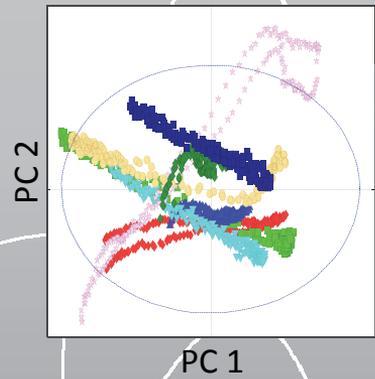
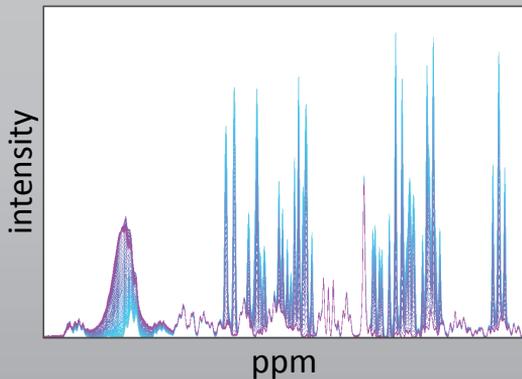
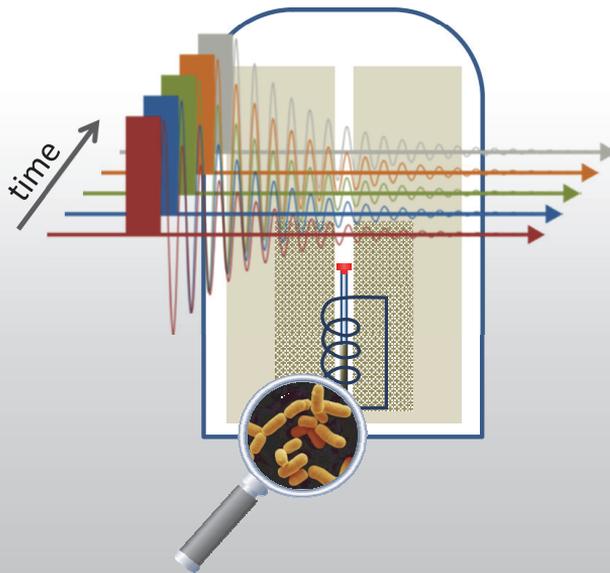




# Metabolic profiling of food protective cultures by *in vitro* NMR spectroscopy

PHD THESIS 2015 · PARVANEH EBRAHIMI



# **Metabolic profiling of food protective cultures by *in vitro* NMR spectroscopy**

PhD Thesis

**Parvaneh Ebrahimi**

September 2015

Spectroscopy & Chemometrics, Department of Food Science  
University of Copenhagen  
Denmark

**Title**

Metabolic profiling of food protective cultures by *in vitro* NMR spectroscopy

**Author**

Parvaneh Ebrahimi

**Date of Submission**

04 August 2015

**Date of Defense**

18 September 2015

**Supervisors**

Prof. Søren Balling Engelsen (*principal supervisor*), Spectroscopy & Chemometrics  
Department of Food Science, Faculty of Science  
University of Copenhagen

Assoc. Prof. Flemming Hofmann Larsen (*co-supervisor*), Spectroscopy & Chemometrics  
Department of Food Science, Faculty of Science  
University of Copenhagen

Senior Principal Scientist Henrik Max Jensen (*co-supervisor*), DuPont Nutrition Biosciences  
ApS, Brabrand, Denmark

**Assessment Committee**

Assoc. Prof. Åsmund Rinann (*chairman*), Spectroscopy & Chemometrics  
Department of Food Science, Faculty of Science  
University of Copenhagen

Prof. Royston Goodacre, Manchester Institute of Biotechnology and School of Chemistry,  
University of Manchester

Prof. Jens Øllgaard Duus, Department of Chemistry  
Technical University of Denmark (DTU)

# Preface

This thesis is submitted as a requirement for obtaining the PhD degree at the University of Copenhagen. The thesis is written based on the research by the author, Parvaneh Ebrahimi, that has been carried out at the Spectroscopy & Chemometrics group (SPECC), Department of Food Science. The study has been supervised mainly by Prof. Søren Balling Engelsen, and has been funded by The Danish Council for Strategic Research, and Dupont ApS (formerly Danisco A/S).

I am utterly grateful to my supervisor, Søren, for giving me the opportunity to perform this PhD, and for being such a knowledgeable and nice person to work with. He has supported the project endlessly during all the phases, and I have greatly benefited from his comments and novel ideas. I am also very grateful to my co-supervisors, Flemming Hofmann Larsen and Henrik Max Jensen. Special thanks to Flemming for all of his help with NMR parts of the study, and the scientific discussions. He has been always available to help me, which I highly appreciate. I would also like to thank Mathias Nilsson for introducing me to reference deconvolution and his scientific supports in this context.

My colleagues have made working at SPECC a splendid experience for me, and I am grateful to all of them. Special thanks go to Frans van den Berg for the nice collaboration that we had on some parts of the project, and for always being open to scientific discussions. I am also grateful to Rasmus Bro for the scientific discussions and his kind attitude that made me interested in the group in the first place.

I also take this opportunity to thank Maja H. Kamstrup-Nielsen, Tine Ringsted, and Gerrard Starr, for being good friends and company when we shared offices. I am also thankful to Stina D. Aunbjerg for her kind help and the discussions in the microbiology field.

And finally, thank you Hassan for your patience and understanding. Thanks for believing in me more than myself, and for being there for me unconditionally.

*Parvaneh Ebrahimi*  
*August 2015*

# Abstract

Food spoilage is of major concern to the food industry, because it leads to considerable economic losses, a deteriorated environmental food-print, and to possible public health hazards. In order to limit food spoilage, research on the preservation of food products has always received particular attention by the food industry. Traditionally, such efforts have mainly relied on the application of chemical preservatives or drastic physical treatments. However, chemical preservatives are becoming increasingly unpopular by the consumers, and some have even proven to be toxic and linked to cancer and other health problems. Physical treatments of the products, on the other hand, can deteriorate the sensory properties of the products, and may even destroy some of the nutrients and vitamins. In this context, biopreservation, which is defined as the use of safe antibacterial/antifungal microorganism (so-called protective cultures) has unexploited potential to inhibit the growth of pathogenic microorganisms and enhance the shelf life of the final food product. In order to apply biopreservation in food products effectively, detailed knowledge on the metabolism of protective cultures is required. The present PhD project is mainly focused on the application of *in vitro* NMR spectroscopy for studying the metabolism of protective cultures. As an important part of this work, an analytical protocol was developed for real-time *in vitro* NMR measurements of bacterial fermentation, which includes guidelines from the sample preparation to the data processing and the modelling of the metabolic profiles. The protocol is applied in an experimental design with two strains of lactic acid bacteria. The results highlight some of the metabolic differences between the strains, in terms of nutrients consumption and metabolites kinetics. As a part of this work, an NMR data preprocessing technique, called 'Reference Deconvolution', was employed for the first time to improve the multivariate analysis of the *in vitro* real-time metabolomics data and proved a necessary and elegant solution to the inherent inhomogeneity problem of the samples in the *in vitro* NMR measurements of cells. A second objective of the project was to develop an accurate approach for quantifying mold growth and inhibition. A new method was presented for quantifying mold growth and measuring different segments of mold colonies, based on multispectral images and *k*-means clustering. The method was developed into a software package called 'PCLUSTER', and was demonstrated to be very helpful in two other biopreservation related metabolomic studies. In one case, PCLUSTER was used to quantify how the concentration of diacetyl affects inhibition of the indicator molds and in the second case PCLUSTER served as an efficient tool for quantifying inhibition assays, and finding antifungal metabolites and metabolites that correlated positively/negatively with the

inhibition. The developed analytical tools are expected to be very beneficial in the studies related to the biopreservation, and will be used in the future investigations of the protective cultures.

# List of Publications

## Paper I

“Cleaning up NMR spectra with reference deconvolution for improving multivariate analysis of complex mixture spectra”, P. Ebrahimi, M. Nilsson, G. Morris, H. Jensen, and S.B. Engelsen. *Journal of Chemometrics*. Volume 28, Issue 8, pages 656–662.

