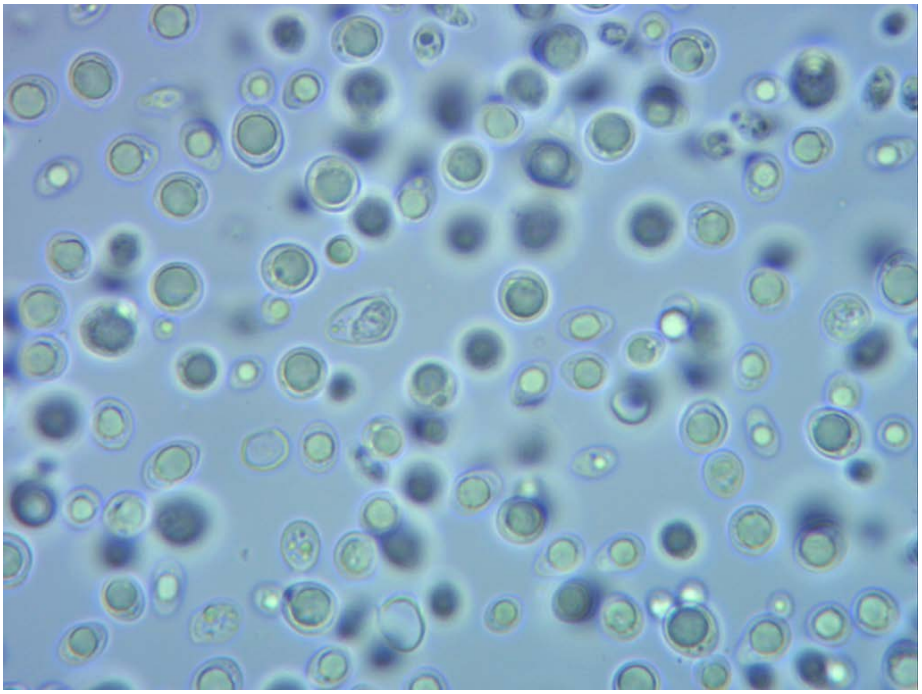
Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Natural Resources and
Agricultural Sciences

Establishing a predictive model for total fat content in *Lipomyces starkeyi* CBS 1807 by FT-NIR analysis

Irina Ehnström



Department of Molecular Sciences
Independent project • 15 hec • First cycle, G2E
Biology with specialization in Biotechnology
Uppsala 2018

Establishing a predictive model for total fat content in *Lipomyces starkeyi* CBS 1807 by FT-NIR analysis

Irina Ehnström

Supervisor: Mikolaj Chmielarz, Swedish University of Agricultural Sciences, Department of Molecular Sciences

Co- supervisor: Nils Mikkelsen, Swedish University of Agricultural Sciences, Department of Molecular Sciences

Examiner: Volkmar Passoth, Swedish University of Agricultural Sciences, Department of Molecular Sciences

Credits: 15 hec
Level: First cycle, G2E

Course title: Independent project in Biology - Bachelor project
Course code: EX0689
Programme/education: Biology with specialization in Biotechnology - Bachelor's Programme

Place of publication: Uppsala
Year of publication: 2018
Cover picture: Irina Ehnström, *Lipomyces strakeyi* CBS 1807
Title of series: Molecular Sciences

Part number: 2018:10

Online publication: http://stud.epsilon.slu.se

Keywords: *Oleaginous microorganisms, SCO, biofuels, FT-NIR, lipid extraction, prediction, validation*

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences
Faculty of Natural Resources and Agricultural Sciences
Department of Molecular Sciences

Abstract

This study investigates the potential of using near- infrared spectroscopy (FT-NIR) to establish a predictive model for total fat content in the oleaginous yeast *Lipomyces strakeyi* CBS 1807. FT-NIR-based quantification allows for rapid lipid determination compared to traditional extraction methods. The advantages of FT-NIR is not only rapid analysis, but also the ease of sample preparation resulting in little or no chemical waste. As FT-NIR is a chemometric analysis technique, it is possible to use a complete spectral structure in contrast to univariate analysis techniques, which only use one spectral datapoint. The spectra examined was within the wavelength range of 3600- 12800 cm^{-1} and two regions of the NIR spectra were chosen for the construction of the model (8771.2 cm^{-1} – 7922.6 cm^{-1}) and (5986.3 cm^{-1} – 5322.9 cm^{-1}). A calibration model was created based on the best RMSECV and R^2 values (RMSECV= 3.17, R^2 = 92.72) and used for further analysis of lipid content. Validation of the model was carried out by comparing predicted concentrations of lipids, using the model, to actual concentrations obtained from lipid extraction. The result from the calibration curve showed an average percentage error of ~ 24 %. These results show that further improvements are needed to increase the reliability of the model by the addition of a more representative set of test samples.

