact on ethanol production. The blue lines represent the effect of varying zinc concentration on ethanol production when the specific ADH isozyme is present (WT = wild type), the red lines when the specific ADH isozyme is deleted (Δ = gene deletion).

The function of Adh5 remains unclear with its putative role in ethanol production proposed due to the production of ethanol in an Δ*adh1/2/3/4* gene deletion *S. cerevisiae* strain (Drewke *et al.*, 1990). The prediction profiler (Figure 4.13) for the modelled response of ethanol production indicates that the *ADH5* gene does function in the production of ethanol in *S. cerevisiae*, however this function occurs predominantly at pH 4.5 with the activity reduced by 76 % at pH 5.5. This may indicate why its function has yet to be fully elucidated; laboratory evaluation of yeast strains often occurs at the standardised pH of 5.5.

Figure 4.9 indicates that the *SFA1* gene is involved in ethanol metabolism, however the explanatory model does not clarify the dynamics of its function. The interaction of Δ*sfa1* and nitrogen is shown to be an important predictor for the model (Figure 4.9 (**13**); VIP = 1.78) that increases ethanol production. Confounding this result is the interaction of the intact gene, *SFA1* and nitrogen (Figure 4.9 (**32**); VIP = 1.34) also shown, although to a lesser extent, to increase ethanol production. This may be due to further complexity regulating its function either by additional genotypic or environmental factors that were not included in the analysis, or by more complex factor interactions such as 3-level factor interactions. Evaluation of additional level interaction combinations is possible using the DoE methodology and the current model would require augmentation to provide a model estimate for these combinations. The current model does hint towards other interactions possibly impacting the effect of *SFA1* gene on ethanol production. The presence of the *ADH3* gene in combination with *SFA1* produces more ethanol under anaerobic conditions (data not shown). Both these ADH isozymes have been shown to localise in the

mitochondria (Sickmann *et al.*, 2003) and the information together indicates a spatial interaction which may be the reason as to why the isolated Sfa1 protein showed no *in vitro* catalytic activity with either ethanol or acetaldehyde (Wehner *et al.*, 1993).

The current model is congruent with the literature. Additionally it has shown interactions that have furthered the understanding of ethanol production under different environmental conditions in *S. cerevisiae* and highlighted some interesting areas for future research. The research shown evaluates the involvement of the ADH isozymes in metabolism of the substrate glucose, however it does not consider the reassimilation of ethanol or the impact of the ADH isozymes on fusel alcohol production (Dickinson *et al.*, 2003). Augmentation of the current model to include an evaluation using mixtures of glucose, ethanol, amino acids and other carbon substrates is a research area that would require the multifactorial DoE methodology appraised here.

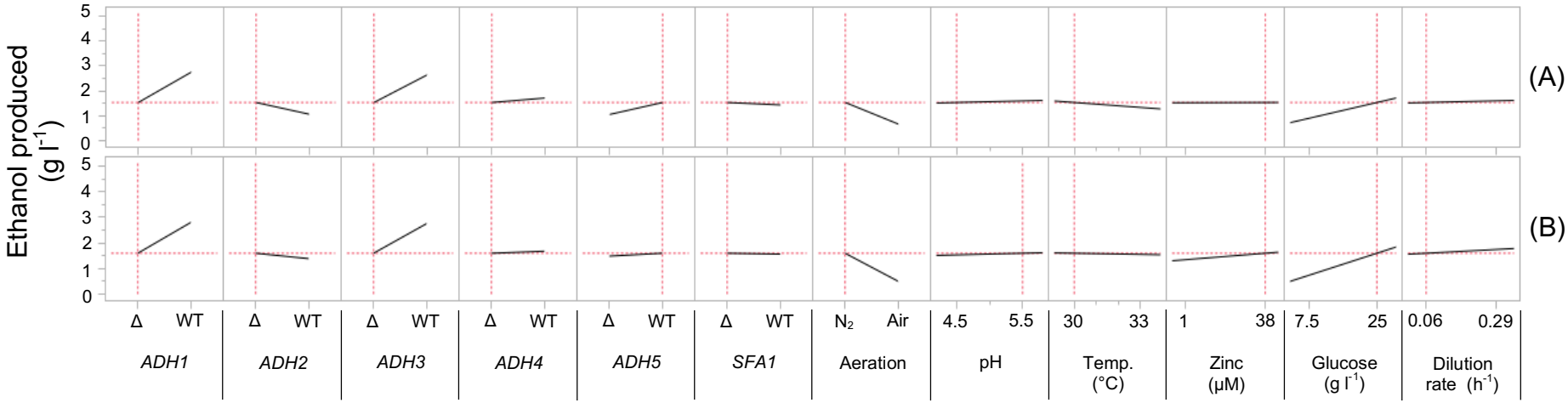


Figure 4.13 The prediction profiler for the model projection of ethanol production in the ADH gene deletion library. The red dashed vertical lines highlight the selected set point for each of the experimental factors; their selection affects the horizontal red dashed line that indicates ethanol production dependent upon the validated model prediction formula. The set points for each of the factors have been maintained apart from initial media pH set 4.5 (in A) and 5.5 (in B). The data support the function of Adh2 corresponding to the regulation of the *ADH2* gene, which is oxygen inducible and thought to be catabolite repressed. The data for Adh5 show it functions in producing ethanol in *S. cerevisiae*, although at a lesser extent at pH 5.5.

4.2.4 Trade off evaluation of the ADH knockout library

A further advantage of the DoE methodology is being able to generate data-derived models that permit a dynamic, unifying evaluation of the impact of factors on multiple measured responses. Data can be visualised providing an evaluation of the system with specific targets in mind; this targeted trade-off analysis may dictate future experimental direction or indeed condition set points relevant to industrial processes of a variety of products made using this industrially relevant cell factory. In addition to the measurement of ethanol production (investigated in the previous section) the following model includes the measurements of free fatty acids (an important pool of energy rich molecules used as the substrate for the development of biofuels, discussed in Section 5.1.2), cell count and glucose consumed (both of which are important metrics for carbon distribution).

A PLS model fitted using 6 factors to describe the relationship between X and Y matrices had a root mean PRESS of 0.9253 and explained 28.9 % of the cumulative variation in the X score and 76.6 % of the cumulative variation in the Y score. The actual by predicted diagnostic plots for the measured response of ethanol production, total free fatty acid production, cell count and consumed glucose shows that data were predicted well by the validated models (Figure 4.14). The modelling software allows desirabilities to be set for each of the measured responses. The experiment factors are then systematically adjusted with the goal to maximise desirability for the performance of the system under evaluation (a trade-off evaluation). A desirability value of 1.0 indicates all

defined desirabilities have been met using the set points of the factors shown in the prediction profiler. Figure 4.15 is the prediction profiler for the model, and shows the outcome of the maximise desirability function set to perform the trade-off evaluation of the system when the desirabilities were set as such that ethanol production was minimised, cell count was undefined and both the total free fatty acids produced and glucose consumed were maximised. The selected factor set points had the maximised desirability value of 0.79, with the model determining the most desirable conditions as *S. cerevisiae* CEN.PK113-7D Δ*adh1/2/3/sfa1* grown aerobically at a slow dilution rate, pH 5.5, 33 °C, with initial 25 g l^{-1} glucose and zinc-limited conditions. The conditions relevant to the specific desirability are in agreement with the previously discussed interactions of ADH isozyme genotypes and environmental conditions; the combination of aeration (air) and Δ*adh1* has been shown to have a big impact in reducing ethanol production (Table 12), with aeration negating the requirement for Adh3 to be present to reoxidise mitochondrial NADH (Figure 4.10; Bakker *et al.*, 2000). The deletion of *ADH2* is important to maximise free fatty acid production under specific environmental conditions that have been shown to reduce the impact of the Crabtree effect: operating the ministat at a slow dilution rate (Sonnleitner and Kappeli, 1986) and using zinc limiting media (Figure 4.12), both of which may not be applicable to the industrial implementation of the strain for the manufacture of products derived from free fatty acids (Flikweert *et al.*, 1997).

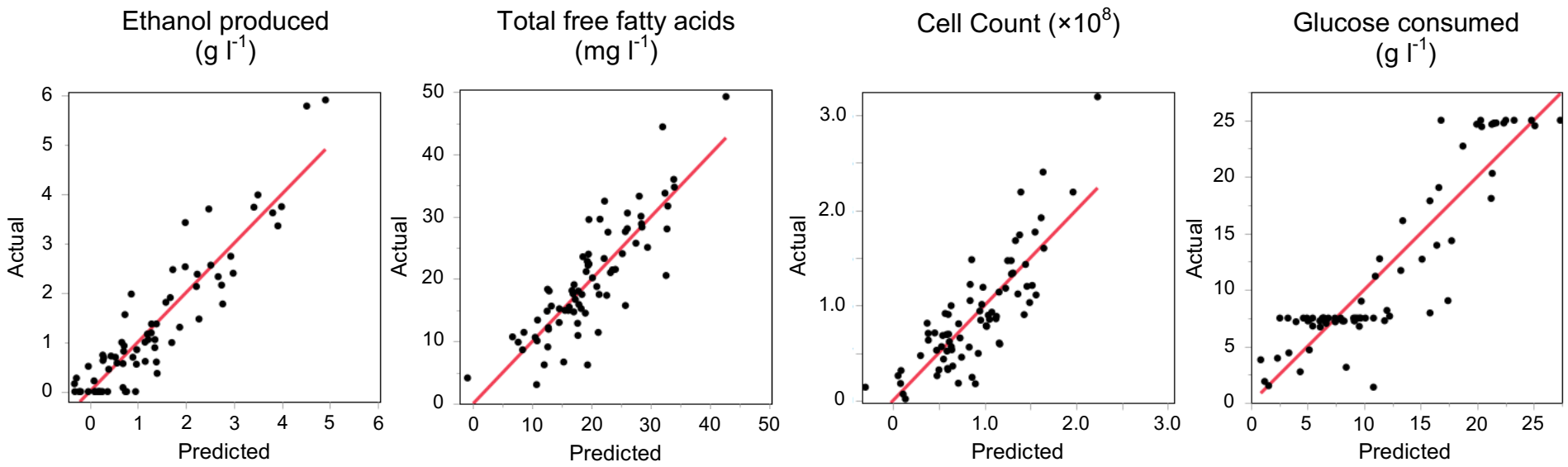


Figure 4.14 The actual by predicted diagnostic plots for the PLS model evaluating the ADH gene deletion library. The measured responses include cell count at the time of culture harvest, glucose consumed, total free fatty acids produced and ethanol produced. In addition to multivariate analysis, the DoE methodology can provide a model that includes multiple measured responses permitting a dynamic trade-off analysis of the system under investigation.

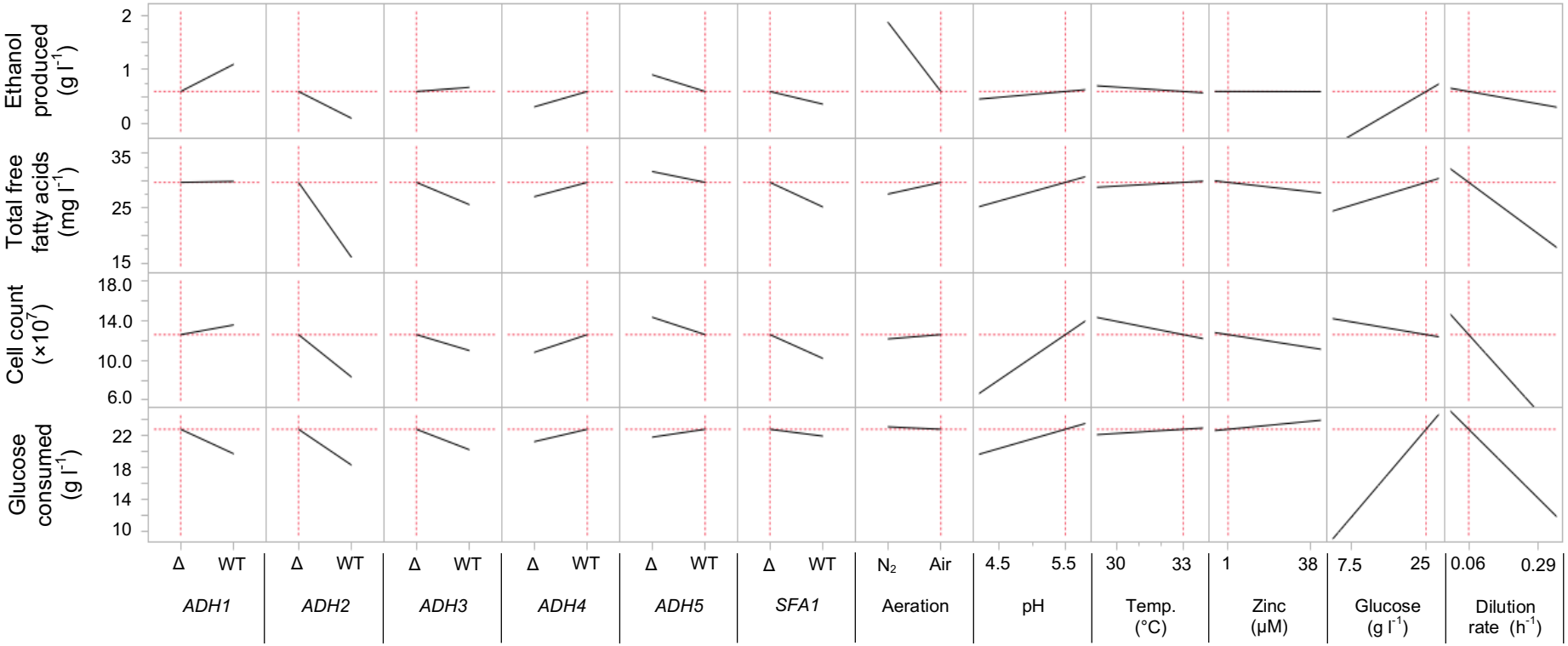


Figure 4.15 The linked prediction profiler for the model projection of ethanol and total free fatty acid production, cell count and glucose consumed. The profiler permits a targeted trade-off analysis of the ADH genes in combination with environmental perturbations in the *S. cerevisiae* strain CEN.PK113-7D. Desirabilities were set such that ethanol production was minimised, cell count was undefined and both the total free fatty acids produced and glucose consumed were maximised. The selected factor set points had the maximised desirability value of 0.79, with the model determining the most desirable conditions as *S. cerevisiae* CEN.PK113-7D Δ*adh1/2/3/sfa1* grown aerobically at a slow dilution rate, pH 5.5, 33 °C, with initial 25 g l^{-1} glucose and zinc-limited conditions.

4.3 Conclusion

The ministat bioprocess equipment in combination with the Design of Experiments (DoE) methodology are essential tools in order to investigate the flux of carbon toward ethanol production in the prototrophic *S. cerevisiae* ADH gene deletion library. Advantageously, the small-scale adaptation of the chemostat, the ministat, ensured experiments were performed in a medium throughput manner and permitted steady state measurements to be made providing a direct comparative evaluation of the ADH gene deletion library.

The PLS (Partial Least Squares) modelling platform including KFold (10) cross validation and the SIMPLS algorithm were applied to a dataset derived from 88 experiment runs (a significant reduction compared to the 4096 runs covering all factor combinations), providing a multivariate analysis that highlighted 60 genotype and environmental factors, including their interactions, as being important predictors (VIP > 1.0) for modelling ethanol production. The model of ethanol production generated is congruent with the literature showing the dynamics of the ADH isozymes with regard to metabolism in *S. cerevisiae*, for example the role of the mitochondrial Adh3 isozyme in the ethanol-acetaldehyde shuttle required to reoxidise mitochondrial NADH under anaerobic conditions resulting in increased production of ethanol.

Data visualisation, particularly that of factor interactions, is a key benefit of using the DoE methodology, furthering our current understanding of ethanol production under different environmental conditions in the well-studied yeast model organism, *S. cerevisiae*. The model prediction profiler provides

empirically derived evidence that the *ADH2* gene is not linked with the production of ethanol under anaerobic conditions in combination with higher glucose concentrations in the culture medium. The *ADH5* gene is linked with ethanol production, contributing to this function predominantly at pH 4.5. Moreover, the interaction profiler provides empirically derived evidence that *ADH4* is associated with increasing ethanol production when the cell is confronted with a zinc-limited environment.

Additionally, data can be visualised in a dynamic way permitting trade-off analyses which, based on a particular set of desirabilities, have the advantage of being able to dictate future experimental direction or indeed condition set points relevant to industrial processes. A multivariate analysis of several measured responses, including free fatty acid production, provides a useful model of the industrially relevant CENPK.113-7D ADH gene deletion library. This library is now amenable to further metabolic engineering, particularly using free fatty acids as a substrate pool, in order to produce chemicals within a biorefinery concept, further developing the industrial use of the yeast *S. cerevisiae*.

The advantages of the DoE methodology, including the ability to dynamically evaluate interactions within a multifactorial research space (both genotypic and environmental factors), make it a method that would enhance the aim of generating a genome scale model of *S. cerevisiae*.

5 Hydrocarbon production in *S. cerevisiae*

5.1 Introduction

The *S. cerevisiae* ADH gene deletion library has been characterised (Chapter 4) using combinations of both genotype and environment to assess carbon flux toward ethanol production. This was performed in the prototrophic CEN.PK113-7D strain that is concurrently being developed for use within industrial processes (Ito *et al.*, 2014; Ahn *et al.*, 2016). However whilst it is possible to relatively rapidly develop a proof of principle strain that produces a product of interest, it is far more challenging to develop a strain that meets commercial targets and fits within a biorefinery concept (Runguphan and Keasling, 2014). The CEN.PK113-7D strain is both genetically amenable (conducive to the relatively rapid development of products as proof of principle) and industrially tractable (van Dijken *et al.*, 2000), permitting an appropriate assessment of developed product pathways based on commercial targets and within a large scale biorefinery concept.

5.1.1 Industrial use of *S. cerevisiae*

The choice of cell factory is critical in the assessment of the industrial production of chemicals and recently there has been a consolidation and focus on a few industrial cell factories which include the yeast *S. cerevisiae* (Hong and Nielsen, 2012). *S. cerevisiae* is generally regarded as safe (GRAS) and has

been used both domestically and industrially for centuries in the leavening of dough and for brewing. Beneficial traits of *S. cerevisiae* compared to other cell factories (Table 15) include high rate of glycolytic flux and reduced biomass yield. Other industrial process advantages which contribute to the capability of *S. cerevisiae* to perform robust large-scale fermentations (and implicate reduced capital and operational expenditure) include osmotolerance, growth at low pH and lack of phage contamination.

Table 15. Comparison of industrial process metrics of several cell factories (adapted from Rumbold *et al.*, 2009).

Cell Factory	Growth rate (h^{-1})	Glucose uptake rate ($g\ g^{-1}\ h^{-1}$)	Biomass yield ($g\ g^{-1}$)
Saccharomyces cerevisiae	0.20	2.52	0.09
Pichia stiptis	0.10	0.34	0.14
Aspergillus niger	0.03	0.20	0.34
Corynebacterium glutamicum	0.33	0.79	0.31
Escherchia coli	0.49	1.56	0.28

The proven ability of *S. cerevisiae* to perform robustly within industrial processes highlights its potential as an industrial cell factory that may replace the current production of chemicals from crude oil in a sustainable way. Developing an industrially relevant yeast platform (i.e. *S. cerevisiae* strain CEN.PK113-7D) in combination with different production pathways would provide biorefineries with a plug-and-play range of products allowing adaptability to market requirements (Tang *et al.*, 2013; Hong and Nielsen, 2012). The range of engineered pathways in *S. cerevisiae* is ever increasing and

examples of the production of biofuels, protein drugs, bulk and fine chemicals are shown in Table 16.

Table 16. Examples of the product range of industrially applicable engineered pathways in the yeast *S. cerevisiae* (adapted from Hong and Nielsen, 2012).

Category	Product	Strain	Reference
Biofuels	Ethanol	CEN.PK102-3A	Guadalupe Medina *et al.* (2010)
	Biobutanol	CEN.PK 2-1C	Chen *et al.* (2011)
Bulk chemicals	Succinic acid	CEN.PK113-7D	Ahn *et al.* (2016)
	Pyruvic acid	CEN.PK113-7D	van Maris *et al.* (2004)
	1,2-propanediol	BY4742	Peralta-Yahya *et al.* (2011)
Fine chemicals	β-amyrin	CEN.PK113-7D	Madsen *et al.* (2011)
	Artemisinic acid	CEN.PK 2	Paddon *et al.* (2013)
	L-ascorbic acid	W303 1B	Sauer *et al.* (2004)
Protein drugs	Single-chain antibodies	BJ5464	Hackel *et al.* (2006)
	Epidermal growth factor	W303	Chigira *et al.* (2008)

The large scale industrial production of cellulosic ethanol using the yeast *S. cerevisiae* ("Project Liberty": a joint venture between POET and DSM in the USA) shows the economic feasibility of using this cell factory for the production of a relatively cheap, high volume commodity and is a success story related to the production of biofuels.

5.1.2 Fatty acid-derived biofuel

An integral part of all living systems, fatty acids are energy rich molecules utilised in both chemical and energy storage functions within the cell, particularly as the energy contained in the C-C bonds can be efficiently released by β-oxidation (Tehlivets *et al.*, 2007). The fatty acid biosynthesis pathway is well characterised in eukaryotes and the molecules produced have suitable chain lengths as precursors for the production of biofuel molecules, typically comprising chain lengths of 12 (C_{12}) to 20 (C_{20}). These can be used to produce biofuel molecules within the diesel range of chain length and energy densities (Rude *et al.*, 2011; Runguphan and Keasling, 2014).

The fatty acids produced in yeast are more defined, consisting mostly of C_{16} and C_{18} chain lengths. Figure 5.1 shows the flux through the native fatty acid pathway of yeast occurring in either the cytosol or mitochondria. Carbon dioxide is required for the carboxylation of acetyl-CoA to malonyl-CoA in an ATP-dependent reaction, however there is no net requirement for CO_2 as it is released during the condensation reaction catalysed by fatty acid synthase (Tehlivets *et al.*, 2007). Yeast fatty acid synthase is composed of two subunits (Fas1 β-subunit, Fas2 α-subunit) forming a hexameric complex that performs seven to eight cycles of fatty acid elongation producing fatty acyl-CoAs for phospholipid and triacylglycerol biosynthesis. Fatty acids serve multiple functions in yeast and are potentially toxic to the cell (Tehlivets *et al.*, 2007), therefore fatty acid biosynthesis (including processing of exogenous fatty acids) is tightly regulated at multiple levels. In order to increase fatty acid-derived

biofuels, these regulatory elements must be mitigated (Runguphan and Keasling, 2014).

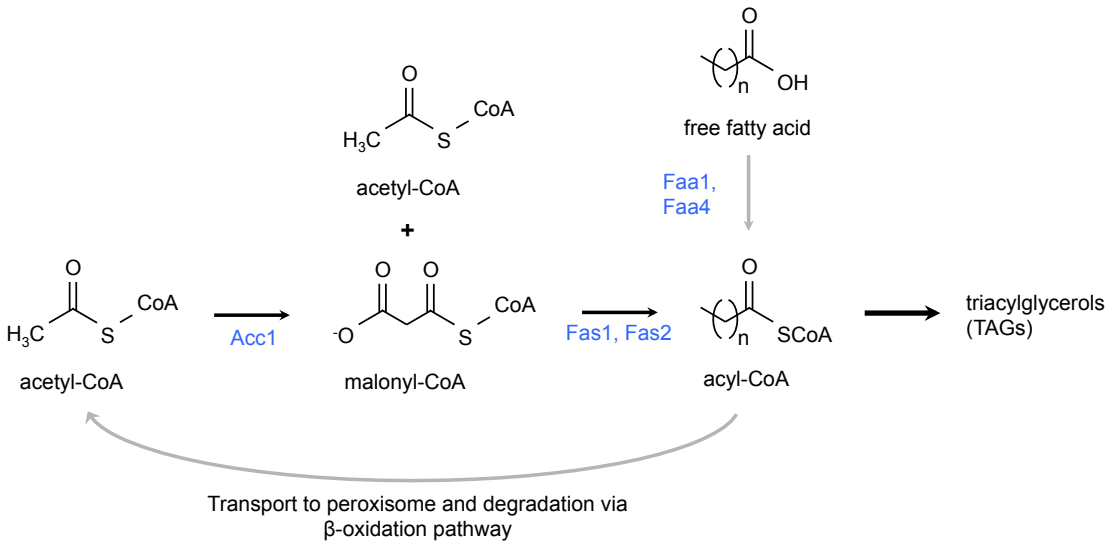


Figure 5.1 Overview of the native fatty acid biosynthesis pathway of *S. cerevisiae.* Enzymes include acetyl-CoA carboxylase (Acc1) and the fatty acid synthase complex composed of two subunits (Fas1 β subunit and Fas2 α subunit). The fatty-acyl-CoA synthetases (Faa1 and Faa4) are responsible for 99 % of activity involved in the breakdown of exogenous free fatty acids (grey arrows) (Trotter, 2001). (Adapted from Runguphan and Keasling, 2014).

The products of fatty acid biosynthesis cannot be directly used as fuel and must be chemically processed prior to utilisation as a biofuel compatible with existing infrastructure. There are, however, natural metabolic pathways that exist that can convert fatty acid intermediates into hydrocarbons (Figure 5.2). Enzymes involved in these pathways have been identified and heterologously expressed, predominantly in *E. coli*, in order to assess the feasibility of this biofuel production strategy (Schirmer *et al.*, 2010; Rude *et al.*, 2011; Harger *et al.*, 2012; Akhtar *et al.*, 2013; Howard *et al.*, 2013). These studies have demonstrated the real potential of genetically manipulating metabolism in order

to produce precise chemical replacements to fossil fuels (alkanes and alkenes). The challenge remains to demonstrate that this potential biofuel strategy is economically and sustainably feasible, and assessment of these enzymes within the industrially relevant host *S. cerevisiae* will allow a more appropriate evaluation of suitability.

