A proteomics and metabolomics analysis of the mechanism of lipid accumulation in yeast

Shi, Jiahua

2015

Shi, J. (2015). A proteomics and metabolomics analysis of the mechanism of lipid accumulation in yeast. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/62924>

<https://doi.org/10.32657/10356/62924>

NANYANG
TECHNOLOGICAL
UNIVERSITY

**A PROTEOMICS AND METABOLOMICS
ANALYSIS OF THE MECHANISM OF LIPID
ACCUMULATION IN YEAST**

SHI JIAHUA

SCHOOL OF CHEMICAL AND BIOMEDICAL ENGINEERING

2015

**A PROTEOMICS AND METABOLOMICS
ANALYSIS OF THE MECHANISM OF LIPID
ACCUMULATION IN YEAST**

SHI JIAHUA

School of Chemical and Biomedical Engineering

A thesis submitted to the Nanyang Technological University

in partial fulfilment of the requirement for the degree of

Doctor of Philosophy

2015

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and deep appreciation to the following people, without their patient support and guidance, this dissertation would not have been completed.

My sincere and deepest gratitude goes first and foremost to my supervisor, Professor William Chen Wei Ning, who offers me insightful guidance, precious encouragement and fully trust and support in the process of my Ph. D. program. His advisable foresight and passion on research, as well as the positive attitude towards life really set an impressive example for me. He teaches me how to appreciate what we have been given and being thankful. To me, he is not only a supervisor for my Ph. D. study, but also a lifetime advisor.

Secondly, I would thank School of Chemical and Biomedical Engineering and CRP Program, which give me an excellent study and research environment to explore this prospective topic and the opportunity to finish my Ph. D. study with full research scholarship.

Thirdly, I would like to express my deep appreciation to my helpful group members: Dr. Feng Huixing and Dr. Zhang Jianhua for their continuous guidance on the experiment and instrument; Ms Chen Liwei, Tang Xiaoling, Li Xiang, Zhao Guili, Jaslyn Lee, Toh Shi Hui and Mr. Tan Kee Yang for their warm help, useful discussion and encouragement during my lab work.

Also, I am grateful to the professional officer Dr Ong Teng Teng for her patient instrumental training and guidance. Technical executives, Ms. Tan Lay Yian and Ms. Jessica Gan Chew Mei, are also appreciated for their guidance and assistance during chemical purchasing and characterization.

Finally I wish to express my gratitude to my beloved family for their fully support, warm encouragement and continuous confidence and loving consideration in me all these years. Special thanks to Mr. Du Bo for his continuous support, understanding, company and love for me, which has enlightened my life.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
TABLE OF CONTENTS	iii
LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xvii
SUMMARY	xxi
CHAPTER 1 INTRODUCTION	1
1.1 Biofuel.....	1
1.2 Single cell oils for biofuel production	2
1.3 Yeast strains with high lipid content productivity	5
1.4 Lipid accumulation mechanism in yeast.....	8
1.4.1 Intermediate metabolism in the oleaginous microorganisms.....	8
1.4.2 The lipid synthesis of non-oleaginous microorganisms.....	10
1.4.3 The de novo lipid accumulation mechanism of oleaginous yeast...	10
1.4.4 Importance of several crucial enzymes in oleaginous yeast	13
1.5 Protein profiling based on LC-MS.....	14
1.5.1 Introduction about proteomics	14
1.5.2 Basic introduction of LC-MS.....	15

1.5.3 Quantification method based on iTRAQ	20
1.6 Metabolite profiling based on GC-MS	24
1.6.1 Introduction about metabolomics	24
1.6.2 Comparison of analytical methods for metabolites	25
1.6.3 General background of GC-MS.....	28
1.7 Proteomics and metabolomics study in yeast	31
CHAPTER 2 OBJECTIVES	32
CHAPTER 3 MATERIALS AND METHODS	34
3.1 Yeast strains and growth conditions.....	34
3.1.1 <i>Saccharomyces cerevisiae</i> and its culture condition	34
3.1.2 <i>Rhodospiridium toruloides</i> and its culture condition	34
3.1.3 <i>Cryptococcus albidus</i> and its culture condition.....	35
3.2 Growth measurement.....	36
3.2.1 Growth curve test	36
3.2.2 Dry weight test	36
3.2.3 Residual glucose detection	37
3.2.3.1 Standard curve test of glucose	37
3.2.3.2 Sample preparation for residual glucose determination	37
3.2.4 Lipid content measurement	38
3.3 Cell lysis, protein digestion and labelling with iTRAQ reagents	39
3.4 On-line 2D nano-LC-MS/MS analysis for protein.....	43

3.5 Protein identification and data analysis	45
3.6 Western blot analysis	46
3.6.1 Cell Extracts Preparation	46
3.6.2 Gel Electrophoresis	46
3.6.3 Transfer	47
3.6.4 Blocking	48
3.6.5 Detection	48
3.7 Real-time RT-PCR.....	50
3.7.1 RNA extraction	50
3.7.2 RNA quantitation	51
3.7.3 Real-time RT-PCR.....	51
3.8 Cell lysis and metabolites extraction	53
3.9 Derivatization of soluble metabolites	54
3.10 Derivatization for fatty acids	54
3.11 GC-MS analysis	54
3.12 Metabolites detection and quantification	55
3.12 Statistical analysis	56
CHAPTER 4 BIOSYNTHESIS OF LIPIDS IN DIFFERENT YEAST STRAINS	58
4.1 Introduction.....	58
4.2 Experimental procedures	60
4.3 Results and discussion	61

4.3.1 Substrate inhibition study of oleaginous yeast	61
4.3.2 Cell density optimization.....	64
4.3.3 Lipid content comparison	66
4.4 Chapter conclusions.....	70
CHAPTER 5 COMPARATIVE PROTEOMICS PROFILE OF LIPID- CUMULATING OLEAGINOUS YEAST: AN ITRAQ-COUPLED 2-D LC- MS/MS ANALYSIS	71
5.1 Introduction	71
5.2 Experimental procedures	73
5.3 Results and discussion	75
5.3.1 ITRAQ analysis and protein identification.....	75
5.3.2 Differentially expressed proteins within comparison.....	79
5.3.3 Western blot analysis of differentially expressed proteins	89
5.3.4 Real-time RT-PCR based investigation for mRNA level determination	92
5.3.5 Influence of different culture conditions	94
5.4 Chapter conclusions.....	95
CHAPTER 6 COMPARATIVE METABOLIC PROFILING OF YEAST STRAINS DURING DIFFERENT GROWTH PHASES IN LIPID PRODUCTION	96
6.1 Introduction	96
6.2 Experimental procedures	97
6.3 Results and discussion.....	98

6.3.1 Metabolite profiling of each yeast strain during different lipid accumulation phases	98
6.3.1.1 Time-course based metabolite profiling of <i>S. cerevisiae</i>	98
6.3.1.3 Time-course based metabolite profiling of <i>R. toruloides</i>	105
6.3.2 Comparison of metabolite changes among <i>S. cerevisiae</i> , <i>C. albidus</i> and <i>R. toruloides</i> at each stage.....	108
6.3.3 Analysis of changes of various metabolites along comparison	110
6.3.4 Cellular lipid composition analysis of yeast strains by GC-MS ...	112
6.3.5 Mechanism of fatty acid composition changes in different yeast strains	114
6.4 Chapter conclusions	115
CHAPTER 7 CONCLUSIONS AND FUTURE DIRECTIONS	117
7.1 Conclusions.....	117
7.1.1 Background and motivation.....	117
7.1.2 Biosynthesis of lipids in different yeast strains	118
7.1.3 Proteomic analysis of yeast strains at various lipid accumulation stages.....	119
7.1.4 Metabolic profiling of yeast strains at different lipid accumulation stages.....	120
7.1.5 Overall summary of dynamic changes in different yeast strains ..	123
7.1.6 Limitations of the results	125
7.2 Future directions	127

LIST OF PUBLICATIONS.....	129
REFERENCES	130
APPENDIX	140

LIST OF TABLES

Table 1.1 Single cell oil comparison as feedstock for biofuel production.

Table 1.2 Fatty acid composition of lipid produced by various yeast strains growing on sugar substrates in culture conditions favouring the accumulation of microbial lipid.

Table 3.1 Dilution ranges of different primary antibodies.

Table 3.2 List of primer sequences for real-time RT-PCR.

Table 5.1 Experimental description for time course based iTRAQ labelling for yeast strains.

Table 6.1 Fold change of metabolites in *C. albidus* (C) and *R. toruloides* (R) comparing with *S. cerevisiae* (S) at different lipid accumulation stages.

Table 6.2 Fatty acid composition of lipid produced by yeast strains at different lipid accumulation stages.

LIST OF FIGURES

CHAPTER 1

Figure 1.1 Intermediate metabolism pathways in oleaginous yeast. Enzymes: Ac, acotinase; ACC, acetyl-CoA carboxylase; ACL, ATP-citrate lyase; FAS, fatty acid synthetase; ICDH, iso-citrate dehydrogenase; MD_c, malate dehydrogenase (cytoplasmic); MD_m, malate dehydrogenase (mitochondrial); PD, pyruvate kinase.

Figure 1.2 Consecutive steps leading to the de novo lipid biosynthesis in oleaginous microorganisms.

Figure 1.3 Schematic diagram of HPLC/MS machine.

Figure 1.4 Schematic of ESI and MALDI ionization sources.

Figure 1.5 Schematic of four types of mass analyzers. (A. Quadrupole; B. Ion trap; C. TOF; D. FT-ICR).

Figure 1.6 Structure of Q-TOF mass analyzer.

Figure 1.7 Working Mechanism of iTRAQ Reagent. A, B and C represent three different samples.

Figure 1.8 Protein profiling workflow design.

Figure 1.9 A schematic diagram of the “Omics” cascade.

Figure 1.10 Summary of metabolites compatible with different separation methods (LC-MS and GC-MS).

CHAPTER 3

Figure 3.1 Layout for iTRAQ-based quantitative proteomics experimental design. Proteins in yeast strains with low, medium and high lipid accumulation abilities were extracted, quantified, reduced and digested. Respective iTRAQ reagent 114, 115 and 116 were used to label *S. cerevisiae*, *C. albidus* and *R. toruloides*. The samples were then combined into one and injected into online 2D LC/MS system and differences in proteins levels were identified by analysing the raw data by software.

Figure 3.2 Workflow of on-line 2-D nano-LC. (a) The peptides eluted through SCX column and trapped onto enrichment column. (b) HPLC-chip was changed to analytical mode and previously trapped peptides were eluted and analyzed by analytical column.

Figure 3.3 Layout of semidry transfer system.

CHAPTER 4

Figure 4.1 Effect of initial glucose concentrations in *C. albidus*. (A: Effect on cell growth; B: Effect on biomass and lipid accumulation).

Figure 4.2 Effect of initial glucose concentrations in *R. toruloides*. (A: Effect on cell growth; B: Effect on biomass and lipid accumulation).

Figure 4.3 Effects of cell density on lipid accumulation in oleaginous yeast. (A: *C. albidus*; B: *R. toruloides*).

Figure 4.4 Growth curve of *S. cerevisiae*, *R. toruloides* and *C. albidus* in YEPD medium. Optical density was measured at wavelength of 600 nm by taking 1 ml culture medium and diluting into proper concentrations. Data from three replications were analyzed to calculate means and standard errors, indicated by thin lines.

Figure 4.5 Substance level dynamic change of oleaginous yeast. (A: *C. albidus*; B: *R. toruloides*).

CHAPTER 5

Figure 5.1 Total intensity chromatogram of peptides of early lipid accumulation stage sample. The ‘X’ axis is the acquisition time, and the ‘Y’ axis is the chromatogram of peptides. Samples were eluted with gradient diluted ammonium formate, and totally ten run were accomplished. Each color represented the chromatogram of peptides eluted at respective concentration of ammonium formate. Black, 20 mM; red, 40, mM; dark green, 60 mM; dark blue, 80 mM; purple, 100 mM; brown, 300 mM; light green, 500 mM; light blue, 500 mM; orange, 1000 mM; pink, 1000 mM.

Figure 5.2 Representative MS spectrum indicating four peptide precursors (524.74, 540, 579.34 and 655.66) marked by red diamond.

Figure 5.3 One representative peptide fragmentation spectrum of pyruvate kinase: GVEVVL PVDFIIADAFSADANTK.

Figure 5.4 Functional distribution of proteins identified during each lipid accumulation stage.

Figure 5.5 Fold change of proteins of different functional distribution at early, middle and late lipid accumulation stages. Black color represented that the protein was not identified at that stage.

Figure 5.6 Key protein changes in yeast strains mapped onto metabolic pathways. Red color represented up-regulation, whereas green color represented down-regulation in oleaginous yeast strains as compared to *S. cerevisiae*. The blue arrow represented the general flow of metabolic flux. Abbreviations for the enzymes were GLK, glucokinase; PGI, glucose-6-phosphate isomerase; FBA, fructose-bisphosphate aldolase; TPI, triosephosphate isomerase; TDH, glyceraldehyde-3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; GPM, phosphoglycerate mutase; ENO, enolase; PYK, pyruvate kinase; PDC, pyruvate decarboxylase isozyme; ADH, alcohol dehydrogenase; CIT, citrate synthase; IDH, isocitrate dehydrogenase [NAD] subunit; ICL, isocitrate lyase; MDH, malate dehydrogenase; GND, 6-phosphogluconate dehydrogenase; AAC, ADP/ATP carrier protein.

Figure 5.7 Western blot result of protein expressions in *S. cerevisiae* (S), *C. albidus* (C) and *R. toruloides* (R) at early lipid accumulation stage. (A) Protein expression changes. (B) Quantification of protein level changes based on western blot analysis. The ratio between *C. albidus* or *R. toruloides* and *S. cerevisiae* was shown.

Figure 5.8 Western blot result of protein expressions in *S. cerevisiae* (S), *C. albidus* (C) and *R. toruloides* (R) at middle lipid accumulation stage. (A) Protein

expression changes. (B) Quantification of protein level changes based on western blot analysis. The ratio between *C. albidus* or *R. toruloides* and *S. cerevisiae* was shown.

Figure 5.9 Western blot result of protein expressions in *S. cerevisiae* (S), *C. albidus* (C) and *R. toruloides* (R) at late lipid accumulation stage. (A) Protein expression changes. (B) Quantification of protein level changes based on western blot analysis. The ratio between *C. albidus* or *R. toruloides* and *S. cerevisiae* was shown.

Figure 5.10 Real-time RT-PCR analysis of mRNA levels of differentially expressed proteins in *S. cerevisiae*, *C. albidus* and *R. toruloides*. (A: Early lipid accumulation stage; B: Middle stage; C: Late stage).

CHAPTER 6

Figure 6.1 Differential expression levels of intracellular metabolites at different lipid accumulation stages of *S. cerevisiae*. The ratios were normalized to the values obtained for the early stage sample.

Figure 6.2 Principal component analysis (PCA) of the GC-MS spectra from intracellular metabolites in *S. cerevisiae*. A: PCA scores plot. Samples with low (L), medium (M) and high (H) lipid content were included; t1 and t2 represented scores. B: PCA loading plot; p1 and p2 represented loadings.

Figure 6.3 Differential expression levels of intracellular metabolites at different lipid accumulation stages of *C. albidus*. The ratios were normalized to the values obtained for the early stage sample.

Figure 6.4 Principal component analysis (PCA) of the GC-MS spectra from intracellular metabolites in *C. albidus*. A: PCA scores plot. Samples with low (L), medium (M) and high (H) lipid content were included; t1 and t2 represented scores. B: PCA loading plot; p1 and p2 represented loadings.

Figure 6.5 Differential expression levels of intracellular metabolites at different lipid accumulation stages of *R. toruloides*. The ratios were normalized to the values obtained for the early stage sample.

Figure 6.6 Principal component analysis (PCA) of the GC-MS spectra from intracellular metabolites in *R. toruloides*. A: PCA scores plot. Samples with low (L), medium (M) and high (H) lipid content were included; t1 and t2 represented scores. B: PCA loading plot; p1 and p2 represented loadings.

CHAPTER 7

Figure 7.1 Differentially expressed proteins in three yeast strains at different lipid accumulation stages. *S. cerevisiae* was used as control sample, comparison of protein level changes was carried out in *C. albidus* and *R. toruloides* against *S. cerevisiae*. The red arrow represented up-regulation of protein levels; the blue arrow indicated the down-regulation of protein levels.

Figure 7.2 Schematic diagram summarizing the metabolic alterations in three yeast strains at different lipid accumulation stages. For soluble intracellular metabolites, *S. cerevisiae* was used as control sample, comparison of metabolite level

changes was carried out in *C. albidus* and *R. toruloides* against *S. cerevisiae*. The red arrow represented up-regulation; the blue arrow indicated down-regulation. For fatty acid analysis, composition differences were presented respectively, with majority labelled by red color and minority labelled by blue color.

Figure 7.3 Schematic diagram summarizing the dynamic changes in yeast strains with different lipid accumulation abilities. Red arrow indicated up-regulation, blue arrow represented down-regulation. Higher changes were revealed by two arrows.

LIST OF ABBREVIATIONS

AMP	Adenosine monophosphate
APS	Ammonium persulfate
ATCC	American Type Culture Collection
ATP	Adenosine triphosphate
ATP-CL	ATP citrate lyase
BSA	Bovine serum albumin
CE	Capillary electrophoresis
CE/MS	Capillary electrophoresis coupled with mass spectrometry
CID	Collision-induced dissociation
CP	Cloud point
2-D	Two-dimensional
ECL	Enhanced chemiluminescence
ESI	Electrospray ionization
EUROSCARF	European <i>Saccharomyces cerevisiae</i> Archive for Functional analysis
FAME	Fatty acid methyl ester

FAS	Fatty acid synthase
FT-ICR	Fourier transform ion cyclotron resonance
FT-IR	Fourier transform infrared spectroscopy
GC	Gas chromatography
GC/MS	Gas chromatography mass spectrometry
HPLC	High-performance liquid chromatography
HRP	Horseradish peroxidase
IMP	Inosine monophosphate
iTRAQ	Isobaric tag for relative and absolute quantitation
LC	Liquid chromatography
LC/MS	Liquid chromatography mass spectrometry
LC-MS/MS	Liquid chromatography with tandem mass spectrometry
LC-NMR	Liquid chromatography coupled with nuclear magnetic resonance
MALDI-TOF	Matrix-assisted laser desorption/ionization time of flight
ME	Malic enzyme
MMTS	Methyl methanethiosulfonate
MS	Mass spectrometry

MSTFA	N-methyl-N-(trimethylsilyl) trifluoroacetamide
NHS	N-Hydroxysuccinimide
NIST	National Institute of Standards and Technology
NMR	Nuclear magnetic resonance
OD	Optical density
PLS-DA	Partial least squares-discrimination analysis
PD	Pyruvate-dehydrogenase
PBS	Phosphate-buffered saline
PBST	Phosphate buffered saline-tween 20
PCA	Principal Component Analysis
PCR	Polymerase chain reaction
PP	Pour point
PVDF	Polyvinylidene fluoride
QIT	Quadrupole ion trap
QQQ	Triple quadrupole
Q-TOF	Quadrupole time-of-flight
RP	Reverse-phase

RT-PCR	Reverse transcription polymerase chain reaction
SCO	Single cell oil
SCX	Strong cation-exchange
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE	Standard error
TCA	Trichloroacetic acid
TCA cycle	Tricarboxylic acid cycle
TCEP	Tris-(2-carboxyethyl) phosphine
TEAB	Triethylammonium bicarbonate
TEMED	N, N, N', N' - tetramethyl-ethylenediamine
TMCS	Trimethylchlorosilane
TOF	Time-of-flight
TOF/TOF	Time-of-flight/time-of-flight

SUMMARY

Energy shortage and environmental concerns have become urgent problems all over the world. Microbial production of high-energy fuels represents one of the viable options for sustainable energy supply due to its considerable advantages, such as being renewable, biodegradable and nontoxic. Oleaginous yeast strains are capable of accumulating lipid with a yield of more than 20% in their cells. Accumulation of intracellular lipid in oleaginous yeast cells has been studied for providing an alternative supply for energy, biofuel. Numerous studies have been conducted on increasing lipid content in oleaginous yeasts. However, few explore the mechanism of the high lipid accumulation ability of oleaginous yeast strains. A comprehensive understanding and exploration of lipid accumulation mechanism in oleaginous yeast remains insufficient. Using iTRAQ-coupled two-dimensional liquid chromatography mass spectrometry (2D LC-MS/MS) analysis and gas chromatography-mass spectrometry, we report here a time course comparative study of protein profile and metabolite profile of oleaginous yeast strains. The development of proteomics and metabolomics platforms will greatly facilitate understanding of lipid accumulation mechanisms in oleaginous yeast.

Two oleaginous yeast strains and one non-oleaginous yeast strain were selected as our investigated subjects. *Rhodospiridium toruloides* and *Cryptococcus albidus* are reported oleaginous yeast strains, producing up to 65.2% and 46.3% lipid of their cell dry weight, respectively. Non-oleaginous yeast *Saccharomyces cerevisiae*, whose lipid content is merely less than 15%, is selected as a control for comparison. Each

strain was cultured in optimized medium and collected at early, middle and late lipid accumulation stages. Two dimensional LC-MS/MS approach has been applied for protein profiling together with isobaric tag for relative and absolute quantitation (iTRAQ) labelling method. 132 proteins were identified when three yeast strains were all at early lipid accumulation stage; 122 and 116 proteins were found respectively within cells of three strains collected at middle and late lipid accumulation stages. Significant up-regulation or down-regulation synthesis of proteins was observed among yeast strains. Essential proteins correlated to lipid synthesis and regulation were detected. For example, the up-regulated proteins, ADP, ATP carrier protein 2 (AAC2p), ATP synthase subunit beta, mitochondrial (ATP2p), are closely relevant with the energy production and transportation, providing steady supplement of energy of lipid synthesis.

To complement our proteomics analysis, the metabolomics analysis was carried out to establish the changes in metabolites which are truly reflective of the physiological state of cells. The intracellular metabolites of cells during different lipid accumulation phases were compared using gas chromatography-mass spectrometry. Fifteen metabolites exhibited marked differences along with lipid accumulation in *S. cerevisiae*. Thirteen metabolites displayed different expression levels in *C. albidus* along lipid content change from 10.73% to 27.74% of its cell dry weight. 10 metabolites were identified to have remarkable changes for *R. toruloides* when lipid content accumulated from 14% to 45%. When time-course based comparison among the three strains was carried out, 10, 14 and 9 differentially expressed metabolites were found in the cells at early, middle and late lipid accumulating stages,

respectively. Significant changes in metabolite levels correlated to lipid synthesis were identified, such as lysine, succinate and citrate.

Our results highlight the potential of using proteomics and metabolomics platforms to study lipid accumulation mechanisms in oleaginous yeast. Comparison within cells between non-oleaginous yeast and oleaginous yeast strains demonstrates characteristic proteomic and metabolic signatures. Taken together, these findings demonstrate the feasibility of using LC-MS-based proteomics and GC-MS-based metabolomics platforms to understand lipid accumulation mechanisms. It is to be hoped the platforms developed in this study can also be applied to the investigation of other microorganisms and would contribute to further genetic engineering for higher lipid accumulation in yeast.

CHAPTER 1

INTRODUCTION

1.1 Biofuel

Being the key principle in the management of natural resource, sustainability involves the following interdependent factors: efficiency of operation, minimized environmental impact and socio-economic considerations. Not only the depleting of reserves all over the world, the greenhouse gas emissions also contribute to the undesirability of reliance on the fossil fuel energy resources [1]. As a nontoxic, biodegradable and renewable fuel, biofuels can contribute to minimizing fossil fuel burning and CO₂ production, and has attracted considerable attention in the past few years [2,3]. Biofuel can be produced from biological waste and renewable resources, which is of great importance as it could offer an alternative way to economically supply oils as compared to traditional fossil fuels [4].

The ‘first generation’ or conventional biofuels (i.e. biodiesel, bioethanol and biogas) are mainly made from food crops, like cereals, oil seeds and sugar crops. An increase in the need for first generation biofuels has been observed during the past few years, reaching 22.5 billion litres in 2012 [5,6]. However, together with the changes in government support policies, this increase has raised the price of fossil fuel and various foodstuffs, and increased biofuel generation competes for the water and land used for food production. This situation has led to the emergence of ‘second generation’ of biofuels, which are derived from non-food biomass, such as

lignocellulosic biomass or woody crops, agricultural residues and non-edible parts of plants. Second generation biofuels could help solve the problems encountered with first generation fuels and also provide a larger proportion of global fuel supply, with greater benefits for environment. It is reported that by 2022, the renewable fuels estimated to be used in the United States will reach 36 billion gallons, and 16 billion gallons are supposed to be generated from cellulosic sources and residual non-food parts of current crops [7]. The problem that second generation biofuel encounters is the difficulty of extracting useful feedstock from cellulosic material, and the technology to change it to biofuels is at an early stage of exploration. Further research has been undertaken and oleaginous microorganisms are potential candidates to process these agro-industrial residues or wastes to produce lipids, to be further converted into biofuels [8,9].

1.2 Single cell oils for biofuel production

Single cell oils (SCOs) are oils and fats produced from oleaginous microorganisms, and they have attracted considerable industrial interests as they viable feedstock to produce fatty acid methyl esters (FAMES) [10]. “Oleaginous microorganisms” refers to microorganisms that are able to accumulate lipids to a yield of more than 20% w/w of their cell dry weight. Another advantage of SCO is that it can be produced from residual biomass as the carbon source, which reduces the competition for resources with land and food production. Microalgae, bacteria, yeast and fungi are the main oleaginous microorganisms being investigated for biofuel production [2]. Comparisons between these SCO sources are listed in Table 1.1 below [11] (page 3).