Keywords: Oleaginous microorganisms, SCO, biofuels, FT-NIR, lipid extraction, prediction, validation

Table of contents

1 Introduction

1.1 The challenge

As the agricultural expansion and global population has increased significantly, the application of new types of energy sources have become one of the most significant challenges in our present times (Sarris and Papanikolaou, 2016). Fossil resources are utilized for the production of various products such as chemicals, fuels and materials (e.g. plastic). The major usage of fossil resources is found in the transport sector that alone utilizes approximately two thirds of the total crude oil production world-wide (Karlsson, 2018). These factors in combination with the decrease in petroleum stock has shifted the focus from the conventional non-renewable energy source of fossil fuels to renewable ones such as biofuels. To this day there are two important types of liquid renewable biofuels found on the market, namely biodiesel and bioethanol.

Biodiesel is generated through the trans-esterification of plant oils, animal fats, waste products and short chained oils. Even though the demands for biofuel production are continuously increasing, the lack of oil feedstocks available has a negative impact on the price of conventional plant oil commodities, which could be seen as a two-fold increase in prices between the years 2007 and 2008. (Sarris and Papanikolaou, 2016). One approach taken to resolve this issue can be found with the aid of microorganisms, or more precisely in their ability to produce intracellular lipids, also known as single-cell oils (SCO) which can be further converted into valuable chemicals such as biofuels.

The production of bioethanol is considered the most industrialized process of applied microbiology world-wide, and in its' wake follows both economic and environmental benefits as ethanol has a high- octane number. This means that even small amounts of ethanol mixed with gasoline have great impact on the overall octane number. The combustion efficiency is also increased by the addition of ethanol due to its higher oxygen content. Moreover, a reduction in greenhouse emissions, carbon monoxide (CO), sulfur oxides, volatile organic compounds etc. are observed when compared to combustion of fossil fuels.

Bioethanol is the largest volume of biofuel used in the transportation sector world-wide and the production of starch-based (maize) ethanol mainly occurs in the US (Sarris and Papanikolaou, 2016) with a production of around 56 million m^3 of bioethanol in the year of 2015. The second largest bioethanol producing country the same year was Brazil, with an output production of 28 million m^3. Though the most important feedstock used for producing biodiesel is palm oil in East Asia (29%), soybean oil in North and South America (26%) followed by rapeseed in Europe (24%) (UFOP, 2016/2017).

The increasing interest in biofuel production using microorganisms requires the construction of genetically engineered microorganisms together with the discovery of new natural microorganisms for SCO production. Moreover, the production process of these SCO for further processing into biofuels must be optimized to yield high lipid contents at low costs. Optimization refers to the development of fermentation configurations as well as an increase in the utilization of various substrates, such as raw renewable materials, waste and by-products (Sarris and Papanikolaou, 2016).

1.2 The solution?

Microorganisms that are capable of accumulating over 20% of their dry cell weight (DCW) as lipids are defined as oleaginous microorganisms. An increasing interest in their potential to produce different valuable chemicals for e.g. biofuel production has led to efforts in optimization of bioprocesses as well as metabolic engineering. One of these microorganisms is the yeast *Lipomyces starkeyi* that has shown great potential as a host for lipid production as it is capable of utilizing a variety of carbon sources such as glucose and xylose. Previous studies have shown that *L. starkeyi* have accumulated lipid contents of 60.5 % of DCW when utilizing hemicellulose

hydrolysate with a xylose and acetic acid mixture (Brandenburg et al. 2016), and 55 % lipid content of DCW when cultivated on glucose. Moreover, the combination of complete genome sequencing of *L. starkeyi* and advancements in transformation techniques, genetic engineering of oleaginous yeast species that were not possible before will further improve the utilization of oleaginous yeasts species for biofuel production (Calvey *et. al*., 2016).

1.3 Fourier Transform Near Infrared spectroscopy (FT-NIR)

Demands of improved product quality in e.g. petrochemical, pharmaceutical, chemical, food and agricultural industries have led to a gradual replacement of gravimetric methods with methods that are non-invasive, create less chemical waste and are less time- consuming. For these reasons, together with ongoing chemometric software development, the interest for near infrared (NIR) spectroscopy is increasing (Siesler *et. al*, 2002). NIR spectroscopy is a valuable analytical technique for industrial applications due to its non-destructive nature, speed of analysis, and ability to perform both on-, at- and in- line analyses (Andersson, 2011). It has been demonstrated that FT-NIR serves as a potential technique capable of dealing with a multitude of problems such as determining the content of moisture, carbohydrate, protein and fat in different types of food. The ability to analyze food content provides the tools for optimization of product quality, as well as the ability to follow regulation requirements. Traditionally, these analyses have been performed using methods that are laborious and result in chemical waste, but FT-NIR spectroscopy requires little or no sample preparation which avoids generation of chemical waste and therefore can be seen as a more environmentally friendly approach (Ferreira et. al, 2013).

Chemometrics evaluation techniques have become important and widespread in modern analytical chemistry, and today chemometrics is used to refer to all multivariate calibration methods within this area. What differs from classical univariate calibration is that multivariate calibration methods make it possible to utilize a complete spectral structure in contrast to univariate calibration which uses only one spectral data point. Multivariate calibration techniques require quite an extensive number of results from analytical trials prior to the development of a mathematical model, thus making the process of creating a calibration method somewhat laborious. Moreover, when the mathematical model has been constructed,

it must be validated by independent samples i.e. samples not part of the constructed model. On the other hand, when the model has been developed and validated, it serves as a rapid tool in analytical chemistry (Conzen, 2014).