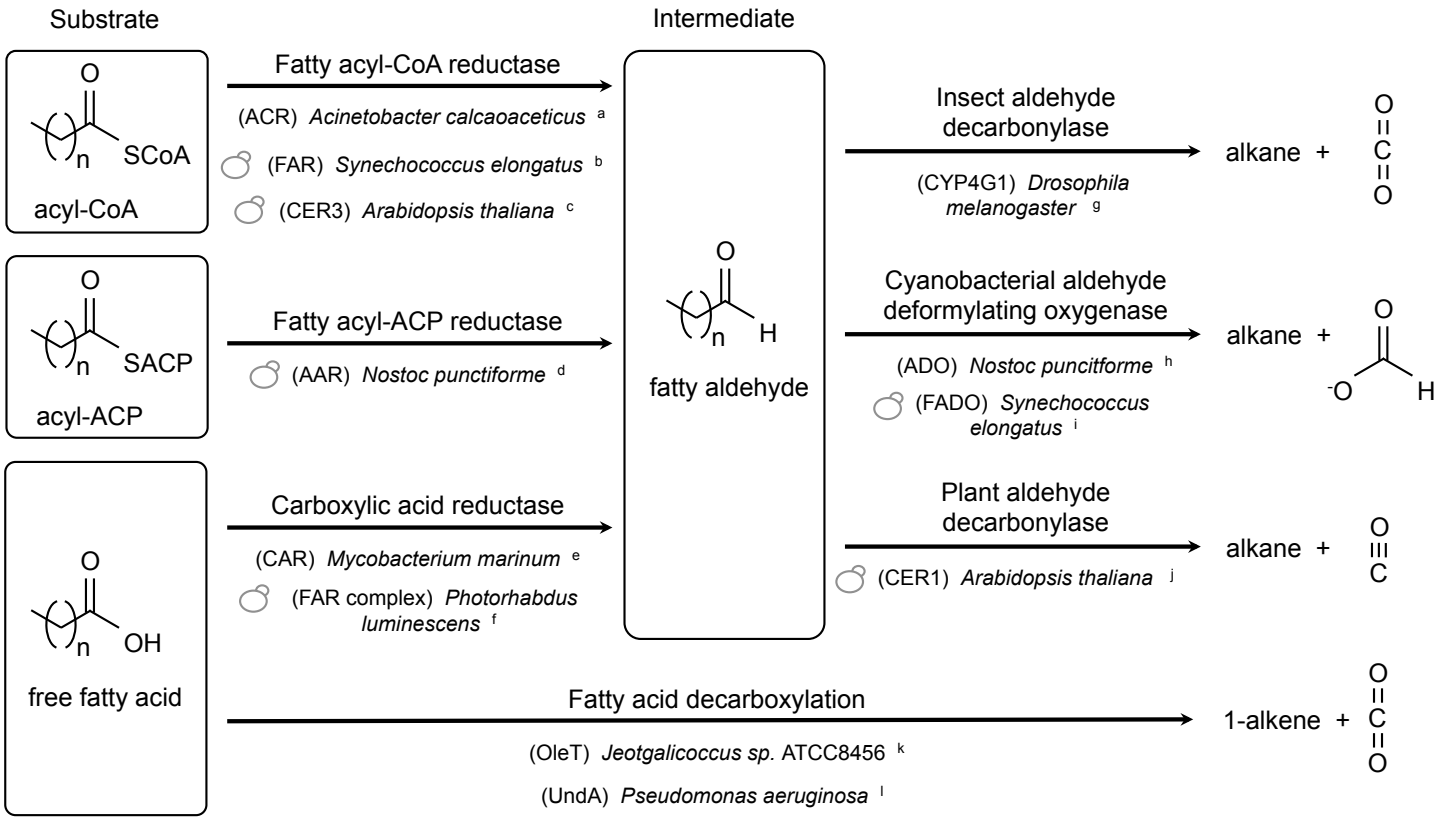


Figure 5.2 Examples of enzymes involved in the biosynthesis of alkanes and alkenes derived from fatty acid precursors. Enzymes that have been heterologously expressed in *S. cerevisiae* are represented by (adapted from Marsh and Waugh, 2013). [a] (Reiser and Somerville, 1997); [b] (Buijs *et al.*, 2015); [c] (Bernard *et al.*, 2012); [d] (Schirmer *et al.*, 2010; Howard *et al.*, 2013); [e] (Akhtar *et al.*, 2013); [f] (Meighen, 1991; Howard *et al.*, 2013); [g] (Qiu *et al.*, 2012); [h] (Schirmer *et al.*, 2010; Das *et al.*, 2011; Howard *et al.*, 2013); [i] (Buijs *et al.*, 2015); [j] (Bernard *et al.*, 2012); [k] (Rude *et al.*, 2011; Liu *et al.*, 2014); [l] (Rui *et al.*, 2014).

5.1.3 Aims

The aims of experiments described in this chapter are to provide a proof of principle genetic engineering strategy, converting fatty acid precursors into hydrocarbons in the industrially relevant *S. cerevisiae* prototrophic strain CEN.PK113-7D.

5.2 Results and discussion

5.2.1 Development of a metabolic engineering strategy for production of hydrocarbons in *S. cerevisiae*

In order to produce alkenes in *S. cerevisiae* three main factors were considered: i) increasing free fatty acid precursors; ii) redox balance; and iii) conversion of free fatty acid precursors to alkenes.

(i) *increasing free fatty acid precursors*. Free fatty acids are potentially toxic to the cell and due to their high energy state they are tightly regulated at multiple levels within the multiple subcellular organelles involved in fatty acid metabolism (Tehlivets *et al.*, 2007). In order to increase fatty acid-derived biofuels in *S. cerevisiae*, the regulatory elements involved in fatty acid homeostasis must be overcome. Metabolic engineering strategies to increase free fatty acid production in the yeast *S. cerevisiae* have been published and have thus far resulted in a 670-fold improvement (Runguphan and Keasling, 2014). Gratifyingly, the production of 400.2 mg l^{-1} free fatty acids is not toxic to the cell. Figure 5.3 summarises the metabolic engineering strategy that included heterologous expression of an *E. coli* thioesterase, deletion of genes associated with fatty acid degradation and modification of promoter sequences in order to overexpress fatty acid biosynthesis genes.

In light of the published data, an engineering strategy was chosen comprising heterologous expression of the leaderless sequence version of the thioesterase from *E. coli* (Steen *et al.*, 2010) and deletion of *FAA1*, to increase free fatty acids in *S. cerevisiae*. This combination has been shown to increase free fatty

acid production from 0.6 mg l^{-1} to 164.3 mg l^{-1} (Runguphan and Keasling, 2014) and was chosen as a trade off between the free fatty acid increase and the required genetic engineering effort.

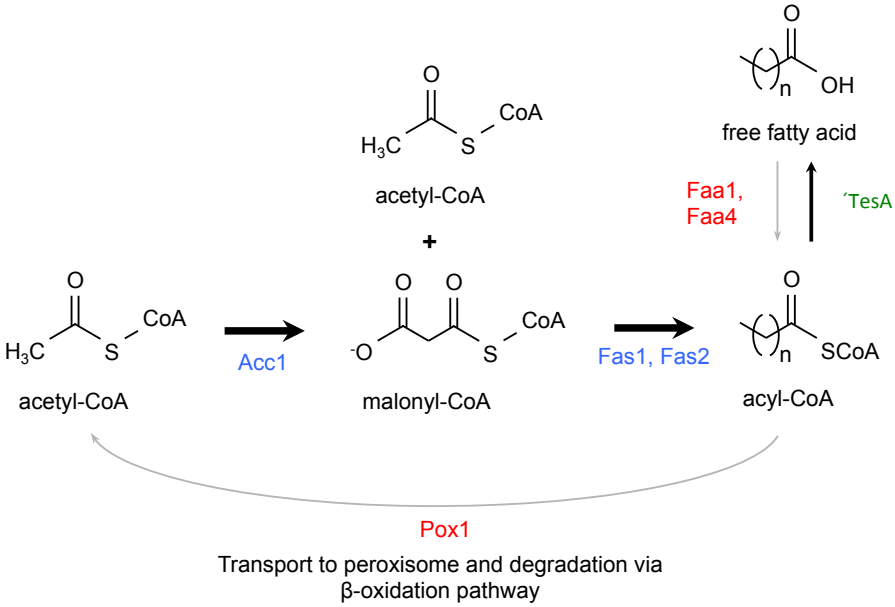


Figure 5.3 The engineered pathway for overproduction of free fatty acids in *S. cerevisiae* (adapted from Runguphan and Keasling, 2014). Increased flux through the yeast fatty acid biosynthesis pathway comprised changing the promoters of *ACC1*, *FAS1* and *FAS2* to the strong constitutive promoter *TEF1*, deleting genes (in red) associated with degradation and heterologous expression of a leaderless version of the *E. coli* thioesterase ´TesA.

The leaderless sequence version of the *E. coli* ´TesA lacks the membrane signal peptide at the N-terminal end, providing cytosolic expression of the thioesterase. Appropriately for heterologous expression in yeast, ´TesA has been shown to exhibit both acyl-ACP and acyl-CoA activities with a range of chain lengths between C_8-C_{18} (Steen *et al.*, 2010).

(ii) *redox balance*. The maintenance of redox balance between the cofactors $NADH/NAD^+$ and $NADPH/NADP^+$ is important for the metabolism and cell

growth of all living organisms (Wang *et al.*, 2012). The redox pairs within *S. cerevisiae* have distinct roles in metabolism, where NADH predominantly functions as a reducing equivalent in catabolic reactions such as glycolysis, and NADPH functions in anabolic reactions such as fatty acid biosynthesis (Nissen *et al.*, 2001). Additionally, the redox balance must be locally maintained within cellular compartments, as redox pairs cannot cross the mitochondrial inner membrane (Heux *et al.*, 2006).

The potential impact on redox balance of alkene production in *S. cerevisiae* strains with a reduced carbon flux to ethanol production is shown in Figure 5.4B, resulting in the production of glycerol and loss of carbon. A solution to the system's redox imbalance may be through expression of a transhydrogenase, an activity absent in yeast (Nissen *et al.*, 2001). There are two structurally unrelated types of transhydrogenases; the BB-specific transhydrogenases are soluble energy-independent flavoproteins (Boonstra *et al.*, 1999; Cao *et al.*, 2011) and catalyse the following reaction with a standard Gibbs free energy close to zero but favouring the conversion of NADPH to NADH (Hou *et al.*, 2009):

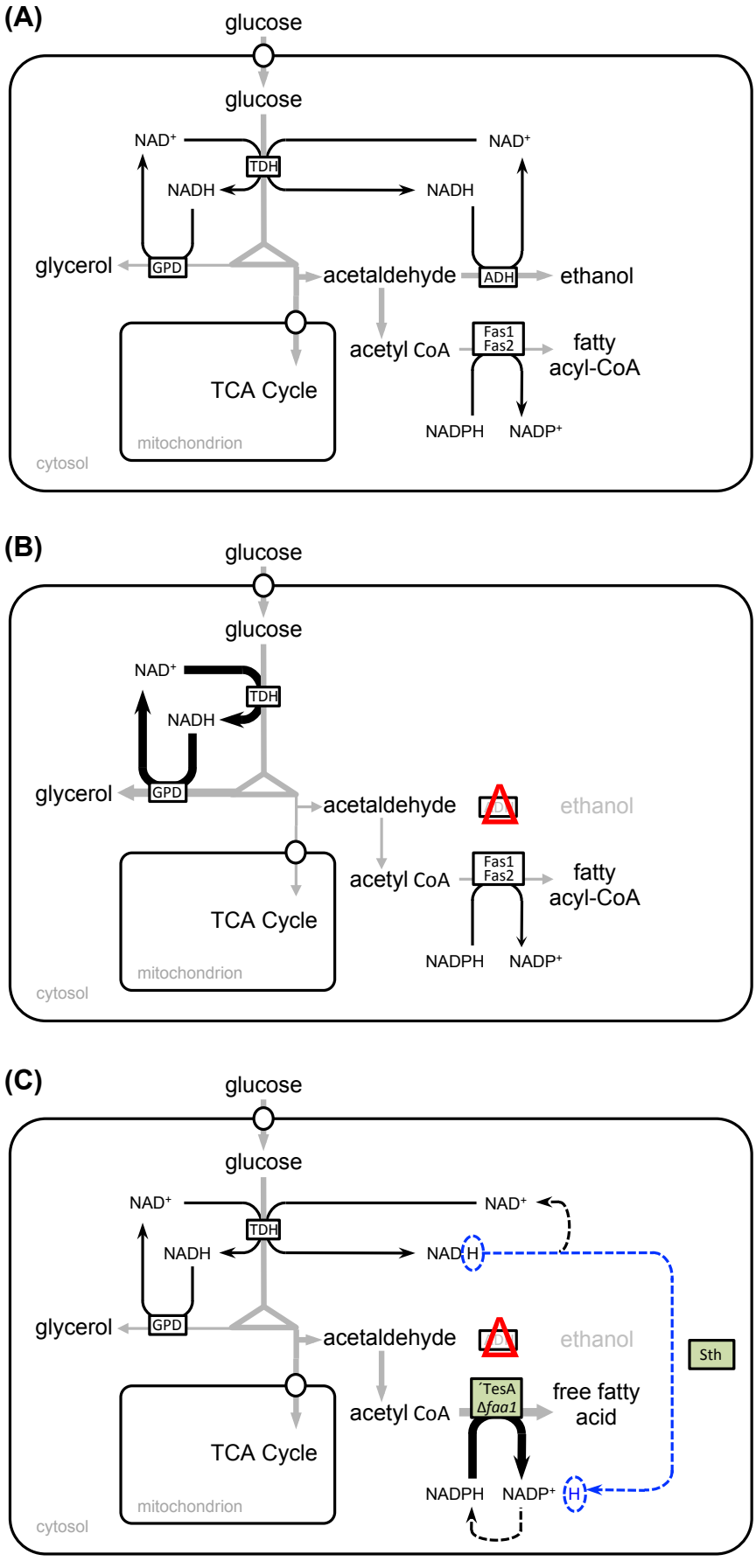
$$NADH + NADP^+ \rightleftharpoons NAD^+ + NADPH$$

Hydrogen transfer between the NAD(H) and NADP(H) cofactor systems has been evaluated previously both *in vitro* and heterologously expressed in *S. cerevisiae*. This analysis indicates that the reaction favours the conversion of NADPH to NADH (Chung, 1970; Nissen *et al.*, 2001; Hou *et al.*, 2009). This would be undesirable for the redox balance of the engineered strain shown in Figure 5.4, however during growth on ethanol the reaction was shown to favour

the production of NADPH (Hou *et al.*, 2009). It was postulated that this enzyme would be capable of the reversible reaction converting NADH to NADPH in an *S. cerevisiae* strain engineered with a reduced flux of carbon toward ethanol production, notwithstanding the previous evaluation of the transhydrogenase in yeast (Nissen *et al.*, 2001).

The soluble transhydrogenase gene (*Sth*) from *Azotobacter vinelandii* was chosen to be heterologously expressed as it has a higher specific transhydrogenase activity (4.53 U mg^{-1} protein (Nissen *et al.*, 2001)) compared to the soluble transhydrogenase (undA) from *E. coli* (2.9 U mg^{-1} protein (Boonstra *et al.*, 1999)) and the enzyme has been previously expressed in *S. cerevisiae* (Nissen *et al.*, 2001).

Figure 5.4 Overview of glucose fermentation, highlighting cofactor redox balance during glycolysis and fatty acid biosynthesis, in genotypic variants of *S. cerevisiae*. (**A**) wild type. (**B**) engineered strain in which ethanol production is perturbed (Δ). Compared to the situation in (A), (B) demonstrates increased production of glycerol. This has been proposed to be the mechanism to maintain redox homeostasis to the detriment of biomass production (shown by Ida *et al.*, 2012). (**C**) proposed engineered strain where ethanol production is perturbed (Δ) and the carbon flux toward free fatty acid biosynthesis is increased (´TesA and Δfaa1). The heterologous expression of the cytosolic transhydrogenase (Sth) from *A. vinelandii* has the potential to maintain redox balance and ensure optimal carbon flux within the cell. The mode of action of the transhydrogenase is shown by the blue dashed line, where hydrogen from NADH produced during glycolysis is transferred to the cofactor $NADP^+$ produced during the synthesis of free fatty acids.


(A)
glucose
glucose
NAD+
NAD+
TDH
NADH
NADH
glycerol
GPD
acetaldehyde
ADH
ethanol
acetyl CoA
Fas1
Fas2
fatty
acyl-CoA
TCA Cycle
mitochondrion
NADPH
NADP+
cytosol
(B)
glucose
glucose
NAD+
TDH
NADH
glycerol
GPD
acetaldehyde
ethanol
acetyl CoA
Fas1
Fas2
fatty
acyl-CoA
TCA Cycle
mitochondrion
NADPH
NADP+
cytosol
(C)
glucose
glucose
NAD+
NAD+
TDH
NADH
NADH
glycerol
GPD
acetaldehyde
ethanol
Sth
acetyl CoA
´TesA
Δfaa1
free fatty
acid
TCA Cycle
mitochondrion
NADPH
NADP+
H
cytosol

(iii) *conversion of free fatty acid precursors to alkenes*. The production of hydrocarbons from fatty acid precursors either directly or with the intermediary production of an aldehyde precursor is shown in Figure 5.2. The application of several of these pathways within the industrial cell factory *S. cerevisiae* has been demonstrated including the production of very long chain alkanes with the expression of the plant aldehyde decarbonylase (Bernard *et al.*, 2012), and the production of C_{13}, C_{15} and C_{17} alkanes via the heterologous expression of a cyanobacterial aldehyde deformylating oxygenase (Buijs *et al.*, 2015). Using *OleT*$_{JE}$ from *Jeotgalicoccus* sp. to produce terminal olefins (1-alkenes) from free fatty acid precursors is an interesting single-enzyme engineering strategy that would increase the heterologous product portfolio of the yeast *S. cerevisiae*.

The *OleT*$_{JE}$ enzyme is a cytochrome P450 fatty acid peroxygenase (Rude *et al.*, 2011). It catalyses the formation of *n*-1 alkenes through the H_2O_2-dependent decarboxylation of C_{12}, C_{14}, C_{16}, C_{18} and C_{20} saturated fatty acids (Belcher *et al.*, 2014). H_2O_2 is a reactive oxygen species and can cause cellular damage and result in apoptosis. The cell elicits many responses leading to rapid detoxification of reactive oxygen species (Perrone *et al.*, 2008). For commercial production it is preferable to avoid the toxic electron and oxygen donor, H_2O_2. OleT$_{JE}$ is capable of performing H_2O_2-independent catalysis using either a reducing system (*in vitro* assay including the *E. coli* flavodoxin/flavodoxin reductase) or fused to the RhFRED reductase domain from *Rhodococcus* spp. NCIMB 9784 (Roberts *et al.*, 2002; Liu *et al.*, 2014). *S. cerevisiae* does possess homologs of the *E. coli* reducing system (*Yah1* and *Arh1*) however they are present in the mitochondria and not available for cytosolic production of alkenes

(Buijs *et al.*, 2015). Therefore it was decided to use *OleT*$_{JE}$ fused to the RhFRED reductase; the hypothesised cofactor balance is shown in Figure 5.5. OleT-RhFRED has been shown to perform the same decarboxylation activity supported with NADPH and oxygen in an engineered fatty acid-overproducing strain of *E. coli* (Liu *et al.*, 2014), with a substrate preference spectrum (C_{12} to C_{18}) appropriate for alkene production from free fatty acid precursors in yeast.

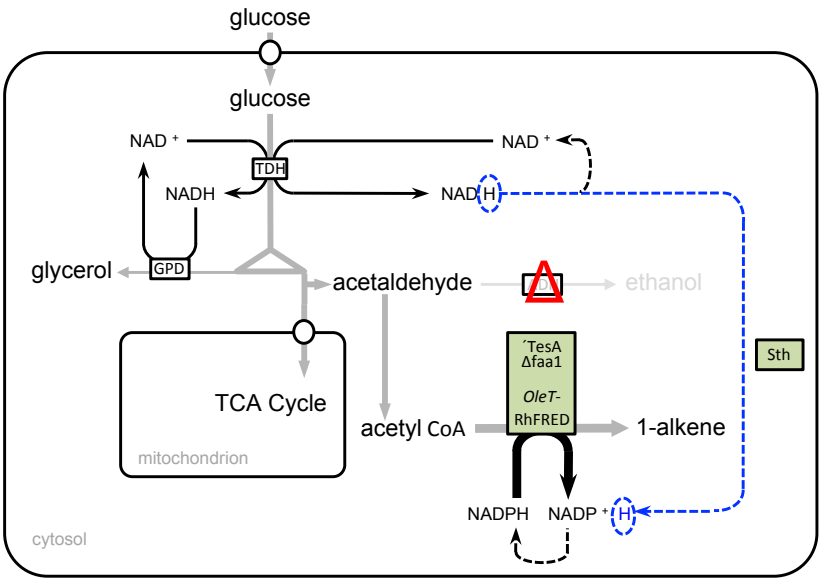


Figure 5.5 Overview of the metabolic engineering strategy for the production of terminal olefins from free fatty acid precursors in *S. cerevisiae* using OleT-RhFRED. The engineered yeast strain includes perturbation of ethanol production, increased carbon flux toward free fatty acid biosynthesis and a strategy to balance the anticipated redox imbalance of the system.

5.2.2 Generation of *S. cerevisiae* ADH gene variants with a Δ*faa1* genotype

The metabolic engineering strategy for increased production of free fatty acids in *S. cerevisiae* includes deletion of *FAA1*, which is involved in free fatty acid degradation. Figure 5.6 shows deletion of *FAA1* in the *S. cerevisiae* strain CEN.PK113-7D, by integration of a specific *amdSYM* deletion cassette (100 % efficiency) and subsequent scarless marker removal. Additionally, in order to evaluate the production of hydrocarbons in strains with a reduced flux of carbon toward ethanol, the genotypes CEN.PK113-7D *Δadh1 Δfaa1* and CEN.PK113-7D *Δadh1/2/3/4/5/sfa1 Δfaa1* were also engineered. Adh1 is the main cytosolic ADH isozyme of yeast (Dickinson, 2004) and *Δadh1* is the genetic background used in the commercialisation of succinic acid production from renewable sources (Jansen *et al.*, 2012) by the company Reverdia (Geleen, Netherlands).

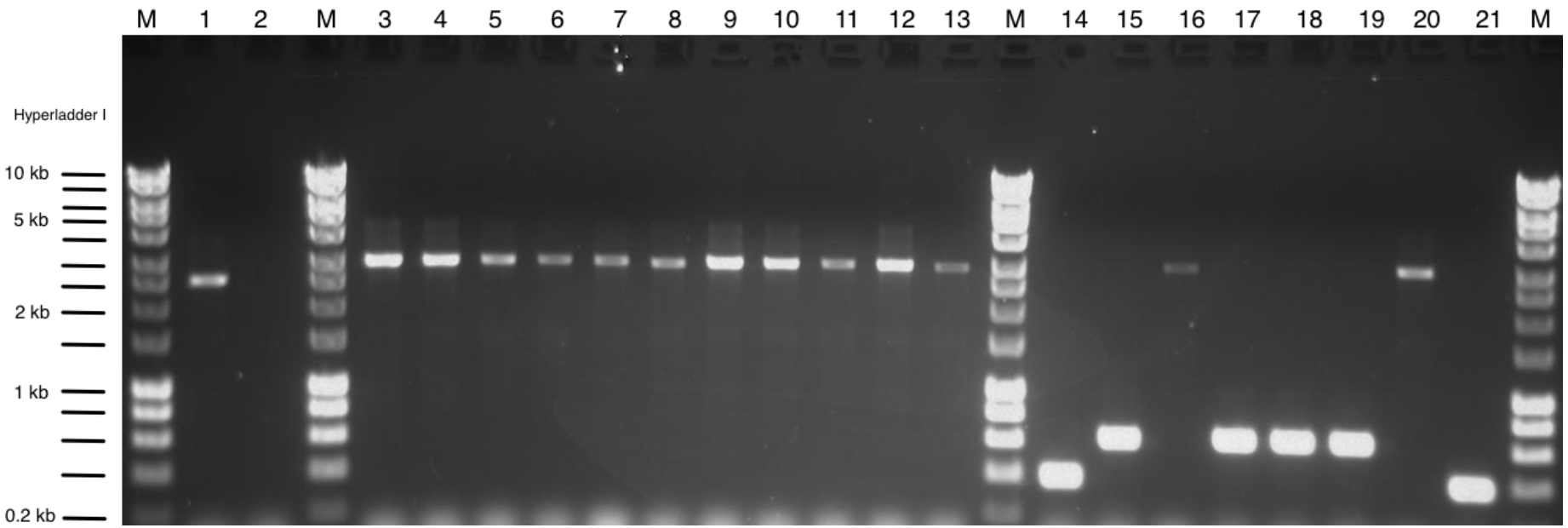


Figure 5.6 Diagnostic colony PCR verification of a CEN.PK113-7D *Δfaa1*::*amdSYM* genotype and subsequent excision of the *amdSYM* selectable marker. All PCRs were performed at the standardised annealing temperature of 63.3 °C. Genes, genotype variants and expected fragment sizes are: Lane 1, *FAA1* (2501 bp); Lane 2, negative control; Lanes 3-13, *Δfaa1*::*amdSYM* (2940 bp), 47 transformants per µg of DNA; Lanes 14 and 21, *Δfaa1* (400 bp), the correct excision frequency was 25 % of the eight screened colonies. The band positions in Lanes 16 and 20 indicate the intact *Δfaa1*::*amdSYM* deletion cassette (2940 bp); Lanes 15, 17-19 indicate incomplete excision of the *amdSYM* selection marker (~ 650 bp).

5.2.3 Generation of the pAlkene_3genes yeast expression vector variants

The yeast expression vector pAlkene_3genes was designed for the heterologous expression of genes as required by the metabolic engineering strategy (Figure 2.2 materials and methods) and was synthesised by DNA 2.0 Inc. (California, US). The genes were codon optimised for expression in the yeast *S. cerevisiae* and the promoters were chosen to be constitutively expressed at similar high levels during glucose metabolism by *S. cerevisiae* (Partow *et al.*, 2010). The variants of the pAlkene_3genes plasmid required for evaluation within *S. cerevisiae* were generated as detailed in Table 5 (materials and methods), and contained different combinations of the genes encoding the OleT peroxygenase (with or without the RhFRED reductase domain fusion), the ´TesA thioesterase, and the Sth transhydrogenase. Figure 5.7 shows the results of diagnostic *Apa*LI digests of the pAlkene_3genes plasmid variants confirming the desired expression vectors for analysis.

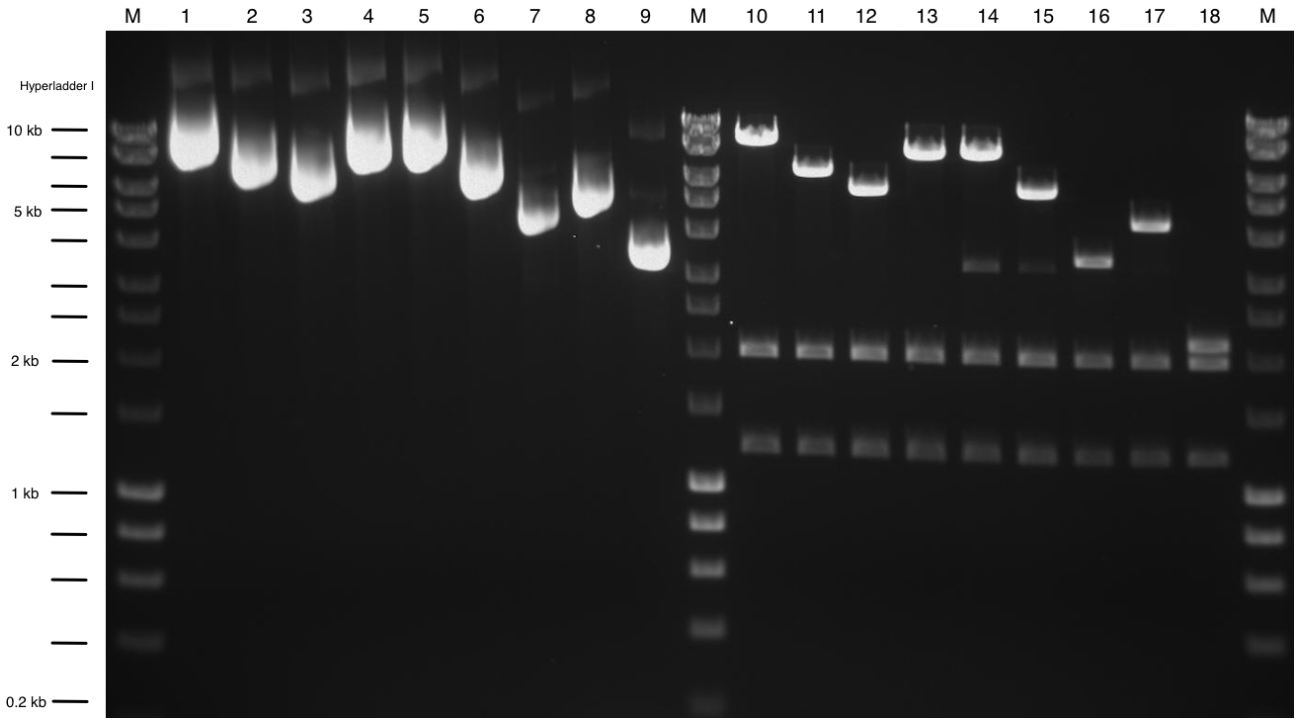


Figure 5.7 Diagnostic digest (*Apa*LI) verification of the yeast expression plasmid pAlkene_3genes variants.