## Paper II

“Real-time metabolomic analysis of lactic acid bacteria as monitored by *in vitro* NMR and chemometrics”, P. Ebrahimi, F. H. Larsen, H. M. Jensen, F. K. Vogensen, S.B. Engelsen. *Metabolomics*, Submitted.

## Paper III

“Quantitative determination of mold growth and inhibition by multispectral imaging”, P. Ebrahimi, F. van den Berg, S. Aunbjerg, A. Honoré, C. Benfeldt, H. Jensen, and S.B. Engelsen. *Food Control*. Volume 55, Pages 82–89.

## Paper IV

“Contribution of volatiles to the antifungal effect of *Lactobacillus paracasei* in defined medium and yogurt”, S. Aunbjerg, A. Honoré, J. Marcussen, P. Ebrahimi, F. Vogensen, C. Benfeldt, T. Skov, S. Knøchel. *International Journal of Food Microbiology*. Volume 194, Pages 46–53.

## Paper V

“Metabolic Footprinting for Investigation of antifungal properties of *Lactobacillus paracasei*”, A. Honoré, S. Aunbjerg, P. Ebrahimi, M. Thorsen, C. Benfeldt, S. Knøchel, T. Skov. Submitted to the *Journal of Analytical and Bioanalytical Chemistry*.

# List of Abbreviations

ASCA	Analysis of Variance-Simultaneous Component Analysis
LAB	Lactic Acid Bacteria
MCR-ALS	Multivariate Curve Resolution-Alternating Least Squares
NMR	Nuclear Magnetic Resonance spectroscopy
PARAFAC	Parallel Factor Analysis
PAT	Process Analytical Technology
PCA	Principal Component Analysis
PLS	Partial Least Squares regression
RD	Reference Deconvolution

# Table of Contents

<i>Preface</i> .....	<i>I</i>
<i>Abstract</i> .....	<i>II</i>
<i>List of Publications</i> .....	<i>IV</i>
<i>List of Abbreviations</i> .....	<i>V</i>
<b>1. Introduction</b> .....	<b>1</b>
1.1. The concept of food biopreservation .....	1
1.2. The main objectives of the project .....	2
1.3. A brief description of the outline of the thesis .....	2
<b>2. Food biopreservation</b> .....	<b>5</b>
2.1. Food spoilage and preservation .....	5
2.2. Biopreservation by lactic acid bacteria .....	6
2.3. History of biopreservation by lactic acid bacteria .....	10
2.4. Metabolism of lactic acid bacteria .....	11
2.5. Antimicrobial metabolites of lactic acid bacteria .....	17
2.6. The effects of stress factors on the metabolism of lactic acid bacteria .....	20
<b>3. Microbial metabolomics by NMR</b> .....	<b>23</b>
3.1. Principles of Nuclear Magnetic Resonance spectroscopy .....	23
3.2. NMR data processing .....	29
3.3. Enhancing quality of NMR spectra (Paper I) .....	35
3.4. Sensitivity in NMR spectroscopy .....	41
3.5. Quantitative NMR spectroscopy .....	41
3.6. Metabolomics .....	42
3.7. The role of NMR in metabolomics .....	45
3.8. NMR for studying microbial metabolism .....	46
3.9. <i>In vivo/in vitro</i> NMR spectroscopy for studying microbial metabolism .....	47
3.10. The application of <i>in vitro</i> NMR in micro-PAT .....	50
<b>4. Multivariate data analysis and chemometrics in metabolomics</b> .....	<b>51</b>
4.1. Principal Component Analysis (PCA) .....	51
4.2. Partial Least Squares regression (PLS) .....	53
4.3. Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS) .....	54
4.4. Analysis of Variance-Simultaneous Component Analysis (ASCA) .....	56

4.5.	Clustering by <i>k</i> -means algorithm.....	57
5.	<i>In vitro</i> NMR studies of the fermentation of lactic acid bacteria (Paper II).....	59
5.1.	Experimental design, sample preparation, and data acquisition.....	59
5.2.	Data processing .....	63
5.3.	Data analysis.....	65
5.4.	Outcome.....	71
6.	<i>Quantifying mold growth by multispectral imaging</i> .....	75
6.1.	Quantifying mold growth.....	75
6.2.	Multispectral imaging .....	77
6.3.	PCLUSTER software-- <i>a new approach for quantifying mold growth</i> (Paper III) .....	77
6.4.	Applications of PCLUSTER to metabolomics studies (Paper IV&V) .....	81
7.	<i>Conclusions</i> .....	87
8.	<i>Outreach</i> .....	89
	<i>References</i> .....	91



# 1. Introduction

This chapter provides a general overview and presentation of the objectives and the main directions of the thesis work, including a brief description of the thesis chapters, and the publications.

## 1.1. The concept of food biopreservation

Food spoilage is of major concern to the food industry, because it leads to considerable economic losses, and can also bring about public health problems. Therefore, the development and application of efficient preservation techniques is an obvious area of research when the shelf life of the food products needs to be extended. The common practice to preserve food relies on the application of chemical preservatives or drastic physical treatments. Although these methods are efficient in preserving food products, they may not, from other perspectives, be the best solutions. From consumers' health point of view, chemical preservatives are not desirable as some have proven to be toxic and have been linked to cancer, cardiovascular disease, and aging (Parke and Lewis 1992). Drastic physical treatments such as the application of high temperatures, on the other hand, can deteriorate the sensory properties of the products and destroy some of the nutrients and vitamins. One of the other solutions for the preservation of food products, which is relatively new to the food industry, is biopreservation. Biopreservation can be defined as the use of safe antibacterial/antifungal microorganism in food products to inhibit the growth of pathogenic microorganism and enhance the shelf-life (Annou S 2007). Biopreservation is natural and can obviate the concerns that are associated with other preservation techniques. The food industry is experiencing an increasing demand from the customers' part for products that are preserved naturally (i.e. having neither E-number nor any chemical preservatives on the ingredients list). Due to this consumers' trend, and to the fact that replacing the common means of preservation by biopreservation can significantly benefit the public health, expanding knowledge and insight into developing efficient biopreservation systems for different food products is of great importance.

In the context of food biopreservation, lactic acid bacteria (LAB) are definitely in the spotlight. Owing to their antagonistic properties and their metabolites, specific strains of LAB can preserve food. Numerous scientific studies have been carried out on the antimicrobial metabolism of LAB, but still many questions remain elusive. This demands further research

in the field and investigations on how optimum biopreservation can be provided for different food products.

## **1.2. The main objectives of the project**

The goal of the present PhD project was to investigate the potential of using NMR spectroscopy for studying the metabolism of antifungal strains of LAB, or so-called protective cultures. The project included real-time *in vitro* NMR measurements of the fermentation of LAB and developing an analytical protocol for the future metabolomic studies of protective cultures. The other aspect of the project was developing an approach for quantifying inhibition, as this would allow evaluating the antifungal properties of different strains or samples in an experimental design, and discovering metabolites that contribute to the antifungal effect.

## **1.3. A brief description of the outline of the thesis**

The different chapters in this PhD thesis present the research that has been performed to meet the objectives of the project, in addition to a concise presentation of the theory behind them. The thesis includes eight chapters and five peer-reviewed scientific journal papers. In the following a brief description on the content of the chapters and the publications is given.

*Chapter 1*, the current chapter, defines the project and its objectives. It explains the motive behind the project and its significance. *Chapter 2* describes the importance of preserving food, biopreservation, and the role of LAB in this context. It also describes some of the main metabolic pathways in LAB. *Chapter 3* presents the principles of NMR spectroscopy and metabolomics, and the reason for which NMR was selected as an analytical technique to investigate the microbial metabolism. *Chapter 4* provides brief theory and definitions of the multivariate data analysis and chemometrics methods that have been used throughout the project. *Chapter 5* presents information, figures, and discussion regarding the *in vitro* NMR measurements of LAB fermentation. The chapter further elaborates on some of the practical aspects that are not included in the corresponding scientific paper. *Chapter 6* presents the theory of multispectral imaging and explains the methodology behind ‘PCLUSTER’, the software that was developed for quantifying mold growth and inhibition, as well as more detail on the instructions of the software which are not included in the corresponding scientific paper. This chapter also describes two applications of PCLUSTER in biopreservation related metabolomic studies of LAB. It is demonstrated how PCLUSTER was successfully used to identify antifungal metabolites and investigate their effect on mold growth in a quantitative manner. *Chapter 7* presents the main conclusions of the research performed, and *Chapter 8* give the author’s personal account of the perspectives and outreach of the research, with

emphasis on how we can further improve our insight into the detailed molecular knowledge about biopreservation of food products.