1.3.1 Multivariate calibration method and PLS- regression

The goal when working with quantitative analysis is the ability to determine a system property Y in a quantitatively manner from a measured system parameter X. For this to be possible, the correlation between Y and X is searched for during what is known as a calibration. The correlation between the parameters can be seen in equation 1 (Eq. 1).

$$Y_{(Analysis)} = X_{(Analysis)} * b \quad \text{(Eq. 1)}$$

Where $X_{(Analysis)}$ is the spectral data of the sample being analyzed, $Y_{(Analysis)}$ is the obtained concentration of said sample and *b* is usually referred to as the regression coefficient.

When the calibration has been performed it becomes possible to determine the system property Y of an unknown sample. When doing evaluations quantitatively, especially in near infrared spectrum, the value that is measured is usually an emission or absorbance spectrum, and the system value that is to be determined is often a concentration. A calibration method can be set up in two ways, either univariate (using a single variable) or multivariate. In the case of univariate calibration, information is gathered and correlated to the reference value from e.g. the area of one peak in the spectra, whereas in the multivariate calibration method, the combination of a large set of both spectral data and reference data is measured. One such multivariate calibration method is the increasingly popular PLS (Partial Least Squares) - regression method. This method requires a large set (20-200) of measured absorbance values, it becomes possible to correlate the system not only to one specific spectral peak but to the whole structure of the spectra (Conzen, 2014).

1.3.2 Validation of the calibration model

The quality of the analysis is dependent on the correlation between the combined concentration and spectral data. This means that good analytical

results can never be obtained by a bad correlation. Therefore, it is of outmost importance to find the function b that shows the best correlation and therefore a validation of the acquired model must be performed. This can be done by comparing what the model predicts for a set of samples with known concentrations, with the actual concentrations of the same samples. The samples used for validation is preferably “independent” which means that they have not been included as part of the model, but stand completely separate (Conzen, 2014).

2 Aim

The aim of this project was to establish a predictive model for estimation of total fat content in *Lipomyces starkey*i CBS 1807 by FT-NIR analysis (MPA, Bruker), as FT-NIR based quantification allows for rapid lipid determination, little or no sample preparation and less hazardous work.

3 Materials and methods

3.1 Yeast strain

The yeast strain used in this project was *Lipomyces starkeyi* CBS 1807.

3.2 Sterilization

Prior to laboratory work, all utensils were sterilized by autoclaving. YNB media components were sterile-filtered with 0.2 μm vacuum filters.

3.3 Media and inoculates

3.3.1 Starter culture set-up

Components for pre-culture (YPD) media can be seen in table 1.

Table 1. *Pre-culture components*

Media	**Composition**
YPD	• Glucose 20g/L • Yeast Extract 10g/L • Peptone 20g/L • pH= 6 (adjusted with 1 M HCl)

3.3.2 Preparation of pre-culture

Lipomyces starkeyi CBS 1807 colonies were collected from a YPD plate and transferred to a 500 ml baffled flask containing 100 mL of pre-culture and placed in a shaker (25 °C at 130 rpm) for 24h. This process was repeated, resulting in another pre-culture that was put in a shaker (25 °C at 130 rpm) for 72h.

Prior to inoculation, optical density (OD_{600}) was measured on a CO8000 Cell Density Meter (WPA Biowave) to confirm cell growth. The pre-culture was then transferred into sterile falcon tubes and centrifuged on a Soervall ST 8 Centrifuge (Thermo Scientific) at 4500 x g for 5 minutes, and the supernatant was discarded. The pellets were washed with NaCl-solution (9 g/l), vortexed and spun down a second time and the supernatant was discarded. dH_2O was added to dissolve the pellet and OD_{600} was measured to calculate the volume of cell suspension needed to reach a starting OD_{600} of ~1.

3.3.3 Preparation of test media and inoculation

YNB media was prepared according to table 2 and filtered.

Table 2. *Inoculum culture components*

Media	Composition
YNB	• Glucose 20g/L • Yeast Extract 0.75g/L • YNB 1.7g/L • Ammonium phosphate 2g/L • pH= 6 (adjusted with 1 M HCl)

250 ml shake flasks were prepared with 50 ml of YNB media, 2 mL of 1.5 M Potassium Phosphate Buffer and cell suspension. Initial OD_{600} was measured and the shaking flasks were placed on a shake table (25 °C at 130 rpm) for 72h.

After inoculation, OD_{600} was measured to confirm cell growth prior to cell collection. The media was transferred into 50 mL falcon tubes and centrifuged (4400 x g for 5 min). After centrifugation, the supernatant was discarded and the pellet was washed with dH_2O, vortexed and spun down

(4400 x g for 5 min). The supernatant was removed and the pellets were stored in the freezer at ~ (-20 °C).
This process of inoculation in 250 mL shake flasks and cell harvesting was repeated until 25 pellets were collected.

3.4 Fermentation in bioreactors

For quantification of lipid production and assessment of dry mass, cultivation in two 2.0 L Minifors Bioreactors (Infors AG, Switzerland) was performed with the starting volume of 1.8 L, with the following parameters : 25°C, pH=6, stirrer speed 300 rpm and aeration at 0.3 L/min. The composition of media used can be seen in table 3.