Lanes 1-9 show uncut plasmids and lanes 10-18 show the respective *Apa*LI digests. Plasmid variants and expected fragment sizes are: Lanes 1 and 10 (8358/1928/1199 bp) **pO1** (**OleT-RhFRED/´TesA/Sth**); Lanes 2 and 11 (6282/1928/1199 bp) **pO2** (**OleT-RhFRED/´TesA**); Lanes 3 and 12 (5383/1928/1199 bp) **pO3** (**´TesA/Sth**); Lanes 4 and 13 (7362/1928/1199 bp) **pO4** (**OleT/´TesA/Sth**); Lanes 5 and 14 (7390/1928/1199 bp) **pO5** (**OleT-RhFRED/Sth**); Lanes 6 and 15 (5314/1928/1199 bp) **pO6** (**OleT-RhFRED**); Lanes 7 and 16 (3307/1928/1199 bp) **pO7** (**´TesA**); Lanes 8 and 17 (4189/1928/1199 bp) **pO8** (**Sth**); Lanes 9 and 18 (2113/1928/1199 bp) **pO9** (**Empty Vector**).

5.2.4 Empirical evaluation of the metabolic engineering strategy for production of alkenes in *S. cerevisiae*

A Design of Experiments (DoE) approach was used to evaluate the pAlkene_3genes plasmid variants in different ADH mutant genotype backgrounds of *S. cerevisiae.* The experimental aim was to evaluate the metabolic engineering strategy for the production of alkenes in yeast and include environmental impacts which could be easily assessed using the shake flask experimental format, including incubation temperature, starting glucose concentration and the impact of sample time on the system (either 2 or 4 days after inoculation of the culture).

The custom design function was used as part of the DoE modelling software package (JMP Pro v.12, SAS Institute Inc. USA). A custom design allows the greatest flexibility in the development of experimental design and modelling strategies and considers constraints such as the number of experiments that can be performed and factor combination constraints. Subsequently, empirically derived models provide an estimation of the effects of the chosen experimental parameters. Table 17 shows the associated levels of experimental parameters (or factors) that were identified as a part, or potentially influencing the evaluation, of the metabolic engineering strategy for production of alkenes in yeast (section 5.2.1). The temperatures represented the laboratory standard growth temperature for yeast culture and a higher value balancing ethanol yield and culture viability (Torija *et al.*, 2003). The glucose concentration levels were chosen providing a range in order to evaluate concentrations shown to affect the rate of sugar uptake in *S. cerevisiae* (Rossi *et al.*, 2010).

Table 17. List of the experimental factors, and associated levels of each, for evaluation of the metabolic engineering strategy for production of alkenes in yeast.

Factor	Role	Factor level		
Alkene enzyme	Categorical	OFF	*OleT*	*OleT-RhFRED*
Thioesterase	Categorical	OFF	ON (´TesA)	
Transhydrogenase	Categorical	OFF	ON (Sth)	
ADH genotype	Categorical	Wild type	Δ*adh1*	Δ1/2/3/4/5/6
FAA1 genotype	Categorical	OFF (Δ*faa1*)	ON (Wild type)	
Initial [glucose]	Continuous	10 g l^{-1}	70 g l^{-1}	
Temperature	Continuous	30 °C	33 °C	
Sample time	Continuous	2 days	4 days	

The custom experimental design provided by the DoE methodology is shown in Table 18 with the defined 12 experimental runs providing an estimate for all the main effects (factors) listed in Table 17; interactions between two factors are aliased and cannot be evaluated using this design. The full factorial design would require 576 experiments. A OFAT approach evaluating the same design space would require considerably more experimentation. The experimental design incorporates the availability of plasmid variants (Figure 5.7), which did not include all possible combinations, for example, the plasmid variants expressing the enzyme OleT and either the thioesterase or transhydrogenase. This constraint is reflected in the power analysis of the design (i.e. the

probability of detecting a difference in an effect at a significance level of 0.05). The assessment of the effect of OleT has a value of 0.188 compared, for example, to the temperature effect (unconstrained) with a power value of 0.406. Moreover, the constrained experimental design results in correlation of the experimental variables. Correlation of variables can be troublesome for explanatory models, as parameter estimates for predictor effects on the response can be highly variable. The extent of the correlation was visually summarised using the map of correlations (Figure 5.8), and shows that the maximum absolute correlation value of certain predictors for the design is 0.333. The experimental variables are not highly correlated, however, the reduced power of analysis in determining an estimation of an effect of certain predictors on a response needs to be accounted for when evaluating the models and generating conclusions.

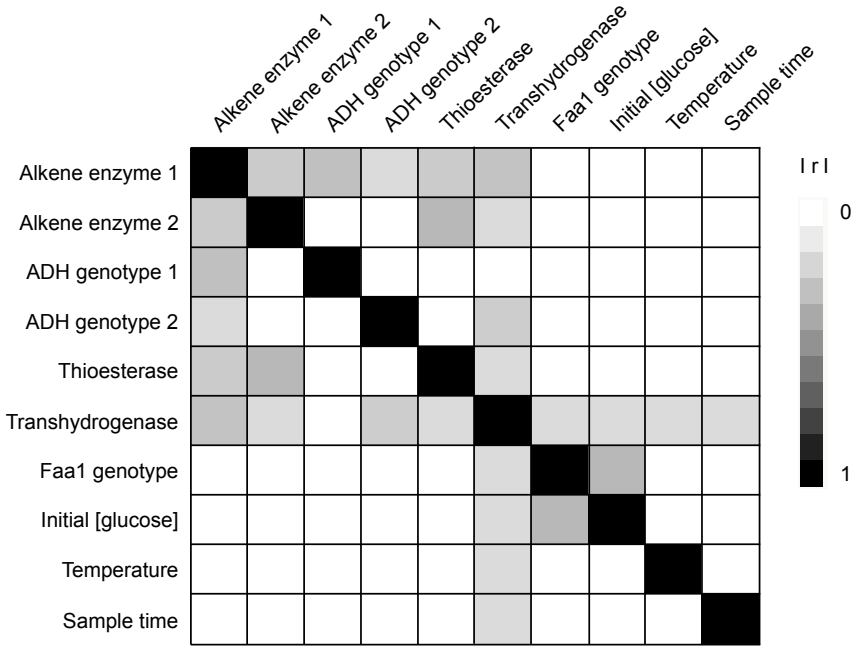


Figure 5.8 Map of correlation for the experimental design evaluating alkene production in yeast. The scale (r) for the absolute correlation values range from no correlation in white (0 value) through to black indicating an absolute correlation value of one; the black diagonal shows the correlation of the model terms with themselves.

Table 18. Experimental design providing a model estimate for all the main effects of the metabolic engineering strategy factors evaluated for production of alkenes in yeast.

	Enzymes				Genotypes		Environmental conditions		
Run	Alkene enzyme	Thioesterase	Trans-hydrogenase	Plasmid variant	ADH genotype	*FAA1* genotype	Initial [Glucose] (g l^{-1})	Temperature (°C)	Sample time (days)
1	OleT	ON (´TesA)	ON (Sth)	pO4	Δ1/2/3/4/5/6	Wild type	10	33	2
2	OFF	OFF	OFF	pO9	Wild type	OFF (Δ*faa1*)	10	30	2
3	OleT-RhFRED	OFF	OFF	pO6	Δ1/2/3/4/5/6	OFF (Δ*faa1*)	70	33	2
4	OFF	ON (´TesA)	OFF	pO7	Δ1/2/3/4/5/6	Wild type	10	30	4
5	OleT	ON (´TesA)	ON (Sth)	pO4	Δ1	OFF (Δ*faa1*)	70	30	4
6	OleT-RhFRED	OFF	ON (Sth)	pO5	Δ1/2/3/4/5/6	OFF (Δ*faa1*)	70	30	4
7	OFF	ON (´TesA)	ON (Sth)	pO3	Δ1	OFF (Δ*faa1*)	70	33	2
8	OleT-RhFRED	OFF	OFF	pO6	Δ1	Wild type	10	33	4
9	OleT-RhFRED	ON (´TesA)	OFF	pO2	Wild type	Wild type	70	30	2
10	OFF	OFF	ON (Sth)	pO8	Wild type	Wild type	70	33	4
11	OleT-RhFRED	ON (´TesA)	ON (Sth)	pO1	Wild type	OFF (Δ*faa1*)	10	33	4
12	OleT-RhFRED	OFF	ON (Sth)	pO5	Δ1	Wild type	10	30	2

Table 19 lists the responses measured during the evaluation of the metabolic strategy for alkene production in yeast. Included are measurements of alkenes and their precursors (fatty acids), and also response measurements of both carbon utilisation and distribution.

Table 19. Measured responses of experiment to evaluate the production of alkenes in yeast.

Response	Units	Analytical method
Glucose	$g\ l^{-1}$	HPLC
Glycerol	$g\ l^{-1}$	HPLC
Ethanol	$g\ l^{-1}$	HPLC
Cell count	$cells\ ml^{-1}$	Flow cytometry
Culture viability	%	Flow cytometry
Fatty acids (C_{14}, C_{16}, C_{18}, $C_{18:1}$)	$mg\ l^{-1}$	GC FID
Alkenes (1-C_{13}, 1-C_{15}, 1-C_{17})	$mg\ l^{-1}$	GC FID

The initial analysis of the results included both the measured responses (Table 19) and the following transformed data responses: i) total free fatty acids standardised to cell count (1.0×10^8), and ii) glucose utilised. The PLS modelling platform was used for initial screening as it is a flexible, effective method for modelling experimentation where the experimental variables are correlated (multicollinearity) and noisy (Cox and Gaudard, 2013).

The modelling validation was performed using KFold (10) cross validation and the SIMPLS algorithm applied to the data set. The coefficient estimates of X

and Y for models were comparatively evaluated using the root mean PRESS statistic.

A PLS model fitted using 1 factor to describe the relationship between X and Y matrices had a minimum root mean PRESS of 1.0613 and explained 16.1 % of the cumulative variation in the X score and 39.8 % of the cumulative variation in the Y score. The capacity of the model to predict each of the measured responses was evaluated; Figure 5.9 shows the actual by predicted diagnostic plots of two measured responses. We can see from Figure 5.9 that “glucose consumed’ and “ethanol produced” data were not predicted well by the validated model. Experimentally, the use of shake flasks and a single sample point prevented the accurate analysis of both glucose consumption and ethanol production in the system. Ethanol is produced and consumed by *S. cerevisiae* at different stages of the batch experimentation; as with the ADH knockout library evaluation, a steady state analysis would be required to assess ethanol production. The glucose consumption and ethanol production data were therefore removed from subsequent model evaluations as they negatively impacted the explanatory power of the models. Additionally, the data for alkene production were removed as a measured response, as no alkenes were detected during this set of experiments (data not shown). The implications of this result are discussed in section 5.2.5.

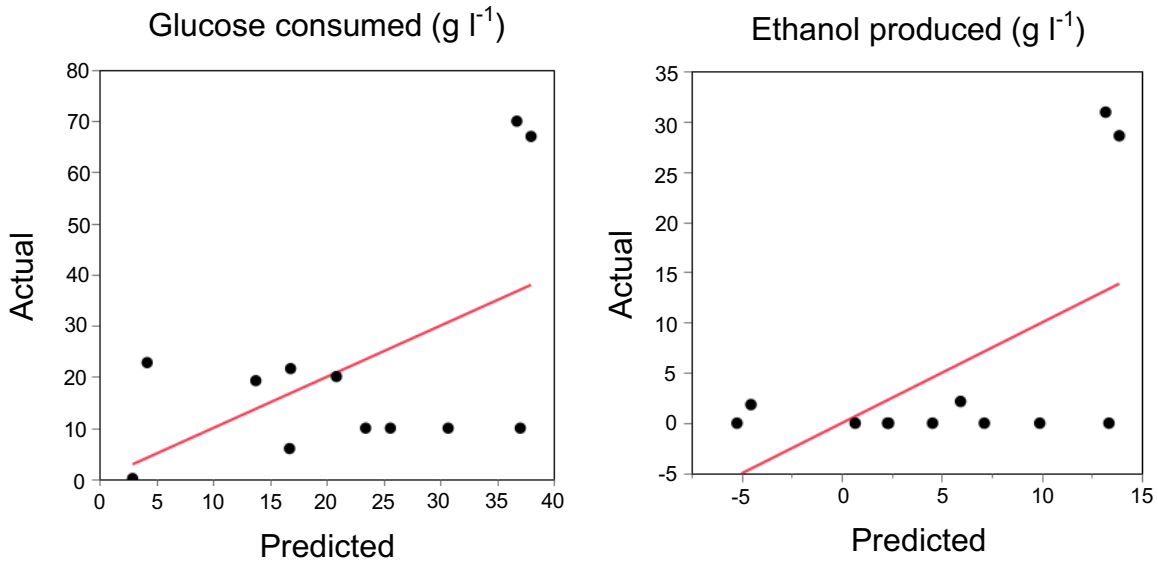


Figure 5.9 Actual by predicted diagnostic plots showing the transformed measured responses, glucose consumed and ethanol produced, that are not predicted well by the validated PLS model using SIMPLS fit with 1 factor.

Accordingly, the responses that could be measured reliably were assessed using the PLS modelling methodology. The modified aim was to elucidate the impact of elements of the metabolic engineering strategy on the biosynthesis of the alkene precursors (fatty acids) and carbon distribution within the cell. Table 20 shows the modelling parameters for each of the PLS models including a measure of the variance of the model prediction for that response (root mean PRESS) and the percentage of the dataset explained. Model selection was subject to a trade-off between the root mean PRESS and the number of factors involved in the PLS model fit that described the most variation in the response, Y. This assessment was required to ensure a useful model was chosen that sufficiently explains the response (compared to a cross validated data set) without over-fitting the data set.

Table 20. Evaluation of the KFold cross-validated PLS models for each of the measured responses including values of the root mean PRESS (predicted residual sum of squares) statistic and percentage of the dataset explained. The individual fatty acid measured responses were analysed using a single model.

Measured Response	Root Mean PRESS	Variation explained for cumulative X (%)	Variation explained for cumulative Y (%)
Glycerol	1.1000	28.0	97.2
Cell count	0.8340	41.0	98.6
Culture viability	1.1271	42.9	98.5
Myristic acid (C_{14})	0.9249	54.4	93.6
Palmitic acid (C_{16})			
Stearic acid (C_{18})			
Oleic acid ($C_{18:1}$)			
Total free fatty acids standardised to cell count	1.0047	18.3	65.3

The multiple explanatory predictors (factors) included in modelling the responses (Table 20) are standardised in the following analysis, permitting a direct comparison of individual predictor effects on the model projections. The centred and scaled data results in predictors having a mean of 0 and a standard deviation of 1 (Cox and Gaudard, 2013); this can help clarify the assessment of what - and to what extent - a predictor(s) is having on a response.

Figures 5.10, 5.12, 5.13 and 5.18 are an analysis of the PLS models, and include the variable importance plots for a measured response. The variable importance plot includes the summary statistic VIP (variable importance for the projection), a weighted sum of squares score calculated from the amount of

variance attributed to each predictor defining the model projection (Wold *et al.*, 2001). If a predictor has a small VIP (a value of ≤ 0.8) then it is a candidate for removal from the model, whereas a VIP value > 1.0 indicates predictors that are important to the generation of the model (Cox and Gaudard, 2013). The VIP cut-off values provide a criterion to simplify the factor terms used to predict the measured response, however unimportant predictors may still be biologically significant and should be assessed using the model prediction profiler as well as the raw experimental data.

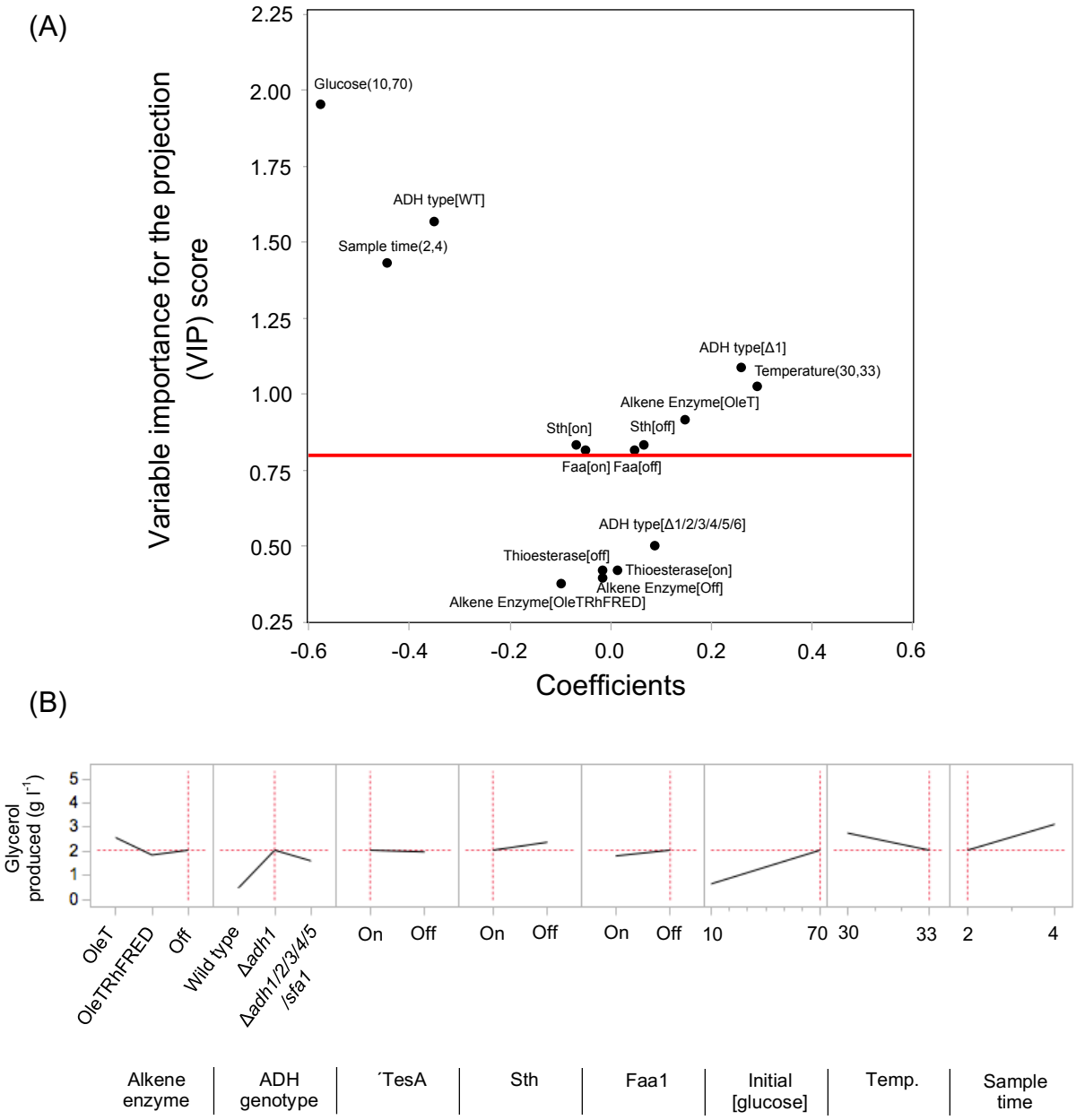


Figure 5.10 Model analysis for glycerol production. (A) Variable importance plot of VIP (variable importance for the projection) scores for experimental predictors against the centred and scaled data coefficients. Standardised coefficients highlight if a predictor is having a positive or negative impact, as well as the magnitude of that impact on the measured response. Predictors below the VIP value of 0.8 (red line) are unimportant to the explanatory model for glycerol production (Cox and Gaudard, 2013); (B) The prediction profiler of the model projection for glycerol production. The red dashed vertical lines highlight the selected set point for each of the experimental factors; their selection affects the horizontal red dashed line that indicates the amount of glycerol produced dependent upon the validated model prediction formula.

The PLS model evaluation of glycerol produced in *S. cerevisiae* (Figure 5.10) provides inconclusive evidence that the inclusion of Sth in the constructs influences glycerol production. The variable importance of the Sth predictor on the model projection (VIP = 0.83) is between the cut-off values of the predictor either being removed or seen as important for generation of the model ($0.8 \leq 0.83 \leq 1.0$). Glycerol production in *S. cerevisiae* plays important physiological roles in both osmoregulation (preventing cell lysis) and maintenance of the cellular redox balance (Cao *et al.*, 2007; Navarrete *et al.*, 2014). During the industrial bioproduction of ethanol, glycerol is a by-product accounting for as much as 4 to 5 % of the total carbon flux. This negatively affects the economics of the process (Nissen *et al.*, 2000). It had been previously postulated that the use of a soluble transhydrogenase (Sth) would maintain the redox balance in strains of *S. cerevisiae* where ethanol production had been perturbed and fatty acids were overproduced. In turn this would reduce the metabolic requirement to produce glycerol and concomitant loss of carbon. Encouragingly, the empirically derived model projection shows that expression of the Sth reduces the amount of glycerol produced by 11-12 % dependent upon the ADH type of the *S. cerevisiae* strain. The role of the Sth reaction in the reduction of glycerol remains speculative, however, particularly without associated protein expression data.

Evaluation of the effect of different combinations of ADH knockouts on glycerol production was of interest as ethanol and glycerol production function in the maintenance of the cellular redox balance, shifting between the first and second forms of Neuberg's fermentation (Cao *et al.*, 2007; Drewke *et al.*, 1990). The model shows that wild type and Δ*adh1* genotypes were important predictors for

glycerol production, with VIPs of 1.57 (reducing glycerol production) and 1.09 (increasing glycerol production) respectively. The data profile from the model projection shows that the Δ*adh1* genotype significantly increases glycerol production by ~63 % compared to the wild type ADH genotype. The Δ*adh1/2/3/4/5/sfa1* (complete ADH knockout) genotype meanwhile increases the production of glycerol by a comparable ~45 %. However, a full ADH knockout genotype is not a factor that is important to the model for predicting glycerol production (VIP = 0.50). This is because the largest influence on glycerol production is removal of ADH1, with further gains by removal of additional ADHs being less significant. The PLS modelling methodology describes the variation in the empirical measurements for glycerol production which can be attributed to or explained by the levels of the ADH genotype categorical factor (modelled within the context of the other experimental factors). It uses only wild type and Δ*adh1* genotypes in order to sufficiently explain this variation. This highlights the requirement to evaluate the data profile of the model prediction to fully understand the impact of experimental factors on a measured response. The Δ*adh1/2/3/4/5/sfa1* genotype therefore is not an important predictor in the model for the glycerol production, however strains with this ADH genotype will produce more glycerol compared to wild type strains.

The availability of carbon (initial glucose concentration) has the largest impact as a predictor on the amount of glycerol produced (VIP = 1.95); the lower the amount of carbon (glucose) available, the less glycerol is made. The defined experimental lower limit of 10 g l^{-1} initial glucose concentration was selected to provide a glucose concentration which may perturb ethanol metabolism (as

discussed in Chapter 3), however this concentration became limiting before the fourth day sample point leading to the unintentional evaluation of carbon limitation in experimental runs 4, 8 and 11 (Table 18). This did not occur in the other experimental runs. Increasing the lower concentration of glucose for subsequent experiments may reduce or remove the impact of initial glucose concentration on the model projection and refine the model used for evaluation of the metabolic engineering strategy on glycerol production.

The sampling time points also had a significant impact (VIP = 1.43) on the glycerol production projection, with less glycerol produced after two days compared to sampling cultures after four days. The increasing production of glycerol as the time of cell culture increases is intuitive providing substrates and metabolic precursors are available, however glycerol can be used as a carbon source by *S. cerevisiae* with assimilation linked to respiration and occurring in the absence of glucose (Scanes *et al.*, 1998). Figure 5.11 shows that a large proportion of the experiments with lower amounts of glycerol had consumed all the initial glucose provided (blue to red scale of data points showing percentage glucose utilised) and strains in these experiments were likely to have re-assimilated the produced glycerol as a carbon source. Therefore, whilst this result is broadly intuitive, it must be born in mind that glycerol concentration does not equal carbon flux to glycerol, which may be underestimated.

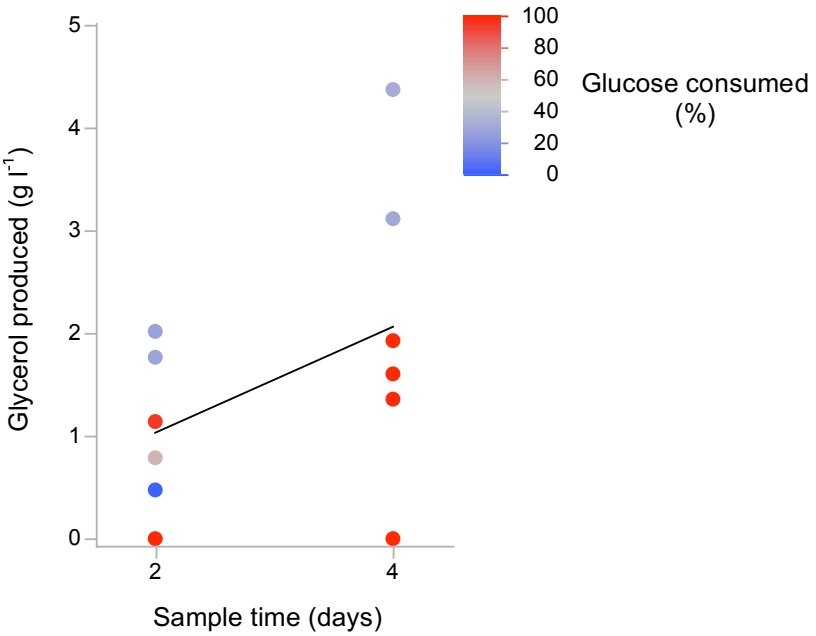


Figure 5.11 Experiment measurements of glycerol produced (g l^{-1}) at two different sample times (2 or 4 days). The data points are coloured according to the percentage of glucose consumed at the point of measurement, ranging from dark blue = 0 % to dark red = 100 %. The lower glycerol measurements for each of the sample times coincide with depletion of glucose substrate (dark red data points) and is likely to be due to assimilation of glycerol by *S. cerevisiae* under these conditions, leading to an underestimate for prediction of glycerol production using the PLS model (Figure 5.10). The black line connects the mean value of glycerol produced for each of the sample times and highlights the positive correlation between glycerol production and sample time.

Culture temperature was also an important predictor for glycerol production (VIP = 1.02). Figure 5.10(B) shows the impact of temperature which is in agreement with the literature. Increasing temperature increases glycerol production in *S. cerevisiae* up to a maximum at 30 °C. Beyond this further increase in temperature causes a reduction in the amount of glycerol produced (Yalcin and Yesim Ozbas, 2008).

Assessment of the PLS model projection for glycerol production has highlighted experimental developments required to improve its accuracy of prediction. The

model has, however, proved useful in assessing the general contributions of the main experimental factors on glycerol production. The model projection is specific to the industrially relevant *S. cerevisiae* strain developed during this project and the general results are supported by other studies represented in the literature.