Table 1.1 Single cell oil comparison as feedstock for biofuel production (permission from ref. 11 was obtained from publisher to use this table) [11].

Microorganism	Lipid content (% dry weight)	Main lipid composition	Advantages	Disadvantages
Microalgae	20-70	Fatty acid	High lipid accumulation yield	Expensive, complicated scale-up
		Free fatty acid	Some species can produce hydrocarbons directly	Necessity of light
		Hydrocarbon	Possibility of heterotrophic growth	Low growth rate
		Unusual lipid	Convert carbon dioxide to potential biofuel	
Bacteria	18-70		Can provide several different types of renewable biofuels, including methane produced by anaerobic digestion of the algal biomass	
		Free fatty acid	High growth rate	Production of complex lipid
		Unusual lipid	Easy culture method	Few species able to accumulate triacylglycerol
		Wax esters	Relatively easy-to-use biological engineering technology, genetic engineering and metabolic engineering to modify bacteria performance to improve oil accumulation	
Yeast	25-70	Triacylglycerols	Can be grown in conventional bioreactor	
		Polyhydroxyalkanoates	High growth rate	Need organic carbon sources
		Triacylglycerols	High lipid accumulation	
		Free fatty acid	Easy culture method	
Fungi	25		Can be grown in conventional bioreactor	
		Unusual lipid	High growth rate	Need organic carbon sources
		Fatty acid	Easy culture method	
			Can be grown in conventional bioreactor	

Microalgae are able to use carbon dioxide and sunlight as their carbon source and energy to accumulate oils under specific conditions. The yield of oil production depends on the species of microalgae and also the culture conditions, for example nitrogen sources, pH, and salinity. The degree of nitrogen limitation and intensity of light are the key parameters in oil production [12].

Bacteria are capable of accumulating specialized lipids, such as fatty alcohols, polyhydroxybutyrate, wax esters, hydrocarbons, triacylglycerols and peptidolipids [13]. Bacteria present advantages in producing biodiesel when compared with microalgae, as they have a higher growth rate and culturing methods are relatively simple.

Under proper culture conditions, several yeast species also accumulate lipids, which mainly consist of triacylglycerols and free fatty acids. Adjusted C/N ratios, pH, oxygen and temperature with proper selection of nitrogen resources and inorganic salt can be manipulated to alter lipid accumulation in yeast strains [2].

Although various species of fungi are reported to accumulate lipids, such as *Mucor circinelloides*, *Aspergillus oryzae*, *Humicola lanuginosa* and *Mortierella vinacea*, few reports describe the use of these fungal species for biodiesel production. Most reports describe the use of these fungi for the production of special lipids, such as eicosapentaenoic acid, arachidonic acid, decosahexaenoic acid and γ -linolenic acid [2,14].

1.3 Yeast strains with high lipid content productivity

After comprehensive comparison of lipid synthesis among different microorganisms, yeast exhibits considerable advantages in lipid production, such as high growth rate, high ability of lipid accumulation, easy culture methods and ability to be grown in the conventional bioreactors [7]. The majority of lipids produced by oleaginous yeast are triacylglycerols, which can be easily converted into fatty acids. The rest of the lipid components produced consist of a small quantity of free fatty acid together with other neutral lipids, sterols and polar fractions [8]. Apart from producing high levels of lipids, the variety of substrates that oleaginous yeasts can utilize is also an enormous advantage. Carbohydrates, hydrocarbons and waste materials can be effectively utilized by oleaginous yeast and converted into lipids [9-11]. Hence, all these properties make oleaginous yeast more useful for lipid production as compared to other oleaginous microorganisms.

Numerous yeast species are reported to be oleaginous yeasts, such as *Rhodospiridium toruloides* [12], *Trichosporon fermentans* [13], *Cryptococcus albidus* [14], *Yarrowia lipolytica* [15] and *Lipomyces starkeyi* [16]. They are able to accumulate lipids ranging from 25% to 70% of their dry weight under appropriate culture conditions using optimized carbon and nitrogen sources, a relatively high C/N ratio, proper temperature, pH, dissolved oxygen concentration and inorganic salts concentration [7].

Lipid accumulation is favoured when a higher C/N ratio is utilized in cultivation. An increase of lipid content from 18 to 46% was observed when the C/N ratio was raised from 25 to 70% [15]. Different organic nitrogen sources also affect oil production

[16]. Although both organic and inorganic nitrogen sources can be utilized to culture yeast, cell growth is favoured over lipid production when inorganic nitrogen sources are used, whereas organic nitrogen sources favour oil production [17]. Trace metal ions are also reported to influence biomass and lipid accumulation. Increase in oil content could be induced with optimized concentrations of Mn^{2+} , Mg^{2+} , Cu^{2+} , Zn^{2+} , and Ca^{2+} as reported by Y. Li [18].

The composition of synthesized lipid in different oleaginous yeast strains varies though the majority of lipid in all yeasts is composed of C_{16} and C_{18} fatty acids, among which palmitic acid ($\text{C}_{16:0}$) and oleic acid ($\Delta^9\text{C}_{18:1}$) are found in higher percentages, followed by linoleic acid ($\Delta^9, ^{12}\text{C}_{18:2}$), stearic acid ($\text{C}_{18:0}$) and palmitoleic acid ($\Delta^9\text{C}_{16:1}$). The fatty acid composition of lipid synthesized by different oleaginous yeast strains is listed in Table 1.2 (page 7).

Table 1.2 Fatty acid composition of lipid produced by various yeast strains growing on sugar substrates in culture conditions favouring the accumulation of microbial lipid (permission from ref. 9 was obtained from publisher to use this table) [9].

Strain	Lipid (% w/w)	C16:0	C16:1	C18:0	C18:1	C18:2	C18:3
<i>Candida sp. 107</i> ^{a)}	37.1	37	1	14	36	7	T.
<i>Candida sp. 107</i>	n.r.	28	n.r.	8	41	17	17
<i>Candida sp.</i>	40.3	23	13	3	54	5	2
<i>Rhodotorula gracilis</i>	41.0	21	T.	13	51	11	3
<i>Candida curvata</i> ^{b)}	29.1	36	T.	14	40	7	T.
<i>Candida curvata</i> ^{b)}	28.0	37	T.	10	44	6	T.
<i>Apiotrichum curvarum</i> ^{b)}	31.0	34	T.	10	43	7	2
<i>Cryptococcus curvatus</i> ^{b)}	38.0	24	T.	10	46	9	6
<i>Cryptococcus curvatus</i> ^{b)}	25.0	18	T.	16	50	16	T.
<i>Cryptococcus curvatus</i> ^{b)}	50.0	31	-	22	42	1	n.r.
<i>Cryptococcus albidus</i>	46.3	14	T.	9	53	18	2
<i>Cryptococcus albidus</i>	n.r.	20	n.r.	1	59	6	6
<i>Yarrowia lipolytica</i>	43.2	15	2	1	47	21	3
<i>Yarrowia lipolytica</i>	30.7	12	11	9	57	11	T.
<i>Yarrowia lipolytica</i> ^{a)}	22.3	13	17	6	55	7	n.r.
<i>Rhodospiridium toruloides</i>	67.5	20	1	15	47	13	3
<i>Rhodospiridium toruloides</i>	65.2	34	T.	13	48	1	T.
<i>Rhodospiridium toruloides</i>	62.1	26	2	5	62	3	T.
<i>Rhodospiridium toruloides</i>	55.6	43	T.	16	35	2	T.
<i>Lipomyces starkeyi</i>	68.0	56	2	14	26	T.	T.
<i>Lipomyces starkeyi</i>	61.5	37	4	6	49	1	T.
<i>Rhodotorula mucilaginosa</i>	48.6	22	2	9	55	11	T.
<i>Trichosporon capitatum</i>	37.6	12	1	2	74	9	n.r.
<i>Rhodotorula sp.</i>	22.0	22	1	7	56	12	n.r.
<i>Candida oleophila</i>	15.3	13	3	7	66	11	n.r.

T. <0.5%, w/w; n.r.: not reported.

^{a)} Representation of the neutral fraction of microbial lipids produced.

^{b)} *Cryptococcus curvatus* was formerly *Candida curvata* and then *Apiotrichum curvatum*; thus these microorganisms in fact are the same species.

1.4 Lipid accumulation mechanism in yeast

1.4.1 Intermediate metabolism in the oleaginous microorganisms

Lipids are synthesized in all microorganisms, whereas only oleaginous microorganisms accumulate significant quantities of lipids, to an extent of more than 20% of their cell dry biomass. It has been suggested that instead of possessing a hyperactive system to synthesize fatty acids, oleaginous microorganisms produce large quantities of acetyl-CoA, which is the crucial component of fatty acid synthesis [19,20]. The intermediate metabolism and lipid biosynthesis pathways of oleaginous yeast strains are shown in Figure 1.1 (page 9). A-C pathways represents the pyruvate transportation from cytoplasm to mitochondrion and transportation for malate inversely. D stands for the citrate and malate transportation between cytoplasm and mitochondrion.

The generation of acetyl-CoA is catalyzed by the enzyme Acc1p (acetyl-CoA carboxylase), which is a biotin-dependent enzyme [21]. Acc1p catalyzes the limiting step and first reaction in the biosynthesis of fatty acid, which is described as follows:

Figure 1.1 Intermediate metabolism pathways in oleaginous yeast (permission from ref. 9 was obtained from publisher to use this figure)

[9]. Enzymes: Ac, acotinase; ACC, acetyl-CoA carboxylase; ACL, ATP-citrate lyase; FAS, fatty acid synthetase; ICDH, iso-citrate dehydrogenase; MD_c, malate dehydrogenase (cytoplasmic); MD_m, malate dehydrogenase (mitochondrial); PD, pyruvate kinase.

When the generation of acetyl-CoA and malonyl-CoA is finished, fatty acid synthesis is carried out with the catalysis of fatty acid synthase (FAS), which is a multi-enzymatic complex. This series of reaction can be described as follows [19,22]:

1.4.2 The lipid synthesis of non-oleaginous microorganisms

In non-oleaginous microorganisms, glycolysis produces pyruvic acid as the net product, which is concentrated within the mitochondrial matrix after passing through the mitochondrial membrane. Pyruvic acid is a substrate for further catalysis by pyruvate-dehydrogenase (PD) and produces acetyl-CoA. The generation of acetyl-CoA only comes from the pyruvate-acetaldehyde-acetate pathway. Some of the acetyl-CoA will enter citric acid cycle or the Krebs cycle, whereas some acetyl-CoA will be used to synthesize acetyl-carnitine by carnitine-acyl-transferase and transported into cytoplasm to further enhance the cellular fatty acid biosynthesis [19-21,23]. However, in the non-oleaginous microorganisms, only minimal amounts of acetyl-carnitine is able to go through mitochondrial matrix for lipid synthesis in the cytoplasm [20,23,24].

1.4.3 The de novo lipid accumulation mechanism of oleaginous yeast

In the oleaginous microorganisms, apart from the pyruvate-acetaldehyde-acetate pathway, acetyl-CoA is produced from excess citric acid in the cytoplasm. The excess citric acid from citric acid cycle accumulates in mitochondria and is then exported into cytoplasm, where it is then cleaved into acetyl-CoA and becomes the

precursor of intracellular fatty acids [20-23]. The reaction is catalyzed by ATP citrate lyase and described as follows [25]:

The critical step for lipid accumulation in oleaginous yeast strains is the change of intracellular metabolite concentrations after nutrient exhaustion in the culture medium [20,21,23,26]. Various studies has been carried out to study the induction of lipid accumulation when nitrogen is in depletion conditions and have proved that extracellular nitrogen depletion is a limiting and introducing factor for lipid accumulation in oleaginous yeast [19,23,24,26-28].

Under nitrogen depletion condition, the AMP-desaminase enzyme would be activated, which cleaves AMP into inosine monophosphate (IMP) and NH_4^+ ions, and thus perform a complementary nitrogen source with NH_4^+ ions for further essential cell material synthesis [29]. As a result, the concentration of intracellular adenosine monophosphate (AMP) would be decreasing rapidly, which leads to an alteration in citric acid cycle. For example, NAD^+ - and in some cases NADP^+ - isocitrate dehydrogenase, will lose their activity because they are allosterically activated by intracellular AMP. As NAD^+ - and NADP^+ - isocitrate dehydrogenase are enzymes that cleave isocitric acid into α -ketoglutaric acid, their activity reduction leads to the accumulation of iso-citric acid inside mitochondrion [30-32]. Iso-citric acid could be reversibly transformed into citrate and is considered equilibrium with citrate. Consequently, the over accumulation of citrate inside mitochondrion will lead the citrate transportation into cytoplasm and exchange with malate [33].

ATP-citrate lyase enzyme (ATP-CL), which is key in lipid accumulation and uniquely exists in oleaginous yeast, catalyzes citric acid into acetyl-CoA and oxaloacetate [29]. After that, acetyl-CoA is formed into malonyl-CoA and generate cellular fatty acid through β -oxidation process [23]. Figure 1.2 presents the intermediate cellular metabolism and lipid accumulation in the oleaginous yeast under nitrogen depletion condition in the culture medium.

Figure 1.2 Consecutive steps leading to the de novo lipid biosynthesis in oleaginous microorganisms growing under nitrogen-limited conditions (permission from ref. 9 was obtained from publisher to use this figure) [9,26].

1.4.4 Importance of several crucial enzymes in oleaginous yeast

There are two main reasons to illustrate the outstanding lipid accumulation ability in oleaginous yeast. One is they could provide acetyl-CoA continuously, which is the crucial precursor to synthesize intracellular lipids [34]. The other one is that they produce adequate supply of NADPH and use it as the reductant for intracellular lipid synthesis [25]. During these processes, several crucial enzymes play important roles and the malic enzyme (ME) and ATP citrate lyase (ACL) were reported to be two of these essential enzymes, which could respectively generate sufficient acetyl-CoA and NADPH.

ACL enzyme exists uniquely in the oleaginous yeast strains and reported to be absent in the non-oleaginous ones. It is an enzymatic complex and regarded as one of the most influential factors that counts for lipid accumulation ability of oleaginous yeast [21]. It catalyzes citric acid into acetyl-CoA and oxaloacetate and the net product could be further transformed into fatty acid.

NADPH is required by each step in fatty acid synthesis. In oleaginous yeast, malic enzyme (ME) has been proposed to be essential NADPH provider according to the reaction:

Malic enzyme is also responsible for fatty acid synthase activity through direct channelling of NADPH, which in the end leads to the formation of a multi-enzyme complex of ME, ACL, and FAS. Importance of malic enzyme in the biosynthesis of lipids is further supported by researchers who terminate ME activity through gene

mutation [35] or inhibition of the enzyme [36] and observed decrease of lipid content in oleaginous yeast strains.

1.5 Protein profiling based on LC-MS

1.5.1 Introduction about proteomics

Proteome is the entire complement of proteins that express within a cell, tissue or organism. It is the study on proteome, which is the large scale identification and characterization of proteins, especially the structures and functions [37-39]. Proteomics is considered to be the next step in biological system study after genomics and transcriptomics and acts as a 'bridge' which links genomic function and complex cellular behaviour together. However, proteomics is more complicated than genomics as proteome differs significantly between different cells and time points, whereas genome of organisms is relatively constant.

Aiming to identify differentially expressed proteins from crude biological samples, proteomics makes it possible to characterize biological roles of proteins and clarify biological mechanisms. In order to set up proteomics platform, the following steps are commonly built. Firstly, the target organisms, which is yeast strains in our research, are cultured and proteins are extracted. Second step would be protein separation with gel electrophoresis or liquid chromatography techniques. Thirdly, protein information would be detected using mass spectrometry. Last step is the protein identification based on database and bioinformatics analysis. [38]

In order to further validate the identified proteins, techniques should be introduced to confirm the identification using mass spectrometry, such as immunoblotting, real-time RT-PCR or N-terminal sequencing.

1.5.2 Basic introduction of LC-MS

Liquid chromatography-mass spectrometry (LC-MS) is a very powerful technique of critical sensitivity and selectivity and being used for various applications. Most of the applications of LC-MS is to detect and identify chemicals in a complex mixture. Numerous LC techniques, such as reversed phase chromatography, size-exclusion chromatography, affinity chromatography and ion exchange chromatography, have been reported these years to be applied both independently or combined with proteomics study [40,41]. One advantage of liquid chromatography is that it could safely separate organic compounds, which could have a wide range from small molecule metabolites to peptides and even to proteins. Whereas gas chromatography would not be suitable for non-volatile and molecules that are thermally fragile.

In order to build a complete LC-MS system, the following procedures are needed. Firstly, physicochemical properties of target peptide mixtures should be studied to acquire the differences between them and suitable chromatography columns should be selected to separate peptide mixtures according to their different properties. Secondly, the ionization source would be built to elute peptides for identification, normally used ionization sources are matrix-assisted laser desorption/ionization (MALDI) or electrospray ionization (ESI). Next, mass analyzer should be connected to separate the ions based on their different mass to charge ratios (m/z). Lastly, a detector is utilized to detect the relative abundance of the ions which are at discrete

m/z . In general, samples are firstly pumped into high performance liquid chromatography (HPLC) by auto sampler and flow through columns for separation. Then the separated peptides will be ionized by ionization sources and then detected in mass spectrometer. (Figure 1.3).

Figure 1.3 Schematic diagram of HPLC/MS machine.

Mass spectrometry is a paramount technique to characterize and identify the sequences of protein and MS-based proteomics study have improved significantly these years [42]. There are three components of mass spectrometer: ionization source, mass analyzer and ion detection system. In order to analyse peptides with mass spectrometry, the peptides are firstly ionized to produce gas-phase ions, which will be separated according to different mass to charge ratio (m/z) and detected by ion detection system at last.

Together with available mass spectrometers of proper performance and mass range, Electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) are two primary methods to ionize and characterize whole proteins (Figure 1.4). The purpose of ionization source is to originate charges on the eluted peptides. For ESI method, a high voltage is provided to form a plume droplet, which will be evaporated and charges will be left on the peptides [43,44]. As for the MALDI method, it is a two-step process. The first step is to incorporate the molecules into a suitable matrix and protonate or deprotonate the molecules in a hot plume as a second step.

Figure 1.4 Schematic of ESI and MALDI ionization sources.

After being ionized by ionization source, the sample will flow through mass analyzer, which is a key part of mass spectrometry. There are four commonly used mass analyzers in proteomics study (Figure 1.5): time-of-flight (TOF), ion trap, quadrupole, and Fourier transform ion cyclotron resonance (FT-ICR) [45]. TOF and FT-ICR are preferable because of their wide mass range and higher mass accuracy as for FT-ICR.

Figure 1.5 Schematic of four types of mass analyzers (A. Quadrupole; B. Ion trap; C. TOF; D. FT-ICR).

In most cases, a combination of mass analyzers, which is called tandem mass, will be used for better structural information acquisition, such as triple quadrupole (QQQ), quadrupole ion trap (QIT), quadrupole-time-of-flight (Q-TOF), and time-of-flight/time-of-flight (TOF/TOF). The combination of different mass analyzers could

strengthen the advantage of individual mass analyzer and expand proteomics analysis capabilities [45].

In our study, Q-TOF mass analyzer is used as it could qualitatively and quantitatively analyze complex samples with very high sensitivity and accuracy (Figure 1.6). Normally three elements are included in Q-TOF analyzers: quadrupole analyzer Q1, which will be used to select and transit the precursor ions of interest in mass filter mode; quadrupolar collision cell Q2 which fragments the ions from Q1 by collision induced dissociation (CID) under nitrogen or argon gas molecules; and TOF mass analyzer which separate the fragments from Q2 and the remaining precursor ions from Q1 and finally get to the detector and generate MS/MS data.

Figure 1.6 Structure of Q-TOF mass analyzer.

HPLC coupled mass spectrometry system has become an effective platform for proteomics study, especially LC with nano-flow rates which provides higher sensitivity. Tandem mass systems coupled with nano-flow HPLC platform has been introduced in a number of proteomics analysis studies. Mass analysis of proteolytic peptides is of great popularity in protein characterization study. It requires cheaper instrument designs and easier sample preparation procedures.

After trypsin digestion, proteins turn into peptides and are fractionated through liquid chromatography separation in tandem MS. From the data collected, peptide sequences could be identified with on-line libraries or databases, such as SwissProt, NCBI and PIR. Various computer software have been developed to predict the sequence of peptides, such as Spectrum Mill MS Proteomics Workbench (Agilent Technologies) and Protein Pilot TM (AB Sciex).

1.5.3 Quantification method based on iTRAQ

The degree of protein expression changes could be illustrated if the quantitative information could be obtained. For this purpose, the technique of isobaric tags for relative and absolute quantification (iTRAQ), which is a stable isotope labelling method has been widely used and gained popularity recently with its advantage of simplicity in labelling procedures and ability to label up to eight samples in a single run.

The basic labelling principle of iTRAQ is to use tags with different mass to label the N-terminus of peptides. Each iTRAQ labelling reagent is a tag with a total molecular weight of 145 and is composed of three parts: an amino reactive N-

Hydroxysuccinimide (NHS) group, a balance group and a reporter group. The NHS group could label all the digested peptides at their N-terminus and lysine side chain. The reporter groups are of different masses and are defined m/z 114, 115, 116 and 117 for 4-plex iTRAQ and m/z 113, 114, 115, 116, 117, 118, 119 and 121 (m/z 120 is absent as the mass for phenylalanine immonium ion is also 120) for 8-plex. The balance groups aim to make sure that all the samples display the same mass after labelling. Each iTRAQ reagent would label one sample and all the targeted samples will be labelled with different iTRAQ reagent. Before analysis, the multiple iTRAQ labelled samples will be combined into one mixture and injected into LC-MS/MS machine for analysis.

In the first dimension of MS, all the labelled samples with different iTRAQ reagents remain the same position as the molecular weights are of no difference. During the second dimension MS/MS, fragmentation would be performed and the reporter groups will be generated whereas the balance groups will be cleaved. As a result, a relative quantification of different samples will be achieved simultaneously according to the relative peak area that the reporter groups contribute to each sample in the total peptides. Figure 1.7 (page 22) shows the schematic diagram of iTRAQ labelling mechanism.

Figure 1.7 Working Mechanism of iTRAQ Reagent. A, B and C represent three different samples.

The iTRAQ labelling method is a quantitative and simplified analysis method that has various advantages, which is able to multiplex different samples and increases accuracy of analysis [46,47]. Figure 1.8 (page 23) shows the design of protein profiling workflow. The feature to quantify multiple samples at the same time could increase the efficiency and make comparative proteomics analysis possible. However, as enzymatic digestion is performed prior to iTRAQ labelling, the complexity of analysis is also increased and an effective fractionation method for peptide separation is required as well.

Figure 1.8 Protein profiling workflow design.

1.6 Metabolite profiling based on GC-MS

1.6.1 Introduction about metabolomics

Metabolomics is defined as the study of metabolites involving in chemical processes in any cell, tissue or organism. To be more specific, metabolomics is the profiling study of a large scale collection of small molecule metabolites, the final products of the cellular processes [48,49]. In 1971, Horning, *et al.* first introduced the term “metabolic profile” and the usage of gas chromatography-mass spectrometry (GC-MS) for measuring the compounds in human urine and tissue extracts [50,51].

Metabolite are the intermediates and products of gene expression and could be a part of the regulatory system in an integrated manner [52]. The levels of metabolites in a biological system could be generally considered as the ultimate response and reflection of genetic or environmental changes [53,54]. It is difficult to profile the metabolites in an organism comprehensively, but successful identification and quantitation of a large number of metabolites could still provide essential insight into the biochemical phenotype of an organism and help monitor the biological response. Metabolomics could provide a complementary picture and an instantaneous description of the entire physiology in an organism, together with mRNA gene expression data and the proteomics analysis. (Figure 1.9, page 25)

Figure 1.9 A schematic diagram of the “Omics” cascade.

Metabolomics presents several advantages comparing with other “omics” (genomics, transcriptomics, and proteomics) technologies. The estimated number of metabolites in an organism is far less than that of genes, transcripts, and proteins. What is more, metabolomics measures the end products of genomic, transcriptomics, and proteomic variability, which makes it the most integrated profile when the organism is perturbed and response to it.

1.6.2 Comparison of analytical methods for metabolites

Comprehensive analytical methods and tools for metabolomics are still under development, although the technology for other ‘omics’, such as genomics, transcriptomics and proteomics have already made great strides [55]. As the metabolites have diverse chemical properties and a large dynamic range, there is no single platform for analysis of all metabolites in a biological sample. The whole metabolome could be divided into several of chemical compound classes with

different chemical properties and wide range of concentrations, such as organic acids, sugars, lipids and amino acids, etc. This has brought a big challenge for metabolomics study as current analytical instruments could not offer analysis in true sense. What is more, the generation of complete metabolite databases is also challenging, which makes the data analysis a problem in metabolic profiling study.

In order to analyze the metabolites with a wide range of structures, changing concentrations and complex chemical properties such as volatility and solubility, tools with high sensitivity and specificity is a necessity for metabolite detection.

