Optimization of Lignocellulosic Bioethanol Production Using Pentose Fermenting Yeasts and Raman Spectroscopy

Iversen, Jens Asmus

Publication date:
2014

Document Version
Early version, also known as pre-print

OPTIMIZATION OF LIGNOCELLULOSIC BIOETHANOL PRODUCTION USING PENTOSE FERMENTING YEASTS AND RAMAN SPECTROSCOPY

PhD thesis by

Jens Asmus Iversen

AALBORG UNIVERSITY
DENMARK

Section for Sustainable Biotechnology

Aalborg University

Copenhagen 2014

Aalborg University
Faculty of Engineering and Science
Department of Biotechnology, Chemistry and Environmental Engineering
Section for Sustainable Biotechnology

PhD student:
Jens Asmus Iversen (contact: jensive@gmail.com)

Academic Supervisor:
Professor Birgitte Kiær Ahring

PhD assessment committee:
Professor Erik Gydesen Søgaard, Aalborg University, Denmark (Chairman)
Professor Isabel Sá-Correia, Universidade de Lisboa, Portugal
Associate Professor Bo Jensen, Copenhagen University, Denmark

ISBN: 978-87-93100-56-5

Published by:
Aalborg University Press
Skjernvej 4A, 2nd floor
DK – 9220 Aalborg Ø
Phone: +45 99407140
aauf@forlag.aau.dk
forlag.aau.dk

Printed at Uniprint Aalborg

Cover illustration: Raman sapphire ball probe

PREFACE

This PhD thesis is the result of a research project carried out at the Section for Sustainable Biotechnology (SSB), Aalborg University Copenhagen, Denmark throughout the period lasting from November 2009 to March 2014. The research was done under the supervision of professor Birgitte K. Ahring and financially supported by a PhD Scholarship from Aalborg University and by the Danish Council for Strategic Research, grant numbers: 2104-05-0017 and 09-065165, and was part of the projects "Bioethanol from Important Foreign Biomasses" and "Bioref: A novel biorefinery concept".

I was given the opportunity to carry out most of the work at the new Bioproducts, Sciences & Engineering Laboratory at Washington State University in Richland, Washington State, where I was fortunate to have access to state of the art research facilities. I would like to thank all the wonderful colleagues at WSU for their hospitality and the great friendships made during my stay. Also, I would like to thank Peter S. Lübeck at SSB for helping with formalities initially and at the end of the project as well as Rolf W. Berg at DTU for helping me through the process of writing my first science paper for publication. I am grateful for the help I got from my dear Marie Cimmiyotti for making final corrections. Finally, I would like to thank my family and especially my parents for their love, support and patience.

Jens Asmus Iversen
Sønderborg, March 2014

SUMMARY

The unfortunate impacts arising from the continued use of fossil fuels has in recent years given rise to an increased focus on the development of biofuels. Bioethanol has in particular been considered as a viable alternative to transportation fuels derived from oil. However, the current controversial production systems and processes based on sugar and starch has to undergo change to enable lignocellulosic biomass to be utilized as feedstock instead of food. Comprehensive research and development is needed in order to optimize the various processes involved in the conversion of biomass to 2nd generation bioethanol. The recalcitrant structure of lignocellulose requires a combination of thermochemical and enzymatic pretreatment to release sugars to be used for the subsequent fermentation process. The fermentation step, with its many process parameters, has endless opportunities for improvements and to optimize the process. Therefore, one way to be able to create optimal conditions for fermentation, or the prior pretreatment steps for that matter, is by being able to monitor these conversions in a simple manner to retrieve as much process information as possible. The development of 2nd generation bioethanol production is also being challenged by the fact that only few yeast or bacteria species in the wild are capable of producing ethanol efficiently from all the different types of sugars found in lignocellulosic biomass.

The aim of the this research project was to study the possibilities and limitations arising from applying monitoring techniques for controlling fermentations or other processes in a biorefinery setting. Such methods may help improve process control or reduce labor costs associated with sampling and analysis. Monitoring was attempted using Raman spectroscopy to follow the progress of a yeast fermentations in real-time. The factors affecting Raman measurements such as attenuation of signal by the suspended particulates were measured and build into a quantification models. This made it possible with Raman to continuously measure the concentrations of ethanol, glucose and cell biomass in the growth media during fermentations with *Saccharomyces cerevisiae*. Concentration quantifications were performed without using complex statistics involving chemometric or multivariate methods commonly used with other spectroscopy methods such as NIR. Overall, monitoring a fermentation process containing biomass residues in the hydrolysate using this method yielded reasonable concentration values to be collected, as long as the amount of particulates surrounding the immersion probe could be kept at a low level. An attempt to monitor a pretreatment reaction, however, seemed to give more inconsistent results, while applying Raman to monitor a hydrolysis process did not appear applicable at all.

In another part of the project, I focused on working with the two xylose-fermenting yeasts *Scheffersomyces* (*Pichia*) *stipitis* and the more recent isolated *Spathaspora passalidarum*. The yeasts were grown on hydrolysates prepared from three abundant tropical biomasses: Sugarcane bagasse, eucalyptus tree fiber and empty fruit bunches (EFB), residues from the palm oil production. The comparison between the two yeasts resulted in similar yields, but the conversion rate of *S. passalidarum* was slightly better under the micro-aerobic conditions. The higher level of acetic acid found in the hydrolysate from EFB (6 g/L) resulted in a longer lag phase, which initially seemed to inhibit *S. passalidarum* the most. Despite inhibition in the beginning, this fermentation eventually yielded the highest ethanol concentration of about 20 g/L. Under strict anaerobic conditions, neither of two yeasts where capable to overcome the inhibition caused by the acetic acid.

RESUME

De uønskede virkninger ved den fortsatte anvendelse af fossile brændstoffer har i de senere år givet anledning til en øget fokus på udviklingen af biobrændstoffer. Bioethanol forventes især at have potentiale til at blive et bæredygtigt alternativ til brændstoffer fremstillet af olie. De eksisterende kontroversielle første generations produktionssystemer baseret på sukker og stivelse må dog ændres nok til at kunne bruge biomasse af lignocellulose som råvare i stedet for fødevarer. Det nødvendigt med en omfattende forskningsindsats for at opnå dette med henblik på at optimere de forskellige delprocesser, der indgår i omdannelsen af biomasse til anden generations bioethanol. Det svært nedbrydelige lignocellulose kræver en kombination af termokemisk og enzymatisk forbehandling for at kunne frigive sukre til den efterfølgende gæringsproces. Gæringsprocessen med sine mange procesparametre har uendelig mange forbedringsmuligheder. De optimale procesbetingelser for en gæringen eller det forudgående forbehandlingstrin skabes, hvis omdannelsen hele tiden kan overvåges, så flest mulig procesoplysninger nemt kan opsamles. Udviklingen af anden generations bioethanol produktion udfordres også af at kun få gær- eller bakteriearter i naturen er i stand til at producere ethanol effektivt ud fra alle de forskellige sukre, som findes i det lignocellulose biomasse.

Formålet med dette forskningsprojekt var at undersøge de muligheder og begrænsninger, som opstår ved brug monitoreringstekniker, der anvendes til at styre gæringer eller andre processer i et bioraffinaderi. Herigennem kan der muligvis opnås en forbedret processtyring eller reduceret lønomkostning i forbindelse med prøvetagning og analysearbejde. Muligheden for at anvende procesovervågning blev afprøvet ved at monitorer gærfermenteringer i realtid med Raman spektroskopi. De faktorer, som påvirker Ramanmålinger, såsom omfanget af dæmpning af lyssignal forårsaget af lysspredende partikler i reaktionsblandingen, blev målt og integreret i kvantificeringsmodellen. Herved blev det muligt at foretage løbende koncentrationsbestemmelser af ætanol, glucose og cellebiomasse i vækstmedie ved gæringer med *Saccharomyces cerevisiae*. Koncentrationsbestem-melserne blev gjort uden brug af kompliceret statistik med kemometri eller multivariate analysemetoder, der for det meste anvendes i forbindelse med andre spektroskopiske metoder, såsom NIR.

Fremgangsmåden til monitorering af en fermenteringsproces indeholdende rester fra biomasse i hydrolysatet gav overordnet set fornuftige koncentrationsmålinger, så længe mængden af partikler omkring sensoren kunne holdes på et relativt lavt niveau. Et forsøg på at overvåge en forbehandlingsreaktion syntes dog at give unøjagtige resultater, mens anvendelse Raman til overvågning af en hydrolyseproces på ingen vis forekom brugbar.

I en anden del af projektet fokuserede jeg på at arbejde med de to xylosefermenterende gærarter; *Scheffersomyces* (*Pichia*) *stipitis* og den senest isolerede *Spathaspora passalidarum*. De blev opformeret på hydrolysater fremstillet af tre biomasser, som forefindes i store mængder i troperne: sukkerrørbagasse, eukalyptus træfiber og restprodukter fra produktionen af palmeolie - empty fruit bunches (EFB). Ved sammenligning af de to gærarter fremkom lignende udbytter, men stofomsætningshastigheden for *S. passalidarum* var lidt bedre under de mikroaerobe forhold. Den højere mængde af eddikesyre, som fandtes i hydrolysatet fra EFB (6 g/l), syntes mest at inhibere *S. passalidarum*, hvilket resulterede i en længere nølefase. På trods af denne hæmning i starten endte denne fermentering med at give den højeste ætanolkoncentration på omkring 20 g/L. Under fuldstændige iltfrie forhold syntes ingen af de to gærarter at kunne overvinde den hæmmende virkning forårsaget af eddikesyren.

List of manuscripts included in the thesis:

Manuscript I
Quantitative Monitoring of Yeast Fermentation using Raman Spectroscopy
Jens A. Iversen, Rolf W. Berg, Birgitte K. Ahring
Accepted for publication in Analytical and Bioanalytical Chemistry

Manuscript II
Monitoring Lignocellulosic Bioethanol Production Processes using Raman Spectroscopy
Jens A. Iversen, Birgitte K. Ahring
Submitted for publication in Bioresource Technology

Manuscript III
Comparison between the pentose fermenting yeasts *Scheffersomyces stipitis* and *Spathaspora passalidarum* grown in hydrolysates of tropical biomasses
Jens A. Iversen, Birgitte K. Ahring
Intended for submission in Applied Microbiology and Biotechnology

Additional publications from project period, which is not part of the thesis

Production of ethanol from hemicellulose fraction of cocksfoot grass using *pichia stipitis* Sustainable Chemical Processes 2013, 1:13
Stephen Ikechukwu Njoku, Jens Asmus Iversen, Hinrich Uellendahl and Birgitte Kiær Ahring

Poster Presentation at Pacific Rim Summit, Vancouver, Canada, October 2012 Poster title:
Monitoring Yeast Fermentation Real-time Using Raman Spectroscopy with Sapphire Ball Probe

TABLE OF CONTENTS

INTRODUCTION

The current concerns about climate change have urged both governments and in the private sector to find viable solutions to reduce the green house gas emissions. The combustion of coal, oil and gas has caused carbon dioxide in the atmosphere to reach record levels at 400 ppmv in 2013 from originally 280 ppmv at preindustrial times – more than a 40 percent increase. The green house gasses, predominantly carbon dioxide, absorb infrared thermal radiation re-emitted from the planets surface when warmed by the visible sunlight, which easily passes through the atmosphere. Almost half the carbon dioxide released into the atmosphere by human activity is furthermore dissolved in the oceans, lakes and rivers of the planet, where it contributes to acidification. Decreasing pH reduces the concentration of available carbonate needed by calcifying organisms (1, 2). According to latest IPPC assessment report on climate change, it is more than 95 % certain that human influence has been the dominant cause of the global warming observed since the mid-20th century. The temperatures of the last three decades have probably been the warmest for 1,400 years, which calls for essential precautionary actions. Even if emissions would completely cease, the elevated temperatures will remain almost constant for centuries (3, 4).

Currently, the global consumption of energy arising from electricity production, the transportation sector and power needed for heating, is mainly covered by the use of fossil fuels. Carbon-free alternative sources of energy for electricity production such as; hydro-, nuclear-, solar-, wind power and through combustion of biomass or biogas are widespread. In contrast, it seems more difficult to replace the petrochemical fuels used for transportation. Only surpassed by emissions from power plants, the transportation sector is responsible for almost 25 percent of the annual global emissions of carbon dioxide (5). The use of electric vehicles, based on battery technology, may be promising; provided the needed electricity is produced without using fossil fuels. However, this would require an expensive modification of the current infrastructure. Ideally, any method to produce alternative carbon dioxide neutral liquid fuels requiring little or no modifications at the gas station, or of the engines in vehicles currently on our roads, would be highly desirable. In addition, dependency from foreign oil may be reduced in many countries, if petrochemicals can be replaced by domestically produced biofuels.

Energy derived from any sort of combustion of biomass or products hereof is generally regarded to be renewable and carbon neutral for the most part. The carbon dioxide emitted by the biofuel roughly corresponds to the amount assimilated by the biomass during the growth period, driven by photosynthesis with sunlight being the ultimate source of energy (6). Liquid biofuels such as methanol, bioethanol, biodiesel, ethers and biobutanol with properties similar to gasoline or diesel fuels

can be produced from agricultural residues, forestry waste and municipal solid waste through chemical or biochemical conversion processes. An abundant and sustainable alternative source of organic raw material to fossil fuels can be derived from biomass. Forest and agricultural residues alone constitutes an estimated energy resource of 30 EJ/year, and a considerable amount of the global primary energy demand of more than 400 EJ/year could be covered by utilizing dedicated crops in addition to the residues of different sorts (6).

Today, ethanol is the biofuel produced in largest amounts. Production is increasing rapidly with a global production of 17.25 billion liters per year in 2000, to projections exceeding 125 billion liters in 2020. This is the result of government programs around the world promoting the increased use of biofuels. Bioethanol is produced with the yeast *Saccharomyces cerevisiae* using mainly raw materials such as cornstarch in the US or sucrose extracted from sugarcane in places like Brazil (7). Development of 2nd generation biofuels using non-food lignocellulosic biomass for feedstock has gained much attention in recent years because of the controversial use of food for the production of fuel. In order to facilitate the production of bioethanol or other advanced biofuels using biomass as feedstock, research must be carried out to reduce process costs of develop new conversion technologies. Moreover, attempts to increase product yields or energy efficiencies by applying the latest from biotechnology or other fields of technology are also a priority (8).

BIOMASS STRUCTURE AND RESOURCE

A prerequisite for being able to develop efficient production processes for converting the constituted biomass components into biofuels is to obtain knowledge about the general structure of biomass **(Figure 1)**. The main component of plant biomass is lignocellulose, which is constructed as a complex matrix of three interlinked polymers within the cell wall; cellulose (35-50%), hemicellulose (20-35%) and lignin (10-30%) (9,10). The tensile strength of biomass fibers is mainly determined by cellulose and hemicellulose; while lignin provides rigidity to the structure. The abundance of each polymer depends on the type of biomass and varies between hardwoods, softwoods and grasses. Lignin content is highest in woods, whereas hemicellulose is found in higher amounts in grasses rather than in woody biomass. In addition, all plants also contain significant amounts of inorganic minerals (ash), proteins, lipids, soluble sugars and pectins, which is a common polysaccharide in fruits, but is also present in the cell wall of lignocellulosic biomass (11).

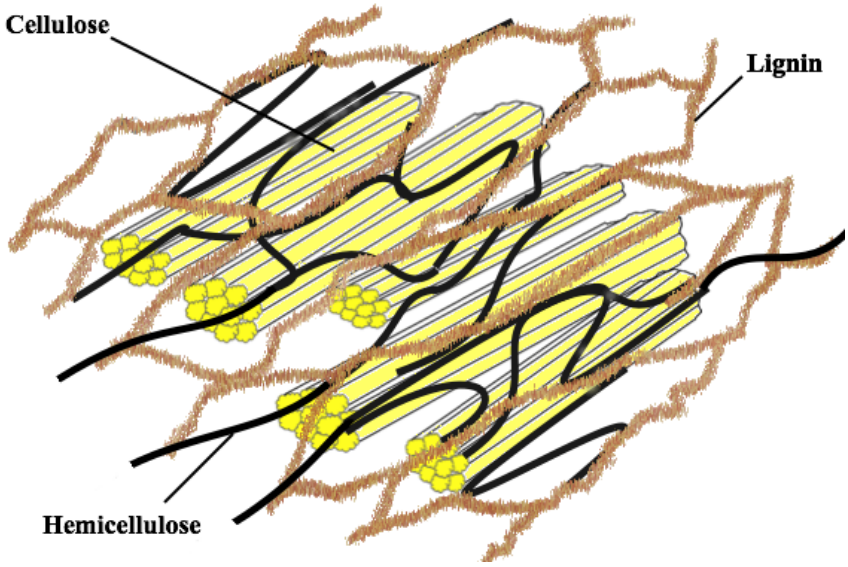

Figure 1: The structure of lignocellulose composed of the three biopolymers.

CELLULOSE

As the most abundant organic compound on Earth, cellulose represents a major potential feedstock for renewable fuels. Cellulose consists of homogeneous linear polymers of up to more than ten thousands of β-D-glucose units linked by β (1→4)-glycosidic bonds with cellobiose as the repeating unit in the glucan chain. Adjacent glucan chains are parallel to each other and bound together by hydrogen bounds and weak van der Wall's forces in hydrophilic micro fibril subunits of 24 to 36 chains. The rigid insoluble cellulose micro fibrils have both an ordered crystalline interior and a more disordered amorphous characteristic at the surface. The polymer is synthesized in the plasma membranes by the honeycomb-arrayed Rosette subunits. The micro-fibril structure, resembling a scaffold, gives the polymer great mechanical strength and recalcitrant properties towards degradation during thermochemical pretreatment or enzymatic hydrolysis (12, 13). By contrast, the highly branched starch (amylose and amylopectin) consisting of (1→4)-α-glucans also has crystalline features, but the transition to the amorphous state in water takes place at only 65 °C compared to 320 °C required for cellulose to undergo this transition in water (14).

HEMICELLULOSE

Hemicellulose, composed of both hexoses and predominantly pentoses, has a more heterogeneous and branched structure than cellulose. It consists of glucose, arabinose, mannose, galactose and rhamnose, but in most biomasses xylose is the predominant sugar monomer in these polymers consisting of up to 3000 sugar units **(Figure 2)**. Sometimes, glucuronic acid and galacturonic acid are also constituents of hemicellulose. These monomers are linked together with (1→4), (1→3) or (1→6) linkage by both β- and α-bonds and the polysaccharide binds non-covalently to the longer cellulose microfibrils. The main categories of hemicellulose are xylans, manans, xyloglucans and β-glucans, which are determined by the backbone of the polymer. The abundance varies among plant species with the xylan backbone (1→4-β linked D-xylose) being the most common polymer in hemicellulose (15):

- Softwood such as spruce and pine (gymnosperms): Galactoglucomannans, arabino-glucuronoxylans and arabinogalactan
- Harwood such as birch and aspen (angiosperms): Glucuronoxylans and glucomannans
- Grasses such as corn (herbaceous plants): Arabinoxylans

The branched polysaccharide pectin differs from hemicellulose mainly by having a backbone of linked α-(1 → 4) D-galacturonic acid, mostly, with a methyl ester group attached at the carboxylic group. Unlike cellulose and lignin, hemicellulose is relatively easy solubilized in hot dilute acids, thus enabling subsequently hydrolysis of the polymer to form single sugar monomers. However, prolonged thermochemical pretreatment can further degrade the released pentose sugars to furfural before the cellulose in the biomass has undergone any decomposition (11).

LIGNIN

Next to cellulose, lignin is the second most abundant biological polymer on earth. This complex branched polymer of phenyl-propanoid units are in contrast to the polysaccharide components in the cell wall hydrophobic and lack a primary structure. Covalently bound to hemicellulose, the brown amorphous lignin fills out the space between the other polymers and provides rigidity to the cell wall. As shown in **Figure 2**, the macromolecule is composed of a three dimensional network of the three aromatic monolignol monomers: p-coumaryl-, coniferyl- and sinapyl alcohol. These monomers, differing in the degree of methoxylation, are joined together by alkyl– aryl, alkyl–alkyl and aryl–aryl ether bonds. Herbaceous plants such as grasses contain all three units, whereas wood lignins mainly consist of guaiacyl and syringyl monomers (16). Biodegradation of lignin into the constituting monolignol units cannot simply occur through hydrolysis, but may involve reaction with radicals. Lignin is thus indigestible by animals, and it helps to shield the polysaccharides in the cell wall from enzyme attack (17). In the absence of an

oxidant, the polymer also resists chemical treatment well. The lignin degradation products can adsorb to enzymes and then inhibit further hydrolysis. Lignin is the initial source of fossil fuels, because of this ability to resist degradation (18, 19).

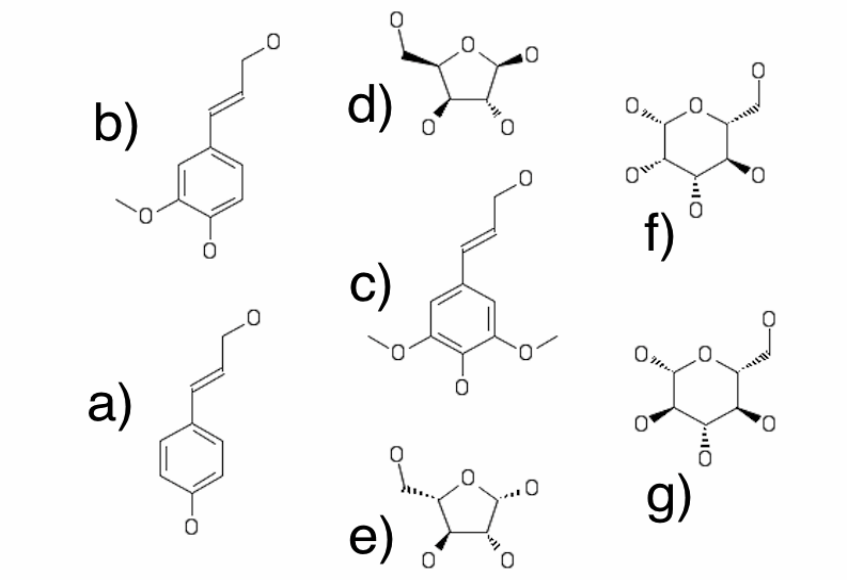

Figure 2: The seven major building blocks of the polymers in lignocellulose: a) *p* -hydroxyphenyl (H), (*p* -coumaryl alcohol), b) Guaiacyl (G), (coniferyl alcohol), c) Syringyl (S), (sinapyl alcohol), d) Xylose (ß-D-Xylofuransoe), e) Arabinose (ß-L-Arabinofuranose), f) Mannose (ß-D-Mannopyranose), g) Glucose (ß-D-glucopyranose).

BIOMASS FEEDSTOCK

Biomass is the largest, if not the only renewable source of carbon for several applications such as food, feed, materials, and it can also be used for chemicals, gaseous fuel, liquid fuel or solid fuel. Most often, the biomass feedstock originates from woody plants or herbaceous plants but the resource can also be derived from aquatic plants, manures or municipal wastes. Some of the general characteristics of ideal energy crops are:

- High yield per hectare
- Decease, pest and weed resistant
- Low energy input to produce and harvest
- Low cost to store and process
- Low nutrient requirements

In regards to the subsequent processing, additional properties of the feedstock may be important such as: moisture content, ash content, cellulose to lignin ratio (6, 20).

The available land for production of biomass on Earth is limited and getting the best out of the resource is a challenge, when considering obstacles such as; a future population of 9 billion people, loss of biodiversity, the controversial use of food for fuel, and soil depletion. Proteins, oils and to some extend starch produced by plants, require more resources than the production of cellulose, hemicellulose or lignin. Oils and proteins should therefore be left for food and feed production, whereas future production of bioenergy ideally should focus on using cellulose, hemicellulose or lignin.

The biomass needed for the production of 2nd generation biofuels may come from both dedicated energy crops or from lignocellulosic residues produced by existing agriculture and forestry. Technologies integrating food and fuel production could minimize the loss of energy and nutrients and assist in facilitating the best use of the available biomass resource. Exploiting marginal lands with limited potential for food production may also serve as a way to supply biorefineries with raw materials. Examples of all ready abundant lignocellulosic residues rendered by existing industrial, agricultural or silvicultural productions are:

- Sugarcane bagasse from sugar and bioethanol industry in especially South America
- Straw from grain and seed production such as corn or wheat
- Residues from palm oil production in South East Asia such as; leaves and empty fruit bunches
- Forest residues from wood and pulp production
- Municipal solid waste

These low value agricultural bi-products are already present in tremendous amounts. In fact some are already available at production facilities, reducing transportation costs. Alternatively, biomass for bioenergy, may be provided by dedicated crops such as:

- Herbaceous perennials grasses (seasonal plants): Switchgrass, *Miscanthus*, sweet sorghum and alfalfa.
- Dedicated production of woody biomasses from *Populus* (poplar), *Salix* (willow) eucalyptus
- Whole crop concepts combining traditional starch crop and their lignocellulosic residues

The perennial grasses, sugarcane, *miscantus*, switchgrass, soghum, *Arundo donax* (giant cane) and maize fixate carbon dioxide via a 4-carbon organic compound by utilizing the faster and more energy efficient C4 photosynthetic pathway compared to the pathway found in cereals and other C3-plants. These C4-plants are therefore potentially more interesting as feedstock for bioenergy purposes by giving greater biomass yields. However, in colder climate, C3-plants may be more productive and tolerant towards low-temperature conditions than C4 photosynthesis (6, 20, 21).

Considerations about sustainability are essential when implementing dedicated bioenergy crops on agricultural land otherwise used for the production of food or feed. There is a potential risk, especially in the tropical regions, that cultivation of bioenergy crops or even the increased production of common agricultural crops delivering residual biomass for biorefineries could threaten pristine forests and wild life. Sugarcane bagasse, eucalyptus tree fibers and empty fruit bunches or other residues from the palm oil production are examples of some biomasses with tremendous potential as raw material for the production of lignocellulosic biofuels in tropical regions, such as Brazil and Thailand (22, 23). The use of such biomass resources in sensitive areas potentially threating tropical forest should thus be approached with caution. **Manuscript III** deals with production of bioethanol using the xylose-fermenting yeasts *Scheffersomyces* (*Pichia*) *stipitis* and *Spathaspora passalidarum* and these three important tropical biomasses as raw material. *S. stipitis* has been able to grow and produce ethanol in hydrolysates from a wide variety of available lignocellulosic biomasses like corn stover or common grasses such as cocksfoot grass (24).

Feedstock is only one of several important features of a biorefinery system. The conversion pathway from the biomass resource to the final product is being determined by the process and platform. The platforms are intermediates from which the products are derived, which could be the C6 sugars or both C6 and C5 released during the pretreatment and hydrolysis processes. Integrating feedstock, products, platform and processes in the early development stages of a biorefinery concept are necessary to estimate the overall impact. An overview of some of these features and classifications of biorefinery systems is presented in **Table 1** (25).

BIOREFINERY

Utilization of sugarcane has been known for millennia, and the first sugar refinery was established in 1801 upon research conducted by A.S. Marggraf on the isolation of crystalline sugar from different beets and roots. Later, hydrolysis of starch was observed when potato starch was cooked in dilute acid, thereby converting starch into sugar. In 1835, J. J. Berzelius in Sweden discovered that hydrolysis of starch could also be catalyzed by enzymes. The process of wood saccharification was discovered in the early19th century; where glucose was released in low yields, as

wood was exposed to either dilute acid at high temperatures or to concentrated acid at lower temperatures. Significant ethanol production could be based on the fermentation of such sugar fractions using yeast in this early biotechnological process and particularly with starch as feedstock. Another use of these sugar feedstocks with fermentation was in the production of acetic acid. Processes to yield other products from lignocellulosic feedstock such as; furfural through distillation from decomposed hemicellulose, levulinic acid from cellulose, cellulose pulp and vanilin from lignin were also developed, but have not achieved the same significance as bioethanol production (26).

Biorefinery concepts have evolved from being based on sucrose from sugarcane or beet to the utilization of glucose or fructose released from starch grains exposed to enzymes. More recently, approaches to use lignocellulosic feedstock, considered to be more sustainable, have received more attention:

1. Sugarcane/ beet → sucrose
2. Starch grains + enzymes → glucose + fructose
3. Lignocellulose + enzymes → glucose and pentose sugar mixture + lignin

The difference in cost of the available fermentable carbohydrate produced through these three concepts has diminished over the last decades. This has partly been due to a dramatic decrease in the cost of the needed enzyme dose as a consequence of better enzymes performance and due to improved production economics implemented by the main industrial enzyme producers, Novozymes and Genencor. The cost of enzymes is now estimated to contribute to approach levels below 50 cent per gallon of lignocellulosic bioethanol produced (27).

Thermochemical decomposition processes include pyrolysis and gasification. Pyrolysis is performed in the absence of oxygen at 350 to 550 °C, possibly increasing up to 700 °C during reaction to produce liquids, gasses and char. In comparison, gasification occurs at 800 to 900 °C through partial oxidation of the biomass, which produces syngas, that consists of a gas mixture of carbon monoxide, carbon dioxide and hydrogen. Methane, methanol or larger hydrocarbons may catalytically be synthesized from syngas (28, 29).

Table 1: Biorefinery categories

Feedstocks	Platforms	Processes	Products
Starch crops	C6 sugars	**Thermochemical:**	**Energy:**
		Combustion	Bioethanol
Sugar crops	C5 sugars	Gasification	Biodiesel
		Hydrothermal upgrading	Biomethane
Lignocellulosic crops	Lipids/oil	Pyrolysis	Biobutanol
		Supercritical extraction	Methanol
Oil crops	Protein		Synthetic biofuels
		Biochemical:	Dimethyl ether
Grasses	Biogas	Fermentation	Electricity and heat
		Anaerobic digestion	
Marine biomass	Syngas	Aerobic digestion	**Material products:**
		Enzymatic processes	Food
Lignocellulosic residue	Hydrogen		Animal feed
		Chemical:	
Oil based residue	Organic acids	Catalytic	Fertilizer
		Pulping	Glycerin
Organic residue	Pyrolytic liquid	Esterification	Biomaterials
		Hydrogenation	Hydrogen
	Lignin	Hydrolysis	Polymers
		Electrolysis	Furans
	Electricty and Heat		
		Mechanical:	
		Extraction	
		Separation	
		Milling	
		Drying	

Biochemical conversion processes have some benefits over chemical conversion of biomass. These benefits include lower energy costs, less risk of catalytic poisoning, less microbial dependency on fixed substrate composition and higher specificity towards the substrate. Enzymatic hydrolysis technology applied during biochemical conversion of biomass can preserve the original carbohydrate structure if required, which is in contrast to fully thermochemical conversion causing decomposition of the sugar units (30). A vast number of organic acids or alcohols with biofuel potential for transportation can be produced in a biorefinery through biochemical conversion of the free carbohydrates released during pretreatment and hydrolysis. Among useful organic acids with potential as high volume platform chemicals produced by fermentation with wild type organisms or genetically modified organisms are: Acetic acid, oxalic acid, lactic acid, propionic acid, butyric acid, fumaric acid, malic acid, succinic acid, levulinic acid, ascorbic acid, citric acid and gluconic acid (31). Choice of conversion method and biofuel is dependent on

factors such as conversion yields and rates, sugar utilization, product and inhibitor tolerance, toxicity and fuel properties in combustion engines. The ability to control fermentation in large scale, to avoid microbial contamination in the fermentation process and to separate the product from the broth after fermentation may also be important requirements.

ANAEROBIC DIGESTION

The production of biogas through anaerobic digestion is the most simple biorefinery process employed to convert biomass into a biofuel having properties similar to a common fossil fuel. In the absence of oxygen, the organic components in biomass will gradually be degraded by a vast number of different microorganisms and will eventually be converted to a gas mixture consisting of mostly methane. The steps involved in the digestion are: hydrolysis, acidogenesis, acetogenesis and methanogenesis. Extra-cellular enzymes from hydrolytic bacteria initially decompose polymers such as polysaccharides, proteins and lipids down to smaller soluble monomers. Then, the available sugars, amino acids and long chain fatty acids are metabolized to intermediate alcohols, lactate, acetic acid and higher volatile fatty acids (VFA) by the fermentative bacteria during acidogenesis. The higher volatile fatty acids, lactate and alcohols, are oxidized by acetogenic bacteria to yield acetic acid, hydrogen, ammonia and carbon dioxide. Finally, methanogens convert these products to methane (typically about 70 %), and the remaining constituent of the biogas is carbon dioxide as well as trace amounts of hydrogen sulfide and hydrogen. Subsequent to the removal of hydrogen sulfide, the biogas can be stored and used for production of electricity. Further processing may be done to strip of the carbon dioxide and isolate the methane to yield fuel for transportation purposes if compressed. Liquid biofuels however seems easier to handle with current infrastructure than compressed biogas (32).