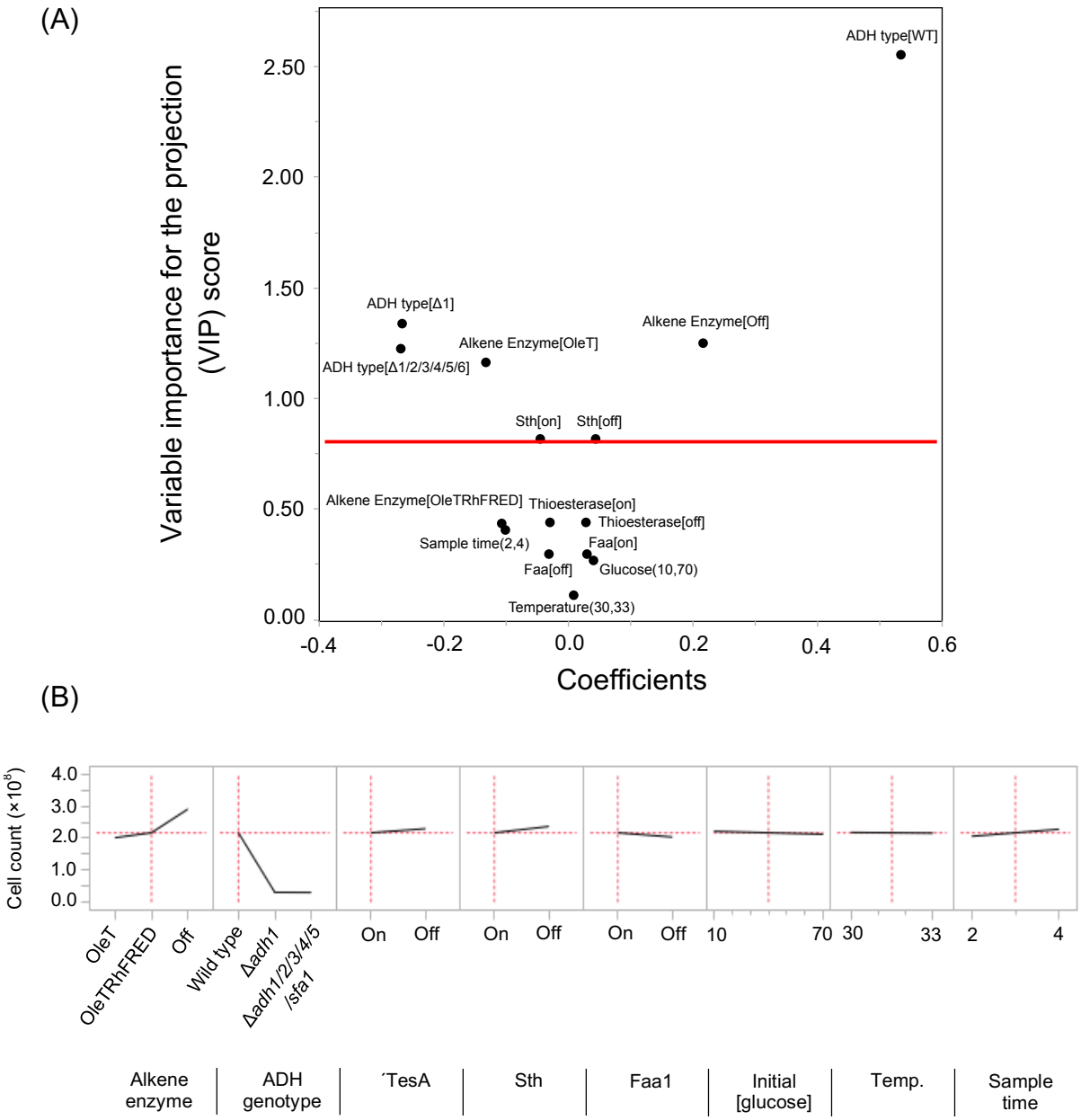


Figure 5.12 Model analysis for cell count. (A) Variable importance plot of VIP (variable importance for the projection) scores for experimental predictors against the centred and scaled data coefficients. Standardised coefficients highlight if a predictor is having a positive or negative impact, as well as the magnitude of that impact on the measured response. Predictors below the VIP value of 0.8 (red line) are unimportant to the explanatory model for cell count (Cox and Gaudard, 2013); (B) The prediction profiler of the model projection for cell count. The red dashed vertical lines highlight the selected set point for each of the experimental factors; their selection affects the horizontal red dashed line that indicates the cell count dependent upon the validated model prediction formula.

The deletion of the genes encoding the alcohol dehydrogenase isozymes of *S. cerevisiae* has been shown to negatively impact cell growth (Chapter 4; Drewke *et al.*, 1990; Ida *et al.*, 2012). The PLS model (Figure 5.12) also indicates that wild type, Δ*adh1* and Δ*adh1/2/3/4/5sfa1* genotypes are important predictors for cell count with VIPs of 2.55, 1.34, 1.22 respectively. Other important predictors for cell count include the presence of alkene enzymes. The prediction profiler for the alkene enzyme (Figure 5.12(B)) indicates that cultures achieve a higher biomass when transformed with a plasmid that does not include a version of the alkene enzyme (Alkene Enzyme [Off]; VIP = 1.25). This may be due to metabolic burden specific to the expression of the alkene enzyme or toxicity of product(s) which may, as with alkane production, interfere with the structural integrity of cell membranes (Chen *et al.*, 2013). Whilst alkenes could not be detected, the OleT enzyme is also capable of hydroxylating fatty acids to give α- and β-hydroxy fatty acids. The products of the hydroxylation of the fatty acids produced in *S. cerevisiae* were not assessed. *In vitro* the activity of the hydroxylation reaction has been measured as 6.3 % of total enzyme activity. Though it may not be the dominant activity, the products of such a reaction would prove toxic and may be hypothesised to contribute to decreased cell densities (Rude *et al.*, 2011; Liu *et al.*, 2014).

Culture viability provides an additional metric by which the impact of the experimental factors can be evaluated. The variable importance plot in Figure 5.13(A) shows the model projection of culture viability is based largely upon the initial glucose concentration predictor (VIP = 2.76). The dominant extent to which the variation in culture viability can be explained using the variation in the

initial glucose concentration is also shown in the analysis of the raw experimental data (Figure 5.14), providing confidence regarding the explanatory power of the model projection.

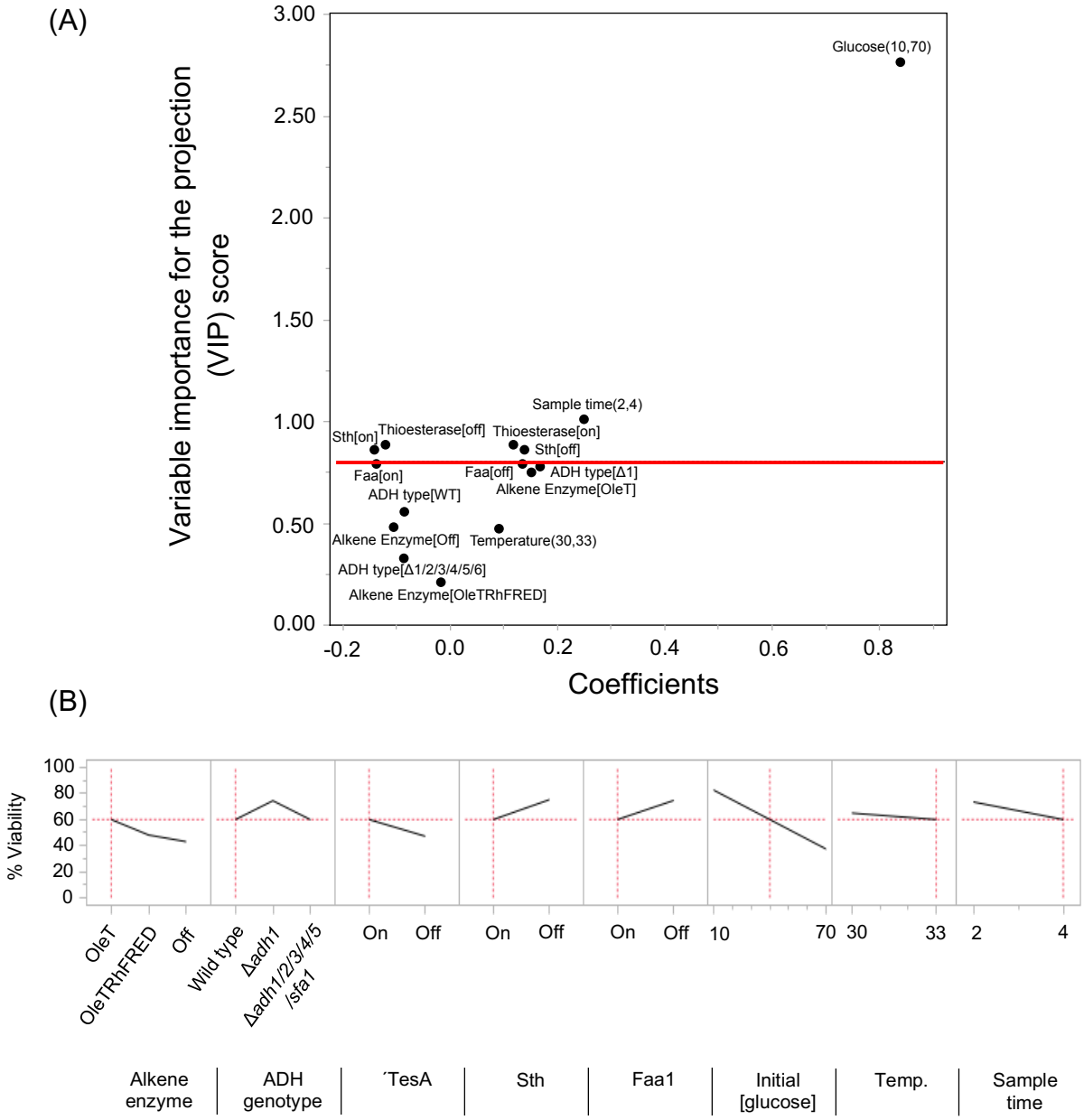


Figure 5.13 Model analysis for culture viability. (A) Variable importance plot of VIP (variable importance for the projection) scores for experimental predictors against the centred and scaled data coefficients. Standardised coefficients highlight if a predictor is having a positive or negative impact, as well as the magnitude of that impact on the measured response. Predictors below the VIP value of 0.8 (red line) are unimportant to the explanatory model for culture viability (Cox and Gaudard, 2013); (B) The prediction profiler of the model projection for culture viability. The red dashed vertical lines highlight the selected set point for each of the experimental factors; their selection affects the horizontal red dashed line that indicates the culture viability dependent upon the validated model prediction formula.

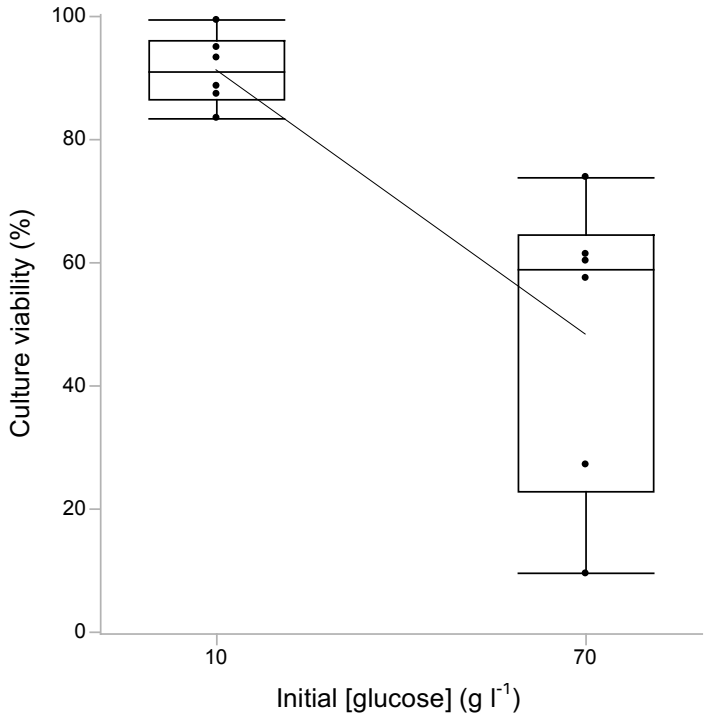


Figure 5.14 The effect of initial glucose concentration on cell viability. The boxes show the interquartile range and median values for the data and the whiskers show the highest and lowest data values. The line connects the mean value of culture viability for each of the initial glucose concentrations and highlights the negative correlation between culture viability and initial glucose concentration.

The screening model in combination with knowledge of the system provides clues as to the potential cause of the impact of initial glucose concentration on cell viability; this subsequently has the potential to dictate future areas of research and associated experiments. The reduction in culture viability associated with the higher initial glucose concentration (70 g l^{-1}) may be due to the availability of carbon which could be metabolised into inhibitory concentrations of metabolites shown to affect both cell growth and viability of *S. cerevisiae,* for example when ethanol production exceeds ~13.4 g l^{-1} or the accumulation of acetaldehyde exceeds ~0.4 g l^{-1} (assuming Neuberg's second

form of fermentation) (Brown *et al.*, 1981; Drewke *et al.*, 1990; Pons *et al.*, 1986). These potentially inhibitory metabolites were not directly measured during experimentation. Stoichiometrically greater than 18 g l^{-1} glucose is required to be converted directly to ethanol in order to achieve an inhibitory concentration. However, there was no correlation between glucose utilised (greater than 18 g l^{-1}) and a reduction in culture viability (data not shown). Additionally, perturbation of ethanol production in *S. cerevisiae* (for example, due to ADH gene deletion) has implications on the redox balance within the cell, which can be addressed by adapting metabolism to Neuberg's second form of fermentation, the equation of which is:

$$C_6H_{12}O_6 \longrightarrow C_3H_8O_3 + CH_3CHO + CO_2$$

The measurement of glycerol may provide an indirect evaluation of acetaldehyde as they are produced in equimolar amounts assuming Neuberg's second form of fermentation (toxic accumulation at concentration > 0.37g l^{-1} acetaldehyde corresponding to 0.77g l^{-1} glycerol). Figure 5.15 shows that at 70 g l^{-1} initial glucose concentration the production of glycerol improves the culture viability, which implies that associated production of acetaldehyde is not accumulating at toxic levels leading to a reduction in culture viability.

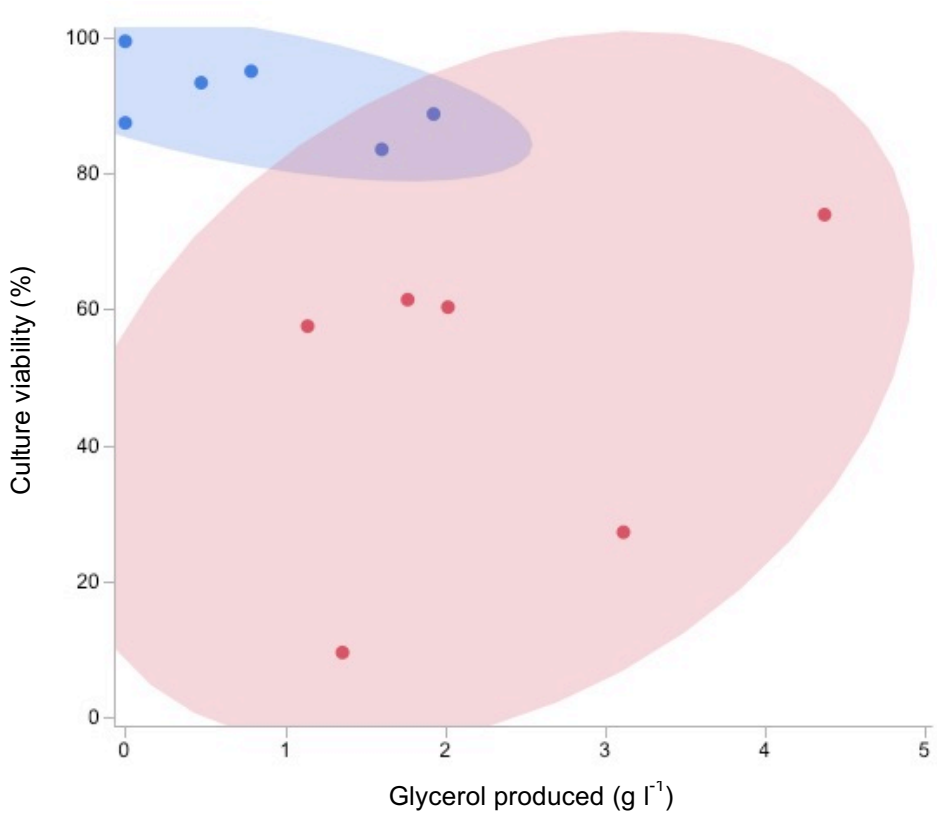


Figure 5.15 Correlation between glycerol biosynthesis and cell viability under different initial glucose concentrations. The setpoint of the initial glucose concentration for each experiment is represented by either blue data points (10 g l^{-1}) or red data points (70 g l^{-1}). The density ellipses shown (colours corresponding to the initial glucose concentration) provide the 90 % confidence curve assuming a bivariate normal distribution. As previously discussed the measurement of glycerol during experimentation was shown to be an underestimate of the true value.

Alternatively, the reduction in culture viability dependent upon the initial glucose concentration could be due to either osmoregulation, requiring glycerol production at the higher concentration of glucose (Cao *et al.*, 2007), which fits the data (Figure 5.15), or due to carbon catabolite repression, the regulation of which has been well studied in the literature providing insight into what should be measured e.g. the Snf1 protein complex, in order to evaluate these hypotheses (Gancedo, 1998).

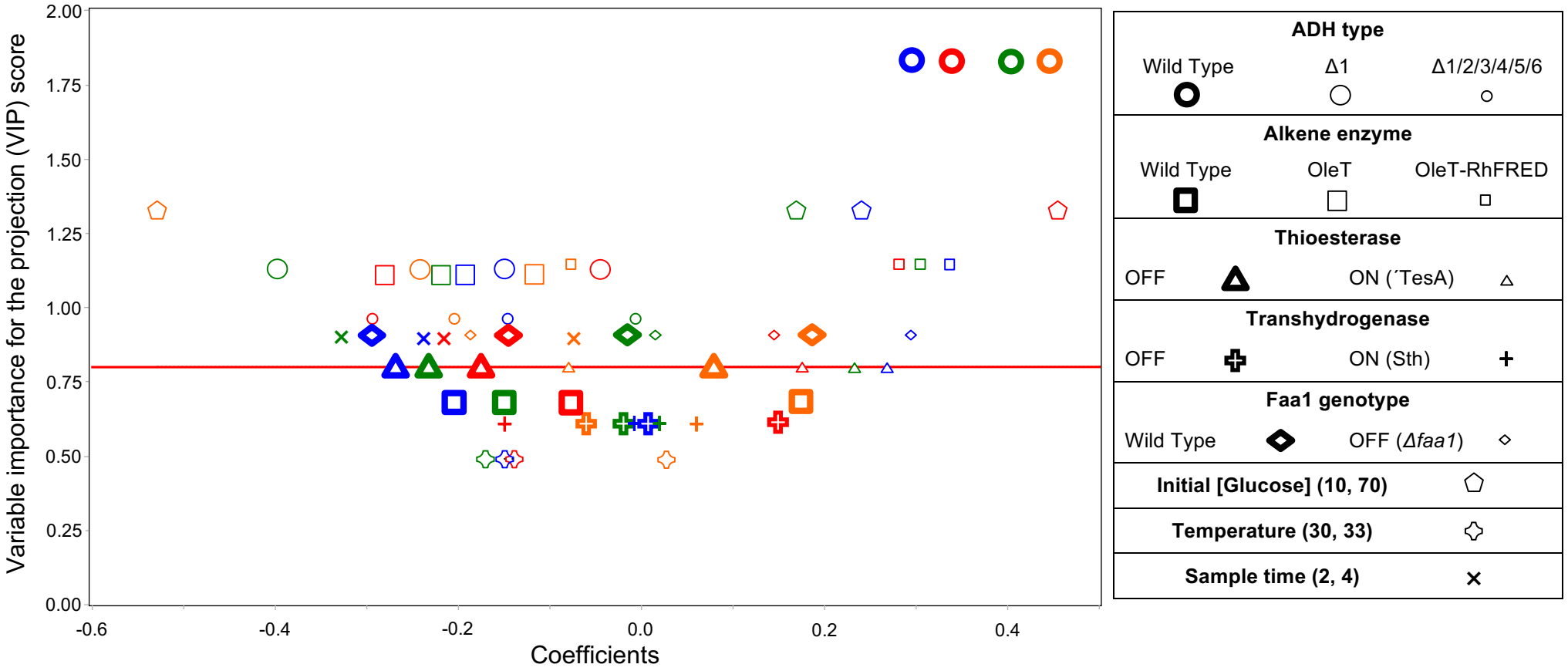


Figure 5.16 Variable importance plot of VIP (variable importance for the projection) scores for experimental predictors (represented by the corresponding symbol referenced in the legend) against the centred and scaled data coefficients for each of the measured free fatty acids produced by *S. cerevisiae* (colour coded as C_{14}, myristic acid; C_{16}, palmitic acid; C_{18}, stearic acid; $C_{18:1}$, oleic acid). Standardised coefficients highlight if a predictor is having a positive or negative impact, as well as the magnitude of that impact on the measured response. Predictors below the VIP value of 0.8 (red line) are unimportant to the explanatory model for culture viability (Cox and Gaudard, 2013).

Next, free fatty acids production (which have been shown to be differentially produced in *S. cerevisiae*) was evaluated using a single PLS model (Figure 5.16). The VIP score of a given experimental predictor remained unchanged for each of the measured free fatty acids, however the variable importance plot detailed the extent and directionality of the associated model coefficient which differed for each of the measured free fatty acids. The exploratory detail of data processing using the PLS methodology is shown in the example of the importance of the initial glucose predictor (VIP = 1.33) on free fatty acid production in *S. cerevisiae* (⬠, Figure 5.16). The effect of the C_{14}, C_{16} and C_{18} free fatty acids is different to that for $C_{18:1}$ free fatty acid. In agreement with the PLS model for fatty acid production, differential production of free fatty acids in *S. cerevisiae* has been shown, where elevated $C_{18:1}$ levels have been implicated with ethanol stress tolerance (You *et al.*, 2003). However, without the experimental data for ethanol production, the described dose-dependent production of oleic acid correlating to the ethanol production cannot be verified.

The wild type ADH genotype was shown to be the most significant predictor for the free fatty acid projection (VIP = 1.83), increasing free fatty acid production. Evaluation of the raw experimental data (Figure 5.17) shows that the modelled impact of the wild type ADH genotype on the concentration of the free fatty acids is likely to be due to the associated increase in biomass.

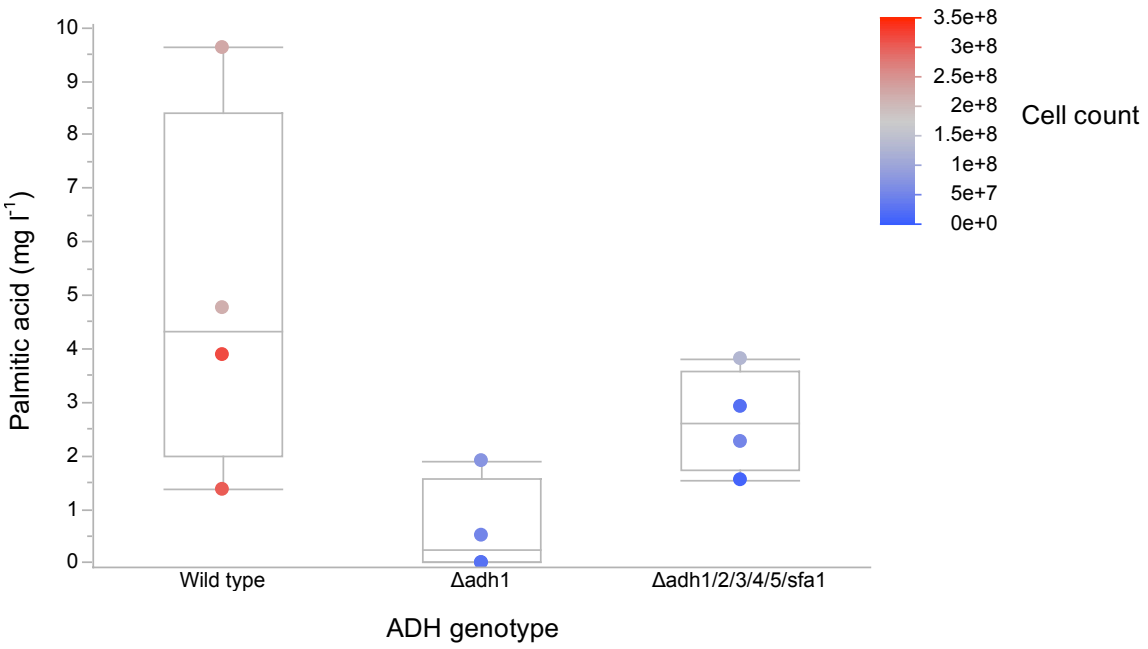


Figure 5.17 Effect of the deletion of ADH genes on fatty acid (palmitic acid) accumulation in yeast cultures including final cell count. The data points are coloured according to the cell count of the culture, ranging from low cell densities (dark blue) to high cell densities (dark red). The boxes show the interquartile range and median values for the data and the whiskers show the highest and lowest data values The data is representative for all fatty acids; the wild type ADH genotype achieves higher cell counts during experimentation and this results in the higher production of fatty acid.

The correlation between fatty acid production and biomass is intuitive, however the information does not easily reflect the impact of the metabolic engineering strategy on the distribution of carbon (an important metric for evaluation of the feasibility of an industrial process). The data for free fatty acid concentrations were therefore transformed, standardising the fatty acid production to biomass. The PLS model evaluation of the transformed measured response total free fatty acids standardised to cell count (1.0×10^8), is shown in Figure 5.18. This again shows the wild type ADH genotype to be the most significant predictor (VIP = 1.67) of fatty acid concentration, however the predictor now reduces total standardised free fatty acid production. The model prediction profiler for the ADH genotype (Figure 5.18(B)) indicates that the deletion of ADH isozymes (and associated perturbation on ethanol metabolism (Chapter 4)) provides a more efficient use of carbon substrate in terms of production of free fatty acids per unit of biomass as compared to the wild type *S. cerevisiae* strain. Reducing carbon flux to ethanol has been shown to increase glycerol production required to rebalance NAD:NADH during glycolysis according Neuberg's second form of fermentation, this additionally results in the production of acetate which can be converted into acetyl-CoA, a precursor for fatty acid production. Increased acetyl-CoA substrate requires an increased supply of the cofactor, NADPH in order to improve fatty acid production yields. The proposed mode of action of the expressed soluble transhydrogenase is to rebalance the cofactors utilised during glycolysis providing NADPH (Figure 5.5). The derived model (Figure 5.18(A)) shows that the soluble transhydrogenase (Sth) is an important predictor (VIP = 1.44), increasing the total free fatty acids accumulated. Although the current experiment allows only a modelling estimation of the main

effects on a measured response, the data highlights a potential interaction between deletion of the ADH isozymes and inclusion of the soluble transhydrogenase in these lines (Figure 5.19). Confirmation of this interaction would be achieved by augmenting the current experiment to evaluate two level interactions, whilst further biochemical investigation (including determining the ratios of the cofactors NADH and NADPH and a measurement of ethanol production) may also provide further insight.