Gas chromatography (GC), capillary electrophoresis (CE) and high performance liquid chromatography (HPLC) are three of the most commonly used metabolite separation methods. Gas chromatography is a powerful methods for separation with high resolution, especially when it is interfaced with mass spectrometry, which could be utilized in metabolic profiling contributing to its high sensitivity, specificity, and positive peak identification [56]. One drawback is that chemical derivatization is required and only volatile chemicals is able to be analyzed without derivatization. Also, metabolites with large size and polar property are not detectable by GC [57]. HPLC could detect analytes with a wider range, however the chromatographic resolution is relatively low [58]. The separation efficiency of CE is better than HPLC, and CE is also suitable to detect metabolites with a wider range than GC. But as for all electrophoresis techniques, CE is more appropriate only for charged metabolites [59].

For analytical tools, most commonly utilized approaches are mass spectrometry (MS) and nuclear magnetic resonance (NMR). MS is one of the most powerful approaches

in metabolomics study for metabolites separation and detection, especially when it is combined with chromatographic separation methods, such as HPLC, GC or CE. In order to perform metabolomics analysis, the approach must be high speed and effective to satisfy the demand of complex metabolites.

Each compound owns a specific retention time, which could be used to identify the metabolite, especially the separation and detection of isomers. However, MS alone is not suitable to distinguish the isobaric compounds and must be combined with other methods.

NMR spectroscopy is first discovered in the 1940s and a powerful tool to analyze metabolites nowadays contributing to its high resolution. One of the advantages for NMR method is that the reproducibility and easiness of quantitation comparing with MS. It could perform absolute quantitation of metabolites with use of single internal or external standard [60,61]. The amount of sample usage is little, sample preparation is easy and no chromatographic separation is required. Another advantage of NMR is the high efficiency and sufficient sensitivity, which could detect a sample within one to fifteen minutes [62].

Other techniques are also developed for metabolomics analysis, such as gas chromatography mass spectrometry (GC-MS), Fourier transform infrared spectroscopy (FT-IR) and Capillary electrophoresis mass spectrometry (CE-MS). Different separation and analytical tools are combined for different detecting purposes.

1.6.3 General background of GC-MS

In order to perform profiling (unknown) of metabolites in our study, GC-MS is selected as a preferred profiling method with good reproducibility and high sensitivity [63,64]. What is more, the availability of mass spectral libraries makes metabolite identification easier [65]. GC-MS is the combination of two powerful techniques for analyzing chemical mixtures. However, it is only suitable for detection of volatile compounds, and non-polar ones need to perform derivatization before GC-MS analysis.

Metabolomics based on GC-MS could cover a unique range of analytes and makes this technique different from NMR and LC-MS. Numerous metabolites are particularly suitable using GC-MS to analyze, such as lipids and esters, which are volatile and undetectable in LC-MS well due to their poor ionization. Other metabolites including amino acids, fatty acids and alkaloids are compatible with GC-MS. Whereas HPLC is amendable for metabolites which are non-volatile and no chemical derivatization is needed. Meanwhile, LC-MS is a powerful tool that commonly used for organic acids, nucleosides and amino acids detection. Figure 1.10 (page 29) has concluded the different metabolites that are compatible for GC-MS and LC-MS analysis.

Figure 1.10 Summary of metabolites compatible with different separation methods (LC-MS and GC-MS).

For better analysis of lipids, fatty acids, amino acids and other metabolites, GC-MS is introduced in our study as the main metabolic profiling method. GC could separate the different components inside a sample mixture and MS is used to individually characterize each component. The combination of GC and MS has successfully established a qualitative and quantitative evaluation platform of a mixture of metabolites.

The first step in metabolic profiling with GC-MS is sample preparation, which includes metabolism quenching, extraction of metabolites and derivatization of metabolites. An efficient quenching step is important so as to stop all the metabolic activity quickly, which is usually achieved by rapid change of temperature by

transferring the samples into ice-cold methanol or liquid nitrogen instantly after sample collection.

Metabolite extraction is usually performed by enzyme digestion or chemical extraction with organic solvents and physical extraction according to different biological targets [66]. Derivatization is required to make sure all the metabolites could be detected inside the sample. Methoxymation is used in order to convert the functional groups, such as amines, imines and hydroxyls into oximes and alkyloximes and stabilize the compounds. Silylation is introduced in the derivatization system to reduce molecule polarity by replacing the active hydrogen with an silyl group [-Si(CH₃)₃] [67].

After derivatization, samples will be separated in a proper column and suitable program setup in the GC-MS. Different molecules will be separated during the travel in the column according to their various chemical properties, which would be retained by the column and eluted at different times. Then the MS part will function by breaking the molecules into ionized fragments and detect the molecules separately according to their mass to charge ratio (m/z). After GC-MS analysis, the acquired raw data will be analyzed by a few software (such as Chem-Station by Agilent, GC-MS Solution by Shimadzu and Mass Frontier by Thermo Fisher) and compare with the database to identify the metabolites [68].

For data variance interpretation, Principal component analysis (PCA) is commonly used through examining the linear combinations of individual variations. It is known as an unsupervised method to illustrate differences among individual samples [69]. The variances between samples is explained by the differences of scores for samples

and the original variations are represented by several principle components, e.g. PC1, PC2, PC3, etc. The importance of each variation to the PC is called loading value and represented by p , and each PC could be defined by loading values and defined variables, for example, $PC1 = p_1 \times 1 + p_2 \times 2 + p_3 \times 3 \dots$, which means the larger the loading value p , the higher importance of variable to PC value. Other statistical methods, e.g. partial least squares-discrimination analysis (PLS-DA) [70] and hierarchical clustering [71] are also applied for different purpose of study.

1.7 Proteomics and metabolomics study in yeast

LC-MS based proteomics study and GC-MS based metabolomics study could contribute to identify and characterize proteins and metabolites in a large scale. Researchers have started to utilize these two ‘omics’ studies in yeast in the late 1990s. Protein expression and metabolite level changes have been detected in the yeast, mostly *Saccharomyces cerevisiae*, to determine the pattern change and further decipher the proteomic and metabolic networks that dictate the cellular functions in yeast [72-76]. However, these ‘omics’ study targeting on oleaginous yeast strains were undertaken only in the last several years [77,78]. Most studies on the oleaginous yeast strains are focusing on the accumulation of lipids for industrial utilization. The only few articles on ‘omics’ study have been making effort to characterize the biochemical events and lead to deeper understanding of the lipid accumulation abilities in oleaginous yeast strains.

CHAPTER 2

OBJECTIVES

Oleaginous yeast, an emerging type of microorganism, is commonly used for biofuel production. It could accumulate lipids to a high content up to 70% of its cell dry weight, expressing potentials to be extracted and derived into biofuels and satisfy the increasing demand of energy in the world as an alternative source. Different kinds and species of oleaginous yeast have been identified or isolated, with various lipid accumulation abilities. However, the mechanism and systematic changes during lipid synthesis are not clear enough. In this project, we aim to find out the reason for the particular and extraordinary high lipid accumulation ability in oleaginous yeast strains and the essential proteins and metabolites involved and affected in the lipid accumulation and metabolism pathways.

Lipid accumulation was studied to optimize the culture conditions in three different yeast strains, *Saccharomyces cerevisiae* with low lipid accumulation ability, *Cryptococcus albidus* with medium lipid accumulation ability and *Rhodospiridium toruloides* with high lipid accumulation ability. Cultivation condition for each strain was optimized to make sure that they express their best status to produce lipids. Early, middle and late lipid accumulation stages for each strain were recorded and cells were collected at each stage for further study.

Proteomics analysis could provide a platform to identify large scale of intracellular proteins. Essential proteins or enzymes may involve in many important pathways and

cause the lipid accumulation, the discovery of these protein level verifications will definitely contribute to a better understanding of the mechanism in lipid assembly. We would utilize LC-MS/MS analysis to investigate the protein levels change and establish the protein profiling platform by LC-MS. Time-course comparison between oleaginous yeast and non-oleaginous yeast would be achieved by iTRAQ-coupled 2D LC-MS/MS quantitative analysis.

Metabolomics analysis could identify the whole set of biological small molecules participated in the metabolic reactions. Global metabolites profiling platform by GC-MS was established to provide valuable information on yeast synthesis and catabolism pathways. In addition to proteomics, metabolomics could be utilized to explore the physiological relevance of the proteomics data.

Our findings could provide molecular insights for lipid production in oleaginous yeast. The combined proteomics and metabolomics information in comparison between oleaginous and non-oleaginous yeast strains should provide a comprehensive understanding of lipid accumulation. It is to be hoped that the investigated important proteins and genes could be utilized in the model yeast strain to increase the amount of biofuel precursors--fatty acids through metabolic engineering and genetic engineering strategies.

CHAPTER 3

MATERIALS AND METHODS

Lipid accumulation

3.1 Yeast strains and growth conditions

3.1.1 *Saccharomyces cerevisiae* and its culture condition

The *Saccharomyces cerevisiae* BY4741 (leu2 Δ , MAT α , ura3 Δ , his3 Δ , met15 Δ) wild type strain was obtained from European *Saccharomyces cerevisiae* Archive for Functional analysis (EUROSCARF) and introduced in this study. The cells were cultured at 30 °C, 250 rpm in YEPD liquid medium (10 g/l yeast extract, 20 g/l peptone, 20 g/l glucose) in 250 ml flasks.

3.1.2 *Rhodospiridium toruloides* and its culture condition

Rhodospiridium toruloides is reported to possess the potential to accumulate merely the highest lipid content among oleaginous yeast species and was purchased from the American Type Culture Collection (ATCC 10788). Two-stage process was conducted for lipid production. Cells were first cultured in YEPD medium for proliferation until stationary phase. After that, cells were centrifuged, washed twice with pre-chilled distilled water and transferred with 10% v/v inoculum into nitrogen-limited culture medium to trigger lipid accumulation phase, containing per litre: yeast extract 0.5 g, (NH₄)₂SO₄ 0.1 g, KH₂PO₄ 1.0 g, MgSO₄·7H₂O 1.5 g. Then the cells

were maintained at 28 °C and agitated at 250 rpm in both stages. Different glucose concentrations (20, 30, 40, 60, 80 and 100 g/l) in nitrogen limited medium were used to grow the oleaginous cells to determine the substrate inhibition. Various cell density of inoculation when transferring cells from YEPD to nitrogen limited medium were also tested, with final cell densities varying at 0.25×10^8 /ml, 0.5×10^8 /ml, 1×10^8 /ml, 2×10^8 /ml, 3×10^8 /ml and 4×10^8 /ml.

3.1.3 *Cryptococcus albidus* and its culture condition

Cryptococcus albidus (ATCC 56297) has the potential to get a medium lipid content around 46% when compared with *S. cerevisiae* (less than 10%) and *R. toruloides* (up to 65.2%). Same as *R. toruloides*, lipid accumulation was initiated by the exhaustion of nitrogen in the culture medium. Two-stage cultivation method was performed on *C. albidus* exactly the same with *R. toruloides*.

The carbon to nitrogen (C/N) ratio in YEPD and nitrogen-limited medium was 4.5 and 135, respectively. Cell proliferation phase had taken place first, and along with nitrogen exhaustion in the culture medium, the lipid accumulation phase is initiated.

All the pH value of culture mediums was adjusted to 6.0 with NaOH. The carbon source and the magnesium solution are autoclaved separately. All of the three strains were cultivated in 250 ml flasks and maintained at 4 °C on YEPD agar (10 g/l yeast extract, 20 g/l peptone, 20 g/l dextrose, 20 g/l agar) plates and transferred to fresh plates every month.

3.2 Growth measurement

3.2.1 Growth curve test

For *S. cerevisiae*, a single colony was picked and cultured in 5 ml YEPD broth in a 50 ml tube overnight to exponential phase. The overnight culture was used as inoculum for 50 ml culture medium in 250 ml flasks and the initial optical density (OD) is 0.2. Optical density at 600 nm (OD₆₀₀) was measured for growth record using Nano Drop 2000 (Thermo Scientific).

For *C. albidus* and *R. toruloides*, the inoculums were cultured for 16 hours in 5 ml YEPD broth in 50 ml tubes separately to develop them to the exponential phase. Calculated amount of the inoculums were injected into the 250 ml flasks containing 50 ml culture medium to make initial OD to be 0.2. After 36 hour cultivation, the residual glucose in the medium was merely zero, cells of each strain were collected and centrifuged, washed twice with distilled water. Different amount of cells were transferred into the nitrogen-limited culture medium in order to achieve various inoculation cell densities. The yeast cells were then cultured and collected every 12 hours.

3.2.2 Dry weight test

Dry cell weight test could give accurate and consistent information about the biomass production, providing a more specific status of cell physiology. Twenty milliliter cells of each strain were collected at proper time point and pellets were washed with pre-chilled distilled water twice. Afterwards, the pellets were transferred into pre-weighed tubes and lyophilized in freeze drier overnight to constant weight.

In the two-stage process for lipid production, when cells were transferred into the second stage, the initial glucose level was comparatively high and decreases with time as consumed by cells to accumulate lipids. The lipid content reached the peak level when the glucose was used up, thus residual glucose detection was undertaken to indicate the lipid content change in another way. The method is according to I. Branyikova and B. Marsalkova, *et al* [79].

3.2.3 Residual glucose detection

3.2.3.1 Standard curve test of glucose

Anthrone solution was prepared fresh daily, dissolving 200 mg anthrone in 10 ml concentrated sulfuric acid. Glucose concentration gradient was 0, 10, 20, 30, 40, 60, 80, 100 mg/ml, respectively. Five milliliter anthrone solution was added into 1 ml glucose solution and heated in boiling water for ten minutes. After cooling to room temperature, absorbance was measured at 620 nm using spectrophotometer; the instrument was zeroed using 1 ml water reacting with 5 ml anthrone. The concentration of residual glucose was calculated based on the standard curve.

3.2.3.2 Sample preparation for residual glucose determination

For measurement of the residual glucose in the culture medium, 1 ml aliquot was taken at proper time interval. Centrifugation was at room temperature under 10,000 rpm for 5 minutes. Then the supernatant was transferred and diluted to a proper concentration as the sensitivity of anthrone method requested the concentration of glucose below 150 µg/ml. One milliliter diluted supernatant was added into 5 ml anthrone-sulfuric acid solution and heated in the boiling water for ten minutes. The

samples were cooled down to room temperature immediately after 10 minutes and the absorbance was measured at 620 nm in spectrophotometer after zeroing the instrument.

3.2.4 Lipid content measurement

Yeast cells were difficult to disrupt due to the rigid cell wall and various methods of extraction have been tried out to set up a feasible and productive platform for successful yeast crush. Different methods such as mechanical grind by mortar or glass beads, ultra sonication chemical treatment by acid digestion [80-82] were tried with same amount of cells and the results showed that method of R. Schneiter & G. Daum and C. Ratledge & R. K. Saxton gave the highest lipid content for the same batch and same amount of yeast cells, the former one was chosen as it is relatively easier to operate and time-saving.

Yeast cells of *S. cerevisiae*, *C. albidus* and *R. toruloides* were cultured in a suitable medium so as to obtain 50 mg dry cell weight and collected by centrifugation at 10,000 g, 5 minutes. The supernatants were discarded and cell pellets were washed with distilled water twice. The harvested cells were mixed with 4 ml methanol and 100 µg heptadecanoic acid (C17:0, which would not be produced in yeast cells naturally) was added as internal control. Eight grams of (6 ml) glass beads (425-600 µm; Sigma) were added and the cells were disrupted with Fast-prep agitator by shaking ten periods of 30 s with 1 minute cooling intervals on ice. Eight millilitre chloroform was added to the suspension to make the final chloroform/methanol ratio to be 2:1 (v/v) and the suspension was stirred for 1 hour on a vortex mixer at room temperature. The supernatant was centrifuged and transferred to a clean tube. The

glass beads were washed with 8ml chloroform/methanol again, centrifuge and combine the supernatant with that of the last step. Four milliliter 0.034% MgCl_2 was added to the supernatant and stirred for 10 minutes. The extract was centrifuged and the supernatant was discarded. The organic phase was washed with 4 ml 2 mol/l KCl/methanol (4:1; v/v), centrifuged again and the supernatant was discarded. The organic phase was washed with 4 ml chloroform/methanol/water (3:48:47; per vol.) and centrifuged in a table-top centrifuge for 5 minutes. The upper aqueous layer, including the protein layer that formed at the phase boundary was aspirated off and the organic phase was washed repeatedly until the phase boundary becomes clear. The organic phase was then transferred to a pre-weighed tube to evaporate the solvent in a centrifugal evaporator. The lipid was weighted after having reached dryness and stored in $-20\text{ }^{\circ}\text{C}$ for further analysis.

Proteomics

3.3 Cell lysis, protein digestion and labelling with iTRAQ reagents

For cell lysis and protein extraction, all steps were carried out on ice to avoid denaturation of proteins. Twenty OD_{600} units of yeast cells were harvested by centrifugation at 10,000 rpm, $4\text{ }^{\circ}\text{C}$ for 5 minutes. The supernatant was discarded and cell pellets were washed twice by pre-chilled water. Lysis buffer containing the following compositions gave the highest and most stable protein level when treating the same amount and batch of cells: 8 M Urea, 50 mM DTT, 100 mM NaCl, 50 mM Tris-Cl (pH7.4), 1 mM EDTA, 0.1% Triton X-100, and 1 mM PMSF.

The pH value of lysis buffer was adjusted to 7.4 with 1 M HCl. Three hundred microliter lysis buffer and approximate 200 μl glass beads were then added to the

pellet and disruption was conducted in the homogenizer by shaking ten periods of 30 seconds with 2 minutes cooling intervals on ice. Centrifugation was then performed for the lysate at 10,000 rpm for 10 minutes at 4 °C. The supernatant was moved to clean tubes and stored at -80 °C.

Lipid contaminants in protein extract were removed prior to further treatment by organic solvents as demonstrated by W. Wang, *et al.* [83]. Briefly, for 100 µl supernatant, 100 µl chloroform/methanol (2:1, v/v) was added together and mixed well, followed by the addition of 100 µl ice cold 10% trichloroacetic acid (TCA)/acetone. The whole mixture was vortexed thoroughly and put on ice for 10 minutes for precipitation. Then the mixture was centrifuged for 2 minutes at 10,000 rpm and the upper layer was aqueous phase, which was discarded. The rest mixture in the tube was then washed with 300 µl 10% TCA for three times and 500 µl ice cold water consecutively. Centrifugation was performed after each washing step and upper aqueous layer was discarded.

After these steps, the lipid contaminant in the protein sample should have been removed efficiently and 600 µl pre-chilled acetone was then added for protein precipitation. The mixture was put in -20 °C for 30 minutes and centrifuged for 5 minutes at 8,000 rpm. The supernatant was removed and the pellets were air-dried and reconstituted in 0.5 M triethylammonium bicarbonate (TEAB).

Protein concentration was determined using 2-D Quant Kit (GE Healthcare). Briefly, a standard curve was plotted using bovine serum albumin (BSA) as a control. A gradient dilution of BSA into 0, 10, 20, 30, 40 and 50 µg/ml was performed so as to control them in a linear range. Different amount of unknown samples were used to make sure they were in this range. After precipitation and co-precipitation step, the

proteins inside control and unknown samples in each tube were pelleted. Copper solution and colour reagent were then added into each sample and incubated for twenty minutes at room temperature. The OD values of protein samples were measured at 480 nm with spectrophotometer (zeroed with copper solution plus colour reagent) and quantitated based on the standard curve.

A total of 100 µg protein from each sample was precipitated after quantification by adding six times ice cold acetone of the sample volume and incubated at -20 °C for two hours. The samples were centrifuged at 10,000 rpm for 10 minutes and the supernatants were carefully removed.

Preparation for labelling steps were performed according to the iTRAQ protocol (AB Sciex). Each protein pellet was dissolved and reconstituted in 20 µl dissolution buffer (TEAB) and 1 µl denaturant (2% SDS) and vortexed to mix thoroughly. Four microliter reducing reagent, which was 50 mM tris-(2-carboxyethyl) phosphine (TCEP) was added, vortexed to mix and spinned slightly. The tubes were incubated at 60 °C for one hour to break the disulphide bonds between proteins. After that, 1 µl cysteine blocking reagent, 200 mM methyl methanethiosulfonate (MMTS) were added into each tube and vortex to mix. The mixture was spun down and incubated for 15 minutes at room temperature.

Protein digestion was performed by adding 20 µl of 0.25 mg/ml sequencing grade modified trypsin (Promega) and incubated at 37 °C overnight. Each sample was labelled with the iTRAQ tags after digestion as follows: *S. cerevisiae*, iTRAQ 114; *C. albicus*, iTRAQ 115; and *R. toruloides*, iTRAQ 116. The labeled samples were then pooled together for further LC-MS/MS analysis. Figure 3.1 (page 42) was the layout for experimental design.

Figure 3.1 Layout for iTRAQ-based quantitative proteomics experimental design. Proteins in yeast strains with low, medium and high lipid accumulation abilities were extracted, quantified, reduced and digested. Respective iTRAQ reagent 114, 115 and 116 were used to label *S. cerevisiae*, *C. albidus* and *R. toruloides*. The samples were then combined into one and injected into online 2D LC/MS system and differences in proteins levels were identified by analysing the raw data by software.

3.4 On-line 2D nano-LC-MS/MS analysis for protein

LC-MS/MS analysis was performed by Agilent 1200 series nano-flow liquid chromatography system (Agilent Technologies) interfaced with 6500 Q-TOF mass spectrometer with HPLC-Chip Cube (Agilent Technologies) for two dimensional protein analysis. The HPLC-Chip was a combination of Zorbax 300SB C₁₈ reversed-phased (RP) column (75 µm x 50 mm, 3.5 µm) packing with Zorbax 300SB C₁₈ enrichment column (0.3 x 5 mm, 5 µm) [84].

In the first dimension, 4 µg combined peptide mixtures was trapped onto the PolySulfoethyl A strong cation exchange (SCX) column (0.32 x 50 mm, 5 µm) and the retained peptides were stepwise eluted by sequential injection of 8µL salt plugs with series of ammonium formate solutions in nine gradient concentrations of 20, 40, 60, 80, 100, 300, 500, 1000 mM, respectively.

In the second dimension, the system was turned into enrichment mode and the peptides were eluted through the SCX column and loaded onto Zorbax 300SB C₁₈ enrichment column by buffer A (5% acetonitrile and 0.1% formic acid) with a flow rate of 4 µl/min. Subsequently, the HPLC-Chip was changed to analytical mode. The previous trapped peptides on the enrichment column were eluted for 60 minutes by buffer B (0.1% formic acid) and buffer C (95% acetonitrile and 0.1% formic acid) with a nano-flow gradient of buffer C changing from 5% to 80% at a flow rate of 300 nl/min and further flowed through the analytical Zorbax 300SB C₁₈ reverse phase column for separation (Figure 3.2). The effluent was directly detected and analysed by 6500 Q-TOF mass spectrometer with a capillary voltage of 1950 V. In total, 10 runs were carried out to accomplish analysis of one sample. For MS analysis, positive

ionization mode was used and survey scans has been acquired within the range of m/z 300 to m/z 2000 with an acquisition rate of 4 spectra per second. Four most abundant ions that exceeded 1000 counts were selected for MS/MS analysis within m/z 50 to m/z 2000. Mass of control compounds were 299.2944 and 1221.9906. They were used for automatic mass correction in each run.

Figure 3.2 Workflow of on-line 2-D nano-LC. (a) The peptides eluted through SCX column and trapped onto enrichment column. (b) HPLC-chip was changed to analytical mode and previously trapped peptides were eluted and analyzed by analytical column.

3.5 Protein identification and data analysis

All the mass data were analyzed by Spectrum Mill MS Proteomics Workbench (Agilent Technologies) to quantitate peptides and identify the proteins. Each MS/MS spectrum was searched based on species of Microorganisms against the UniProt_Swiss_Prot database. Proteins qualified for further statistical analysis should meet the following criteria: two or more unique peptides identified with high confidence (>99%), protein score was more than 11.0, peptide score was at least 6.0 and *p*-value below 0.05 was recorded as significant for protein quantification in this study. Three independent experiments were carried out and data were reported as means $\pm$ SD.

Modification of methylmethanethiosulfate-labelled cysteine and iTRAQ labelling at N-terminal residues were set as fixed modification. Protein relative quantification using iTRAQ was achieved on the MS/MS scans and was represented by the ratio of peak areas of masses of the tags corresponding to iTRAQ reagents, which was 114, 115 and 116 Da, respectively. The relative amount of one peptide was determined by comparing peak areas obtained at *m/z* 115.1 and 116.1 by that of *m/z* 114.1. The area ratios of determined peak were corrected for the contributions of isotopic overlapping, estimating the relative abundances of one particular peptide. Three independent batches were performed to increase statistically evidence of protein expression. Proteins quantification with relative expression ≥ 1.2 or ≤ 0.8 was selected for further analysis.

3.6 Western blot analysis

3.6.1 Cell Extracts Preparation

Proteins for western blot analysis were extracted with the same method in the experiments for LC-MS/MS analysis. Briefly, cell extracts were achieved by lysed in 200 μ M lysis buffer (8 M Urea, 50 mM DTT, 100 mM NaCl, 50 mM Tris-Cl (pH7.6), 1 mM EDTA, 0.1% Triton X-100, and 1 mM PMSF, pH 7.4), followed by protein concentration quantitation by 2-D Quant Kit (GE Healthcare). Loading amount of total proteins was 50 μ g in each sample for western blot analysis.