BIOBUTANOL

Another process considered to have potential for producing industrial biofuel is the ABE fermentation, which produces mainly butanol with *Clostridium* bacteria strains such as *C. pasteurianum*. Both hexoses and pentoses can be utilized and besides butanol significant amounts of ethanol and acetone are also produced giving the name ABE fermentation. The main disadvantages to this process are low yield, low productivity and the low toxicity of butanol to these strains. The poor butanol tolerance requires complicated removal and recovery of the product. Continued solvent removal during fermentation is possible, but has so far made the ABE process too complicated and expensive compared to the production of bioethanol (33).

BIOETHANOL

Production of 2nd generation bioethanol from lignocellulosic biomass is the biochemical conversion process that receives the most attention in a biofuel context. The following main process steps are generally involved in the various concepts of lignocellulosic bioethanol production: size reduction of biomass, thermo-chemical pretreatment, enzymatic hydrolysis, ethanol fermentation and recovery of ethanol. Detoxification, neutralization, separation of solids from liquids and residue processing are often also part of the overall process. The major bottleneck in the conversion of biomass into ethanol is to achieve efficient deconstruction to simple sugars during pretreatment and hydrolysis due to the recalcitrant nature of plant cell walls. Optimizations of sugar yields have been carried out to a great extent by improving enzyme performance and reducing the enzyme production costs. Furthermore, minor changes of pretreatment parameters can easily result in increased formation of inhibitors, affecting the fermentation process and thus the overall performance. Only a limited number of bacteria, yeasts and fungi are able through fermentation to also convert the pentose sugars from the hemicellulose fraction into ethanol. Currently, most concepts heading for commercialization employ a platform based on fermenting only the hexose sugars by using the baker's yeast, *Saccharomyces cerevisiae* (**Figure 3**).

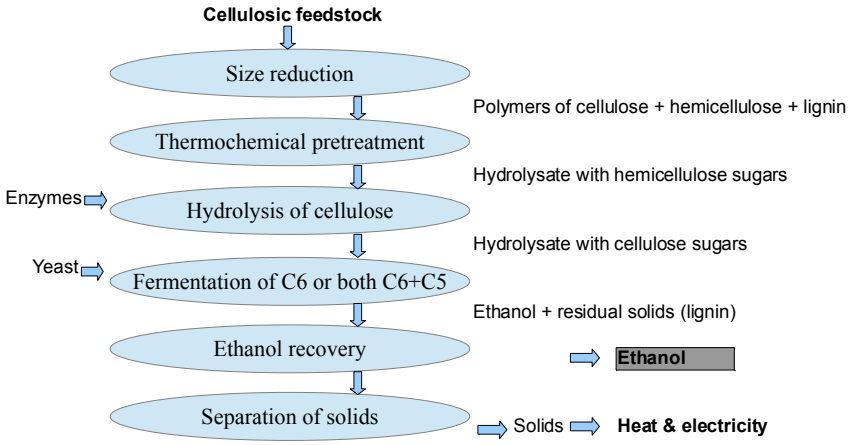

Figure 3: General lignocellulosic bioethanol production process.

Models have shown that future cellulosic biorefineries concepts have the potential to achieve efficiencies comparable to what is reached with the current fossil based production of fuel for transportation. Biochemical production costs that are competitive with gasoline (in terms of cost per gasoline equivalent), may be realized at oil prices as low as $30 per barrel, provided that the development of

effective processes to overcome lignocellulosic recalcitrance are executed for production scale. For instance, merging the enzyme production, hydrolysis and fermentation steps into one single theoretical possible unit process could optimize the production of bioethanol. Biological ethanol production seems a bit more promising than thermochemical processing, but integration of both with lignin-rich residues being converted thermochemically appears to be a scenario with an even better performance. Biorefineries with coproduction of fuel and protein for animal feed, as well as, significant integration of processes to recover heat and water are also assessed to be necessary to achieve economic feasibility. High capital costs are associated with advanced biofuels, compared to starch ethanol plants, which with the current technology and corn prices are still more cost effective (34, 35).

PRETREATMENT

Pretreatment is the first step in deconstructing the rigid structure of the lignocellulose during the production of lignocellulosic bioethanol. An initial mechanical grinding or milling of the biomass usually takes place before pretreatment to increases the surface area and thus facilitate the chemical and enzymatic degradation processes. This mechanical size reduction however adds to the total energy consumption. The subsequent pretreatment reduces crystallinity and increases the porosity of cellulose, thereby contributing additionally to increasing the access of the enzymes to the polymers, hence for a more effective enzymatic hydrolysis of the biomass. Several pretreatment methods have been developed including physical, chemical, biological, and electrical or combinations of these (36, 37).

The most common pretreatment methods require conditions with high temperature and pressure; applying dilute acid, steam explosion, ammonium fiber expansion, wet oxidation, or wet explosion of the biomass. Hemicellulose acetyl groups are cleaved off at the elevated temperatures during pretreatment and the hydrolysis of the polymer to monosaccharides is further aided by acetic acid acting as catalyst on the reaction. Lignin is not easily removed, but sort of melts and is redistributed on the fiber surface from depolymerization and repolymerization reactions, thus enhancing fiber digestibility (38). Higher temperatures are generally needed when more dilute sulfuric acid or other catalyst is used. Pretreatments with dilute sulfuric acid typically release most of the sugars from the hemicellulose in the biomass within 20 min at temperatures above 150 °C. Release yields of the fermentable hemicellulose sugars above 90% of the theoretical possible have been achieved with many of these methods. The process should be optimized to ensure high sugar yields, little formation of fermentation inhibitors and minimal use of energy and chemicals (39).

HYDROLYSIS

The enzymatic digestibility of lignocellulosic biomass or recalcitrance is determined by enzymatic accessibility. The dilute acid pretreatment of lignocellulosic biomass exposes the cellulose by releasing most of the hemicellulose from the biomass. Moreover, the pretreatment hydrolyses most of the hemicellulose into monomers. Acid hydrolysis of the remaining cellulose to glucose is also possible if more severe conditions regarding temperature and acid concentration is applied, but can result in the formation of fermentation inhibitors such as furfural. As a consequence, the hydrolysis of the remaining cellulose polymers is in practice only completed when utilizing cellulytic enzymes. The cooperative action of three types of enzymes is necessary for complete degradation of cellulose into glucose. Endo-ß-glucanases (EG) randomly cleaves bonds in the amorphous parts of the cellulose chain, yielding shorter oligosaccharides. Exo-ß-glucanases, also named cellobiohydrolases (CBHs), cleaves the disaccharide cellobiose off from each end of the cellulose chain, with CBHIs cleaving at the reducing ends and CBHIIs working at the non-reducing ends. Together, with some of the free oligomers cellobiose is finally hydrolyzed down to single glucose monomers by ß–glucosidase (13, 38). Any of these enzymes may exhibit product inhibition, but the overall hydrolysis rate is in particular dependent on the activity of ß–glucosidase - being the last acting enzyme in a chain of cellulose hydrolysis reactions (40).

Strains of filamentous fungi, *Hypocrea jecorina* (*Trichoderma reesei*) or *Humicola insolens* are among the best known producers of these hydrolytic enzymes for industrial applications, in which the optimal hydrolytic activity with enzyme preparations are obtained at 55°C and pH 4.5 (41). More recently, a novel class of fungal glycoside hydrolases (GH61) have been shown to be capable of degrading cellulose by oxidative cleavage of the polysaccharide in its crystalline state, which in combination with the classical cellulases may enhance the enzymatic conversion of biomass (42, 43).

FERMENTATION

In contrast to conventional starch or sugar-based fermentations, a more severe pretreatment prior to enzymatic hydrolysis and fermentation is necessary, when using lignocellulosic biomass as feedstock, because of the more recalcitrant structure. Fermentation and distillation technologies on the other hand would need less modification when upgrading the current bioethanol production processes to use lignocellulose instead of starch or sugar. The yeast *Saccharomyces cerevisiae* has for centuries been used to produce ethanol for consumption. It can grow on simple sugars and disaccharides, but not on pentose sugars. Most production concepts currently heading for commercialization of lignocellulosic bioethanol use *S. cerevisiae*, which is well adapted to industrial processes and have been successfully tested at 2nd generation bioethanol demonstration plants. The main

advantages are an exceptional high ethanol yield close to the theoretical limit of 0.51 g ethanol/ g sugar and high tolerances towards inhibitors and ethanol itself of typically 14 % or even as high as 18 % (w/v) of some of the industrial strains of this yeast. The ethanol tolerance is however significantly lower in the presence of the inhibitors found in pretreated lignocellulosic biomass. A relative fast growth rate also contributes to the ability to avoid contamination by fast growing external microorganisms in the fermentation process, otherwise threatening to outgrow the yeast culture and thereby decreasing ethanol yields. It is not practically possible to perform steam sterilization of the major parts of a biorefinery in the event of a bacterial contamination, since the low value products cannot justify the costly use of high-pressure fermentation vessels that are capable of withstanding stream sterilization. Instead, a concept must be applied where bacteria cannot easily outgrow a robust production organism such as *S. cerevisiae* (50, 74).

As a consequence of the cellulase product inhibition described above, enzymatic hydrolysis of pretreated lignocellulosic biomass is rather slow, compared to hydrolysis of starch during the conventional ethanol production. The concept of simultaneous saccharification and fermentation (SSF) aims to reduce product inhibition by combing hydrolysis and fermentation in a single step, whereby ethanol fermentation maintains glucose concentrations at lower levels. In spite of the drawback presented when the hydrolysis temperature has to be reduced from the optimal 45-50 °C down to below 35 °C, where the yeast can exist, SSF benefits from an overall increased cellulose to ethanol conversion rate and from eliminating the additional reactor needed for the configuration with the original separated hydrolysis and fermentation process (SHF). The promising hybrid hydrolysis and fermentation (HHF) process is a modified SSF where a short saccharification step at optimal hydrolysis temperature initiates SSF (44).

XYLOSE-FERMENTING YEASTS

S. cerevisiae is an excellent ethanol producer when it comes to using glucose from starch, but it is limited to hexose sugars; however, the main component in hemicellulose, xylose, therefore cannot naturally be metabolized. A limited number of bacteria, yeasts and fungi are able through fermentation to convert the pentose sugars from the hemicellulose fraction to ethanol (74). Amongst the naturally xylose fermenting yeasts are *Pachysolen tannophilus, Candida shehateae*, *Scheffersomyces stipitis* (previously known as *Pichia stipitis*), *Kluyveromyces marxianus* (75, 77) and recently *Spathaspora passalidarum*; these yeasts have been shown to ferment xylose efficiently (76). The fungus *Mucor inducus* is also well known for its xylose fermenting abilities. So far, *S. stipitis* has been considered to be one the best xylose fermenting yeasts for industrial application because of its ability to ferment the pentose sugars from hemicellulose efficiently. Generally, this yeast is capable of fermenting both pentose and hexose carbohydrates into ethanol, as well as, metabolizing sugars such as cellobiose and even to hydrolyze xylan (78).

Unfortunately, *S. stipitis* generally has a lower ethanol tolerance, lower yields and a slower sugar consumption rate compared to *S. cerevisiae*. The purpose of this thesis is to assess the industrial potential and limitations of using natural xylose-fermenting yeasts as biocatalyst for the production of lignocellulosic bioethanol. A comparison between the performance of *Scheffersomyces stipitis* and *Spathaspora passalidarum* fermenting pretreated biomass is presented in **Manuscript III**.

Attempts to genetically improve the wild type xylose fermenting yeast *Scheffersomyces stipitis* for bioethanol production have been carried out through numerous cloning experiments (45, 75). Alternatively, more gradual strain improvements may be achieved through the application of classical chemical mutagenesis, techniques to optimize parameters, such as; the range of sugars consumed, the temperature span or the tolerance towards inhibitors and ethanol.

Scheffersomyces stipitis and the other xylose fermenting yeasts cannot completely metabolize xylose to ethanol under strict anaerobic conditions. Microaerophilic conditions must be provided during fermentation in order to maintain NADH balance in the cell. The intermediate xylitol will accumulate without any aeration hence terminating the uptake of xylose (79). In contrast, xylose metabolic pathways in bacteria using xylose isomerase do not need oxygen for the regeneration of cofactors, as illustrated in **Figure 4** (46). Another pentose-fermenting yeast with some interesting beneficial traits in regards to lignocellulosic bioethanol production is the thermotolerant *Kluyveromyces marxianus,* which is considered optimal in SSF processes due to the capability to ferment and grow well at temperatures above 40 °C. However, ethanol yields with *K. marxianus* are unfortunately not optimal (47-49).

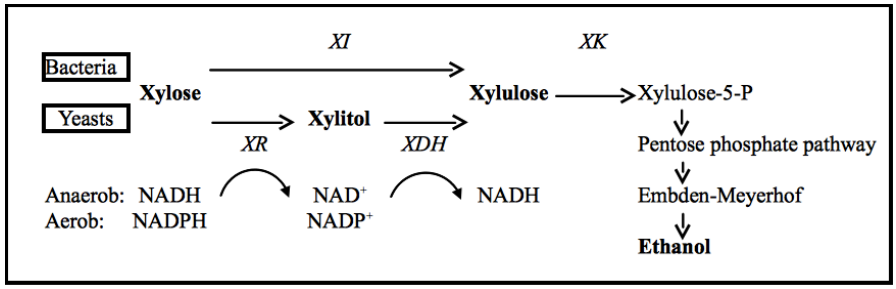

Figure 4: Redox balance during xylose fermentation in yeast and bacteria. *XI* - xylose isomerase, *XR* - xylose reductase, *XDH* - xylitol dehydrogenase, *XK* – xylulokinase.

In this context, many research projects have focused on creating strains that can convert all carbohydrates in lignocellulosic biomass to ethanol by other means than mutagenesis. To achieve this, genes involved in the xylose metabolism in fungi or other yeasts have been inserted into *S. cerevisiae* through genetic engineering. To avoid the need for strictly controlled aeration required with the yeast xylose pathway, some groups have attempted to introduce xylose isomerase genes from other organisms into the genome of *S. cerevisiae*. In the presence of xylose isomerase enzyme expressed by genes inserted from bacteria or other fungi, *S. cerevisiae* has produced ethanol from xylose (75, 77, 80). Genetically, metabolic engineered strains with the xylose isomerase genes inserted into the genome of the yeast, have produced promising results for further scale-up trails in pilot plants and possibly commercialization.

XYLOSE-FERMENTING BACTERIA

Other research efforts have been focused on developing recombinant ethanologenic bacterial strains capable of producing ethanol when fermenting both C-5 and C-6 sugars. Among these are the Gram-negative bacteria *Echerichia coli*, *Zymomonas mobilis and Klebsiella oxytoca.* In contrast to the Embden-Meyerhof (EM) pathway in many microorganisms, *Z. mobilis* metabolizes glucose through the Entner-Doudoroff (ED) pathway producing less biomass, but more fermentation product than *S. cerevisiae* as a consequence of a lower ATP yield. The wild-type bacterium has a narrow substrate range, pentoses not included, and the integration of genes responsible for the assimilation of arabinose and xylose has successfully improved the fermentation capabilities. However, *Z. mobilis* is not considered to be as robust as *S. cerevisiae.* The well-characterized *E. coli* can naturally grow on a wide range of sugars and recombinant strains have been created to produce ethanol, but the fermentation process is still limited by unfortunate pH requirements (6.0-8.0). In addition to the difficulties associated with hydrolysis and fermentation having different optimal pH range, the production of bi-products during fermentation, such as acetate or lactate, is another disadvantage often encountered with natural bacterial ethanol producers (50). Some thermophilic anaerobic bacteria with high ethanol tolerance and the ability to ferment both pentose and hexose sugars efficiently have also drawn some attention as potential ethanol producers. Cultures with recombinant strains of *thermoanaerobacter* species with genetically interrupted lactic acid and acetic acid pathways have shown able to ferment most of the sugars in lignocellulosic hydrolysate to ethanol at high conversion yields and productivities. Recombinant strains of *thermoanaerobacter* with terminated production of bi-products are being tested in pilot scale (51, 52).

INHIBITORS

Lignocellulosic ethanol production is challenged by the wide range of compounds formed at the elevated temperatures during the pretreatment of the biomass, which may inhibit fermentation. The effect of such inhibitors thus has to be included in a

process evaluation. Among inhibitors present in hydrolysate are aliphatic acids such as: acetic, formic and levulinic acids, aromatic phenolics and the furaldehydes, 5-hydroxymethylfurfural (HMF) and furfural. Acetic acid and furfural are considered to be the major inhibitors of the fermentation. Below certain concentrations, the degradation of both acetic acid and furfural by the yeasts is possible, but at a cost of ATP, consequently resulting in extended lag phase, lower productivities, lower ethanol yields or cell growth (53).

In summary, evaluation of certain desirable microbial features is inevitable, whether fungi or bacterial strains are considered as fermentation biocatalyst for lignocellulosic bioethanol production. Among these are:

- Conversion yield
- Ethanol tolerance
- Tolerance to salt and the inhibitors in hydrolysates
- Sugar consumption rate
- GRAS Status (Generally Regarded As Safe)
- Substrate utilization range
- Growth rate
- Low fermentation pH to avoid contamination
- Volumetric productivity
- Minimal nutrient requirements
- Compatibility with SSF
- Cellulase producer

Therefore, the complexity of assessing a process involving so many factors leaves plenty of room for future optimization of fermentation processes considered for the production of bioethanol or any other microbial produced biofuel for that matter (50).

MONITORING

The success of future biorefinery concepts is dependent upon development of economical feasible production processes. Any measures to reduce formation of undesirable byproducts or to decrease energy consumption is therefore crucial in order to be able to implement production of biofuels on a vast scale. Acquiring detailed knowledge about the different biochemical conversion reactions is an essential basis for improvements. Utilizing effective analysis methods may aid achieving a thorough understanding of the processes involved and enable optimal control of the critical process parameters. During biochemical conversion processes, gas chromatography (GC) and High Pressure Liquid Chromatography (HPLC) are among the analysis methods most commonly used for measuring the

components of interest in a biorefinery such as sugars and alcohols. Although reliable and accurate results are expected, a time delay must be considered when applying these well established methods, since sampling, transport of sample, handling in the laboratory and delivery of result to the production line may take hours, or even days. This delay time complicates active process control, and such off-line analysis serves more as tool to guarantee quality parameters for the subsequent process step, rather than a monitoring method allowing active adjustments and feedback for ongoing processes. Production processes can be operated more efficiently or even automated if production parameters can be monitored real-time instead of retrieving vital process information through manual time-consuming sampling and subsequent off-line analysis, despite being designed to account for time delay (54).

PROCESS ANALYTICAL TECHNOLOGIES

Labor intensive processing of samples done off-line in a laboratory may be reduced when using Process Analytical Technologies (PAT), which accommodates analysis right next to the process by monitoring in-line or even on-line by directly measuring in the processes. Various on-line configurations are available, including direct *in situ* sensor insertion, as well as, *in situ* and *ex situ* sampling loops with in-line analyzers. The American Food and Drug Administration (FDA) now recommends applying tools for on-line monitoring to improve production consistency and quality in pharmaceutical manufacturing. PAT may also be applied in research to optimize future production processes. In particular, fast reactions or processes being dependent on many process parameters may benefit from rapid analysis and fast feedback to enable a dynamic production, where variations are constantly adjusted and compensated for in order to assure high quality by avoiding the process steps getting off track. The speed of analysis and the elimination of errors from manual sample processing such as separation of sample components, provide more true quantitative and qualitative information about the process than derived from an external sample, but only if the performance of the monitoring method is robust. Censor technology, representative measurements of the relevant process parameters and data analysis are the three basic elements of a successful PAT approach. Implementation may therefore long term serve to improve use of resources through optimal monitoring and control of production, or even serve to facilitate continuous processes and automation tools. This contributes on-line information to be received directly assuring valuable short response times. Ideally, monitoring requirements for each process parameter applied in a bioreactor control system are met with small and stable in-line sensors (55).

The new concepts for the production of lignocellulosic bioethanol are still under development and the analytical methods required may similar to food and pharmaceutical processes benefit from utilizing effective process monitoring methods based on PAT strategy. Therefore, methods to perform *in situ*

measurements of the chemical conversions during the three main processes, pretreatment, hydrolysis and fermentation are of great interest. The major challenge is to apply reliable and robust censor technologies able to withstand these environments, while delivering measurement results of adequate qualities, in order to be able to replace existing techniques used to analyze samples subsequently. In bioreactors, requirements for industrial applicable probes have been estimated to include:

- Designed with inert materials not contaminating process
- Endure high temperature and pressure, possibly to the point of being fully sterilizable (SIP)
- Robust in corrosive environments with high salt or even acid concentrations
- Adequate sensitivity and resolution for monitoring requirement
- Easy to perform and maintain calibration
- Linear dependency or other well defined correlation
- Noise insensitive

REAL-TIME MEASUREMENT IN BIOREFINERY

Basically every process step of a biorefinery is expected to contain high amounts of particulates from biomass debris. Applying immersed censors in such heterogeneous liquids poses the major obstacle for implementing *in situ* measurement methods. Approaches to overcome fouling of probes or signal disturbance caused by solids must then be considered when monitoring in these conditions. Another issue, likely to disturb during on-line measurements, are the diverse range of compounds not converted during the heterogeneous liquids of the various process steps. Complex background signals of little or no value originating from molecules or matter present in the different process mixtures, but not being processed such as lignin, inert carbohydrates, dissolved inorganic minerals, or perhaps biomass debris must be taken into account when evaluating *in situ* analysis methods. In summary, suitable analytical measurement techniques for on-line monitoring should be able to withstand rough conditions, maintain calibration, handle background noise and the censors should not be prone to settling of particulates causing unreliable measurements.

Several process parameters are of significance when monitoring and controlling bioethanol production. Temperature, pH and dissolved oxygen (DO) were some of the first successfully monitored parameters in the chemical and pharmaceutical industries applying *in situ* sensors inside the occasionally harsh environments of the reaction vessels. Monitoring these parameters using existing standard equipment during the pretreatment, hydrolysis or fermentation processes of lignocellulosic

bioethanol production is therefore considered to be of rather trivial matter where only instrument calibration may prove difficult.

Additional analysis of other important process indicators is however estimated to be more complicated to perform on-line when considering the common analysis methods currently applied for in similar bioprocesses. Information about concentrations of substrates, products, or inhibitors such as acetic acid, furfural or HMF is of major interest during the pretreatment, hydrolysis and fermentation reactions. Chromatography analysis methods with integrated pre-separation of the samples such as HPLC or GC are currently often used to distinguish these components from each other when quantifying concentrations. Sampling and subsequent analysis appears to be necessary when monitoring these concentrations, but utilization of the rapidly evolving optical sensor techniques may ideally enable direct simultaneous monitoring of substrate and product without prior separation (56, 57).

ELECTROCHEMICAL SENSOR

Some of the first technologies available to monitor fermentation processes real-time were based on electrochemical sensors. An electrochemical censor transforms chemical information into a useful signal for analysis. Probes to measure pH or dissolved oxygen are some of the most common electrochemical censors used for monitoring bioprocesses. The censor is basically an electrochemical cell consisting of two or three electrodes, a sensing electrode and a counter electrode separated by electrolyte, wherein there sometimes is a reference electrode to provide a constant potential to the censing electrode. The analyte diffuses over a selective permeable barrier, for instance a hydrophobic membrane, into the electrolyte generating an electrical signal in the form of voltage, current or resistance. The overall performance is determined by selectivity, noise sensitivity, limit and range of detection, response time, and lifetime. Chemosensors are robust but lack specificity compared to biosensors (56, 57).

DIELECTRIC SPECTROSCOPY

As an alternative to measuring optical density (OD) real-time monitoring of microbial cell density is also possible applying a method based on radio-frequency impedance. This method is also referred to as dielectric spectroscopy or simply capacitance. The concentration of viable cells in a fermentation culture can be estimated by utilizing measurements of capacitance and conductance. The principle behind this has been known for 20 years and capacitance is now used to control mammalian cell cultures. In an electric field, ions cannot freely move across the nonconductive cytoplasma membrane of a cell resulting in buildup of charge or induced polarization. The polarized living cells act as capacitors in the culture medium, in contrast to dead cells with a damaged cytoplasma membrane not hindering the movement of charge. **Figure 5** shows a dielectric immersion probe

and illustrates a polarized cell. The amount of living cells or total cell volume is proportional to the capacitance measurable in an alternating electrical field at certain high frequencies. In addition to this, the shape of a dielectric spectrum of the capacitance over a range of frequencies reflects characteristics about cell morphology, size and type (58, 59). At lower frequencies, cells have enough time to become polarized (maximal cell polarization), whereas fewer ions can charge the cells within the shorter charging time at higher frequencies (minimal cell polarization) and the remaining background capacitance is then mostly due to dipoles of water and other molecules in the culture growth medium (60).

Figure 5. Annular type dielectric probe and polarized cell in an electrical field (right corner).

In contrast to biopharmaceutical production, biofuel fermentation cultures contain considerable amounts of suspended solids and dissolved minerals potentially disturbing measurement. Increasing ion concentrations and hence conductivities, give rise to higher biological capacitance and corrections might therefore be needed if ionic strength does not remain constant (58). The technology is challenged by issues regarding calibrations and polarization problems at low cell density in growth media with high conductivities, which have been encountered with commercial probes (59). Considerable capacitance background is expected to

originate from the fiber fraction in the pretreated biomass, compared to the signal generated by the fermentation culture itself. The potential use of dielectric capacitance for monitoring the growth of the microbial cell cultures in biorefineries, may be limited, if the signal to noise ratio will be too large, due to much background signal caused by suspended solids present in the fermentation broth.

BIOSENSORS AND IMMOBILIZED ENZYME ELECTRODES

Conventional ion-selective electrodes measuring pH or redox potentials have been based on detecting the level of only inorganic anions or cations. More recent electrochemical procedures using more advanced enzyme electrodes also enable measurements of the organic compounds. The electrode basically consists of a receptor able to selectively respond to certain molecules, ions and a transducer, which continuously converts the non-electric process information into an electrical signal. The organic molecule to be assayed can simply be censed by selective enzymes immobilized onto an ion-selective electrode if the enzyme reaction involves increasing or decreasing electroactive compounds. A transducer thus continuously converts the non-electric process information into an electrical signal from changing concentration of ions. Instead of enzymes, selective antigens may also be employed in these advanced types of electrodes.

In regards to biofuel production, utilizing commercially available electrodes with a gel containing glucose oxidase is a useful application of this technique to detect glucose at low levels. Carbon dioxide, acetic acid or other relevant sugars such as xylose may as well be monitored using this principle. However, enzyme electrodes seem delicate and sensitive to temperature and pH changes compared to robust optical sensors designed to withstand cleaning and sterilization in place. In addition, measuring multiple compounds simultaneously in a viscous mixture containing particulates with only one optical sensor seems less complicated and costly than using several enzyme electrodes for each compound investigated, or applying one specially designed electrode containing all the needed enzymes while expected to maintain catalytic activity. Superior precision and specificity of these electrodes may in fact be of limited benefit, if the membrane or diffusion gel layer of the enzyme electrode also proves to be more prone to irreversible fouling than the lens surfaces of optical immersion probes. As a consequence, these bioselective sensors are probably most useful for off-line analysis (56, 57).

OPTICAL PROBES AND MONITORING

Spectroscopy conducted with in-line sensors is rapid, less expensive, non-destructive, straightforward and sometimes more accurate than analysis performed with conventional electrodes. Electromagnetic radiation can interact with matter through the mechanism of reflection, absorption or scattering. All energy of a photons just bouncing back from the matter is preserved during reflection, whereas

the energy of the photon is completely transferred to the matter through absorption. The photon is partially or completely reemitted in the case of the scattering process.

Most optical sensors rely on the analytical principles of spectroscopy based on UV/VIS, IR (NIR and FTIR), Raman scattering and fluorescence light undergoing transmission, absorbance or backscattering of the photons. A major parameter affecting the fermentation conversion rate hence performance of the fermentation process, is cell concentration. Optical density (OD), relying on reduced transmission of light due to light scattering, is an easy and widely used optical method to estimate microbial biomass. This traditional method of measuring cell biomass often at 600 nm in off-line samples is now possible *in situ* with the application of new immersed biosensors, as long as cells are the only light scattering particulates present in the suspension. Without dilution, polynomial approximation may be needed at higher cell concentration to account for the shadow effect, which disrupts the linear relationship between concentration and reduction of the transmitted signal. Interference from particulates is less likely in a clean pharmaceutical cell cultivation process, whereas the amount of debris from plant biomass inevitably will remain in the fermentation broth of both 1st and in particular 2nd generation bioethanol production will complicate direct measurements of OD in the slurry. In contrast to OD measurements merely relying on the scattering of light by the cells, spectroscopic techniques may yield more detailed information about components than just the microbial cell concentration (57).

FLUOROMETRY

Fluorescence-active analytes possessing distinct spectra of excitation and emission wavelengths allow measurements and reaction monitoring to be conducted with fluorescence sensors. For instance, estimation of viable cells is based on the amount of nicotinamide adenine dinucleotide phosphate NAD(P)H measured using fluorescence spectroscopy. The cell culture is exposed to UV excitation light, to generate measurable emission light at another wavelength, useful for quantifying culture cells if NAD(P)H amount per cell remains constant and interference from other components also acting as fluorophores is minimal. Proteins, coenzymes and vitamins can also be measured directly as fluorophores. Moreover, fluorescent dyes may be added to cultures to specifically sense other parameters such as oxygen, pH and carbon dioxide (57, 59). Some interference hurdles can be overcome with multi-wavelength techniques. The potential use of this method in a biorefinery may prove difficult, because the high amount of lignin with its aromatic structure is strongly fluorescent.

INFRARED SPECTROSCOPY

Absorption, emission and reflection spectra in the field of infrared spectroscopy all take advantage of the interaction between electromagnetic radiation and the molecular transition from one state of vibrational or rotational energy to another. In contrast, absorption within the ultraviolet-visible spectral range originates from the transition from one orbital state of electrons to another within molecules. In the electromagnetic spectrum infrared light is located in between visible light and microwaves, so IR wavelengths range from about 0.8 to 1000 μm, which is equivalent to a range from 12,500 to 10 cm^{-1}, when using the more convenient wavenumber scale - proportional to frequency. IR is usually further divided into near infrared (NIR) ranging from about 12,500 to 4000 cm^{-1}, mid infrared (MIR) ranging from 4000 to 200 cm^{-1} and far infrared (FIR) ranging from 200 to 10 cm^{-1}. Most useful molecular details are revealed in parts of the MIR extending from 700 – 1200 cm^{-1} (61, 62).

The oscillating electrical field of a photon interacts with the dipole moments of the bonds in a molecule. The dipole moment depends on the charge magnitude and charge distance. For IR absorption to occur, this dipole moment must undergo a net change, at the radiation frequency, that matches some of the vibrational frequencies of the molecule. The vibrational energy state of a bond simplified resembles a mechanical spring oscillating harmonically with a frequency dependent on mass, distance and charge differences. A certain electromagnetic energy quantum is required to change a discrete natural vibrational energy state of a molecule to another. The energy quantum of a photon can sometimes change energy states of two bonds simultaneously rather than only one, thereby creating combination bands, complicating spectra for structural analysis. As a consequence of vibrational oscillations exhibiting some disharmonic characteristics, the combined energy transitions may become twice, three or four times larger than the fundamental vibrational energy of a bond, thereby causing overtone bands and also combination bands hereof to appear, mainly in the NIR region. The occurrence of overtone bands is less likely to happen than the fundamental transition, and overtone intensities are thus generally lower. Water generates fundamental O-H stretching and bending vibrational band around 1900 nm, first overtone O-H stretch band is seen at 1450 nm and second overtone O-H stretch band is displayed around 970 nm in a NIR spectra (54).

Spectra vibrations in a molecule containing more than two atoms can appear basically as a stretching or bending motion due to a change of the interatomic distance or a change in the angel between two bonds respectively. In addition, bending vibrations fall into the categories of twisting, wagging, rocking and scissoring. All six fundamental modes of vibration are displayed in **Figure 6**. Larger molecules with more vibration points, than between neighboring atoms, have more complicated types of vibration. The number of possible vibrations is

determined by degrees of freedom. A linear molecule with N atoms has 3N-5 vibrational modes, whereas a nonlinear molecule has 3N-6 vibrations.