Table 3. Media components for fermentation in bioreactors.

Media	Composition
YNB	• Glucose 20g/L • Yeast Extract 0.75g/L • Ammonium phosphate 2g/L • Glycerol (97%) 20g/L • YNB 1.7g/L • pH= 6 (adjusted with 1 M HCl)

3.4.1 Sampling from bioreactors

Samples of ~50 ml were withdrawn from the bioreactors two-three times every day and OD_{600} was measured using deionized water as a blank. The samples were centrifuged at 4500 x g for 5 minutes, washed with deionized water, centrifuged again at 4500 x g for 5 minutes and frozen until further analysis. This process was repeated until 73 pellets were obtained.

3.4.2 Feeding solutions and anti-foaming agent

Separate feeding solutions were prepared for each bioreactor, see table 4 for all calculations. A first feeding was performed after ~45 h of cultivation to each bioreactor. One feeding solution contained a glucose-ammonium phosphate mixture with a total volume of 200 mL, and second feeding solution contained a glucose-ammonium sulphate mixture with a total volume of 200 mL.

Additional feeding solutions (50 ml each) were prepared and added at ~92h, ~165h and ~189h after initial set up.

3.5 Calculations for feeding solutions

The first feeding solution for Minifors bioreactor 1 (45 h) were calculated for a total volume of 200 mL, the following feeding solutions were calculated for a total volume of 50 mL. The differences in composition of the feeding solutions was to accomplish differences in intracellular lipid accumulation during cultivation.

Table 4. Calculations for feeding solutions prepared for Minifors Bioreactors.

Bioreactor	Feeding time after initial set up (h)	Content	Calculated for a total volume (mL)	Amount (g)	Final concetration in solution (g/L)
Minifors 1	45	Glucose	1650	33	165
		Ammonium phosphate	1650	3.3	16.5
	92	Glucose	1350	27	540
		Ammonium phosphate	1350	2.7	54
	165	Glucose	1200	24	480
		Ammonium phosphate	1200	2.4	48
	189	Glucose	1150	23	460
		Ammonium phosphate	1150	2.3	46
Total amount		Glucose		107	
		Ammonium phosphate		10.7	
Minifors 2	45	Glucose	1650	16.5	82.5
		Ammonium sulphate	1650	1.65	8.25
	165	Glucose	1250	25	500
		Ammonium sulphate	1250	1.25	25
	189	Glucose	1200	24	480
		Ammonium sulphate	1200	1.2	24
Total amount		Glucose		65.5	
		Ammonium sulphate		4.1	

Antifoaming agents PPG and antifoam B were added to the bioreactors when necessary. A total volume of 0.5 mL PPG and a total volume of 0.25 mL Antifoam B was added.

3.6 Freeze drying

Prior to lipid extraction, freeze drying was performed to remove water. The 50 mL falcon tubes containing cells stored at -20 °C (see 3.4.1) were transferred to a freeze dryer and left for 72h. The freeze-dried samples were stored in the freezer at -20 °C.

3.7 FT-NIR measurements

Samples of freeze dried cells were transferred to flat bottomed glass vials (inner diameter of 19.6 mm) and measured in triplicates to get three (3) spectra per sample using a Multipurpose Analyzer Fourier Transform Near Infrared spectrometer (MPA FT-NIR, Bruker Optik GmbH, Ettlinger, Germany). Between each measurement, the pestle was twisted a few times to obtain slight changes in readings. The set up prior to measurement was as follows;

- Measurement was set to “Sphere Macrosample”
- Resolution: 16 cm^{-1}
- Detector setting: RT-PbS (External) 10 KHz
- Sample form: Integrating Sphere (Sphere Macrosample)
- Wavelength range of 3600- 12800 cm^{-1}
- Result spectrum: Absorbance

3.8 Lipid Extraction using modified Folch- method

Lipid extraction was done using a modified Folch- method (Folch *et al,* 1957). Approximately 100 mg of freeze dried cells were transferred to 10 mL glass tube together with 2 mL of 1 M HCL. The vial was left to soak for 15 minutes prior to incubation on a heating block at 75 °C for 1 h. The tube was vortexed every 15 min during incubation. After incubation, 2 mL of KCl-solution (8 g/l) and 6 mL of Folch solution (2:1 chloroform : methanol) was added to the tube. The tube was vortexed for a few seconds then centrifuged (Hettich EBA 12) at 2000 rpm for 5 minutes. The lower layer was transferred to a second 10 mL glass tube using a Pasteur- pipette and an additional 4 mL of chloroform was added to the first glass tube, which was then centrifuged again at 2000 rpm for 5 minutes. The remaining lower layer was transferred to the second glass tube and centrifuged at 2000 rpm for 5 minutes. The lower layer from the second tube was transferred to a

third pre-weighted 10 mL glass tube and placed under a stream of N_2 for chloroform evaporation. The lipid quantity was determined in % of cell dry weight by weighing the pre-weighted tube after evaporation. The sample was re-suspended in 1 mL of hexane and stored in the freezer at -20 °C. This process was performed in duplicates.