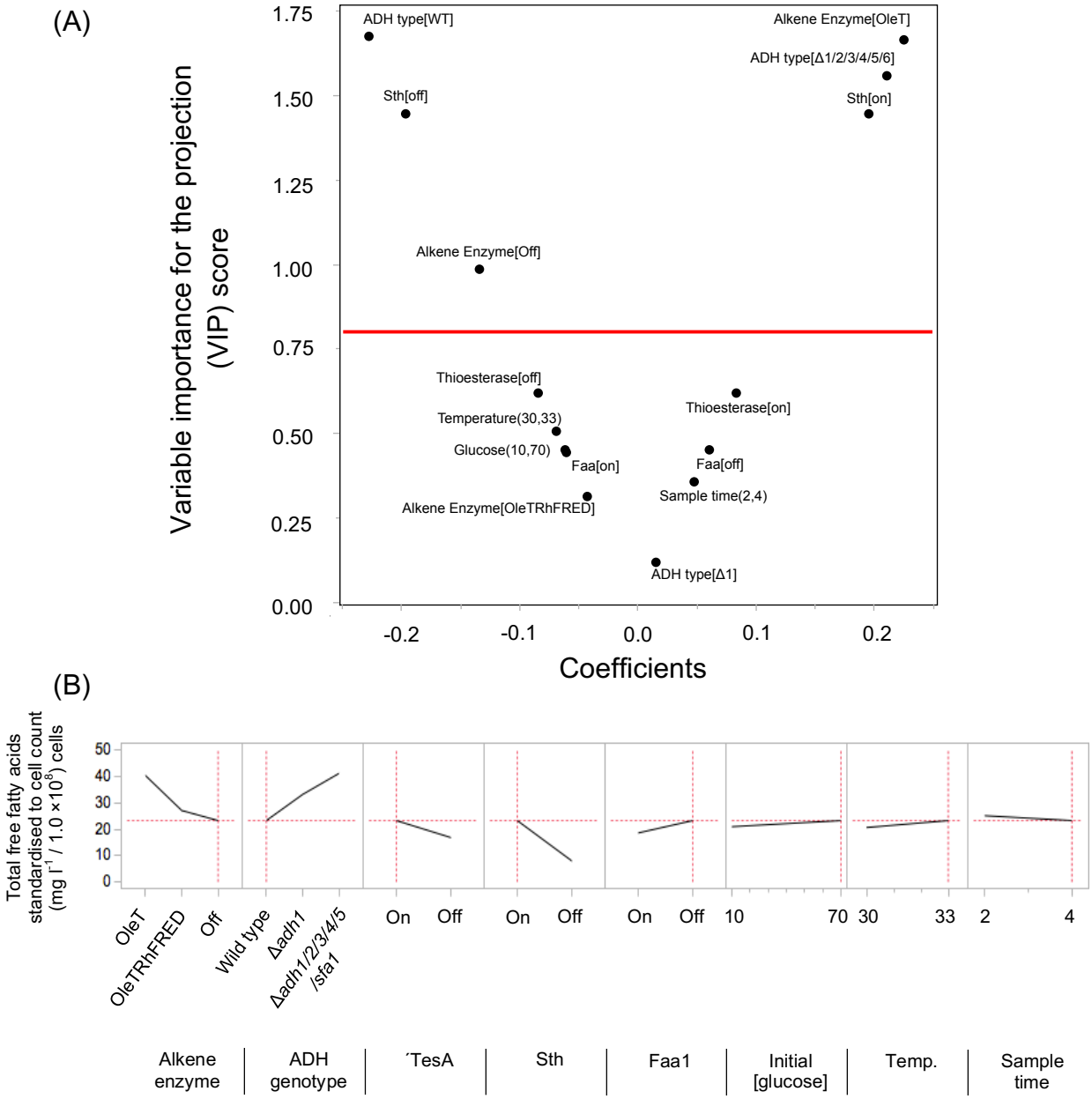


Figure 5.18 Model analysis for total free fatty acids standardised to cell count. (A) Variable importance plot of VIP (variable importance for the projection) scores for experimental predictors against the centred and scaled data coefficients. Standardised coefficients highlight if a predictor is having a positive or negative impact, as well as the magnitude of that impact on the measured response. Predictors below the VIP value of 0.8 (red line) are unimportant to the explanatory model for total free fatty acids standardised to cell count (Cox and Gaudard, 2013); (B) The prediction profiler of the model projection for total free fatty acids standardised to cell count. The red dashed vertical lines highlight the selected set point for each of the experimental factors; their selection affects the horizontal red dashed line that indicates total free fatty acids standardised to cell count dependent upon the validated model prediction formula.

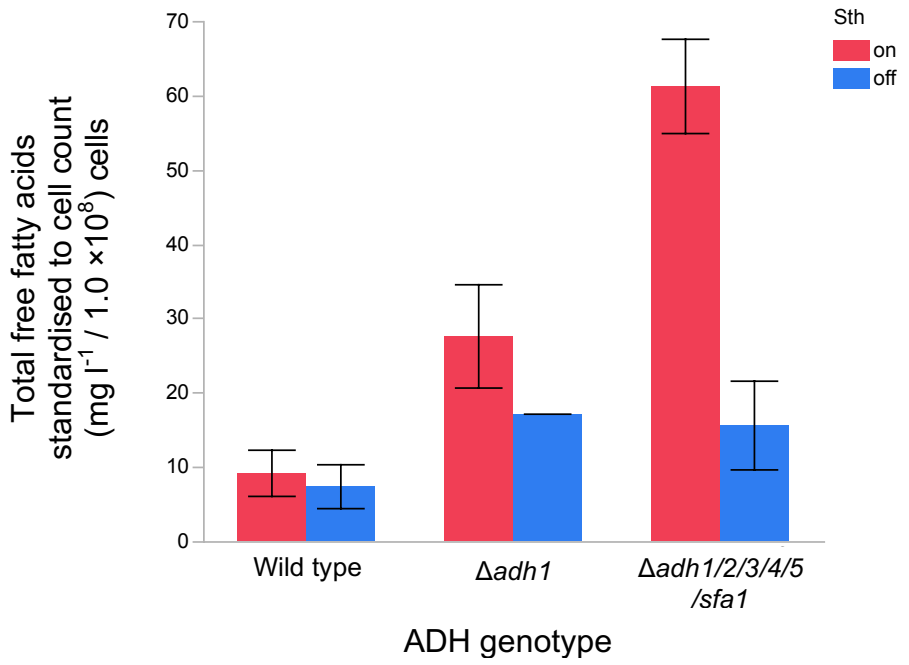


Figure 5.19 The raw experimental measurements of total free fatty acids standardised to cell count for the ADH genotypes. The coloured bars represent whether the soluble transhydrogenase (Sth) is present (in red) or absent (in blue). The data highlight a potential interaction between the ADH genotype and the Sth. The error bars represent standard error.

The aim of the metabolic engineering strategy, specifically the expression of the thioesterase (´TesA) and deletion of Faa1, was to increase carbon flux to free fatty acids. The variable importance of these predictors on the model projection indicates them to be unimportant factors for the generation of the model (VIPs = 0.62 and 0.45 respectively). However, the model prediction profiler (Figure 5.18(B)) does show that both the thioesterase and deletion of Faa1 increase total fatty acid production standardised to biomass. Previous assessment reported that sole expression of the thioesterase increased the concentration of free fatty acids from 0.6 mg l^{-1} to 4.7 mg l^{-1}, whilst the combination of both ´TesA and Δ*faa1* improved free fatty acid production to 164.3 mg l^{-1} (Runguphan and Keasling, 2014). The impact of this combination (and this two

factor interaction) cannot be evaluated using the current experimental design, as only the main effect estimates can be modelled (augmentation of experiments is required).

The inclusion of the gene encoding the OleT alkene enzyme in the construct was shown to have a significant impact (VIP = 1.66) on the model projection, increasing the amount of total free fatty acid production when standardised to biomass. This counterintuitive result, as the enzyme should utilise free fatty acids to produce alkenes, may again indicate a metabolic burden specific to expression of the alkene enzyme leading to reduced biomass formation (Figure 5.18(B) shows this also occurs, although to a lesser extent, when OleT-RhFRED is expressed compared to the absence of the alkene enzyme), or may be due to an as yet undescribed activity of the OleT enzyme increasing free fatty acid production in yeast. However, the result illustrates an advantage of using a structured DoE methodology that includes defined constraints in the design of the experiment, permitting a more reasoned interpretation of empirically derived results. The absence of the factor combination of OleT enzyme without Sth means there is no model estimate of the effect of this combination on the measured response. The associated reduction in the power of analysis for the combination of these two variables decreases the capacity to be able to distinguish between their effects, indicated by the map of correlation for the model (Figure 5.8). Therefore the effect of the OleT enzyme on the model projection could be due to the significant impact of the Sth effect, which cannot be separated due to the constrained experimental design.

Figure 5.20 details an example of linking model profilers for all measured responses permitting a dynamic trade off evaluation of the metabolic engineering strategy, with the potential to dictate future experimental direction or indeed condition set points relevant to industrial processes. The data derived model indicates that lower initial glucose concentrations and increasing the culture time prior to cell harvest benefits the process by reducing carbon flux to glycerol (1.61 to 1.29 g l^{-1}) and increasing both culture viability (48.6 to 80.3 %) as well as total free fatty acids standardised to culture cell count (31.89 to 42.79 mg l^{-1}).

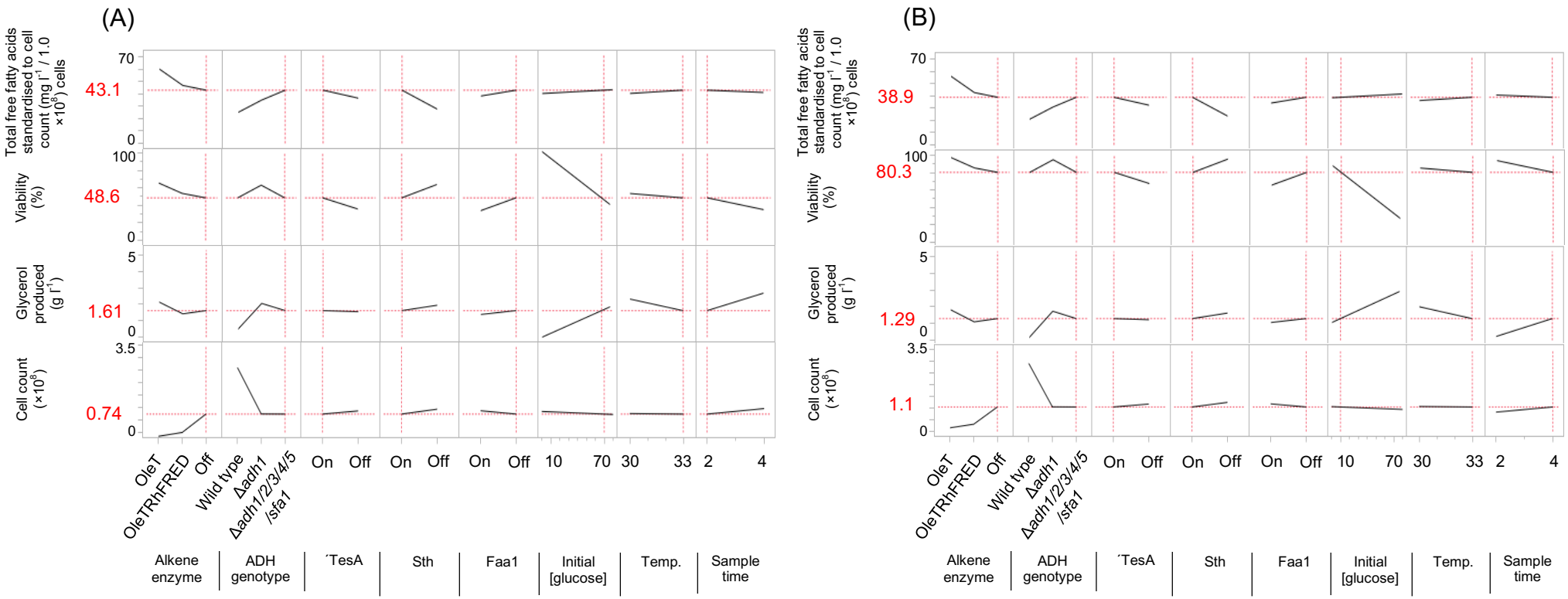


Figure 5.20 Linked prediction profile of the model projections for total free fatty acid production standardised to biomass, glycerol production, culture viability and cell count, which permits a trade off analysis of the metabolic engineering strategy. (A) Levels of the experimental factors have been set to maximise the amount of total free fatty acid production standardised to biomass (alkene enzyme set to [Off] as reduced power of analysis due to the constrained experimental design), the effect on the other measured responses is shown (red values). (B) Changing the experimental factors (initial [glucose] and sample time) benefits the process by reducing carbon flux to glycerol (1.61 to 1.29 g l^{-1}) and increasing both culture viability (48.6 to 80.3 %) as well as total free fatty acids standardised to culture cell count (31.89 to 42.79 mg l^{-1}) (red values).

Although subject to limitations, experimental design using the DoE methodology in order to efficiently model the effect of the metabolic engineering strategy on measured responses has proved useful. The 12 screening experiments (Table 18) provided empirical model estimates for the eight genotypic and environmental factors. The explanatory power of the models (particularly using the variable importance plots) has provided an assessment of the empirically derived results compared to the hypothesised strategy, however alkenes were not detected. The next section details the development of a methodology to extract and measure alkenes from yeast. The method is subsequently used to evaluate a subset of the metabolic engineering strategy factors for heterologous production of alkenes.

5.2.5 Method development for measuring alkenes in *S. cerevisiae*

No alkenes were detected during evaluation of the metabolic engineering strategy for the production of alkenes in yeast. This result poses a number of questions, including whether the OleT enzymes were being expressed correctly, however while this work was in progress a paper was published describing the production of alkenes in *S. cerevisiae* using the OleT enzyme (Chen *et al.*, 2015). Although Chen *et al*. did not characterise the OleT-RhFRED fusion enzyme they did characterise eight homologues of OleT fatty acid decarboxylases and, gratifyingly, confirmed that the codon optimised version of the OleT enzyme from *Jeotgalicoccus* sp. ATCC 8456 (evaluated in this thesis) is expressed in *S. cerevisiae* and produces the highest total alkene titre (54.5 μg l^{-1}). Therefore the successful alkene extraction method of Chen *et al*. was applied to further metabolic engineering strategy experiments designed to identify alkenes. The initial method for measuring alkenes (section 2.5.4) was adapted as follows: 50 ml of yeast cell suspension (cultured for five days) was centrifuged and re-suspended in 1.0 ml of methanol. After adding tetradecene as an internal standard [25 ng μl^{-1}], the cell suspension was disrupted using lysing matrix C tubes and the FastPrep®-24 instrument (MP Biomedicals, California, US) at a speed setting of 6.0 m s^{-1} for 1.0 min and for 8 passes. The homogenised material was transferred to 30 ml glass universals and 1.0 ml of hexane was added prior to mixing for 2.0 hours. After phase separation the upper phase was removed and stored in GC vials prior to analysis.

The variables for the experiment are shown in Table 21. The reported alkene productivity in yeast was linked to cell growth and improved by optimising

culture conditions in bioreactors that included growth on rich media (Chen *et al.*, 2015). Therefore all experiments (1-9) were performed in both the synthetic defined (SD) medium and YEPD (section 2.1.2), a rich medium. The tested strain genotypes included different perturbations of ADH isozymes, providing an assessment of alkene production related to diverting carbon away from the predominant carbon sink in *S. cerevisiae*, ethanol. Experiments 3, 6 and 9 were negative controls for alkene production in each of the defined strain genotypes (each transformed with the empty vector pO9). The concise plasmid variant pO6 was used, which contains only the OleT-RhFRED fusion gene, as well as the larger and possible unstable pO1 plasmid which contains three expressed genes. Experiment 7 evaluates the pO4 vector that contains the codon optimised OleT gene without RhFRED (previously characterised by Chen *et al.*, 2015), the thioesterase and the transhydrogenase, transformed into the Δ*adh1/faa1* strain genotype.

Table 21. Variables for experiments evaluating the production of alkenes in different *S. cerevisiae* genotypes. Experiments 1 to 9 were individually evaluated in defined (SD broth) and complex (YEPD) media.

Exp.	Strain genotype	Alkene enzyme	Thioesterase	Transhydrogenase	Plasmid variant
1	Wild type	OleT-RhFRED	ON (´TesA)	ON (Sth)	pO1
2	Wild type	OleT-RhFRED	OFF	OFF	pO6
3	Wild type	OFF	OFF	OFF	pO9
4	Δ*adh1/faa1*	OleT-RhFRED	ON (´TesA)	ON (Sth)	pO1
5	Δ*adh1/faa1*	OleT-RhFRED	OFF	OFF	pO6
6	Δ*adh1/faa1*	OFF	OFF	OFF	pO9
7	Δ*adh1/faa1*	OleT	ON (´TesA)	ON (Sth)	pO4
8	Δ*adh1/2/3/4/5/sfa1/faa1*	OleT-RhFRED	OFF	OFF	pO6
9	Δ*adh1/2/3/4/5/sfa1/faa1*	OFF	OFF	OFF	pO9

The experiment was performed in 18 × 250 ml volume shake flasks; 50 ml cultures were incubated at 30 °C with 200 rpm shaking for 5 days prior to harvest. Samples were analysed by gas chromatography with a flame ionisation detector (GC-FID) to simultaneously detect and quantify alkene and free fatty acids. The aim of the experiment was to review if the amended alkene extraction method in combination with the strategy described above would enable detection of alkenes in *S. cerevisiae*.

GC-FID chromatography demonstrated that C_{17} alkenes (heptadecene) were present in culture samples and that in the process blank controls no alkenes were detected (i.e. no contaminating alkenes from the culture and extraction protocol) (Figure 5.21). However in the biological negative controls a peak at a similar retention time (10.101 – 10.104 min) to the C_{17} alkene standard (10.097 min) was present. Only C_{17} alkenes were measured in the biological samples (Table 22) despite the presence of C_{14}, C_{16} and C_{18} free fatty acids by *S. cerevisiae* (previous section).

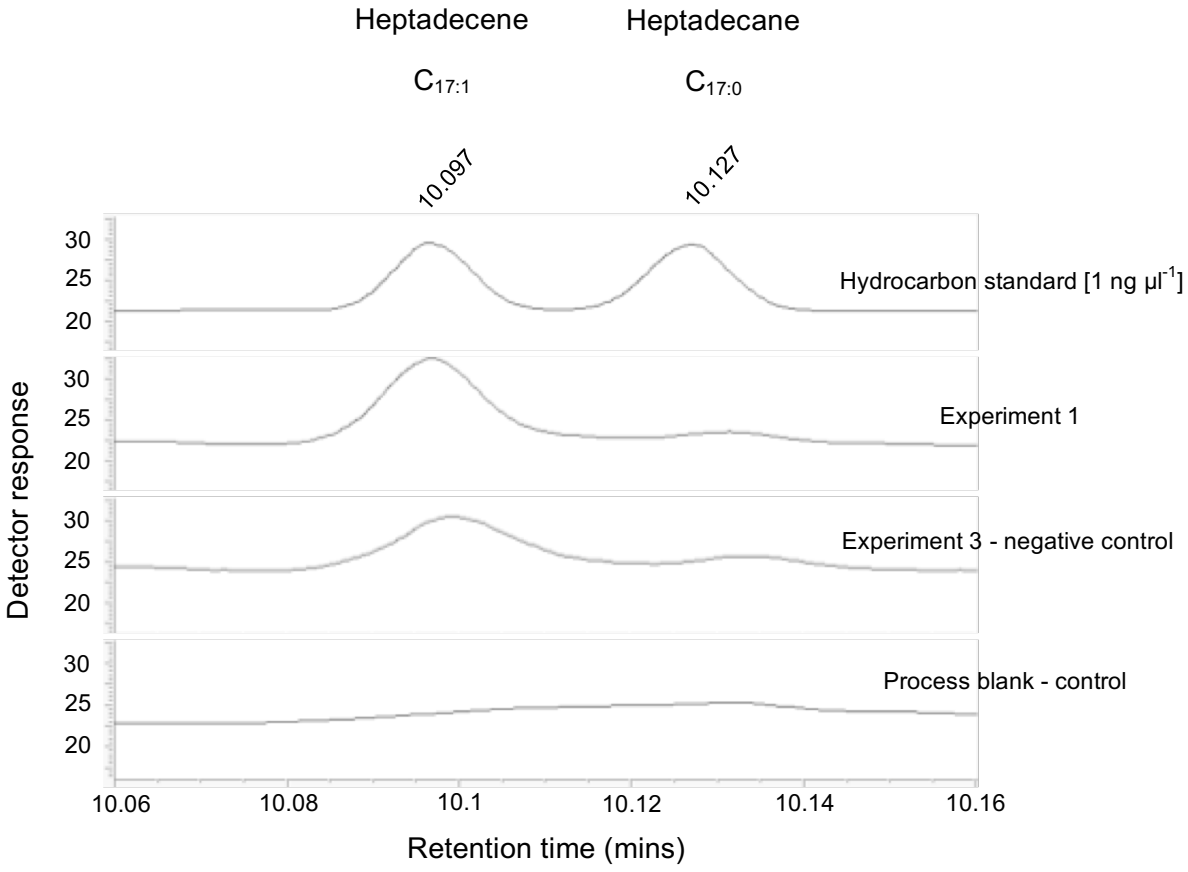


Figure 5.21 Example GC-FID analysis for detection of C_{17} alkene produced in *S. cerevisiae.* GC-FID profiles showing: hydrocarbon standard [1 ng μl^{-1}] (heptadecene at retention time 10.097 min and heptadecane at retention time 10.127 min); extraction sample from Experiment 1 cultured in YEPD, (wild type *S. cerevisiae* transformed with pO1 plasmid containing OleT-RhFRED, thioesterase and transhydrogenase genes); extraction sample from Experiment 3 cultured in YEPD, (wild type *S. cerevisiae* transformed with an empty plasmid); and the process blank control.

Table 22. GC-FID quantification data for alkene production in *S. cerevisiae*. Data in red are for detected peaks at retention time of 10.101 – 10.104 min in the negative controls. "Trace" indicates a peak present at the retention time of interest but which falls below the lower bound of the linear calibration curve for that alkene [< 0.56 ng μl^{-1}].

Exp.	Media type	OD_{600nm}	Alkenes (ng μl^{-1})			Total alkene standardised to OD (ng μl^{-1} / 10 OD_{600nm})
			C_{13}	C_{15}	C_{17}	
1	YEPD	25.0	0	0	2.46	0.98
	SD	7.7	0	0	Trace	Trace
2	YEPD	24.7	0	0	1.67	0.68
	SD	8.0	0	0	Trace	Trace
3	YEPD	31.9	0	0	1.63	0.51
	SD	9.0	0	0	Trace	Trace
4	YEPD	5.2	0	0	2.39	4.60
	SD	2.5	0	0	Trace	Trace
5	YEPD	4.9	0	0	0.84	1.71
	SD	2.6	0	0	0	0
6	YEPD	8.7	0	0	0.61	0.70
	SD	5.3	0	0	Trace	Trace
7	YEPD	8.9	0	0	8.28	9.30
	SD	3.2	0	0	5.52	17.25
8	YEPD	5.3	0	0	Trace	Trace
	SD	2.9	0	0	Trace	Trace
9	YEPD	9.3	0	0	0	0
	SD	4.8	0	0	Trace	Trace

To confirm the identity of the GC-FID peaks, further investigation of extraction samples was performed using a quadrupole time-of-flight gas chromatography with mass spectrophotometer (GC/Q-TOF/MS). The GC/Q-TOF/MS provides high resolution, accurate mass data and permits the separation of target compounds from background interferences (improving the signal to noise ratio). The further analysis also aimed to identify the compound in the negative controls (represented by the GC-FID peak at a retention time of 10.101 - 10.104 mins) and to see if other chain lengths of alkenes could be detected, as reported in the literature (Chen *et al.*, 2015).

The retention times of the hydrocarbon standards are different between the two analytical methods; the example GC/Q-TOF/MS total ion chromatogram (TIC) (Figure 5.22), gives a retention time of 10.515 min for heptadecene. The peak in the negative control sample is now distinguishable from both the alkene peak in the standard and the corresponding peak in the biological sample, Experiment 1. Figure 5.22(B) shows the electron ionisation mass spectra at a retention time 10.515 min, corresponding to each TIC of the samples analysed in Figure 5.22(A). The distribution patterns of ions by mass (mass-to-charge ratio, *m/z*) confirm the presence of heptadecene in the biological sample, Experiment 1, but show there is no heptadecene in the negative or process controls. To improve the signal to noise ratio, subsequent analysis of alkenes in all samples was performed using GC/Q-TOF/MS extracted ion chromatograms (EIC), where the most abundant mass ion for all alkene mass spectra (m/z = 55.0543 ± 50 ppm) was used to quantify peaks, and ratios of the diagnostic qualifier mass ions (m/z = 55.0543 ± 50 ppm, 69.0700 ± 50 ppm and 83.0856 ±

50 ppm) characterised peaks at retention times corresponding to the alkene response from the hydrocarbon standard. The quantification of detected alkenes is shown in Table 23.

The production of alkenes has been ascertained using the developed extraction method and confirmed using GC/Q-TOF/MS. C_{17} heptadecene was detected in biological samples along with trace amounts of the other expected alkenes C_{13} and C_{15}. No alkenes were measured unless the OleT or OleT-RhFRED enzyme was expressed in yeast. The chromatographic peak seen in the negative control that had confused the data interpretation was further analysed. Deconvoultion of the chromatography indicated that the peak represented at a retention time of 10.495 min represented a compound with a mass of 85.0282, and was not an alkene. MassHunter qualitative software identified the likely compound formula to be $C_4H_5O_2$.

Figure 5.22 Example GC/Q-TOF/MS analysis for detection of C_{17} alkene produced in *S. cerevisiae.* (A) total ion chromatograms (TICs) showing: hydrocarbon standard [10 ng μl^{-1}] (heptadecene at retention time 10.515 mins and heptadecane at retention time 10.567 mins); extraction sample from Experiment 1 cultured in YEPD, (wild type *S. cerevisiae* transformed with pO1 plasmid containing OleT-RhFRED, thioesterase and transhydrogenase genes); extraction sample from Experiment 3 cultured in YEPD, (wild type *S. cerevisiae* transformed with an empty plasmid); and the process blank control. (B) electron ionisation mass spectra at a retention time 10.515 min (the retention time of heptadecene in the hydrocarbon standard) for each of the TICs in (A).

The data show that heptadecene is present in the biological sample and absent from both the negative and process blank controls.