## **Papers:**

Paper I demonstrates how reference deconvolution of NMR data can improve the results of multivariate analysis, using the data from a triangular experimental design. It presents the theory of reference deconvolution, and shows how reference deconvolution can improve PCA and PLS models of the data from the designed samples.

Paper II presents the results from using NMR spectroscopy to study real-time *in vitro* bacterial fermentation in a designed metabolomics study. The paper describes the developed protocol -from sample preparation to the kinetic modelling of metabolic changes- and makes suggestions regarding suitable data processing and data analysis techniques for *in vitro* NMR data. The paper addresses some of the challenges of *in vitro* measurements of cells and suggests solutions. Reference deconvolution is applied for the first time to a dynamic NMR study and the results clearly show that metabolomics can benefit from it.

Paper III presents the developed approach for quantifying mold growth and inhibition by using multispectral imaging and *k*-means clustering. The developed freeware is called 'PCLUSTER', and the paper demonstrates the application of PCLUSTER to three different sample sets (multispectral images), with the objective of quantifying mold growth and size of the colony segments of *Penicillium* molds.

Paper IV is a metabolic study on the inhibition of mold growth by one of the strains of *L. paracasei*, focusing on the volatile metabolites. In this study, PCLUSTER was used to analyze multispectral images from the inhibition assays and to investigate how the concentration of diacetyl affects the inhibition of the indicator molds.

Paper V presents another application of PCLUSTER in a biopreservation related study of two strains of *L. paracasei*. In this study, PCLUSTER software served as an efficient tool for quantifying inhibition assays, and was of great help in finding metabolites that contributed to the inhibition.



## 2. Food biopreservation

This chapter aims at defining the concept of biopreservation that is used as a natural, safe and ecological approach for food preservation to reduce or avoid food spoilage. Furthermore, the reasons that necessitate it for the food industry to investigate and use appropriate preservation systems for different food products are discussed.

### 2.1. Food spoilage and preservation

Food spoilage is defined as a metabolic process that can change the sensory characteristics of food products, such as taste, smell and appearance, and make the products undesirable or unsuitable for human consumption. Food spoilage is associated with public health concerns, environmental and resource costs, and finally considerable economic losses for industry. Food spoilage is caused by a variety of microbes that use food as a source of carbon and energy. Among the microorganisms that can spoil food are bacteria, yeasts and molds. Bacteria constitute a large domain of prokaryotic (lack defined nuclei and other organelles) single-celled microorganism. Yeasts are eukaryotic (have nuclei and other organelles) single-celled microorganisms that are classified as fungi, and molds are fungi that grow in the form of multicellular filaments called 'hyphae'. Some of the spoilage microbes are commonly found in many types of spoiled foods, whereas others are specific for certain food systems. Usually, multiple species are identified in a single type of spoiled food item, but there may be only one species (SSO: specific spoilage organism) that is primarily responsible for production of the compounds causing off-odor and flavor. Within a spoiling food system, there is often a succession of different populations that rise and fall, as different nutrients become available or are depleted (Doyle 2007).

Molds and yeasts are important spoilage organisms in food production. Each year, 5 to 10 percent of the world's food products are lost by fungal spoilage (Yang and Chang 2010). When the food is to be stored for a prolonged period, use of additives and preservatives is essential, in order to maintain its quality and flavor. Their use prevents products spoilage due to the growth of bacteria and fungi. If spoiled food is consumed by customers, besides causing health problems for the consumers, it can ruin the reputation of the manufacturing company. Economic losses of food spoilage are also another nontrivial factor which cannot be ignored by industry. So, it is extremely important for food industry to preserve food and extend the shelf-life of the products.

So far, the approaches to enhance food safety and increase its shelf-life have mainly relied on finding more efficient chemical preservatives or on the application of more drastic physical treatments, for instance high temperatures, refrigeration, application of high hydrostatic pressure (HHP), ionizing radiation, pulsed-light, ozone, ultrasound technologies, etc. However, there are various disadvantages associated with these methods. Most of the physical treatments can alter nutritional properties of the products, and chemical preservatives, such as nitrites, have proven to be toxic and have been linked to cancer, cardiovascular disease, and aging (Parke and Lewis 1992). Maybe it is because of these findings, that there has been an increasing consumers' trend in purchasing food products that are safe, minimally processed and free of chemical preservatives (Annou S 2007, Ghanbari, Jami et al. 2013). In this context, biopreservation represents a promising alternative to chemical preservatives, and if properly applied, biopreservation will ensure preservation of even minimally processed food products.

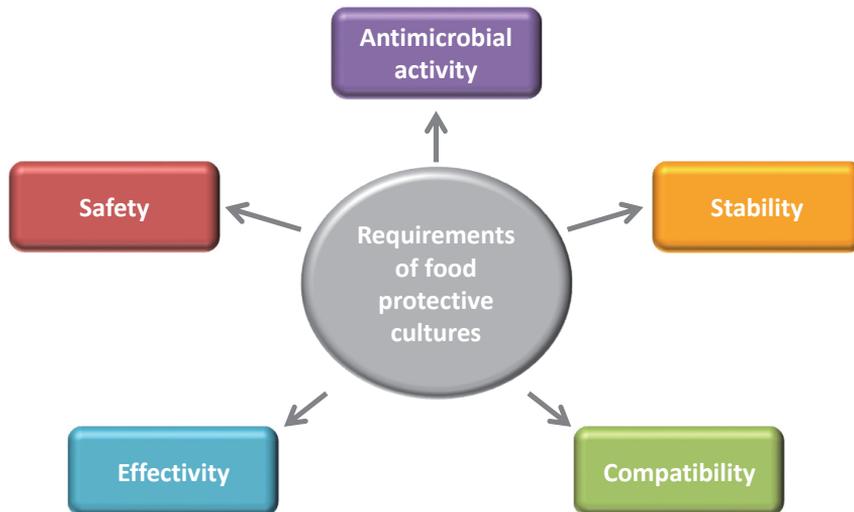
## **2.2. Biopreservation by lactic acid bacteria**

Biopreservation is defined as extension of shelf life and enhanced safety of food products by the use of natural or controlled microbiota and/or their antimicrobial compounds of defined quality and at certain quantities, while changing the sensory properties of the products as little as possible (Stiles 1996, Hugas 1998, Vermeiren, Devlieghere et al. 2004, Annou S 2007). Biopreservation is an innocuous and ecological approach to the problem of food preservation and has gained increasing attention in recent years. Biopreservation permits the application of less severe heat treatments without compromising food safety, provides better preservation of food nutrients and vitamins as well as organoleptic properties of foods, and can serve to satisfy industrial and consumers demands (Galvez, Abriouel et al. 2007).

Antagonistic cultures that are only added to inhibit pathogens and/or prolong the shelf life, while changing the sensory properties of the food product the least possible, are termed 'protective cultures' (Castellano and Vignolo 2006). Protective cultures do not necessarily have to ferment food to provide the preserving effect, and in this regard, they differ from the 'starter cultures' which are used specifically in the fermentation processes and cause a sensory alteration of the food. Although fermentation is a means of biopreservation, the concept of biopreservation is not limited to fermentation and is used in a broader sense.