3.6.2 Gel Electrophoresis

Bio-Rad mini-protein electrophoresis system was utilized to perform western blot analysis. Proteins of different samples were separated using Sodium dodecyl sulphate-poly acrylamide gel electrophoresis (SDS-PAGE) by molecular weight. The SDS-PAGE system consisted of two layers of gels: a stacking gel and a separating gel. A two-layer gel system was prepared according to general protocol [85]. Briefly, separating gel was first prepared and loaded into well assembled gel-casting glass plates with the spacers. Different concentrations of separating gels were used in our experiments to achieve better separation efficiency based on the target protein size. Ten millilitre separating gel contained 30% acrylamide (volume determined by the final concentration), 2.5 ml of 1.5 M Tris-HCl (pH 8.8), 100 μ l of 10% SDS, 100 μ l of 10% APS, 4 ml distilled water and 5 μ l TEMED. The stacking gel was loaded after the separating gel was polymerized. Three millilitre stacking gel consisted of 0.5 ml of 30% acrylamide, 0.38 ml of 0.5 M Tris-HCl (pH 6.8), 30 μ l of 10% SDS, 30 μ l of

10% APS, 2.1 ml distilled water and 3 μ l TEMED. A 1 mm thickness 10-well comb was inserted into the stacked gel subsequently. After thorough gel polarization at room temperature, the whole gel system was established.

The gel run was carried out in a chamber which was filled with 1 $\times$ SDS running buffer containing 0.192 M glycine, 0.025 M Tris, 0.1 % SDS, pH 8.3. Protein samples were mixed with 6 $\times$ SDS loading buffer and boiled at 95 °C for 5 minutes. The samples were loaded slowly into the wells and 8 μ L Novex Sharp Pre-stained Protein Standard with a molecular weight range from 3.5 kDa to 260 kDa (Life technologies) was also loaded as the protein ladder.

Constant voltage was applied to run the gels. The voltage was maintained at 80 V for stacking gel and changed to 150 V when the samples reached the separating gel. The gel run was terminated when the dye reached the bottom of the casted gel.

3.6.3 Transfer

After electrophoresis, the gel was carefully cut, casting out the stacking gel and the redundant part, and transferred onto 0.45 μ m PVDF membrane via a semi-dry transfer system (Bio-Rad). Three layers of filter paper were also added. The layout of semidry transfer system was shown in Figure 3.3.

The transfer was performed at constant voltage (20 V) for 55 minutes. The transfer efficiency could be indicated by loaded marker.

Figure 3.3 Layout of semidry transfer system.

3.6.4 Blocking

The transferred membrane was then blocked by 5% non-fat milk in 1× PBST for one hour at room temperature to prevent the interactions between the membrane and antibodies.

3.6.5 Detection

After blocking, specific primary antibodies were dissolved in 3% non-fat milk/1×PBST and reacted with proteins on the membrane at 4 °C overnight. All the primary antibodies were purchased from ABCam Company and were diluted in different ranges as shown in the following Table 3.1 (page 49). Unbounded primary antibodies were washed off with 1× PBST for 15 minutes (three times). Secondary antibody was tagged to a reporter enzyme horseradish peroxidase (HRP). HRP-conjugated anti-mouse IgG antibody (31430, Pierce) or HRP-conjugated anti-Rabbit

IgG antibody (31460, Pierce) was used with 1/5000 dilution in 3 % non-fat milk in 1× PBS solution and incubated for one hour at room temperature. The membrane was washed for three times at 15 minutes each to rinse off the unbounded secondary antibodies.

Enhanced chemiluminescence (ECL) was carried out using SuperSignal West Pico Chemiluminescent Substrate (Pierce). Briefly, 1 ml of stable luminol/enhancer solution was mixed with equal volume of peroxide solution and applied to the membrane (avoid light after mix together). An image analyzer LAS-4000 (Fujifilm) was used to record the signals. The signals on the membrane were detected and recorded along with exposure every 15 seconds and whole exposure time was usually several minutes. Quantification of western blot results was analyzed by Image Pro Plus 6.0 software (Media Cybernetics).

Table 3.1 Dilution ranges of different primary antibodies.

Name	Series No.	Dilution
Anti-beta Actin	AB170325	1/5000
Anti-PGK1	AB154613	1/1000
Anti-Alcohol dehydrogenase	AB20994	1/10,000
Anti- Histone H4	AB10158	1/1000
Anti-Ubiquitin	AB19247	1/1000

3.7 Real-time RT-PCR

Real-time reverse transcription polymerase chain reaction (RT-PCR) was performed in order to quantify the amount of target mRNA in the samples accurately. Comparing with traditional RT-PCR, the real-time RT-PCR could monitor the polymerase chain reaction (PCR) product amplification during each cycle. SYBR Green could bind double-stranded DNA and be used as a dye to emit fluorescence signal. The fluorescence density was increased during amplification cycles and thus could be used to perform quantification. Beta-actin was used as internal control and non-template control was also included when preparing the samples.

3.7.1 RNA extraction

Total RNAs of different yeast cells were extracted using the RNeasy Mini kit (QIAGEN, USA) according to purification of total RNA from yeast method in the manufacturer's protocol. Briefly, cells were pelleted by centrifugation for 5 minutes at 6,000 rpm at 4 °C. The supernatant and any remaining media were carefully removed. The pellet was resuspended in 600 µl of Buffer RLT and approximately 600 µl of glass beads (425-600 µm; Sigma). The cells were agitated in a Fast-prep agitator by shaking six periods of 30 s with 1 minute cooling intervals. Then 350 µl lysate was transferred to a new tube and centrifuged at 14,800 rpm for two minutes and the supernatant was transferred into a new tube. Equal volume of 70% ethanol was then added to the lysate, and the mixture was loaded to the RNeasy spin columns and centrifuged at 12,000 rpm for 30 seconds. The columns were then washed with 700 µl of Buffer RW1 and centrifuged for 30 seconds at 12,000 rpm. Two additional washes with 500 µl Buffer RPE were performed and additional centrifugation was

applied to remove any residual buffer RPE. The total RNA bound on the membrane of the spin column was eluted by 50 µl RNase-free water. Total RNA was quantified and stored at -80 °C.

3.7.2 RNA quantitation

The quantity and quality of RNA were determined by NanoDrop 2000 machine (Thermo Scientific). Different samples were diluted with RNase-free water to the same concentration for further usage as templates. Quality and purity of the RNA were calculated by division of OD₂₆₀ over OD₂₈₀.

3.7.3 Real-time RT-PCR

To quantify RNA levels in different samples accurately, real-time RT-PCR was achieved using iScript One-Step RT-PCR kit with SYBR Green (Bio-Rad). Briefly, 12.5 µl of 2× SYBR Green RT-PCR Reaction Mix, 0.75 µl of forward primer and reverse primer (10 µM), 1 µl of iScript Reverse Transcriptase, proper amount of RNA template and proper amount of RNase-free water were mixed together and gave a final 25 µl volume. The mixture was then transferred to Low-Profile 0.2 ml tube strips (Bio-Rad), covered by optical flat strip caps (Bio-Rad). Real-time RT-PCR was carried out in IQ5 multicolor Real-time PCR detection system (Bio-Rad) with the following program: 50 °C for 10 minutes to reverse transcript first strand cDNA, 95 °C for 5 minutes to inactivate the reverse transcriptase; after that, 40 cycles of PCR were repeated: 95 °C 10 seconds, 60 °C 30 seconds. The melting curve was achieved within 80 cycles of 95 °C for 60 seconds, 55 °C for 60 seconds, and 55 °C for 10 seconds in each cycle. The disassociation analysis was accomplished by acquiring the

fluorescent reading increasing by 0.5 °C at each cycle from 55 °C to 95 °C as described [86]. The data including analysis of amplification process, threshold cycle number and melting curve were provided by IQ5 Real-time PCR system software Version 2.0 (Bio-Rad) in the form of Microsoft Excel as described [86]. The comparative change in folds of mRNA levels in different samples were calculated with this formula:

$$\text{Sample } \Delta\text{Ct} = \text{Ct}_{\text{sample}} - \text{Ct}_{\beta\text{-actin}}; \Delta\Delta\text{Ct} = \text{Sample } \Delta\text{Ct} - \text{control } \Delta\text{Ct}; \text{the fold of sample versus control} = 2^{-\Delta\Delta\text{Ct}}$$

The sequences of selected proteins were searched against NCBI nucleotide database. Primers were designed by Primer Premier Software and synthesized by Integrated DNA Technologies. The sequences of the primers were listed in Table 3.2.

Table 3.2 List of primer sequences for real-time RT-PCR.

Protein name	Primer sequence	PCR product (bp)
ACT1	Sense: 5'-AACTTTCAACGTTCCAGCCT-3'	216
	Antisense: 5'-CCACGTTCACTCAAGATCTT-3'	
ADH1	Sense: 5'-GAAGGTGCCGGTGTCGT-3'	177
	Antisense: 5'-AGAACCGTCGTGGGTGTAAC-3'	
PGK1	Sense: 5'-CTTGCGTTACCACATC-3'	173
	Antisense: 5'-CGAAACCGACCATAGA-3'	
HHF1	Sense: 5'- ATGGGTAGATTGCTTT-3'	216
	Antisense: 5'- TGCCAACTCGGGTCAG-3'	
UBI1	Sense: 5' CCCACCTGACCAACAA-3'	190
	Antisense: 5'- ACACTTACGGCAAACA-3'	

Metabolomics

3.8 Cell lysis and metabolites extraction

Intracellular metabolites extraction was conducted according to modified M. Wang's method [87]. Same amounts of cells for each strain, *S. cerevisiae*, *C. albidus* and *R. toruloides* were collected during early, middle and late lipid accumulation phase. Then they were centrifuged at 10,000rpm, 4 °C for 5 minutes and washed twice by distilled water. The supernatants were removed thoroughly and the cell pellets were re-suspended in 800 µl pre-chilled methanol. Ribitol (2 mg/ml, 10 µl) and heptadecanoic acid (2 mg/ml, 10 µl) were added as internal standards prior to cell disruption to correct for metabolites losses.

Approximately 600 µl glass beads (425-600 µm; Sigma) were added into the tubes. The cells were disrupted in a Fast-prep agitator by shaking eight periods of 30 s with 2 minutes cooling intervals in liquid nitrogen. After cell disruption, 320 µl chloroform and 320 µl distilled water were added into the tubes to establish a 2:5:2 (v/v/v) ratio mixture of chloroform/ methanol/ water. The samples were shaken vigorously for 1 hour at room temperature. Centrifugation of samples was carried out at 14,800 g for 10 minutes at 4 °C and two layers mix was observed. The extracted metabolites were contained in the aqueous water/methanol phase, whereas the lipid contents were in the organic chloroform phase. Next samples were lyophilised to remove residual liquid.

3.9 Derivatization of soluble metabolites

Sample derivatization for metabolites was carried out prior to GC-MS analysis. For soluble intracellular metabolites in the aqueous phase, a two-step derivatization method was used. The lyophilised samples were dissolved in 50 µl of methoxyamine hydrochloride (20 mg/ml in pyridine) and maintained at 37 °C for 60 minutes for methoximation. Samples were subsequently silylated with 100 µl N-methyl-N-(trimethylsilyl)-trifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (TMCS) at 70 °C for 30 minutes [88]. Samples were shaken gently for two hours after incubation and transferred into vials for further analysis by GC-MS.

3.10 Derivatization for fatty acids

For lipid contents in the organic chloroform phase, boron fluoride-methanol method was used for derivatization according to W. Morrison and L. Smith's [89]. After lyophilisation of organic phase, the dried lipid sample was re-dissolved in 300 µl reagent (25% boron fluoride-methanol, 20% hexane and 55% methanol) and heated at 95 °C for 30 minutes to accomplish trans-esterification. Next, 600 µl hexane and 300 µl distilled water were added for fatty acid methyl ester (FAME) extraction. The mixture was vortexed for one hour at room temperature and the organic phase containing the FAME was transferred into vials for fatty acid analysis in GC-MS.

3.11 GC-MS analysis

The derivatized samples were analyzed with an Agilent Technologies 7890A and 5975C GC-MS system. Metabolites were separated through a HP-5MS capillary

column (30 m $\times$ 0.250 mm i.d.; 0.25 μ m film thickness; Agilent J&W Scientific, Folsom, CA, USA). Each 1 μ l sample was injected into the system by Agilent auto-sampler and separated through column in splitless mode. Helium was utilized as carrier gas and flow rate was set at 1.1 ml/min. Temperature for inlets and MS source were kept at 250 and 230 $^{\circ}$ C, respectively.

For metabolic profiling of soluble intracellular metabolite, the oven temperature was kept at 80 $^{\circ}$ C for 3 minutes and increased to 300 $^{\circ}$ C with a rate of 4 $^{\circ}$ C/min and held for 2 minutes. MS detection was accomplished in electron ionization mode. Mass spectrum was recorded ranging from 35 to 550 m/z with a scan time of 0.3 second.

For lipid profiling of fatty acid, the oven temperature was set at 80 $^{\circ}$ C for 2 minutes and increased to 280 $^{\circ}$ C at a ramp of 8 $^{\circ}$ C/min and held at 280 $^{\circ}$ C for 10 minutes. MS data was acquired in full scan from 35 to 550 m/z with a 0.3 second of scan time.

3.12 Metabolites detection and quantification

Acquisition of the total ion chromatogram and identification of mass spectra identification were carried out by GC-MS solution software (GC/MSD Chemstation Data Analysis, Agilent). NIST08 mass spectral library (National Institute of Standards and Technology) was used as database to identify the detected metabolite peaks. Prior to integration of peak area, reduction of noise and correction of baseline were performed. Normalization of target compounds were accomplished through comparing the peak area relative to that of the added internal control.

For soluble intracellular metabolites peak identification, width at half-height must be more than two seconds, which is the standard to distinguish peaks with noises.

Eligible peaks were selected and subjected to library search. Peaks with similarity index more than 70% were selected to name the candidate compounds. Ribitol was used as internal control to calculate the losses during experimental operation. The selected compounds were relatively quantified by normalizing and comparing the relative chromatograph peak areas.

For fatty acid analysis, heptadecanoic acid (C_{17:0}) was utilized as internal control as it could not be naturally synthesised in yeast. Standard curve of C_{17:0} was obtained by injecting derivative heptadecanoic acid with various concentrations (2, 4, 6, 8, 10, 12 mg/ml). Quantification of FAME could be achieved with reference of Fatty Acid Methyl Esters Standard Mix (C₈-C₂₄, SUPELCO). Series dilutions of FAME standard were obtained as 0.125, 0.25, 0.375, 0.5, 0.625, 0.75 and 1 mg/ml and analyzed by GC/MS. Standard curves of different carbon chain length FAMES were calculated. Identification and quantification of fatty acids in different samples were obtained according to the standard curves.

3.12 Statistical analysis

Three independent experiments were conducted for each sample. The significance of changes of intracellular metabolite levels in three yeast strains at different lipid accumulation phases was determined utilizing Student's t-test. All the detected metabolites were analyzed and a threshold of $p < 0.05$ was used during analysis.

Principal component analysis (PCA) was carried out to determine the changes of patterns of metabolites within comparison and also the altered levels of each metabolite. It has been successfully utilized to set up classification model using

metabolomics data [90,91]. PCA analysis was performed based on three biological replicates for every sample utilizing SIMCA-P+ 12 software (Umetrics, Umea, Sweden).

CHAPTER 4

BIOSYNTHESIS OF LIPIDS IN DIFFERENT YEAST STRAINS

(Part of this chapter was from our paper published in PLoS ONE 8 (12):385532. [92] Permission was obtained from publisher to use this published article in this dissertation.)

4.1 Introduction

Demand for fuels is increasing continually these years and biofuel production has become a research focus to provide an alternative supplement for fuel requirements, which has the advantages of being biodegradable, renewable and environmental friendly. Microbial lipid production has been studied recently as a viable feedstock for the production of fatty acid methyl esters (FAMES).

Oleaginous yeast, which is defined as yeast strains with the ability to accumulate lipids to more than twenty percent of yeast cell dry weight, has caught the attention of research and industrial interest [93]. They exhibit considerable advantages in lipid production. High lipid accumulation ability is one of the most outstanding advantages, some oleaginous yeast strains could produce around 70% lipid of its cell dry weight. They also have high growth rate and culture methods are relatively easy; cultivation could be accomplished in conventional bioreactors [11]. One more advantage is that

the oleaginous yeast could utilize variety of substrates, such as hydrocarbons, carbohydrates and even waste materials, and convert them into lipids.

A number of oleaginous yeast species are reported in literatures, mostly focusing on their industrial lipid production. Several publications have studied the major metabolite pathways and some key enzymes [9,21,94]. The lipid accumulation mechanism is rarely reported from ‘omics’ levels to offer a comprehensive understanding.

In this study, we have performed a comparative proteomics profile and metabolic profile of three yeast strains and explored the possible mechanism of lipid accumulation by comparing the profiles between two oleaginous yeast strains and non-oleaginous strain during the various growth and lipid accumulation phases. *R. toruloides* was widely studied as industrial strain for lipid extraction and usage due to its surprisingly high lipid productivity, which was reported to be 67.5% of their dry cell weight [95]. *C. albidus* was reported to produce up to 46.3% (w/w) lipid [96] and chosen as a medium lipid content strain. In order to further identify the differences between oleaginous yeast and non-oleaginous yeast, *S. cerevisiae* was also selected to be a ‘control’, as it is considered non-oleaginous yeast, due to its low lipid content, which consists of no more than 15% of its biomass [97].

Before proteomics and metabolomics study, the cultivation and lipid accumulation conditions were studied to estimate that the commercial *S. cerevisiae*, *C. albidus* and *R. toruloides* obtained from different vendors were indeed of low, medium and high lipid accumulation abilities. Lipid accumulation could be triggered by nitrogen depletion in the culture medium and detailed mechanism has been described in

Section 1.4. Culture conditions, such as initial glucose concentrations, amount of inoculation were studied. Optimized culture medium constituent and lipid accumulation methods were acquired to ensure that each yeast strain has reached its full capacity in lipid accumulation. Cells at early, middle and late lipid accumulation stages of three strains were collected for further profiling study.

4.2 Experimental procedures

Two-stage cultivation were performed on *C. albidus* and *R. toruloides*, cells were grown in YEPD medium for proliferation and transferred into nitrogen limited medium to trigger lipid accumulation. In order to determine the suitable glucose concentrations in the nitrogen limited medium, different glucose concentrations (20, 30, 40, 60, 80 and 100 g/l) were used to grow the oleaginous cells to determine the substrate inhibition. Cell density of inoculation when transferring cells from YEPD to nitrogen limited medium were also optimized, with final cell densities varying at 0.25×10^8 /ml, 0.5×10^8 /ml, 1×10^8 /ml, 2×10^8 /ml, 3×10^8 /ml and 4×10^8 /ml and cultured for 96 hours. For *S. cerevisiae*, both growth and lipid production were carried out in YEPD medium. Same cultivation procedure for oleaginous yeast strains was tried for *S. cerevisiae*, however, it experienced a poor growth and abnormal lipid production. Growth curve, cell dry weight, residual glucose and lipid content were measured.

4.3 Results and discussion

4.3.1 Substrate inhibition study of oleaginous yeast

C. albidus and *R. toruloides* were firstly cultured in YEPD medium and used as inoculum in nitrogen limited medium to study lipid accumulation. In order to determine the effect of substrate concentration and optimize it for growth, batch experiments in 250 ml flasks were carried out in nitrogen limited medium with different glucose concentrations, ranging from 20 to 100 g/l. Figure 4.1 A and figure 4.2 A (page 62 and 63) has plotted the optical density of *C. albidus* and *R. toruloides* cells against fermentation time. Absorbance were recorded every 4 hours to indicate the growth of cells with glucose as sole carbon source for initial 24 hours. It could be concluded that cell growth was repressed when glucose concentration reached 60 g/l and expressed more severe repression when higher glucose concentrations were utilized.

After 96 hours, the biomass, lipid amount, lipid content and residue glucose of *C. albidus* and *R. toruloides* against different initial glucose concentrations were analyzed respectively.

For *C. albidus*, along with glucose concentration increase from 20 to 40 g/l, the biomass and lipid content of cell dry weight have increased respectively from 9.02 g/l and 22.81% to 10.89 g/l and 26.43% (Figure 4.1B, page 62). They were reduced when higher concentrations of glucose were used and higher residual glucose has been observed.

Figure 4.1 Effect of initial glucose concentrations in *C. albidus*. (A: Effect on cell growth; B: Effect on biomass and lipid accumulation).

For *R. toruloides*, when the initial concentration of glucose increased from 20 to 40 g/l, the biomass and lipid content has accumulated from 7.79 g/l and 33.17% to 9.12 g/l and 46.14% (Figure 4.2B, page 63). They have also experienced a decrease when the initial glucose concentrations increased to higher ones. As a result, initial

concentration of 40 g/l glucose was used as carbon source in the following lipid accumulation experiments.

Figure 4.2 Effect of initial glucose concentrations in *R. toruloides*. (A: Effect on cell growth; B: Effect on biomass and lipid accumulation).

4.3.2 Cell density optimization

As two stage cultivation method was utilized, the amount of cells to inoculate into nitrogen limited medium after proliferation in YEPD medium was important for lipid accumulation in oleaginous yeast.

After 30 hours cultivation in YEPD medium, the cells of *C. albidus* and *R. toruloides* were collected by centrifugation and washed twice by pre-chilled distilled water. Different amount of cells of each strain were transferred into nitrogen limited medium to obtain various cell densities 0.25×10^8 /ml, 0.5×10^8 /ml, 1×10^8 /ml, 2×10^8 /ml, 3×10^8 /ml and 4×10^8 /ml and cultured in 50 ml nitrogen limited medium in 250 ml flasks at 28 °C.

C. albidus and *R. toruloides* cells were cultivated in nitrogen limited medium respectively for 96 hours, and the biomass, lipid amount and lipid content were determined and calculated (Figure 4.3, page 65). It could be concluded that lipid content roughly decreased together with the increase of cell density, reaching a highest content of 27.74% and 45.03% for *C. albidus* and *R. toruloides*, respectively when they were inoculated with cell density of 0.5×10^8 /ml. As for 0.25×10^8 /ml of cell density, the lipid content was not varying a lot, whereas the biomass and lipid amount were slightly lower than that of 0.5×10^8 /ml.

Figure 4.3 Effects of cell density on lipid accumulation in oleaginous yeast. (A: *C. albidus*; B: *R. toruloides*).

For *C. albidus*, when cell density increased from 0.5×10^8 /ml to 4×10^8 /ml, the biomass increased from 10.03 g/l to 26.70 g/l, whereas the lipid content decreased dramatically from 27.74% to 10.30%. As for *R. toruloides*, the biomass has been observed to increase from 9.08 g/l to 23.81 g/l. However, the lipid content also experienced a decrease from 45.03% to 14.85%. Residual glucose was depleting at the end of cultivation for both strains regardless of cell density. As a result, cell density of 0.5×10^8 /ml was utilized as appropriate inoculum index in further lipid accumulation experiments for oleaginous yeast strains.

This indicated the diminishing of ability to accumulate lipid when cell density was above certain range in oleaginous yeast. High cell density cultivation could be beneficial in raising the biomass and eliminating contamination of other microorganisms. However, some disadvantages could come along, such as substrate depression, metabolite depression, insufficient supply of oxygen and high dissolution of CO₂ [98].

4.3.3 Lipid content comparison

After estimation of appropriate initial substrate concentration and cell density, oleaginous yeast strains were cultured to estimate the lipid accumulation ability, together with cultivation of non-oleaginous yeast *S. cerevisiae*.

For *S. cerevisiae*, cells were cultured in YEPD medium and collected at early, middle and late lipid accumulation stages, which were 8 h, 18 h and 36 h, respectively (Figure 4.4, page 67). The lipid content was very low, reaching the highest level at 7.98% at 36 h. Two stage cultivation was also introduced to *S. cerevisiae* cells, however, poor growth and lipid accumulation were experienced after transferring the cells into

nitrogen limited medium. This was due to the lack of certain enzymes in non-oileaginous yeast and will be discussed later in proteomics part. Thus, the rich medium YEPD was used for lipid accumulation study for *S. cerevisiae*.

Figure 4.4 Growth curve of *S. cerevisiae*, *R. toruloides* and *C. albidus* in YEPD medium. Optical density was measured at wavelength of 600 nm by taking 1 ml culture medium and diluting into proper concentrations. Data from three replications were analyzed to calculate means and standard errors, indicated by thin lines.

The two oleaginous yeast strains, *C. albidus* and *R. toruloides*, were cultured in YEPD medium (Figure 4.4, page 67) for proliferation, until the residual glucose approached zero at 30 h. The lipid contents were 10.73% and 14.70% respectively in this step. Then *C. albidus* and *R. toruloides* cells were centrifuged and transferred into nitrogen limited medium with a high C/N ratio of 135, to stimulate lipid accumulation. *C. albidus* showed the ability to grow to a cell density of 10.78 g/L and accumulate lipid up to 27.74% of dry cell weight at 96 hour (Figure 4.5A, page 69). *R. toruloides* reached a cell density of 9.08 g/L and lipid content of 45.03%, when glucose was exhausted at 96 hour (Figure 4.5B, page 69).

Early, middle and late stages denoted different lipid accumulation periods in yeast cells. These three stages were defined according to the lipid content change in each strain. The early stage was determined when cells have adapted to the environment and started to produce lipids. The middle stage was determined by the time point at which lipid content achieved median value of the initial and final lipid content. The late stage was decided when the lipid accumulation came to the end and lipid content decreased thereafter. For *S. cerevisiae*, the first eight hours were mainly for growth adaption, the lipid accumulation started gradually after that. After 36 h, the consumption of lipid started as nutrient in the medium was used up. For oleaginous yeast strains, cells were transferred from rich medium into nitrogen limited medium to trigger the lipid accumulation. The cells have adapted the change in the first 12 h and accumulated lipids fast in the following 12 hours. Further lipid accumulation were undertook till 96 h.