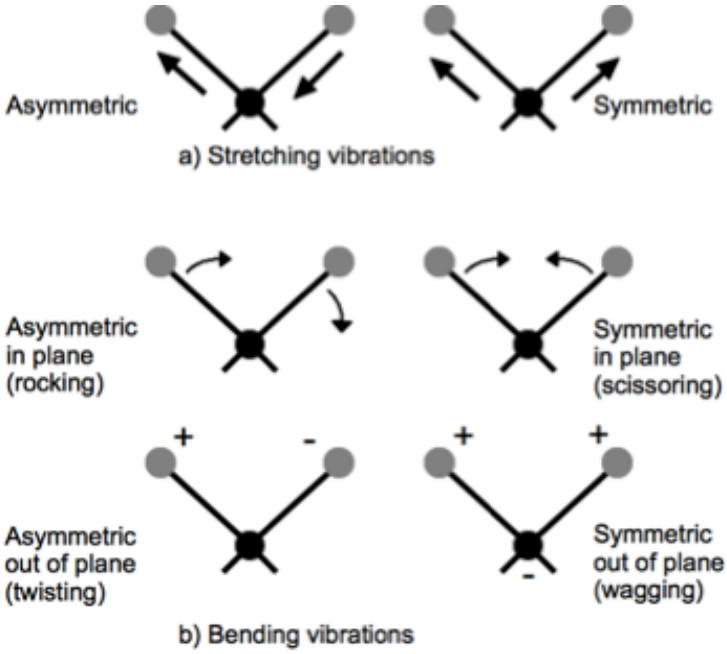

Figure 6: Modes of molecular vibrations, with motion towards (+) or away (-) from the reader.

The spectra displaying these various types of vibrational transitions can be utilized to retrieve information about molecular compounds when irradiating gas mixtures, solutions or solid samples with light at different wavelengths. In the case of the different methods of absorption spectroscopy (NIR and MIR), spectra are generated as a result of the complete absorption of the photons, transferring all electromagnetic energy to the excited molecules. Some incoming photons may however only undergo partial transfer of energy to a molecule or even from an already excited molecule and leave the molecule again having a different wavelength than before the interaction. This inelastic scattering also referred to as Raman scattering form the basis for another branch of analytical spectroscopy with characteristics different from absorption techniques (55, 61).

Altogether, vibrational spectroscopy is a method of chemical analysis where the sample is illuminated with incident radiation in order to excite molecular vibrations. Vibrational excitation occurs when the molecule absorbs, reflects or scatters a

particular discrete amount of energy. There are two major types of vibrational spectroscopy: Infrared (IR) and Raman.

ABSORPTION SPECTROSCOPY

IR absorption spectra are generally acquired through means of transmission, reflectance or techniques hereof. In transmission, a full spectra light beam passes through a transparent sample and absorption (A) or transmission (T) values at the various wavelengths are derived from the light intensities of the incident light (I_0) and of the light intensity remaining after passing through the sample (I):

$$A = \log (1/T) = \log (I_0 /I) = \varepsilon \cdot c \cdot l$$

Expressed by the Lambert-Beer law, the absorbance is dependent by the molar analyte concentration (c), the molar absorptivity or extinction coefficient of the absorber (ε) and the distance through the sample (l). This linear correlation between absorbance (also referred to as optical density) and concentrations of analytes deviates at higher concentration. Nevertheless, size of absorption peaks in an IR or UV/VIS spectra generally correlates well with analyte concentrations. The transmission of light is not always reduced by absorption alone, but light scattering by particulates can also contribute significantly to the extinction of light passing through a suspension. The light extinction exhibited by fermentation cultures are caused by cells scattering light rather than absorbing it, and cell concentrations can also be determined through measurements of optical density (OD) at low concentrations where similar linear correlation exists (55, 62).

The reflection mode is often used for solid samples or powders where light is reflected of the surface instead of passing completely through the sample. The light undergoes sample penetration to some extent, internal diffuse reflections and some degree of absorption at the outer layers of a rough and grainy surface of the sample. The relative intensity of reflected light may be reduced at some wavelengths if the surface material is an infrared absorbent. Reflectance (R) corresponds to transmission (T) depicted above:

$$\log (1/R) = \log (I_0 /I)$$

The measurement of reflectance instead of transmittance is possible due to the low molar absorptivity of absorption bands in the NIR region. This enables recording of solid samples and therefore rapid determination by NIR spectroscopy. Reflectance spectroscopy measures the light reflected from the sample surface. This reflection contains two components: specular and diffuse reflectance. Specular reflectance is the mirror-type reflection occurring at the boundary between the sample surface and air, and contains very little information of the composition. Diffuse reflectance is light reflecting from the interior of the sample. It arises through multiple scatterings of the light by particles near the surface but inside the material. In the process of

diffusion, the radiation becomes entirely depolarized, whereas any specular reflected radiation maintains its state of polarization. The specular component may be eliminated by adjusting the position of the detector relative to the position of the sample and light source. In addition, the method of transflectance has features of both transmission and reflectance modes and may be used for liquids containing solids. The light is directed through the sample to a mirror and reflected back through the sample a second time (62).

Another widely used measurement technique similarly based on internal reflection for rapid qualitative and quantitative analysis of samples with little or no sample preparation is Attenuated Total Reflectance (ATR), which mostly is used with FTIR. ATR allows for spectra to be obtained from solids of limited solubility, pastes, fibers, slurries or other difficult samples. A beam of light reflected of the internal surface of an ATR crystal can for a short distance, as an evanescent wave, penetrate 2 to 4 micrometers into a sample placed on the crystal. The penetration depth depends on the wavelength, the refractive index of the ATR crystal and the sample and the angle of the entering light beam. Absorption of energy from multiple total light reflections of the interface between ATR crystal and sample result in attenuation of the radiation at certain wavenumbers, from which, spectra similar to traditional absorption spectra can be acquired. Sample dilution or path length adjustment may sometimes be necessary to avoid complete light absorption. An ATR crystal dipped into a slurry may furthermore work as an immersion probe for *in situ* monitoring (63).

DISPERSIVE AND FTIR INSTRUMENTS

Dispersive infrared spectrophotometers use prisms or, nowadays, diffraction gratings as a monochromator to separate the light source into different wavelengths. A double beam configuration is generally employed to initially split and direct radiation through both sample and reference before the grating disperses and directs the different wavelengths from both beams onto the detector. Fourier Transform Infrared Spectroscopy (FTIR) is a more rapid alternative measurement technique for acquiring absorption spectra to dispersive instruments. The full spectra of infrared light after interaction with the sample is collected and transformed to an interferogram using moving mirrors inside a Michelson interferometer. A Fourier transformation algorithm then converts this data set to all frequencies of the desired spectrum of a sample (55, 62).

RAMAN SPECTROSCOPY

Raman spectra are acquired from the scattered photons reemitted in all directions from molecules when irradiated with a visible or near-infrared monochromatic beam of laser light. This type of spectroscopy only applies excitation at one wavelength in contrast to absorption spectroscopy methods applying transmission or reflection of the full light spectra. Most of the radiation is during Raman

spectroscopy reemitted with unchanged wavelength in a process called Rayleigh scattering, where the collisions of the incoming photons with the molecules is said to be elastic. The Raman effect on the other hand causes a small proportion of incident photons to scatter inelastically and shift wavelength, due to transition of the intrinsic vibrational energy states of certain molecules (**Figure 7**). This shift, mainly towards smaller wavenumbers than excitation, is called Stokes shift. However, some photons may also gain energy from the molecules and less intense anti-Stokes lines will then appear with a higher wavenumbers, but with same shift in wavelength as seen with the Stokes lines. Raman scattering predominantly takes place with non-polar bonds by creating induced electric dipole moment, while water or other molecules with dipolar bonds generate weaker Raman signals (55, 57).

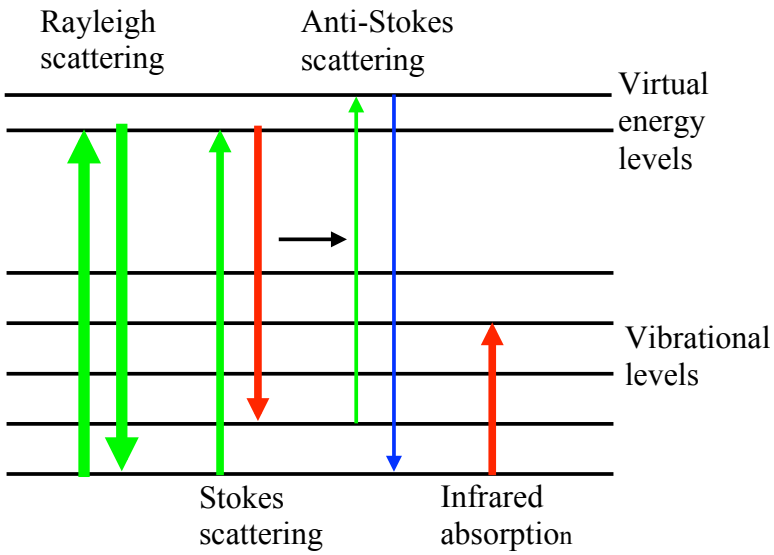

Figure 7. Transition of vibrational states of energy.

Raman spectra are generally displayed in the range from 400 to 2000 cm^{-1} relative to the excitation wavelength, which both the Stokes shift and the corresponding anti-Stokes shift are independent of. Only the weak Raman Stokes scattering is usually transmitted to the detector and displayed in the displayed spectra. Employment of extremely sensitive Charged Coupled Device (CCD) detectors in addition to applying more stable and cheap lasers has dramatically improved Raman spectroscopy.

A major challenge in Raman spectroscopy is dealing with the interference caused by the fluorescence background noise exhibited by many biological molecules. The

Raman effect is similar to fluorescence, since emission occurs at a different wavelength than the incident light, but the absorption of light is more limited to specific resonance frequencies of the fluorescent molecule. The subsequent relaxation to a lower energy state occurs in various ways resulting in broader fluorescence peaks than commonly produced by the Raman effect. Fluorescence is less likely at higher excitation wavelengths and, similarly, Raman signal intensities also tend to decreases with increasing excitation wavelengths. The optimal excitation wavelength may vary according to the compounds measured and impurities present in samples. In recent years, Raman spectrophotometers have been significantly improved by methods implemented to mathematically estimate and reject the fluorescence contribution of the measured signal (64).

Shifted Excitation Raman Difference Spectroscopy (SERDS) is another technique intended to remove fluorescence from the Raman spectra. SERDS requires excitation lasers in pairs with slightly different excitation wavelengths in order to create two spectra with a small shift in Raman bands, but with hardly any change to the broad fluorescence background. The contribution from fluorescence can be eliminated by subtracting one of the spectra from the other and reconstructing a new SERDS spectrum with intact Raman signals (65). Another Raman method, Surface Enhanced Raman Spectroscopy (SERS), strongly increases Raman signals from molecules attached to metal (Ag, Au) nanostructures (10 – 100 nm) (66).

SPECTROSCOPY COMPARISON

Raman and other IR methods complement each other well, so weak signals in IR are typically strong in Raman spectra vice versa. Dipolar and asymmetric functional groups such as carboxyl groups giving the strongest signal in NIR and MIR, whereas stretching vibration from symmetric, non-polar and double or triple bonds such as aromatic groups or hydrocarbon chains yield high Raman peaks.

The scattering principle behind Raman spectroscopy makes it relative easy to place the sample in the excitation beam and collect the scattered light, which in principle can be done using the same lens for both excitation and collection of light using a probe shown in the picture on the thesis cover page. The picture shows a Raman immersion ball probe for focused illumination and collection of scattered light from Enwave Optronics Inc. Less interference from water in aqueous samples is experienced with Raman spectroscopy compared to conventional IR transmission spectroscopy, and with little or no sample preparation required making analysis of gas, solute or solid is easy. Distinct peaks with intensities proportional to the analyte makes Raman spectroscopy suitable for quantification as well as qualitative analysis. Dispersive Raman and Fourier transform Raman are the two techniques commonly used. Dispersive Raman often equipped with grating monochromator and operating with visible excitation lasers splits the collected beam to the constituent wavelengths with a diffraction grating and directs the beam to the

detector as illustrated in **Figure 8**. Generally illuminating at 1064 nm, FT-Raman spectrophotometers employ excitation lasers at higher wavelengths, than dispersive Raman. Just as in FTIR, FT-Raman utilizes interferometer and Fourier transformation to obtain spectra of samples (55).

The major advantage of FT-Raman is inherent wavelength calibration, good resolution, low cost, and reduced fluorescence using NIR excitation, which however, also results in lower Raman scattering intensities. On the other hand, dispersive Raman is faster and can use the more efficient cooled CCD sensor, which have less background noise levels than the NIR detectors required for FT-Raman higher.

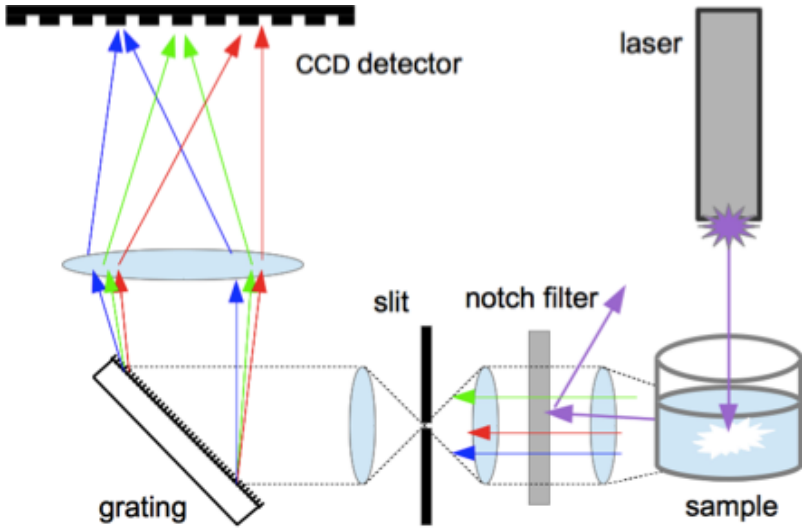

Figure 8. Dispersive Raman spectrophotometer with a notch filter to hold back Rayleigh scattered light and a diffraction grating to split the spectra light and direct light towards the detector.

The major source of background noise in Raman spectra is fluorescence exhibited mainly in biomolecules by aromatic functional groups or compounds containing conjugated double bonds. In this context, monitoring any biomass conversion processes with Raman spectroscopy may prove difficult due to the high content of aromatic lignin. Despite these drawbacks, Raman spectroscopy benefits from less cluttered spectra with narrower and distinct bands than observed with any IR techniques, as shown in **Figure 9**. Furthermore, interference of signals when using fiber optics is more disruptive with IR absorption methods, therefore making Raman well suited for remote monitoring (57).

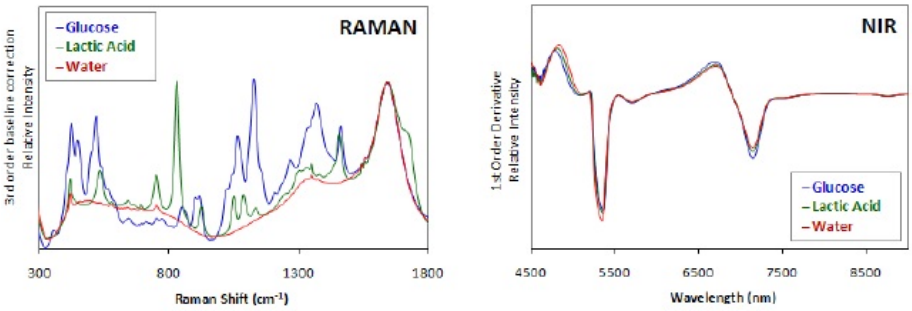

Figure 9: Comparison of baseline corrected Raman spectra (3rd order) and derivative of NIR spectra (1st order) of components often found in fermentation cultures (http://www.azom.com/article.aspx?ArticleID=8442).

The generally broad absorption bands seen in the NIR are the result of molecular overtones and combinations of the more intense fundamental mid-infrared vibrations and mainly originate from the functional groups: -CH, -OH, -SH and -NH. The overlapping combination bands make it more difficult to interpret the subtle spectral variations and to assign specific molecules to features in the NIR spectra (81). Despite the overlapping peaks, NIR has been used more extensively for quantitative monitoring of bioprocesses, because interference from water is more likely to occur using MIR exhibiting stronger molar absorptivity from the fundamental vibrations. NIR on the other hand, facilitates deeper penetration in aqueous samples, and measurement of samples with limited transparency is achievable when using reflectance methods not requiring sample preparation. The introduction of efficient chemometric calibration techniques furthermore enabled simultaneous analysis of multiple variables from NIR spectra, despite the lack of spectral features.

LORENZ-MIE SCATTERING

Applying optical sensors in the analysis and monitoring of the viscous slurries present in the bioprocesses of future biorefineries is challenged by disturbance from the high amounts of suspended biomass debris. Quantification of the components of interest is complicated by the scattering of light caused by particulates in various sizes and cells present during fermentation. Molecules or small particles scatter light elastically without changing wavelength through the Rayleigh process. The wavelength of the reemitted light scattered in all directions is unchanged. Shorter wavelengths tend to undergo Rayleigh scattering more easily, which result in the sky to appear blue.

The properties of the scattering mechanism changes and fall into the Mie-Lorenz model, as the size of the particles increase and approximate the wavelengths of the electromagnetic radiation. The similarly elastically Lorenz-Mie scattering generally

occurs when any spheres larger than molecules are exposed to visible or infrared light, and the interaction between light and spherical surfaces involves some degree of light refraction, as illustrated in **Figure 10**. This scattering effect is seen in the case with water droplets in the clouds or suspended biological cells having sizes close to wavelengths of light. As a consequence of refraction being involved, the scattered photons are redirected less equally in all directions than exhibited by the Rayleigh process and the angular directions of scattered light create resonance patterns in space. Furthermore, Lorenz-Mie scattering is in contrast to Rayleigh not strongly wavelength dependent. Almost all the wavelengths in white light is scattered equally, hence clouds appear white.

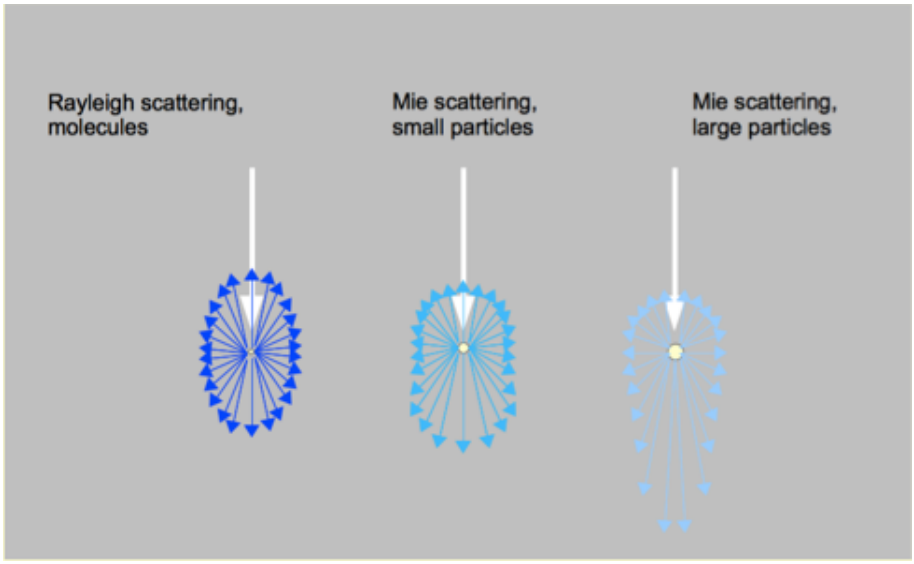

Figure 10: Rayleigh and Lorenz-Mie scattering.

The size of the spheres determines the angular directions of the scattered light. Lorenz-Mie theory can predict the direction of light scattered of suspended particles of equal size. However, modeling the scattering pattern produced in a suspension containing yeast cells and debris of biomass of different sizes is very complicated, especially when photons scatter multiple times. When measuring components in solution using absorption or Raman spectroscopy, it is essential to understand the impact from light extinction occurring when photons scatter out of a beam of light sent into these turbid samples before reaching the censor. This extinction or attenuation of light, is a combined result of absorption and the Lorenz-Mie

scattering caused by the suspended solid particulates. In this context, investigating the course of signal attenuation exhibited in opaque slurries is crucial when considering using vibrational spectroscopy techniques for on-line monitoring of processes in biorefineries (67, 68).

This thesis attempts to address issues concerning scattering during the monitoring of a yeast fermentation process using Raman spectroscopy, which is described in **Manuscript I**: Quantitative Monitoring of Yeast Fermentation using Raman.

CHEMOMETRICS

Information about a chemical parameter may be extracted from one measured variable alone, but the analytical chemical methods used for the complex systems encountered in science most often yield multivariate data. Multivariable models may therefore often prove more effective than classical univariate statistics when applying spectroscopy methods yielding soft overlaying signals as in NIR. Empirical data-driven modeling can improve interpretation of vast datasets where prior knowledge and theory alone insufficiently provides all significant correlations of a complex chemical system. In this context, chemometrics combines empirical and multivariate modeling of chemical data in order to extract hidden data and obtain the relevant chemical information. Chemometric computer-based methods usually involve:

- Exploratory data analysis
- Classification methods
- Regression methods for prediction

Hidden patterns in complex data may be revealed in a comprehensible form with exploratory data analysis using algorithms such as Principal Component Analysis (PCA) or Artificial Neural Networks (ANN) to show which variables affect these patterns the most. A classification model can visualize groupings to find unusual samples in the data set by comparing it to previously analyzed samples. The correlation between the multiple independent variables and the quantifiable desired properties is in the multivariate system expressed by the regression model. The model can include algorithms such as Partial Least Squares (PLS), Multiple Linear Regression (MLR) or Principal Component Regression (PCR) to avoid some correlation between the independent variables in the model (69).

Prior to predicting values upon measurements by using the statistical calibration techniques of chemometrics, preprocessing of the raw spectral data set may in some instances, especially with NIR spectra, be performed to enhance the relevant

information of the data set and improve predictability. Pretreatment of spectral data often include, background subtraction, mean centering, auto scaling, normalization, derivatives, smoothing noisy data (Savitzky–Golay), Standard Normal Variates (SNV) or using Multiplicative Scatter Correction (MSC) especially used with raw NIR spectra (70).

A regression model relates the dependent or response variables y to one or more independent or explanatory variables x. Multiple independent variables is common in science and the multivariate regression model, assuming a linear link, may be expressed by linear algebra notation or matrix-vector product:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \cdots \beta_p x_{i,p} + e$$

$$\mathbf{y} = \mathbf{X}\beta + E$$

where the response variable vector **y** could be a concentration, **X** could be a matrix of spectral data sets, composed of measurements (observations) for each row and columns with wavenumber (variables). β is the parameter vector or regression coefficient, and the residual e or E vector is the error term or noise (71).

MULTIPLE LINEAR REGRESSION

Applying the simple MLR method, a model β may be estimated for data set with a reduced number of variables and properties Y. Least squares is a method to estimate the coefficients to yield the best fit through data points, where the sum of all squared errors (SSE) have minimal value.

$$SSE = \sum(e)^2 = \sum(yi - \beta_0 - \beta 1 Xi,1 - \cdots \beta p Xi,2)^2$$

Assuming the model errors are Gaussian-distributed, the model regression coefficient b can be estimated (the matrix transpose is denoted ^{T}):

$$SSE = (y - Xb)^T (y - Xb)$$

$$\text{if } \partial(SSE)/\partial b = 0$$

$$\text{then } b = (X^T X)^{-1} X^T y$$

hence the fitted squares regression line is:

$$\check{y} = Xb = X(X^T X)^{-1} X^T y$$

Traditional MLR modeling can be used for data sets with few uncorrelated X variables. However, spectra, often composed of overlapping peaks, consist of numerous x-variables, where little is known about the degree of interference or correlation between the x-variables in the data set (69).

PRINCIPAL COMPONENT ANALYSIS

PCA is method for exploratory data analysis and data compression used to visualize relations in only the X-matrix between objects and between variables. The number of variables may be reduced by projecting the sample data into a low dimensional system of coordinates displaying the maximum variance of the variables. The result of PCA performed on spectral data is a transformed matrix, interpreted as the original spectra, with less or equal number of orthogonal and uncorrelated variables called principal components (PC). The compressed matrix contains only important variations from the data set, while the loss is minimal. A PCA model describes the relation between the original data in matrix X (I,J) and principal components in terms of component scores T (I,R), loadings P (J,R) and residual E (I,J), where I represents samples, J is variables and the R row in the loading matrix is the contribution of single variable to the PC:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathbf{T}} + \mathbf{E}$$

An iterative least squares process is used to determine the pairs of scores and loadings for each PC with PC1 having the highest variance followed by PC2 and so forth. Placing the loadings of each PC to be orthogonal to the others eliminates the correlation between the PC's. An increasing number of PC's will reduce the remaining variance after data compression, which is contained in the residual E matrix. Yet too many PC's may, however, result in fitting data more than necessary. PCA is a qualitative discriminant analysis method to extract the most variation of the data set to be represented in a reduced form, but PCA is not sufficient for performing predictions of the dependent variable without any subsequent regression analysis (71).

PRINCIPAL COMPONENT REGRESSION

PCR can be used in natural extension to PCA to estimate regression coefficients of a multivariate model, which makes PCR suited for large data set having correlated input variables. Similarly to MLR, PCR involves least squares computations and is applied to determine the regression values, but instead of performing the linear regression on independent variables directly, the PCA derived principal components are used. X can be substituted with components as long as they originate from independent variables only and not from Y - variables. Subsequent to PCA and

ordinary least squares regression on selected PC's have been carried out, model parameters are determined from the explanatory input variables (72).

PARTIAL LEAST SQUARES

PLS regression can establish a quantitative multivariate model of the X to Y relation despite having only minimal relevant X-variable information available and without involving qualitative visualization as with PCA. Unlike simple MLR, PLS can analyze strongly correlated noisy data having many independent variables compared to the number of observations. The PLS method combines a decomposition with linear regression and enables simultaneous prediction of several dependent variables. PLS uses both variation of X and Y to find the new compressed latent variables or components. In contrast to the two-step PCR with projection of x-variables followed by regression of compressed variables, PLS decomposition of data incorporates the covariance of both X and Y directly through a more complex process. The covariance structure of the X and Y matrices is found by projecting the predicted and observable matrix as latent structures to new spaces. The dependent and independent values are decomposed into scores (T and U), loadings (P and Q) and residuals (E and F) and linked by a linear relationship:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^{\mathbf{T}} + \mathbf{E}$$

$$Y = UQ^T + F$$

$$U = bT$$

where the regression vector of the linear model is

$$b = U^T T / T^T T$$

The model has weights instead of scores for the y-variable: $X = yw^T + E$

The score value from the X component with highest covariance can be used to predict the score value of the Y component with the highest covariance. There are several algorithms to obtain PLS estimators and one of the most popular ones used for decomposition is NIPALS (Nonlinear Interactive Partial Least Squares). The PLS loading weights are iteratively re-estimated by performing simple bivariate regression algorithms. (73). PLS models are often simpler than PCR with fewer condensed variables, because Y-data is included in the compression step.

MODEL VALIDATION

One measure to assess how well the regression line of a model represents the data is the coefficient of determination R^2, which in a model with a perfect fit will approach 1.0. However, the coefficient of determination may not give the best way

estimate of the validity of a regression model, since just adding more variables can increase R^2 hence seemingly improve the fit of the model. R^2 cannot be used as sole validation procedure to verify the effectiveness of a regression model before attempting to apply the calibration for predictions of new independent values such as concentrations.

The Root Mean Square Error of Prediction (RMSEP) is the average prediction error and can express how well the data of a new test measurement fits to the model. In contrast to the similar Root Mean Square Error of Estimation (RMSEE), applying the calibration samples already used to describe the model fit, RMSEP uses external samples to determine the actual validation error of the model. Both calibration fit error, RMSEE providing the rough estimate. The subsequent prediction error RMSEP decrease with increasing number of latent variables, but errors associated with more complex modeling on the other hand increase as more latent variables are integrated.

The Root Mean Square Error of Cross-Validation (RMSECV) is one of several internal validation methods using the calibration data itself instead of new data to obtain optimal number of model variables, while evaluating the model fit at the same time. The model is made upon one or more sub-validations, where the same data is either used for building or testing the calibration model, as the model is constructed (55).

SPECTROSCOPY SUMMARY

The introduction of chemometrics has significantly improved the overall ability to precisely monitor and quantify important bioprocess parameters. In particular, relevant information previously hidden in spectral data sets is now retrievable, thereby drastically enhancing the possible uses of spectroscopy methods such as NIR, MIR, Raman and fluorescence for construction of reliable prediction models. However, the overall combined effort required to perform calibration including full data pretreatments and chemometric analysis is very extensive and difficult to carry out - even by skilled professionals. As long as the process monitored is changed very little subsequent to the construction and validation of the regression model, utilizing chemometrics can be justified. This might not be the case, when dealing with a process being altered regularly, as often occurring in the development phase or, for instance, when media compositions are modified between test runs. It simply may be too time-consuming to adjust the comprehensive multivariate calibration model in such situations. Moreover, a sort of 'black box' is created, when chemometrics is employed, since pre-processing and calibration errors are difficult to recognize without performing full data analysis. Especially with NIR, it is practically impossible to intuitively estimate how minor changes appearing in the raw spectra will affect the different process response variables in the multivariate model.

This might not be the case with Raman spectroscopy due to more distinct and less overlapping peaks in the spectra. In the event a reasonable quantification method can be obtained with fewer pre-processing steps, than currently used with NIR, while also applying a simple univariate regression instead of multivariate regression, Raman spectroscopy may prove more flexible, easy to utilize and overall better for some non-standardized applications. Such method expected to benefit from being more flexible and transparent are however anticipated to come at the cost of precision of prediction. In summary, part of this PhD project intends to investigate at what extend acceptable quantification estimations is obtainable and applicable for monitoring bioethanol production processes, without the application of excessive data processing. Instead, quantifications are attempted applying only zeroth-order models. This matter is addressed in **Manuscript I and II**.

CONCLUSION AND PERSPECTIVES

The outcome of the presented experimental work shows some potential benefits of applying Raman spectroscopy for monitoring biorefinery processes in real-time. **Manuscript I** describes the development and use of a quantification model for measuring concentrations present during fermentation. The model deviates from conventional multiple linear chemometric methods by using an initial scattering correction step to account for the non-linear attenuation of Raman signal caused by cell particulates. Reasonable results measuring yeast fermentation was obtained with this two-step approach involving only a simple final univariate regression to determine concentrations. Extended use of this quantification model in regards to lignocellulosic bioethanol production processes involving opaque hydrolysates with much biomass debris is demonstrated in **Manuscript II.** Attempts to use Raman spectroscopy and the established quantification method resulted in satisfactory measurements of concentrations to be obtained from monitoring in the hydrolysates during especially the fermentation process and to some degree also from the pretreatment process. However, the enzymatic hydrolysis reaction was difficult to monitor using Raman spectroscopy due to only minor spectral changes being observed during the release of sugars from the cellulose polymers. Measurements in all the processes are significantly affected by background noise caused by scattering of signal from the suspended biomass debris. Some degree of signal transparency through the hydrolysates is necessary in order to acquire any useful spectra for quantification. Real-time concentration measurements during the fermentation process were only obtainable, if the stirring speed was reduced enough to allow some precipitation of the suspended particulates around the immersion probe at the top of the reaction mixture, while keeping yeast cells suspended.

Overall, Raman spectroscopy shows some great potential as an analysis tool for future biorefinery processes and additional utilization should be investigated in extension to this study. The simple univariate model presented may become more accurate by integrating the non-linear scattering features for scattering correction

into the conventional chemometric methods using multiple variant regressions for quantification. A natural expansion of the method would be additionally to monitor the conversion of the pentose sugars, which might be used to monitor combined C5 and C6 fermentation or control microaerobic conditions when using xylose-fermenting yeasts. Another application possibility within bioethanol fermentation could be to measure the presence of lactic acid, thereby enabling early detection of bacterial contamination. Raman spectroscopy might also be applied for other biofuel production processes such as measuring different volatile fatty acids during anaerobic digestion or the gas composition during gasification and pyrolysis processes. Strong Raman signals are expected to arise from the many non-polar bonds in these compounds.

Another obvious study possibility is to develop more techniques to combine separation and real-time Raman measurements in order to reduce some of the background noise from cell and biomass particulates expected to be present in any liquid reaction mixture regardless of the bioprocess being monitored. For instance, a system with providing a continuous flow of the process liquid through a cyclone or spin filter before passing by the immersion probe may improve measurement accuracy.