When using food protective cultures for biopreservation, there are a number of criteria and requirements that should be taken into account. Some of the most relevant and important requirements are summarized in Figure 2.1. The most important requirement is that the protective microorganisms should be safe and not pose any risk to the consumers' health. Moreover, they should not alter the sensory characteristics of the food system and be

compatible with the food matrix. Another important requirement for a successful application of protective cultures is their efficiency and the ability to produce sufficiently active antagonistic metabolites against a broad range of relevant food-borne pathogen and/or spoilage bacteria and fungi. In addition, it is very important that the cultures are stable and capable of surviving adverse conditions encountered during the industrial treatment processes and storage of the products, and maintain their inhibitory activities (Annou S 2007, Ghanbari, Jami et al. 2013, Brosnan, Coffey et al. 2014).



**Figure 2.1.** The main requirements of food protective cultures.

Lactic acid bacteria (LAB) comprise a relatively diverse group of Gram-positive, non-sporulating, catalase-lacking, cocci or rods microorganisms. LAB produce lactic acid as the major end product during the fermentation of carbohydrates. They only grow in complex media where fermentable carbohydrates and higher alcohols are used as an energy source, mainly to form lactic acid (Calo-Mata, Arlindo et al. 2008). LAB, which are considered as ‘food-grade’ organisms, have a major potential for use in biopreservation, because they are safe to consume and during storage they naturally dominate the microflora of many foods (Stiles 1996). LAB are used to ensure safety, preserve food quality, develop characteristic new flavors, and to improve the nutritional qualities of food (Calo-Mata, Arlindo et al. 2008). Most LAB, due to their long history of safe use in food and feed fermentations, have received both GRAS (Generally Recognized As Safe) and QPS (Qualified Presumption of Safety) status in the EU, and thus, have good potential for future exploitation as antifungal biocontrol agents (Brosnan, Coffey et al. 2014). Properties that make LAB suitable for food preservation are: (i) they are generally regarded as safe, (ii) are not active and toxic on eukaryotic cells, (iii) become inactivated by digestive proteases, having little influence on the gut microbiota,

(iv) are usually pH and heat-tolerant, (v) have a relatively broad antimicrobial activity range, and can inhibit growth of many food-borne pathogenic and spoilage bacteria, and (vi) genetic manipulation of them is easy (Galvez, Abriouel et al. 2007). LAB are also increasingly being used as probiotics, owing to their contribution to the healthy microflora of human mucosal surfaces. They have also been introduced into animal feed owing to their contribution to the health of farmed animals and as biological control agents in aquaculture (Calo-Mata, Arlindo et al. 2008).

LAB are widespread in most ecosystems and are found in soil, water, plants, and animals. Besides being responsible for many food fermentation processes, they are also commonly found in non-fermented foods such as meat products, seafood, fruits, vegetables, and cereals. They can also be found in the intestinal and respiratory tracts of humans and animals. LAB are widely used as starter cultures in the food industry for the production of fermented foods, including dairy (yogurt, cheese), meat (sausages), fish, cereals (bread and beverages such as beer), fruit (malolactic fermentation processes in wine production), and vegetables (sauerkraut and kimchi), and have associations with many different foods. Table 2.1 provides a list of different food systems and their associated LAB.

It should be noted that not all LAB strains and species are safe. For instance, the genus *Streptococcus* includes many human and animal pathogens. Despite this, *Streptococcus thermophilus* is an important nonpathogenic organism that is used in the manufacture of yogurt and several cheese types. The best characterized LAB are those associated with milk fermentations, especially the subspecies of *Lactococcus lactis* (*L. lactis*). They preserve food by the low pH and lactic acid that they produce, as well as bacteriocins, in particular nisin that has found widespread application as a food preservative (Stiles 1996).

Research into antifungal properties of LAB and biopreservation has increased in the past decade and shows that many LAB strains have the potential to combat the proliferation of fungi in various food and feed systems. In particular, as mentioned before, the demand to reduce fungal contamination in foods has seen an increased demand by the consumer for the replacement of artificial chemical preservatives with natural biopreservatives (Annou S 2007). Studies into the applications of antifungal LAB strains are quite diverse and typically involve both *in vitro* and *in situ* food model studies, where a known fungus is tested against the antifungal LAB strain and where successful, applied to a final food (bread, cheese, yogurt) or beverage (fruit juice/fermented drink) products. The most promising areas where these applications have seen success include: (i) breads, (ii) dairy products, (iii) fruits and vegetables, (iv) silage, (v) seafood, and (vi) beverages like orange juice, and fermented seaweed drinks (Brosnan, Coffey et al. 2014).

**Table 2.1.** Different food systems and their associated lactic acid bacteria.

Food Types	Lactic acid bacteria
Milk and dairy food	
- Hard cheeses and Edam cheese	<i>L. lactis</i> subsp. <i>cremoris</i> , and subsp. <i>lactis</i> <i>L. lactis</i> subsp. <i>cremoris</i> , and subsp. <i>lactis</i> <i>L. mesenteroides</i> subsp. <i>cremoris</i>
- Cultured butter, buttermilk, and Gouda cheese	<i>L. lactis</i> subsp. <i>cremoris</i> , and subsp. <i>lactis</i> , and var. <i>diacetylactis</i> <i>L. mesenteroides</i> subsp. <i>cremoris</i> .
- Swiss type cheeses	<i>L. delbrueckii</i> subsp. <i>bulgaricus</i> <i>L. helveticus</i>
- Dairy products in general	<i>L. brevis</i> <i>L. buchneri</i> <i>L. casei</i> <i>L. paracasei</i> <i>L. fermentum</i> <i>L. plantarum</i> <i>L. mesenteroides</i> subsp. <i>cremoris</i> <i>L. Lactis</i>
Fermented milk products	
- yogurt	<i>Streptococcus thermophiles</i> <i>L. delbrueckii</i> subsp. <i>bulgaricus</i> <i>L. lactis</i> subsp. <i>diacetylactis</i>
- acidophilus milk	<i>L. Acidophilus</i>
- kefir	<i>L. kefir</i> <i>L. kefiranofaciens</i>
Meat products	
- raw	<i>Carnobacterium divergens</i> <i>C. piscicola</i> ( <i>maltaromicus</i> ) <i>L. sake</i> <i>L. curvatus</i> <i>L. carnosum</i> <i>L. gelidum</i>
- Fermented meat	<i>Pediococcus acidilactici</i> <i>Pediococcus pentosaceus</i> (inoculated into semi-dry sausages) <i>L. sake</i> <i>L. curvatus</i> <i>L. farciminis</i> (uninoculated)

<b>Food Types</b>	<b>Lactic acid bacteria</b>
Marinated fish products	<i>L. alimentarius</i> <i>C. piscicola</i>
Fermented vegetables	
- cucumbers, sauerkraut	<i>L. mesenteroides</i> (initial fermentation) <i>L. bavaricus</i> <i>L. brevis</i> <i>L. sake</i>
- olives <sup>22</sup> .	<i>L. plantarum</i>  <i>L. mesenteroides</i> <i>L. pentosus</i>
- fermented vegetables in general	<i>Pediococcus acidilactici</i> <i>P. pentosaceus</i> <i>L. plantarum</i> <i>L. sake</i> <i>L. buchneri</i> <i>L. fermentum</i>
Baked goods	
- sourdough bread	<i>L. sanfrancisco</i> (wheat and rye sourdough) <i>L. farciminis</i> <i>L. fermentum</i> <i>L. brevis</i> <i>L. plantarum</i> <i>L. amylovorus</i> <i>L. reuteri</i>
Wine (malo-lactic fermented)	<i>L. oenos</i>

*Adapted from (Stiles 1996).*

### **2.3. History of biopreservation by lactic acid bacteria**

Fermentation was the first form of biopreservation that was used by human beings. Fermentation seems to have originated from the Indian subcontinent and indications of developed agricultural and animal husbandry have been found that date back to before 1700 BCE. Besides, artifacts discovered from Egypt and Middle East also suggest that fermentation was known to human from ancient times. Green plant material is one of the natural sources of lactic acid bacteria, and in the past the involuntary contamination of raw milk with a variety of lactic acid bacteria must have led to fermented products during the storage. The conversion of lactose to acids and mainly lactic acid prevented the growth of undesirable bacteria, and it

may have been a mere accident when people first experienced the taste of fermented food. Later on, applications of fermentation extended to other food products including bread, cheese, fermented vegetables like sauerkraut, and fermented meats. Fermentation became more and more popular over time, not only because of its preserving ability but also because of the desirable sensory properties that it adds to the fermented products (Ström, Sjögren et al. 2002, Farnworth 2008).