3.9 Chemometrics analysis

3.9.1 Data modeling

A multivariate calibration model was established by using partial least squares (PLS). The obtained NIR-spectra was treated in OPUS Quant 2 Method.

3.9.2 Choice of samples for calibration and validation

Freeze- dried samples were scanned in triplicates and used for constructing a calibration curve. The total fat content (in %) for each sample was assigned to its corresponding spectra and the model was evaluated using a separate data series external validation.

3.9.3 Model evaluation

The model was evaluated based on the coefficient of correlation (R^2) and root mean square error of cross-validation (RMSECV).

4 Results and Discussion

4.1 Construction and evaluation of the calibration model

After harvesting and freeze- drying *L. starkeyi* cells, a FT-NIR spectra were established and a calibration model was constructed by assigning the average lipid content (in %) obtained by gravimetric determination to the corresponding spectra. The FT-NIR calibration curve is shown in Figure 1, constructed out of 108 spectra. Two near- infrared regions (8771.2 cm^{-1} – 7922.6 cm^{-1}) for CH_2 – CH_3 second overtones, and (5986.3 cm^{-1} – 5322.9 cm^{-1}) for CH_2- CH_3 overtones, were used in calibration. The model had a good correlation between the predicted values obtained from the FT-NIR readings and the reference values obtained from lipid extraction, as the coefficient of correlation (R^2) was 92.71 % and root mean square error of cross-validation (RMSECV) was 3.17. See Appendix 1 (4.5) for calibration data.

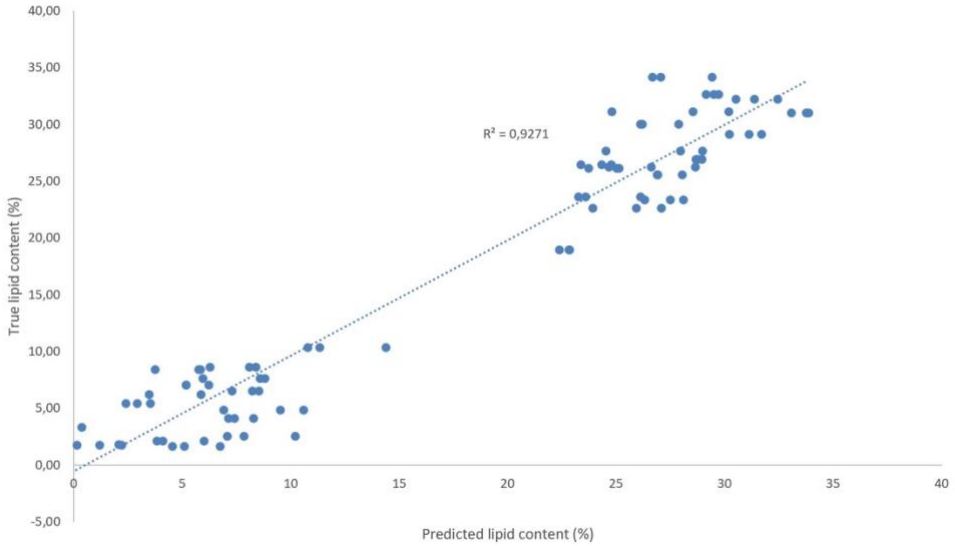


Figure 1. Calibration model constructed out of 108 samples with a coefficient of correlation (R^2) of 92.71 % and a root mean square error of cross-validation (RMSECV) value of 3.17.

Samples included in the calibration model must be built on a set of representative test samples that covers the whole range of concentrations. If this is not the case, the validation of the model will not lead to reliable statements concerning expected errors in analysis due to lack of natural variance (Conzen, 2014). As seen in Figure 1, the calibration model does not cover the whole concentration range homogenously. Therefore, the calibration curve could be constructed continuously rather than late in the experimental stages so that samples can be chosen more selectively. Online construction can be quite difficult if there is a narrow window when it is possible to obtain representative data such as the log-phase of cell growth. Examining freeze-dried samples 72 h after harvesting can lead to errors while constructing the calibration model if this window is not included. Therefore, a different sampling scheme may be more suitable to avoid these errors.

4.1.1 Optimizing the model

When constructing a model, it is necessary to find the most suitable method for the task at hand, such as finding the best combination of frequency ranges and processing of data. Since there is no true answer to what the best combination is, one must search for these parameters through a series of trial and error. This is done with the aid of values such

as the coefficient of determination R^2 displaying the variance present in the component values shown in percentage where values over 90 % are considered good when working with solid materials. Another value to be considered is the RMSECV (Root Mean Square Error of Cross Validation) which measures the preciseness with which the samples are predicted during validation (Conzen, 2014). It would be possible to further improve the model by applying a period of trial and error as the R^2 value is only 92.72 %.

4.2 Validation of FT-NIR model

For validation, twelve separate samples were chosen that were not part of the calibration model. Spectra from these samples were read in triplicates and the model was used to predict the average lipid content of these spectra. Lipid extraction were performed in duplicates. The results from the validation can be seen in Figure 2. See Appendix 1 (4.6) for validation data.