The final part of this project concerning a performance comparison between the two xylose fermenting yeasts *Scheffersomyces stipitis* and *Spathaspora passalidarum* when grown on three important tropical biomasses is disclosed in **Manuscript III**. Fermentation of pretreated sugarcane bagasse, eucalyptus tree fiber and palm oil empty fruit bunches revealed some slightly better yields of up to 20 g/L ethanol in hydrolysate from empty fruit bunches when using *S. passalidarum*. This research furthermore supports results by others showing *S. passalidarum* to be among the best xylose fermenting yeasts. Hydrolysate with a high concentration of acetic acid caused metabolism in *S. passalidarum* initially to be inhibited. One important perspective for future studies on the potential of using *S. passalidarum* for lignocellulosic bioethanol production could thus be to investigate the extent of inhibition by acetic acid on the performance of this yeast.

REFERENCES

(1) Doney SC, Fabry VJ, Feely RA, Kleypas JA. Ocean acidification: the other CO2 problem. Marine Science 2009; 1

(2) Jones N. Troubling milestone for CO2. Nature Geoscience 2013; 6(8): 589-589

(3) Stott P. Highlights from IPCC WGI AR5 report and MOHC contribution. MOSAC-18 BRIEFING PAPERS 2013: 155

(4) Stocker, T. F., D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P. M. Midgley. IPCC, 2013: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*
Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013

(5) Davis SJ, Caldeira K, Matthews HD. Future CO2 emissions and climate change from existing energy infrastructure. Science 2010; 329(5997): 1330-1333

(6) McKendry P. Energy production from biomass (part 1): overview of biomass. Bioresour Technol 2002; 83(1): 37-46

(7) Balat M, Balat H. Recent trends in global production and utilization of bio-ethanol fuel. Appl Energy 2009; 86(11): 2273-2282

(8) Lynd LR, Wyman CE, Gerngross TU. Biocommodity engineering. Biotechnol Prog 1999; 15(5): 777-793

(9) Wei H, Xu Q, Taylor II LE, Baker JO, Tucker MP, Ding S. Natural paradigms of plant cell wall degradation. Curr Opin Biotechnol 2009; 20(3): 330-338

(10) Zhao X, Zhang L, Liu D. Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. Biofuels, Bioproducts and Biorefining 2012; 6(4): 465-482

(11) Cosgrove DJ. Growth of the plant cell wall. Nature reviews molecular cell biology 2005; 6(11): 850-861

(12) Béguin P, Aubert J. The biological degradation of cellulose. FEMS Microbiol Rev 1994; 13(1): 25-58

(13) Lynd LR, Weimer PJ, Van Zyl WH, Pretorius IS. Microbial cellulose utilization: fundamentals and biotechnology. Microbiology and molecular biology reviews 2002; 66(3): 506-577

(14) Deguchi S, Tsujii K, Horikoshi K. Cooking cellulose in hot and compressed water. Chem.Commun. 2006(31): 3293-3295

(15) Saha BC. Hemicellulose bioconversion. J Ind Microbiol Biotechnol 2003; 30(5): 279-291

(16) Buranov AU, Mazza G. Lignin in straw of herbaceous crops. Industrial crops and products 2008; 28(3): 237-259

(17) Akin DE. Plant cell wall aromatics: influence on degradation of biomass. Biofuels, Bioproducts and Biorefining 2008; 2(4): 288-303

(18) Scott AC. Coal petrology and the origin of coal macerals: a way ahead? International Journal of Coal Geology 2002; 50(1): 119-134

(19) Hedges JI, Cowie GL, Ertel JR, James Barbour R, Hatcher PG. Degradation of carbohydrates and lignins in buried woods. Geochim Cosmochim Acta 1985; 49(3): 701-711

(20) Karp A, Shield I. Bioenergy from plants and the sustainable yield challenge. New Phytol 2008; 179(1): 15-32

(21) Zhu X, Long SP, Ort DR. Improving photosynthetic efficiency for greater yield. Annual review of plant biology 2010; 61: 235-261

(22) Papong S, Yuvaniyama C, Lohsomboon P, Malakul P. Overview of biomass utilization in Thailand. The meeting for LCA in ASEAN Biomass Project, Japan 2004; 28

(23) Lora E, Andrade R. Biomass as energy source in Brazil. Renewable and Sustainable Energy Reviews 2009; 13(4): 777-788

(24) Njoku SI, Iversen JA, Uellendahl H, Ahring BK. Production of ethanol from hemicellulose fraction of cocksfoot grass using pichia stipitis. Sustainable Chemical Processes 2013; 1(1): 1-13

(25) Cherubini F, Jungmeier G, Wellisch M, Willke T, Skiadas I, Van Ree R, et al. Toward a common classification approach for biorefinery systems. Biofuels, Bioproducts and Biorefining 2009; 3(5): 534-546

(26) Kamm B, Kamm M, Gruber PR, Kromus S. Biorefinery systems – an overview. Biorefineries-industrial processes and products: Status quo and future directions 2006: 1-40

(27) Klein-Marcuschamer D, Oleskowicz-Popiel P, Simmons BA, Blanch HW. The challenge of enzyme cost in the production of lignocellulosic biofuels. Biotechnol Bioeng 2012; 109(4): 1083-1087

(28) Goyal H, Seal D, Saxena R. Bio-fuels from thermochemical conversion of renewable resources: a review. Renewable and Sustainable Energy Reviews 2008; 12(2): 504-517

(29) Bridgwater A. Renewable fuels and chemicals by thermal processing of biomass. Chem Eng J 2003; 91(2): 87-102

(30) Wackett LP. Microbial-based motor fuels: science and technology. Microbial Biotechnology 2008; 1(3): 211-225

(31) Sauer M, Porro D, Mattanovich D, Branduardi P. Microbial production of organic acids: expanding the markets. Trends Biotechnol 2008; 26(2): 100-108

(32) Weiland P. Biogas production: current state and perspectives. Appl Microbiol Biotechnol 2010; 85(4): 849-860

(33) Lee SY, Park JH, Jang SH, Nielsen LK, Kim J, Jung KS. Fermentative butanol production by clostridia. Biotechnol Bioeng 2008; 101(2): 209-228

(34) Laser M, Larson E, Dale B, Wang M, Greene N, Lynd LR. Comparative analysis of efficiency, environmental impact, and process economics for mature biomass refining scenarios. Biofuels, Bioproducts and Biorefining 2009; 3(2): 247-270

(35) Wright MM, Brown RC. Comparative economics of biorefineries based on the biochemical and thermochemical platforms. Biofuels, Bioproducts and Biorefining 2007; 1(1): 49-56

(36) Mosier N, Wyman C, Dale B, Elander R, Lee Y, Holtzapple M, et al. Features of promising technologies for pretreatment of lignocellulosic biomass. Bioresour Technol 2005; 96(6): 673-686

(37) Alvira P, Tomás-Pejó E, Ballesteros M, Negro M. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review. Bioresour Technol 2010; 101(13): 4851-4861

(38) Jørgensen H, Kristensen JB, Felby C. Enzymatic conversion of lignocellulose into fermentable sugars: challenges and opportunities. Biofuels, Bioproducts and Biorefining 2007; 1(2): 119-134

(39) Yang B, Wyman CE. Pretreatment: the key to unlocking low-cost cellulosic ethanol. Biofuels, bioproducts and biorefining 2008; 2(1): 26-40

(40) Effects of sugar inhibition on cellulases and β-glucosidase during enzymatic hydrolysis of softwood substrates. Proceedings of the Twenty-Fifth Symposium on Biotechnology for Fuels and Chemicals Held May 4–7, 2003, in Breckenridge, CO: Springer; 2004

(41) Wilson DB. Cellulases and biofuels. Curr Opin Biotechnol 2009; 20(3): 295-299

(42) Harris PV, Welner D, McFarland K, Re E, Navarro Poulsen J, Brown K, et al. Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. Biochemistry 2010; 49(15): 3305-3316

(43) Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VG. Novel enzymes for the degradation of cellulose. Biotechnology for Biofuels 2012; 5(1): 1-13

(44) Merino ST, Cherry J. Progress and challenges in enzyme development for biomass utilization. Biofuels: Springer; 2007. p. 95-120

(45) Bajwa PK, Pinel D, Martin VJJ, Trevors JT, Lee H. Strain improvement of the pentose-fermenting yeast *Pichia stipitis* by genome shuffling. J Microbiol Methods 2010; 81(2): 179-186

(46) Chandel AK, Chandrasekhar G, Radhika K, Ravinder R, Ravindra P. Bioconversion of pentose sugars into ethanol: a review and future directions. Biotechnol Mol Biol Rev 2011; 6(1): 008-020

(47) Tomás-Pejó E, Oliva J, González A, Ballesteros I, Ballesteros M. Bioethanol production from wheat straw by the thermotolerant yeast Kluyveromyces marxianus CECT 10875 in a simultaneous saccharification and fermentation fed-batch process. Fuel 2009; 88(11): 2142-2147

(48) Ballesteros M, Oliva J, Negro M, Manzanares P, Ballesteros I. Ethanol from lignocellulosic materials by a simultaneous saccharification and fermentation process (SFS) with *Kluyveromyces marxianus* CECT 10875. Process Biochemistry 2004; 39(12): 1843-1848

(49) Wilkins MR, Mueller M, Eichling S, Banat IM. Fermentation of xylose by the thermotolerant yeast strains *Kluyveromyces marxianus* IMB2, IMB4, and IMB5 under anaerobic conditions. Process Biochemistry 2008; 43(4): 346-350

(50) Lin Y, Tanaka S. Ethanol fermentation from biomass resources: current state and prospects. Appl Microbiol Biotechnol 2006; 69(6): 627-642

(51) Georgieva TI, Mikkelsen MJ, Ahring BK. High ethanol tolerance of the thermophilic anaerobic ethanol producer *Thermoanaerobacter* BG1L1. Central European Journal of Biology 2007; 2(3): 364-377

(52) Chang T, Yao S. Thermophilic, lignocellulolytic bacteria for ethanol production: current state and perspectives. Appl Microbiol Biotechnol 2011; 92(1): 13-27

(53) Palmqvist E, Hahn-Hägerdal B. Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition. Bioresour Technol 2000; 74(1): 25-33

(54) Cervera AE, Petersen N, Lantz AE, Larsen A, Gernaey KV. Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation. Biotechnol Prog 2009; 25(6): 1561-1581

(55) Bakeev KA. Process analytical technology: spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries. Blackwell Publishing Ltd.: Wiley; 2005

(56) Becker T, Hitzmann B, Muffler K, Pörtner R, Reardon K, Stahl F, et al. Future aspects of bioprocess monitoring. White Biotechnology 2007: 249-293

(57) Beutel S, Henkel S. *In situ* sensor techniques in modern bioprocess monitoring. Appl Microbiol Biotechnol 2011; 91(6): 1493-1505

(58) Bryant DN, Morris SM, Leemans D, Fish SA, Taylor S, Carvell J, et al. Modelling real-time simultaneous saccharification and fermentation of lignocellulosic biomass and organic acid accumulation using dielectric spectroscopy. Bioresour Technol 2011; 102(20): 9675-9682

(59) Kiviharju K, Salonen K, Moilanen U, Eerikäinen T. Biomass measurement online: the performance of *in situ* measurements and software sensors. J Ind Microbiol Biotechnol 2008; 35(7): 657-665

(60) Carvell JP, Dowd JE. On-line measurements and control of viable cell density in cell culture manufacturing processes using radio-frequency impedance. Cytotechnology 2006; 50(1-3): 35-48.

(61) Stuart BH. Infrared spectroscopy: fundamentals and applications. John Wiley & Sons, Inc.; 2004

(62) Scarff M, Arnold SA, Harvey LM, McNeil B. Near infrared spectroscopy for bioprocess monitoring and control: current status and future trends. Crit Rev Biotechnol 2006; 26(1): 17-39

(63) Mazarevica G, Diewok J, Baena JR, Rosenberg E, Lendl B. On-line fermentation monitoring by mid-infrared spectroscopy. Appl Spectrosc 2004; 58(7): 804-810

(64) Zhang Z, Chen S, Liang Y, Liu Z, Zhang Q, Ding L, et al. An intelligent background-correction algorithm for highly fluorescent samples in Raman spectroscopy. J Raman Spectrosc 2010; 41(6): 659-669

(65) Shreve AP, Cherepy NJ, Mathies RA. Effective rejection of fluorescence interference in Raman spectroscopy using a shifted excitation difference technique. Appl Spectrosc 1992; 46(4): 707-711

(66) Stiles PL, Dieringer JA, Shah NC, Van Duyne RP. Surface-enhanced Raman spectroscopy. Annu.Rev.Anal.Chem. 2008; 1: 601-626

(67) Cox A, DeWeerd AJ, Linden J. An experiment to measure Mie and Rayleigh total scattering cross sections. American Journal of Physics 2002; 70: 620

(68) McCartney EJ. Optics of the Atmosphere: Scattering by molecules and particles. New York, John Wiley and Sons, Inc., 1976. 421 p.; 1976; 1

(69) Roggo Y, Chalus P, Maurer L, Lema-Martinez C, Edmond A, Jent N. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. J Pharm Biomed Anal 2007; 44(3): 683-700

(70) Rinnan Å, Berg F, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends in Analytical Chemistry 2009; 28(10): 1201-1222

(71) Miller CE. Chemometrics for on-line spectroscopy applications - theory and practice. J Chemometrics 2000; 14(5-6): 513-528

(72) Escandar GM, Damiani PC, Goicoechea HC, Olivieri AC. A review of multivariate calibration methods applied to biomedical analysis. Microchemical journal 2006; 82(1): 29-42

(73) Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. Chemometrics Intellig Lab Syst 2001; 58(2): 109-130

(74) Kumar S, Singh SP, Mishra IM, Adhikari DK. Recent advances in production of bioethanol from lignocellulosic biomass. Chem Eng Technol 2009; 32: 517-526

(75) Agbogbo FK, Coward-Kelly G. Cellulosic ethanol production using the naturally occurring xylose-fermenting yeast, *Pichia stipitis*. Biotechnology letters 2008; 30: 1515-1524

(76) Hou X. Anaerobic xylose fermentation by *Spathaspora passalidarum*. Appl Microbiol Biotechnol 2011; 1-10

(77) Hahn-Hägerdal B, Karhumaa K, Fonseca C, Spencer-Martins I, Gorwa-Grauslund MF. Towards industrial pentose-fermenting yeast strains. Appl Microbiol Biotechnol 2007; 74: 937-953

(78) Lee H, Biely P, Latta R, Barbosa M, Schneider H. Utilization of xylan by yeasts and its conversion to ethanol by *Pichia stipitis* strains. Appl Environ Microbiol 1986; 52: 320–324

(79) Jeffries TW. Engineering yeasts for xylose metabolism. Current Opinion in Biotechnology 2006; 17(3): 320-326

(80) Kuyper M, Toirkens MJ, Diderich JA, Winkler AA, Dijken JP, Pronk JT. Evolutionary engineering of mixed-sugar utilization by a xylose-fermenting Saccharomyces cerevisiae strain. FEMS yeast research 2005; 5(10): 925-934

(81) Blanco M, Coello J, Iturriaga H, Maspoch S, De La Pezuela, C. Near-infrared spectroscopy in the pharmaceutical industry. ANALYST-LONDON-SOCIETY OF PUBLIC ANALYSTS THEN ROYAL SOCIETY OF CHEMISTRY- 1998; 123: 135R-150R.

MANUSCRIPT I

Quantitative Monitoring of Yeast Fermentation using Raman Spectroscopy

Jens A. Iversen, Rolf W. Berg, Birgitte K. Ahring
In press in *Analytical and Bioanalytical Chemistry*

Quantitative Monitoring of Yeast Fermentation using Raman Spectroscopy

Jens A. Iversen[1, 3], Rolf W. Berg[2] and Birgitte K. Ahring[1, 3,*]

[1]Section for Sustainable Biotechnology, Aalborg University, A. C. Meyers Vænge 15, 2450 Copenhagen SV, Denmark
[2]Department of Chemistry, Technical University of Denmark, Bygning 207, 2800 Lyngby, Denmark
[3]Center for Bioproducts and Bioenergy, Washington State University Tri-Cities, 2710 Crimson Way, Richland, WA 99354, USA

* Corresponding author. Phone: +1 509-372-7682; Email: bka@wsu.edu

Abstract Compared to traditional IR methods, Raman spectroscopy has the advantage of only minimal interference from water when measuring aqueous samples, which makes this method potentially useful for *in-situ* monitoring of important industrial bioprocesses. This study demonstrates real-time monitoring of a *Saccharomyces cerevisiae* fermentation process using a Raman spectroscopy instrument equipped with a robust sapphire ball probe. A method was developed to correct the Raman signal for the attenuation caused by light scattering cell particulate, hence enabling quantification of reaction components and possibly measurement of yeast cell concentrations. Extinction of Raman intensities to more than 50% during fermentation were normalized with approximated extinction expressions using Raman signal of water around 1627 cm^{-1} as internal standard to correct for the effect of scattering. Complicated standard multi variant chemometric techniques, such as PLS, were avoided in the quantification model, as an attempt to keep the monitoring method as simple as possible and still get satisfactory estimations. Instead, estimations were made with a two-step approach, where initial scattering correction of attenuated signals was followed by linear regression. *In-situ* quantification measurements of the fermentation resulted in root mean square errors of prediction (RMSEP) of 2.357, 1.611 and 0.633 g/L for glucose, ethanol and yeast concentrations respectively.

Keywords Raman spectroscopy · sapphire ball probe · on-line monitoring · yeast fermentation · quantitative analysis · scattering correction.

Introduction

Temperature, pressure, pH and dissolved oxygen are common parameters monitored *in-situ* during chemical and biological production processes. Retrieving additional information about concentrations of reagents and products often requires labor-intensive sample processing, delaying analysis and process control. Real-time monitoring using process analytical technologies (PAT) to ensure effective operation, automated control and the analysis of critical quality parameters in industrial bioprocesses are becoming increasingly important [1, 2]. Vibrational spectroscopy techniques in near-infrared (NIR) and mid-infrared (MIR) or light absorption within the ultraviolet or visible ranges enable real-time measurement and use of a feedback mechanism for computer automated process control. Non-destructive optical methods can complement or substitute time-consuming off-line measuring methods for monitoring chemical reactions, such as high performance liquid chromatography (HPLC) or gas chromatography (GC). In particular, monitoring of fermentations or other bioprocesses can benefit from spectroscopy methods, since several components can be analyzed simultaneously. The risk of microbial contamination is reduced by non-invasively measuring with optic components able to withstand steam sterilization [3]. Attempts of *in situ* non-invasive spectroscopic monitoring of fermentation processes have been reported using NIR to analyze biomass, glucose and ethanol during fermentation [4]. NIR is based on the molecular overtone vibrational transition in contrast to the fundamental vibrational mode transitions of molecular bonds measured in the MIR range. As a consequence, the more narrow peaks seen in MIR are more useful and capable of identifying unknown compounds. However, water is a strong infrared absorber and use of IR spectroscopy is difficult even when using attenuated total reflection (ATR) elements [5]. Alternatively, Raman spectroscopy can be used, because the Raman scattering of water is weak, especially at 0-3000 cm^{-1}. In addition narrower and cleaner peaks are typically obtained than with MIR absorption, hence making better resolution, and direct spectral interpretation possible [6]. Raman spectroscopy is based on the inelastic scattering of monochromatic light from molecules, reemitting photons with Stokes energy differences equal to the vibrational transitions of the molecule. Raman scattering is generated by induced dipole moments and strong Raman intensities are exhibited by symmetric molecular bonds with low polarity such as carbon-carbon double bonds [7]. Raman signals are thus weak compared to the Rayleigh emitted light created by elastic scattering. Raman spectroscopy has occasionally troublesome fluorescence background noise created by laser illumination [8]. Instrument advancement has recently been achieved with the design of a chemically robust sapphire ball probe with focal point optimized for opaque solutions (US Patent 6,831,745) that has the advantage of being functional with only one lens for both the excitation light beam entering the sample and for collecting the backscattering Raman signal.

An early attempt to analyze fermentation products applying ATR Raman spectroscopy was done in 1987 by Shope [9]. Signal-to-noise improvements have been achieved applying mathematical algorithms on the basis of polynomial curve fitting or wavelength shifting of the excitation laser to estimate and reject the noise from the fluorescence [10]. Shaw [11] used dispersive Raman to determine concentrations of glucose and ethanol during yeast fermentation by monitoring on a bypass line free of cells obtained by filtration. A study to simultaneously monitor glucose, ethanol and optical cell density during yeast fermentation has been presented by Sivakesava [12] who used both FT-MIR and FT-Raman spectroscopy with multivariate techniques. FT-Raman proved to be less accurate at low concentrations than FT-MIR. Recently, real time monitoring of multiple process parameters in the increasingly important mammalian cell culture bioprocesses has been achieved using *in situ* Raman spectroscopy [13, 14].

Chemometrics combines spectroscopy with statistical methods to enable quantitative analysis of complex data with multiple response variables by incorporating the combined effect into the model. Multivariate data analysis methods are particularly beneficial when signals are weak. The simplest multivariate regression method to find the correlation between one dependent variable and multiple independent variables is multi-linear regression (MLR) estimated by partial least square (PLS). Principal component regression (PCR) - using principal component analysis (PCA) instead of least squares to estimate regression coefficients - is more suitable to explain the variance in data sets [15, 16]. The Raman intensity is directly proportional to the number of molecules [17]. Exceptions to this exist when the laser wavelength is close to an electronic absorption band resulting in resonance Raman scattering [18]. If significant non-linear characteristics are present in spectral data sets the situation is more complicated.

The presence of light scattering microbial cells in a fermentation broth further complicates data analysis. Molecules are subject to Rayleigh scattering, whereas larger particles, bubbles, droplets and cells with a size closer to the wavelength of the incident light fall into the domain of Lorenz-Mie scattering and are responsible for turbid waters. Lorenz-Mie scattering is more difficult than Rayleigh scattering to calculate in terms of direction and magnitude [19, 20]. This diffuse scattering is responsible for attenuation of both the laser intensity reaching the sample and the Raman light exiting from the sample, thereby reducing the signals. Both scattering and absorption contribute to extinction of the light in particulate media and this kind of extinction is also referred to as optical density [21]. Scattering is regarded as predominant compared to absorption in the cell suspensions [19]. Extinction of signal intensity as a linear function of concentration does, however, only apply to single scattering events, while higher turbidity with diffuse light and shadow effects proved this expression to be more complicated [22]. Raman data sets obtained from cell suspensions are expected to involve multiple scattering, so a non-linear extinction function of cell concentration *E(c)*

should be used in a quantification model covering the scattering range above the dilute Beer-Lambert domain:

$$I = I_0 \cdot 10^{-E(c)} \quad (1)$$

All Raman signals are reduced when the turbidity increases, but not all peaks in spectrum are changed by variations in analyte concentrations. Therefore, variations of Raman intensities of components - constant during reaction - can be used as internal standards to reveal and correct for all disturbances affecting the Raman signals from reaction components [23]. Quantitative Raman monitoring of ethanol fermentation has been reported by Picard [24], who used sulfate as an internal standard to correct for Raman signal disturbances generated in the sample or the instrument. In addition, concentrations of glucose and fermentation products in an *Escherichia coli* culture have been measured with Raman spectroscopy by correcting the light attenuation from the biomass and bubbles using the estimated water concentration as an internal reference [25]. In anaerobic fermentations where cell particulates are the only source of light attenuation, the cell concentration may thus be monitored during fermentation by use of the intensity reduction of the internal standard signal.

In this study we demonstrate real-time monitoring of a *Saccharomyces cerevisiae* fermentation process by use of Raman spectroscopy with a sapphire ball probe designed to withstand the harsh environment during the sterilization and to minimize the interference from particulates. We examined linear and non-linear features in the data sets for quantification purposes, and the scattering correction procedure was attempted using the internal standard approach to develop working process control technologies for biofuel production or other bioprocesses.

Materials and Methods

Fermentation

A culture medium containing 10 g/L yeast extract, 20 g/L protease peptone and 40 g/L glucose was autoclaved and kanamycin (50 mg/L) was added before inoculation to 2 g/L *Saccharomyces cerevisiae* dry yeast (Thermosacc, Lallemand Ethanol Technology, Denmark). Fermentation was performed at 35 °C for 7 hours in 1L Multifors bioreactor (Infors-HT, Switzerland). Yeast concentration was determined by optical density measurements at 600 nm on a UV/Vis spectrophotometer (Jenway, UK).

HPLC Analysis

Sample concentrations were determined with High Performance Liquid Chromatography (HPLC) (Ultimate 3000, Dionex,CA, USA). 10 µL of sample supernatant were run through the Aminex HPX-87H column (300 X 7.8 mm) combined with a Cation-H guard column (30 X 4.6 mm), both Bio-Rad Laboratories, CA, USA. The components were separated isocratically with 4 mM H_2SO_4 as eluent for 26 min with a flow rate of 0.6 mL/min at 60 °C and detected with the refractive index (RI). Mixtures containing glucose, glycerol, ethanol and lactate were used as standards (concentrations of 0.5 to 20 g/L). 1 mL samples were added with 10 µL 1 M H_2SO_4 and supernatant was used for HPLC.

Raman Analysis

All measurements were done with maximum frequency-stabilized laser output at 785 nm (deep red, 300 mW power) using a ProRaman-I instrument from Enwave Optronics Inc. The instrument had a 250-2350 cm^{-1} spectral coverage and a 6 cm^{-1} spectral resolution, was equipped with a cooled charge-coupled device (CCD) detector operating at -60 °C and used an immersion fiber optic probe (12 mm diameter) with ball sapphire lens. The cm^{-1} axis scale was calibrated by means of recording Raman lines of the provided Enwave calibrating sample with given bands at 382, 537, 996, 1140, 1590 and 2220 cm^{-1}. Data were processed with the EZRaman Reader ver. 8.2.2 spectra managing software. References, supernatant samples and the fermentation process were all measured or monitored by acquiring Raman data for 20 seconds and averaging 45 times, thereby giving data points for every 15 min during the real-time continuous measurements. Samples taken every hour during fermentation were centrifuged for 60 seconds at 5000 rpm and typically 2-10 mL supernatant samples were subsequently measured with Raman. Signals from ambient light were avoided by covering samples or the bioreactor during all measurements.

Except for spectra used for measurements of the magnitude of fluorescence or the impact from particles or yeast cells, all shown spectra were manipulated to subtract the fluorescence background. Baseline fit was done automatically with a polynomial algorithm using the autobaseline 2 setting of the software. In order to quantify the measured Raman signal of the components, the baseline corrected values were normalized according to zero values and then correlated to the standard curves.

Reference solutions (1 % w/w in water) of glucose, ethanol, glycerol and lactic acid were made to identify the peaks relevant in fermentation with yeast. Glucose and ethanol calibration curves were based on a simulated fermentation with the gradual conversion of 100 g/L glucose to 50 g/L ethanol in 10 % increments. Mixtures of 20 g/L ethanol and 70 g/L glucose with various

concentrations of yeast were used to quantify the extinction of the Raman signal due to cell particulate scattering. Both calibration standards and the extinction solutions were made as culture media with 10 g/L yeast extract and 20 g/L protease peptone in the reference solutions.

Quantitative Raman analysis method

Quantification of glucose, ethanol and cell biomass concentrations using Raman measurements were done in a two-step approach. First, extinction of the Raman signal using the water peak at ~1627 cm^{-1} as an internal standard was determined to get the cell concentrations applying equation 3. Extinction of Raman signals of the analytes were then determined and corrected according to the obtained cell concentration. Subsequent to the scattering corrections, the concentrations of glucose and ethanol were predicted without using any multiple linear regressions or partial least squares analysis, but only standard linear regression.

Results and Discussions

Reference Overview

Raman spectra of relevant components likely to be formed or present during yeast fermentation were collected. Reference solutions of glucose, ethanol, glycerol and water are shown in Fig. 1a In addition to these substrate and product molecules, lactic acid produced as a result of bacterial contamination is also an important compound to monitor during yeast fermentation.

Each compound has some distinct peaks in wavenumber regions with only minor interference from other compounds. Thus, the main peak of glucose is at 1123 cm^{-1} and a smaller one is present at 514 cm^{-1}, whereas a Stokes shifted Raman band from ethanol is seen strongly at 877 cm^{-1}, but bands are also present at 1046 and 1455 cm^{-1}. Lactic acid Raman scatters at 822 cm^{-1}, whereas distinct peaks at 480, 840 and 1055 cm^{-1} can be observed with glycerol and a broad Raman signal around 1620-1645 cm^{-1} is the bending water molecule. The strong signals at 378, 416, 642 and 749 cm^{-1} is Raman scattering from the sapphire crystal in the probe.

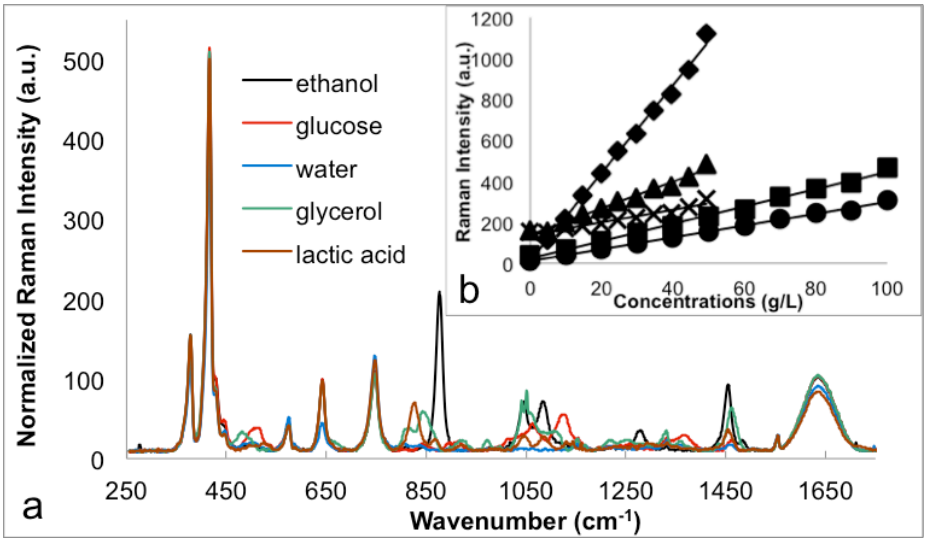

Fig. 1 Fermentation components. Raman spectra of references upon fluorescence corrections (**a**) and Standard plot of baseline corrected spectra for a simulated fermentation of 100 g/L glucose up to 50 g/L (**b**). Ethanol measured at 877 cm^{-1} (♦), y = 21.406x + 12.553, R^2 = 0.9959; 1046 cm^{-1} (×), y = 3.0716x + 138.07, R^2 = 0.9569 at 1455 cm^{-1} (▲), y = 6.4871x + 138.52, R^2 = 0.984. Glucose measured at 514 cm^{-1} (●), y = 2.8793x + 11.362, R^2 = 0.9973 and 1123 cm^{-1} (■), y = 4.2116x + 24.177, R^2 = 0.994.

A strong linearity between the concentrations and corresponding Raman intensities - subsequent to the fluorescence baseline correction - is seen in Fig. 1b. The intensity calibration was made on the basis of the simulated fermentation of glucose to ethanol and is displayed in 10% increments of conversion. The Raman intensity from ethanol at 877 cm^{-1} is particularly strong, even at low concentrations, compared to any glucose signals, which can be explained by the non-polar feature of the C-C bond.

The effect of yeast cells on Raman measurement was seen when a raw spectrum of a fermentation broth sample containing 7 g/L yeast cells was compared with the spectra of the same sample after centrifugation (Fig. 2). Most of the directly measured signal prior to baseline correction is from background fluorescence with Raman scattering contributing with less than 15% in both cases. The presence of the cells reduces the Raman intensities with more than 40% for some peaks, whereas the removal of yeast cells resulted in an overall 10% reduction of the fluorescence background noise contribution to the signal. The change in yeast content thus seems to have a high impact on the Raman signal attenuation, but little on the fluorescence.