## **2.4. Metabolism of lactic acid bacteria**

The metabolism of LAB is very diverse, and can vary significantly between different strains. This definitely cannot be fully covered by this section, but some of the most important and relevant metabolic pathways of the commonly used strains of LAB in food industry will be briefly discussed in the following.

### **2.4.1. Fermentation pathways**

LAB do not possess a functional respiratory system, and obtain their energy by substrate-level phosphorylation. LAB can utilize sugars like glucose through either the homo- or heterofermentative pathways. In the homofermentative pathway, that is also called glycolysis, homolactic fermentation, or the Embden–Meyerhof–Parnas pathway (EMP), as shown in Figure 2.2, the only end-product is lactic acid. Oxidation of glyceraldehyde-3-phosphate to 1,3-bisphosphoglycerate produces NADH, which is oxidized to  $\text{NAD}^+$  when pyruvate is reduced to lactate by lactate dehydrogenases (LDH). Through glycolysis, 1 mole of glucose yields 2 moles of lactic acid, and 2 moles of ATP (Fugelsang and Edwards 2006, Von Wright and Axelsson 2011).

In the heterofermentative or heterolactic fermentation pathway, also known as the pentose phosphoketolase pathway, the hexose monophosphate shunt, or the 6-phosphogluconate pathway, in addition to lactic acid, significant amounts of  $\text{CO}_2$ , ethanol or acetate are produced from sugars. In the heterolactic fermentation, 1 mole of glucose yields only 1 mole of ATP if the intermediate acetyl phosphate is reduced to ethanol, and 2 moles of ATP if the acetyl phosphate is converted to acetic acid (Fugelsang and Edwards 2006, Von Wright and Axelsson 2011).

Hexoses other than glucose (mannose, galactose, and fructose) can also be utilized through these major pathways, after different isomerization and phosphorylation steps and being converted to either glucose-6-phosphate or fructose-6-phosphate. For galactose there are two different pathways, depending on whether it enters the cell as galactose-6-phosphate, or as

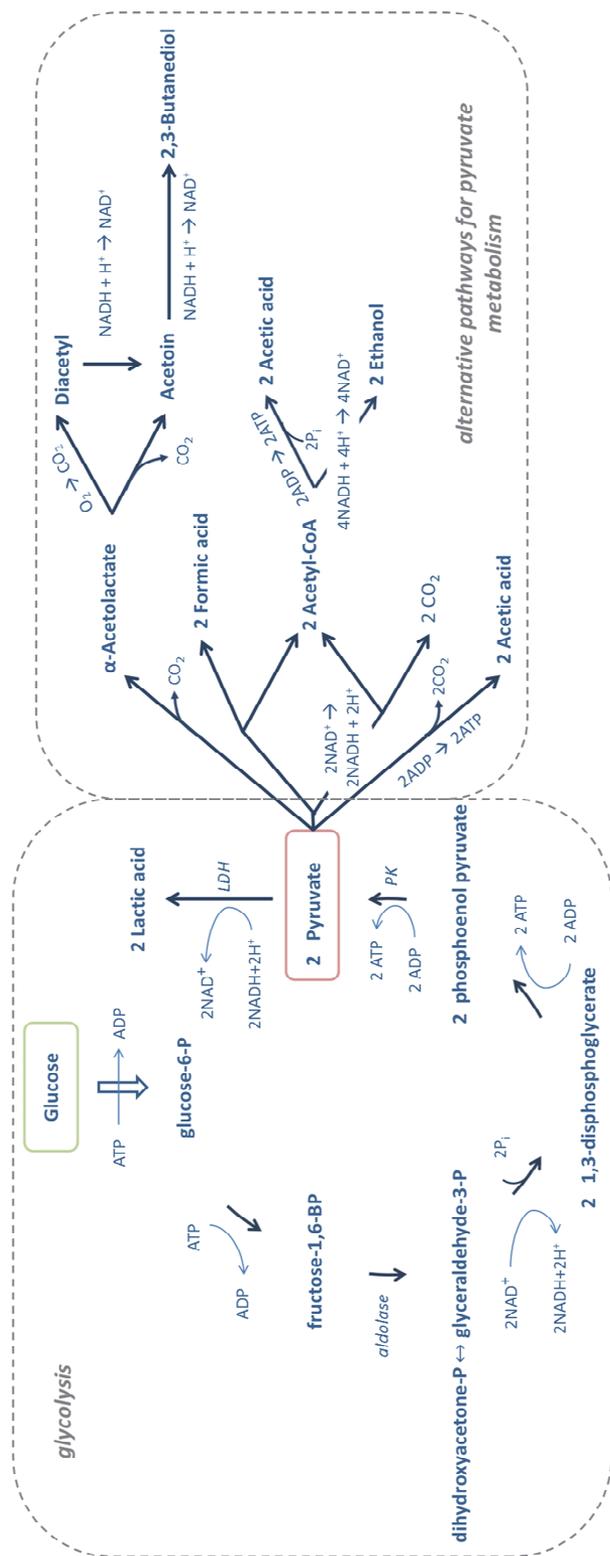
free galactose, but this will not be discussed in detail (Fugelsang and Edwards 2006, Von Wright and Axelsson 2011).

The type of fermentation is an important taxonomic criterion for LAB. LAB can be obligate homofermentative (Group I), facultative heterofermentative (Group II), or obligate heterofermentative (Group III). Obligate homofermentative lactobacilli (e.g. *L. acidophilus*, and *L. delbrueckii*) ferment hexoses homofermentatively, but cannot metabolize pentoses. The characteristic enzyme in microorganisms that can metabolize hexoses by glycolysis is 'aldolase', which as Figure 2.2 shows, catalyzes the conversion of 1 mole of fructose-1,6-bisphosphate to 2 moles of glyceraldehyde-3-phosphate. Obligate heterofermentative lactobacilli (e.g. *L. brevis*, and *L. fructivoransi*) lack aldolase, and divert carbon flux through the pentose phosphate, or phosphoketolase (heterolactic fermentation) pathways. These lactobacilli possess phosphoketolase, the enzyme responsible for the formation of glyceraldehyde-3-phosphate and acetyl phosphate from xylulose-5-phosphate. Facultatively heterofermentative lactobacilli (including *L. casei*, and *L. plantarum* that is investigated in Paper V) ferment hexoses homofermentatively, and also ferment pentoses. These bacteria contain aldolase, and besides that, production of phosphoketolase is induced by pentoses (Fugelsang and Edwards 2006, Von Wright and Axelsson 2011).

#### **2.4.2. Alternative pathways for pyruvate**

Pyruvate has a key role in the fermentation pathways, and also helps to maintain the oxidation-reduction (redox) balance in the cell, by acting as an electron acceptor to form lactic acid. Depending on the type of the bacteria and the physiological conditions, pyruvate can be converted into other metabolites than merely lactate, through other biochemical pathways. Especially, under carbohydrate source limitation, the metabolism can shift towards mixed-acid fermentation, in which acetic acid, formic acid, succinic acid and also ethanol can be produced besides lactic acid (Liu 2003, Lahtinen, Ouwehand et al. 2011). Some of the alternative pathways for pyruvate consumption are shown in Figure 2.2.

Under carbohydrate source limitation and under anaerobic conditions, LAB can metabolize pyruvate to formic acid and acetyl-CoA in a reaction catalyzed by pyruvate-formate lyase (Thomas, Ellwood et al. 1979, Kandler 1983). The acetyl-CoA formed can act as an electron acceptor to yield ethanol, or it can be used for substrate-level phosphorylation and subsequent ATP synthesis, giving acetate as the end-product. Therefore, even LAB species with homolactic hexose metabolism can under certain conditions have lactate, acetate, formate, and ethanol as the final metabolic end-products. This phenomenon is called 'mixed acid fermentation' and is different from the normal heterolactic fermentation (Von Wright and Axelsson 2011).