Fig 4.5 Substance level dynamic change of oleaginous yeast. (A: *C. albidus*; B: *R. toruloides*).

In summary, the three strains were compared in the following 3 phases: early lipid accumulation stage comparison (*S. cerevisiae* 8 h, *C. albidus* 12 h in nitrogen limited medium, *R. toruloides* 12 h in nitrogen limited medium); middle lipid accumulation stage comparison (*S. cerevisiae* 18 h, *C. albidus* 24 h in nitrogen limited medium, *R. toruloides* 24 h in nitrogen limited medium) and late lipid accumulation stage comparison (*S. cerevisiae* 36 h, *C. albidus* 96 h in nitrogen limited medium, *R. toruloides* 96 h in nitrogen limited medium).

The data above clearly showed lipid accumulation differences between the non-oleaginous yeast *S. cerevisiae*, and oleaginous yeast strains *C. albidus* and *R. toruloides*.

4.4 Chapter conclusions

Lipid accumulation conditions were successfully optimized and proper cultivation parameters were obtained. Non-oleaginous yeast *S. cerevisiae* and oleaginous yeast *C. albidus* and *R. toruloides* were cultured and estimated that they were of low, medium and high levels of lipid contents. Although the lipid content of each strain was lower than that of previously published in literature, it might be due to the fact that the cells were cultured using bioreactors and fed-batch method in the literatures, as compared to shake flasks in a laboratory scale in this study, which was sufficient for further proteomics and metabolomics study.

CHAPTER 5

COMPARATIVE PROTEOMICS PROFILE OF LIPID-CUMULATING OLEAGINOUS YEAST: AN ITRAQ-COUPLED 2-D LC-MS/MS ANALYSIS

(Part of this section was from our paper published in PLoS ONE 8 (12):385532. [92] Permission was obtained from publisher to use this published article in this dissertation.)

5.1 Introduction

Accumulation of intracellular lipid in oleaginous yeast cells has been studied for providing an alternative supply for energy, biofuel. Numerous studies have been conducted on increasing lipid content in oleaginous yeasts. The industrial profit of derived biofuels has led to more research, exploring the mechanism of the prominent ability of lipid storage in oleaginous yeast. Metabolic investigations suggest that under nitrogen-limited condition, a series of reactions would be activated in oleaginous yeast strains, contributing to a large scale production of acetyl-CoA, precursor for fatty acid synthesis [20]. Several key enzymes are also found to be crucial, for example, ATP-citrate lyase (ATP-CL) is found to be uniquely existing in oleaginous yeast strains and has never been found in non-oleaginous ones [99]. It

could catalyze conversion of citric acid to acetyl-CoA and thus continuously provides basic subunit for fatty acid synthesis.

Although major metabolite pathways and some key enzymes have been determined, lipid accumulation mechanism is rarely reported from a proteomics level, which would contribute to a more comprehensive understanding. In this chapter, we performed a comparative proteomics profile of three yeast strains and explored the possible mechanism of lipid accumulation by comparing the protein profile between two oleaginous yeast strains *C. albidus* and *R. toruloides* and non-oleaginous strain *S. cerevisiae* during the various growth and lipid accumulation phase.

Lipid accumulation was studied in Chapter 3 and *S. cerevisiae*, *C. albidus* and *R. toruloides* has been estimated to own low, medium and high lipid accumulation abilities, respectively. In order to study the protein expression differences in these three strains, a comparative proteomics study was performed using two dimensional LC-MS/MS. A time course research was introduced to investigate protein changes during different lipid accumulation stages in various yeast strains.

Two dimensional (2-D) LC-MS/MS analysis has been more preferable in proteomics profiling analysis due to its higher efficiency and accuracy comparing to the traditional 2-D gel electrophoresis, especially when coupled with iTRAQ labelling method. Isobaric tag for relative and absolute quantitation (iTRAQ) labelling method is a MS-based approach allowing multiplexing of samples and the quantification of multiple peptides of each protein relatively, which relies on the derivatization of primary amino groups in the intact proteins [100,101].

After labelling with different isotopes, the derived peptides in each strain would experience an identical mass and retention time during LC-MS analysis. The precursor ions are analyzed by the following MS/MS analysis. The intensities of reporter ions would relatively indicate the concentration of peptides correspondingly [102]. The relative amount of one single peptide was determined by dividing the peak areas obtained with different isotope tags. Overlapping isotopic contributions were corrected for calculation of area ratios of peaks, estimating the relative abundances of one particular peptide. The results of protein changes were also verified by western blot analysis and real-time RT-PCR.

5.2 Experimental procedures

Cultivation of *S. cerevisiae*, *C. albidus* and *R. toruloides* cells for lipid accumulation was described in Chapter 3. Under optimized culture conditions for lipid accumulation for each strain, cells were collected at early, middle and late lipid accumulation stages, respectively for each yeast strain. Specifically, *S. cerevisiae* cells were collected at 8 h, 18 h and 36 h, and *C. albidus* and *R. toruloides* cells were harvested at 12 h, 24 h and 96 h. They were centrifuged at 10,000 rpm for 5 minutes at 4 °C. Protein extraction, digestion and peptide labelling were described in Section 2.3. Specific labelling and experimental design was demonstrated in Table 5.1. *S. cerevisiae*, *C. albidus* and *R. toruloides* were labelled by iTRAQ tag 114, 115 and 116, respectively in each experiment.

Table 5.1 Experimental description for time course based iTRAQ labelling for yeast strains.

	<i>S. cerevisiae</i>	<i>C. albidus</i>	<i>R. toruloides</i>
	iTRAQ 114	iTRAQ 115	iTRAQ 116
Early stage	8 h	12 h	12 h
Middle stage	18 h	24 h	24 h
Late stage	36 h	96 h	96 h

In total, three time course based experiments were performed on 2-D LC-MS/MS system, which was described in Section 2.4 in detail. Spectrum Mill MS Proteomics Workbench was introduced to analyze all the mass data to quantitate peptides and identify the proteins. Each MS/MS spectrum was searched based on species of Microorganisms against the UniProt_Swiss_Prot database. Parameter restrictions were set according to Section 2.5.

Three independent batches were performed for each lipid accumulation stage sample to increase statistically evidence of protein expression. Proteins quantification with relative expression ≥ 1.2 or ≤ 0.8 was selected for further analysis. The proteomics differences were further confirmed by western blot, methods and antibodies utilized were detailed described in Section 2.6. Real-time RT-PCR was also accomplished to partially confirm the protein expression in each strain.

5.3 Results and discussion

5.3.1 ITRAQ analysis and protein identification

Protein extraction was difficult due to the rigid yeast cell wall and it took effort to maintain the activities of proteins, as they were easily denatured when treated too harsh during cell wall disruption. Considering that the lipid content was high in oleaginous yeast strains, a higher concentration of denaturing reagents was used in the lysis buffer to help efficiently break lipid-protein and lipid-lipid interactions and increase protein solubility. Excess addition of detergent Triton X-100 in lysis buffer successfully overcame this problem without denaturing the proteins, as Triton X-100 was more efficient at breaking the lipid-protein and lipid-lipid interactions, rather than protein-protein interactions [103,104].

Agitation with glass beads could efficiently break the cells with proper lysis buffer, but the excess lipid in oleaginous yeast cells affected the protein charge and molecular weight seriously as they would bind proteins through hydrophobic interactions, resulting in insoluble protein-lipid complex formation. Hence organic solvents treatment was further used to thoroughly remove the lipid contaminants. In the presence of chloroform/ methanol (2:1, v/v), the combination of 10% TCA/acetone and aqueous TCA wash steps showed effective removal of lipids.

After extraction and lipid removal, protein samples were quantified and digested into peptides and labelled as follows: *S. cerevisiae* with iTRAQ 114; *C. albidus* with iTRAQ 115 and *R. toruloides* with iTRAQ 116. Samples from early, middle and late lipid accumulation stages were analyzed respectively. Three independent

experiments of protein profile by 2-D nano-LC-MS/MS analysis were conducted for each group of sample.

Figure 5.1 (page 77) has shown the results of total intensity chromatogram of peptides of early lipid accumulation stage sample, which were eluted by gradient diluted ammonium formate. In addition, one representative MS spectrum was also presented in Figure 5.2 (page 78), indicating four peptide precursors and marked by red diamond. Figure 5.3 (page 79) was one representative peptide fragmentation spectrum of pyruvate kinase: GVEVLPVDFIADAFSADANTK.

Figure 5.1 Total intensity chromatogram of peptides of early lipid accumulation stage sample. The ‘X’ axis is the acquisition time, and the ‘Y’ axis is the chromatogram of peptides. Samples were eluted with gradient diluted ammonium formate, and totally ten run were accomplished. Each color represented the chromatogram of peptides eluted at respective concentration of ammonium formate. Black, 20 mM; red, 40, mM; dark green, 60 mM; dark blue, 80 mM; purple, 100 mM; brown, 300 mM; light green, 500 mM; light blue, 500 mM; orange, 1000 mM; pink, 1000 mM.

Figure 5.2 Representative MS spectrum indicating four peptide precursors (524.74, 540, 579.34 and 655.66) marked by red diamond.

Figure 5.3 One representative peptide fragmentation spectrum of pyruvate kinase: GVEVLPVDFIADAFSADANTK.

5.3.2 Differentially expressed proteins within comparison

Differential protein expression patterns were assessed through analysing the changes in protein expression among comparison of different lipid accumulation stages of *S. cerevisiae*, *C. albidus* and *R. toruloides*. The identified proteins were classified into eight functional groups: metabolism, biosynthesis, transportation, signal transduction, stress response, structural proteins, ribosomal proteins, and unclassified proteins with other functions.

Based on three independent batches of experiments for each group, a total of 132, 122 and 116 proteins were constantly identified at early, middle and late lipid accumulation stages, respectively.

At the early lipid accumulation stage, 71 out of the 132 proteins were found to meet the strict criteria for reliable statistical analysis and have considerable fold change (relative expression ≥ 1.2 or ≤ 0.8). Majority of these proteins had functions involving in metabolism and ribosomal biosynthesis. At the middle stage, 62 among the 122 proteins passed through the screening; meanwhile, 55 of 116 proteins at the late stage of lipid accumulation went through the filtration. Higher percentages of protein expression in these two stages were found in metabolism, biosynthesis and stress response (Figure 5.4; Appendix 1). Fold change of proteins of different functional distribution at early, middle and late lipid accumulation stages was also intuitively figured out in Figure 5.5 (page 81).

Figure 5.4 Functional distribution of proteins identified during each lipid accumulation stage.

Figure 5.5 Fold change of proteins of different functional distribution at early, middle and late lipid accumulation stages. Black color represented that the protein was not identified at that stage.

Among all the listed proteins, 15 of them were found up-regulated comparing oleaginous yeast strains with *S. cerevisiae*. Five of these 15 proteins were identified in each time point and these included elongation factor 1-alpha, ATP synthase subunit beta, cytochrome c iso-1, 78 kDa glucose-regulated protein and ubiquitin. ADP, ATP carrier protein 2 was only found in 24 h and 96 h samples. Four were identified only in 96 h sample, which were 6-phosphogluconate dehydrogenase decarboxylating 1, NADP-specific glutamate dehydrogenase 2, heat shock protein 70 and thioredoxin-2. The other five proteins were found during 12 h or 24 h samples; including histone H4, heat shock protein HSP90-alpha, phosphomannomutase, luminal-binding protein 4 and cytochrome c. One protein, citrate synthase-mitochondrial (Cit1p) was found up-regulated only at early lipid accumulation stage and decreased thereafter, indicating the acceleration of TCA cycle to meet the excess demand for energy, when lipid was being accumulated. When they were compared among three strains, higher concentrations were detected in *R. toruloides*, following by *C. albidus* and indicated larger supply of energy for oleaginous yeast.

Proteins of particular importance were illustrated as follows (Figure 5.6, page 87). NADP-specific glutamate dehydrogenase 2 (NADP-Gdh3p) is involved in glutamate biosynthesis pathway, catalyzing the reversible reaction between 2-oxoglutarate and glutamate. It was suggested that under low level of nitrogen concentration, this enzyme was found to increase and served as cellular nitrogen supply [105-107]. This would help explain the up-regulation of Gdh3p in the oleaginous yeast strains, especially its unique existence in late lipid accumulation stage, as the nitrogen source was extraordinarily low at 96 h.

6-phosphogluconate dehydrogenase, decarboxylating 1 (Gnd1p) catalyzes the oxidative decarboxylation of 6-phosphogluconate to ribulose 5-phosphate, reducing NADP to NADPH at the same time. The generation of NADPH is critical for various of reductive biosynthetic reactions, especially lipid production, as well as protecting the cells against oxidative stress [108]. The up-regulation of Gnd1p in this study was thus important as it indicated sufficient supplement of NADPH was given in oleaginous yeast in lipid accumulation process.

ADP, ATP carrier protein 2 (Aac2p) catalyzes the exchange of ADP and ATP across mitochondrial membrane, which imports ADP into mitochondria and pumps the ATP out into cytosol [109]. ATP synthase subunit beta, mitochondrial (Atp2p) converts the energy of proton gradient and produces ATP from ADP across the membrane initiated by electron transport complexes of the respiratory chain [110] and its activity and coupling exhibits strict lipid requirement [111]. The up-regulation of Aac2p and Atp2p in oleaginous yeast strains would not only provide steady supplement of energy, but also was the response for the excess lipid in oleaginous yeast strains. Moreover, higher degree of up-regulation of Gnd1p, Aac2p and Atp2p were detected in *R. toruloides* comparing with *C. albidus*, which helped explain the higher lipid accumulation ability of *R. toruloides*.

The majority of identified proteins with remarkable change were down-regulated. Interestingly, almost all the enzymes of glycolysis (Fba1p, Pyk1p, Tdh3p, Pgi1p, Pfkfb3p, Glk1p, Tpi1p, Eno2p and Gpm1p) were found to be down-regulated during comparison between oleaginous strains and *S. cerevisiae*, as well as two gluconeogenesis proteins (Pyc2p and Pck1p). Detailed explanation was difficult, as

in the ordinary course of events, the glycolysis and gluconeogenesis pathways should be quite active so as to apply sufficient energy source for fatty acid generation. However, this down-regulation would be reasonable when the growth environment difference between oleaginous strains and non-oleaginous one was taken into consideration. As for lipid accumulation, *R. toruloides* and *C. albidus* strains were transferred from proliferation culture medium into the nitrogen-limited medium. The substrate and energy for lipid formation were accumulated gradually in early phase of growth. Along with lipid accumulation, metabolic flux was experiencing a shift to lipid biosynthesis.

Idh2p, Mdh1p, Icl1p and Cit1p were engaged in tricarboxylic acid cycle (TCA cycle) and also found to be down-regulated seriously at middle and late lipid accumulation stages, indicating the down shift of TCA cycle when cells were experiencing lipid accumulation. When the oleaginous yeast cells were transferred into nitrogen-limited medium and experienced gradual depletion of nitrogen source in the environment, the activity of particular enzyme, adenosine monophosphate (AMP) desaminase was sharply increased, leading to the cleavage of AMP [112]. The activity of Idh2p depended on the existence of AMP and was thus degraded critically, leading to a downturn of TCA cycle. This would help explain the down-regulation of enzymes in TCA cycle.

Pyruvate decarboxylase isozyme 1 (Pdc1p) and alcohol dehydrogenase 1 (Adh1p) catalyzes the pathways from pyruvate to ethanol. They were both down-regulated significantly and it would be easily found that the inhibition of these enzymes contributed to the shift down of the pathways competing with TCA cycle. In other

words, the down-regulation of these enzymes in oleaginous yeasts would decrease the energy 'lose', lead more metabolites from glucose into TCA cycle, promote pyruvate metabolism and produce more acetyl-CoA, which is the precursor for fatty acid.

Meanwhile, it was noticed that most of the proteins with metabolism functions were further down-regulated in *R. toruloides* than *C. albidus*, when comparing with *S. cerevisiae*. This indicated a higher degree of shift from metabolic flux to lipid biosynthesis along lipid accumulation in oleaginous yeast strains.

Many proteins found were involved in stress response, such as heat shock proteins which were found up-regulated, in the early and medium lipid accumulation stage. This would be explained as the cells were grown in a high glucose and low nitrogen medium, which was considered restrictive environment, thus triggering an up-regulation of stress response related proteins. Oleaginous yeast strains produce lipids as a protective mechanism under the stressed circumstances. Meanwhile, at 96 h, it was shown that majority of stress response proteins were down-regulated comparing with those at 12 h or 24 h.

The degree of down-regulation was higher in *R. toruloides* than *C. albidus*. This would be clarified as the yeast cells containing high level of unsaturated fatty acid at late stage and decreased stress sensitivity according to M. Chatterjee and S. Khalawan [113]. At late lipid accumulation stage, higher lipid contents were experienced and less stress could be encountered, which explained the down-regulation of stress response proteins in oleaginous yeast.

Numerous ribosomal proteins were also identified, whereas reduced levels of them were reflected over time; most of them were down-regulated during 12 h and 24 h and even undetectable at 96 h, consistent with reduced protein yield along lipid accumulation. The initial period of lipid production should experience the highest protein level, so as to prepare for lipid accumulation. At late stage, many proteins were degraded and protein synthesis was not as vigorous as starting stage in order to preserve energy for lipid production.

Figure 5.6 Key protein changes in yeast strains mapped onto metabolic pathways. Red color represented up-regulation, whereas green color represented down-regulation in oleaginous yeast strains as compared to *S. cerevisiae*. The blue arrow represented the general flow of metabolic flux. Abbreviations for the enzymes were GLK, glucokinase; PGI, glucose-6-phosphate isomerase; FBA, fructose-bisphosphate aldolase; TPI, triosephosphate isomerase; TDH, glyceraldehyde-3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; GPM, phosphoglycerate mutase; ENO, enolase; PYK, pyruvate kinase; PDC, pyruvate decarboxylase isozyme; ADH, alcohol dehydrogenase; CIT, citrate synthase; IDH, isocitrate dehydrogenase [NAD] subunit; ICL, isocitrate lyase; MDH, malate dehydrogenase; GND, 6-phosphogluconate dehydrogenase; AAC, ADP/ATP carrier protein.

Although the function of proteins with particular importance in lipid accumulation process were discussed above, it is still difficult to explain the functions of some proteins in detail. Also, few proteins were found in relate to lipid metabolism directly. This was understandable as many proteins were hydrophobic and the abundance was low, which have low solubility and the separation behaviour is poor. Moreover, the comparison and discussion of proteomic data of these three yeast strains were based on database searched against 'microorganism', rather than specific model microbes, remarkable challenges existed for accurate data analysis. These problems were also reported by Z. Zhao *et al.* who studied the comparative proteomics analysis of *R. toruloides* [104].

Many essential genes and proteins were reported to be relevant in lipid accumulation process, whereas they were not identified in our results. Kamisaka *et al.* has reported the overexpression of diacylglycerol acyltransferase (Dga1p) in *S. cerevisiae* and changed it into an oleaginous yeast [114]. He has also studied the increase of lipid content in *S. cerevisiae* by disruption of particular genes, such as SNF2 which encodes the transcription regulatory protein Snf2p [115]. W. Runguphan and J. Keasling also reported the increase of lipids in *S. cerevisiae* by overexpression of acetyl-CoA carboxylase (Acc1p), fatty acid synthase 1 (Fas1p) and fatty acid synthase 2 (Fas2p) [116].

However, these proteins were not successfully identified in our proteomics study. This has revealed a weakness of our iTRAQ labelled proteomics study that the detailed protein data could only be interpreted by the software. Only the proteins that were found simultaneously in all three strains would give a spectrum and be identified

by the system. When labelling the samples of three yeast strains, the amino active part of iTRAQ reagent was bind to the N-terminus and lysine side chain of the digested peptides from protein samples. The more proteins inside the samples, the more iTRAQ reagent would be bound to them. Rather than identifying and quantifying proteins individually in each strain, the samples of three strains were combined together to obtain fast determination.

After 2D LC-MS detection, a relative quantification of proteins in different samples was achieved by comparing the relative peak area of each labelled one. A ratio indicating protein fold change was given instead of examining the exact protein amount in each strain and calculating the variations thereafter. Thus, if a protein was absent in any of these three strains, it would not be reported. More efforts will be devoted into further improvement of the methods to expand the amount of proteins that could be identified.

5.3.3 Western blot analysis of differentially expressed proteins

In order to validate the differentially expressed proteins detected by LC-MS/MS, western blot analysis was performed to determine the degree of protein expressions. Pyruvate kinase 1 (Pgk1p), alcohol dehydrogenase (Adh1p), Histone H4 (Hhf1p) and Ubiquitin (Ubi1p) expressions in each strain at different lipid accumulation stages were estimated.

Results were shown in Figure 5.7-5.9 (A) (page 90-91), indicating the changes of protein levels, which were consistent with the variations of LC-MS/MS analysis generally. For example, the Pgk1p and Adh1p levels were decreased in *C. albidus* and *R. toruloides*, comparing with those in *S. cerevisiae* at early lipid accumulation

stage. While Hhf1p and Ubi1p both showed up-regulation in oleaginous yeast than *S. cerevisiae*. Quantification of protein levels of Pgc1p, Adh1p, Hhf1p and Ubi1p based on western blot was calculated by Image Pro plus 6.0 software and shown in Figure 5.7-5.9 (B) (page 91).

Figure 5.7 Western blot result of protein expressions in *S. cerevisiae* (S), *C. albidus* (C) and *R. toruloides* (R) at early lipid accumulation stage. (A) Protein expression changes. (B) Quantification of protein level changes based on western blot analysis. The ratio between *C. albidus* or *R. toruloides* and *S. cerevisiae* was shown.

Figure 5.8 Western blot result of protein expressions in *S. cerevisiae* (S), *C. albicus* (C) and *R. toruloides* (R) at middle lipid accumulation stage. (A) Protein expression changes. (B) Quantification of protein level changes based on western blot analysis. The ratio between *C. albicus* or *R. toruloides* and *S. cerevisiae* was shown.

Figure 5.9 Western blot result of protein expressions in *S. cerevisiae* (S), *C. albicus* (C) and *R. toruloides* (R) at late lipid accumulation stage. (A) Protein expression changes. (B) Quantification of protein level changes based on western blot analysis. The ratio between *C. albicus* or *R. toruloides* and *S. cerevisiae* was shown.

5.3.4 Real-time RT-PCR based investigation for mRNA level determination

As protein precursor, mRNA is of much importance and the cellular level changes of mRNA could reflect the dynamic protein expression. Real-time RT-PCR was accomplished to quantify the expression at mRNA level of four selected genes. Actin (Act1p) from *S. cerevisiae* was used as endogenous control. Primers were designed using Primer Premier Software and final product lengths were controlled to be between 100-250 bp. Three independent experiments for three strains at different lipid accumulation stages were analyzed. In each sample, the total mRNA were extracted and utilized at same concentrations as templates. The mRNA levels in *S. cerevisiae* were taken as control value (100%), the mRNA levels of *C. albidus* and *R. toruloides* were compared with that of *S. cerevisiae* (Figure 5.10, page 93).

Overall, the validation results were consistent with LC-MS/MS data. Pgk1p and Adh1p showed similar down-regulation in different lipid accumulation periods. Hhf1p in early and middle lipid accumulation stages expressed the same potential changes with that of LC-MS/MS result. However, down regulation was detected in late stage, whereas Hhf1p was absent in LC-MS/MS analysis. Ubi1p showed down-regulation at every stage, whereas its up-regulation was detected at early and middle lipid accumulation stages by LC-MS/MS. This inconsistency may be explained by the post-translational regulation and modification, such as glycosylation, phosphorylation and acetylation [117-119].

Figure 5.10 Real-time RT-PCR analysis of mRNA levels of differentially expressed proteins in *S. cerevisiae*, *C. albidus* and *R. toruloides*. (A: Early lipid accumulation stage; B: Middle stage; C: Late stage).

5.3.5 Influence of different culture conditions

Culture conditions were different for *S. cerevisiae* and the two oleaginous yeast strains. The former was cultured only in rich medium for both growth and lipid production, whereas the latter was transferred to nitrogen limited medium to further trigger lipid accumulation. It was a challenge as a slight change in the cultural environment would affect the metabolism of yeast cells. However, we tried to culture *S. cerevisiae* using the same nitrogen limited medium of oleaginous yeast strains. They cannot produce lipids normally and experienced poor growth instead, whereas the two oleaginous yeast strains grew vigorously along lipid accumulation.

When the yeast cells encountered with nitrogen exhaustion, the intracellular AMP (adenosine monophosphate) concentration will be decreased since the AMP-desaminase will cleave AMP into IMP (inosine monophosphate) and NH_4^+ ions, which could constitute a complementary nitrogen source to support the cell [29]. Meanwhile, the activity of NAD^+ isocitrate dehydrogenase, which is activated by AMP, will decrease if the intra-cellular AMP concentration drops [32]. As a result, the iso-citric acid, which is found to be in equilibrium with citrate, will accumulate inside the mitochondrion. In exchange with malate, citrate will enter the cytoplasm as long as the intra-mitochondrial citric acid concentration reaches a critical value [33]. Citric acid will finally be cleaved to acetyl-CoA and oxaloacetate by ATP-citrate lyase (ATP-CL), generate cellular fatty acids and provide building blocks for lipid accumulation [26,112,120]. However, the ATP-CL gene only exists in oleaginous microorganisms and being absent in non-oleaginous ones. If ATP-CL enzyme cannot be found inside of yeast cells (in our case, the *S. cerevisiae*) which

are cultivated in nitrogen limited mediums, the citric acid will either be secreted into culture medium or provoke the inhibition of the 6-phosphoro-fructokinase, resulting in the accumulation of intracellular polysaccharides [121]. Therefore, different culture mediums were used to culture the three strains in order to prompt them to produce the highest contents of lipid respectively. Comparison among the strains of low, medium and high lipid accumulation ability at different stages was carried out afterwards.