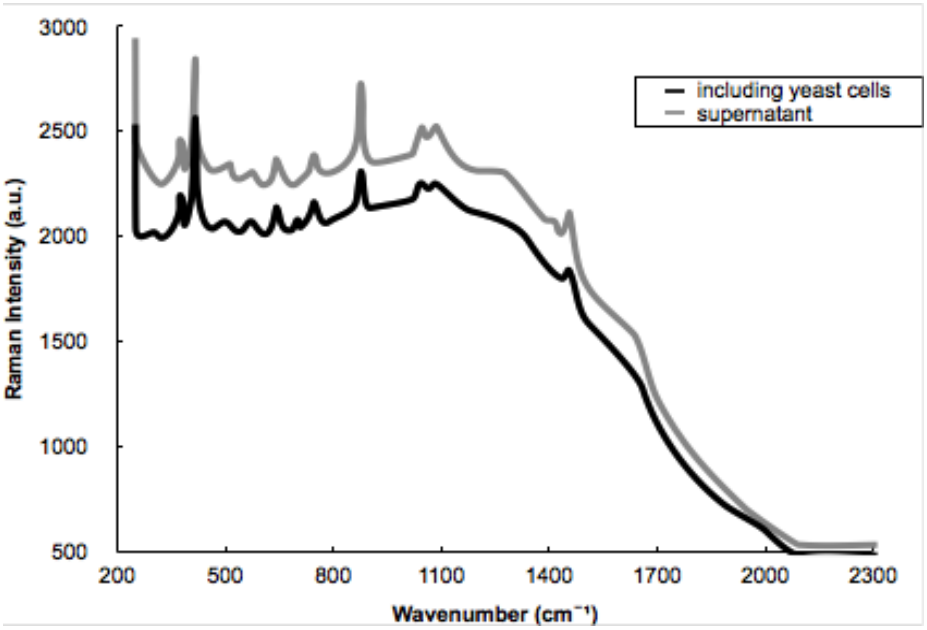

Fig. 2 Non-baseline corrected Raman spectra for fermentation with *S. cerevisiae*. The sample was obtained after 20 hours of fermentation and measured with yeast cells and without after separation.

Effect of Scattering

Fig. 3 illustrates spectra acquired from suspensions of viable yeast in water. In general, Lorenz-Mie scattering is not considered to be wavelength dependent compared to Rayleigh emission, but the directionally dependent scattering property - also referred to as anisotropy – of the cell particulates, resulting in a wave pattern of the Raman signal throughout an extended range of wavenumbers in the spectra. These periodic bands of significant magnitude, observed at 300, 363, 427, 498, 571, 649 cm^{-1} and seeming to continue for every 65 to 70 cm^{-1} at diminishing strengths, do not correlate with any expected structural component in the yeast cells or any previously found band in spectra of a single yeast cell acquired by Raman micro spectroscopy as reported by Xie [26]. One way to explain this is to let it result from the change in angular scattering intensity of a particle in accordance with Lorenz-Mie theory. Accordingly, the scattering cross section of a suspension of spheres exhibits a pattern with periodic maxima and minima as the wavelength changes and the angular configuration is assumed to remain constant [27, 28]. These oscillating features are caused by the constructive and destructive interference of the diffracted

light expected to come from both the fluorescence background and the Stokes signal itself and can be expressed by a Bessel function, which describes the scattering fluctuations within the Lorenz-Mie regime as the parameters (particle size, frequency or wavelength) are gradually altered. The spectra may however be partially at the boundary of the Mie domain, since a spectral range from 200 to 2000 cm^{-1} corresponds to wavelengths of 50000 to 5000 nm, whereas the diameter of the yeast cells could be somewhere between 10000 and 40000 nm. Another explanation for the pattern can be some kind of resonance fluorescence.

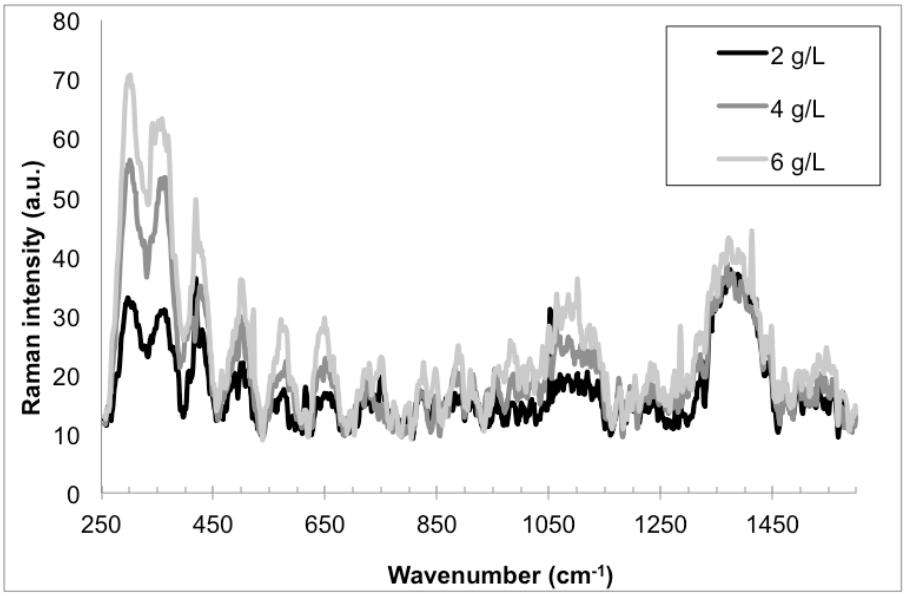

Fig. 3 Oscillating patterns in the Raman spectra exhibited by increasing concentrations of yeast cells suspended in water. Spectra derived upon the subtraction of a reference Raman spectrum of water.

To further investigate the behavior of the spectral signals, Raman spectra of glucose and ethanol solutions mixed with yeast at various concentrations were acquired. Fig. 4a displays how the intensities of some of the distinct Raman peaks - originating from solutes or water throughout the spectrum generally decline at a decreasing rate as the yeast content increases. This seemed to be true for all Raman bands except the one at around 300 cm^{-1}, which was the only oscillating band exhibited by the yeast, described above, having strong enough intensity to also be distinguishable in the spectra from mixtures. Furthermore and remarkably, it was the only signal in the spectra found to increase linearly with the growing cell content, thus overcoming the general attenuation effect.

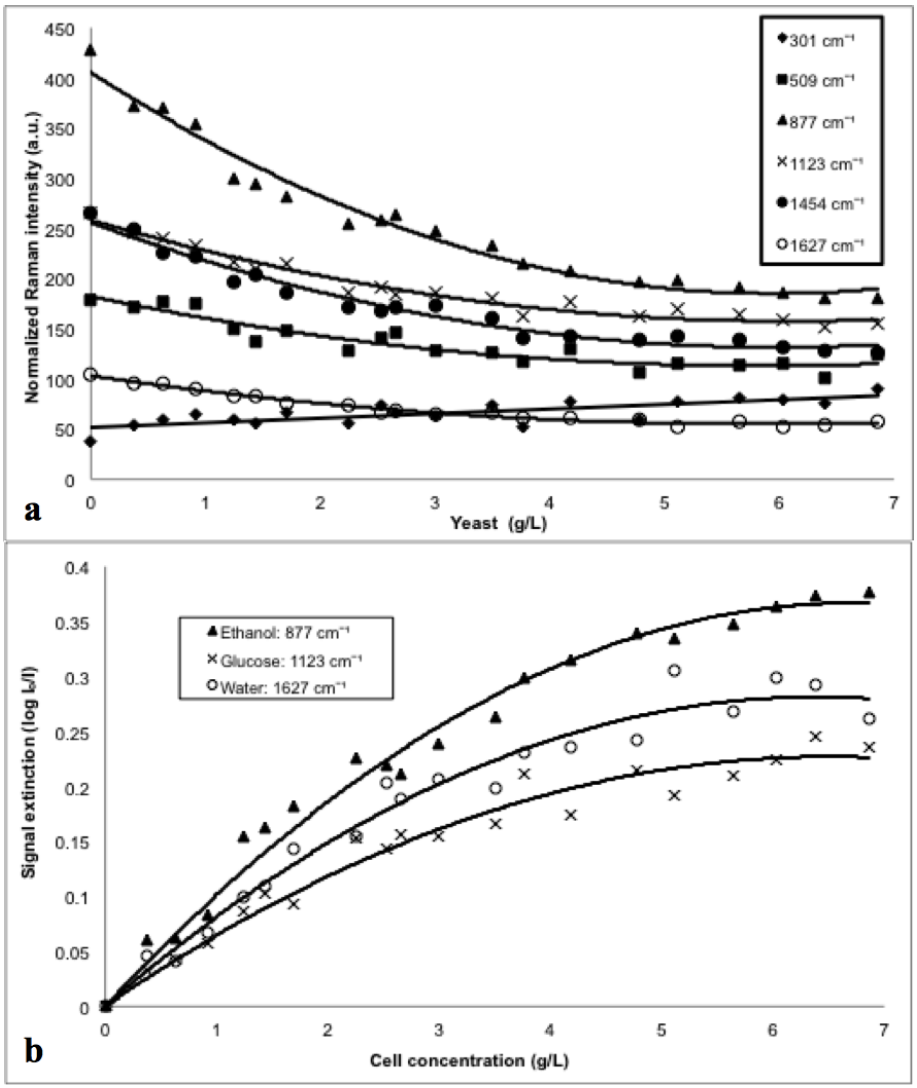

Fig. 4 The effect of gradually increasing content of light scattering yeast cells on Raman intensities **a)** originating from glucose (■, ×), ethanol (▲, ●), water (o) and yeast cells (♦) and the consequently attenuation of the signals **b)** seen in the samples containing 20g/L glucose and 70 g/L ethanol.

The scattering effect is also a common phenomenon in NIR spectra and corrections are usually done using Multiplicative Scatter Correction (MSC) or

the similar Standard Normal Variate (SNV) transformation. Both of these correction models are based on regressing spectral values on the average spectrum and used to estimate a wavelength dependent shifting baseline. For MSC or SNV to be used for spectral pre-processing, the scattering is expected to shift the spectra gradually with changing wavelengths [20]. A similar pre-processing method could in principle be used for correction of Raman scattering spectra, but such a relationship could not be deduced from our Raman data sets, due to the greater variance in the extinction of the Raman signal (Fig. 4b). The extinction was higher than typically seen for NIR, where the scattering basically introduces a uniform baseline shift. Some of this variance may be explained as the oscillating wavelength dependent Mie scattering (Fig. 3), which however is likely to change in pattern for a different excitation setup. For this reason of the expected variance in attenuation coefficient, the scattering effects of each of the few bands used for quantification were estimated individually instead of trying to apply some sort of smoothed curve or single value for calibration.

The changes in Raman signal extinction with increasing cell concentration are shown in Fig. 4b. Extinction values were determined from values displayed in Fig. 4a. Increasing amounts of cell particulates significantly attenuated the water band at ~1627 cm^{-1} as well as the most distinct Raman signals of ethanol at 877 cm^{-1} and glucose at 1123 cm^{-1}, despite of constant analyte concentrations. The Raman intensities, shown in Fig. 4a, were as expected not declining linearly as the amount of cell particulates increased. The relationship between yeast concentration and the extinction were moreover far from following the Beer-Lambert correlation. Light attenuation, due to the loss of photons in the light path, was significantly non-linear, throughout this range of yeast concentration. Instead, the correlation between particle concentration and light attenuation is complex. On the basis of the data, a quadratic equation gives an acceptable approximation describing the extinction as a simple quadratic function of the concentration of the scattering cell particulates *E(c)*, as in equation (3):

Ethanol, 877 cm^{-1}: $E(c) = 0.0081 \cdot c^2 - 0.1093 \cdot c$ $R^2 = 0.9823$ (2)

Glucose, 1123 cm^{-1}: $E(c) = 0.0054 \cdot c^2 - 0.0698 \cdot c$ $R^2 = 0.9650$ (3)

Water, ~1627 cm^{-1}: $E(c) = 0.0069 \cdot c^2 - 0.0881 \cdot c$ $R^2 = 0.9712$ (4)

Different logarithmic approaches were tested, but did not give better fit to the data set. Despite vigorous stirring of the cell suspensions during the measurements, the Raman intensity data at various yeast concentrations showed greater deviation from the trend line, no matter which mathematical correlation was chosen, than compared to the non-particulate calibration standard measurements of reference analytes shown in Fig. 1. Some inaccuracy may hence have to be taken into account

when using the internal standard method to correct for light extinction caused by Mie scattering from the cells. The displayed extinction changes at increasing cell concentrations are not equal, and the ethanol signal measured at 877 cm^{-1} is attenuated more than 50% relative to the glucose band at 1123 cm^{-1} in the data set. This is why the same extinction values for the internal water standard was not directly used to correct for all the other attenuated Raman signals in the model, but only to find cell concentrations needed to determine the attenuation for each analyte. This method is in contrast to some of the approaches in the literature [23, 24], where the extinction magnitude of the internal standard was assumed to be valid for all other peaks.

Fermentation Monitoring

Raman spectra of supernatant samples after removal of fluorescence background, taken throughout the yeast formation and fermentation process, are shown in Fig. 5a. The displayed changes in Raman intensities show only the effects of the conversion process, since the cells active in Mie scattering have been removed and the internal standard signal at around 1627 cm^{-1} from water was constant. The resulting signals at 1123, 514 and 1360 cm^{-1} were diminishing due to glucose consumption and the increasing ethanol concentrations giving rise to growing intensities at 877, 1046, 1080 and 1455 cm^{-1} are remarkably distinct and having minimal overlap. Spectra of the same samples with yeast cells, measured before separation, revealed light attenuation of the signals at all these wavenumber positions, including the internal standard Raman band from water (spectra not shown). This proves that the internal standard peak only was affected by scattering and not by the biochemical conversion. The signal induced by the growth of the yeast cell - expected to show up at 300 cm^{-1} - was unfortunately too weak and hidden in noise to retrieve any information about the yeast growth.

The attempt to estimate yeast cell concentrations on the basis of this internal standard Raman signal attenuation, at ~1627 cm^{-1} using the extinction equation for water, is depicted in Fig. 5b. After the inoculation (2 g/L) both concentrations measured by HPLC and by Raman showed the yeast growth from 2 g/L at inoculation to 7 g/L after 7 hours of fermentation. The effect of scattering is demonstrated by the disappearance of the signal when stirring of the reaction was stopped between 7 and 17 hours, resulting in cell precipitation. Compared to the cell concentration values measured by UV-VIS, the Raman determined growth curve seems less stable. Further noise reduction and improvement of cell estimations are likely to be achievable applying smoothening algorithms to the spectra. It must be emphasized, that this method to estimate cell concentrations only is applicable, as long as the cells are the only type of Mie scattering particulates present when a bioprocess is monitored, which is usually not the case. Nevertheless,

extinction of the internal standard Raman band can still be used to estimate the particulate content and to correct the attenuated analyte signals.

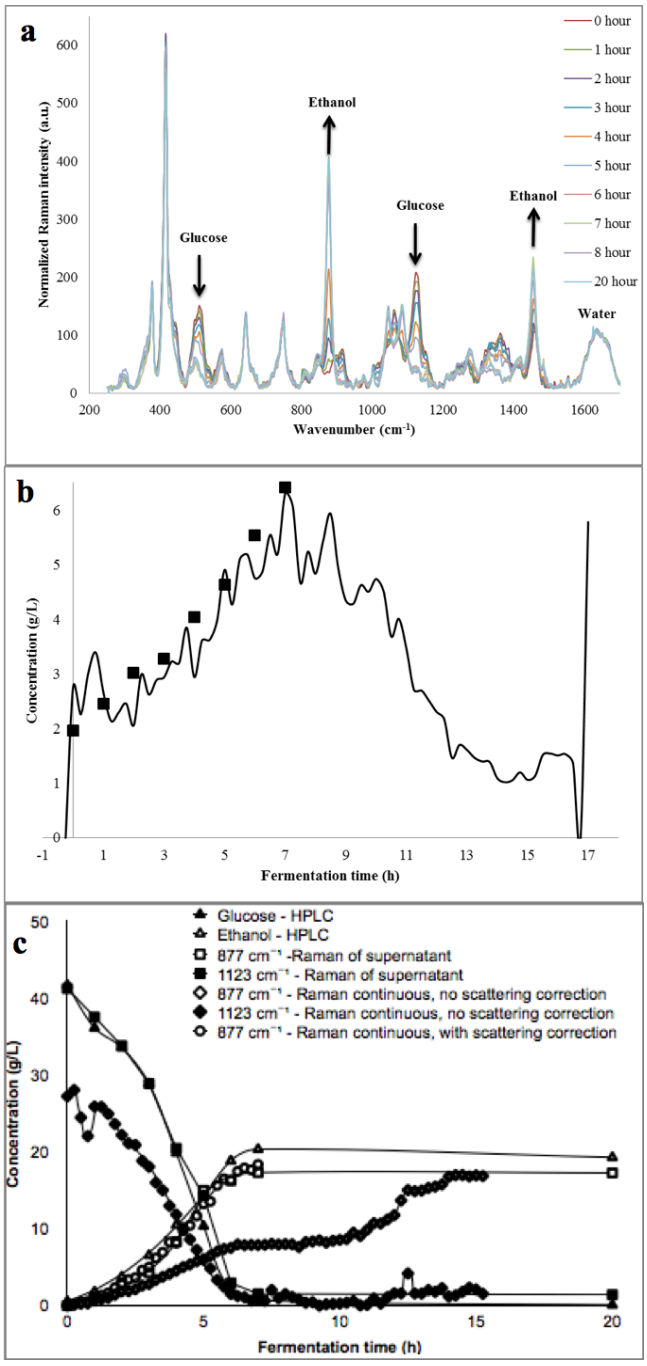

Fig. 5 Results of *S. cerevisiae* fermentation in YPD media containing 40 g/L glucose (reactor stirring was stopped between hours 8-17). Raman spectra of baseline-corrected supernatants of samples obtained during fermentation (**a**). The development of cell concentration determined by UV-VIS (■) and *in-situ* Raman (line) (**b**). The concentrations of solutes determined by HPLC and upon Raman signals at 877 cm^{-1} for ethanol (open symbols) and at 1123 cm^{-1} for glucose (closed symbols) (**c**). Compo-nent measurements were done by *in-situ* Raman (♦), scattering normalized *in-situ* Raman (●), subsequent Raman of supernatants (■) and HPLC (▲).

The significant improvements obtainable by using particle scattering correction of the Raman signals, applying the estimated cell concentrations, is seen in Fig. 5c. The method corrects the glucose and ethanol values with more than 50 % concentration offsets, thus correlating well with the values derived from Raman spectra of supernatant samples as well as from the external HPLC data. The benefit of this new scattering correction method is furthermore confirmed by the plots of the correlation between the measured and predicted values presented in Fig. 6.

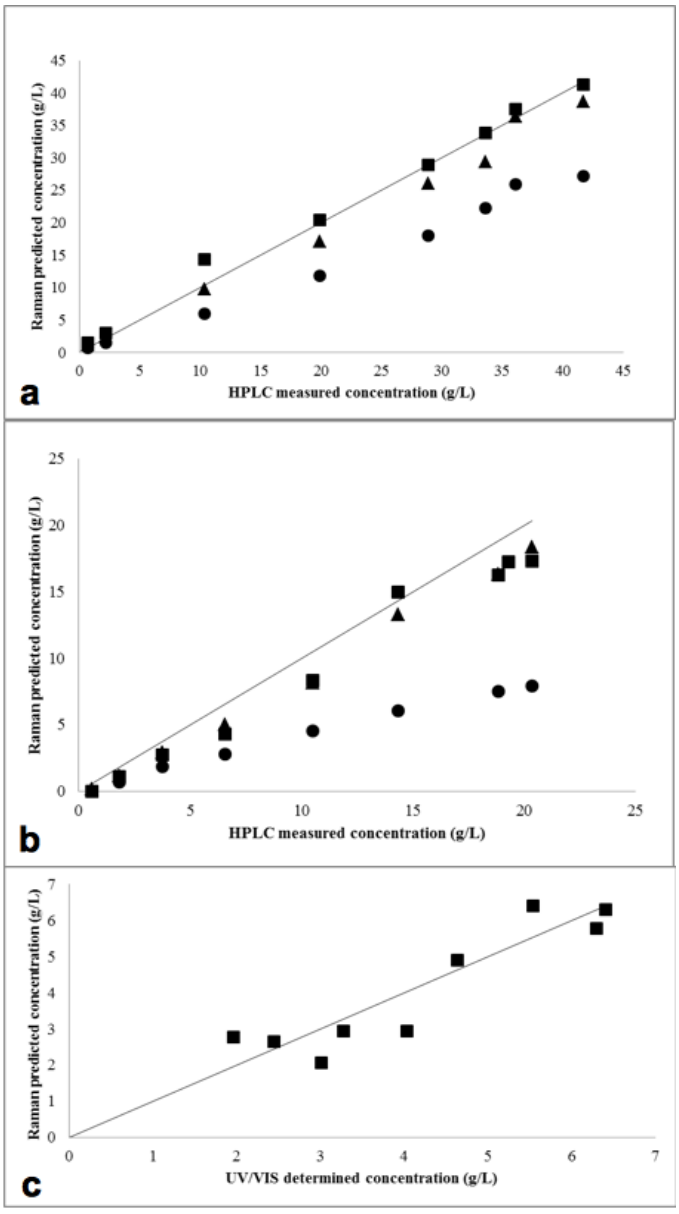

Fig. 6 Prediction of the content of glucose (**a**), ethanol (**b**) and yeast cells (**c**) during fermentation as obtained by use of Raman compared to HPLC or UV-VIS analysis. Measurement done by *in-situ* Raman (▲), scattering normalized *in-situ* Raman (●) and Raman of supernatants analyzed subsequently (■). Perfect correlations, are shown by the line of where predicted values equal measured values.

Thus simple scattering normalization significantly improves the quantification, although similar fermentation monitoring as conducted by other methods using PLS with Raman [14, 29] gave better overall accuracies, when compared to our RMSEP values (Table 1). Further accuracy improvements of the method can most likely be gained by also incorporating conventional preprocessing such as smoothening or by use of spectra derivatives. Chemometric models can account for non-linear correlations by for instance applying quadratic PLS instead of linear regression as described by Blanco [30]. Similar use of quadratic regressions in a multivariate model may also further improve the quantification, if scattering estimations and quantifications are still done separately.

Table 1 Results of the applied quantification model using Raman spectroscopy to predict the concentrations of glucose, ethanol and yeast cells by correcting for signal attenuation caused by Mie-scattering

	R^2	RMSEP
Glucose, 1123 cm^{-1}		
Supernatant	0.995	1.535
In-situ, no correction	0.991	8.973
Scattering corrected	0.990	2.357
Ethanol, 877 cm^{-1}		
Supernatant	0.981	1.892
In-situ, no correction	0.993	7.133
Scattering corrected	0.995	1.611
Yeast (water), ~1627 cm^{-1}	0.840	0.633

Conclusion

The presented study of Raman spectroscopic monitoring and quantification of a fermentation process showed significant extinctions of the Raman signals due to Lorenz-Mie scattering by the yeast cell particulates. The true scattering effect cannot be accounted for applying traditional chemometric models without initial pre-processing, since quantification of analytes by linear regression is not permitted in the presence of interfering components. The attenuation impact on the Raman intensities was shown not to be uniform throughout the examined spectral range and regions with oscillating extinction were observed. The simply procedure of

employing the same intensity reduction ratio of an internal reference to correct the scattering effects of all Raman signals is thus problematic. Instead, we are suggesting a two-step algorithm with an initial determination of the scattering factor and a subsequent quantification based on the peak intensity extinction calibration curve for each analyte. We have shown that robust concentration estimations and real-time monitoring can be achieved, without the need of further pre-processing of the Raman spectral data, such as calculating derivatives. It is even not needed to extend this simple univariate model in order to get robust results, as shown above.

Acknowledgements This work was part of the project "Development of a biorefinery concept for integrated production of biomedical, biochemical, feed and fuel from selected plant materials (BIOREF)" supported by the Danish Council for Strategic Research, grant no.: 09-065165.

References

1. Bakeev KA (2005) Process analytical technology: spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries. Wiley, Blackwell Publishing Ltd.

2. Beutel S, Henkel S (2011) In situ sensor techniques in modern bioprocess monitoring. Appl Microbiol Biotechnol 91:1493-1505. DOI 10.1007/s00253-011-3470-5

3. Becker T, Hitzmann B, Muffler K, Pörtner R, Reardon K, Stahl F, Ulber R (2007) Future aspects of bioprocess monitoring. White Biotechnology 105:249-293. DOI 10.1007/10_2006_036

4. Finn B, Harvey LM, McNeil B (2006) Near-infrared spectroscopic monitoring of biomass, glucose, ethanol and protein content in a high cell density baker's yeast fed-batch bioprocess. Yeast 23:507-517

5. Mazarevica G, Diewok J, Baena JR, Rosenberg E, Lendl B (2004) On-line fermentation monitoring by mid-infrared spectroscopy. Appl Spectrosc 58:804-810

6. Shih CJ, Lupoi JS, Smith EA (2011) Raman spectroscopy measurements of glucose and xylose in hydrolysate: Role of corn stover pretreatment and enzyme composition. Bioresour Technol 102:5169–5176

7. Stewart D, Wilson H, Hendra P, Morrison I (1995) Fourier-transform infrared and Raman spectroscopic study of biochemical and chemical treatments of oak wood (*Quercus rubra*) and barley (*Hordeum vulgare*) straw. J Agric Food Chem 43:2219-2225

8. Vankeirsbilck T, Vercauteren A, Baeyens W, Van der Weken G, Verpoort F, Vergote G, Remon JP (2002) Applications of Raman spectroscopy in pharmaceutical analysis. TrAC trends in analytical chemistry 21:869-877

9. Shope TB, Vickers TJ, Mann CK (1987) The direct analysis of fermentation products by Raman spectroscopy. Appl Spectrosc 41:908-912

10. Lieber CA, Mahadevan-Jansen A (2003) Automated method for subtraction of fluorescence from biological Raman spectra. Appl Spectrosc 57:1363-1367

11. Shaw AD, Kaderbhai N, Jones A, Woodward AM, Goodacre R, Rowland JJ, Kell DB (1999) Noninvasive, on-line monitoring of the biotransformation by yeast of glucose to ethanol using dispersive Raman spectroscopy and chemometrics. Appl Spectrosc 53:1419-1428

12. Sivakesava S, Irudayaraj J, Demirci A (2001) Monitoring a bioprocess for ethanol production using FT-MIR and FT-Raman spectroscopy. J Ind Microbiol Biotechnol 26:185-190

13. Abu-Absi NR, Kenty BM, Cuellar ME, Borys MC, Sakhamuri S, Strachan DJ, Hausladen MC, Li ZJ (2010) Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe. Biotechnol Bioeng 108:1215-1221

14. Whelan J, Craven S, Glennon B (2012) In situ Raman spectroscopy for simultaneous monitoring of multiple process parameters in mammalian cell culture bioreactors. Biotechnol Prog 28:1355-1362. DOI 10.1002/btpr.1590

15. Roggo Y, Chalus P, Maurer L, Lema-Martinez C, Edmond A, Jent N (2007) A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. J Pharm Biomed Anal 44:683-700

16. Wold S, Sjöström M, Eriksson L (2001) PLS-regression: a basic tool of chemometrics. Chemometrics Intellig Lab Syst 58:109-130

17. Pelletier M (2003) Quantitative analysis using Raman spectrometry. Appl Spectrosc 57:20-42

18. Liu C, Berg RW (2012) Non-linearity in Intensity versus Concentration Dependence for the Deep UV Resonance Raman Spectra of Toluene and Heptane. Applied Spectroscopy Reviews 48, issue 3:159-172

19. Latimer P (1982) Light scattering and absorption as methods of studying cell population parameters. Annu Rev Biophys Bioeng 11:129-150

20. Rinnan Å, Berg F, Engelsen SB (2009) Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends in Analytical Chemistry 28:1201-1222

21. Butler WL (1964) Absorption spectroscopy *in vivo* theory and application. Annual Review of Plant Physiology 15:451-460

22. Amesz J, Duysens L, Brandt D (1961) Methods for measuring and correcting the absorption spectrum of scattering suspensions. J Theor Biol 1:59-74

23. Aarnoutse PJ, Westerhuis JA (2005) Quantitative Raman reaction monitoring using the solvent as internal standard. Anal Chem 77:1228-1236

24. Picard A, Daniel I, Montagnac G, Oger P (2007) In situ monitoring by quantitative Raman spectroscopy of alcoholic fermentation by Saccharomyces cerevisiae under high pressure. Extremophiles 11:445-452

25. Lee HLT, Boccazzi P, Gorret N, Ram RJ, Sinskey AJ (2004) In situ bioprocess monitoring of *Escherichia coli* bioreactions using Raman spectroscopy. Vibrational spectroscopy 35:131-137

26. Xie C, Dinno MA, Li Y (2002) Near-infrared Raman spectroscopy of single optically trapped biological cells. Opt Lett 27:249-251

27. Graaff R, Aarnoudse J, Zijp JR, Sloot P, De Mul F, Greve J, Koelink M (1992) Reduced light-scattering properties for mixtures of spherical particles: a simple approximation derived from Mie calculations. Appl Opt 31:1370-1376

28. Walstra P (1964) Approximation formulae for the light scattering coefficient of dielectric spheres. British Journal of Applied Physics 15:1545-1552

29. Ávila TC, Poppi RJ, Lunardi I, Tizei PAG, Pereira GAG (2012) Raman spectroscopy and chemometrics for on-line control of glucose fermentation by *Saccharomyces cerevisiae*. Biotechnol Prog 28:1598-1604. DOI 10.1002/btpr.1615

30. Blanco M, Coello J, Iturriaga H, Maspoch S, Pages J (2000) NIR calibration in non-linear systems: different PLS approaches and artificial neural networks. Chemometrics Intellig Lab Syst 50:75-82

MANUSCRIPT II

Monitoring Lignocellulosic Bioethanol Production Processes using Raman Spectroscopy

Jens A. Iversen, Birgitte K. Ahring
Submitted for publication in Bioresource Technology

Monitoring lignocellulosic bioethanol production processes using Raman spectroscopy

Jens A. Iversen[1,2], Birgitte K. Ahring[1,2,*]

[1]Section for Sustainable Biotechnology, Aalborg University, A. C. Meyers Vænge 15, 2450 Copenhagen SV, Denmark
[2]Center for Bioproducts and Bioenergy, Washington State University Tri-Cities, 2710 Crimson Way, Richland, WA 99354, USA

*Corresponding author. Phone: +1 509-372-7682; Email: bka@wsu.edu

ABSTRACT

Process control automation in the emerging biorefinery industry may be achieved by applying effective methods for monitoring compound concentrations during the production processes. This study examines the application of Raman spectroscopy with an excitation wavelength of 785 nm and an immersion probe for *in situ* monitoring the progression of pretreatment, hydrolysis and fermentation processes in the production of lignocellulosic ethanol. Raman signals were attenuated by light scattering cells and lignocellulosic particulates, which the quantification method to some degree could correct for by using an internal standard in the spectra. Allowing particulates to settle by using a slow stirring speed further improved results, suggesting that Raman spectroscopy should be used in combination with continuous separation when used to monitor process mixtures with large amounts of particulates. The root mean square error of prediction (RMSE) of ethanol and glucose measured in real-time was determined to be 0.98 g/L and 1.91 g/L respectively.

Keywords:
Raman spectroscopy, lignocellulose, pretreatment, hydrolysis, ethanol fermentation.

1. Introduction

Concerns about increasing carbon dioxide emissions and the depletion of the remaining global oil reserves provide ample reasons to develop alternatives to fossil fuels. A promising renewable resource for the production of fuel for transportation exists in the form of bioethanol produced from the carbohydrates present in lignocellulosic biomass from abundant agricultural residues (Menon and Rao, 2012). The production of 2nd generation bioethanol involves some sort of pretreatment of biomass at elevated temperatures followed by enzymatic hydrolysis to release the sugars bound in cellulose and hemicellulose for the subsequent fermentation process. These processes can be operated more efficiently or even automated if production parameters can be monitored real-time instead of retrieving vital process information through manual time-consuming sampling and subsequent off-line analysis.

Various forms of absorption spectroscopy technologies present alternative *in situ* analysis methods to the commonly used high performance liquid chromatography (HPLC) or gas chromatography (GC) for measuring concentrations of multiple reaction components simultaneously (Bakeev, 2005; Beutel and Henkel, 2011). Near-infrared (NIR) and mid-infrared (MIR) spectroscopy have been used to monitor the fermentation processes (Finn, Harvey and McNeil, 2006; Mazarevica et al., 2004; Sivakesava, Irudayaraj and Demirci, 2001). However, these analysis techniques are complicated by the broad overlapping peaks seen in NIR spectra or by the strong absorption of water in the MIR spectral range, unless used with an attenuated total reflection unit (ATR). In contrast, Raman spectroscopy is not based on absorption but on inelastic scattering of monochromatic light with a wavelength of the reemitted photons determined by the Stokes energy shift of the molecule (Bakeev, 2005; Stewart et al., 1995). The need for a probe with a fixed light path within a gap is circumvented if the spectroscopy method is not based upon light transmission, since the same lens can be used for both the excitation light beam entering the sample and collecting the backscattering Raman signals. This makes Raman spectroscopy capable of analyzing viscous reaction mixtures or ones with high solids content. Raman spectra generally have distinct narrow peaks and the interference from water in aqueous samples is weak, unlike MIR. However, Raman signals are very weak and the laser-generated Rayleigh scattered light and fluorescence background noise must be eliminated from the collected backscattered signal before spectral interpretation (Bocklitz et al., 2011; Vankeirsbilck et al., 2002). Process monitoring using Raman spectroscopy has not yet gained the same popularity as NIR, but this may change with; recent instrument improvements through the implementation of better lasers, new spectrometer designs, improved detectors with increased sensitivity and applying computers and software, which are able to overcome the challenge of fluorescence (Bakeev, 2005).