**Figure 2.2.** Glucose metabolism in lactic acid bacteria. Homofermentative (glycolysis) pathway, as well as the alternative pathways for pyruvate metabolism are shown.

Under aerobic condition, pyruvate can be converted to acetate by the pyruvate oxidase, where  $H_2O_2$  is also formed. This pathway may lead to the significant aerobic formation of acetic acid. It has been shown that *L. lactis* can undergo homoacetic fermentation in cultures under substrate limitation and aerobic conditions (Smart and Thomas 1987, Von Wright and Axelsson 2011).

In many strains of LAB, pyruvate can also be converted to diacetyl (butter aroma), acetoin, or 2,3-butanediol. Diacetyl is a product of citrate metabolism and is responsible for the aroma and flavor of certain fermented dairy products. Pyruvate can be converted to  $\alpha$ -acetolactate by acetolactate synthase (Lahtinen, Ouwehand et al. 2011), which according to the pathways shown in Figure 2.2, can be subsequently converted to diacetyl and acetoin.  $\alpha$ -acetolactate is converted to diacetyl spontaneously through a slow chemical, and non-enzymatic reaction and diacetyl can be subsequently reduced to acetoin by diacetyl reductase.  $\alpha$ -Acetolactate can also be directly converted into acetoin by acetolactate decarboxylase (Caspi, Altman et al. 2014). This pathway and synthesis of  $\alpha$ -acetolactate is active only under the condition of surplus pyruvate, and as reported in literature, will be enhanced at lower pH values (Le Bars and Yvon 2008, Von Wright and Axelsson 2011).

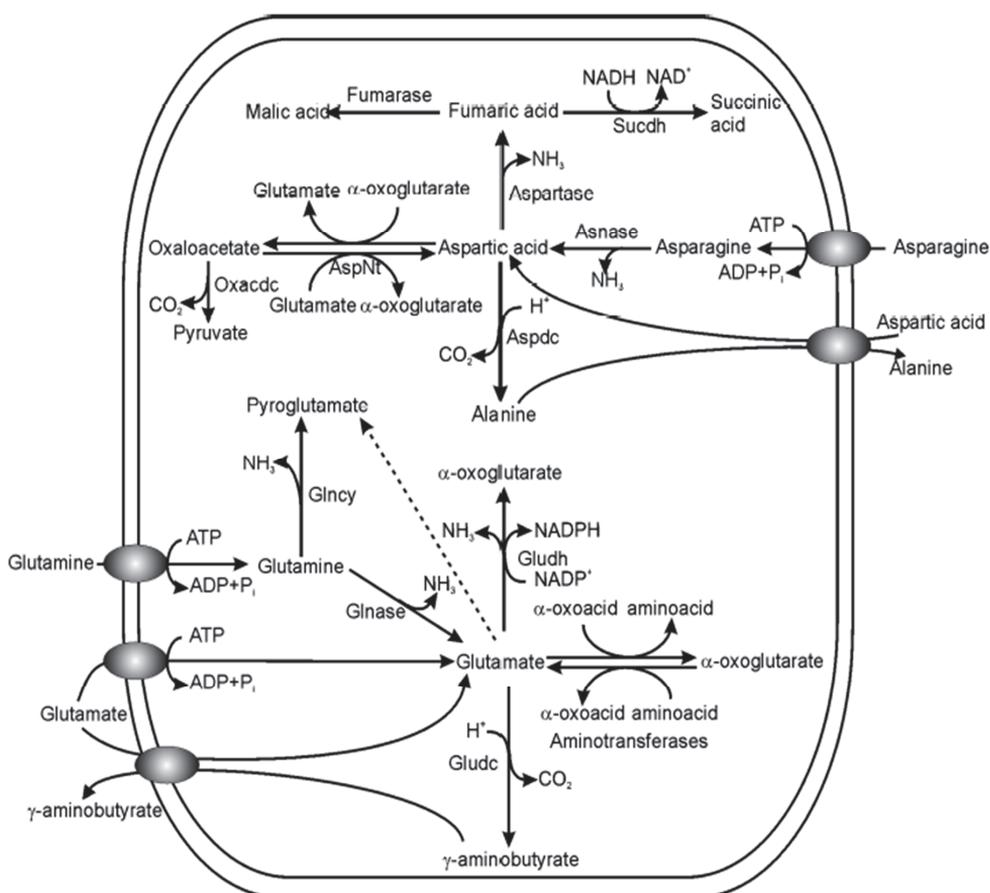
### **2.4.3. Catabolism of amino acids**

Metabolism of amino acids by LAB have important physiological roles including intracellular pH control, controlling the redox state of the cells, and being involved in metabolic stress responses. A wide range of general and specific regulators are involved in the catabolism of amino acids by LAB, which varies significantly between different strains. As LAB are widely used for the production of fermented food products by the industry, studying the catabolism of amino acids by LAB is relevant for the safety and the quality of fermented products. Studying the metabolism of amino acids by LAB has received special attention, because the products from amino acids catabolism can significantly enhance the sensory properties of food products. It is reported that in general, all amino acids can be metabolized by LAB, however there are significant differences between strains in their ability to degrade amino acids. Different enzymes are involved in the catabolism of LAB, including aminotransferases, decarboxylases, lyases, and dehydrogenases (Tammam, Williams et al. 2000, Williams, Noble et al. 2001, Liu, Holland et al. 2003). In the following, the catabolism of several amino acids are briefly described.

Figure 2.3 shows the catabolic pathways of asparagine, aspartic acid, glutamine, and glutamic acid. Asparagine degradation generally starts by the asparaginases activity that catalyzes the hydrolysis of asparagine to aspartic acid and ammonia. For aspartic acid, three catabolic pathways have been reported that are catalyzed by three different enzymes: aspartate aminotransferase, aspartase, and aspartate decarboxylase. The pathway catalyzed by the

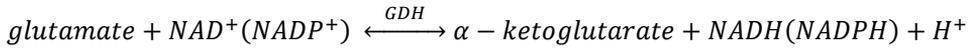
aspartate aminotransferase produces oxaloacetate and pyruvate, the pathway catalyzed by aspartase, depending on the physiological conditions can produce fumaric acid, succinic acid or malic acid, and finally the pathway catalyzed by aspartate decarboxylase can synthesize alanine (Fernández and Zúñiga 2006).

There is not detailed knowledge on the catabolism of glutamine by LAB, but it has been observed that several strains of LAB can metabolize it (Kieronczyk, Skeie et al. 2001, Williams, Noble et al. 2001). Glutamine cyclotransferase activity that converts glutamine into pyroglutamate and ammonium has been reported for thermophilic LAB such as *L. delbrueckii*, *L. helveticus*, and *Streptococcus thermophiles* (Mucchetti, Locci et al. 2002).



**Figure 2.3.** Catabolic pathways of asparagine, aspartic acid, glutamine, and glutamic acid by LAB. The following abbreviations are used: Aspdc, Asp decarboxylase; AspNt, Asp aminotransferase; Asnase, asparaginase; Glnase, Gln cyclase; Gludc, Glu decarboxylase; Gludh, Glu dehydrogenase; Oxadc, oxaloacetate decarboxylase; Sucdh, succinate dehydrogenase Reprinted from (Fernández and Zúñiga 2006).

Glutamic acid plays an important role as an intermediate in the catabolism of amino acids by LAB, as all the aminotransferases utilize glutamic acid as the donor substrate of amino groups. LAB strains that have active glutamate dehydrogenase (GDH), can produce  $\alpha$ -oxoglutarate from the deamination of glutamic acid, by the biochemical reaction (Tanous, Kieronczyk et al. 2002):



$\alpha$ -oxoglutarate ( $\alpha$ -ketoglutarate) has a key role in amino acid catabolic pathways and a number of studies have shown that the addition of exogenous  $\alpha$ -oxoglutarate can significantly increase the metabolism of amino acids by several LAB (Rijnen, Courtin et al. 2000, Tammam, Williams et al. 2000, Williams, Withers et al. 2000, Tanous, Kieronczyk et al. 2002, Helinck, Le Bars et al. 2004).