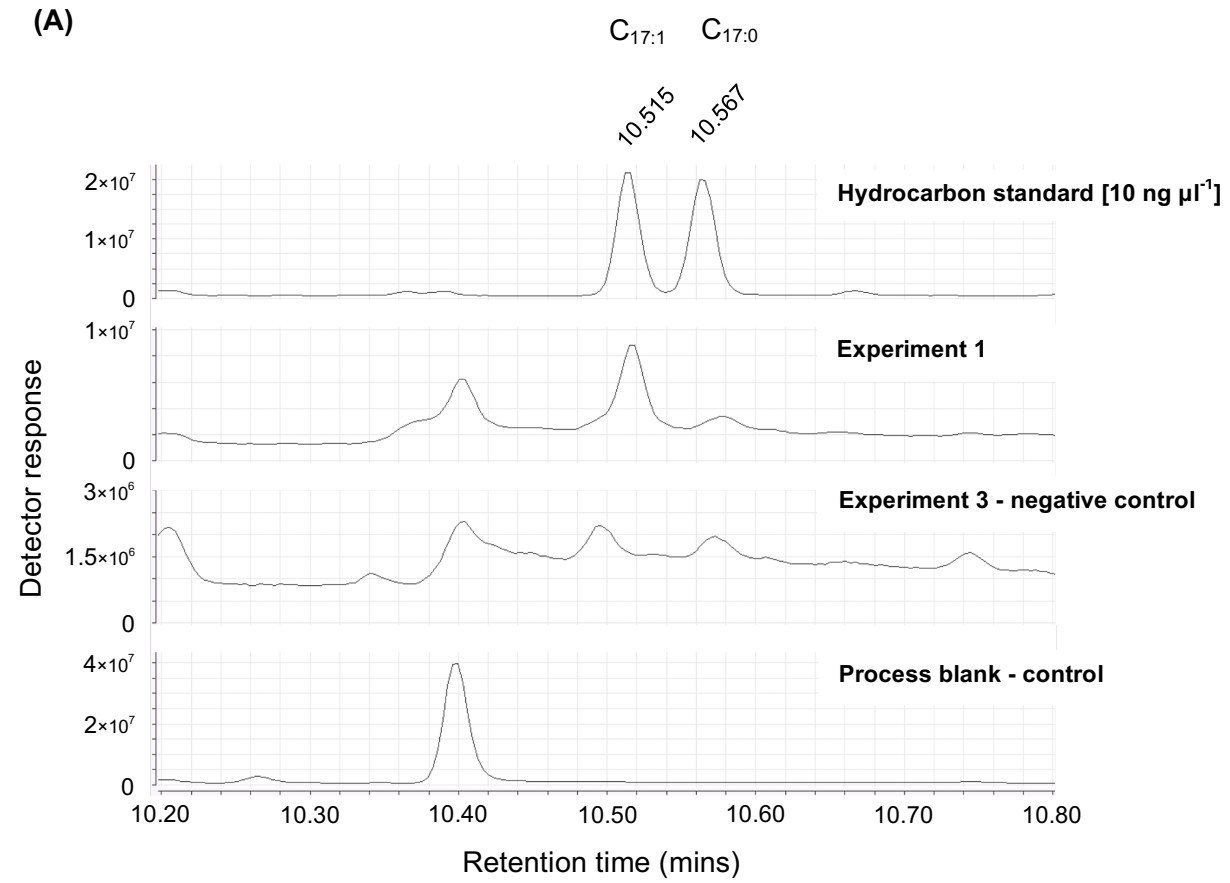

(A)
C17:1
C17:0
10.515
10.567
Hydrocarbon standard [10 ng µl-1]
Experiment 1
Experiment 3 - negative control
Process blank - control
Detector response
Retention time (mins)
(B)
Hydrocarbon standard [10 ng µl-1]
Experiment 1
Experiment 3 - negative control
Process blank - control
Counts
Mass-to-charge (m/z)

Table 23. GC/Q-TOF/MS quantification data for alkene production in *S. cerevisiae*. "Trace" indicates an alkene peak that falls below the lower bound of the linear calibration curve for that alkene [< 0.56 ng μl^{-1}]. No alkenes were detected in the negative control experiments 3, 6 and 9.

Exp.	Strain genotype	Alkene enzyme	Thioesterase	Trans-hydrogenase	Media type	Alkenes (ng μl^{-1})		
						C_{13}	C_{15}	C_{17}
1	Wild type	OleT-RhFRED	ON (´TesA)	ON (Sth)	YEPD	Trace	Trace	3.65
					SD	0	0	Trace
2	Wild type	OleT-RhFRED	OFF	OFF	YEPD	0	0	Trace
					SD	0	0	0
4	Δ*adh1/faa1*	OleT-RhFRED	ON (´TesA)	ON (Sth)	YEPD	0	Trace	3.81
					SD	0	0	Trace
5	Δ*adh1/faa1*	OleT-RhFRED	OFF	OFF	YEPD	0	0	0.57
					SD	0	0	0.21
7	Δ*adh1/faa1*	OleT	ON (´TesA)	ON (Sth)	YEPD	0	Trace	14.78
					SD	0	Trace	10.77
8	Δ*adh1/2/3/4/5/ sfa1/faa1*	OleT-RhFRED	OFF	OFF	YEPD	0	0	0.91
					SD	0	0	Trace

Total concentration of alkenes produced by the OleT enzyme construct is comparable to the amount reported in the literature using the same enzyme, prior to process optimisation (Chen *et al.*, 2015). The OleT enzyme construct is associated with a 3.8 fold increase in the production of total alkenes as compared to the OleT-RhFRED construct. The impact of the ADH genotype (varying carbon flux to ethanol) on alkene production is difficult to assess, as the experimental design (Table 21), which is not based on DoE, evaluates the OleT enzyme only in the Δ*adh1* genotype; this remains an interesting combination to research. Moreover application of the DoE methodology for experimental design, including evaluation of environmental factors, would enable efficient process optimisation for the heterologous production of alkenes in yeast. Process optimisation and the associated understanding of how to culture a specific microorganism for heterologous production of chemicals has been shown to have a large impact on product yield, for example, media development, feeding strategies and culture harvest time optimisation resulted in an incremental 67.4 fold increase in total alkene production (Chen *et al.*, 2015). It is important to note that subsequent optimisation would be performed on a yeast strain that is amenable to industrial scale manufacture.

5.3 Conclusion

A Design of Experiments methodology utilising the PLS modelling algorithm and a developed alkene extraction method have together provided the proof of principle for a designed metabolic engineering strategy, which converts free fatty acid precursors into alkenes using the industrially relevant *S. cerevisiae* prototrophic strain CEN.PK113-7D. An optimised experiment design resulted in model estimates of the impact of each of the factors of the metabolic engineering strategy on the production of free fatty acids, the substrate for alkene production using OleT. The explanatory power of the models has provided a dynamic trade-off evaluation of the overall strategy, with the potential to dictate condition set points relevant to industrial process; however alkenes were not measured. The successful development of an alkene extraction protocol now can be combined with the optimised design of experiments and PLS modelling algorithm with the aim of providing a dynamic trade-off evaluation that includes the production of alkenes. The production of alkenes in *S. cerevisiae* by heterologous expression of the OleT enzyme from *Jeotgalicoccus* sp. has previously been reported (Chen *et al.*, 2015), however this has now been expressed in a strain amenable to industrial bioprocess and includes a preliminary assessment of the enzyme fused to a reductase domain (OleT-RhFRED), which may be more relevant to large scale manufacture.

6 General conclusion

Global energy production and chemical manufacturing using natural systems may facilitate a move away from humanity's dependence upon petrochemicals and fuels. The yeast *Saccharomyces cerevisiae* is a prominent cell factory used in current large-scale commercial biorefineries; increasing the product portfolio of this yeast biocatalyst has been a focus of this research and has been demonstrated by evaluation of a metabolic engineering strategy designed to produce alkenes. To aid future development of the strategy within a biorefinery concept, evaluation of alkene production was performed in a prototrophic, industrially relevant strain of *S. cerevisiae*.

Characteristically, ethanol production is a major carbon sink that affects the efficiency of product synthesis, such as alkenes, in *S. cerevisiae* (Ida *et al.*, 2013; Abbott *et al.*, 2009). The yeast ADH isozymes are well studied, however the current literature does not fully characterise the impact of genotype-by-genotype and genotype-by-environment interactions on ethanol metabolism. A genetic engineering strategy involving deletion of ADH isozymes was used to decouple the predominant flux of carbon toward ethanol production. Compared to other gene deletion methods for yeast such as the Cre-*loxP* site-specific recombination system of bacteriophage P1 (Sauer, 1987), the dominant and counter-selectable *amdSYM* method was shown to have the advantages of being applicable for rapid, multiple and scarless gene deletions in a prototrophic *S. cerevisiae* strain. The method was used for the successful production of the 64 strain genotypes required for a complete combinatorial ADH knockout library.

Design-based engineering and evaluation of complex biological systems, such as the interactions of the ADH isozymes within different environments, has many associated challenges. These include optimising experimental design and visualisation of multivariate data leading to appropriate conclusions; limitations of these have the potential to negatively impact scientific replicability. The Design of Experiments (DoE) methodology was implemented to evaluate the ADH knockout library as a potential solution to some of these challenges. This DoE methodology is appraised in this thesis as essential for detailed empirical evaluation of complex systems and as a method that would enhance the aim of generating a genome scale model of *S. cerevisiae*.

The specific DoE method detailed includes data analysis using the multivariate PLS (Partial Least Squares) modelling platform, a powerful statistical analysis approach made readily accessible to the biologist using the commercially available JMP software package. The model of carbon flux to ethanol shown in Chapter 4 details 60 different genotype-by-genotype and genotype-by-environment interactions important to the model projection. The model highlights both the complexity of the single reaction step catalysed by the six ADH isozymes in yeast and the power of the multivariate DoE methodology for understanding natural systems. The model furthers the current knowledge of the ADH isozymes in *S. cerevisiae* providing empirical evidence that in *S. cerevisiae*:

- *ADH2* is not associated with producing ethanol under anaerobic culture conditions in combination with 25 g l^{-1} glucose substrate concentrations;

- *ADH4* is associated with increased ethanol production when the cell is confronted with a zinc-limited environment;
- *ADH5* is linked with the production of ethanol, predominantly at pH 4.5.

Moreover, the efficiency of DoE experiments in providing statistically relevant data contributes to the power of the DoE approach; 88 experiments (a significant reduction compared to the full factorial 4096 experiment combinations) were able to provide data for a model of carbon flux to ethanol that permits detailed understanding of the ADH isozymes within different environments.

The trade-off evaluation of the industrially relevant CEN.PK113-7D *S. cerevisiae* ADH knockout library provided a dynamic assessment of the impact of condition set points on carbon distribution. The analysis demonstrated the potential of the DoE approach to evaluate the production of other biologically-derived chemicals or fuels, particularly using free fatty acids as a substrate, further developing the industrial use of the yeast *S. cerevisiae*. The future direction of this research should include DoE methodology applied to the evaluation of combining the CEN.PK113-7D *S. cerevisiae* ADH knockout library with the metabolic engineering strategy shown to produce alkenes. Additionally, using biomass feedstocks (such as bagasse) and other associated industrial bioprocess parameters would evaluate alkene production appropriate to industrial implementation within a biorefinery.

Other future research aims could include the following:

- The current model of carbon flux to ethanol evaluates main effects and 2-level factor interactions. It is intriguing to ascertain how complex the single reaction step catalysed by the six ADH isozymes in yeast may be. Performing additional experiments to augment the current model would permit the evaluation of higher-level factor interactions on ethanol production.
- Development of accurate assays for the quantification of other responses including NAD:NADH and NADP:NADPH cofactor ratios, ADH gene expression levels and ADH activity of cell extracts. These could provide further detail of yeast metabolism involving the ADH isozymes.
- The ADH isozymes have been shown to catalyse the reassimilation of ethanol as well as being involved in fusel alcohol production (de Smidt *et al.*, 2012; Hazelwood *et al.*, 2008). Evaluation of culture conditions corresponding to these catalytic functions would provide a unifying model representative of all the currently known metabolic functions of the ADH isozymes in *S. cerevisiae*. Indeed, if successful, this would enhance the application of the DoE methodology to contribute toward a genome scale model of yeast applicable to the dynamic evaluation of changing environments.
- Quantification of ADH activity would provide a numerical factor to model ethanol production, as opposed to the categorical factors of the ADH gene (either wild type or deleted) used for the current model evaluation. Modelling a numerical factor has the potential to develop a knockdown model of ADH activity in yeast, which may permit fine-tuning and optimisation of the metabolic pathway.

In conclusion, the Design of Experiments approach is beneficial to the rigour of scientific method required to evaluate complex biological systems, and its use would be advantageous for all such multivariate research areas.

Appendices

Appendix A: Experimental data from the evaluation of the ADH gene deletion library at standardised laboratory growth conditions using the high-throughput 96 well plate equipment format.

ADH genotype (Wild Type or deleted)						Harvest OD	Maximum specific growth rate μmax	Ethanol produced
ADH1	ADH2	ADH3	ADH4	ADH5	SFA1	OD_{600nm}	(h^{-1})	(g l^{-1})
WT	WT	WT	Δ	Δ	Δ	1.60	0.35	5.80
WT	Δ	Δ	WT	WT	Δ	1.56	0.36	6.66
WT	WT	Δ	Δ	WT	Δ	1.59	0.35	7.09
WT	WT	Δ	Δ	Δ	WT	1.61	0.36	5.92
WT	WT	WT	WT	Δ	Δ	1.62	0.36	5.77
WT	WT	WT	Δ	WT	WT	1.68	0.36	4.93
WT	WT	WT	WT	WT	Δ	1.64	0.36	6.08
WT	WT	Δ	WT	WT	WT	1.64	0.36	6.86
WT	Δ	Δ	Δ	Δ	Δ	1.61	0.36	6.12
WT	Δ	WT	WT	WT	WT	1.67	0.36	5.57
WT	Δ	WT	Δ	WT	Δ	1.68	0.37	6.15
WT	WT	WT	WT	Δ	WT	1.68	0.37	5.89
WT	WT	Δ	Δ	WT	WT	1.68	0.38	5.51
WT	Δ	Δ	WT	Δ	WT	1.70	0.37	5.67
WT	WT	Δ	WT	Δ	WT	1.74	0.38	5.53
WT	WT	WT	Δ	Δ	WT	1.71	0.36	6.36
WT	WT	Δ	WT	Δ	Δ	1.65	0.35	6.95
WT	Δ	Δ	Δ	WT	WT	1.59	0.34	3.80
WT	Δ	WT	Δ	WT	WT	1.53	0.35	5.33
WT	Δ	Δ	Δ	Δ	WT	1.43	0.35	5.27
WT	Δ	WT	Δ	WT	Δ	1.65	0.35	4.44
WT	Δ	Δ	WT	Δ	Δ	1.59	0.35	3.50
WT	Δ	WT	Δ	Δ	WT	1.60	0.35	5.15
WT	Δ	WT	WT	Δ	Δ	1.64	0.36	5.89
WT	WT	WT	Δ	WT	Δ	1.62	0.35	6.72
WT	WT	WT	WT	WT	WT	1.66	0.37	6.05
WT	WT	Δ	WT	Δ	WT	1.66	0.36	5.50
WT	Δ	WT	WT	WT	Δ	1.65	0.37	6.59
WT	Δ	Δ	Δ	WT	Δ	1.58	0.34	6.47
WT	WT	Δ	Δ	Δ	Δ	1.65	0.37	6.08
WT	Δ	WT	Δ	Δ	Δ	1.66	0.37	6.38
WT	WT	Δ	WT	WT	Δ	1.71	0.37	6.13
WT	Δ	Δ	WT	WT	WT	1.66	0.36	5.35
WT	Δ	WT	WT	Δ	WT	1.70	0.36	5.26

ADH genotype (Wild Type or deleted)						Harvest OD	Maximum specific growth rate μmax	Ethanol produced
ADH1	ADH2	ADH3	ADH4	ADH5	SFA1	OD_{600nm}	(h^{-1})	(g l^{-1})
Δ	Δ	Δ	WT	WT	WT	0.52	0.09	0
Δ	WT	Δ	Δ	WT	Δ	0.49	0.10	0
Δ	WT	Δ	Δ	Δ	Δ	0.56	0.09	0
Δ	WT	WT	Δ	WT	Δ	0.42	0.09	0
Δ	WT	Δ	Δ	WT	WT	0.50	0.09	0
Δ	WT	WT	WT	WT	WT	0.52	0.09	0
Δ	Δ	WT	WT	Δ	WT	0.78	0.11	0
Δ	WT	WT	Δ	Δ	WT	0.58	0.10	0
Δ	WT	Δ	WT	WT	Δ	0.50	0.12	0
Δ	Δ	WT	WT	WT	Δ	0.59	0.12	0
Δ	WT	WT	WT	Δ	Δ	0.59	0.09	0
Δ	WT	Δ	WT	Δ	WT	0.65	0.10	0
Δ	WT	WT	Δ	WT	WT	0.53	0.11	0
Δ	Δ	Δ	WT	Δ	Δ	0.53	0.10	0
Δ	WT	WT	Δ	Δ	Δ	0.60	0.09	0
Δ	WT	Δ	Δ	Δ	WT	0.54	0.08	0
Δ	WT	Δ	WT	WT	WT	0.69	0.09	0
Δ	Δ	WT	Δ	WT	WT	0.44	0.09	0
Δ	WT	Δ	WT	Δ	Δ	0.58	0.10	0
Δ	Δ	Δ	WT	Δ	WT	0.63	0.10	0
Δ	WT	WT	WT	WT	Δ	0.47	0.09	0
Δ	Δ	Δ	Δ	Δ	Δ	0.42	0.10	0
Δ	Δ	WT	Δ	WT	Δ	0.44	0.08	0
Δ	Δ	WT	Δ	Δ	Δ	0.59	0.09	0
Δ	Δ	Δ	WT	WT	Δ	0.59	0.10	0
Δ	WT	Δ	WT	WT	WT	0.76	0.11	0
Δ	Δ	Δ	Δ	WT	WT	0.46	0.10	0
Δ	Δ	Δ	WT	Δ	Δ	0.52	0.10	0
Δ	Δ	Δ	Δ	WT	Δ	0.49	0.09	0
Δ	WT	WT	WT	Δ	WT	0.65	0.10	0
Δ	Δ	WT	WT	WT	WT	0.62	0.09	0
Δ	Δ	WT	Δ	Δ	WT	0.54	0.09	0
Δ	Δ	WT	WT	Δ	Δ	0.57	0.09	0
Δ	Δ	Δ	Δ	Δ	WT	0.49	0.08	0

Appendix B: Raw data evaluating the ADH gene deletion library using the ministat bioprocess equipment including the effect of the environmental factor, aeration, on ethanol production.

ADH genotype (Wild Type or deleted)									
ADH1	ADH2	ADH3	ADH4	ADH5	SFA1	Aeration	Cell count at culture harvest ($\times 10^8$)	Glycerol produced ($g\ l^{-1}$)	Ethanol produced ($g\ l^{-1}$)
Δ	WT	WT	Δ	Δ	WT	Air	2.67	2.85	1.37
WT	Δ	Δ	Δ	Δ	Δ	Air	2.44	0.65	1.09
Δ	Δ	Δ	WT	Δ	WT	Air	1.57	2.88	0.00
Δ	Δ	WT	WT	Δ	Δ	Air	2.82	3.78	0.80
Δ	Δ	Δ	WT	WT	Δ	Nitrogen	0.00	0.26	0.00
Δ	WT	WT	WT	WT	Δ	Nitrogen	0.11	2.58	0.43
WT	Δ	Δ	WT	WT	Δ	Air	2.31	0.35	1.19
WT	Δ	WT	Δ	WT	WT	Air	2.93	0.90	1.15
Δ	Δ	WT	Δ	WT	Δ	Air	1.69	4.78	0.00
WT	Δ	Δ	WT	Δ	Δ	Nitrogen	1.12	0.89	2.12
WT	WT	WT	Δ	WT	Δ	Air	2.00	0.32	1.36
WT	Δ	WT	Δ	Δ	Δ	Nitrogen	1.42	0.97	2.41
WT	Δ	Δ	WT	WT	WT	Nitrogen	1.17	0.61	1.68
WT	WT	WT	WT	Δ	Δ	Air	2.40	0.35	1.40
WT	WT	Δ	WT	Δ	WT	Air	1.65	0.18	0.37
Δ	WT	WT	WT	Δ	WT	Nitrogen	0.00	0.52	0.00
Δ	WT	Δ	WT	Δ	Δ	Air	2.76	2.49	0.42
Δ	WT	WT	WT	WT	WT	Air	1.70	0.73	0.39
WT	Δ	WT	WT	Δ	WT	Air	2.61	0.19	1.67
WT	WT	WT	Δ	WT	WT	Nitrogen	1.77	0.94	2.33
Δ	WT	Δ	Δ	WT	WT	Nitrogen	0.00	0.81	0.00
Δ	Δ	Δ	Δ	WT	WT	Air	1.82	2.50	0.00
Δ	WT	WT	Δ	Δ	Δ	Nitrogen	0.71	1.67	2.59
Δ	Δ	Δ	Δ	Δ	Δ	Nitrogen	0.00	0.27	0.00
WT	WT	Δ	Δ	Δ	Δ	Nitrogen	1.31	0.93	2.30
WT	Δ	Δ	Δ	WT	Δ	Nitrogen	1.04	0.83	4.18
WT	WT	Δ	WT	WT	Δ	Nitrogen	0.91	0.96	1.89
WT	Δ	Δ	Δ	Δ	WT	Nitrogen	1.46	0.91	2.06
WT	Δ	WT	WT	WT	Δ	Nitrogen	1.97	0.90	2.27
WT	WT	Δ	Δ	WT	WT	Air	3.42	0.37	1.02
Δ	Δ	WT	Δ	WT	WT	Nitrogen	0.00	0.58	0.00
Δ	WT	Δ	Δ	WT	Δ	Air	1.92	1.01	2.26

Appendix C: Raw data evaluating the impact on ethanol metabolism of the ADH isozymes and environmental factors using the ministat bioprocess equipment.

		ADH genotype (Wild Type or deleted)															
Whole Plots	Subplots	ADH1	ADH2	ADH3	ADH4	ADH5	SFA1	Aeration	Media pH	Temp. (ºC)	Zinc (µM)	Glucose (g l^{-1})	Dilution rate	Ethanol produced (g l^{-1})	Glucose consumed (g l^{-1})	Cell count at culture harvest ($\times 10^8$)	Total free fatty acids (mg l^{-1})
1	1	Δ	Δ	Δ	Δ	Δ	Δ	Air	5.5	33	0.1	7.5	0.29	0.00	4.42	0.62	15.91
1	1	WT	Δ	Δ	WT	WT	Δ	Air	4.5	33	0.1	7.5	0.29	1.00	7.38	1.19	24.05
1	1	WT	WT	Δ	WT	Δ	Δ	Nitrogen	5.5	33	38.0	7.5	0.29	2.16	7.50	1.12	27.59
1	1	WT	Δ	WT	WT	Δ	Δ	Nitrogen	5.5	33	0.1	25.0	0.29	5.79	20.30	2.19	49.24
1	1	Δ	Δ	Δ	Δ	WT	Δ	Nitrogen	4.5	33	38.0	7.5	0.29	0.00	0.02	0.00	1.74
1	1	Δ	WT	WT	WT	WT	Δ	Air	5.5	33	38.0	25.0	0.29	0.72	8.18	0.32	11.44
1	1	Δ	Δ	WT	Δ	WT	WT	Nitrogen	5.5	33	0.1	25.0	0.29	0.00	0.00	0.01	3.85
1	1	Δ	Δ	Δ	Δ	WT	WT	Air	4.5	33	0.1	7.5	0.29	0.58	2.75	0.36	6.68
1	1	Δ	WT	Δ	WT	WT	WT	Nitrogen	4.5	33	0.1	25.0	0.29	0.00	0.00	0.01	2.78
1	1	Δ	WT	WT	Δ	Δ	Δ	Air	4.5	33	38.0	7.5	0.29	0.74	6.79	0.71	14.95
1	1	Δ	WT	Δ	Δ	WT	WT	Nitrogen	5.5	33	38.0	7.5	0.29	0.00	0.00	0.01	2.96
1	2	WT	Δ	WT	WT	WT	WT	Nitrogen	4.5	30	0.1	7.5	0.29	2.40	7.50	1.60	35.97
1	2	WT	WT	Δ	Δ	Δ	Δ	Air	5.5	30	38.0	25.0	0.29	3.43	12.71	0.69	22.19
1	2	WT	Δ	Δ	WT	WT	Δ	Air	5.5	30	0.1	25.0	0.29	0.85	1.39	0.34	6.21
1	2	Δ	WT	Δ	WT	Δ	WT	Air	4.5	30	38.0	7.5	0.29	0.70	7.02	0.56	18.06
1	2	WT	WT	Δ	WT	WT	WT	Nitrogen	5.5	30	38.0	7.5	0.29	1.47	7.50	0.93	21.19
1	2	WT	Δ	Δ	Δ	Δ	WT	Air	5.5	30	0.1	7.5	0.29	0.69	3.81	0.64	9.10
1	2	Δ	WT	WT	Δ	WT	Δ	Nitrogen	5.5	30	0.1	7.5	0.29	0.00	0.06	0.00	2.92
1	2	WT	WT	Δ	Δ	WT	WT	Nitrogen	5.5	30	0.1	25.0	0.29	2.74	25.00	0.66	17.51
1	2	WT	Δ	WT	Δ	Δ	WT	Nitrogen	4.5	30	0.1	25.0	0.29	3.36	14.34	1.33	33.76
1	2	Δ	WT	WT	Δ	WT	WT	Nitrogen	4.5	30	38.0	25.0	0.29	0.00	0.00	0.00	2.86
1	2	Δ	WT	Δ	Δ	Δ	Δ	Air	4.5	30	0.1	25.0	0.29	0.00	3.15	0.26	10.64
2	3	Δ	WT	Δ	Δ	Δ	WT	Nitrogen	4.5	33	38.0	25.0	0.06	0.00	0.00	0.01	1.67
2	3	Δ	Δ	WT	Δ	WT	Δ	Air	5.5	33	38.0	7.5	0.06	0.00	7.50	1.22	15.48
2	3	Δ	WT	WT	Δ	WT	WT	Air	4.5	33	0.1	25.0	0.06	0.00	7.95	0.92	15.22
2	3	WT	Δ	WT	Δ	WT	Δ	Air	4.5	33	0.1	7.5	0.06	0.00	7.18	0.57	14.66
2	3	WT	WT	WT	WT	Δ	WT	Air	4.5	33	38.0	7.5	0.06	0.00	7.50	1.05	12.20
2	3	WT	Δ	Δ	WT	WT	WT	Air	5.5	33	38.0	7.5	0.06	0.00	7.50	2.19	17.37
2	3	Δ	Δ	Δ	WT	Δ	Δ	Nitrogen	4.5	33	0.1	25.0	0.06	0.00	0.00	0.00	1.02
2	3	WT	WT	Δ	Δ	WT	Δ	Air	5.5	33	0.1	25.0	0.06	0.00	13.94	0.94	16.71
2	3	Δ	WT	WT	Δ	Δ	Δ	Nitrogen	5.5	33	0.1	25.0	0.06	2.13	25.00	0.85	10.93
2	3	WT	WT	Δ	Δ	Δ	Δ	Nitrogen	4.5	33	0.1	7.5	0.06	1.37	7.50	1.48	15.65
2	3	WT	Δ	WT	Δ	WT	WT	Nitrogen	4.5	33	38.0	25.0	0.06	3.99	25.00	2.40	20.55
2	4	Δ	WT	Δ	WT	Δ	Δ	Nitrogen	4.5	30	0.1	7.5	0.06	1.56	7.50	0.50	6.23
2	4	Δ	WT	WT	WT	WT	WT	Nitrogen	4.5	30	38.0	7.5	0.06	1.98	7.50	0.57	9.86
2	4	WT	WT	Δ	Δ	Δ	WT	Air	4.5	30	0.1	25.0	0.06	0.00	7.67	0.81	13.02
2	4	WT	WT	WT	WT	Δ	Δ	Air	4.5	30	0.1	25.0	0.06	0.82	11.21	0.70	11.93
2	4	Δ	WT	WT	WT	Δ	Δ	Air	5.5	30	38.0	7.5	0.06	0.63	7.50	1.77	14.52
2	4	Δ	WT	Δ	WT	WT	WT	Air	5.5	30	38.0	25.0	0.06	0.57	25.00	0.90	15.25
2	4	WT	Δ	Δ	WT	Δ	WT	Nitrogen	4.5	30	0.1	7.5	0.06	1.05	7.50	1.01	18.13
2	4	WT	WT	WT	Δ	WT	Δ	Nitrogen	5.5	30	38.0	25.0	0.06	3.62	25.00	3.19	28.02
2	4	Δ	Δ	WT	Δ	Δ	WT	Air	5.5	30	38.0	25.0	0.06	0.61	25.00	0.89	15.74
2	4	Δ	WT	Δ	Δ	WT	Δ	Air	5.5	30	0.1	7.5	0.06	0.00	6.77	1.43	12.89
2	4	WT	Δ	WT	WT	WT	Δ	Air	4.5	30	38.0	7.5	0.06	0.00	7.50	0.18	23.56