Altogether, Raman spectroscopy is a promising monitoring technology, with various configurations capable of analyzing concentrations of substrates and products during fermentation processes (Sivakesava, Irudayaraj and Demirci, 2001; Shaw et al., 1999; Shope, Vickers and Mann, 1987). Quantifications with high degrees of accuracy have recently been achieved by monitoring multiple process parameters simultaneously in yeast fermentations and mammalian cell culture bioprocesses using *in situ* Raman spectroscopy and chemometrics (Abu-Absi et al., 2010; Ávila et al., 2012; Gray, Peretti and Lamb, 2013; Li et al., 2013; Uysal et al., 2013; Whelan, Craven and Glennon, 2012). Chemometrics combines spectroscopy with statistical methods and most often includes multivariate data analysis, which enables quantitative analysis of complex data with multiple independent variables for each dependent variable. It has been used extensively with NIR where absorption bands tend to overlap (Roggo et al., 2007; Wold, Sjöström and Eriksson, 2001). Raman spectra with its narrower peaks should in some cases enable sufficient quantification using simple univariate calibration models, thus reducing the need for complicated chemometrics. Quantification of concentrations by applying Raman data with single or multi-linear regression is reasonable, since the intensity of Raman scattered light is directly proportional to the number of molecules (Pelletier, 2003). This linear correlation between concentration and Raman intensity can; however, deviate by resonance Raman scattering if the excitation wavelength is near an electronic absorption band (Liu and Berg, 2012). The linear relationship between Raman intensity and concentration can also be disturbed by the presence of microbial cells or signal-attenuating particles due to diffuse Lorenz-Mie scattering responsible for optical density (Butler, 1964). For a long time the challenge of removing the fluorescence background signal was considered a major obstacle in working with Raman spectroscopy. However, with better software and algorithms to deal with this; the next important issue to solve may be the development of ways to overcome the effects of light scattering particulates.

The relationship between light extinction and the concentrations of light-scattering particulates can be expressed by a simple linear correlation in dilute suspensions. In more turbid media, attenuations of Raman signals are more mathematically complicated to quantify and correct, because of multiple scattering of photons and shadow effects. Instead, the overall magnitude of Mie scattering can be revealed by using a component unchanged through the reaction and with peaks not shared by other components in the reaction mixture as an internal standard. The changes in intensity of the internal standard can only be explained by changes in scattering, if the concentration remains constant; which can be used to estimate the reductions of all other Raman signals through a relative simple scattering correction method (Aarnoutse and Westerhuis, 2005). This method of correcting for signal extinction using an attenuation coefficient was demonstrated by using sulfate in the media as an internal standard when monitoring yeast fermentation (Picard et al., 2007).

The amount of light-scattering particulates is limited in fermentations with defined media compared to lignocellulosic bioethanol production processes, where large amounts of biomass debris must be anticipated. Therefore, monitoring highly heterogeneous lignocellulosic biofuels production processes with Raman spectroscopy is complicated by the attenuation of the Raman signals due to particulates. There have been a number of publications investigating biomass structures or spectral features from components in lignocellulosic biomass using Raman spectroscopy (Stewart et al., 1995; Adapa et al., 2011; Agarwal, Reiner and Ralph, 2012; Barsberg et al., 2006; Xu et al., 2013). Until recently, minimal research has been done on how to quantify component concentrations during biorefinery processes using Raman spectroscopy. Glucose and xylose have been measured off-line simultaneously in samples taken from corn stover hydrolysate applying extraction procedures to improve detection (Shih, Lupoi and Smith, 2011). In another recently published study, HPLC data of ethanol and glucose concentrations obtained from a lignocellulosic fermentation of switch grass hydrolysate correlated well with real-time Raman derived results (Ewanick et al., 2013). These two studies were based on monitoring fermentation of the separated liquid fractions or filtered samples without interference from light scattering biomass debris. In future large scale biorefineries, continuous removal of all scattering solids in the production processes is not a realistic scenario and some degree of particulates should be expected and accounted for in the method used for real-time monitoring.

The purpose of our research was to examine the application of Raman spectroscopy in monitoring conversion processes involved in the production of lignocellulosic bioethanol, considering disturbance from particulates. Experiments were conducted with the aim of developing methods for real-time quantification of substrate, product, and inhibitor concentrations during dilute acid pretreatment, enzymatic hydrolysis, and fermentation with *Saccharomyces cerevisiae*. The microcrystalline cellulose Avicel was used as a model substrate for the initial hydrolysis experiment. The analysis of the model substrate was followed by the analysis of more complicated lignocellulosic material by monitoring pretreatment, hydrolysis, and fermentation of sugarcane bagasse. Applying complicated multiple linear regression or partial least squares analysis was avoided; and quantification was carried out with only simple linear regression with one independent variable after initial correction of attenuated Raman intensities. We demonstrate real-time monitoring of a *Saccharomyces cerevisiae* fermentation process using a Raman instrument equipped with a sapphire ball probe designed to minimize interference by particulate as well as withstand the harsh environment during sterilization. We examined the prospect of using Raman spectroscopy for monitoring concentration changes during the pretreatment, hydrolysis, and fermentation processes. Scattering correction of the spectra by using an internal standard in the media was performed before quantification of concentrations.

2. Materials and Methods

2.1. Avicel Hydrolysis and Fermentation

A one liter suspension of 40 g/L Avicel® PH-101 (~50 μm particle size) microcrystalline cellulose, (Sigma-Aldrich, USA), 20 g/L proteose peptone, and 10 g/L yeast extract, was adjusted to pH 5.5. This suspension was hydrolyzed by adding 10 mL (0.36 mL product/g cellulose) of Accelerase® 1500 (DuPont, USA) at 50 °C and 100 rpm for 48 hours in a 1.4 L Multifors bioreactor (Infors-HT, Switzerland). Subsequent to hydrolysis, the bioreactor with the suspension was sanitized at 95 °C for 30 min and kanamycin (50 mg/L) was added before inoculation with *Saccharomyces cerevisiae* dry yeast (3 g/L) for fuel ethanol (Thermosacc Dry, Lallemand Ethanol Technology A/S, Denmark). The initial fermentation was carried out at 35 °C and 100 rpm for 10 hours with sampling every hour. The fermentation was prolonged to 30 hours by adding 10 g of glucose at 6, 10 and 21 hours fermentation time. During hydrolysis and fermentation, real-time monitoring of glucose and ethanol concentrations was carried out using Raman spectroscopy acquiring data points for every 15 min.

2.2. Pretreatment hydrolysis and fermentation of sugarcane bagasse

Sugarcane bagasse was from Lafourche Sugars, LLC, Thibodaux, LA, via Dr. Edward Richard from the USDA. Hammer milled (1 mm particle size) sugarcane bagasse was pretreated utilizing a low-temperature method with dilute sulfuric acid as catalyst. A 1 L suspension containing 10% w/v total solids of sugarcane bagasse and 0.4% w/v sulfuric acid was pretreated on a heated magnetic stir plate (RCT Basic, Ika®-Werke GmbH, Staufen, Germany) at 95 °C and 300 rpm for 48 hours in the 1.4 L Multifors bioreactor. Real-time Raman monitoring during pretreatment and Raman measurements were also performed on samples taken from the pretreatment three times a day.

Subsequent to pretreatment, 3.5 g sodium hydroxide pellets and 5 M sodium hydroxide were added to the one liter suspension for final adjustment to pH 4.8. Protease peptone was added (10 g/L) and the suspension was sanitized at 90 °C for 60 min before hydrolysis. The pretreatment was followed by hydrolysis in the Multifors bioreactor and was performed at 50 °C for 100 hours with 2 mL Cellic CTec2 (Novozymes A/S, Bagsvaerd, Denmark) added, corresponding to 10 mg protein/g cellulose. To limit the disturbance from light scattering particles during continuous Raman monitoring, the stirring speed was reduced to 100 rpm during hydrolysis, thereby continuously precipitating the non-solubilized particulates and increasing transparency in the top 5cm of the bioreactor. The hydrolysate was inoculated (3 g/L) with dry yeast and fermentation was conducted at 35 °C and 100

rpm for 7 hours with samples taken once every hour. The Raman spectra were acquired every 15 minutes from the top layer of the hydrolysate, an area free of most lignocellulosic particulates.

2.3. HPLC Analysis

A High Performance Liquid Chromatography instrument (HPLC) equipped with a refractive index (RI) detector was used to measure sample concentrations. The HPLC (Ultimate 3000, Dionex, CA, USA) was operated with an Aminex HPX-87H column (300 x 7.8 mm) combined with a Cation-H guard column (30 X 4.6 mm) from Bio-Rad Laboratories, CA, USA. Sample supernatants of 10 µL were run through the column isocratically for 26 min with a flow rate of 0.6 mL/min at 60 °C using 4 mM H_2SO_4 as eluent. Supernatants of 1 mL samples with 10 µL 1 M H_2SO_4 were run on the HPLC and concentrations of separated sample component were determined using mixtures containing glucose, xylose, arabinose, glycerol, ethanol, acetic acid and lactate (concentrations of 0.5 to 20 g/L).

2.4. Raman Analysis and Quantification

Spectroscopy analysis and monitoring was carried out with a ProRaman-I instrument (Enwave Optronics Inc.) equipped with a cooled charge-coupled device (CCD) detector operated at -60 °C and a deep red 785 nm laser providing a power output during measurement of 300 mW. Sample measurements or real-time monitoring was conducted with an immersion fiber optic probe (12 mm diameter) equipped with a ball sapphire lens designed to withstand harsh chemical and physical environments. The acquired spectra (250-2350 cm^{-1} coverage and 6 cm^{-1} resolution) were processed with the EZRaman Reader ver. 8.2.2 spectra managing software. Calibration of the wavenumber axis scale was carried out using the solid calibrating sample provided by Enwave Optronics. The rate of stirring was kept high enough to avoid settling of biomass debris or yeast cells on the probe lens.

The spectra were generally obtained from references, samples or *in situ* reactions by averaging 45 Raman spectral data sets, each acquired through 20 seconds of exposure. This resulted in acquisition of data points for every 15 min during the real-time measurements of hydrolysis and fermentation of both the Avicel and the sugarcane bagasse. Raman measurements of samples taken during the pretreatment of bagasse were, however, only obtained through 5 seconds of acquisitions in order to reduce the increased fluorescence background due to particulates and thereby maintaining the maximum peak signals below the saturation level of the detector. These samples were allowed to settle for a few minutes before measurement – thereby reducing most of the impact from the biomass debris. The samples taken during fermentation of bagasse hydrolysate were centrifuged at 5000 rpm for 60 seconds before acquiring Raman from the particulate-free supernatants. This allowed for minimal fluorescence interference. The bioreactor and samples were

shielded during all measurements to avoid interfering signals from ambient light. All spectra were normalized by subtracting the fluorescence background through the application of a polynomial algorithm through the EZRaman software's autobaseline 2 setting, except when displaying the impact of fluorescence (Fig. 2).

Subsequent to normalization of spectra, quantification of individual components in the spectra was achieved by correlating Raman signals originating from single components in the sample mixtures to standard curves derived from normalized spectra. However, this method is without further processing of spectra only accurate for samples absent from light scattering particulates. Thus an initial scattering correction was applied when cells or biomass particles were present in samples or during real-time monitoring of hydrolysis or fermentation. Scattering correction was carried out using either the water peak at 1630 cm^{-1} or sulfate as internal standards – with the assumption that the concentration of these would be constant during biochemical conversions. The ratio between the attenuated signal and the unscattered Raman signal of the internal standard was used to estimate the extinction magnitude of all other Raman signals. After the scattering correction, concentrations of glucose and ethanol were predicted by standard linear regression in accordance to glucose and ethanol calibration curves prepared in the pretreated sugarcane bagasse. In summary, concentration quantification of samples and standards was conducted through the following steps:

1. Normalization of spectra through the removal of the fluorescence background with auto-polynomial subtraction
2. Estimation and correction of the attenuated Raman signals using the reduction of the normalized Raman signal of the internal standard as the common correction factor
3. Baseline adjustment of the scattering corrected Raman values by subtracting intensity values corresponding to concentrations equivalent to zero
4. Determination of concentrations through univariate linear regression of Raman intensities using only one independent variable

No further preprocessing typically applied with NIR quantification methods such as Savitzky-Golay smoothening of the spectra or spectral derivative pretreatment was performed with the Raman data sets. Chemometric calibration techniques involving multivariate analysis using principal component analysis (PCA) or partial least squares (PLS) were also omitted.

3. Results and discussion

3.1. Identification of important components in lignocellulosic material

In order to use Raman spectroscopy to analyze and monitor the conversion of lignocellulosic material into biofuel, initial identification of the Raman signal profile of single components is necessary. Fig. 1 displays Raman spectra of solutions of some of the most important components present during pretreatment, hydrolysis and fermentation of lignocellulosic material, including glucose, xylose, ethanol, acetic acid and furfural as well as

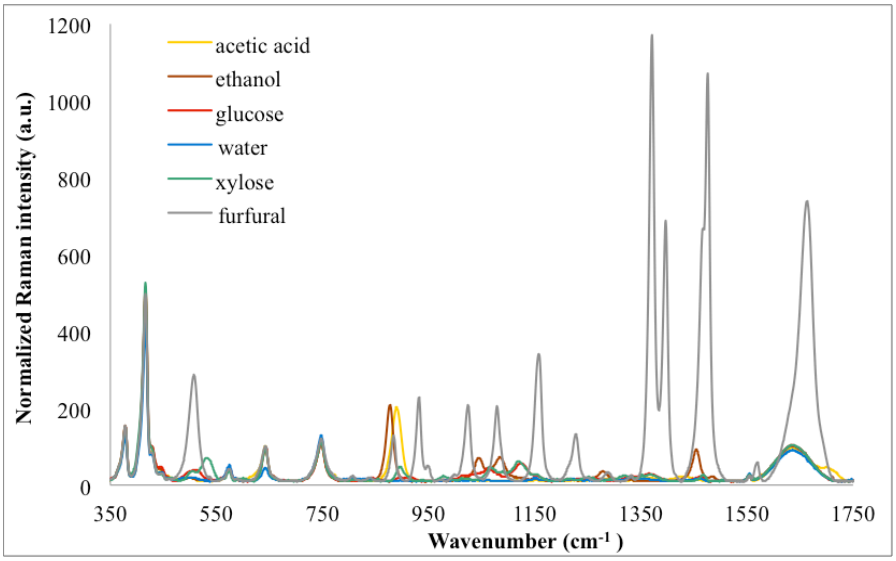

Fig. 1. Fluorescence subtracted Raman spectra in the fingerprint region of water and 10 g/L reference solutions of glucose, xylose, ethanol, acetic acid and furfural.

water without solutes. Each of these components has some bands in their Raman spectra, not shared by the other molecules, thereby yielding nearly ideal distribution of Raman peaks in the spectra of biomass hydrolysate. However, xylose and glucose share most of their main Raman scattering wavenumber regions in the spectra except peak present at 530 cm^{-1} from xylose. Raman scattering from the sapphire crystal in the probe is responsible for the signals seen at 378, 416, 642 and 749 cm^{-1} (Colomban and Havel, 2002). The most important and distinct bands of these molecules are listed in Table 1.

Furfural stands out as having many peaks, some with the strongest Raman intensities among all peaks found in the reference spectra, more than 5 times higher than ethanol or acetic acid, and more than 10 times the intensities of the sugars at the same concentrations. This is because Raman scattering originating from the

vibrations of less polar and symmetric molecular bonds results in the strongest intensities. The carbon-carbon double bonds in furfural thus give particularly strong Raman signals (Stewart et al., 1995). Similarly, Raman signals originating from C-C bonds in ethanol and acetic acid at 877 and 890 cm^{-1}, respectively, also generate strong Raman bands compared to peaks emitted from the polar bonds in glucose or xylose. This demonstrates why Raman spectroscopy has some advantages compared to NIR in relation to analysis of biofuels and inhibitors, since detection and measurement of non-polar hydrocarbon chains is possible at relatively low concentrations using Raman. Linear correlation between the concentrations of molecules of interest for quantification and the Raman intensities of corresponding main Raman bands was demonstrated in standard curves (curves not shown). The standards curves were prepared in the supernatant derived from the pretreated sugarcane bagasse and the linear regressions of the compounds showing strong correlation (Table 1) were used for concentration determinations of Raman measurements. The coefficients are determined by the Raman cross-sections of the components.

Table 1. Raman bands and intensities of important components in lignocellulosic hydrolysate*

Component	Wavenumber, cm^{-1}	Linear regression coefficient of the correlation between intensity and concentration, each from 10 references samples	
Glucose	**1124**, 510	I (c) = 4.2·c	R^2 = 0.9940
Xylose	**530**, 1120, 897		
Ethanol	**877**, 1455, 1045	I (c) = 21.8·c	R^2 = 0.9968
Acetic acid	**890**	I (c) = 18.4·c	R^2 = 0.9928
Furfural	**1371**, 1476, 1663		
Water	**1630**		

* Band intensities derived from the average of 45 spectra, each with 20 second exposure time

Comparisons of the raw spectra and the normalized spectra of samples with and without debris show how both the fluorescence and the Raman signal was affected by biomass particulates in the dark brown colored lignocellulosic hydrolysate (Fig. 2). The overall intensities in the raw spectrum of a pretreated sugarcane bagasse

sample was more than 50 times higher when it contained lignocellulosic particulates as compared to the raw spectrum given when measuring the supernatant of the same sample (both spectra shown at the top of the figure). This difference in signal intensity is mainly due to fluorescence originating from the suspended particulates, which is revealed by the normalization process, whereby the fluorescence part is subtracted from both spectra. Prior to separation, the fluorescence constituted for more than 95% of the full signal in the raw spectra, which was reduced to about 50% by removing the biomass debris. After the subtraction of the fluorescence contribution to the raw spectra, the remaining Raman signal itself was also shown to be affected by particulates, with Raman signal reduction of almost 95%, due to light scattering and hence attenuation of signal (not shown). For this reason, initial attenuation correction of the normalized Raman signals was conducted before quantification of the concentrations of the compounds of interest. This was done by using peaks originating from compounds maintaining constant concentrations during reactions, such as water or sulfate. In order to avoid exceeding the saturation level of the CCD detector of 60,000 a.u., acquisition times were reduced from 20 seconds, most commonly used throughout this study, to 5 seconds, when measuring mixtures containing particulates.

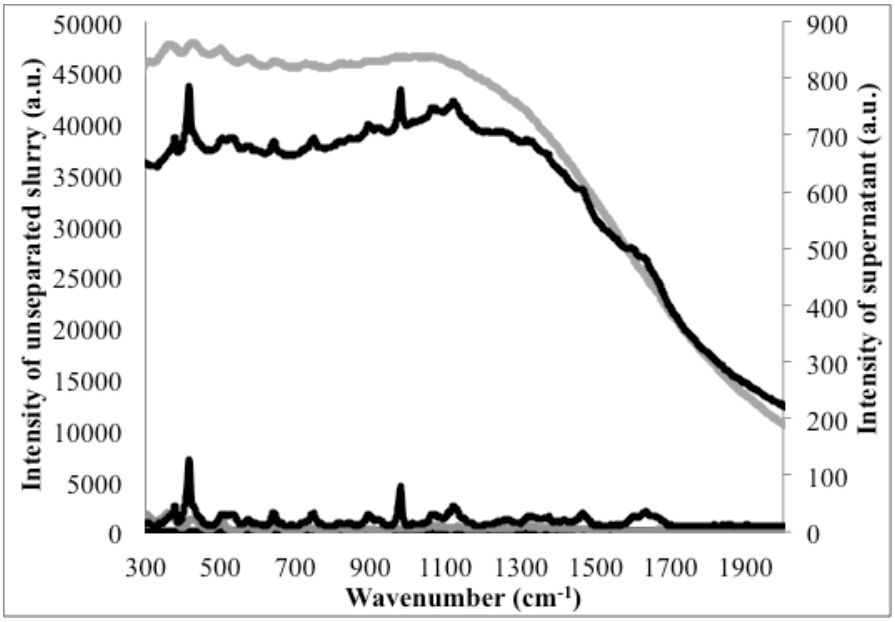

Fig. 2. Raman spectra of pretreated and hydrolyzed sugarcane bagasse from samples with biomass debris (grey, left y-axis) and the supernatant of the same sample without particles (black, right y-axis) with the raw spectra including the fluorescence contribution depicted at the top of the figure and the normalized spectra without the background fluorescence at the bottom. Raman spectra were derived from averages of 45 acquisitions, each with 4 seconds of exposure time.

The undesirable fluorescence background signal shown above to mainly be exhibited by the particulates is assumed in most part to be caused by lignin remaining in the solid biomass debris during pretreatment. Lignin with its complex phenolic heteropolymer structure exhibits strong auto-fluorescence in the visible and as well as IR regions compared to carbohydrates (Radotić et al., 2006). Only small amounts of soluble lignin were expected to be present in the hydrolysate after the pretreatment at low temperature below 100 °C, which is shown through the significant reduction in fluorescence after removal of the lignin-containing solids. At the higher temperatures above 150 °C used during more common pretreatment processes, more lignin is expected to be solubilized and present in the liquid fraction. More studies using lignin references are however needed to give a more comprehensive understanding of the full extend of the fluorescence background noise caused by lignin in regards to monitoring using Raman. As a consequence of the suspended solids causing higher fluorescence and attenuating scattering of Raman signal, real-time monitoring by Raman may nevertheless be improved by somehow continuously reducing the amount of disturbing biomass particulates during measurement.

3.2. Avicel Hydrolysis and Fermentation

Avicel, microcrystalline cellulose, was used as a model substrate for the monitoring and analysis of hydrolysis applying Raman spectroscopy (Fig. 3). The glucose concentration reached 22 g/L after the 50 hours of hydrolysis. Only modest changes were seen in the supernatant spectra of samples taken while the polymers were hydrolyzed to sugar monomers. Decreasing Raman intensity at 1095 cm^{-1} was the most apparent change (not shown). However, this was not apparent in spectra derived when the unseparated opaque mixtures were monitored real-time. The progress of the hydrolysis could be monitored to some degree, but quantity values could not be obtained. The few covalent bonds modified during hydrolysis of cellulose result in only minor spectral changes and the presence of particulates easily disturbs any spectral trends. Therefore, real-time measurement of a lignocellulosic hydrolysis reaction using Raman spectroscopy is expected to be difficult and inaccurate. Extensive use of multivariate spectral analysis combined with chemometrics may give some monitoring capabilities, but a low level of quantification inaccuracy is still expected to be difficult to overcome. In contrast, the subsequent Raman monitoring during fermentation of the hydrolysate revealed more promising results. The light-scattering particles that attenuated all Raman signals were corrected for by using the reduction of the water peak at 1630 cm^{-1} as correction factor for the other peak intensities. Ethanol concentrations determined by Raman correlated well with HPLC data upon scattering correction of spectra, as seen in Fig. 3. However, monitoring glucose at any of the corresponding wavenumbers was more difficult in the presence of particles and the estimated concentrations determined by Raman did not align well with HPLC data, in particular below 10 g/L. The prediction of glucose with its Raman peaks tending to

be partly overlapped by peaks originating from other compounds (Fig. 1) would likely benefit from applying multivariate analysis techniques. This is in contrast to the reasonable predictions of ethanol when applying only its distinct peak (877 cm^{-1}) in the univariate regression model.

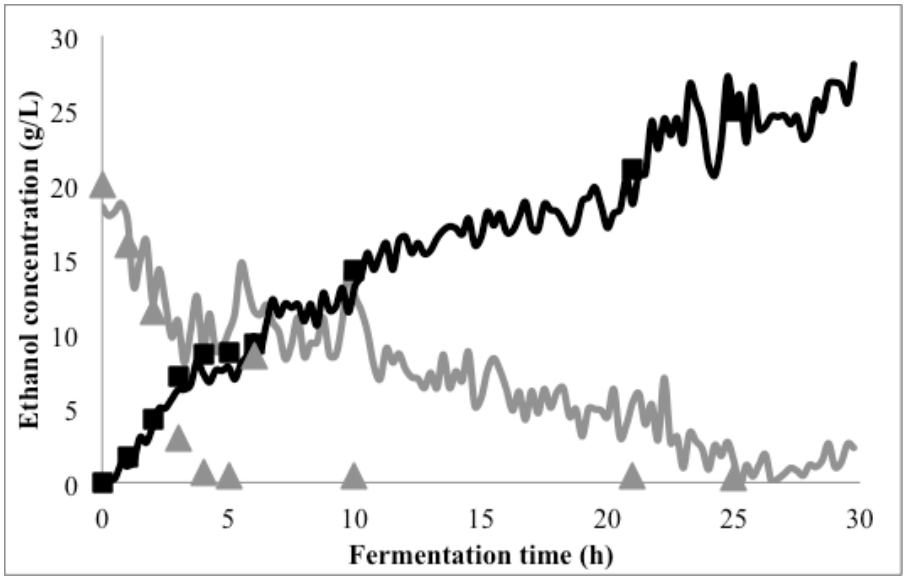

Fig. 3. Fermentation of the 10% Avicel hydrolysate by *S. cerevisiae*. Concentrations were monitored real-time applying scattering corrected and normalized Raman signals measured at 877 cm^{-1} for ethanol (black) and at 1123 cm^{-1} for glucose (grey). The subsequent off-line supernatant HPLC measurements of are displayed as (■) for ethanol and (▲) for glucose. Fermentation time was extended by addition of 10 g glucose at 10 and 21 hours of fermentation, each time after sampling.

3.3. Pretreatment of Sugarcane Bagasse

During dilute acid pretreatment of lignocellulosic biomass, acetic acid and xylose are the main components released in the suspension as the hemicellulose is solubilized. Real-time pretreatment monitoring would be possible if the released xylose or acetic acid can be measured using Raman spectroscopy. After a 48-hour pretreatment of sugarcane bagasse, concentrations of xylose and acetic acid reached final concentrations of 15.4 g/L and 3.9 g/L respectively. Increasing Raman intensities from xylose at mainly 530 cm^{-1} or acetic acid at 890 cm^{-1} were thus anticipated. Fig. 4a shows some of the spectra of precipitated samples taken throughout the pretreatment and concentrations of acetic acid. Intensities of xylose at 530 cm^{-1} seemed to increase, but the increase from acetic acid at 890 cm^{-1} seemed even more significant and consistent with the progress of the reaction. The

Raman signal observed at 980 cm^{-1} originated from sulfate present in the hydrolysate due to the addition of sulfuric acid. Sulfate was not involved in the pretreatment, hydrolysis or fermentation processes and the sulfate concentration and corresponding Raman intensity was thus assumed to remain constant throughout these processes. The Raman intensity from sulfate was only expected to change due to overall attenuation of Raman signals in the presence of light scattering particles, thereby allowing the Raman signal from sulfate to be used as an internal standard for general scattering correction of spectral values before quantification. Applying the distinct peak at 980 cm^{-1} originating from sulfate as internal standard is estimated to be more reliable instead of applying the broad water band at 1630 cm with possible interference from the furfural formed during pretreatment. This method of scattering correction cannot give as accurate quantification as obtainable when supernatants of samples exempt from all particulates is measured, but this form of in-line monitoring still gives valuable information about concentrations and the progress of the reaction.

Following the scattering correction, the calculated acetic acid concentrations yielded values resembling the progress of HPLC data (Fig. 4b). A similar attempt to find spectral changes during the pretreatment to estimate concentrations of the released xylose or formed furfural was not successful. The high amount of particulate matter generated too much noise for exact quantification of xylose to be obtained from the unseparated samples. Improved monitoring of xylose release could have been achieved if supernatants of the samples were measured instead. Acetic acid concentrations may seem difficult to estimate accurately, Raman spectroscopy with a probe submerged directly into a pretreatment reaction would nevertheless be a useful tool to follow the progression of the pretreatment reaction in real-time. In contrast to hydrolysis or fermentation, the pretreatment reaction is a significantly more complicated degradation process, with alterations of numerous covalent bonds in more unpredictable ways; which implies many spectral changes during the reaction. These spectral changes were also seen in spectra of fully separated supernatants of the samples and cannot be contributed merely as noise from particulates. The presented monitoring results were made upon measurements carried out on pretreatment samples after being cooled down to room temperature and thus cannot be considered real-time monitoring at the actual pretreatment temperature. However, this first attempt to monitor the progress of pretreatment serves as a preliminary step to assess the potential application of Raman spectroscopy in more common pretreatment reactions at elevated temperatures. A typical pretreatment of lignocellulosic biomass taking place at 140 – 180 °C would require calibration to take place in similar temperature range, if a real-time monitoring setup with the immersion probe in the pretreatment reactor would be applied, since the intensity of Raman is temperature dependent (Pelletier, 2003). This is only possible if the immersion probe over time can prove capable to withstand the environment in the pretreatment system.

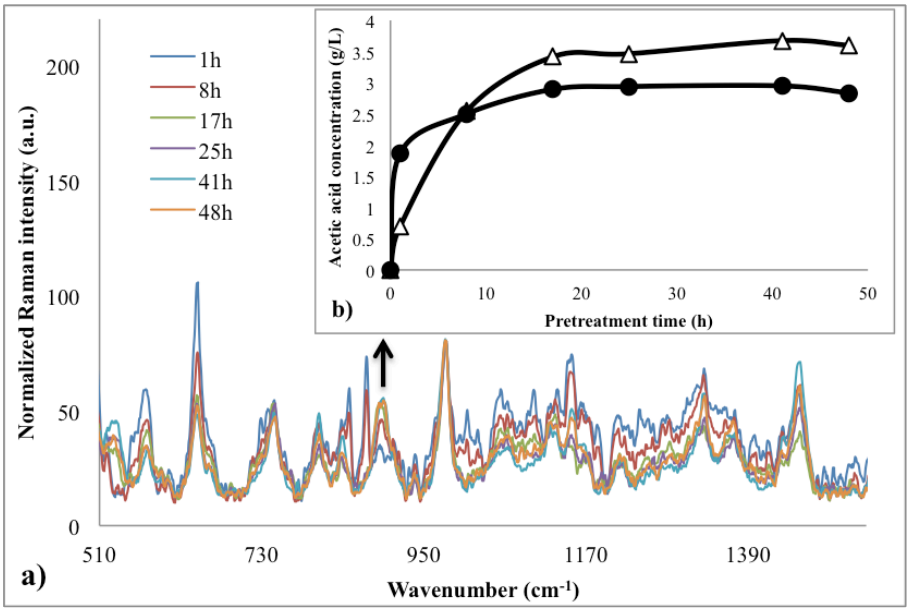

Fig. 4. Progress of the sugarcane bagasse pretreatment. a) Normalized and scattering corrected Raman spectra of unseparated samples taken during dilute acid pretreatment of sugarcane bagasse performed at 95 °C. Raman spectral data derived from acquisitions lasting 5 seconds with 45 exposures for an average spectra. b) Acetic acid concentrations determined by HPLC (Δ) compared to the increasing Raman estimated values at 890 cm^{-1} (●).

3.4. Fermentation of Sugarcane Bagasse Hydrolysate

In light of the Avicel hydrolysis results, *in situ* Raman measurement of the pretreated sugarcane bagasse hydrolysis was not attempted and focus was on the real-time monitoring of the fermentation after hydrolysis. The bioreactor was set up with a reduced stirring speed, low enough to precipitate particulates and hence limit the amount of light scattering particles in a ~5 cm top layer of the broth, but vigorous enough to keep yeast suspended. In this way, spectra with acceptable noise levels were obtained from the less opaque zone around the tip of the probe, allowing for real-time monitoring shown in Fig. 5a. In contrast, smoother and less attenuated spectra with less noise were obtained from the fermentation sample supernatants, where complete removal of all particulates had been achieved, as seen in Fig. 5b.