It has been reported that some strains of LAB can metabolize alanine when  $\alpha$ -oxoglutarate is present, suggesting that alanine is metabolized by transamination (Tammam, Williams et al. 2000, Williams, Noble et al. 2001). On the other hand, it has been reported that some other strains of LAB do not require  $\alpha$ -oxoglutarate to utilize alanine (Liu, Holland et al. 2003), that can be explained by strain variations. Alanine racemase activity has been observed in *L. reuterii* (Thompson, Griffin et al. 2002), and *L. plantarum* (Hols, Defrenne et al. 1997). Alanine racemase plays an important role in alanine catabolism and is essential for cell wall biosynthesis in *L. plantarum* (Palumbo, Favier et al. 2004) and *L. lactis* (Steen, Palumbo et al. 2005).

It is shown by several studies that some strains of LAB can utilize proline (Tammam, Williams et al. 2000, Williams, Noble et al. 2001, Liu, Holland et al. 2003), but the catabolic pathways are not known. Proline can be converted to 1-pyrroline-5-carboxylate which can then be oxidized to glutamic acid, by proline dehydrogenase activity. Proline can also be utilized by pyrroline-5-carboxylate reductase, but the reaction is reversible and it can also be used in the biosynthesis of proline (Kenklies, Ziehn et al. 1999). It has been observed that some strains of LAB accumulate proline in response to osmotic stress (Jewell and Kashket 1991, Molenaar, Hagting et al. 1993, Glaasker, Konings et al. 1996, Baliarda, Robert et al. 2003). Therefore, proline's main role may be in osmotic stress response rather than a source of energy (Poolman and Glaasker 1998, Fernández and Zúñiga 2006).

## 2.5. Antimicrobial metabolites of lactic acid bacteria

The antagonistic and inhibitory properties of LAB refer to the inhibition of other microorganisms through competition for nutrients and/or by the production of one or more antimicrobial metabolites.

The antimicrobial metabolites can be divided into the low molecular mass non-proteinaceous compounds and the proteinaceous bacteriocins that are antimicrobial peptides. Low molecular weight compounds have a wide range of activity against both Gram-positive and Gram-negative bacteria and fungi, whereas bacteriocins mainly inhibit Gram-positive bacteria (Niku-Paavola, Laitila et al. 1999). The diversity of compounds that have been shown to provide antifungal activity include organic acids (lactic and acetic acids), diacetyl, acetoin, phenyllactic acid, fatty acids, cyclic dipeptides, reuterin, reutericyclin, antifungal peptides, bacteriocines, lactones, hydrogen peroxide, and others (Galvez, Abriouel et al. 2007, Calo-Mata, Arlindo et al. 2008, Ghanbari, Jami et al. 2013, Brosnan, Coffey et al. 2014). A summary of some of the LAB antimicrobial compounds that have been identified and their inhibition spectrum is presented in Table 2.2.

One of the main factors that lead to the preservation effect of LAB is definitely the production of organic acids. Acidification is highly used for preservation of food products like milk, vegetables, and sausages. LAB are characterized by their tolerance to low pH, allowing their growth under circumstances in which other bacteria are unable to grow.

Organic acids have different modes of action for exerting their antimicrobial effects. The inhibition mechanism of lactic acid is probably related to the different solubility properties of its acid-base conjugates. The non-dissociated lactic acid can dissolve within the cytoplasm membrane, whereas its dissociated form is insoluble. This results in the acidification of cytoplasm and influences the trans-membrane pH gradient which in turn decreases the amount of available energy for the cells of other bacteria (Oda, Saito et al. 2002). Lactic acid can also interfere with metabolic processes such as oxidative phosphorylation (Calo-Mata, Arlindo et al. 2008). Acetic acid has proved to inhibit the growth of Gram-positive and Gram-negative bacteria, as well as yeasts and fungi. This effect is pH dependent and is more pronounced at pH values below 4.5. It is the un-dissociated form of acetic acid that contributes to the antimicrobial effect. Although its mechanism of action is not fully known, un-dissociated acetic acid may enhance lipid solubility, resulting in increased fatty acid accumulation on the cell membrane and consequently the rapid dissolution of the membrane. Lactic acid and acetic acid mixtures have a synergistic effect that reduces the minimum inhibitory concentration (MIC) of each acid in the mixture.

**Table 2.2.** A summary of some of the antimicrobial compounds identified from lactic acid bacteria.

Class	LAB	Source	Compound(s)	Inhibition spectrum
Protein	<i>L. coryniformis</i> Si3	sourdough	proteinaceous compounds (approximately 3 kDa)	mold: <i>Aspergillus</i> , <i>Penicillium</i> , <i>Mucor</i> , <i>Talaromyces</i> , <i>Fusarium</i>  yeast: <i>Debaryomyces</i> , <i>Kluyveromyces</i>
	<i>L. paracasei</i> subsp. <i>paracasei</i> M3	cheese	proteinaceous compounds (43 kDa)	<i>Saccharomyces cerevisiae</i>
Other compounds	<i>L. reuteri</i>	pig intestine	reuterin	mold: <i>Aspergillus</i> , <i>Fusarium</i>  yeast: <i>Saccharomyces</i> , <i>Candida</i> , <i>Torulopsis</i>
	<i>L. sanfrancisco</i> CB1	sourdough	caproic acid, propionic acid, butyric acid, valeric acid	<i>Fusarium</i> , <i>Penicillium</i> , <i>Aspergillus</i> , <i>Monilia</i>
	<i>L. plantarum</i> VTT E-78076	beer	benzoic acid, methylhydantoin, mevalonolactone, cyclo (Gly-Leu)	<i>Fusarium avenaceum</i>
	<i>L. plantarum</i> 21B	sourdough	phenyllactic acid, 4-hydroxyphenyllactic acid	<i>Aspergillus</i> , <i>Eurotium</i> , <i>Endomyces</i> , <i>Penicillium</i> , <i>Monilia</i>
	<i>L. plantarum</i> MiLAB 393	grass silage	phenyllactic acid, cyclo(Phe-Pro), cyclo(Phe-OH-Pro)	mold: <i>Aspergillus</i> , <i>Penicillium</i> , <i>Fusarium</i>  yeast: <i>Candida</i> , <i>Debaryomyces</i> , <i>Kluyveromyces</i> , <i>Rhodotorula</i> , <i>Saccharomyces</i> , <i>Phichia</i>
	<i>L. plantarum</i> MiLAB 14	lilac flowers	hydroxy fatty acids (3-hydroxydecanoic acid, 3-hydroxy-5-cis-dodecenoic acid, 3-hydroxydodecanoic acid, 3-hydroxytetradecanoic acid)	mold: <i>Aspergillus</i> , <i>Penicillium</i>  yeast: <i>Kluyveromyces</i> , <i>Phichia</i> , <i>Rhodotorula</i>
	<i>L. plantarum</i> FST 1.7	malted barley	phenyllactic acid, cyclo(Leu-Pro), cyclo(Phe-Pro)	<i>Aspergillus niger</i> , <i>Fusarium</i>
	<i>L. plantarum</i> AF1	kimchi	cyclo(Leu-Leu), unidentified compounds	<i>Penicillium</i> , <i>Aspergillus</i> , <i>Epicoccum</i> , <i>Cladosporium</i>

Table modified from Ref. (Yang and Chang 2010).

Benzoic acid and its sodium salt are commonly used in the food industry as antifungal agents (Hazan, Levine et al. 2004). Dairy products can contain benzoic acid as some LAB are able to produce it (Garmiene, Salomskiene et al. 2010). The mechanism of benzoic acid action seems to be hindering macroautophagy. Macroautophagy is thought to allow cellular physiology to continue in the absence of external nitrogen resources, by the degradation of cellular contents (Hazan, Levine et al. 2004). Propionic acid is used in food as an antifungal agent, but can also hinder the growth of Gram-positive and Gram-negative bacteria. Propionic acid is produced by heterofermentative LAB in trace amounts. Propionic acid interacts with cell membranes and neutralizes the electrochemical proton gradient. The antifungal inhibition effect of propionic acid is often dependent on the decrease in pH caused by lactic acid, and is higher at lower pH values. Propionic acid also inhibits amino acid uptake. Salts of propionic acid, such as sodium propionate and ammonium propionate also show similar antifungal and anti-yeast effects at low pH (Reis, Paula et al. 2012).