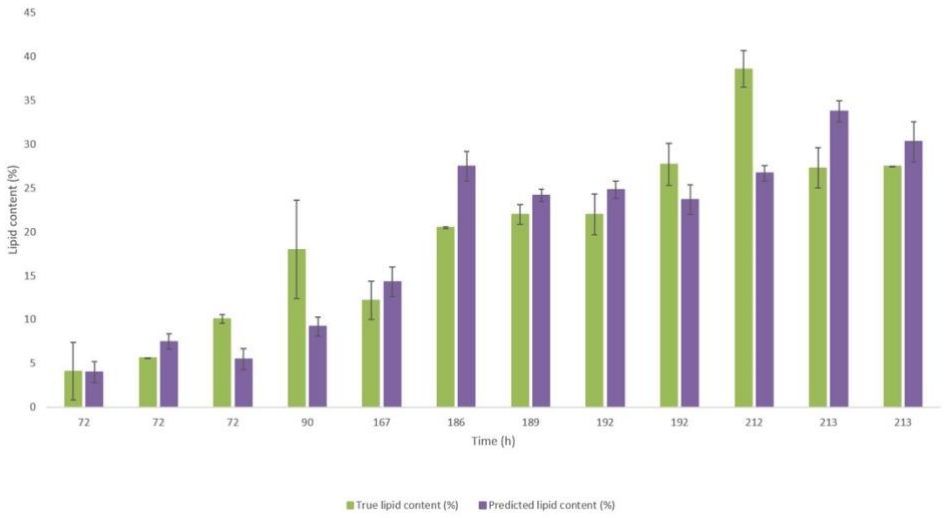


Figure 2. Validation of the FT-NIR model represented as a bar chart. Green bars indicate the average lipid content in % extracted and the purple bars represent the predicted lipid content in %.

From the observed results, an average percentage error of ~ 24 % was calculated. These results show that the model requires further improvements. As mentioned above, this can be achieved by collecting a more homogenous set of test samples but also performing lipid extractions in triplicates rather than duplicates to receive a more representative average lipid content. In addition to the sources of error mentioned above, performing lipid extractions using the Folch method always includes the risk of contamination during transfers of the chloroform layers.

4.3 Unexpected findings during validation of the FT-NIR model

An unknown amount of 15% NaOH was pumped into one of the Minifors Bioreactors during early stages of fermentation, increasing the pH from around 6 to 12. At this point it was believed that the cells had not survived, though the OD was measured continuously at regular intervals to see if the cells would recover. After ~42 h, the OD increased from 22 to 26 and cell harvesting was resumed.

The results from lipid extraction of the freeze-dried samples chosen for validation that originated from this bioreactor showed suspiciously low lipid content. These samples were therefore examined under microscope and the observed content showed that the extraction method was not working properly. An increase in molarity of HCl from 1 M to 5 M resulted in an extracted lipid content corresponding more closely to the observed lipid content. These finding indicate that the stress response of the increased pH might have affected the resilience in the yeast towards lysis. If the thickening of the cell wall is a response to pH and osmotic stress can only be confirmed by further examination.

4.4 Conclusion

The aim of the studies were reached as a predictive model was created. However, the results show that the reliability of the model can be enhanced by the addition of more representative test samples. Therefore, establishing a more reliable model will need improvements like including representative test samples, the ability to construct the model continuously and

performing lipid extractions in triplicates. If these considerations are addressed, near-infrared spectroscopy and multivariate calibration shows great potential for predicting total fat content in *Lipomyces starkeyi* CBS 1807.

5 References

Andersson F. 2011. Near infrared spectroscopy as an analytical tool in hyaluronan processes, Master Thesis, Teknik-Naturvetenskaplig fakultet UTH- enheten, ISSN: 1650-8297, K11014.

Brandenburg J., J Blomqvist, J. Pickova, N. Bonturi, M. Sandgren, V. Passoth, 2016. Lipid production from hemicellulose with *Lipomyces starkeyi* in a pH regulated fed-batch cultivation" Yeast 33: 451–462. DOI: 10.1002/yea.3160

Calvey CH, Su YK, Willis LB, et al. 2016. Nitrogen limitation, oxygen limitation, and lipid accumulation in *Lipomyces starkeyi*. Bioresource Technology **200**: 780–788.

Conzen J-P. Multivariate Calibration, A practical guide for developing methods in the quantitative analytical chemistry, 3rd ed, Bruker Optik GmbH (2014)

Karlsson H. 2018. Climate Impact and energy balance of emerging biorefinery systems. Doctoral Thesis No.2018:03. Faculty of Natural resources and Agricultural Sciences, Uppsala, Sweden.