Whole Plots	Subplots	ADH genotype (Wild Type or deleted) ADH1	ADH2	ADH3	ADH4	ADH5	SFA1	Aeration	Media pH	Temp. (ºC)	Zinc (μM)	Glucose (g l^{-1})	Dilution rate	Ethanol produced (g l^{-1})	Glucose consumed (g l^{-1})	Cell count at culture harvest ($\times 10^6$)	Total free fatty acids (mg l^{-1})
3	5	Δ	WT	WT	WT	WT	WT	Air	5.5	33	0.1	7.5	0.29	0.27	3.95	0.53	8.64
3	5	WT	Δ	WT	Δ	WT	Δ	Nitrogen	5.5	33	38.0	7.5	0.29	2.33	7.50	1.20	25.06
3	5	WT	WT	WT	Δ	WT	Δ	Nitrogen	4.5	33	0.1	25.0	0.29	3.75	18.09	1.03	31.69
3	5	WT	Δ	WT	WT	Δ	WT	Air	5.5	33	38.0	7.5	0.29	1.05	7.50	1.74	28.07
3	5	Δ	Δ	Δ	WT	Δ	WT	Nitrogen	5.5	33	38.0	25.0	0.29	0.00	4.10	0.00	2.49
3	5	Δ	Δ	WT	Δ	WT	WT	Air	4.5	33	38.0	7.5	0.29	0.22	7.50	0.68	22.41
3	5	WT	WT	WT	Δ	Δ	WT	Nitrogen	5.5	33	38.0	25.0	0.29	5.91	24.51	0.90	34.73
3	5	WT	Δ	Δ	Δ	Δ	Δ	Air	4.5	33	38.0	25.0	0.29	1.90	16.11	0.47	21.54
3	5	Δ	Δ	WT	WT	Δ	WT	Air	4.5	33	0.1	25.0	0.29	0.55	7.30	0.53	11.43
3	5	Δ	WT	WT	WT	Δ	WT	Nitrogen	4.5	33	0.1	7.5	0.29	1.37	7.50	1.00	18.35
3	5	WT	WT	Δ	WT	Δ	WT	Air	5.5	33	0.1	25.0	0.29	0.89	12.75	0.91	14.96
3	6	Δ	Δ	Δ	Δ	Δ	Δ	Nitrogen	5.5	30	38.0	25.0	0.29	0.14	3.25	0.01	1.39
3	6	Δ	WT	Δ	WT	WT	Δ	Nitrogen	4.5	30	38.0	25.0	0.29	0.37	11.71	0.14	10.71
3	6	Δ	Δ	Δ	Δ	WT	Δ	Air	5.5	30	38.0	7.5	0.29	0.16	4.68	0.44	14.72
3	6	Δ	Δ	WT	WT	Δ	Δ	Air	4.5	30	0.1	7.5	0.29	0.93	6.71	1.34	25.73
3	6	WT	WT	WT	WT	WT	WT	Air	4.5	30	38.0	25.0	0.29	2.53	19.05	0.52	20.18
3	6	WT	Δ	Δ	Δ	WT	Δ	Nitrogen	4.5	30	0.1	7.5	0.29	2.38	7.50	1.47	33.28
3	6	WT	Δ	Δ	Δ	WT	WT	Air	4.5	30	38.0	25.0	0.29	2.47	17.87	0.71	23.28
3	6	WT	WT	Δ	Δ	Δ	WT	Nitrogen	4.5	30	38.0	7.5	0.29	1.78	7.50	0.85	21.46
3	6	Δ	WT	WT	WT	Δ	WT	Nitrogen	5.5	30	0.1	25.0	0.29	0.11	11.71	0.00	5.90
3	6	WT	Δ	WT	Δ	Δ	Δ	Air	5.5	30	0.1	25.0	0.29	0.60	8.99	0.81	19.08
3	6	Δ	Δ	Δ	Δ	WT	WT	Nitrogen	4.5	30	0.1	7.5	0.29	0.08	1.52	0.02	3.05
4	7	Δ	WT	WT	Δ	Δ	Δ	Air	4.5	30	38.0	25.0	0.06	0.70	24.65	0.18	29.59
4	7	WT	Δ	WT	WT	Δ	WT	Nitrogen	5.5	30	38.0	25.0	0.06	3.74	24.75	1.21	28.32
4	7	WT	Δ	WT	Δ	WT	WT	Air	5.5	30	38.0	7.5	0.06	0.45	7.26	0.86	23.98
4	7	Δ	Δ	WT	WT	WT	Δ	Nitrogen	4.5	30	0.1	25.0	0.06	3.70	24.76	1.68	30.05
4	7	Δ	Δ	WT	Δ	Δ	WT	Nitrogen	5.5	30	0.1	7.5	0.06	0.00	0.41	0.01	4.62
4	7	Δ	Δ	Δ	Δ	Δ	Δ	Air	4.5	30	38.0	7.5	0.06	0.00	7.26	1.14	28.84
4	7	Δ	Δ	Δ	WT	WT	WT	Nitrogen	4.5	30	38.0	25.0	0.06	1.17	9.04	0.18	10.08
4	7	Δ	WT	WT	Δ	Δ	WT	Air	4.5	30	0.1	7.5	0.06	0.00	7.25	1.47	29.53
4	7	Δ	Δ	Δ	WT	Δ	Δ	Air	5.5	30	0.1	25.0	0.06	0.00	24.76	1.92	44.40
4	7	WT	WT	WT	WT	WT	Δ	Nitrogen	5.5	30	0.1	7.5	0.06	1.30	7.26	1.11	22.83
4	7	WT	WT	Δ	WT	WT	WT	Air	4.5	30	0.1	7.5	0.06	0.00	7.16	1.18	17.63
4	8	Δ	Δ	Δ	Δ	WT	WT	Air	5.5	33	0.1	25.0	0.06	0.00	24.44	0.46	32.48
4	8	Δ	WT	Δ	WT	Δ	WT	Nitrogen	5.5	33	0.1	7.5	0.06	0.51	1.89	0.07	4.12
4	8	Δ	Δ	Δ	WT	WT	Δ	Nitrogen	5.5	33	0.1	7.5	0.06	0.00	0.00	0.01	3.47
4	8	WT	Δ	Δ	Δ	Δ	WT	Nitrogen	5.5	33	38.0	7.5	0.06	1.00	7.24	0.32	13.41
4	8	WT	WT	WT	Δ	WT	WT	Nitrogen	5.5	33	0.1	7.5	0.06	1.81	7.24	0.25	14.84
4	8	Δ	Δ	Δ	WT	WT	Δ	Air	4.5	33	38.0	25.0	0.06	0.00	24.64	0.26	21.01
4	8	WT	WT	Δ	WT	WT	Δ	Nitrogen	4.5	33	38.0	25.0	0.06	2.56	24.70	0.79	27.51
4	8	WT	WT	WT	Δ	WT	Δ	Air	4.5	33	38.0	7.5	0.06	0.00	7.25	0.33	18.07
4	8	WT	Δ	WT	WT	WT	WT	Air	5.5	33	0.1	25.0	0.06	1.19	22.72	0.61	30.57
4	8	WT	Δ	WT	WT	Δ	Δ	Nitrogen	4.5	33	38.0	7.5	0.06	0.99	7.25	0.78	18.78
4	8	Δ	Δ	Δ	WT	Δ	WT	Air	4.5	33	0.1	7.5	0.06	0.00	7.16	0.59	17.51

Bibliography

ABBOTT, D. A., ZELLE, R. M., PRONK, J. T. & VAN MARIS, A. J. 2009. Metabolic engineering of *Saccharomyces cerevisiae* for production of carboxylic acids: current status and challenges. *FEMS Yeast Res,* 9**,** 1123-36.

AHN, J. H., JANG, Y.-S. & LEE, S. Y. 2016. Production of succinic acid by metabolically engineered microorganisms. *Curr Opin Biotechnol,* 42**,** 54-66.

AKHTAR, M. K., TURNER, N. J. & JONES, P. R. 2013. Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proc Natl Acad Sci,* 110**,** 87-92.

ANDREASEN, A. A. & STIER, T. J. B. 1954. Anaerobic nutrition of *Saccharomyces cerevisiae*. II. Unsaturated fatty and requirement for growth in a defined medium. *J Cell Compar Physiol,* 43**,** 271-281.

ANISIMOVA, M. & GASCUEL, O. 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol,* 55**,** 539-52.

ASLANKOOHI, E., VOORDECKERS, K., SUN, H., SANCHEZ-RODRIGUEZ, A., VAN DER ZANDE, E., MARCHAL, K. & VERSTREPEN, K. J. 2012. Nucleosomes affect local transformation efficiency. *Nucleic Acids Res*.

AUNG, H. W., HENRY, S. A. & WALKER, L. P. 2013. Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. *Ind Biotechnol (New Rochelle N Y),* 9**,** 215-228.

BAKKER, B. M., BRO, C., KOTTER, P., LUTTIK, M. A., VAN DIJKEN, J. P. & PRONK, J. T. 2000. The mitochondrial alcohol dehydrogenase Adh3p is involved in a redox shuttle in *Saccharomyces cerevisiae*. *J Bacteriol,* 182**,** 4730-7.

BASSO, L. C., DE AMORIM, H. V., DE OLIVEIRA, A. J. & LOPES, M. L. 2008. Yeast selection for fuel ethanol production in Brazil. *FEMS Yeast Res,* 8**,** 1155-63.

BELCHER, J., MCLEAN, K. J., MATTHEWS, S., WOODWARD, L. S., FISHER, K., RIGBY, S. E. J., NELSON, D. R., POTTS, D., BAYNHAM, M. T., PARKER, D. A., LEYS, D. & MUNRO, A. W. 2014. Structure and biochemical properties of the alkene producing cytochrome P450 OleTJE (CYP152L1) from the *Jeotgalicoccus sp*. 8456 bacterium. *J Biol Chem,* 289**,** 6535-6550.

BENNETZEN, J. L. & HALL, B. D. 1982. The primary structure of the *Saccharomyces cerevisiae* gene for alcohol dehydrogenase. *J Biol Chem,* 257**,** 3018-3025.

BERNARD, A., DOMERGUE, F., PASCAL, S., JETTER, R., RENNE, C., FAURE, J. D., HASLAM, R. P., NAPIER, J. A., LESSIRE, R. & JOUBES, J. 2012. Reconstitution of plant alkane biosynthesis in yeast demonstrates that Arabidopsis ECERIFERUM1 and ECERIFERUM3 are

core components of a very-long-chain alkane synthesis complex. *Plant Cell,* 24**,** 3106-18.

BERTANI, G. 1951. Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *J Bacteriol,* 62**,** 293-300.

BIRD, A. J., GORDON, M., EIDE, D. J. & WINGE, D. R. 2006. Repression of *ADH1* and *ADH3* during zinc deficiency by Zap1-induced intergenic RNA transcripts. *Embo j,* 25**,** 5726-34.

BOONSTRA, B., FRENCH, C. E., WAINWRIGHT, I. & BRUCE, N. C. 1999. The *udhA* gene of *Escherichia coli* encodes a soluble pyridine nucleotide transhydrogenase. *J Bacteriol,* 181**,** 1030-4.

BRACHMANN, C. B., DAVIES, A., COST, G. J., CAPUTO, E., LI, J., HIETER, P. & BOEKE, J. D. 1998. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast,* 14**,** 115-32.

BRANDEN, C., JORNVALL, H., EKLUND, H., AND FURUGREN, B. 1975. *Alcohol Dehydrogenases, in The Enzymes,* New York, Academic Press.

BRO, C., REGENBERG, B., FORSTER, J. & NIELSEN, J. 2006. In silico aided metabolic engineering of *Saccharomyces cerevisiae* for improved bioethanol production. *Metab Eng,* 8**,** 102-11.

BROWN, S. W., OLIVER, S. G., HARRISON, D. E. F. & RIGHELATO, R. C. 1981. Ethanol inhibition of yeast growth and fermentation: Differences in the magnitude and complexity of the effect. *Appl Microbiol Biotechnol,* 11**,** 151-155.

BUIJS, N. A., ZHOU, Y. J., SIEWERS, V. & NIELSEN, J. 2015. Long-chain alkane production by the yeast *Saccharomyces cerevisiae*. *Biotechnol Bioeng,* 112**,** 1275-9.

BYRNE, D., DUMITRIU, A. & SEGRE, D. 2012. Comparative multi-goal tradeoffs in systems engineering of microbial metabolism. *BMC Syst Biol,* 6**,** 127.

CAO, L., ZHANG, A., KONG, Q., XU, X., JOSINE, T. L. & CHEN, X. 2007. Overexpression of *GLT1* in *fps1*Δ*gpd*Δ mutant for optimum ethanol formation by *Saccharomyces cerevisiae*. *Biomol Eng,* 24**,** 638-642.

CAO, Z., SONG, P., XU, Q., SU, R. & ZHU, G. 2011. Overexpression and biochemical characterization of soluble pyridine nucleotide transhydrogenase from *Escherichia coli*. *FEMS Microbiol Lett,* 320**,** 9-14.

CASTRESANA, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol,* 17**,** 540-52.

CHEN, B., LEE, D. Y. & CHANG, M. W. 2015. Combinatorial metabolic engineering of *Saccharomyces cerevisiae* for terminal alkene production. *Metab Eng,* 31**,** 53-61.

CHEN, B., LING, H. & CHANG, M. W. 2013. Transporter engineering for improved tolerance against alkane biofuels in *Saccharomyces cerevisiae*. *Biotechnol Biofuels,* 6**,** 21.

CHEN, X., NIELSEN, K. F., BORODINA, I., KIELLAND-BRANDT, M. C. & KARHUMAA, K. 2011. Increased isobutanol production in *Saccharomyces cerevisiae* by overexpression of genes in valine metabolism. *Biotechnol Biofuels,* 4**,** 21.

CHEVENET, F., BRUN, C., BANULS, A. L., JACQ, B. & CHRISTEN, R. 2006. TreeDyn: towards dynamic graphics and annotations for analyses of trees. *BMC Bioinformatics,* 7**,** 439.

CHIGIRA, Y., OKA, T., OKAJIMA, T. & JIGAMI, Y. 2008. Engineering of a mammalian O-glycosylation pathway in the yeast *Saccharomyces cerevisiae*: production of O-fucosylated epidermal growth factor domains. *Glycobiology,* 18**,** 303-314.

CHUNG, A. E. 1970. Pyridine nucleotide transhydrogenase from *Azotobacter vinelandii*. *J Bacteriol,* 102**,** 438-447.

CIRIACY, M. 1979. Isolation and characterization of further Cis- and Trans-acting regulatory elements involved in the synthesis of glucose-repressible alcohol dehydrogenase (ADHII) in *Saccharomyces cerevisiae*. *Mol Gen Genet,* 176**,** 427-431.

COX, I. & GAUDARD, M. 2013. *Discovering Partial Least Squares with JMP,* Cary, NC, USA, SAS Institute.

CRABTREE, H. G. 1928. The carbohydrate metabolism of certain pathological overgrowths. *Biochem J,* 22**,** 1289-98.

DA SILVA, G. P., MACK, M. & CONTIERO, J. 2009. Glycerol: A promising and abundant carbon source for industrial microbiology. *Biotechnol Adv,* 27**,** 30-39.

DAS, D., ESER, B. E., SCIORE, A., MARSH, E. N. G. & HAN, J. 2011. Oxygen-independent decarbonylation of aldehydes by cyano-bacterial aldehyde decarbonylase: a new reaction of di-iron enzymes. *Angewandte Chemie (International Ed. in English),* 50**,** 7148-7152.

DE DEKEN, R. H. 1966. The Crabtree effect: a regulatory system in yeast. *J Gen Microbiol,* 44**,** 149-56.

DE JONG, B., SIEWERS, V. & NIELSEN, J. 2012. Systems biology of yeast: enabling technology for development of cell factories for production of advanced biofuels. *Curr Opin Biotech,* 23**,** 624-630.

DE SMIDT, O., DU PREEZ, J. C. & ALBERTYN, J. 2008. The alcohol dehydrogenases of *Saccharomyces cerevisiae*: a comprehensive review. *FEMS Yeast Res,* 8**,** 967-78.

DE SMIDT, O., DU PREEZ, J. C. & ALBERTYN, J. 2012. Molecular and physiological aspects of alcohol dehydrogenases in the ethanol metabolism of *Saccharomyces cerevisiae*. *FEMS Yeast Res,* 12**,** 33-47.

DEERE, D., SHEN, J., VESEY, G., BELL, P., BISSINGER, P. & VEAL, D. 1998. Flow cytometry and cell sorting for yeast viability assessment and cell selection. *Yeast,* 14**,** 147-160.

DENIS, C. L., FERGUSON, J. & YOUNG, E. T. 1983. mRNA levels for the fermentative alcohol dehydrogenase of *Saccharomyces cerevisiae* decrease upon growth on a nonfermentable carbon source. *J Biol Chem,* 258**,** 1165-71.

DEREEPER, A., GUIGNON, V., BLANC, G., AUDIC, S., BUFFET, S., CHEVENET, F., DUFAYARD, J. F., GUINDON, S., LEFORT, V., LESCOT, M., CLAVERIE, J. M. & GASCUEL, O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res,* 36**,** W465-9.

DICKINSON, J. R., SALGADO, L. E. J. & HEWLINS, M. J. E. 2003. The catabolism of amino acids to long chain and complex alcohols in *Saccharomyces cerevisiae*. *J Biol Chem,* 278**,** 8028-8034.

DICKINSON, J. R. S., M. 2004. *The Metabolism and Molecular Physiology of Saccharmoyces cerevisiae,* UK, CRC Press.

DIDERICH, J. A., RAAMSDONK, L. M., KRUCKEBERG, A. L., BERDEN, J. A. & VAN DAM, K. 2001. Physiological properties of *Saccharomyces cerevisiae* from which hexokinase II has been deleted. *Appl Environ Microbiol,* 67**,** 1587-93.

DOBSON, P. D., SMALLBONE, K., JAMESON, D., SIMEONIDIS, E., LANTHALER, K., PIR, P., LU, C., SWAINSTON, N., DUNN, W. B., FISHER, P., HULL, D., BROWN, M., OSHOTA, O., STANFORD, N. J., KELL, D. B., KING, R. D., OLIVER, S. G., STEVENS, R. D. & MENDES, P. 2010. Further developments towards a genome-scale metabolic model of yeast. *BMC Syst Biol,* 4**,** 145.

DREWKE, C. & CIRIACY, M. 1988. Overexpression, purification and properties of alcohol dehydrogenase IV from *Saccharomyces cerevisiae*. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression,* 950**,** 54-60.

DREWKE, C., THIELEN, J. & CIRIACY, M. 1990. Ethanol formation in *adh*0 mutants reveals the existence of a novel acetaldehyde-reducing activity in *Saccharomyces cerevisiae*. *J Bacteriol,* 172**,** 3909-17.

DUARTE, N. C., HERRGÅRD, M. J. & PALSSON, B. Ø. 2004. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res,* 14**,** 1298-1309.

EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32**,** 1792-7.

FELDMANN, H., AIGLE, M., ALJINOVIC, G., ANDRE, B., BACLET, M. C., BARTHE, C., BAUR, A., BECAM, A. M., BITEAU, N., BOLES, E., BRANDT, T., BRENDEL, M., BRUCKNER, M., BUSSEREAU, F., CHRISTIANSEN, C., CONTRERAS, R., CROUZET, M., CZIEPLUCH, C., DEMOLIS, N., DELAVEAU, T., DOIGNON, F., DOMDEY, H., DUSTERHUS, S., DUBOIS, E., DUJON, B., EL BAKKOURY, M., ENTIAN, K. D., FEURMANN, M., FIERS, W., FOBO, G. M., FRITZ, C., GASSENHUBER, H., GLANDSDORFF, N., GOFFEAU, A., GRIVELL, L. A., DE HAAN, M., HEIN, C., HERBERT, C. J., HOLLENBERG, C. P., HOLMSTROM, K., JACQ, C., JACQUET, M., JAUNIAUX, J. C., JONNIAUX, J. L., KALLESOE, T., KIESAU, P., KIRCHRATH, L., KOTTER, P., KOROL, S., LIEBL, S., LOGGHE, M., LOHAN, A. J., LOUIS, E. J., LI, Z. Y., MAAT, M. J., MALLET, L., MANNHAUPT, G., MESSENGUY, F., MIOSGA, T., MOLEMANS, F., MULLER, S., NASR, F., OBERMAIER, B., PEREA, J., PIERARD, A., PIRAVANDI, E., POHL, F. M., POHL, T. M., POTIER, S., PROFT, M., PURNELLE, B., RAMEZANI RAD, M., RIEGER, M., ROSE, M., SCHAAFF-GERSTENSCHLAGER, I., SCHERENS, B., SCHWARZLOSE, C., SKALA, J., SLONIMSKI, P. P., SMITS, P. H., SOUCIET, J. L., STEENSMA, H. Y., STUCKA, R., URRESTARAZU, A., VAN DER AART, Q. J., VAN DYCK, L., VASSAROTTI, A., VETTER, I., VIERENDEELS, F.,

VISSERS, S., WAGNER, G., DE WERGIFOSSE, P., WOLFE, K. H., ZAGULSKI, M., ZIMMERMANN, F. K., MEWES, H. W. & KLEINE, K. 1994. Complete DNA sequence of yeast chromosome II. *Embo j,* 13**,** 5795-809.

FISHMAN-LOBELL, J., RUDIN, N. & HABER, J. E. 1992. Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Mol Cell Biol,* 12**,** 1292-303.

FLIKWEERT, M. T., VAN DER ZANDEN, L., JANSSEN, W. M., STEENSMA, H. Y., VAN DIJKEN, J. P. & PRONK, J. T. 1996. Pyruvate decarboxylase: an indispensable enzyme for growth of *Saccharomyces cerevisiae* on glucose. *Yeast,* 12**,** 247-57.

FLIKWEERT, M. T., VAN DIJKEN, J. P. & PRONK, J. T. 1997. Metabolic responses of pyruvate decarboxylase-negative *Saccharomyces cerevisiae* to glucose excess. *Appl Environ Microbiol,* 63**,** 3399-404.

FORSTER, J., FAMILI, I., FU, P., PALSSON, B. O. & NIELSEN, J. 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res,* 13**,** 244-53.

FREEDMAN, L. P., COCKBURN, I. M. & SIMCOE, T. S. 2015. The economics of reproducibility in preclinical research. *PLoS Biol,* 13**,** e1002165.

GANCEDO, J. M. 1998. Yeast carbon catabolite repression. *Microbiol Mol Biol Rev,* 62**,** 334-361.

GANZHORN, A. J., GREEN, D. W., HERSHEY, A. D., GOULD, R. M. & PLAPP, B. V. 1987. Kinetic characterization of yeast alcohol dehydrogenases. Amino acid residue 294 and substrate specificity. *J Biol Chem,* 262**,** 3754-61.

GOFFEAU, A., BARRELL, B. G., BUSSEY, H., DAVIS, R. W., DUJON, B., FELDMANN, H., GALIBERT, F., HOHEISEL, J. D., JACQ, C., JOHNSTON, M., LOUIS, E. J., MEWES, H. W., MURAKAMI, Y., PHILIPPSEN, P., TETTELIN, H. & OLIVER, S. G. 1996. Life with 6000 genes. *Science,* 274**,** 546, 563-7.

GREEN, M. R., SAMBROOK, J. & SAMBROOK, J. 2012. *Molecular cloning : a laboratory manual,* Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press.

GUADALUPE MEDINA, V., ALMERING, M. J., VAN MARIS, A. J. & PRONK, J. T. 2010. Elimination of glycerol production in anaerobic cultures of a *Saccharomyces cerevisiae* strain engineered to use acetic acid as an electron acceptor. *Appl Environ Microbiol,* 76**,** 190-5.

GUINDON, S., DUFAYARD, J. F., LEFORT, V., ANISIMOVA, M., HORDIJK, W. & GASCUEL, O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol,* 59**,** 307-21.

HACKEL, B. J., HUANG, D., BUBOLZ, J. C., WANG, X. X. & SHUSTA, E. V. 2006. Production of soluble and active transferrin receptor-targeting single-chain antibody using *Saccharomyces cerevisiae*. *Pharm Res,* 23**,** 790-797.

HALL, C., BRACHAT, S. & DIETRICH, F. S. 2005. contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. *Eukaryot Cell,* 4**,** 1102-1115.

HARGER, M., ZHENG, L., MOON, A., AGER, C., AN, J. H., CHOE, C., LAI, Y.-L., MO, B., ZONG, D., SMITH, M. D., EGBERT, R. G., MILLS, J. H., BAKER, D., PULTZ, I. S. & SIEGEL, J. B. 2012. Expanding the product profile of a microbial alkane biosynthetic pathway. *ACS Synth Biol,* 2**,** 59-62.

HASUNUMA, T. & KONDO, A. 2012. Development of yeast cell factories for consolidated bioprocessing of lignocellulose to bioethanol through cell surface engineering. *Biotechnol Adv,* 30**,** 1207-1218.

HAZELWOOD, L. A., DARAN, J. M., VAN MARIS, A. J., PRONK, J. T. & DICKINSON, J. R. 2008. The Ehrlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* metabolism. *Appl Environ Microbiol,* 74**,** 2259-66.

HEAVNER, B. D. & PRICE, N. D. 2015. Comparative analysis of yeast metabolic network models highlights progress, opportunities for metabolic reconstruction. *PLoS Comput Biol,* 11**,** e1004530.

HEAVNER, B. D., SMALLBONE, K., BARKER, B., MENDES, P. & WALKER, L. P. 2012. Yeast 5 - an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network. *BMC Syst Biol,* 6**,** 55.

HEAVNER, B. D., SMALLBONE, K., PRICE, N. D. & WALKER, L. P. 2013. Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database (Oxford),* 2013**,** bat059.

HERRGARD, M. J., SWAINSTON, N., DOBSON, P., DUNN, W. B., ARGA, K. Y., ARVAS, M., BLUTHGEN, N., BORGER, S., COSTENOBLE, R., HEINEMANN, M., HUCKA, M., LE NOVERE, N., LI, P., LIEBERMEISTER, W., MO, M. L., OLIVEIRA, A. P., PETRANOVIC, D., PETTIFER, S., SIMEONIDIS, E., SMALLBONE, K., SPASIC, I., WEICHART, D., BRENT, R., BROOMHEAD, D. S., WESTERHOFF, H. V., KIRDAR, B., PENTTILA, M., KLIPP, E., PALSSON, B. O., SAUER, U., OLIVER, S. G., MENDES, P., NIELSEN, J. & KELL, D. B. 2008. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol,* 26**,** 1155-60.

HEUX, S., CACHON, R. & DEQUIN, S. 2006. Cofactor engineering in *Saccharomyces cerevisiae*: Expression of a H2O-forming NADH oxidase and impact on redox metabolism. *Metab Eng,* 8**,** 303-14.

HONG, K. K. & NIELSEN, J. 2012. Metabolic engineering of *Saccharomyces cerevisiae:* a key cell factory platform for future biorefineries. *Cell Mol Life Sci,* 69**,** 2671-90.