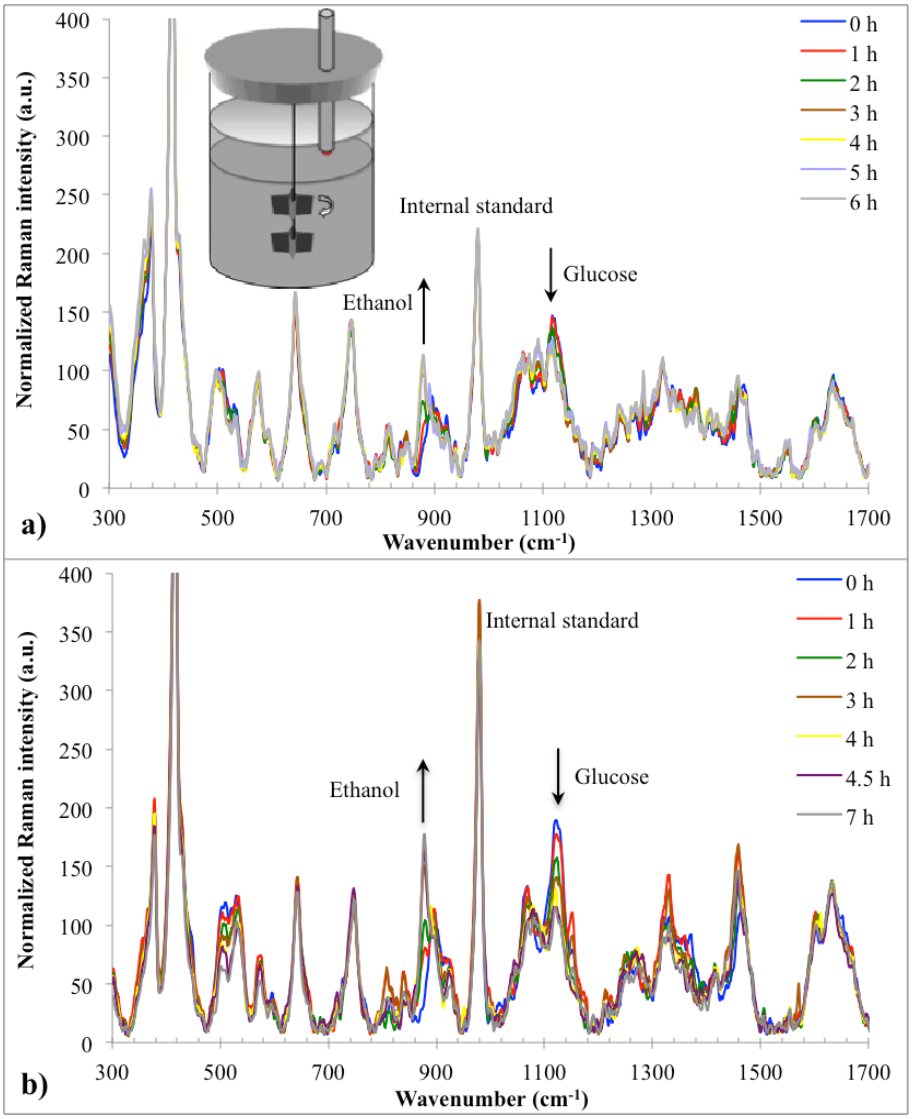

Fig. 5. Fluorescence corrected Raman spectra acquired from measurement of the fermentation of sugarcane hydrolysate with *S. cerevisiae* resulting in a decreasing of the glucose Raman signal at 1123 cm^{-1} and increasing ethanol values at 877 cm^{-1} in contrast to the unaffected peak of the internal standard at 980 cm^{-1} (sulfate). Spectra obtained by data acquisition for 20 seconds with 45 acquisitions for the average spectra. a) Real-time spectra from *in-situ* measurements made with the immersion probe in the top of the fermentation broth. Yeast cells kept suspended by stirring, while most of the biomass debris was allowed to precipitate to minimize attenuation of the scattered Raman signals. b) Off-line spectra of supernatant samples without any particulates.

Monitoring the fermentation real-time applying the spectral data without prior full separation to quantify ethanol and glucose concentrations yielded reasonable results as shown in Fig. 6. The curves obtained from directly monitoring the fermentation containing a limited amount of particles were nevertheless not as smooth as the Raman derived curves from sample supernatants (not shown). Raman estimated glucose concentration especially deviated from the glucose concentrations subsequently determined by HPLC measurements. In regard to external HPLC measurement, the root mean square error of prediction (RMSE) of glucose was calculated to be 1.62 g/L based on supernatants and 1.91 g/L when results were derived from scattering corrected *in situ* spectra, whereas the corresponding values for ethanol prediction was 1.06 g/L and 0.98 g/L respectively. However, inaccurate predictions of glucose may not have been fully expressed by the RMSE value, since fewer samples was taken at the end of fermentation where more fluctuating glucose estimations was seen. The consumption of acetic acid revealed by HPLC measurements from the short fermentation, however, was not significant enough to cause notable reduction of the related Raman band at 890 cm^{-1}. Monitoring additional components such as some of the sugars released during pretreatment and hydrolysis: cellobiose, xylose, arabinose, mannose, or lactic acid in case of bacterial contamination, as well as inhibiting furfural, would also be valuable information during fermentation. The removal of all scattering particles would likely be even more inevitable to be able to quantify these constituents, present only at low concentrations.

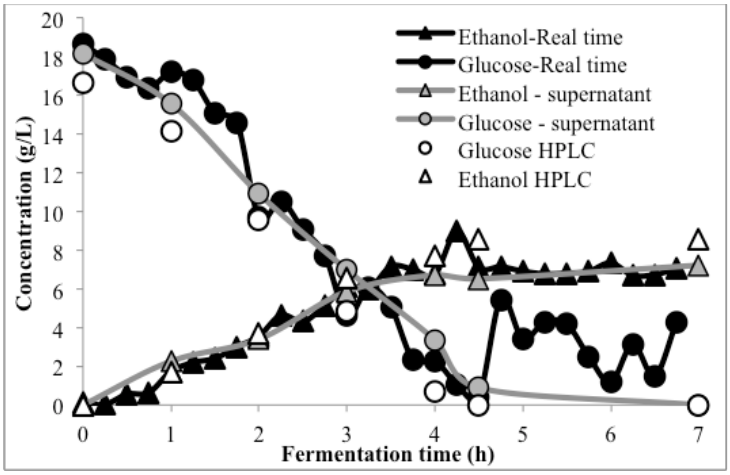

Fig. 6. Concentrations obtained during fermentation by *S. cerevisiae* of the sugarcane bagasse hydrolysate. Prior scattering correction performed using sulfate from the sulfuric acid catalyst as internal standard.

3.5. Separation

Based on these results, Raman is assessed useful for real-time *in situ* monitoring of lignocellulosic bioethanol fermentation processes as long as the inaccuracy caused by particulate matter can be limited. This can be accomplished by creating a zone at the probe with less light-scattering particulates through simple precipitation or by utilizing an integrated method inside to the reactor to reduce particulates continuously through centrifugation or filtration considering the risk of fouling. Precipitation removed most of the suspended particulates and thus scattering during measurement, but even better accuracies could most likely have been achieved using continuous separation technique in an *ex situ* monitoring mode such as in-line spin filter in a by-pass loop.

3.6. Quantitative model

The simple quantification method used in this study can easily be implemented. However, in order to improve accuracy of quantifications and thoroughly assess the potential and limitations utilizing Raman in this context, more extensive analysis of each process and development of the method is recommended. In regards to model validity, the same correction factor was used to perform scattering correction of the attenuated Raman signals at all applied wavenumbers, assuming the scale of signal reduction caused by particulates was the same for all Raman bands as for the Raman signal originating from the internal standard. However, a more accurate but complex model may be developed upon considering the magnitude of scattering hence attenuation is not necessarily constant at the various wavenumbers and each correction of wavenumber used would thus be determined separately. Future investigations could include extending this univariate model to include multiple variables and incorporate chemometric methods such as Principal Component Analysis (PCA) and with calibration and validation test reactions using numerous samples, as is commonly done with NIR spectroscopy. A multivariate model could be used where concentration accuracy is most important, whereas the univariate approach could make more sense, where model flexibility and fast process feedback is more important than obtaining exact concentrations.

The overall advantage of using Raman compared to NIR is the ease of retrieving useful information about process trends by using fewer and less complicated preprocessing steps as a result of the more distinguishable Raman peaks. Adapting Raman spectroscopy to monitor a new process may be faster, since typical chemometrics involving preliminary time-consuming setup and tests of large calibration and validation sets may not be strictly necessary. Multivariate analysis can be avoided if single Raman bands can be found in the spectra of the process mixtures as the only independent variable to estimate concentrations. As a consequence of applying a model exempt of a statistically complicated ‘black box’, fewer skills are required by professionals in order to understand a simpler

quantification method and to be able to rapidly assess the possibility of using Raman spectroscopy for real-time monitoring of an untested process and ultimately replacing off-line methods such as HPLC would be desirable. Such convenience and flexibility can prove valuable despite limitations to accuracy likely to occur from applying only simple univariate linear regression compared to NIR in combination with chemometrics. Conversely, Raman instruments tend to be more expensive than NIR spectrometers due the cost of light sources (Bakeev, 2005). Considerable research has been done using diffuse reflectance or other NIR techniques to estimate biomass composition and thereby the ethanol potential of various lignocellulosic feed stocks (Xu et al., 2013; Cervera et al., 2009). However, the various infrared absorption spectroscopy methods have only to a limited extend proven applicable as analytical tool to monitor the pretreatment, hydrolysis or fermentation processes utilizing lignocellulosic feedstock, despite numerous publications about monitoring ethanol fermentations applying these techniques (Lantz et al., 2010).

Our RMSE result were almost twice as high as compared to the root mean square error of cross-validation (RMSECV) results of 0.60 g/L for ethanol and 1.06 g/L for glucose from a similar attempt to monitor of concentrations during fermentation of separated lignocellulosic hydrolysate, where a more complicated chemometric model involving principal component analysis was applied (Ewanick et al., 2013). A more simple quantification model may however prove to be a powerful tool to monitor process progress, despite a higher limit of detection, where ease of use is more important than accuracy.

4. Conclusion

This study demonstrated the potential of using Raman spectroscopy to monitor the pretreatment, hydrolysis and fermentation process steps employed in the production of lignocellulosic bioethanol. Removal of most particulates through a simple continuous separation method enabled Raman to be used as an effective analysis tool for real-time quantification during fermentation - even without using complicated calibration methods involving chemometrics. However, analysis of the pretreatment process – monitored through the release of acetic acid – indicated possibilities, but seemed more difficult to carry out in real-time, while of the hydrolysis process was unsuccessfully monitored due to minimal spectral changes.

Acknowledgements

This work was part of the project "Bioethanol from Important Foreign Biomasses" financially supported by the Danish Council for Strategic Research, grant no.: 2104-05-0017 and the project "Development of a biorefinery concept for integrated production of biomedical, biochemical, feed and fuel from selected plant materials (BIOREF)" supported by the Danish Council for Strategic Research, grant no.: 09-065165.

References

Aarnoutse, P.J., Westerhuis, J.A., 2005. Quantitative Raman reaction monitoring using the solvent as internal standard. Anal. Chem. 77, 1228-1236.

Abu-Absi, N.R., Kenty, B.M., Cuellar, M.E., Borys, M.C., Sakhamuri, S., Strachan, D.J., Hausladen, M.C., Li, Z.J., 2010. Real time monitoring of multiple parameters in mammalian cell culture bioreactors using an in-line Raman spectroscopy probe. Biotechnol. Bioeng. 108, 1215-1221.

Adapa, P., Schonenau, L., Canam, T., Dumonceaux, T., 2011. Quantitative Analysis of Lignocellulosic Components of Non-Treated and Steam Exploded Barley, Canola, Oat and Wheat Straw using Fourier Transform Infrared Spectroscopy. J. Agr. Sci. Tech. 1, 177-188.

Agarwal, U.P., Reiner, R.R., Ralph, S.A., 2012. Estimation of Cellulose Crystallinity of Lignocelluloses Using Near-IR FT-Raman Spectroscopy and Comparison of the Raman and Segal-WAXS Methods. J. Agric. Food Chem. 61, 103–113.

Ávila, T.C., Poppi, R.J., Lunardi, I., Tizei, P.A.G., Pereira, G.A.G., 2012. Raman spectroscopy and chemometrics for on-line control of glucose fermentation by *Saccharomyces cerevisiae*. Biotechnol. Prog. 28, 1598-1604.

Bakeev, K.A., 2005. Process analytical technology: spectroscopic tools and implementation strategies for the chemical and pharmaceutical industries. Wiley-Blackwell, Oxford.

Barsberg, S., Matousek, P., Towrie, M., Jørgensen, H., Felby, C., 2006. Lignin radicals in the plant cell wall probed by Kerr-gated resonance Raman spectroscopy. Biophys. J. 90, 2978-2986.

Beutel, S., Henkel, S., 2011. *In situ* sensor techniques in modern bioprocess monitoring. Appl. Microbiol. Biotechnol. 91, 1493-1505.

Bocklitz, T., Walter, A., Hartmann, K., Rösch, P., Popp, J., 2011. How to pre-process Raman spectra for reliable and stable models?. Anal. Chim. Acta. 704, 47-56.

Butler, W.L., 1964. Absorption spectroscopy *in vivo*: Theory and application. Annu. Rev. Plant. Physiol. 15, 451-470.

Cervera, A.E., Petersen, N., Lantz, A.E., Larsen, A., Gernaey, K.V., 2009. Application of near-infrared spectroscopy for monitoring and control of cell culture and fermentation. Biotechnol. Prog. 25, 1561-1581.

Colomban, P., Havel, M., 2002. Raman imaging of stress-induced phase transformation in transparent ZnSe ceramic and sapphire single crystals. J. Raman Spectrosc. 33, 789-795.

Ewanick, S.M., Thompson, W.J., Marquardt, B.J., Bura, R., 2013. Real-time understanding of lignocellulosic bioethanol fermentation by Raman spectroscopy. Biotechnology for biofuels. 6, 1-8.

Finn, B., Harvey, L.M., McNeil, B., 2006. Near-infrared spectroscopic monitoring of biomass, glucose, ethanol and protein content in a high cell density baker's yeast fed-batch bioprocess. Yeast. 23, 507-517.

Gray, S.R., Peretti, S.W., Lamb, H.H., 2013. Real-time monitoring of high-gravity corn mash fermentation using *in situ* raman spectroscopy. Biotechnol. Bioeng. 110, 1654-1662.

Li, B., Ray, B.H., Leister, K.J., Ryder, A.G., 2013. Performance monitoring of a mammalian cell based bioprocess using Raman spectroscopy. Anal. Chim. Acta. 796, 84-91.

Liu, C., Berg, R.W., 2012. Non-linearity in Intensity versus Concentration Dependence for the Deep UV Resonance Raman Spectra of Toluene and Heptane. Appl. Spectros. Rev. 48, issue 3, 159-172.

Mazarevica, G., Diewok, J., Baena, J.R., Rosenberg, E., Lendl, B., 2004. On-line fermentation monitoring by mid-infrared spectroscopy. Appl. Spectrosc. 58, 804-810.

Menon, V., Rao, M., 2012. Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & biorefinery concept. Progr. Energ. Combust. Sci. 38, 522-562.

Lantz, A.E., Gernaey, K., Franzén, C.J., Olsson, L., 2010. Online monitoring of fermentation processes in lignocelluloses-to-bioalcohol production, in: Waldron, K., (Ed.), Bioalcohol Production: Biochemical Conversion of Lignocellulosic Biomass. Woodhead Publishing Ltd., Oxford, pp. 315-339.

Pelletier, M., 2003. Quantitative analysis using Raman spectrometry. Appl. Spectrosc. 57, 20-42.

Picard, A., Daniel, I., Montagnac, G., Oger, P., 2007. *In situ* monitoring by quantitative Raman spectroscopy of alcoholic fermentation by *Saccharomyces cerevisiae* under high pressure. Extremophiles. 11, 445-452.

Radotić, K., Kalauzi, A., Djikanović, D., Jeremić, M., Leblanc, R.M., Cerović, Z.G., 2006. Component analysis of the fluorescence spectra of a lignin model compound. J. Photochem. Photobiol. B Biol. 83, 1-10.

Roggo, Y., Chalus, P., Maurer, L., Lema-Martinez, C., Edmond, A., Jent, N., 2007. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. J. Pharm. Biomed. Anal. 44, 683-700.

Shaw, A.D., Kaderbhai, N., Jones, A., Woodward, A.M., Goodacre, R., Rowland, J.J., Kell, D.B., 1999. Noninvasive, on-line monitoring of the biotransformation by yeast of glucose to ethanol using dispersive Raman spectroscopy and chemometrics. Appl. Spectrosc. 53, 1419-1428.

Shih, C.J., Lupoi, J.S., Smith, E.A., 2011. Raman spectroscopy measurements of glucose and xylose in hydrolysate: Role of corn stover pretreatment and enzyme composition. Bioresour. Technol. 102, 5169–5176.

Shope, T.B., Vickers, T.J., Mann, C.K., 1987. The direct analysis of fermentation products by Raman spectroscopy. Appl. Spectrosc. 41, 908-912.

Sivakesava, S., Irudayaraj, J., Demirci, A., 2001. Monitoring a bioprocess for ethanol production using FT-MIR and FT-Raman spectroscopy. J. Ind. Microbiol. Biotechnol. 26, 185-190.

Stewart, D., Wilson, H., Hendra, P., Morrison, I., 1995. Fourier-transform infrared and Raman spectroscopic study of biochemical and chemical treatments of oak wood (*Quercus rubra*) and barley (*Hordeum vulgare*) straw. J. Agric. Food Chem. 43, 2219-2225.

Uysal, R.S., Soykut, E.A., Boyaci, I.H., Topcu, A., 2013. Monitoring multiple components in vinegar fermentation using Raman spectroscopy. Food Chem. 141, 4333-4343.

Vankeirsbilck, T., Vercauteren, A., Baeyens, W., Van der Weken, G., Verpoort, F., Vergote, G., Remon, J.P., 2002. Applications of Raman spectroscopy in pharmaceutical analysis. TrAC trends in analytical chemistry. 21, 869-877.

Whelan, J., Craven, S., Glennon, B., 2012. *In situ* Raman spectroscopy for simultaneous monitoring of multiple process parameters in mammalian cell culture bioreactors. Biotechnol. Prog. 28, 1355-1362.

Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics Intellig. Lab. Syst. 58, 109-130.

Xu, F., Yu, J., Tesso, T., Dowell, F., Wang, D., 2013. Qualitative and quantitative analysis of lignocellulosic biomass using infrared techniques: A mini-review. Appl. Energy. 104, 801-809.

MANUSCRIPT III

Comparison between the pentose fermenting yeasts *Scheffersomyces stipitis* and *Spathaspora passalidarum* grown in hydrolysates of tropical biomasses

Jens A. Iversen, Birgitte K. Ahring
Intended for submission in Applied Microbiology and Biotechnology

Comparison between the pentose fermenting yeasts *Scheffersomyces stipitis* and *Spathaspora passalidarum* grown in hydrolysates of tropical biomasses

Jens A. Iversen[1,2] and Birgitte K. Ahring[1,2,*]

[1]Section for Sustainable Biotechnology, Aalborg University, A. C. Meyers Vænge 15, 2450 Copenhagen SV, Denmark
[2]Center for Bioproducts and Bioenergy, Washington State University Tri-Cities, 2710 Crimson Way, Richland, WA 99354, USA

*Corresponding author. Phone: +1 509-372-7682; Email: bka@wsu.edu

Abstract Sugarcane bagasse (SCB), *Eucalyptus* wood (EW) and empty fruit bunches (EFB) from palm oil production are three abundant lignocellulosic raw materials from tropical regions with major potential as renewable feedstock for the production of lignocellulosic ethanol. To examine this potential we pretreated each biomass material at 10 % w/v TS, enzymatically hydrolyzed the pretreated material and further did fermentation of the hydrolysates using *Saccharomyces cerevisiae* and the pentose fermenting yeasts *Scheffersomyces* (aka *Pichia*) *stipitis* and the ascomyceteous yeast *Spathaspora passalidarum*. The highest xylose consumption rate was observed with *S. passalidarum*, but the final ethanol yield was similar for both the pentose fermenting yeasts. The previous reported ability of *S. passalidarum* to use cellobiose, glucose and xylose simultaneous or to ferment under anaerobic conditions was not observed during our fermentations in pretreated lignocellulosic hydrolysate. *S. passalidarum* was further found to be somewhat inhibited by the high level of acetic acid of more than 6 g/L in the EFB hydrolysate, resulting in a long lag phase. However, the final ethanol concentration of 20.6 g/l in the EFB hydrolysate with the highest amount of total sugars was the highest achieved with the organism. *S. passalidarum* was however not able to overcome the inhibitory effect of acetic acid fermenting in the same hydrolysate under fully anaerobic condition, despite the ability to ferment xylose anaerobically.

Keywords Xylose fermentation · *Scheffersomyces stipitis* CBS 6054 · *Spathaspora passalidarum* · *Eucalyptus* · sugar cane bagasse · oil palm empty fruit bunch fiber

Introduction

Waste materials from corn or grain production in Europe and USA have drawn much attention as possible raw materials for 2nd generation bioethanol production. However, important lignocellulosic biomasses found in developing countries also have tremendous potential as raw materials due to availability and low cost. Sugar cane bagasse residue from the 1st generation bioethanol industry based on sugar, empty fruit bunches from the palm oil industry and *Eucalyptus* tree fibers linked to paper production are some promising lignocellulosic biomasses in the tropical or subtropical world present in abundant amounts and assumed to be ideal for low cost cultivation on a vast scale (Cardona et al. 2010; McIntosh et al. 2012; Sumathi et al. 2008; Yang et al. 2011).

Eucalyptus

Eucalyptus trees are used globally for production paper and energy production as they are among the world fastest growing trees and are cultivated throughout the tropics and subtropics. Almost 20 Mio ha are used to cultivate *Eucalyptus* worldwide in countries such as Argentina, Australia, Brazil, China, India, South Africa, Uruguay and Vietnam. *Eucalyptus* has a very effective transpiration making it very drought resistant and the average annual growth rates of wood volume can reach 70 m^3 per ha per year. Harvesting the trees can happens little as every six years, making cultivation easy and reducing the environmental impact created by disturbances at harvest and planting (Forrester et al. 2010).

Sugar cane bagasse

Sugarcane is a perennial grass cultivated in tropical or temperate climates, mainly for the production of sugar. Sugar cane bagasse (SCB) is the fibrous residue of the sugar cane plant after the juice is extracted. Mainly cultivated in Brazil on about 10 % of the agricultural area, the total annual sugar cane bagasse production in the country is 182 Mt/y and the average sugarcane productivity has reached 6500 ton/km^2. In general, 280 kg of wet bagasse is generated from each ton of sugar cane (Cardona et al. 2010). Sugarcane production in Brazil is known for relatively low soil erosion and some areas have been producing sugar cane for more than 200 years with increasing yields (Goldemberg et al. 2008). Some of the excess bagasse is currently burned at the sugar mill to meet the energy needs, but the remaining fraction could be made available for fuel and chemical production.

Oil palm empty fruit bunch fiber

Palm oil is used for food vegetable oil, soaps and biodiesel production. Originating in the tropical parts of West Africa, oil palm (*Elaeis guineensis*) is most intensively cultivated in areas with high precipitation and more than 90 % is produces in Indonesia and Malaysia. Oil yields per hectare are five to ten times higher than what is obtained by growing rapeseed or soybean (Sumathi et al. 2008). In 2011-2012, the World supply of palm oil was 50.6 million tons increasing by almost 25 % over the last four years (USDA 2012). Empty fruit bunches (EFB) is the main oil palm lignocellulosic residues generated after the oil fruits are removed from the harvested fruit bunches. About fifteen million tons of EFB is annually generated in Malaysia alone (Rahman et al. 2007). Mesocarp fiber is the waste residue from extracting the palm oil from the palm fruit itself. The kernel of the fruit also contains oil and extracting palm kernel oil produces the palm kernel press cake (PKC) in smaller amounts and the kernel shells. Trunks and fronds of the palm tree itself are further fibrous materials besides EFB, PKC, mesocarp fibers and kernel shells, which currently are mostly burned at the oil mills (Yano et al. 2009). The PKC contains 50 % fermentable hexose sugars, which have successfully been hydrolyzed and fermented with *S. cerevisiae* without any pretreatment resulting in a yield of 200 g ethanol per kg PKC (Jørgensen et al. 2010). The lignocellulosic waste products, EFB, mesocarp fiber, trunks and fronds also represent a major resource for ethanol production.

Lignocellulosic ethanol

In contrast to starch or sugar based fermentation, lignocellulosic biomass in general needs severe pretreatment before enzymatic hydrolysis and fermentation in order to release the sugars for fermentation (Mosier et al. 2005). Lignocellulosic ethanol production is also challenged by a wide range of compounds formed during pretreatment of biomass that may inhibit fermentation. Among inhibitors found in hydrolysate are aliphatic acids such as acetic, formic and levulinic acids, aromatic phenolics and the furaldehydes 5-hydroxymethylfurfural (HMF) and furfural. Acetic acid and furfural are considered to be the major inhibitors in pretreated biomass, which a lignocellulosic fermentation process should be able to deal with (Bellido et al. 2011; Klinke et al. 2004). Currently, most bioethanol production concepts successfully tested in demonstration plants and heading for commercialization use *S. cerevisiae* due to the ability to control fermentation in large scale and to avoid microbial contamination in the fermentation process. However, integrated xylose fermentation processes needs yet to prove fully ready for production scale. *S. cerevisiae* is an excellent ethanol producer when it comes to using glucose from starch, but cannot grow on xylose, the main component in hemicellulose. A limited number of bacteria, yeasts and fungi are able through fermentation to convert the pentose sugars from the hemicellulose fraction to ethanol (Kumar et al. 2009).

Xylose fermenting yeasts

To overcome this hurdle many research projects have focused on creating strains that can convert all carbohydrates in lignocellulosic biomass into ethanol. A number of yeast species are capable of fermenting both pentose and hexose carbohydrates into ethanol. However, ethanol tolerance or yields are low or xylose uptake will stop under strict anaerobic conditions. Fermentation under microaerophilic condition is hence required to maintain redox balance (Skoog and Hahn-Hägerdal 1990). Amongst xylose fermenting yeast are *Candida shehateae*, *Scheffersomyces stipitis* (previously known as *Pichia stipitis*), Kluyveromyces marxianus (Hahn-Hägerdal et al. 2007). So far, S. stipitis strains have been considered to be among the best xylose fermenting yeasts. Generally, this robust yeast is capable of using several different monomeric sugars, cellobiose and even to hydrolyze xylan (du Preez et al. 1986; Lee et al. 1986). Recently, the yeast *Spathaspora passalidarum* isolated from the midgut of passalid bettle was shown able to coferment cellobiose, xylose and glucose efficiently with a remarkable fast utilization of xylose (Long et al. 2012; Nguyen et al. 2006). Furthermore, *S. passalidarum* can ferment xylose to ethanol under anaerobic conditions - an unusual property among pentose fermenting yeasts (Hou 2012).

The objective of this work was to investigate the use of the selected important biomasses: sugar cane bagasse, *Eucalyptus* tree fiber and empty fruit bunches from palm oil production found in the tropics as substrate for lignocellulosic bioethanol production using *S. stipitis* and *S. passalidarum*. For better assessment, fermentations by *S. stipitis* and *S. passalidarum* were to some extend compared with fermentation by an industrial strain of *S. cerevisiae*. All fermentations were carried out without costly process steps such as; detoxification, use of any pH buffers or separation of hydrolysates in fractions. This overall simple process approach - suitable for industrial scale - was used to simulate the performance of the yeast strains under microaerobic conditions and in viscous lignocellulosic hydrolysates containing inhibitors.

Materials and Methods

Raw material

The empty fruit bunch fiber from palm oil plant (*Elaeis guineensis*) was acquired from United Plantations BHD in Malaysia, and the *Eucalyptus* wood (*Eucalyptus globulus*) was collected from Burleigh Murray Ranch State Park**,** San Francisco, CA. WSU Tri-Cities obtained the sugar cane bagasse (*Saccharum officinarum* L.) from Lafourche Sugars, LLC, Thibodaux, LA, via Dr. Edward Richard from the

USDA. All biomasses were hammer milled to a particle size of 1 mm with a MF 10 basic Microfine grinder from Ika®-Werke GmbH, Staufen, Germany.

Composition analysis

The amount of structural carbohydrates and lignin in the three biomasses were determined according to the NREL laboratory analytical procedure for standard biomass analysis (Sluiter et al. 2008). The method is based on a two-step hydrolysis at 30 °C in strong sulfuric acid followed by dilution and hydrolysis at 121 °C.

Pretreatment and hydrolysis

A low temperature dilute acid pretreatment method was used for all three biomasses. Suspensions of 625 mL with 10 % w/v total solids and 0.5 % w/v sulfuric acid in 2 L tightly closed blue cap flasks were prepared for each biomass and pretreated in an incubator at 95 °C for 45-82 hours with samples taken twice a day. Prior to hydrolysis, the three pretreated biomasses were neutralized to pH 4.8 adding 1M sodium hydroxide, but mostly sodium hydroxide pellets were used to avoid dilution. Hydrolysis were performed with Cellic CTec2 (210 mg protein/mL) from Novozymes A/S, Bagsvaerd, Denmark at 50 °C at 120 rpm for 72 hours.

Microorganisms and fermentations

Inocula of an industrial strain of *Saccharomyces cerevisiae* intended for fuel ethanol (Thermosacc Dry, Lallemand Ethanol Technology A/S, Denmark), *Scheffersomyces (Pichia) stipitis* CBS 6054 (ATCC 58785) and *Spathaspora passalidarum*; 11-Y1 (ATCC MYA-4345) were propagated in autoclaved 100 mL YPX (pentose-fermenting yeasts) or YPD (*S. cerevisiae*) medium with 20 g/L yeast extract, 20 g/L protease peptone, 10 g/L xylose or glucose. The inoculum cultures were grown in a 250 ml baffled Erlenmeyer fitted with cotton plugs flasks at 32 °C and 160 rpm for 60 hours or 27 hours for *S. cerevisiae* inoculum. The harvested centrifuged cell pellets were washed and resuspended to viscous and concentrated inocula. Cotton plugs were fitted flasks for the aerated cultures, whereas flasks with anaerobic cultures were equipped with a swan neck containing paraffin oil. Baffled 50 mL Erlenmeyer flask were added 30 mL unseparated hydrolysates and sanitized by heating to 100 °C for 30 min (to avoid autoclaving and furfural formation) before adding filter sterilized protease-peptone media stock to produce media containing 0.7 % w/v protease peptone and 0.3 % w/v urea. Media was inoculated to initiate fermentations with targeting 1.0 g/L yeast. An optical density of 1.0 measured at 600 nm of was assumed to correspond to a cell inoculum of 0.15 g/L. The cultures were grown at 35 °C and 140 rpm, whereas the aerated cultures with reduced oxygen transfer were cultivated at 100 rpm. All fermentations were performed in duplicates and samples were taken in a clean bench.

HPLC Analysis

High Performance Liquid Chromatography (HPLC) was used to determine sample concentrations (Ultimate 3000, Dionex, CA, USA) using a refractive index (RI) detection module. The instrument was equipped with an Aminex HPX-87H column (300 X 7.8 mm) and a Cation-H guard column (30 x 4.6 mm), both Bio-Rad Laboratories, CA, USA. Applying 10 µL of sample supernatant, the components were separated isocratically at a flow rate of 0.6 mL/min at 60 °C for 26 min with 4mM H_2SO_4 as eluent. Mixtures containing glucose, xylose, arabinose, glycerol, ethanol, acetic acid and lactate were used as standards (concentrations of 0.5 to 20 g/L), as well as xylitol, furfural and HMF at a lower concentration range. 10 µL 1M H_2SO_4 was added to the 1 mL samples before measuring the supernatants with the HPLC. In order to measure other sugars such as galactose and to confirm results from the composition analysis, some samples were also separated through an Aminex HPX-87P column (300 X 7.8 mm) with a Micro-Guard Carbo-P column (30 x 4.6 mm) at pH 5-9, 83 °C and 0.6 mL/min, using MiliQ water as eluent on the Ultima.

Results

Pretreatment and hydrolysis

Table 1 shows the results of structural composition analysis of the three biomasses. SCB has the highest cellulose content, but the lowest lignin and ash content, whereas the highest ash and lignin along with the lowest cellulose and hemicellulose content was found in *Eucalyptus* tree fiber. Almost all lignin was present as the acid insoluble filtrate and the UV-determined acid soluble lignin was only detectable in insignificant low amounts.

Table 1 Biomass composition (% w/w of DM). Values are given as duplicates with standard deviations (±).