Ferreira S. D, J.A.L Pallone, R.J Poppi, 2013. Fourier transform near- infrared spectroscopy (FT-NIRS) application to estimate Brazilian soybean [*Glycine max* (L.) Merril] compostion. Food Research International **51** (2013) 53-58

Folch, J., M. Lees and G. H. S Stanley, 1957. A simple method for the isolation and purification of total lipides from animal tissues, Journal of Biological Chemistry **226**(1): 497-509

Sarris D, S. Papanikolaou, 2016. Biotechnological production of ethanol: Biochemistry, processes and technologies. *Eng. Life Sci.* **2016**, *16*, 307–329

Siesler H.W, Y Ozaki, S. Kawata, H.M Heise. Near-infrared spectroscopy principles, instruments, applications, 1st ed, Wiley- VCH, Weinheim, DE, (2002)

UFOP, Supply report 2016/2017, European and world demand for biomass for the purpose of biofuel production in relation to supply in the food and feedstuff markets from https://www.ufop.de/files/7814/9977/4144/UFOP_supply_report_20162017.pdf

Acknowledgements

Thanks to everyone at the Department of Molecular Sciences for making this independent project an experience to remember. Special thanks to my supervisor Mikolaj Chmielarz for guidance in the practical work and in the writing process, and for throwing a nice barbecue *"na zdrowie"!*

Appendix 1

5.1 Calibration curve data

Crossed out samples were excluded from the model, values shown in () in column “Avrage lipid (%) “ are included in the model.

MF1 = Minifors 1
MF2 = Minifors 2
YNB = Shake flask

Sample name	Tube Pre-weight (mg)	Tube Post-weight (mg)	Sample weight (mg)	Total lipid (mg)	Lipid (%)	Avrage lipid (%)
MF1.20 (1)	12000.4	12035.2	101.2	34.8	34.2	34.2
MF1.21 (1)	11375.9	11408.3	101.2	32.4	32.0	
MF1.21 (2)	11423.4	11456.9	102.7	33.5	32.6	32.3
MF1.22 (1)	11455.1	11485.2	101.4	30.1	29.7	
MF1.22 (2)	11154.0	11190.0	101.0	36.0	35.6	32.7
MF1.23 (1)	11967.5	11994.2	101.1	26.7	26.4	
MF1.23 (2)	11911.8	11939.7	101.3	27.9	27.5	27.0
MF1.24 (1)	11489.3	11516.6	101.0	27.3	27.0	
MF1.24 (2)	11408.8	11437.4	101.2	28.6	28.3	27.7
MF1.25 (1)	11433.0	11465.9	100.8	32.9	32.6	
MF1.25 (2)	10949.5	10979.6	101.4	30.1	29.7	31.2
MF1.26 (1)	11271.6	11311.4	101.1	39.8	39.4	
MF1.26 (2)	11412.9	11449.9	100.9	37.0	36.7	38.1
MF1.27 (1)	11330.6	11356.9	101.4	26.3	25.9	
MF1.27 (2)	11421.3	11454.0	100.9	32.7	32.4	29.2
MF1.28 (1)	12675.4	12689.6	100.1	14.2	14.2	
MF1.28 (2)	12733.6	12757.3	100.3	23.7	23.6	19.0
MF1.29 (1)	12811.2	12834.3	100.8	23.1	23.0	
MF1.29 (2)	12686.8	12710.5	100.2	23.7	23.7	23.4
MF1.30 (1)	12601.4	12631.3	100.3	29.9	29.8	

MF1.30 (2)	12679.2	12709.7	100.4	30.5	30.4	30.1
MF1.31 (1)	12633.4	12658.7	99.7	25.3	25.4	
MF1.31 (2)	12706.1	12732.0	100.7	25.9	25.7	25.6
MF1.32 (1)	12846.1	12871.1	100.3	25.0	24.9	
MF1.32 (2)	12590.8	12611.3	100.0	20.5	20.5	22.7
MF1.33 (1)	12529.5	12555.4	99.8	25.9	26.0	
MF1.33 (2)	12704.5	12731.6	100.5	27.1	27.0	26.5
MF1.34 (1)	11255.8	11278.6	101.1	22.8	22.6	
~~MF1.34 (2)~~	~~11489.1~~	~~11515.4~~	~~102.1~~	~~26.3~~	~~25.8~~	~~24.2~~
~~MF1.35 (1)~~	~~11584.9~~	~~11610.5~~	~~100.5~~	~~25.6~~	~~25.5~~	
MF1.35 (2)	11493.9	11521.3	101.2	27.4	27.1	26.3
~~MF2.10 (1)~~	~~12032.5~~	~~12050.0~~	~~100.5~~	~~17.5~~	~~17.4~~	
~~MF2.10 (2)~~	~~11911.1~~	~~11967.4~~	~~100.9~~	~~56.3~~	~~55.8~~	~~36.6~~
MF1.11 (1)	10949.1	10954.3	99.8	5.2	5.2	
MF1.11 (2)	11248.9	11254.7	101.4	5.8	5.7	5.5
MF1 10 (1)	11958.6	11963.8	101.5	5.2	5.1	
MF1.10 (2)	11572.6	11572.7	100.2	0.1	0.1	2.6*
MF2.11 (1)	11669.2	11671.3	100.0	2.1	2.1	
MF2.11 (2)	11331.2	11332.7	100.0	1.5	1.5	1.8
MF1.13 (1)	12529.6	12553.1	100.1	23.2	23.2	
MF1.13 (2)	12590.8	12615.3	101.6	24.5	24.1	23.7
MF2.12 (1)	12633.2	12644.2	101.5	11	10.8	
MF2.12 (2)	12812.5	12822.5	100.4	10	10.0	10.4
MF1.12 (1)	12679.3	12701.7	100.7	22.4	22.2	
MF1.12 (2)	12846.2	12876.7	101.0	30.5	30.2	26.2
~~MF2.9 (1)~~	~~12732.6~~	~~12744.3~~	~~100.2~~	~~11.7~~	~~11.7~~	
~~MF2.9 (2)~~	~~12686.6~~	~~12700.0~~	~~100.6~~	~~13.4~~	~~13.3~~	~~12.5~~
MF1.5 (1)	12773.7	12784.2	100.7	10.5	10.4	
MF1.5 (2)	12706.2	12712.7	100.4	6.5	6.5	8.5
MF1.6 (1)	12661.2	12668.6	101.2	7.4	7.3	
MF1.6 (2)	12601.0	12609.2	101.7	8.2	8.1	7.7
MF1.4 (1)	12750.2	12756.0	101.0	5.8	5.7	
MF1.4 (2)	12564.4	12570.9	96.0	6.5	6.8	6.3
MF1.7 (1)	12627.3	12633.2	103.5	5.9	5.7	
MF1.7 (2)	12432.9	12437.0	103.1	4.1	4.0	4.9
MF1.8 (1)	12671.3	12676.4	99.4	5.1	5.1	
MF1.8 (2)	12748.8	12758.1	101.8	9.3	9.1	7.1