5.4 Chapter conclusions

This study successfully set up a platform of time-course proteomic profiling in yeast and to our knowledge, the first one comparing between non-oleaginous with oleaginous yeast strains on proteomics level using iTRAQ-coupled 2-D LC-MS/MS method. It provided an overall reflection from proteomic level, contributing to a better understanding of excessive lipid storage in oleaginous yeast strains and hopefully would contribute to further genetic engineering for higher lipid accumulation in yeast.

CHAPTER 6

COMPARATIVE METABOLIC PROFILING OF YEAST STRAINS DURING DIFFERENT GROWTH PHASES IN LIPID PRODUCTION

6.1 Introduction

Metabolomics is a systematic study of small molecule metabolites in organisms. Small molecules take part in biochemical processes and belong to various categories, such as organic acids, amino acids, fatty acids, carbohydrates and so on. They could provide abundance of information about the functioning status in a living system. Metabolites can interact with proteins or work together with enzymes to affect the cellular pathways and form the metabolism networking.

Oleaginous yeast has been proved to be a competitive microorganism to produce lipids and potentially utilized for biofuel production to apply an alternative energy supply. We have previously studied the lipid accumulation by oleaginous yeast *C. albidus* and *R. toruloides* under appropriate culture conditions. An LC-MS/MS based proteomics platform was also described in Chapter 5 to investigate the protein level changes among comparison of non-oleaginous yeast *S. cerevisiae* and these two oleaginous yeast strains [92]. Our previous study has identified many differentially expressed proteins during lipid accumulation process in different strains.

The proteomics data has offered insights to recognize the lipid accumulation in oleaginous yeast, while metabolomics could help shed light on understanding the possible mechanism of lipid accumulation as it could reveal the downstream changes of proteome. GC-MS has been proved to effectively identify and quantify metabolites in organisms with high resolution and sensitivity and utilized in this study as major analytical method. The soluble intracellular metabolites and fatty acids in yeast cells were successfully analyzed through GC-MS system.

Comparative metabolic profile of yeast strains with different lipid producing abilities was performed using *S. cerevisiae*, *C. albidus* and *R. toruloides*, which were estimated to produce low, medium and high lipid contents, respectively. Relationship of lipid content and metabolite changes during different growth and lipid accumulation phases was also investigated by comparing the metabolic profile between oleaginous yeast strains and non-oleaginous strain. Metabolomics analysis was carried out and complemented the proteomics analysis, revealing the changes in metabolites which are truly reflective of the physiological state of cells.

6.2 Experimental procedures

Non-oleaginous yeast strain *S. cerevisiae* and oleaginous yeast strains *C. albidus* and *R. toruloides* were cultured under optimized conditions for lipid accumulation and described in detail in Section 4.2. Cells were collected at early, middle and late lipid accumulation stages for each yeast strain. For *S. cerevisiae*, samples were collected at 8 h, 18 h and 36 h, and for *C. albidus* and *R. toruloides* cells were harvested at 12 h, 24 h and 96 h, respectively. Cell growth and dry weight were determined and presented in previous Section 4.3.3. Same amount of cells in each sample were used

for extraction of metabolites. Soluble intracellular metabolites and lipids were extracted, lyophilized to dryness and derivatized respectively according to the method in Section 4.8-4.10.

GC-MS analysis was performed to analyze soluble metabolites and fatty acid composition changes in different strains. Three independent experiments were performed for each sample. T-test was introduced to determine the significance of changes and $p < 0.05$ was utilized. Principle component analysis (PCA) was carried out for each sample based on three biological replicates using SIMCA-P+ 12 software.

6.3 Results and discussion

6.3.1 Metabolite profiling of each yeast strain during different lipid accumulation phases

In order to identify the differences of metabolites related to lipid accumulation in different yeast strains, GC-MS-based metabolomics profiling in each yeast strain was performed at various time point with different lipid content and analyzed using NIST08 mass spectral library. The metabolite was identified when it had a similarity index above 90%.

6.3.1.1 Time-course based metabolite profiling of *S. cerevisiae*

S. cerevisiae cells were incubated in optimized conditions and samples were taken during three growth phases with different lipid content: exponential phase defined as early lipid accumulation stage (8h), early stationary phase defined as middle lipid accumulation stage (18h) and late stationary phase defined as late lipid accumulation stage (36h).

Within the 63 peaks obtained, 48 intracellular metabolites were identified using the NIST08 MS library. 15 of these metabolites expressed marked differences at various time points (Figure 6.1, page 100). Majority of them were found to be involved in amino acid metabolism, citric acid cycle and lipid metabolism. Five of them were found up-regulated during early stationary phase, whereas most of them were down-regulated during late stationary phase. Three independent replicate experiments were carried out and multivariate statistical modelling was performed using principle component analysis (PCA). The scores plot (Figure 6.2A, page 101) from PCA clearly revealed a discrimination in *S. cerevisiae* cells at early, middle and late lipid accumulation stages. The loadings plot analysis also found out that valine, asparate and glutamine contributed more to distinguish cells with different lipid content (Figure 6.2B, page 101).

Figure 6.1 Differential expression levels of intracellular metabolites at different lipid accumulation stages of *S. cerevisiae*. The ratios were normalized to the values obtained for the early stage sample.

Figure 6.2 Principal component analysis (PCA) of the GC-MS spectra from intracellular metabolites in *S. cerevisiae*. A: PCA scores plot. Samples with low (L), medium (M) and high (H) lipid content were included; t_1 and t_2 represented

scores. B: PCA loading plot; p1 and p2 represented loadings.6.3.1.2 Time-course based metabolite profiling of *C. albidus*

C. albidus cells were incubated in two-stage optimized conditions and samples were harvested at early lipid accumulation stage (12 h), middle lipid accumulation stage (24 h) and late lipid accumulation stage (96 h).

Within the 57 peaks obtained, 42 intracellular metabolites were identified using the NIST08 MS library. 11 of these metabolites expressed marked differences at various time points (Figure 6.3, page 103). Majority of them were found to be involved in amino acid metabolism, citric acid cycle and lipid metabolism. Four of them were found up-regulated during early stationary phase, whereas only two of them were up-regulated during late lipid accumulation stage. Three independent replicate experiments were carried out and multivariate statistical modelling was performed using principle component analysis (PCA). The scores plot (Figure 6.4A, page 104) from PCA clearly revealed a discrimination in *C. albidus* cells at early, middle and late lipid accumulation stages. The loadings plot analysis also found out that malate, lysine and succinate contributed more to distinguish cells with different lipid content (Figure 6.4B, page 104).

Figure 6.3 Differential expression levels of intracellular metabolites at different lipid accumulation stages of *C. albidus*. The ratios were normalized to the values obtained for the early stage sample.

Figure 6.4 Principal component analysis (PCA) of the GC-MS spectra from intracellular metabolites in *C. albidus*. A: PCA scores plot. Samples with low (L), medium (M) and high (H) lipid content were included; t_1 and t_2 represented scores. B: PCA loading plot; p_1 and p_2 represented loadings.

6.3.1.3 Time-course based metabolite profiling of *R. toruloides*

Under optimized conditions, *R. toruloides* successfully accumulated lipids and cells were collected at three time points, early lipid accumulation stage (12 h), middle lipid accumulation stage (24 h) and late lipid accumulation stage (96 h). During the metabolite profiling with GC-MS, 49 peaks were detected and 10 metabolites were identified with remarkable changes (Figure 6.5, page 106) in three independent replicate experiments. More metabolites involved in lipid metabolism and citric acid cycle were found as compared to *S. cerevisiae*. Multivariate statistical modelling was performed using PCA based on the GC-MS spectra of three replicates at each time point and the scores plot (Figure 6.6A, page 107) clearly showed a discrimination among *R. toruloides* cells at different lipid accumulation stages. The loadings plot analysis also determined that proline, succinate and citrate as major discriminators between cells with different lipid content (Figure 6.6B, page 107).

Figure 6.5 Differential expression levels of intracellular metabolites at different lipid accumulation stages of *R. toruloides*. The ratios were normalized to the values obtained for the early stage sample.

Figure 6.6 Principal component analysis (PCA) of the GC-MS spectra from intracellular metabolites in *R. toruloides*. A: PCA scores plot. Samples with low (L), medium (M) and high (H) lipid content were included; t_1 and t_2 represented scores. B: PCA loading plot; p_1 and p_2 represented loadings.

6.3.2 Comparison of metabolite changes among *S. cerevisiae*, *C. albidus* and *R. toruloides* at each stage

Differential metabolite expression patterns were assessed by analysing the changes in metabolites among *S. cerevisiae*, *C. albidus* and *R. toruloides* at different lipid accumulation stages.

Based on three independent batches of experiments for each group, a total of 67, 40 and 39 peaks were constantly identified at early (8 h for *S. cerevisiae* and 12 h for *C. albidus* and *R. toruloides*), middle (18 h for *S. cerevisiae* and 24 h for *C. albidus* and *R. toruloides*) and late (36 h for *S. cerevisiae* and 96 h for *C. albidus* and *R. toruloides*) lipid accumulation stage, respectively. Table 6.1 (page 109) showed the metabolites that were identified with important changes from three independent experiments. More metabolites were found during middle lipid accumulation stage, less were detected with changes at late stage as the cells entered death phase as metabolism of the cell was sharply decreased.

Table 6.1 Fold change of metabolites in *C. albidus* (C) and *R. toruloides* (R) comparing with *S. cerevisiae* (S) at different lipid accumulation stages.

Metabolites	Fold change*					
	Early stage**		Middle stage**		Late stage**	
	C:S ¹⁾	R:S ²⁾	C:S ²⁾	R:S ¹⁾	C:S ¹⁾	R:S ²⁾
Succinate	1.21±0.03	1.19±0.04	1.03±0.08	0.92±0.05	0.77±0.05	0.83±0.07
Isoleucine	0.87±0.05	0.74±0.07	0.76±0.07	0.58±0.05	0.81±0.05	0.68±0.04
Glycine	1.09±0.11	1.11±0.08	0.79±0.09	0.68±0.05	0.80±0.05	0.93±0.05
Proline	0.87±0.07	0.79±0.05	0.72±0.04	0.57±0.05	0.91±0.05	0.68±0.09
Ornithine	0.92±0.05	0.87±0.04	0.82±0.05	0.79±0.05	0.71±0.05	0.86±0.04
Lysine	1.18±0.09	0.92±0.03	1.21±0.03	0.77±0.05	0.89±0.05	1.09±0.11
Throsine	0.73±0.03	0.65±0.08	1.03±0.06	0.78±0.05	0.76±0.05	0.68±0.04
Inositol	1.06±0.06	0.92±0.06	1.18±0.08	1.01±0.05	0.88±0.05	0.79±0.05
Citrate	1.31±0.09	1.22±0.07	0.89±0.11	0.81±0.05	0.83±0.05	1.06±0.09
Valine	0.89±0.06	0.79±0.02	N.D.	N.D.	N.D.	N.D.
Serine	N.D.	N.D.	1.09±0.05	0.93±0.05	N.D.	N.D.
Threonine	N.D.	N.D.	0.88±0.07	0.72±0.05	N.D.	N.D.
Malate	1.21±0.05	N.D.	0.92±0.08	0.78±0.05	N.D.	N.D.
Aspartate	N.D.	N.D.	0.80±0.03	0.63±0.05	N.D.	N.D.
Glutamine	N.D.	N.D.	0.75±0.04	0.78±0.05	N.D.	N.D.

*Data represented mean±S.D. based on three independent replicate experiments. "N.D." indicated the metabolite was not detected.

**Early stage, Middle stage and Late stage indicated samples were from early, middle and late lipid accumulation stages.

C: S represented the ratio of metabolite expression level in *C. albidus* (C) to *S. cerevisiae* (S).

R: S represented the ratio of metabolite expression level in *R. toruloides* (R) to *S. cerevisiae* (S).

6.3.3 Analysis of changes of various metabolites along comparison

Numerous metabolites were identified in this study and the changes in their levels during three different lipid accumulation stages; low, medium and high were analyzed. Citric acid cycle is an essential metabolic pathway in yeast for energy (e.g. NADH and NADPH) generation. Lipid production is a process that requires a great deal of energy [122]. Succinate, malate and citrate are key components in TCA cycle. The concentrations of these three metabolites were detected to be up-regulated during early lipid accumulation stage, indicating the acceleration of TCA cycle to meet the excess demand for energy, when lipid was being accumulated. When they were compared among three strains at each time point, higher concentrations were detected in *R. toruloides*, following by *C. albidus* and indicated larger supply of energy for oleaginous yeast. However, when the cells were shifted to later lipid accumulation stage, the levels of metabolites involved in TCA cycle in *C. albidus* and *R. toruloides* were found to be decreased. It has been shown that the TCA cycle is significantly inhibited when oleaginous yeast cells are gradually depleted of a nitrogen source. The metabolic flux shift to lipid biosynthesis [112] would explain the down regulation of the TCA cycle in *C. albidus* and *R. toruloides*.

Apart from metabolites involved in the TCA cycle, amino acid metabolites were also identified. Valine and isoleucine are essential amino acids and synthesized via several steps beginning from pyruvic acid. Reduction of concentrations of valine and isoleucine were found along with lipid accumulation, which led to fewer consumption of pyruvate and more substrate for acetyl-CoA, the precursor for fatty acid. Proline, glutamine and ornithine are derived from glutamate and the concentrations of them

were found decreased during lipid accumulation, which led to less competition with the TCA cycle and relatively less energy consumption.

Serine is a critical component of phosphatidylserine synthesis and also precursor of glycine. Threonine is converted to pyruvate under the catalysis of threonine dehydrogenase. It can undergo thiolysis with CoA to produce acetyl-CoA and glycine and also essential for serine synthesis [123]. Aspartate is the precursor for many amino acids, including threonine, isoleucine, methionine and lysine. It also takes part in gluconeogenesis and carries reducing equivalents in malate-aspartate shuttle and could interconvert with oxaloacetate. In addition, it also acts as hydrogen acceptor in the ATP synthase chain. The concentrations of serine, glycine, threonine and aspartate were found to be increased during transition from low to medium lipid accumulating phase, and decreased thereafter. This would probably be due to the metabolic reprogramming in order to meet the energy balance needs according to S. Tadi and S. Kim [124].

The level of lysine was found to be slightly decreased in three yeast strains during growth. However, higher level of lysine was detected in *C. albidus* and *R. toruloides* than that of *S. cerevisiae* in the same lipid accumulation stage. It is suggested that lysine can be synthesized via two pathways in oleaginous yeast, which are α -aminoadipic acid and α , δ -diaminopimelic acid pathways. However, in non-oleaginous strains, lysine could only be synthesized via α -aminoadipate pathway. A high lipid content would be experienced if more lysine is synthesized via α , δ -diaminopimelic acid pathway as a higher level of acetyl-CoA substrate is available, and channelled for lipid biosynthesis [125]. As a result, more lysine was synthesized

through this particular pathway in *C. albidus* and *R. toruloides* and experienced a higher lipid content.

6.3.4 Cellular lipid composition analysis of yeast strains by GC-MS

Apart from the different lipid content, *S. cerevisiae*, *C. albidus* and *R. toruloides* also experienced a difference in lipid composition. Table 6.2 (page 113) showed the differences in fatty acid composition at different lipid accumulation stages for each yeast strain. The majority of fatty acids in *S. cerevisiae* were palmitic acid, palmitoleic acid, stearic acid and oleic acid, whereas for *C. albidus* and *R. toruloides*, more palmitic acid, stearic acid, oleic acid and linoleic acid were detected. The percentage of unsaturated fatty acid was higher in oleaginous yeast strain than that of *S. cerevisiae*, reaching around 60-70% of the total fatty acids. It was reported that unsaturated fatty compounds have significantly lower melting points than the saturated ones. The biofuels derived from large amounts of unsaturated fatty compounds would experience lower cloud point and pour point [126]. Cloud point (CP) refers to the temperature at which the liquid material becomes cloudy and pour point (PP) means the lowest temperature at which the materials will still flow. The highly unsaturated fatty compounds in oleaginous yeast strains made them beneficial for biodiesel application in low-temperature environment due to their relatively low CP and PP.

Table 6.2 Fatty acid composition of lipid produced by yeast strains at different lipid accumulation stages.

	Fatty acid composition (w/w)*				
	Palmitic acid	Palmitoleic acid	Stearic acid	Oleic acid	Linoleic acid
	(C16:0)	(C16:1)	(C18:0)	(C18:1)	(C18:2)
<i>S. cerevisiae</i> with low lipid accumulation ability					
Early stage	47.5±0.8	16.7±0.2	28.4±0.8	3.7±0.1	N.D.
Middle stage	44.1±0.9	17.3±0.5	29.8±1.1	6.3±0.3	N.D.
Late stage	48.6±0.6	15.9±0.4	26.3±0.9	7.1±0.2	N.D.
<i>C. albidus</i> with medium lipid accumulation ability					
Early stage	23.0±0.4	N.D.	11.0±0.5	43.2±0.6	23.8±0.5
Middle stage	17.1±0.6	0.9±0.1	9.2±0.5	51.7±0.7	21.1±0.3
Late stage	14.0±0.8	N.D.	8.7±0.4	59.2±0.5	18.1±0.6
<i>R. toruloides</i> with high lipid accumulation ability					
Early stage	25.6±0.5	N.D.	25.7±0.3	41.2±0.8	5.7±0.1
Middle stage	17.8±0.2	1.1±0.1	14.4±0.2	52.3±1.0	13.9±0.5
Late stage	17.3±0.3	0.9±0.1	10.4±0.5	62.3±1.4	6.7±0.2

*Data represented mean±S.D. based on three independent replicate experiments. "N.D." indicated the metabolite was not detected.

Early stage, Middle stage and Late stage indicated samples were from early, middle and late lipid accumulation stages.

6.3.5 Mechanism of fatty acid composition changes in different yeast strains

Lipid was accumulated along with cultivation in appropriate conditions and the fatty acid composition of accumulated intracellular lipids was analyzed by GC-MS analysis. Apart from different lipid content at different lipid accumulation stages in *S. cerevisiae*, *C. albidus* and *R. toruloides*, the composition of synthesized lipids in different yeast strains were also varied.

Majority of lipid components in oleaginous yeast strains *C. albidus* and *R. toruloides* was composed of C₁₆ and C₁₈ fatty acids, among which palmitic acid (C16:0) and oleic acid (Δ^9 C18:1) possess higher percentage, followed by linoleic acid ($\Delta^9, 12$ C18:2), stearic acid (C18:0) and palmitoleic acid (Δ^9 C16:1). Whereas in *S. cerevisiae*, most of the lipids was composed of saturated fatty acids, including palmitic acid (C16:0) and stearic acid (C18:0). Only a small quantity of unsaturated fatty acids were detected, which was around 15% of palmitoleic acid (Δ^9 C16:1) and no more than 10% oleic acid (Δ^9 C18:1) of total fatty acids. What is more, along with the lipid accumulation, the degree of unsaturation was also increased, changing more saturated fatty acids into unsaturated ones. Higher content of unsaturated fatty acids were detected in *R. toruloides* than that of *C. albidus*.

This interesting phenomenon has brought advantages for oleaginous yeast, especially when they are intended to be used for biodiesel working in low-temperature environment. This composition change could be explained by the stress experienced by oleaginous yeast, especially when they entered lipid accumulation phases. Nitrogen was depleted gradually along with lipid accumulation and thus cells were facing higher stress to struggle for growth and metabolism.

The proteomics data also indicated that a lot of stress response proteins were detected and up-regulated when middle and late lipid accumulation stages were undertaken. The unsaturation of fatty acids in the cells could increase the fluidity of cell membrane and decrease the cell sensitivity to heat stress as described in C. Mahua and K. Seunath's work [113,127,128].

There are two major independently regulated stress responses in yeast. Heat shock response (HSR) which is induced by heat shock in cells; another one is general stress response (GSR) and induced by various stressing agents, such as oxidative stress, osmotic stress, nitrogen depletion and pH abnormal change. The mechanisms inducing the activation of these two stress transduction systems are still unclear.

As our results suggested, the yeast cells could produce higher levels of unsaturated fatty acids to strengthen the adaptability in high stress environments and reduce the stress sensitivity of cells. This conclusion was also supported by other reports [129,130], illustrating that membrane fluidity was well increased when cells experienced higher degree of unsaturated fatty acid compositions.

6.4 Chapter conclusions

The oleaginous yeast *C. albidus* and *R. toruloides* was able to produce marked amounts of lipid, which could be 3-5 times higher than non-oleaginous yeast *S. cerevisiae*, using economical medium component. Metabolomics investigations of three yeast strains revealed the changes in levels of intracellular soluble metabolites and fatty acids which were related to lipid accumulation during different lipid accumulation stages. This study successfully set up a platform for time-course based

metabolomics profiling in yeast using GC-MS and enhanced our understanding of differences between oleaginous and non-oleaginous yeast strains and may shed light to future mechanism study of lipid accumulation to facilitate biofuel production.

CHAPTER 7

CONCLUSIONS AND FUTURE DIRECTIONS

7.1 Conclusions

7.1.1 Background and motivation

Biofuel production is a developing field to alternatively supplement the increasing energy demand all over the world. Oleaginous microorganisms are emerging research targets with high lipid accumulation abilities and easy cultivation methods. A better understanding of dynamic states in oleaginous microorganisms will provide valuable insights into further research in condition optimization and genetic engineering for preferable lipid accumulation. Protein and metabolite are sensitive cellular responders and their concentration shifts could indicate the dynamic response of cells to environment changes. Consequently, the development of proteomics and metabolomics platforms would greatly facilitate the understanding of lipid accumulation during different periods.

Oleaginous yeast strains account for a large proportion of oleaginous microorganisms available and many of them show high potential in lipid production. *C. albidus* and *R. toruloides* belong to oleaginous yeast category and have been utilized in industrial lipid production because of their outstanding lipid accumulation abilities, reaching up to 46% and 65.2% of their cell dry weight, respectively. Previous studies have

indicated the optimization and application of their lipids, however, the dynamic variations along lipid accumulation is not well understood.

To achieve a better understanding of lipid accumulation in different yeast strains, we developed 2-D LC-MS/MS based proteomics and GC-MS based metabolomics platforms to obtain a comprehensive view of their dynamic changes through comparison among two oleaginous yeast strains *C. albidus* and *R. toruloides* and one non-oleaginous yeast strain *S. cerevisiae*. Through analysing the altered proteins and metabolites, we wish to obtain a better understanding of the lipid production and aid their incorporation into future improvement of biofuel production. It is to be hoped that the platforms we developed in this study will be used to elucidate lipid accumulation mechanisms for oleaginous yeast strains.

7.1.2 Biosynthesis of lipids in different yeast strains

It has been reported that *S. cerevisiae*, *C. albidus* and *R. toruloides* exhibit different lipid accumulation abilities. In order to investigate the differences in lipid production, it is necessary to approach appropriate culture conditions to ensure their fully expression of lipid accumulation ability.

In this chapter, optimization of substrate concentrations and inoculum cell densities were developed for best accumulation of lipids. Cell dry biomass, lipid content and residual glucose concentrations were used to record and determine the collection of lipids. Our results indicated that *S. cerevisiae*, *C. albidus* and *R. toruloides* were of different lipid-cumulating abilities, storing 7.98%, 27.74% and 45.03% (w/w) lipid of their cell dry weight and defined as low, medium and high lipid content strains in

our study. According to the different periods of lipid production, the time points with different lipid concentrations were defined as early, middle and late lipid accumulation stages.

7.1.3 Proteomic analysis of yeast strains at various lipid accumulation stages

In this chapter, the sophisticated iTRAQ-coupled 2-D LC-MS/MS platform was employed to identify and quantify the intracellular protein profile in *S. cerevisiae*, *C. albidus* and *R. toruloides*. Comparisons of three strains at each lipid accumulation stage were performed. A total of 132, 122 and 116 proteins were constantly identified at early, middle and late lipid accumulation stages, respectively. Our results indicated that a panel of 71 out of the 132 proteins were differentially expressed in three strains at early lipid accumulation stage. And majority of these proteins had functions involving in metabolism and ribosomal biosynthesis. At the middle stage, 62 among the 122 proteins passed through the screening; meanwhile, 55 of 116 proteins at the late stage of lipid accumulation went through the filtration. More proteins involved in metabolism, biosynthesis and stress response were identified in these two stages. A schematic diagram depicting the differentially expressed proteins at different stages was shown in Figure 7.1 (page 120). The protein changes in different samples were validated by western blot and real-time RT-PCR and the results were consistent with protein level changes observed from LC-MS/MS analysis. Taken together, our study highlights the potential of LC-MS/MS-based proteomic profiling of yeast strains at different lipid accumulation stages. Our results indicated that three yeast cells at each stage exhibited characteristic proteomic signatures and enhanced the understanding of dynamic changes during lipid accumulation.