	Cellulose	Hemi-	Lignin	Ash
EFB	33.87±2.21	33.51±0.68	17.65±1.10	3.04±0.05
SCB	35.40±0.84	31.23±0.61	19.90±0.70	5.99±0.44
Eucalyptus	27.57±0.06	19.40±0.28	20.41±2.13	9.58±0.94

The progress of the low temperature pretreatments of the three used biomasses is shown in Fig. 1. The simple slow low temperature pretreatment is basically able to

solubilize all the available xylose. After 45 hours of pretreatment of sugar cane bagasse and empty fruit bunches, most of the xylose had been released from the hemicellulose and the reactions were stopped. Pretreatment of *Eucalyptus* seemed more difficult and to achieve near complete release of xylose, the reaction was continued for a total reaction time of 84 hours. The concentration of xylose in the *Eucalyptus* hydrolysate was almost half of the amounts found in the EFB and SCB hydrolysates, however, at levels corresponding to the results given by the composition analysis (Table 2).

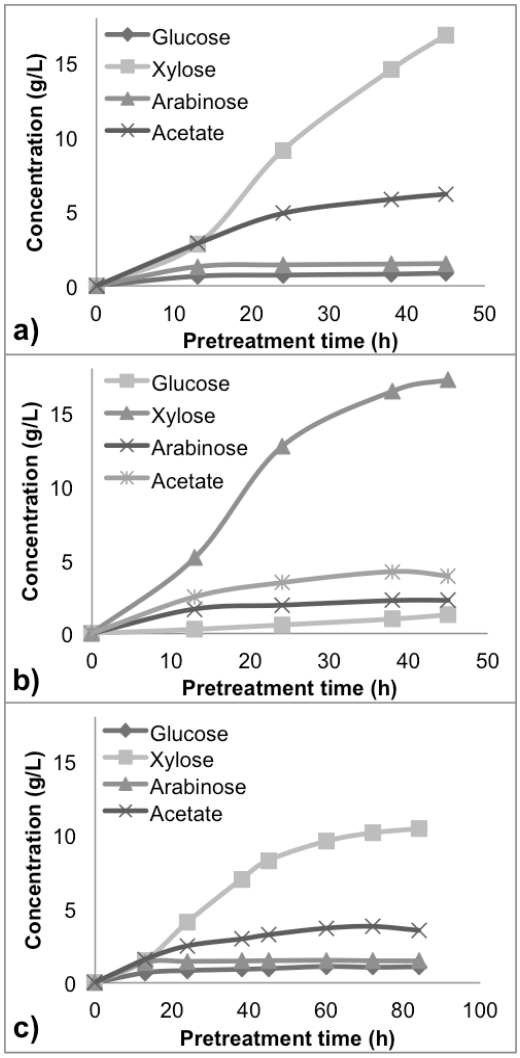

Fig. 1 Pretreatments of the respective biomasses in 0.5% sulfuric acid at 95°C, 10% w/w TS. a) Empty fruit bunches, b) Sugar cane bagasse c) *Eucalyptus* tree fiber

Release of acetic acid was considerably higher from EFB than from the other two materials, as shown in Table 2. The table lists the amounts of sugar obtained after pretreatment and enzymatic hydrolysis against the results of the analysis of the composition of the individual carbohydrates and acetic acid according to the NREL method. The enzymatic hydrolysis of hemicellulose ensured the complete release of the small amounts of xylose left in the biomasses after pretreatment. However, the hydrolysis of the more recalcitrant cellulose fibers was not fully achieved, with 24 %, 30 % and 46 % glucose left in the EFB, *Eucalyptus* and SCB hydrolysates respectively. SCB also had the lowest share of its acetate released from the lignocellulosic structure. A better cellulose hydrolysis may likely have been accomplished applying a more severe pretreatment method at a higher temperature and pressure, in order to break apart more of the lignin. Estimations of the potential ethanol yield of each of the three biomasses could also have been made if a more optimal pretreatment with higher sugar release at conditions more likely to be used in an industrial process had been applied.

Table 2 Monomer composition of biomasses and released hereof (% w/w of DM)

	EFB		SCB		*Eucalyptus*	
	Composition	Hydrolysis	Composition	Hydrolysis	Composition	Hydrolysis
Glucose	37.6	28.7	39.3	21.2	30.6	21.4
Galactose	1.7	1.0	2.0	1.2	2.2	1.8
Xylose	23.5	22.9	21.8	20.9	12.5	13.4
Arabinose	1.3	1.4	2.0	2.4	1.2	1.6
Acetic acid	7.2	6.3	5.8	3.9	5.1	3.8

Microaerobic fermentations

The results of fermentation of SCB, EFB and *Eucalyptus* tree fiber hydrolysates by *S. stipitis* and *S. passalidarum* illustrated in Fig. 2 are all avarage of two parallel samples. Fermentation by *S. cerevisiae* showed full use of fermentable sugars to ethanol within 9 hours (data not shown). Glucose was generally consumed within 40 hours of fermentation by the pentose fermenting yeasts, except the fermentation of EFB hydrolysate by *S. passalidarum*, which lasted for 93 hours (Fig. 2f). Both pentose fermenting yeasts exhibited a strong glucose preference and xylose concentrations changed very little until no glucose was left. Hereafter, the consumption of xylose increased, but with some exceptions at significantly lower

rates than the corresponding metabolisms rates for glucose. Overall, *S. passalidarum* seemed to utilize the remaining pentose sugars after depletion of glucose more rapidly than *S. stipitis*. Ethanol production predominantly occurred during glucose metabolism, with only little or no additional increase in ethanol concentrations during pentose uptake and in many cultures reassimilation of ethanol started before complete consumption of xylose. Almost no consumption of arabinose was seen (not shown) and about half of the initially low amounts of arabinose of 1.5 to 2.5 g/L was left at the end of most fermentations. *S. stipitis* can produce cell biomass, but ethanol from arabinose (Agbogbo and Coward-Kelly 2008). The initial cellobiose levels of 1.5 to 2.0 g/L in the *Eucalyptus* and SCB hydrolysates were generally consumed later than xylose. Assimilation of xylose resulted in formation of xylitol to final concentrations in the range of 1.1 to 3.6 g/L proportional to initial xylose level. Generally, xylitol formation was 30 % higher in the cultures with *S. stipitis* than *S. passalidarum* growing in the same substrates with the final yields of 0.10 - 0.11 and 0.07 - 0.08 gram xylitol per gram xylose consumed respectively. However, significantly more xylitol was produced in EFB hydrolysate cultures, which however was not remarkably higher in proportion to ethanol formation. Xylitol did not seem to be reassimilated in any of the cultures despite the continued supply of oxygen. The relative high temperature of 35°C may just be within the range where *S. stipitis* is known to have optimal biomass and ethanol productivities, but less accumulation of residual xylitol and xylose occurs at lower temperatures (Slininger et al. 1990). Therefore, better ethanol yields might have been obtained if fermentation would have been carried out at temperatures below 30°C.

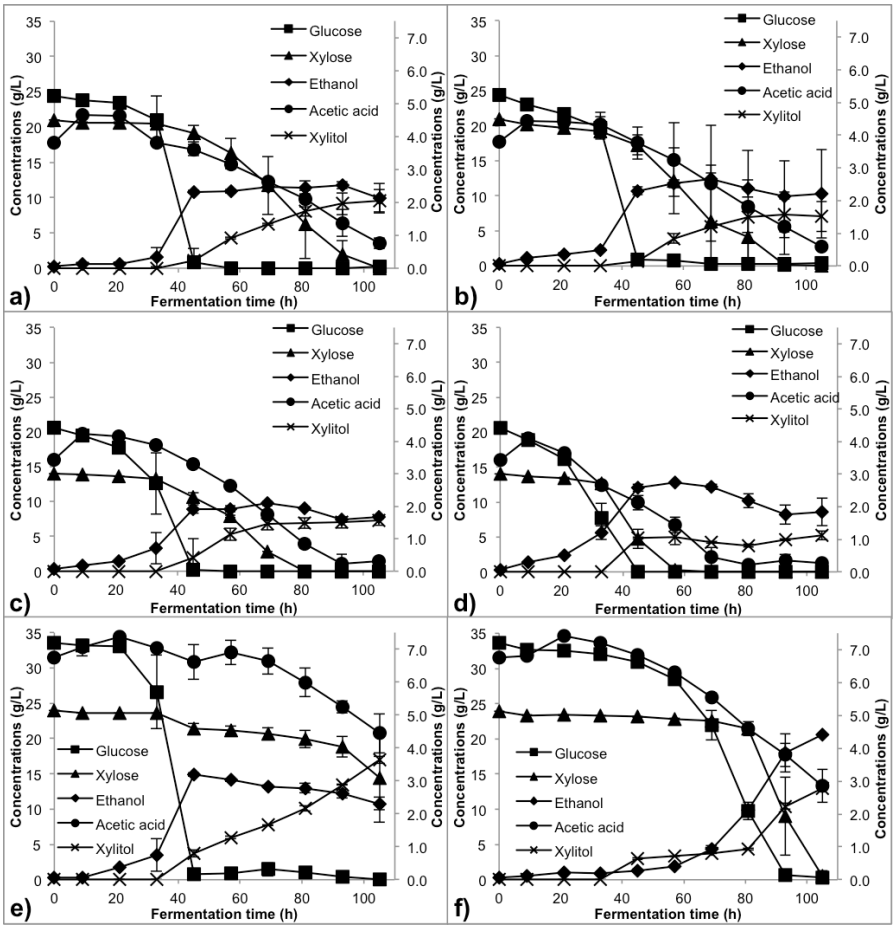

Fig. 2 Fermentation of hydrolysates of a) sugarcane bagasse by *S. stipitis*, b) sugarcane bagasse by *S. passalidarum*, c) *Eucalyptus* tree fiber by *S. stipitis*, d) *Eucalyptus* tree fiber by *S. passalidarum*, e) palm oil empty fruit bunches by *S. stipitis*, f) palm oil empty fruit bunches by *S. passalidarum*. Duplicate fermentations were conducted in shake flasks at 35 °C and 140 rpm. Left vertical axis: glucose, xylose and ethanol; Right vertical axis: acetic acid and xylitol

In all hydrolysates, the yeasts assimilated acetic acid and the concentrations were reduced by 3-4 g/L throughout the fermentations. The high acetic acid concentrations above 6 g/L in EFB hydrolysates were not enough to inhibit fermentations completely. However, initial delay in acetic acid and pentose consumption occurred in all fermentations with EFB except when cultivated with *S. cerevisiae*. The higher acetic acid content in EFB hydrolysate seemed only to delay

the sugar uptake in *S. passalidarum* until much of the acetic acid had been assimilated, but not the rate of xylose consumption (Fig. 2f), unless other unknown inhibitors had an inhibition impact. The rate of acetic acid and xylose uptake by *S. passalidarum* in SCB (Fig. 2b) exhibited some varience, but deviations were not enough to change the overall picture. As a consequence of acetic acid assimilation, pH increased from about 5.0 to 7.5 throughout the course of the fermentations (Fig. 3). In general, minor change was observed for the first 40 hours and the most rapid acetic acid consumption and pH increase was seen in the *S. passalidarum* fermentation of *Eucalyptus* hydrolysate. Initial concentrations of furfural was 0.17 g/L, 0.29 g/L and 0.29 g/L for *Eucalyptus*, SCB and EFB hydrolysate respectivly, while HMF was barely detectable. Amounts were not measured continuesly, but a final measurement revealed complete degradation of furfural and HMF as well as full consumption of galactose during the fermentations with the pentose fermenting yeasts.

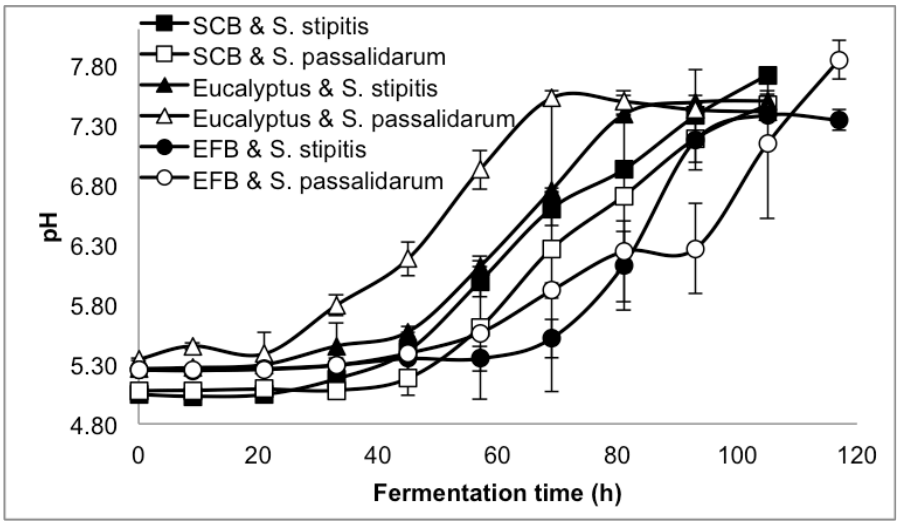

Fig. 3 Effect of acetic acid consumption on pH during fermentation of hydrolysates

Ethanol concentrations and yields are displayed in Table 3. *S. cerevisiae* showed a fast consumption of glucose (less than 9 hours), compared to 40 hours or more for glucose alone and at least 50 hours for fermentation of glucose and xylose by *S. passalidarum* or *S. stipitis*. Fermentations by *S. passalidarum* generally gave higher maximum ethanol concentrations than *S. stipitis*, whereas the highest ethanol yield at the point of maximum ethanol concentration was obtained with *S. stipitis* in EFB (0.42 g/g). Further micro aerobic fermentations of the hydrolysate by both *S. stipitis*

and *S. passalidarum* were conducted at a lower stirring rate to test if the reassimilation of ethanol, following glucose depletion, could be reduced by decreasing the oxygen transfer rate. Reducing the stirring rate during fermentations from initially 140 rpm to 100 rpm had no effect on the consumption of ethanol during pentose consumption (data not shown). The only changes of the fermentation results were that the rates of sugar and acetic acid metabolism was decreased, without affecting the order of substrate uptake.

Table 3 Summary of fermentations carried out in shake incubator at 140 rpm at 35°C. Values are given as duplicates with standard deviations (±).

	Maximum ethanol (g/L)	Ethanol yield[a] (g ethanol/g sugar consumed)	Fermentation time[b] (h)
Eucalyptus			
S. stipitis	9.7±0.31	0.29±0.01	69
S. passalidarum	12.9±0.12	0.35±0.01	57
S. cerevisiae	8.4±0.24	0.41±0.01	< 9
EFB			
S. stipitis	14.9±0.26	0.41±0.02	45
S. passalidarum	20.6±0.13	0.36±0.00	105
S. cerevisiae	12.2±0.01	0.36±0.00	< 9
SCB			
S. stipitis	11.7±0.50	0.25±0.02	93
S. passalidarum	12.5±0.88	0.30±0.03	69
S. cerevisiae	8.7±0.02	0.35±0.00	< 9

[a] Glucose was the only sugar accounted for to determine ethanol yield of fermentations by

[b] Fermentation time for consumption of glucose and xylose, but glucose only with *S.*

Anaerobic fermentations and glucose metabolism

In addition to the micro aerobic fermentations, EFB hydrolysates were further fermented under anaerobic conditions fermented by *S. passalidarum* and for comparison, *S. cerevisiae* (data not shown). The results of *S. cerevisiae* yeast showed complete consumption of glucose within 12 hours before assimilation of acetic acid during the fermentation. In contrast, no metabolism of pentoses, acetic acid or even glucose was observed in the hydrolysates cultivated with the *S. passalidarum*, indicating the need for oxygen to be present for this yeast to overcome inhibition before being able to metabolize the sugars and produce ethanol.

Discussion

Generally we found that all the microaerobic fermentations of hydrolysates with both of the xylose fermenting yeasts showed sequential utilization of glucose before consumption of pentoses, and with slower pentoses utilization rates than the corresponding glucose utilization rates. The efficiency of xylose metabolism is determined by the combined effect of sugar transport mechanisms, enzyme activities during sugar metabolism, and redox balance. First step in xylose metabolism is the conversion of xylose to xylitol by xylose reductase utilizing NADH or NADPH as cofactors. Xylitol is oxidized to xylulose by NAD^+ cofactor specific xylitol dehydrogenase (XDH) before entering the non-oxidative pentose phosphate pathway (Skoog and Hahn-Hägerdal 1990). The redox balance of this XR/XDR- pathway in yeasts would be stabile as long as all NAD^+ needed for XDH would continuously be regenerate by XR. However, as a consequence of XR preference for NADPH over NADH in *S. stipitis*, NADH thus accumulates during anaerobic conditions - resulting in formation of xylitol being favored over xylulose. In contrast, *S. passalidarum* also utilized the same cofactors, but the preference of XR for NADH instead of NADPH enables sufficient regeneration of NAD^+ without the need for oxygen. During xylose metabolism, *S. passalidarum* exhibit a more desirable redox balance compared to *S. stipitis*. *S. passalidarum* is able to ferment xylose to ethanol anaerobically in YPD media, which is explained by the better cofactor balance (Hou 2012). Generally, xylitol formation was 30 % higher in hydrolysates with *S. stipitis* compared to *S. passalidarum* due to cofactor differences in the XR/XDH pathway.

Xylose transport is based on the ATP driven proton xylose symports in *S. stipitis*. Glucose inhibits xylose consumption, since this transport mechanism also mediates the transport of glucose. Other yeasts can also transport xylose by facilited diffusion with concentration gradients as only driving force without using ATP in the process (Kilian and Uden 1988; Webb and Lee 1990). Hou demonstrated how the repression of xylose transport by glucose in *S. passalidarum* is much stronger under anaerobic conditions, suggesting the two

different transport systems responsible for xylose uptake, resulting in simultaneous consumption of glucose and xylose during aerobic growth and sequential sugar assimilation under anaerobic conditions (Hou 2012). Long *et al.* also reported how *S. passalidarum* is capable of utilizing cellobiose, glucose and xylose similtanouosly at similar rates under fully aerobic or oxygen limited conditions and the ability to ferment xylose faster than glucose during single sugar cultivations. However, this unprecedented finding was achived with defined minimal media. In the presence of acetic acid in corn stover hydrolysate, fermentation of xylose was delayed until the depletion of both glucose and acetic acid (Long et al. 2012). Similarly in our study, only minor difference in sugar uptake during microaerobic fermentation of biomass hydrolysate was observed between *S. stipitis* and *S. passalidarum* and anaerobic metabolism did not occur during fermentation of EFB hydrolysate by *S. passalidarum*. These results might be due to the presence of acetic acid.

Acetic acid inhibits both cellular growth and ethanol production during fermentation with *S. stipitis* and xylose consumption has been shown to be more affected than glucose consumption by the presence of acetic acid (Bellido et al. 2011). Nigam reported how a *S. stipitis* fermentation in synthetic media with a constant pH of 5.0 was inhibited by an increase in acetic acid concentration from 1 to 9.5 g/L resulting in reduced productivity and yield by 97 % and 82 %, respectively (Nigam 2001). The inhibitory effect of acetic acid is shown to be more severe at lower pH, where the undissociated uncharged form of acetic acid is membrane permeable. In order to avoid a decrease in the intracellular pH arising from the acetic acid protons must be pumped out of the cell by plasma membrane ATPase at the cost of ATP and cell growth (Casey et al. 2010). Acetic acid enters the Krebs cycle and undergoes respiration in the presence of oxygen in order to be metabolized (Duggan 1964). In the absence of oxygen, acetic acid is not assimilated by *S. stipitis* (Van Zyl et al. 1988). The rate of ATP generation from sugar catabolism must counter the ATP requirement for eliminating the toxic effects of acetic acid. Acetic acid may have a more severe impact on xylose metabolism, since ATP generation from xylose utilization is lower than during glucose fermentation (Bellissimi et al. 2009). Low content of sugar relative to acetic acid in the hydrolysate can thus explain unsuccessful anaerobic fermentation with *S. passalidarum* in EFB hydrolysate. The same hydrolysate did, however, show no inhibition of the *S. cerevisiae* cultures, since *S. cerevisiae* possesses higher tolerance towards acetic acid than a number of xylose fermenting yeasts (Limtong et al. 2000; Palmqvist et al. 2000). Considering the increasing toxicity of acetic acid with decreasing culture pH, better results may be achieved, if yeast fermentations in hydrolysates from lignocellulosic materials would be carried out at higher pH than what has been found optimal for artificial media without acetic acid.

Glucose inhibits uptake of xylose in yeasts not only by occupying the transporters, but glucose also acts as a repressor of xylose induced synthesis of

enzymes involved in metabolism of xylose. In *S. stipitis,* the presence of glucose is known to repress activities of xylose reductase (XR) and xylitol dehydrogenase (XDH) and similar mechanism may exist in *S. passalidarum* (Bicho et al. 1988). The ability to utilize pentose sugars is not the only metabolic feature that sets *S. stipitis* apart from *S. cerevisiae*. The regulation of glucose metabolism also differs fundamentally due to the Crabtree effect, exhibited only in *S. cerevisiae*. Elevated glucose levels in media, induces the repression of the tricarboxylic acid (TCA) cycle even under fully aerobic conditions, which results in simultaneously fermentation and respiration. In contrast, the Crabtree-negative *S. stipitis* more easily favors respiration at the same aeration conditions and the same high glucose concentrations (Fiaux et al. 2003; Van Urk et al. 1989). This can explain why the micro aerobic conditions did not predominantly result in respiration in the *S. cerevisiae* cultures and lower ethanol yields than observed with *S. stipitis*. In excess glucose, aeration is irrelevant to *S. cerevisiae*. *S. passalidarum* is, with similar metabolic features as exhibited by *S. stipitis,* assumed to be Crabtree negative as well. However, more thorough experiments are needed to verify, whether or not respirative metabolism also occurs at increased glucose levels.

The NADH-dependent alcohol dehydrogenase (ADH) metabolizes furfural or HMF to less harmful furfuryl alcohols in yeasts. This enzyme also reduces acetaldehyde to ethanol and a decrease in the production of ethanol in the presence of furfural is thus expected due to substrate competition. Furaldehyde reduction can also be achieved with XR in *S. stipitis,* thereby affecting the xylose pathway (Almeida et al. 2008). A less favored oxidation to furoic acid also occurs in yeast (Taherzadeh et al. 1999). The inhibitory effect of furfural on the ethanol production rate has been observed at furfural concentrations above 2 g/L in fermentations with *S. stipitis* (Díaz et al. 2009). Below this level, the inhibitory effect tends only to result in delayed sugar consumption and a low productivity until the furfural has been degraded. Therefore, inhibition by only HMF and furfural at concentrations not exceeding 0.3 g/L in any of our fermentations is not believed to be of major importance, but a synergetic effect between acetic acid and other compounds such as levolininc acid, which is not measured, may have increased the inhibitory effects of the single components by themselves.

Even though ethanol concentration would not exceed 5 % with the known species of pentose fermenting yeasts, utilizing an easy to control anaerobic fermentation process with efficient metabolism of xylose could help developing a simple method for the industrial production of lignocellulosic bioethanol. The direct comparison of *S. passalidarum* with *S. stipitis* during fermentation of pretreated and hydrolyzed lignocellulosic materials displayed only minor performance differences. This is surprising considering the benefit of *S. passalidarum* possessing a uniqe low degree of glucose represion of xylose consumption and ability to metabolise glucose, xylose and cellobiose simultaneously in defined media as demonstrated in previous studies (Long et al. 2012; Nguyen et al. 2006). Long et al. furthermore

conducted micro aerobic fermentation by adapted *S. passalidarum* on enzymatic hydrolysate of AFEX-pretreated corn stover with an initial acetic acid content of 1.5 g/L and experienced a significant delay in the utilization xylose, while the acetic acid and most rapidly glucose were consumed. The observed sequential consumption of sugars in the presence of acetic acid limits the potential of using this promising yeast in its native form as organism for the production of 2nd generation bioethanol. Adding to this, the initial concentration of biomass hence acetic acid would likely have to be close to twice the amount as we used to approach ethanol concentrations of 5% assumed necessary for economical distillation process. Future studies with *S. passalidarum* could focus on investigating the impact of acetic acid on the order of sugar consumption, yields and productivities under various conditions and ways to reduce acetic acid sensitivity applying thorough adaption procedures or perhaps through genetic modifications. Nevertheless, *S. passalidarum* seems to be a promising pentose fermenting yeast performing similarly as *S. stipitis* and further genetic improvement could make this an industrial relevant micro organism.

Acknowledgement This research was financially supported by the Danish Council for Strategic Research, grant numbers: 2104-05-0017 and 09-065165, and was part of the project "Bioethanol from Important Foreign Biomasses" and "Bioref: A novel biorefinery concept".

References

Agbogbo FK, Coward-Kelly G (2008) Cellulosic ethanol production using the naturally occurring xylose-fermenting yeast, *Pichia stipitis*. Biotechnol Lett 30:1515-1524

Almeida J, Modig T, Röder A., Lidén G., Gorwa-Grauslund M. (2008) *Pichia stipitis* xylose reductase helps detoxifying lignocellulosic hydrolysate by reducing 5-hydroxymethyl-furfural (HMF). Biotechnol for Biofuels 1:12. doi:10.1186/1754-6834-1-12

Bellido C, Bolado S, González-Benito G, García-Cubero MT (2011) Effect of inhibitors formed during wheat straw pretreatment on ethanol fermentation by *Pichia stipitis*. Bioresour Technol 102:10868-10874

Bellissimi E, Van Dijken JP, Pronk JT, Van Maris AJA (2009) Effects of acetic acid on the kinetics of xylose fermentation by an engineered, xylose-isomerase-based *Saccharomyces cerevisiae* strain. FEMS yeast research 9:358-364

Bicho PA, Runnals PL, Cunningham JD, Lee H (1988) Induction of xylose reductase and xylitol dehydrogenase activities in *Pachysolen tannophilus* and *Pichia stipitis* on mixed sugars. Appl Environ Microbiol 54: 50-54

Cardona C, Quintero J, Paz I (2010) Production of bioethanol from sugarcane bagasse: status and perspectives. Bioresour Technol 101:4754-4766

Casey E, Sedlak M, Ho N, Mosier N (2010) Effect of acetic acid and pH on the cofermentation of glucose and xylose to ethanol by a genetically engineered strain of *Saccharomyces cerevisiae*. FEMS yeast research 10:385–393

Díaz MJ, Ruiz E, Romero I, Cara C, Moya M, Castro E (2009) Inhibition of *Pichia stipitis* fermentation of hydrolysates from olive tree cuttings. World Journal of Microbiology and Biotechnology 25:891-899

du Preez JC, Bosch M, Prior BA (1986) Xylose fermentation by *Candida shehatae* and *Pichia stipitis*: effects of pH, temperature and substrate concentration. Enzyme Microb Technol 8:360-364

Duggan P (1964) Acetate and ethanol oxidation by yeast. Irish Journal of Medical Science (1926-1967) 39:19-30

Fiaux J, Çakar Z, Sonderegger M, Wüthrich K, Szyperski T, Sauer U (2003) Metabolic flux profiling of the yeasts *Saccharomyces cerevisiae* and *Pichia stipitis*. Eukar Cell 2, No. 1:170–180

Forrester DI, Medhurst JL, Wood M, Beadle CL, Valencia JC (2010) Growth and physiological responses to silviculture for producing solid-wood products from *Eucalyptus* plantations: An Australian perspective. For Ecol Manage 259:1819-1835

Goldemberg J, Coelho ST, Guardabassi P (2008) The sustainability of ethanol production from sugarcane. Energy Policy 36:2086-2097

Hahn-Hägerdal B, Karhumaa K, Fonseca C, Spencer-Martins I, Gorwa-Grauslund MF (2007) Towards industrial pentose-fermenting yeast strains. Appl Microbiol Biotechnol 74:937-953

Hou X (2012) Anaerobic xylose fermentation by *Spathaspora passalidarum*. Appl Microbiol Biotechnol 94:205-214. doi: 10.1007/s00253-011-3694-4

Jørgensen H, Sanadi AR, Felby C, Lange NEK, Fischer M, Ernst S (2010) Production of ethanol and feed by high dry matter hydrolysis and fermentation of palm kernel press cake. Appl Biochem Biotechnol 161:318-332

Kilian S, Uden N (1988) Transport of xylose and glucose in the xylose-fermenting yeast *Pichia stipitis*. Appl Microbiol Biotechnol 27:545-548

Klinke HB, Thomsen A, Ahring BK (2004) Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. Appl Microbiol Biotechnol 66:10-26

Kumar S, Singh SP, Mishra IM, Adhikari DK (2009) Recent advances in production of bioethanol from lignocellulosic biomass. Chem Eng Technol 32:517-526

Lee H, Biely P, Latta R, Barbosa M, Schneider H (1986) Utilization of xylan by yeasts and its conversion to ethanol by *Pichia stipitis* strains. Appl Environ Microbiol 52:320–324

Limtong S, Sumpradit T, Kitpreechavanich V, Tuntirungkij M, Seki T, Yoshida T (2000) Effect of acetic acid on growth and ethanol fermentation of xylose fermenting yeast and *Saccharomyces cerevisiae*. Kasetsart Journal (Nat. Sci.) 34:64-73

Long TM, Su YK, Headman J, Higbee A, Willis LB, Jeffries TW (2012) Cofermentation of Glucose, Xylose, and Cellobiose by the Beetle-Associated Yeast *Spathaspora passalidarum*. Appl Environ Microbiol 78:5492-5500

McIntosh S, Vancov T, Palmer J, Spain M (2012) Ethanol production from *Eucalyptus* plantation thinnings. Bioresour Technol 110:264-272

Mosier N, Wyman C, Dale B, Elander R, Lee Y, Holtzapple M, Ladisch M (2005) Features of promising technologies for pretreatment of lignocellulosic biomass. Bioresour Technol 96:673-686

Nguyen NH, Suh SO, Marshall CJ, Blackwell M (2006) Morphological and ecological similarities: wood-boring beetles associated with novel xylose-fermenting yeasts, *Spathaspora passalidarum* gen. sp. nov. and *Candida jeffriesii* sp. nov. Mycol Res 110:1232-1241

Nigam J (2001) Ethanol production from hardwood spent sulfite liquor using an adapted stain of *Pichia stipitis*. J Indust Microbiol Biotechnol 26(3):145-150

Palmqvist E, Grage H, Meinander NQ, Hahn-Hägerdal B (2000) Main and interaction effects of acetic acid, furfural, and p-hydroxybenzoic acid on growth and ethanol productivity of yeasts. Biotechnol Bioeng 63:46-55

Rahman S, Choudhury J, Ahmad A, Kamaruddin A (2007) Optimization studies on acid hydrolysis of oil palm empty fruit bunch fiber for production of xylose. Bioresour Technol 98:554-559

Skoog K, Hahn-Hägerdal B (1990) Effect of oxygenation on xylose fermentation by *Pichia stipitis*. Appl Environ Microbiol 56:3389-3394

Slininger P J, Bothast R J, Ladisch M R, Okos M R (1990) Optimum pH and temperature conditions for xylose fermentation by *Pichia stipitis*. Biotechnol Bioeng 35(7):727–731

Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, Crocker D (2008) Determination of structural carbohydrates and lignin in biomass. NREL Laboratory Analytical Procedure 2008. Available at http://www.nrel.gov/biomass/pdfs/42618.pdf

Sumathi S, Chai S, Mohamed A (2008) Utilization of oil palm as a source of renewable energy in Malaysia. Renewable and Sustainable Energy Reviews 12:2404-2421

Taherzadeh MJ, Gustafsson L, Niklasson C, Lidén G (1999) Conversion of furfural in aerobic and anaerobic batch fermentation of glucose by Saccharomyces cerevisiae. Journal of bioscience and bioengineering 87:169-174

USDA (2012) Economic Research Service: Oil Crops Yearbook. http://usda.mannlib.cornell.edu/MannUsda/viewDocumentInfo.do?documentID=1289. Accessed December 2012.

Van Urk H, Postma E, Scheffers WA, Van Dijken JP (1989) Glucose transport in Crabtree-positive and Crabtree-negative yeasts. J Gen Microbiol 135:2399-2406

Van Zyl C, Prior BA, Du Preez JC (1988) Production of ethanol from sugar cane bagasse hemicellulose hydrolyzate by *Pichia stipitis*. Appl Biochem Biotechnol 17:357-369

Webb SR, Lee H (1990) Regulation of D-xylose utilization by hexoses in pentose-fermenting yeasts. Biotechnol Adv 8:685-697

Yang X, Zhang S, Zuo Z, Men X, Tian S (2011) Ethanol production from the enzymatic hydrolysis of non-detoxified steam-exploded corn stalk. Bioresour Technol 102:7840-7844

Yano S, Murakami K, Sawayama S, Imou K, Yokoyama S (2009) Ethanol production potential from oil palm empty fruit bunches in southeast asian countries considering xylose utilization. J Jpn Inst Energy 88:923-926