MF2.6 (1)	12661.9	12664.1	98.8	2.2	2.2	
MF2.6 (2)	12635.3	12637.3	92.3	2.0	2.2	2.2
MF2.4 (1)	12752.2	12754.3	101.8	2.1	2.1	
MF2.4 (2)	12642.6	12644.2	101.1	1.6	1.6	1.9
MF2.7 (1)	12653.7	12655.6	100.5	1.9	1.9	
MF2.7 (2)	12704.4	12705.8	97.8	1.4	1.4	1.7
MF2.8 (1)	12704.6	12708.2	99.0	3.6	3.6	
MF2.8 (2)	12747.4	12750.6	100.1	3.2	3.2	3.4
YNB.A (1)	12737.2	12743.4	100.0	6.2	6.2	
YNB.A (2)	12632.8	12636.1	100.8	3.3	3.3	4.8
YNB.O (1)	12529.5	12535.2	102.1	5.7	5.6	
YNB.O (2)	12600.8	12608.4	100.2	7.6	7.6	6.6
YNB.I (1)	12810.6	12821.5	99.2	10.9	11.0	
YNB.I (2)	12679.0	12685.1	97.7	6.1	6.3	8.7
YNB.E (1)	12657.5	12660.8	101.3	3.3	3.3	
YNB.E (2)	12686.3	12696.2	99.9	9.9	9.9	6.6 (3.3)
YNB.N (1)	12590.5	12594.3	101.0	3.8	3.8	
YNB.N (2)	12845.9	12850.4	100.1	4.5	4.5	4.2

5.2 Validation data

Samples chosen for validation of FT-NIR model, extracted with 5 M HCl
MF1 = Minifors 1
MF2 = Minifors 2
YNB = Shake flask

Sample	Pre-weight tube (mg)	Post-weight tube (mg)	Sample weight (mg)	Total Lipid (mg)	Lipid (%)	Avrage lipid in sample (%)	Avrage predicted (%)
YNB.B (1)	11059,6	11065,3	102,9	5,7	5,5		
YNB.B (2)	10729,6	10735,2	98,8	5,6	5,6	5,6	7,5
YNB.K (1)	11422,5	11433,3	111.3	10.8	9,7		
YNB.K (2)	12302,2	12310,5	78,5	8,2	10,4	10,1	5,5
MF1.9 (1)	11411,4	11435,7	111	24,3	21,9		
MF1.9 (2)	11299,8	11314,8	107,3	15	14	18	9,2
MF2.18 (1)	11502,2	11531,9	125,6	29,7	23,6		
MF2.18 (2)	12009,2	12033,1	117,5	23,9	20,3	22	24,8
MF2.22 (1)	10631,8	10662,6	111,1	30,4	27,4		
MF2.22 (2)	10841,3	10872	111,5	30,7	27,5	27,5	30,3
MF2.31 (1)	11702,5	11733,6	107,6	31,1	28,9		
MF2.31 (2)	11218,3	11246,3	109,5	28	25,6	27,3	33,8
MF2.14 (1)	12670,5	12685,6	110,4	15,1	13,7		
MF2.14 (2)	12747,3	12758,6	106,3	11,3	10,6	12,2	14,3
MF2.17 (1)	12846	12873,3	119,5	27,3	22,8		
MF2.17 (2)	12590	12612,6	106,6	22,6	21,2	22	24,2
MF2.16 (1)	12602,2	12625,1	112,3	22,9	20,4		
MF2.16 (2)	12727,9	12751	112,1	23,1	20,6	20,5	27,5
MF1.18 (1)	12679,9	12712,6	111,3	32,7	29,4		
MF1.18 (2)	11472,6	11502,6	115,4	30	26	27,7	23,7
MF1.19 (1)	11600,8	11644,9	110,1	44,1	40,1		
MF1.19 (2)	11547,9	11592	118,9	44,1	37,1	38,6	26,7
YNB.G (1)	11290,7	11292,5	102	1,8	1,8		
YNB.G (1)	11151,6	11156,1	70,3	4,5	6,4	4,1	4