Figure 7.1 Differentially expressed proteins in three yeast strains at different lipid accumulation stages. *S. cerevisiae* was used as control sample, comparison of protein level changes was carried out in *C. albidus* and *R. toruloides* against *S. cerevisiae*. The red arrow represented up-regulation of protein levels; the blue arrow indicated the down-regulation of protein levels.

7.1.4 Metabolic profiling of yeast strains at different lipid accumulation stages

In this chapter, we employed GC-MS system to investigate the alterations of metabolic profile in *S. cerevisiae*, *C. albidus* and *R. toruloides* at different lipid accumulation stages. The soluble intracellular metabolites and fatty acids were identified and quantified separately. Three independent biological experiments were conducted and a total of 67, 40 and 39 soluble metabolites were constantly identified in all three strains at early, middle and late lipid accumulation stages. Fatty acid quantification and composition analysis was also performed and great variations were

detected when comparing between oleaginous yeast strains with non-oleaginous yeast strain.

Our results revealed a panel of intracellular metabolites displaying statistically significant difference ($p < 0.05$) in their expression levels in oleaginous yeast strains relative to the control cells, which was *S. cerevisiae*. 10, 14 and 9 metabolites were differentially expressed in all three strains at early, middle and late stages. Key components in TCA cycle such as succinate, malate and citrate were detected to be up-regulated during early lipid accumulation stage and decreased thereafter, indicating the acceleration of TCA cycle to meet the excess demand for energy during initial lipid accumulation process and shift of metabolic flux to lipid biosynthesis in later stages. Amino acid metabolites were also identified. The concentrations of serine, glycine, threonine, lysine and aspartate were found to be increased during transition from low to medium lipid accumulating phase, and decreased thereafter due to the metabolic reprogramming to meet the energy balance needs. Valine, isoleucine, proline, glutamine and ornithine were found to decrease all the way along lipid accumulation, leading to less competition with fatty acid precursor or TCA cycle and less consumption of energy.

Fatty acid composition also experienced big differences along comparison, with more unsaturated fatty acid produced in oleaginous yeast strains, while majority of fatty acids in *S. cerevisiae* was saturated ones. This was due to the high stress environments faced by the oleaginous yeast cells, and in order to adapt to the environment, increase the membrane fluidity and reduce the sensitivity to stress, the oleaginous yeast strains produced high levels of unsaturated fatty acids. It led to proper adaption and more

profitable products as unsaturated fatty acids possesses low cloud point and pour point, which is very beneficial to be derived into biofuels and utilized in low temperature environments. The specific metabolic alterations were illustrated in Figure 7.2 (page 123).

Taken together, our work illustrates the feasibility of metabolic profiling to understand differences along lipid accumulation in different yeast cells based on GC-MS platform. Our results indicated that each yeast strain exhibited characteristic metabolic signatures and consistent with their differential biological activities and previous proteomics results. Our study enhances the understanding of differences between oleaginous and non-oleaginous yeast strains and opens a new perspective to apply GC-MS based profiling platform to shed light to future mechanism study of lipid accumulation to facilitate biofuel production.

Figure 7.2 Schematic diagram summarizing the metabolic alterations in three yeast strains at different lipid accumulation stages. For soluble intracellular metabolites, *S. cerevisiae* was used as control sample, comparison of metabolite level changes was carried out in *C. albidus* and *R. toruloides* against *S. cerevisiae*. The red arrow represented up-regulation; the blue arrow indicated down-regulation. For fatty acid analysis, composition differences were presented respectively, with majority labelled by red color and minority labelled by blue color.

7.1.5 Overall summary of dynamic changes in different yeast strains

Based on the proteomics and metabolomics analysis of oleaginous and non-oleaginous yeast strains during various lipid accumulation periods, dynamic changes of pathways and mechanisms were recognized. According to the level changes of proteins and metabolites, the following pathways could be summarised with changes, relating to the lipid synthesis.

TCA cycle and glycolysis were down-regulated in oleaginous yeast strains, especially during the middle and late lipid accumulation stages indicating a shift of metabolic flux to lipid biosynthesis. Energy and fatty acid precursor supply were up-regulated in oleaginous strains, providing steady supplement of energy for lipid synthesis. Moreover, higher degree of up-regulation of these pathways were detected in *R. toruloides* comparing with *C. albidus*, which helped explain the higher lipid accumulation ability of *R. toruloides*.

When yeast cells were transferred into nitrogen limited medium, higher stress was experienced along with lipid accumulation. The oleaginous yeast cells could regulate their sensitivities to these stresses and lower sensitivity was expressed in high lipid content strain. Whereas *S. cerevisiae* lacks this ability and could not grow or produce lipids normally when treated the same way as oleaginous strains.

What is more, the lipid composition analysis revealed that more unsaturated fatty acids were accumulated in oleaginous yeast strains, which would be beneficial in further biofuel exploration for usage in low temperature environment.

Figure 7.3 summarized the dynamic changes of pathways within three strains along lipid accumulation (page 125).

Figure 7.3 Schematic diagram summarizing the dynamic changes in yeast strains with different lipid accumulation abilities. Red arrow indicated up-regulation, blue arrow represented down-regulation. Higher changes were revealed by two arrows.

7.1.6 Limitations of the results

Our preliminary results have demonstrated the feasibility of using LC-MS/MS-based proteomics and GC-MS-based metabolomics platforms to understand lipid accumulation in different yeast strains. They could provide insights on understanding dynamic changes in oleaginous and non-oleaginous yeast cells along lipid production

using proteomics and metabolomics approach. However, certain limitations still exist in this study and need to be taken into account.

Although iTRAQ-coupled LC-MS/MS analysis was successfully carried out and numerous proteins were identified along comparison of oleaginous and non-oleaginous yeast at different lipid accumulation stages, few proteins were found in relate to lipid metabolism directly. This may be caused by the fact that many proteins were hydrophobic and the abundance was low, leading to a weak solubility and poor separation behaviour. Moreover, few literatures have been published on proteomics study of oleaginous yeast strains and genomic information is not integrate. The proteome database for oleaginous yeast strains in our study is not well established yet, so the comparison and discussion of proteomic data of these three yeast strains were based on database searched against ‘microorganism’, rather than specific model microbes, remarkable challenges existed for accurate data analysis.

GC-MS-based metabolic profiling platform offered a relatively broad coverage of metabolites, however, there is no single analytical approach that can give a complete “picture” of metabolome. In our analysis of GC-MS-based metabolome profile, limited metabolites were observed and most of them were amino acids. This may possibly due to the instability of metabolites and their rapid dynamic changes. What is more, some metabolites may not be able to be derivatized and analyzed by GC-MS. Alternative analytical technologies such as LC-MS or NMR may be utilized as complementary methods to provide a more comprehensive identification of metabolites.

7.2 Future directions

Our research has successfully set up proteomics and metabolomics based platforms to understand dynamic changes in lipid accumulation of oleaginous yeast strains. The differentially expressed proteins and metabolites observed in different yeast cells enhanced our understanding of changing of processes in oleaginous yeast, contributing to a better understanding of excessive lipid storage in oleaginous yeast strains. It is hoped that these platforms can shed light to future mechanism study and facilitate biofuel production.

In this study, we developed proteomics platform to understand protein level changes in oleaginous yeast. We identified several proteins and pathways that are essential in lipid accumulation, further genetic engineering will be conducted to overexpress or delete the genes in the related pathways in the engineering yeast species to obtain more lipids and follow-up products. The lack of genetic information and experiences may be a challenge in the genetic engineering study, but a recently published research about *R. toruloides* [131] would shed light to the genetic engineering of oleaginous yeast strains.

Metabolic profiling was achieved by GC-MS comparing each yeast strain individually along lipid accumulation and also comparatively studied between oleaginous yeast and non-oleaginous yeast. Although GC-MS is a relatively mature analytical tool for metabolomics analysis, it could not cover the entire metabolic picture alone. Combination of multiparallel technologies are still necessary to provide more comprehensive coverage of metabolome. Different analytical tools should be

employed, such as LC-MS and NMR. Suitable corresponding analysis protocols should be developed as well.

The accumulated lipids in our study could be utilized as precursors for biofuel production. But most of the fatty acids constituting the lipids were long chain fatty acids. Medium chain fatty acids are intermediary metabolites in yeast and can be converted into medium chain alkanes, which have the potential to be used as transportation fuels, especially in aircraft area. Our work has aimed to find out the reason for the particular and extraordinary high lipid accumulation ability in oleaginous yeast strains and the essential proteins involved and affected in the lipid accumulation and metabolism pathways. The lipids in oleaginous yeasts mostly exist in form of triacylglycerol (TAG) and a little free fatty acid. Under the function of yeast lipase, TAG will transform into fatty acid, which is long chain in yeast strains. Combination with other lab mates' work would help 'cut off' the long chain fatty acid in yeast and obtain medium chain fatty acid for further derivatization to achieve medium chain alkane.

LIST OF PUBLICATIONS

1. **Shi JH**, Feng HX, Lee JL and Chen W. (2013) Comparative proteomics profile of lipid-cumulating oleaginous yeast: An iTRAQ-coupled 2-D LC-MS/MS analysis. *PLoS ONE*, 8(12): e85532
2. Zhang JH, **Shi JH**, Lee BJH, Chen LW, Tan KY, Tang XL, Tan JYL, Li X, Feng HX, and Chen W. (2013) Proteomic analysis of vascular smooth muscle cells with S- and R- Enantiomers of atenolol by iTRAQ and LC-MS/MS. *Method Mol Biol.* Humana Press, NJ, USA. 1000:45-52
3. Lee J, Chen LW, **Shi JH**, Trzcinski A, and Chen W. (2014) Metabolomic profiling of *Rhodospiridium toruloides* grown on glycerol for carotenoids production during different growth phases. *J Agr Food Chem*, 62(41):10203-9
4. **Shi JH**, Chen LW, Lee JL and Chen W*. Comparative metabolic profiling of yeast strains during different growth phases in lipid production. Revised.

Conference Presentations

1. Jiahua Shi, Huixing Feng, William Chen Wei Ning. Comparative proteomics profile of different yeast strains for lipid accumulation mechanism exploration using iTRAQ-coupled 2-D LC-MS/MS method. **3rd International Congress on Analytical Proteomics**, 28th- 31st July, 2013, Sao Pedro, Brazil.

REFERENCES

1. Brennan L, Owende P (2010) Biofuels from microalgae—A review of technologies for production, processing, and extractions of biofuels and co-products. *Renew Sust Energ Rev* 14: 557-577.
2. Li Q, Du W, Liu D (2008) Perspectives of microbial oils for biodiesel production. *Appl Microbiol Biot* 80: 749-756.
3. Naik SN, Goud VV, Rout PK, Dalai AK (2010) Production of first and second generation biofuels: A comprehensive review. *Renew Sust Energ Rev* 14: 578-597.
4. Meher LC, Vidya Sagar D, Naik SN (2006) Technical aspects of biodiesel production by transesterification—a review. *Renew Sust Energ Rev* 10: 248-268.
5. Tolmac D, Prulovic S, Lambic M, Radovanovic L, Tolmac J (2014) Global trends on production and utilization of biodiesel. *Energ Source Part B* 9: 130-139.
6. Ren21 (2014) Renewables 2014 global status report. Paris: REN21 Secretariat. p. 43.
7. Schnepf R, Yacobucci BD (2013) Renewable fuel standard (RFS): overview and issues. USA: Congressional Research Service. p. 1.
8. Subramaniam R, Dufreche S, Zappi M, Bajpai R (2010) Microbial lipids from renewable resources: production and characterization. *J Ind Microbiol Biot* 37: 1271-1287.
9. Papanikolaou S, Aggelis G (2011) Lipids of oleaginous yeasts. Part I: Biochemistry of single cell oil production. *Eur J Lipid Sci Tech* 113: 1031-1051.
10. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, *et al.* (2008) Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *Plant J* 54: 621-639.
11. Azócar L, Ciudad G, Heipieper HJ, Navia R (2010) Biotechnological processes for biodiesel production using alternative oils. *Appl Microbiol Biotechnol* 88: 621-636.
12. Solovchenko A, Khozin-Goldberg I, Didi-Cohen S, Cohen Z, Merzlyak M (2008) Effects of light and nitrogen starvation on the content and composition of carotenoids of the green microalga *Parietochloris incisa*. *Russ J Plant Physiol* 55: 455-462.
13. Leman J (2009) Lipids, Production. In: Schaechter M, editor. *Encyclopedia of Microbiology* (Third Edition). Oxford: Academic Press. pp. 393-406.

14. Papanikolaou S, Komaitis M, Aggelis G (2004) Single cell oil (SCO) production by *Mortierella isabellina* grown on high-sugar content media. *Bioresource Technol* 95: 287-291.
15. Hassan M, Blanc PJ, Granger L-M, Pareilleux A, Goma G (1996) Influence of nitrogen and iron limitations on lipid production by *Cryptococcus curvatus* grown in batch and fed-batch culture. *Process Biochem* 31: 355-361.
16. Huang G, Chen F, Wei D, Zhang X, Chen G (2010) Biodiesel production by microalgal biotechnology. *Appl Energ* 87: 38-46.
17. Liu SJ, Yang WB, Shi AH (2000) Screening of the high lipid production strains and studies on its flask culture conditions. *Microbiology* 27: 93-97.
18. Li YH, Liu B, Zhao ZB, Bai FW (2006) Optimized culture medium and fermentation conditions for lipid production by *Rhodospiridium toruloides*. *China J Biotechnol* 22: 650-656.
19. Davies RJ, Holdsworth JE (1992) Synthesis of lipids in yeasts: Biochemistry, physiology and production. *Adv Appl Lipid Res* 1: 119-159.
20. Ratledge C, editor (1994) Yeasts, moulds, algae and bacteria as sources of lipids. London: Blackie Academic and Professional. pp. 235-291.
21. Ratledge C (1988) Biochemistry, stoichiometry, substrates and economics, in single cell oil. Longman Scientific and Technical. pp. 33-70.
22. Fakas S, Papanikolaou S, Galiotou-Panayotou M, Komaitis M, Aggelis G, editors (2009) Biochemistry and biotechnology of single cell oil. New Delhi: AsiaTech Publishers Inc. pp. 38-60.
23. Ratledge C, Wynn JP (2002) The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms. *Adv Appl Microbiol: Academic Press*. pp. 1-51.
24. Ratledge C, editor (1997) Microbial lipids. 2 ed. Weinheim: Wiley-VCH. pp. 135-197.
25. Ratledge C (2004) Fatty acid biosynthesis in microorganisms being used for Single Cell Oil production. *Biochimie* 86: 807-815.
26. Papanikolaou S, Aggelis G (2009) Biotechnological valorization of biodiesel derived glycerol waste through production of single cell oil and citric acid by *Yarrowia lipolytica*. *Lipid Technol* 21: 83-87.
27. Ratledge C, Cohen Z (2008) Microbial and algal oils: Do they have a future for biodiesel or as commodity oils? *Lipid Technol* 20: 155-160.

28. Papanikolaou S, Aggelis G (2010) *Yarrowia lipolytica*: A model microorganism used for the production of tailor-made lipids. *Eur J Lipid Sci Tech* 112: 639-654.
29. Evans CT, Ratledge C (1985) The role of the mitochondrial NAD⁺: isocitrate dehydrogenase in lipid accumulation by the oleaginous yeast *Rhodospiridium toruloides* CBS 14. *Can J Microbiol* 31: 845-850.
30. Boulton CA, Ratledge C (1980) Regulatory studies on citrate synthase in *Candida 107*, an oleaginous yeast. *J Gen Microbiol* 121: 441-447.
31. Ratledge C, Bowater M, Taylor P (1997) Correlation of ATP/citrate lyase activity with lipid accumulation in developing seeds of *Brassica napus* L. *Lipids* 32: 7-12.
32. Papanikolaou S, Sarantou S, Komaitis M, Aggelis G (2004) Repression of reserve lipid turnover in *Cunninghamella echinulata* and *Mortierella isabellina* cultivated in multiple-limited media. *J Appl Microbiol* 97: 867-875.
33. Evans CT, Scragg AH, Ratledge C (1983) A comparative study of citrate efflux from mitochondria of oleaginous and non-oleaginous yeasts. *Eur J Biochem* 130: 195-204.
34. Boulton CA RC (1983) Use of transition studies in continuous cultures of *Lipomyces starkeyi*, an oleaginous yeast, to investigate the physiology of lipid accumulation. *J Microbiol* 1299: 2871-2876.
35. McCullough W, Roberts CF (1974) The role of malic enzyme in *Aspergillus nidulans*. *FEBS Lett* 41: 238-242.
36. Wynn JP, Ratledge C (1997) Malic enzyme is a major source of NADPH for lipid accumulation by *Aspergillus nidulans*. *Microbiology* 143: 253-257.
37. Blackstock WP, Weir MP (1999) Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol* 17: 121-127.
38. Anderson NL, Anderson NG (1998) Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis* 19: 1853-1861.
39. Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, *et al.* (1996) From proteins to proteomes: Large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Nat Biotech* 14: 61-65.
40. Leinweber FC, Schmid DG, Lubda D, Wiesmuller KH, Jung G, *et al.* (2003) Silica-based monoliths for rapid peptide screening by capillary liquid chromatography hyphenated with electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun Mass Spectrom* 17: 1180-1188.

41. Wang H, Hanash S (2003) Multi-dimensional liquid phase based separations in proteomics. *J Chromatogr B* 787: 11-18.
42. Domon B, Aebersold R (2006) Mass spectrometry and protein analysis. *Science* 312: 212-217.
43. Koomen J, Hawke D, Kobayashi R (2005) Developing an understanding of proteomics: an introduction to biological mass spectrometry. *Cancer Invest* 23: 47-59.
44. Jackson PJ, Brownsill RD, Taylor AR, Walther B (1995) Use of electrospray ionization and neutral loss liquid chromatography/tandem mass spectrometry in drug metabolism studies. *J Mass Spectrom* 30: 446-451.
45. Mann M, Pandey A (2001) Use of mass spectrometry-derived data to annotate nucleotide and protein sequence databases. *Trends Biochem Sci* 26: 54-61.
46. Lund TC, Anderson LB, McCullar V, Higgins L, Yun GH, *et al.* (2007) iTRAQ is a useful method to screen for membrane-bound proteins differentially expressed in human natural killer cell types. *J Proteome Res* 6: 644-653.
47. Zieske LR (2006) A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *J Exp Bot* 57: 1501-1508.
48. Daviss B (2005) Growing pains for metabolomics - The newest 'omic science is producing results - and more data than researchers know what to do with. *The Scientist*. pp. 25-28.
49. Jordan KW, Nordenstam J, Lauwers GY, Rothenberger DA, Alavi K, *et al.* (2009) Metabolomic characterization of human rectal adenocarcinoma with intact tissue magnetic resonance spectroscopy. *Dis Colon Rectum* 52: 520-525.
50. van der Greef J, Smilde AK (2005) Symbiosis of chemometrics and metabolomics: past, present, and future. *J Chemometr* 19: 376-386.
51. Novotny MV, Soini HA, Mechref Y (2008) Biochemical individuality reflected in chromatographic, electrophoretic and mass-spectrometric profiles. *J Chromatogr B* 866: 26-47.
52. Samuelsson LM, Larsson DGJ (2008) Contributions from metabolomics to fish research. *Mol Bio Syst* 4: 974-979.
53. Nicholson JK, J.C. L, E. H (1999) 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 29: 1181-1189.
54. Nicholson JK, Connelly J, Lindon JC, Holmes E (2002) Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov* 1: 153-161.

55. Bino RJ, Hall RD, Fiehn O, Kopka J, Saito K, *et al.* (2004) Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* 9: 418-425.
56. Dias AP, Brown J, Bonello P, Grotewold E (2003) Metabolite profiling as a functional genomics tool. *Methods Mol Biol* 236: 415-426.
57. Schauer N, Steinhauser D, Strelkov S, Schomburg D, Allison G, *et al.* (2005) GC–MS libraries for the rapid identification of metabolites in complex biological samples. *FEBS Lett* 579: 1332-1337.
58. Gika HG, Theodoridis GA, Wingate JE, Wilson ID (2007) Within-day reproducibility of an HPLC–MS-based method for metabonomic analysis: application to human urine. *J Proteome Res* 6: 3291-3303.
59. Soga T, Ohashi Y, Ueno Y, Naraoka H, Tomita M, *et al.* (2003) Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J Proteome Res* 2: 488-494.
60. Griffin JL, Bollard ME (2004) Metabonomics: its potential as a tool in toxicology for safety assessment and data integration. *Curr Drug Metab* 5: 389-398.
61. Griffin JL (2003) Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Curr Opin Chem Biol* 7: 648-654.
62. Viant MR, Rosenblum ES, Tjeerdema RS (2003) NMR-based metabolomics: A powerful approach for characterizing the effects of environmental stressors on organism health. *Environ Sci Technol* 37: 4982-4989.
63. Jonsson P, Gullberg J, Nordström A, Kusano M, Kowalczyk M, *et al.* (2004) A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal Chem* 76: 1738-1745.
64. Jonsson P, Johansson AI, Gullberg J, Trygg J, A J, *et al.* (2005) High-throughput data analysis for detecting and identifying differences between samples in GC/MS-based metabolomic analyses. *Anal Chem* 77: 5635-5642.
65. Grün CH, van Dorsten FA, Jacobs DM, Le Belleguic M, van Velzen EJJ, *et al.* (2008) GC–MS methods for metabolic profiling of microbial fermentation products of dietary polyphenols in human and in vitro intervention studies. *J Chromatogr B* 871: 212-219.
66. Villas-Bôas SG, Roessner U, Hansen MAE, Smedsgaard J, Nielsen J (2006) Microbial metabolomics: rapid sampling techniques to investigate intracellular metabolite dynamics—An overview. *Metabolome Analysis: An Introduction*. Hoboken, NJ, USA: John Wiley & Sons, Inc. pp. 203-214.
67. Halket JM, Zaikin VG (2003) Derivatization in mass spectrometry--1. Silylation. *Eur J Mass Spectrom* 9: 1-21.

68. Fordham PJ, Chamot-Rooke J, Giudice E, Tortajada J, Morizur JP (1999) Analysis of alkenes by copper ion chemical ionization gas chromatography/mass spectrometry and gas chromatography/tandem mass spectrometry. *J Mass Spectrom* 34: 1007-1017.
69. Wold S, Esbensen K, Geladi P (1987) Principal component analysis. *Chemometr Intell Lab* 2: 37-52.
70. Trygg J, Wold S (2002) Orthogonal projections to latent structures (O-PLS). *J Chemometr* 16: 119-128.
71. Trygg J, Holmes E, Lundstedt T (2007) Chemometrics in metabolomics. *J Proteome Res* 6: 469-479.
72. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422: 198-207.
73. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, *et al.* (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415: 180-183.
74. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, *et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotech* 17: 994-999.
75. Raamsdonk L, Teusink B, Broadhurst D, Zhang N, Hayes A, *et al.* (2001) A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotech* 19: 45-50.
76. Villas-Bôas SG, Højer-Pedersen J, Åkesson M, Smedsgaard J, Nielsen J (2005) Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* 22: 1155-1169.
77. Liu H, Zhao X, Wang F, Jiang X, Zhang S, *et al.* (2011) The proteome analysis of oleaginous yeast *Lipomyces starkeyi*. *FEMS Yeast Research* 11: 42-51.
78. Zhao C, Nambou K, Wei L, Chen J, Imanaka T, *et al.* (2014) Evaluation of metabolome sample preparation methods regarding leakage reduction for the oleaginous yeast *Yarrowia lipolytica*. *Biochem Eng J* 82: 63-70.
79. Branyikova I, Marsalkova B, Doucha J, Branyik T, Bisova K, *et al.* (2011) Microalgae--novel highly efficient starch producers. *Biotechnol Bioeng* 108: 766-776.
80. Lamacka M, Sajbidor J, Bohov P (1998) Lipid isolation and fatty acid analysis in *Saccharomyces cerevisiae*. Comparison of different methods. *Biotechnol Tech* 12: 621-625.
81. Ratledge C, Saxton RK (1968) Quantitative extraction of lipid and fatty acids from a *Candida* species. *Anal Biochem* 26: 288-294.

82. Schneiter R, Daum G (2006) Extraction of yeast lipids. *Methods Mol Biol* 313: 41-45.
83. Wang W, Vignani R, Scali M, Sensi E, Tiberi P, *et al.* (2004) Removal of lipid contaminants by organic solvents from oilseed protein extract prior to electrophoresis. *Anal Biochem* 329: 139-141.
84. Zhou Y, Chen WN (2011) iTRAQ-coupled 2-D LC-MS/MS analysis of cytoplasmic protein profile in *Escherichia coli* incubated with apidaecin IB. *J Proteomics* 75: 511-516.
85. Manns JM (2011) SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) of Proteins. *Curr Protoc Microbiol*: A.3M.1-A.3M.13.
86. Zhang J, Sui J, Ching CB, Chen WN (2008) Protein profile in neuroblastoma cells incubated with S- and R-enantiomers of ibuprofen by iTRAQ-coupled 2-D LC-MS/MS analysis: possible action of induced proteins on Alzheimer's disease. *Proteomics* 8: 1595-1607.
87. Wang MX, Bai J, Chen WN, Ching CB (2010) Metabolomic profiling of cellular responses to carvedilol enantiomers in vascular smooth muscle cells. *PloS ONE* 5.
88. Fernandez C, Fransson U, Hallgard E, Spegel P, Holm C, *et al.* (2008) Metabolomic and proteomic analysis of a clonal insulin-producing beta-cell line (INS-1 832/13). *J Proteome Res* 7: 400-411.
89. Morrison W, Smith L (1964) Preparation of fatty acid methyl esters and dimethylacetals from lipids with boron fluoride-methanol. *J Lipid Res* 5: 600-608.
90. Chrysanthopoulos PK, Goudar CT, Klapa MI (2010) Metabolomics for high-resolution monitoring of the cellular physiological state in cell culture engineering. *Metab Eng* 12: 212-222.
91. Devantier R, Scheithauer B, Villas-Boas SG, Pedersen S, Olsson L (2005) Metabolite profiling for analysis of yeast stress response during very high gravity ethanol fermentations. *Biotechnol Bioeng* 90: 703-714.
92. Shi JH, Feng HX, Lee J, Chen WN (2013) Comparative proteomics profile of lipid-cumulating oleaginous yeast: An iTRAQ-coupled 2-D LC-MS/MS analysis. *PLoS ONE* 8: e85532.
93. Meng X, Yang J, Xu X, Zhang L, Nie Q, *et al.* (2009) Biodiesel production from oleaginous microorganisms. *Renew Energ* 34: 1-5.
94. Hamid A, Mokhtar N, Taha E, Omar O, Yusoff W (2011) The role of ATP citrate lyase, malic enzyme and fatty acid synthase in the regulation of lipid accumulation in *Cunninghamella sp. 2A1*. *Ann Microbiol* 61: 463-468.