HOU, J., LAGES, N. F., OLDIGES, M. & VEMURI, G. N. 2009. Metabolic impact of redox cofactor perturbations in *Saccharomyces cerevisiae*. *Metab Eng,* 11**,** 253-61.

HOWARD, T. P., MIDDELHAUFE, S., MOORE, K., EDNER, C., KOLAK, D. M., TAYLOR, G. N., PARKER, D. A., LEE, R., SMIRNOFF, N., AVES, S. J. & LOVE, J. 2013. Synthesis of customized petroleum-replica fuel molecules by targeted modification of free fatty acid pools in *Escherichia coli*. *Proc Natl Acad Sci,* 110**,** 7636-7641.

HUH, W.-K., FALVO, J. V., GERKE, L. C., CARROLL, A. S., HOWSON, R. W., WEISSMAN, J. S. & O'SHEA, E. K. 2003. Global analysis of protein localization in budding yeast. *Nature,* 425**,** 686-691.

IDA, Y., FURUSAWA, C., HIRASAWA, T. & SHIMIZU, H. 2012. Stable disruption of ethanol production by deletion of the genes encoding alcohol dehydrogenase isozymes in *Saccharomyces cerevisiae*. *J Biosci Bioeng,* 113**,** 192-195.

IDA, Y., HIRASAWA, T., FURUSAWA, C. & SHIMIZU, H. 2013. Utilization of Saccharomyces cerevisiae recombinant strain incapable of both ethanol and glycerol biosynthesis for anaerobic bioproduction. *Appl Microbiol Biotechnol,* 97**,** 4811-9.

International Energy Agency 2008. World Energy Outlook (Paris, France).

International Energy Agency 2015a. Climate pledges for COP21 slow energy sector emissions growth dramatically. Paris.

International Energy Agency 2015b. World Energy Outlook (London, UK).

International Energy Agency. 2016. *Oil Market Report* [Online]. Available: https://www.iea.org/aboutus/faqs/oil/ [Accessed 17.08.2016].

ISHTAR SNOEK, I. S. & YDE STEENSMA, H. 2007. Factors involved in anaerobic growth of *Saccharomyces cerevisiae*. *Yeast,* 24**,** 1-10.

ITO, Y., HIRASAWA, T. & SHIMIZU, H. 2014. Metabolic engineering of *Saccharomyces cerevisiae* to improve succinic acid production based on metabolic profiling. *Biosci Biotechnol Biochem,* 78**,** 151-9.

JANSEN, M. L. A., VAN DE GRAAF, M. J. & VERWAAL, R. 2012. Dicarboxylic acid production process. *US 2012/0040422 A1*.

JOUHTEN, P., RINTALA, E., HUUSKONEN, A., TAMMINEN, A., TOIVARI, M., WIEBE, M., RUOHONEN, L., PENTTILA, M. & MAAHEIMO, H. 2008. Oxygen dependence of metabolic fluxes and energy generation of *Saccharomyces cerevisiae* CEN.PK113-1A. *BMC Syst Biol,* 2**,** 60.

JUNGMEIER, G. 2013. *The possible role of biorefineries in a bioeconomy – Activities of IEA bioenergy task 42 "Biorefining"* [Online]. Available: http://www.iea-bioenergy.task42-biorefineries.com/en/ieabiorefinery/Publications-2.htm [Accessed 18.08.2016].

KANEHISA, M., GOTO, S., SATO, Y., KAWASHIMA, M., FURUMICHI, M. & TANABE, M. 2014. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res,* 42**,** D199-D205.

KAPPELI, O. 1986. Regulation of carbon metabolism in *Saccharomyces cerevisiae* and related yeasts. *Adv Microb Physiol,* 28**,** 181-209.

KELLIS, M., BIRREN, B. W. & LANDER, E. S. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature,* 428**,** 617-624.

KING, Z. A., LU, J., DRÄGER, A., MILLER, P., FEDEROWICZ, S., LERMAN, J. A., EBRAHIM, A., PALSSON, B. O. & LEWIS, N. E. 2016. BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res,* 44**,** D515-D522.

KUEPFER, L., SAUER, U. & BLANK, L. M. 2005. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res,* 15**,** 1421-1430.

KUMAR, A., SUTHERS, P. F. & MARANAS, C. D. 2012. MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. *BMC Bioinformatics,* 13**,** 6.

LARROY, C., FERNANDEZ, M. R., GONZALEZ, E., PARES, X. & BIOSCA, J. A. 2002. Characterization of the S*accharomyces cerevisiae* YMR318C (*ADH6*) gene product as a broad specificity NADPH-dependent alcohol dehydrogenase: relevance in aldehyde reduction. *Biochem J,* 361**,** 163-72.

LEE, S. Y., KIM, H. M. & CHEON, S. 2015. Metabolic engineering for the production of hydrocarbon fuels. *Curr Opin Biotech,* 33**,** 15-22.

LENDREM, D. W., LENDREM, B. C., ROWLAND-JONES, R., D'AGOSTINO, F., LINSLEY, M., OWEN, M. R. & ISAACS, J. D. 2016. Teaching examples for the design of experiments: geographical sensitivity and the self-fulfilling prophecy. *Pharm Stat,* 15**,** 90-2.

LENDREM, D. W., LENDREM, B. C., WOODS, D., ROWLAND-JONES, R., BURKE, M., CHATFIELD, M., ISAACS, J. D. & OWEN, M. R. 2015. Lost in space: design of experiments and scientific exploration in a Hogarth Universe. *Drug Discov Today,* 20**,** 1365-1371.

LESKOVAC, V., TRIVIĆ, S. & PERIČIN, D. 2002. The three zinc-containing alcohol dehydrogenases from baker's yeast, *Saccharomyces cerevisiae*. *FEMS Yeast Res,* 2**,** 481-494.

LIU, Y., WANG, C., YAN, J., ZHANG, W., GUAN, W., LU, X. & LI, S. 2014. Hydrogen peroxide-independent production of alpha-alkenes by *OleTJE* P450 fatty acid decarboxylase. *Biotechnol Biofuels,* 7**,** 28.

MADSEN, K. M., UDATHA, G. D. B. R. K., SEMBA, S., OTERO, J. M., KOETTER, P., NIELSEN, J., EBIZUKA, Y., KUSHIRO, T. & PANAGIOTOU, G. 2011. Linking genotype and phenotype of *Saccharomyces cerevisiae* strains reveals metabolic engineering targets and leads to triterpene hyper-producers. *PLoS ONE,* 6**,** e14763.

MARSH, E. N. & WAUGH, M. W. 2013. Aldehyde decarbonylases: enigmatic enzymes of hydrocarbon biosynthesis. *ACS Catal,* 3.

MATSUDA, F., SHIRAI, T., ISHII, J. & KONDO, A. 2013. Regulation of central carbon metabolism in *Saccharomyces cerevisiae* by metabolic inhibitors. *J Biosci Bioeng,* 116**,** 59-64.

MEIGHEN, E. A. 1991. Molecular biology of bacterial bioluminescence. *Microbiol Rev,* 55**,** 123-42.

MILLER, A. W., BEFORT, C., KERR, E. O. & DUNHAM, M. J. 2013. Design and use of multiplexed chemostat arrays. *J Vis Exp***,** 50262.

MIZUNO, A., TABEI, H. & IWAHUTI, M. 2006. Characterization of low-acetic-acid-producing yeast isolated from 2-deoxyglucose-resistant mutants and its application to high-gravity brewing. *J Biosci Bioeng,* 101**,** 31-7.

MO, M. L., PALSSON, B. O. & HERRGARD, M. J. 2009. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol,* 3**,** 37.

MORTLOCK, R. 2013. *Microorganisms as Model Systems for Studying Evolution*, Springer US.

MOURET, J. R., JACOBSEN, J. N. & GUILLOUET, S. E. 2006. Kinetic analysis of a trehalase-overexpressing strain grown on trehalose: a new tool for respiro-fermentative transition studies in *Saccharomyces cerevisiae*. *Lett Appl Microbiol,* 42**,** 363-8.

NAVARRETE, C., NIELSEN, J. & SIEWERS, V. 2014. Enhanced ethanol production and reduced glycerol formation in *fps1*Δ mutants of

Saccharomyces cerevisiae engineered for improved redox balancing. *AMB Express,* 4**,** 1-8.

NIJKAMP, J. F., VAN DEN BROEK, M., DATEMA, E., DE KOK, S., BOSMAN, L., LUTTIK, M. A., DARAN-LAPUJADE, P., VONGSANGNAK, W., NIELSEN, J., HEIJNE, W. H., KLAASSEN, P., PADDON, C. J., PLATT, D., KOTTER, P., VAN HAM, R. C., REINDERS, M. J., PRONK, J. T., DE RIDDER, D. & DARAN, J. M. 2012. De novo sequencing, assembly and analysis of the genome of the laboratory strain *Saccharomyces cerevisiae* CEN.PK113-7D, a model for modern industrial biotechnology. *Microb Cell Fact,* 11**,** 36.

NISSEN, T. L., ANDERLUND, M., NIELSEN, J., VILLADSEN, J. & KIELLAND-BRANDT, M. C. 2001. Expression of a cytoplasmic transhydrogenase in *Saccharomyces cerevisiae* results in formation of 2-oxoglutarate due to depletion of the NADPH pool. *Yeast,* 18**,** 19-32.

NISSEN, T. L., KIELLAND-BRANDT, M. C., NIELSEN, J. & VILLADSEN, J. 2000. Optimization of ethanol production in *Saccharomyces cerevisiae* by metabolic engineering of the ammonium assimilation. *Metab Eng,* 2**,** 69-77.

NOOKAEW, I., JEWETT, M. C., MEECHAI, A., THAMMARONGTHAM, C., LAOTENG, K., CHEEVADHANARAK, S., NIELSEN, J. & BHUMIRATANA, S. 2008. The genome-scale metabolic model iIN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol,* 2**,** 71.

NORTH, M., STEFFEN, J., LOGUINOV, A. V., ZIMMERMAN, G. R., VULPE, C. D. & EIDE, D. J. 2012. Genome-wide functional profiling identifies genes and processes important for zinc-limited growth of *Saccharomyces cerevisiae*. *PLoS Genet,* 8**,** e1002699.

OCHOA-ESTOPIER, A., LESAGE, J., GORRET, N. & GUILLOUET, S. E. 2011. Kinetic analysis of a *Saccharomyces cerevisiae* strain adapted for improved growth on glycerol: Implications for the development of yeast bioprocesses on glycerol. *Bioresour Technol,* 102**,** 1521-1527.

ÖSTERLUND, T., NOOKAEW, I., BORDEL, S. & NIELSEN, J. 2013. Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst Biol,* 7**,** 1-10.

OTTERSTEDT, K., LARSSON, C., BILL, R. M., STAHLBERG, A., BOLES, E., HOHMANN, S. & GUSTAFSSON, L. 2004. Switching the mode of metabolism in the yeast *Saccharomyces cerevisiae*. *EMBO Rep,* 5**,** 532-7.

PADDON, C. J., WESTFALL, P. J., PITERA, D. J., BENJAMIN, K., FISHER, K., MCPHEE, D., LEAVELL, M. D., TAI, A., MAIN, A., ENG, D., POLICHUK, D. R., TEOH, K. H., REED, D. W., TREYNOR, T., LENIHAN, J., FLECK, M., BAJAD, S., DANG, G., DENGROVE, D., DIOLA, D., DORIN, G., ELLENS, K. W., FICKES, S., GALAZZO, J., GAUCHER, S. P., GEISTLINGER, T., HENRY, R., HEPP, M., HORNING, T., IQBAL, T., JIANG, H., KIZER, L., LIEU, B., MELIS, D., MOSS, N., REGENTIN, R., SECREST, S., TSURUTA, H., VAZQUEZ, R., WESTBLADE, L. F., XU, L., YU, M., ZHANG, Y., ZHAO, L., LIEVENSE, J., COVELLO, P. S., KEASLING, J. D., REILING, K. K., RENNINGER, N. S. & NEWMAN, J. D.

2013. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature,* 496**,** 528-32.

PARTOW, S., SIEWERS, V., BJØRN, S., NIELSEN, J. & MAURY, J. 2010. Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast,* 27**,** 955-964.

PERALTA-YAHYA, P. P., OUELLET, M., CHAN, R., MUKHOPADHYAY, A., KEASLING, J. D. & LEE, T. S. 2011. Identification and microbial production of a terpene-based advanced biofuel. *Nat Commun,* 2**,** 483.

PERRONE, G. G., TAN, S.-X. & DAWES, I. W. 2008. Reactive oxygen species and yeast apoptosis. *Mol Cell Res,* 1783**,** 1354-1368.

PIŠKUR, J., ROZPĘDOWSKA, E., POLAKOVA, S., MERICO, A. & COMPAGNO, C. 2006. How did *Saccharomyces* evolve to become a good brewer? *Trends Genet,* 22**,** 183-186.

PONS, M.-N., RAJAB, A. & ENGASSER, J.-M. 1986. Influence of acetate on growth kinetics and production control of *Saccharomyces cerevisiae* on glucose and ethanol. *Appl Microbiol Biotechnol,* 24**,** 193-198.

POSTMA, E., VERDUYN, C., SCHEFFERS, W. A. & VAN DIJKEN, J. P. 1989. Enzymic analysis of the crabtree effect in glucose-limited chemostat cultures of *Saccharomyces cerevisiae*. *Appl Environ Microbiol,* 55**,** 468-77.

PRONK, J. T. 2002. Auxotrophic yeast strains in fundamental and applied research. *Appl Environ Microbiol,* 68**,** 2095-100.

PSAC. 2016. *Products made from oil and natural gas* [Online]. Available: http://www.oilandgasinfo.ca/oil-gas-you/products/ [Accessed 25.08.2016].

QIU, Y., TITTIGER, C., WICKER-THOMAS, C., LE GOFF, G., YOUNG, S., WAJNBERG, E., FRICAUX, T., TAQUET, N., BLOMQUIST, G. J. & FEYEREISEN, R. 2012. An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proc Natl Acad Sci U S A,* 109**,** 14858-63.

REISER, S. & SOMERVILLE, C. 1997. Isolation of mutants of *Acinetobacter calcoaceticus* deficient in wax ester synthesis and complementation of one mutation with a gene encoding a fatty acyl coenzyme A reductase. *J Bacteriol,* 179**,** 2969-75.

ROBERTS, G. A., GROGAN, G., GRETER, A., FLITSCH, S. L. & TURNER, N. J. 2002. Identification of a new class of cytochrome P450 from a *Rhodococcus sp*. *J Bacteriol,* 184**,** 3898-908.

ROSSI, G., SAUER, M., PORRO, D. & BRANDUARDI, P. 2010. Effect of *HXT1* and *HXT7* hexose transporter overexpression on wild-type and lactic acid producing *Saccharomyces cerevisiae* cells. *Microb Cell Fact,* 9**,** 15-15.

RUDE, M. A., BARON, T. S., BRUBAKER, S., ALIBHAI, M., DEL CARDAYRE, S. B. & SCHIRMER, A. 2011. Terminal olefin (1-alkene) biosynthesis by a novel p450 fatty acid decarboxylase from *Jeotgalicoccus* species. *Appl Environ Microbiol,* 77**,** 1718-27.

RUI, Z., LI, X., ZHU, X., LIU, J., DOMIGAN, B., BARR, I., CATE, J. H. D. & ZHANG, W. 2014. Microbial biosynthesis of medium-chain 1-alkenes by a nonheme iron oxidase. *Proc Natl Acad Sci*.

RUMBOLD, K., VAN BUIJSEN, H. J., OVERKAMP, K. M., VAN GROENESTIJN, J. W., PUNT, P. J. & VAN DER WERF, M. J. 2009. Microbial production

host selection for converting second-generation feedstocks into bioproducts. *Microb Cell Fact,* 8**,** 64.

RUNGUPHAN, W. & KEASLING, J. D. 2014. Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metab Eng,* 21**,** 103-13.

SANCHEZ, B. J. & NIELSEN, J. 2015. Genome scale models of yeast: towards standardized evaluation and consistent omic integration. *Integr Biol,* 7**,** 846-858.

SANGER, F., NICKLEN, S. & COULSON, A. R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A,* 74**,** 5463-5467.

SAUER, B. 1987. Functional expression of the cre-lox site-specific recombination system in the yeast *Saccharomyces cerevisiae*. *Mol Cell Biol,* 7**,** 2087-2096.

SAUER, M., BRANDUARDI, P., VALLI, M. & PORRO, D. 2004. Production of L-ascorbic acid by metabolically engineered *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii*. *Appl Environ Microbiol,* 70**,** 6086-91.

SCANES, K., HOHMANN, S. & PRIOR, B. 1998. Glycerol production by the yeast *Saccharomyces cerevisiae* and its relevance to wine: a review. *S Afr J Enol,* 19**,** 17-24.

SCHIESTL, R. H. & GIETZ, R. D. 1989. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr Genet,* 16**,** 339-46.

SCHIRMER, A., RUDE, M. A., LI, X., POPOVA, E. & DEL CARDAYRE, S. B. 2010. Microbial biosynthesis of alkanes. *Science,* 329**,** 559-562.

SHELL. 2008. *Shell energy scenarios to 2050* [Online]. Available: http://www.shell.com/energy-and-innovation/the-energy-future/scenarios/new-lenses-on-the-future/earlier-scenarios.html [Accessed 17.08.2016].

SICKMANN, A., REINDERS, J., WAGNER, Y., JOPPICH, C., ZAHEDI, R., MEYER, H. E., SCHONFISCH, B., PERSCHIL, I., CHACINSKA, A., GUIARD, B., REHLING, P., PFANNER, N. & MEISINGER, C. 2003. The proteome of *Saccharomyces cerevisiae* mitochondria. *Proc Natl Acad Sci U S A,* 100**,** 13207-12.

SOH, K. C., MISKOVIC, L. & HATZIMANIKATIS, V. 2012. From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks. *FEMS Yeast Res,* 12**,** 129-143.

SOLIS-ESCALANTE, D., KUIJPERS, N. G., BONGAERTS, N., BOLAT, I., BOSMAN, L., PRONK, J. T., DARAN, J. M. & DARAN-LAPUJADE, P. 2013. *amdSYM*, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res,* 13**,** 126-39.

SOLIS-ESCALANTE, D., VAN DEN BROEK, M., KUIJPERS, N. G., PRONK, J. T., BOLES, E., DARAN, J. M. & DARAN-LAPUJADE, P. 2015. The genome sequence of the popular hexose-transport-deficient *Saccharomyces cerevisiae* strain EBY.VW4000 reveals LoxP/Cre-induced translocations and gene loss. *FEMS Yeast Res,* 15.

SONNLEITNER, B. & KAPPELI, O. 1986. Growth of *Saccharomyces cerevisiae* is controlled by its limited respiratory capacity: Formulation and verification of a hypothesis. *Biotechnol Bioeng,* 28**,** 927-37.

STEEN, E. J., KANG, Y., BOKINSKY, G., HU, Z., SCHIRMER, A., MCCLURE, A., DEL CARDAYRE, S. B. & KEASLING, J. D. 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature,* 463**,** 559-62.

SWALLEY, S. E., FULGHUM, J. R. & CHAMBERS, S. P. 2006. Screening factors effecting a response in soluble protein expression: Formalized approach using design of experiments. *Anal Biochem,* 351**,** 122-127.

TANG, X., FENG, H. & CHEN, W. N. 2013. Metabolic engineering for enhanced fatty acids synthesis in *Saccharomyces cerevisiae*. *Metab Eng,* 16**,** 95-102.

TEHLIVETS, O., SCHEURINGER, K. & KOHLWEIN, S. D. 2007. Fatty acid synthesis and elongation in yeast. *BBA - Molecular and Cell Biology of Lipids,* 1771**,** 255-270.

THIELE, I. & PALSSON, B. O. 2010. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc,* 5**,** 93-121.

THOMSON, J. M., GAUCHER, E. A., BURGAN, M. F., DE KEE, D. W., LI, T., ARIS, J. P. & BENNER, S. A. 2005. Resurrecting ancestral alcohol dehydrogenases from yeast. *Nat Genet,* 37**,** 630-5.

TORIJA, M. J., ROZÈS, N., POBLET, M., GUILLAMÓN, J. M. & MAS, A. 2003. Effects of fermentation temperature on the strain population of *Saccharomyces cerevisiae*. *Int J Food Microbiol,* 80**,** 47-53.

TROTTER, P. J. 2001. The genetics of fatty acid metabolism in *Saccharomyces cerevisiae*. *Annu Rev Nutr,* 21**,** 97-119.

TYO, K. E., KOCHARIN, K. & NIELSEN, J. 2010. Toward design-based engineering of industrial microbes. *Curr Opin Microbiol,* 13**,** 255-62.

VAN DEN BERG, M. A., DE JONG-GUBBELS, P. & STEENSMA, H. Y. 1998. Transient mRNA responses in chemostat cultures as a method of defining putative regulatory elements: application to genes involved in *Saccharomyces cerevisiae* acetyl-coenzyme A metabolism. *Yeast,* 14**,** 1089-104.

VAN DIJKEN, J. P., BAUER, J., BRAMBILLA, L., DUBOC, P., FRANCOIS, J. M., GANCEDO, C., GIUSEPPIN, M. L. F., HEIJNEN, J. J., HOARE, M., LANGE, H. C., MADDEN, E. A., NIEDERBERGER, P., NIELSEN, J., PARROU, J. L., PETIT, T., PORRO, D., REUSS, M., VAN RIEL, N., RIZZI, M., STEENSMA, H. Y., VERRIPS, C. T., VINDELØV, J. & PRONK, J. T. 2000. An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb Technol,* 26**,** 706-714.

VAN DIJKEN, J. P., WEUSTHUIS, R. A. & PRONK, J. T. 1993. Kinetics of growth and sugar consumption in yeasts. *Antonie Van Leeuwenhoek,* 63**,** 343-52.

VAN HOEK, P., VAN DIJKEN, J. P. & PRONK, J. T. 2000. Regulation of fermentative capacity and levels of glycolytic enzymes in chemostat cultures of *Saccharomyces cerevisiae*. *Enzyme Microb Technol,* 26**,** 724-736.

VAN MARIS, A. J., GEERTMAN, J. M., VERMEULEN, A., GROOTHUIZEN, M. K., WINKLER, A. A., PIPER, M. D., VAN DIJKEN, J. P. & PRONK, J. T. 2004. Directed evolution of pyruvate decarboxylase-negative

Saccharomyces cerevisiae, yielding a C2-independent, glucose-tolerant, and pyruvate-hyperproducing yeast. *Appl Environ Microbiol,* 70**,** 159-66.

VANDERSLUIS, B., HESS, D. C., PESYNA, C., KRUMHOLZ, E. W., SYED, T., SZAPPANOS, B., NISLOW, C., PAPP, B., TROYANSKAYA, O. G., MYERS, C. L. & CAUDY, A. A. 2014. Broad metabolic sensitivity profiling of a prototrophic yeast deletion collection. *Genome Biol,* 15**,** R64.

VEMURI, G. N., EITEMAN, M. A., MCEWEN, J. E., OLSSON, L. & NIELSEN, J. 2007. Increasing NADH oxidation reduces overflow metabolism in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A,* 104**,** 2402-7.

VERDUYN, C., POSTMA, E., SCHEFFERS, W. A. & VAN DIJKEN, J. P. 1992. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast,* 8**,** 501-17.

VERDUYN, C., ZOMERDIJK, T. L., VAN DIJKEN, J. & SCHEFFERS, W. A. 1984. Continuous measurement of ethanol production by aerobic yeast suspensions with an enzyme electrode. *Appl Microbiol Biotechnol,* 19**,** 181-185.

VIEIRA, É. D., DA GRAÇA STUPIELLO ANDRIETTA, M. & ANDRIETTA, S. R. 2013. Yeast biomass production: a new approach in glucose-limited feeding strategy. *Braz J Microbiol,* 44**,** 551-558.

VISSER, W., SCHEFFERS, W. A., BATENBURG-VAN DER VEGTE, W. H. & VAN DIJKEN, J. P. 1990. Oxygen requirements of yeasts. *Appl Environ Microbiol,* 56**,** 3785-92.

WANG, Z., GAO, C., WANG, Q., LIANG, Q. & QI, Q. 2012. Production of pyruvate in *Saccharomyces cerevisiae* through adaptive evolution and rational cofactor metabolic engineering. *J Biochem Eng,* 67**,** 126-131.

WEHNER, E. P., RAO, E. & BRENDEL, M. 1993. Molecular structure and genetic regulation of *SFA*, a gene responsible for resistance to formaldehyde in *Saccharomyces cerevisiae*, and characterization of its protein product. *Mol Gen Genet,* 237**,** 351-8.

WESTFALL, P. J., PITERA, D. J., LENIHAN, J. R., ENG, D., WOOLARD, F. X., REGENTIN, R., HORNING, T., TSURUTA, H., MELIS, D. J., OWENS, A., FICKES, S., DIOLA, D., BENJAMIN, K. R., KEASLING, J. D., LEAVELL, M. D., MCPHEE, D. J., RENNINGER, N. S., NEWMAN, J. D. & PADDON, C. J. 2012. Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. *Proc Natl Acad Sci,* 109**,** E111-E118.

WILLIAMSON, V. M. & PAQUIN, C. E. 1987. Homology of *Saccharomyces cerevisiae* ADH4 to an iron-activated alcohol dehydrogenase from *Zymomonas mobilis*. *Mol Gen Genet,* 209**,** 374-81.

WILLS, C. 1976a. Controlling protein evolution. *Fed Proc,* 35**,** 2098-101.

WILLS, C. 1976b. Production of yeast alcohol dehydrogenase isoenzymes by selection. *Nature,* 261**,** 26-9.

WILLS, C. & JÖRNVALL, H. 1979. The two major isozymes of yeast alcohol dehydrogenase. *Eur J Biochem,* 99**,** 323-332.

WOLD, S., SJÖSTRÖM, M. & ERIKSSON, L. 2001. PLS-regression: a basic tool of chemometrics. *Chemometr Intell Lab Syst,* 58**,** 109-130.

YALCIN, S. K. & YESIM OZBAS, Z. 2008. Effects of pH and temperature on growth and glycerol production kinetics of two indigenous wine strains of *Saccharomyces cerevisiae* from Turkey. *Braz J Microbiol,* 39**,** 325-332.

YOU, K. M., ROSENFIELD, C.-L. & KNIPPLE, D. C. 2003. Ethanol tolerance in the yeast *Saccharomyces cerevisiae* is dependent on cellular oleic acid content. *Appl Environ Microbiol,* 69**,** 1499-1503.

YOUNG, E. T. & PILGRIM, D. 1985. Isolation and DNA sequence of *ADH3*, a nuclear gene encoding the mitochondrial isozyme of alcohol dehydrogenase in *Saccharomyces cerevisiae*. *Mol Cell Biol,* 5**,** 3024-34.

YU, K. O., KIM, S. W. & HAN, S. O. 2010. Engineering of glycerol utilization pathway for ethanol production by *Saccharomyces cerevisiae*. *Bioresour Technol,* 101**,** 4157-4161.

YUAN, D. S. 2000. Zinc-regulated genes in *Saccharomyces cerevisiae* revealed by transposon tagging. *Genetics,* 156**,** 45-58.

ZIMMERMANN, F. K. & ENTIAN, K. D. 1997. *Yeast Sugar Metabolism,* Lancaster, Pennsylvania, USA, Technomic publication.

ZOMORRODI, A. R. & MARANAS, C. D. 2010. Improving the iMM904 *S. cerevisiae* metabolic model using essentiality and synthetic lethality data. *BMC Syst Biol,* 4**,** 178.