95. Li YH, Zhao ZB, Bai FW (2007) High-density cultivation of oleaginous yeast *Rhodospiridium toruloides* Y4 in fed-batch culture. *Enzyme Microb Technol* 41: 312-317.
96. Hansson L, Dostalek M (1986) Influence of cultivation conditions on lipid production by *Cryptococcus albidus*. *Appl Microbiol Biot* 24: 12-18.
97. Dyer JM, Chapital DC, Kuan JW, Mullen RT, Pepperman AB (2002) Metabolic engineering of *Saccharomyces cerevisiae* for production of novel lipid compounds. *Appl Microbiol Biot* 59: 224-230.
98. Ordaz L, López R, Melchy O, Torre M (2001) Effect of high-cell-density fermentation of *Candida utilis* on kinetic parameters and the shift to respiration-fermentative metabolism. *Appl Microbiol Biot* 57: 374-378.
99. Papanikolaou S (2012) Oleaginous yeasts: Biochemical events related with lipid synthesis and potential biotechnological applications. *Fermentat Technol* 1.
100. Wiese S, Reidegeld KA, Meyer HE, Warscheid B (2007) Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research. *Proteomics* 7: 340-350.
101. Abdallah C, Sergeant K, Guillier C, Dumas-Gaudot E, Leclercq CC, *et al.* (2012) Optimization of iTRAQ labelling coupled to OFFGEL fractionation as a proteomic workflow to the analysis of microsomal proteins of *Medicago truncatula* roots. *Proteome Sci* 10: 37.
102. Xu J, Khor KA, Sui J, Zhang J, Tan TL, *et al.* (2008) Comparative proteomics profile of osteoblasts cultured on dissimilar hydroxyapatite biomaterials: an iTRAQ-coupled 2-D LC-MS/MS analysis. *Proteomics* 8: 4249-4258.
103. Shaw MM, Riederer BM (2003) Sample preparation for two-dimensional gel electrophoresis. *Proteomics* 3: 1408-1417.
104. Liu H, Zhao X, Wang F, Li Y, Jiang X, *et al.* (2009) Comparative proteomic analysis of *Rhodospiridium toruloides* during lipid accumulation. *Yeast* 26: 553-566.
105. Benachenhou-Lahfa N, Forterre P, Labedan B (1993) Evolution of glutamate dehydrogenase genes: evidence for two paralogous protein families and unusual branching patterns of the archaebacteria in the universal tree of life. *J Mol Evol* 36: 335-346.
106. Pagliarulo C, Salvatore P, De Vitis LR, Colicchio R, Monaco C, *et al.* (2004) Regulation and differential expression of *gdhA* encoding NADP-specific glutamate dehydrogenase in *Neisseria meningitidis* clinical isolates. *Mol Microbiol* 51: 1757-1772.

107. DeLuna A, Avendano A, Riego L, Gonzalez A (2001) NADP-glutamate dehydrogenase isoenzymes of *Saccharomyces cerevisiae*. Purification, kinetic properties, and physiological roles. *J Biol Chem* 276: 43775-43783.
108. Tehlivets O, Scheuringer K, Kohlwein SD (2007) Fatty acid synthesis and elongation in yeast. *Biochem Biophys Acta* 1771: 255-270.
109. Endres M, Neupert W, Brunner M (1999) Transport of the ADP/ATP carrier of mitochondria from the TOM complex to the TIM22.54 complex. *EMBO J* 18: 3214-3221.
110. Rondelez Y, Tresset G, Nakashima T, Kato Y, Fujita H, *et al.* (2005) Highly coupled ATP synthesis by F1-ATPase single molecules. *Nature* 433: 773-777.
111. Pagadala V, Vistain L, Symersky J, Mueller D (2011) Characterization of the mitochondrial ATP synthase from yeast *Saccharomyces cerevisiae*. *J Bioenerg Biomembr* 43: 333-347.
112. Ratledge C, Wynn JP (2002) The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms. *Adv Appl Microbiol* 51: 1-51.
113. Chatterjee MT, Khalawan SA, Curran BPG (2000) Cellular lipid composition influences stress activation of the yeast general stress response element (STRE). *Microbiology* 146: 877-884.
114. Kamisaka Y, Kimura K, Uemura H, Yamaoka M (2013) Overexpression of the active diacylglycerol acyltransferase variant transforms *Saccharomyces cerevisiae* into an oleaginous yeast. *Appl Microbiol and Biot* 97: 7345-7355.
115. Kamisaka Y, Noda N, Tomita N, Kimura K, Kodaki T, *et al.* (2006) Identification of Genes Affecting Lipid Content Using Transposon Mutagenesis in *Saccharomyces cerevisiae*. *Biosci, Biotechn Bioch* 70: 646-653.
116. Runguphan W, Keasling JD (2014) Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metab Eng* 21: 103-113.
117. Yao S, Huang J, Tan J (2012) The surface display of phytase on yeast cell and activity assay of the displayed protein. *J Yeast and Fungal Res* 3: 58-64.
118. Dolence JM, Poulter CD (1995) A mechanism for posttranslational modifications of proteins by yeast protein farnesyltransferase. *P Natl Acad Sci USA* 92: 5008-5011.
119. Mukherjee S, Hao Y-H, Orth K (2007) A newly discovered post-translational modification – the acetylation of serine and threonine residues. *Trends Biochem Sci* 32: 210-216.

120. Bellou S, Aggelis G (2012) Biochemical activities in *Chlorella sp.* and *Nannochloropsis salina* during lipid and sugar synthesis in a lab-scale open pond simulating reactor. *J Biotechnol* 164: 318-329.
121. Galiotou PM, Kalantzi O, Aggelis G (1998) Modelling of simultaneous production of polygalacturonase and exopolysaccharide by *Aureobasidium pullulans* ATHUM 2915. *Antonie van Leeuwenhoek* 73: 155-162.
122. Morgunov IG, Solodovnikova NY, Sharyshev AA, Kamzolova SV, Finogenova TV (2004) Regulation of NAD(+)-dependent isocitrate dehydrogenase in the citrate producing yeast *Yarrowia lipolytica*. *Biochem Biokhimiia* 69: 1391-1398.
123. Ravnikar PF, Somerville RL (1987) Genetic characterization of a highly efficient alternate pathway of serine biosynthesis in *Escherichia coli*. *J Bacteriol* 169: 2611-2617.
124. Tadi S, Kim SJ, Ryu MJ, Park T, Jeong JS, *et al.* (2013) Metabolic rebalancing of CR6 interaction factor 1-deficient mouse embryonic fibroblasts: A mass spectrometry-based metabolic analysis. *Bull Korean Chem Soc* 34: 35-41.
125. Vorapreeda T, Thammarongtham C, Cheevadhanarak S, Laoteng K (2012) Alternative routes of acetyl-CoA synthesis identified by comparative genomic analysis: involvement in the lipid production of oleaginous yeast and fungi. *Microbiology* 158.
126. Knothe G (2005) Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. *Fuel Process Technol* 86: 1059-1070.
127. Hill S, Hirano K, Shmanai VV, Marbois BN, Vidovic D, *et al.* (2011) Isotope-reinforced polyunsaturated fatty acids protect yeast cells from oxidative stress. *Free Radic Biol Med* 50: 130-138.
128. Landolfo S, Zara G, Zara S, Budroni M, Ciani M, *et al.* (2010) Oleic acid and ergosterol supplementation mitigates oxidative stress in wine strains of *Saccharomyces cerevisiae*. *Int J Food Microbiol* 141: 229-235.
129. Thomas DS, Hossack JA, Rose AH (1978) Plasma-membrane lipid composition and ethanol tolerance in *Saccharomyces cerevisiae*. *Arch Microbiol* 117: 239-245.
130. Yazawa H, Iwahashi H, Kamisaka Y, Uemura H (2009) Production of polyunsaturated fatty acids in yeast *Saccharomyces cerevisiae* and its relation to alkaline pH tolerance. *Yeast* 26: 167-184.
131. Zhu ZW, Zhang SF, Liu HW, Shen HW, Lin XP, *et al.* (2013) A multi-omic map of the lipid-producing yeast *Rhodospiridium toruloides*. *Nature Commun* 3.

APPENDIX

Appendix 1. List of differentially expressed proteins in target yeast strains.

Gene name	Protein name	Regulation	%AA coverage	MS/MS score	Fold change*						
					Early**			Middle**			Late**
					115:114 ⁽¹⁾	116:114 ⁽²⁾	115:114 ⁽¹⁾	116:114 ⁽²⁾	115:114 ⁽¹⁾	116:114 ⁽²⁾	
Metabolism											
SEC53	Phosphomannomutase	Up	9	13	0.79±0.04	2.22±0.09	--	--	--	--	
CIT1	Citrate synthase, mitochondrial	Down	26	39	1.25±0.05	1.78±0.11	0.76±0.04	0.57±0.09	0.80±0.05	0.69±0.06	
ENO2	Enolase 2	Down	43	181	0.18±0.06	0.29±0.02	0.26±0.02	0.29±0.04	0.87±0.1	0.42±0.08	
FBA1	Fructose-bisphosphate aldolase	Down	11	15	0.32±0.09	0.35±0.04	0.19±0.03	0.25±0.05	0.38±0.03	0.32±0.04	
TDH3	Glyceraldehyde-3-phosphate dehydrogenase 3	Down	52	139	0.11±0.01	0.21±0.02	0.16±0.03	0.19±0.02	0.19±0.02	0.24±0.04	
PYK1	Pyruvate kinase 1	Down	25	99	0.14±0.03	0.39±0.05	0.28±0.04	0.28±0.03	0.31±0.05	0.19±0.02	
PGK1	Phosphoglycerate kinase	Down	26	85	0.33±0.02	0.30±0.03	0.34±0.02	0.37±0.06	0.74±0.06	0.34±0.05	
TP11	Triosephosphate isomerase	Down	32	67	0.20±0.01	0.44±0.07	0.22±0.01	0.29±0.04	0.4±0.03	0.26±0.02	
GPM1	Phosphoglycerate mutase 1	Down	15	43	0.20±0.03	0.35±0.04	0.19±0.06	0.25±0.09	0.29±0.02	0.15±0.01	
PGI1	Glucose-6-phosphate isomerase	Down	16	53	0.28±0.01	0.48±0.05	0.36±0.05	0.31±0.07	0.46±0.05	0.30±0.03	
GLK1	Glucokinase-1	Down	7	26	--	--	--	--	0.40±0.07	0.17±0.03	
PCK1	Phosphoenolpyruvate carboxykinase [ATP]	Down	9	35	--	--	0.31±0.04	0.32±0.04	0.60±0.03	0.53±0.04	
PYC2	Pyruvate carboxylase 2	Down	14	26	--	--	--	--	0.49±0.03	0.5±0.07	
PDC1	Pyruvate decarboxylase isozyme 1	Down	23	93	0.12±0.01	0.27±0.04	0.16±0.01	0.20±0.04	0.14±0.02	0.12±0.02	
ADH1	Alcohol dehydrogenase 1	Down	20	23	0.19±0.02	0.28±0.05	0.42±0.02	0.34±0.03	0.36±0.02	0.2±0.03	

OYE2	NADPH dehydrogenase 2	Down	22	42	0.34±0.02	0.33±0.03	0.42±0.05	0.59±0.06	--	--
ICL1	Isocitrate lyase	Down	16	20	--	--	0.31±0.08	0.28±0.03	0.34±0.04	0.21±0.05
MDH1	Malate dehydrogenase, mitochondrial	Down	23	23	--	--	--	--	0.44±0.09	0.62±0.08
IDH2	Isocitrate dehydrogenase [NAD] subunit 2	Down	16	20	--	--	--	--	0.51±0.07	0.34±0.02
ADK1	Adenylate kinase 1	Down	14	25	--	--	0.36±0.07	0.52±0.09	--	--
HXK2	Hexokinase-2	Down	16	18	0.42±0.04	0.53±0.04	0.41±0.04	0.38±0.06	0.19±0.03	0.37±0.09
ALD6	Magnesium activated aldehyde dehydrogenase, cytosolic	Down	15	30	0.39±0.01	0.72±0.04	--	--	--	--
IPPI	Inorganic pyrophosphatase	Down	12	27	--	--	0.51±0.06	0.35±0.03	0.28±0.03	0.24±0.03
HXK1	Hexokinase-1	Down	6	43	0.15±0.03	0.44±0.05	--	--	--	--
HOM6	Homoserine dehydrogenase	Down	10	35	--	--	0.18±0.02	0.18±0.01	--	--
ACO1	Aconitate hydratase, mitochondrial	Down	9	44	--	--	0.33±0.05	0.33±0.05	--	--
ROCA2	1-pyrroline-5-carboxylate dehydrogenase 2	Down	16	36	--	--	--	--	0.74±0.06	0.4±0.02

Biosynthesis

TEF1	Elongation factor 1-alpha	Up	20	37	1.10±0.04	2.15±0.10	1.13±0.05	1.43±0.09	1.11±0.11	1.37±0.08
TMA19	Translationally-controlled tumor protein homolog	Down	20	22	0.56±0.04	0.43±0.05	--	--	--	--
ILV5	Ketol-acid reductoisomerase, mitochondrial	Down	18	17	0.25±0.02	0.42±0.07	0.31±0.03	0.27±0.04	--	--
TEF4	Elongation factor1-gamma	Down	12	22	--	--	--	--	0.22±0.04	0.25±0.03
EFB1	Elongation factor 1-beta	Down	11	23	0.13±0.03	0.26±0.05	0.19±0.02	0.21±0.01	--	--
EF11	Elongation factor 2	Down	10	25	0.28±0.03	0.40±0.02	0.85±0.07	0.35±0.06	0.18±0.02	0.06±0.02
YEF3	Elongation factor 3A	Down	19	16	0.19±0.01	0.56±0.03	0.37±0.03	0.45±0.05	0.21±0.04	0.12±0.04
GUS1	Glutamyl-tRNA synthetase	Down	21	40	--	--	--	--	0.43±0.04	0.32±0.09
FPP1	Farnesyl pyrophosphate synthase	Down	11	31	--	--	--	--	0.15±0.02	0.14±0.04

Transportation

ATP2	ATP synthase subunit beta	Up	16	63	2.52±0.07	3.18±0.12	2.49±0.08	2.02±0.13	1.79±0.14	2.09±0.11
AAC2	ADP, ATP carrier protein 2	Up	10	17	--	--	1.3±0.05	1.74±0.01	2.48±0.05	3.65±0.17
CYCK	Cytochrome c	Up	34	57	--	--	1.17±0.06	0.99±0.07	--	--

TRX2	Thioredoxin-2	Up	36	38	--	--	--	2.41±0.09	1.11±0.05
Signal transduction									
GDH3	NADP-specific glutamate dehydrogenase 2	Up	24	20	--	--	--	1.74±0.09	1.34±0.11
GND1	6-phosphogluconate dehydrogenase, decarboxylating 1	Up	11	43	--	--	--	1.24±0.08	1.92±0.09
CYC1	Cytochrome c iso-1	Up	14	26	1.47±0.06	1.2±0.07	1.36±0.06	1.63±0.09	1.15±0.04
PMA2	Plasma membrane ATPase 2	Down	17	26	0.29±0.03	0.36±0.03	0.64±0.06	0.27±0.03	0.5±0.07
POR1	Mitochondrial outer membrane protein porin 1	Down	13	34	--	--	--	0.17±0.02	0.18±0.07
NQM1	Transaldolase NQM1	Down	24	27	--	--	--	0.55±0.07	3.13±0.21
Ribosomal proteins									
RPS2	40S ribosomal protein S2	Down	13	39	0.21±0.02	0.21±0.01	0.21±0.02	0.24±0.03	0.29±0.03
RPS7A	40S ribosomal protein S7-A	Down	10	28	0.23±0.03	0.24±0.07	0.22±0.01	0.26±0.04	0.53±0.08
RPS5	40S ribosomal protein S5	Down	8	33	0.10±0.04	0.24±0.03	--	--	--
RPS18A	40S ribosomal protein S18	Down	5	30	0.16±0.03	0.24±0.02	--	--	--
RPS12	40S ribosomal protein S12	Down	16	49	0.34±0.08	0.55±0.09	0.36±0.06	0.29±0.02	--
RPS9B	40S ribosomal protein S9-B	Down	21	41	0.17±0.02	0.43±0.05	--	--	--
RPS20	40S ribosomal protein S20	Down	12	19	0.18±0.06	0.25±0.04	--	--	--
RPS4A	40S ribosomal protein S14	Down	14	40	0.28±0.02	0.52±0.06	--	--	0.35±0.07
RPS19A	40S ribosomal protein S19-A	Down	19	30	0.30±0.06	0.40±0.03	--	--	--
RPS1B	40S ribosomal protein S1-B	Down	22	36	0.17±0.01	0.48±0.06	0.32±0.04	0.29±0.02	0.24±0.02
RPS14B	40S ribosomal protein S14-B	Down	13	32	0.44±0.09	0.57±0.04	0.37±0.03	0.46±0.03	--
RPS9A	40S ribosomal protein S9-A	Down	11	47	--	--	--	--	0.91±0.10
RPS16A	40S ribosomal protein S16	Down	10	34	0.08±0.07	0.54±0.05	--	--	--
RPS0	40S ribosomal protein S0	Down	6	45	--	--	0.58±0.06	0.68±0.07	--
RPS3	40S ribosomal protein S3	Down	9	26	0.12±0.03	0.57±0.02	--	--	--
RPS10A	40S ribosomal protein S10-A	Down	12	33	0.38±0.09	0.66±0.9	0.67±0.07	0.49±0.05	--
RPL5	60S ribosomal protein L5	Down	15	34	0.04±0.03	1.63±0.13	0.21±0.02	0.60±0.09	--

RPP2A	60S acidic ribosomal protein P2-alpha	Down	20	19	0.17±0.06	0.31±0.04	--	--	--	--
RPL15A	60S ribosomal protein L15-A	Down	13	21	0.30±0.04	0.62±0.09	--	--	--	--
RPL17A	60S ribosomal protein L17-A	Down	16	43	0.27±0.03	0.44±0.07	--	--	--	--
RPL2A	60S ribosomal protein L2	Down	14	39	0.65±0.08	0.66±0.08	--	--	2.92±0.22	1.40±0.12
RPL12A	60S ribosomal protein L12	Down	9	22	0.25±0.07	0.40±0.05	0.26±0.03	0.28±0.07	--	--
RPL4A	60S ribosomal protein L4-A	Down	10	43	0.30±0.03	0.56±0.09	--	--	--	--
RPL3	60S ribosomal protein L3	Down	11	34	0.21±0.04	0.48±0.03	0.90±0.08	0.58±0.04	0.56±0.04	0.37±0.08
RPL31B	60S ribosomal protein L31-B	Down	4	19	0.25±0.06	0.29±0.04	0.42±0.05	0.20±0.03	--	--
RPP0	60S acidic ribosomal protein P0	Down	9	24	0.38±0.09	0.39±0.06	0.34±0.04	0.44±0.05	0.22±0.07	0.12±0.04
RPL8B	60S ribosomal protein L8-B	Down	13	34	0.38±0.10	0.32±0.03	--	--	--	--
RPL23A	60S ribosomal protein L23	Down	22	22	--	--	0.65±0.08	0.56±0.09	--	--
RPL28	60S ribosomal protein L28	Down	20	30	--	--	--	--	0.82±0.05	0.37±0.05
RPP2B	60S acidic ribosomal protein P2-beta	Down	12	29	--	--	--	--	0.44±0.08	0.31±0.03
Stress response										
HSP90AA1	Heat shock protein HSP90-alpha	Up	12	45	1.04±0.09	1.55±0.09	1.25±0.05	0.93±0.03	--	--
UBI1	Ubiquitin	Up	19	35	1.48±0.11	1.35±0.09	1.3±0.04	1.35±0.09	0.63±0.04	0.58±0.08
HSA1	Heat shock protein 70	Up	11	23	--	--	--	--	1.52±0.07	0.92±0.05
HSPA5	78kDa glucose-regulated protein	Up	19	46	1.51±0.10	1.37±0.07	1.71±0.04	1.63±0.02	0.84±0.06	0.65±0.09
AHP1	Peroxiredoxin type-2	Down	9	32	--	--	0.06±0.03	0.19±0.02	0.25±0.08	0.1±0.02
TSA1	Peroxiredoxin TSA1	Down	10	24	0.33±0.02	0.63±0.07	0.45±0.06	0.37±0.06	0.45±0.10	0.37±0.05
HSC82	ATP-dependant molecular chaperone HSC82	Down	23	26	0.28±0.02	0.34±0.06	0.61±0.08	0.43±0.08	0.46±0.09	0.42±0.08
SSB1	Heat shock protein SSB1	Down	12	32	0.26±0.02	0.36±0.03	0.5±0.09	0.52±0.09	0.54±0.01	0.26±0.03
SSA2	Heat shock protein SSA2	Down	14	22	--	--	0.75±0.04	0.85±0.11	--	--
SSC1	Heat shock protein SSC1	Down	19	13	--	--	0.67±0.06	0.61±0.04	0.51±0.09	0.51±0.05
HSP12	12 kDa heat shock protein	Down	9	19	--	--	0.24±0.03	0.25±0.06	0.43±0.08	0.43±0.04
HSC71	Heat shock protein cognate 70kDa protein	Down	6	11	0.46±0.05	0.93±0.09	1.35±0.08	0.68±0.07	--	--

HSP26	Heat shock protein 26	Down	11	32	--	--	0.21±0.06	0.36±0.02	--	--
SSE1	Heat shock protein homolog SSE1	Down	16	16	0.16±0.01	0.39±0.04	--	--	--	--
GPP1	(DL)-glycerol-3-phosphatase 1	Down	11	22	0.48±0.07	0.6±0.09	0.48±0.05	0.38±0.04	--	--
Structural proteins										
HHF1	Histone H4	Up	10	26	0.6±0.05	1.28±0.06	2.13±0.07	1.78±0.05	--	--
HTB1	Histone H2B	Down	17	27	0.42±0.06	0.69±0.08	0.89±0.06	1.15±0.07	--	--
Others										
BIP4	Luminal-binding protein 4	Up	32	34	--	--	0.89±0.03	1.15±0.08	--	--
ASC1	Guanine nucleotide-binding protein subunit β protein	Down	8	12	0.10±0.04	0.38±0.03	--	--	--	--
CPR1	Peptidyl-prolyl cis-trans isomerase	Down	9	33	0.86±0.08	0.28±0.02	0.37±0.03	0.61±0.07	0.17±0.02	0.21±0.02
TIF1	ATP-dependent RNA helicase eIF4A	Down	8	31	0.16±0.01	0.35±0.04	--	--	--	--
HYP2	Eukaryotic translation initiation factor 5A-2	Down	9	21	0.20±0.02	0.45±0.05	0.36±0.02	0.34±0.02	--	--
STM1	Suppressor protein STM1	Down	14	30	0.27±0.04	0.38±0.06	--	--	--	--
CDC48	Cell division control protein 48	Down	13	29	0.22±0.03	0.39±0.07	--	--	0.74±0.07	0.54±0.05
dnaE	DNA polymerase III subunit alpha	Down	10	20	0.6±0.09	0.86±0.09	--	--	--	--
CYS3	Cystathionine gamma-lyase	Down	8	32	0.16±0.04	0.97±0.1	--	--	--	--
ZEO1	Protein ZEO1	Down	11	39	0.12±0.01	0.42±0.07	0.36±0.03	0.23±0.05	--	--
BMH2	Protein BMH2	Down	15	28	0.69±0.08	0.71±0.05	0.32±0.04	0.24±0.03	--	--
SOD2	Superoxide dismutase [Mn], mitochondrial	Down	5	39	--	--	0.62±0.07	0.29±0.04	--	--
DUG1	Cys-Gly metalloproteinase	Down	14	37	--	--	--	--	0.24±0.03	0.28±0.05

*Data represented mean±S.D. based on three independent replicate experiments. “—” indicated the protein was absent.

**Early, Middle and Late indicated samples were from early, middle and late lipid accumulation stages.

1) The ratio of protein expression level in *C. albidus* (115) to *S. cerevisiae* (114).

2) The ratio of protein expression level in *R. toruloides* (116) to *S. cerevisiae* (114).