Most Lactobacilli species are able to produce hydrogen peroxide by the oxidation of lactate.  $H_2O_2$  has antimicrobial effect even at refrigeration temperatures and can be beneficial in food preservation (Thomas, Milligan et al. 1994). In the presence of lactoperoxidase,  $H_2O_2$  can oxidize thiocyanate ion ( $SCN^-$ ) into hypothiocyanite ( $OSCN^-$ ) which has antimicrobial characteristics. This mode of action is mainly active against microorganisms that produce  $H_2O_2$  as it requires both hydrogen peroxide and thiocyanate for optimum activity. However, hydrogen peroxide on its own is also known to be bactericidal, depending on the concentrations applied and the environmental factors such as pH and temperature (Reis, Paula et al. 2012).

Bacteriocins are in general, cationic, amphiphilic, and membrane-permeabilizing antimicrobial peptides. Four classes of bacteriocins (I, II, III and IV) have been defined based on their chemical structure, molecular weight and thermal stability. Bacteriocins can also be classified based on their mode of action. Some of the members of the class I (lantibiotic) bacteriocins, such as nisin have shown a dual mode of action. They can prevent correct cell wall synthesis of bacteria by binding to peptidoglycan transporter lipids, and leading to the death of the cells. Moreover, they can also cause cells death by the formation of pores in the cell membrane. Large bacteriolytic proteins (bacteriolysins, formerly Class III), such as lysostaphin, can act directly on the cell wall of Gram-positive targets, leading to death and lysis of the target cell (Klaenhammer 1988, Reis, Paula et al. 2012).

To expand the knowledge of antimicrobial properties of LAB, it is important to continuously expand our understanding of the influences that environmental factors have on the implantation and survival of protective cultures. Investigating the effect of different internal and external factors -such as pH, carbohydrate concentration, and temperature- on the metabolism of the strains and the excretion of antifungal compounds will allow optimizing all

the key factors and parameters in a way that the most efficient biopreservation system is ensured. It is also very important to select the optimum strain(s) for the desired food system and its specific conditions (e.g. its composition and storage condition), and to gain insight into the antifungal mechanisms of the selected protective cultures. Studying synergies between LAB and other bacterial cultures that can lead to enhanced biopreservation systems are also of interest.

## **2.6. The effects of stress factors on the metabolism of lactic acid bacteria**

One of the areas that metabolomics has also been used in, has been investigating the metabolic responses of microorganisms to different environmental stress factors. The pattern of gene expression and the metabolome can alter considerably in response to external factors such as temperature, pH, ionic strength, and organic pollutants. These changes aim at adapting the cells to the new environment and increase viability (Lankadurai, Nagato et al. 2013). Stress factors usually induce a broader range of metabolites than just those that result from the major metabolic routes. In a food system, the stress derived metabolites, besides improving the survival and viability of cells under stress conditions in, can improve the sensory characteristics of food products.

NMR has a great potential for investigating the effect of different stress factors on the metabolism of microorganisms and the metabolic shifts, as it allows *in vitro* measurements of bacterial fermentation. It allows rapid and quantitative measurement of the metabolites and is a very useful tool for discovering the stress response mechanism of microorganisms.

As in other bacteria, LAB also use adaptive responses to protect their cells under stress responses. Except for some species, LAB in general are neutrophils and the optimum pH for their growth is between 5 and 9. At low pH values, the intracellular pH decreases and this affects the transmembrane pH gradient that is the motive force for many of the transmembrane transfer processes. The low intracellular pH value also reduces the activity of acid sensitive enzymes and can also damage proteins and DNA (De Angelis and Gobbetti 2004). Acid stress can occur during fermentation of LAB or in the gastronomical tract when they are used as probiotics. When bacteria encounter acid stress, different metabolic reactions are initiated to maintain the neutral intracellular pH. The most important mechanism for regulating the homeostasis of internal pH in fermentative bacteria is proton ATPase that exports proton from the cytoplasm (Hutkins and Nannen 1993). For some of the strains including *L. plantarum*, the optimum pH value for ATPase is 5.0-5.5, whereas for other strains like *L. lactis* the optimum interval is 7.0-7.5 (Nannen and Hutkins 1991). The other important acid-stress response mechanism is arginine deaminase (ADI) pathway. The catabolism of arginine and the production of ammonium through this pathway help to achieve

optimum internal pH. Besides, the extra ATP that is produced enables the extrusion of cytoplasmic protons. Selecting strains that can express ADI pathway, in addition to the protection against acid stress, can enhance the sensory properties of food products. Decarboxylation of amino acids can also help to resist acid stress by consuming proton. Conversion of glutamine to glutamate can also help to maintain intracellular pH close to neutral. Glutamine can further be metabolized to  $\gamma$ -amino butyrate or glutamine. Malolactic fermentation can also help bacteria to cope with acid stress (Serrazanetti, Guerzoni et al. 2009). It is shown that if the cells are pre-exposed to sublethal pH, the viability of the cells will be higher during subsequent exposure to the lethal pH. This phenomenon is known as the acid tolerance response (ATR) (Jin, Zhang et al. 2012). ATR is used in fermented foods with probiotic strains to increase the viability of the cells in gastronomic tract.

High temperatures can denature proteins and damage cell membrane and nucleic acids. Heat stress can also lead to the decrease in the intracellular pH by disturbing the transmembrane proton gradient (Piper, Ortiz-Calderon et al. 1997). The heat shock response is mediated by the synthesis of specific proteins (Hecker, Schumann et al. 1996). The genetic variation, the physiological state and environmental factors such as the growth medium, pH, and ionic strength can affect the heat resistance of lactobacilli. It is shown that for some of the species exposing the cells to a relatively high temperature before the lethal temperature challenge can considerably increase the viability of the cells (De Angelis and Gobbetti 2004).

Depletion of the carbohydrate source or other essential nutrients, as well as the accumulation of the fermentation end products, can limit the exponential growth of bacterial cells and enforce the growth to enter the stationary phase. Other stress factors such as heat, cold, osmotic, and oxidative stress can also enforce the growth to enter the stationary phase. The adaptive responses during the exponential phase of growth and the stationary phase are different. The adaptive responses during the exponential phase of growth involve only particular groups of genes to cope with specific stress factors. However, the stress responses in the stationary phase involve many groups of genes to cope with different stress factors. Besides, the responses in the stationary phase are developed without exposure to the stress conditions.

Stress factors can induce metabolic shifts and change the metabolic profiles, as the result of the changes in gene expression patterns. For instance, for *L. sanfranciscensis*, it was shown that acid stress can cause overproduction of 2- and 3-methylbutanoic acids, reduced sugar consumption, and reduced primary carbohydrate pathway metabolites. Moreover, the consumption of branched-chain amino acids increased up to seven times under acid stress (Serrazanetti, Ndagijimana et al. 2011), which can enhance the sensory properties of food products. Investigating the metabolic rerouting of the microorganisms under stress conditions

can lead to valuable knowledge that can be used for improving the quality of fermented food by using stress factors for the overproduction of desired metabolites.

## 3. Microbial metabolomics by NMR

This chapter aims at providing a short overview of the principles of Nuclear Magnetic Resonance (NMR) spectroscopy and applications of this analytical technique in metabolomics studies, with a focus on bacterial metabolomics. In the last section, *in vivo* NMR spectroscopy is introduced and possibilities that it can offer the metabolomics studies of live organisms and microorganisms are highlighted.

### 3.1. Principles of Nuclear Magnetic Resonance spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful analytical technique and can be used for qualitative and quantitative analysis of complex mixtures, kinetic studies, determination of the isotopic distribution within molecules, and structural elucidation of molecules (Ramos, Neves et al. 2002). Like any other spectroscopic method, NMR relies on the interaction between energy and matter (Weber and Thiele 2008), but unlike other spectroscopic methods it requires a strong static magnet field. The origin of NMR dates back to 1946 when two independent groups of scientists at Harvard University and Stanford University observed proton resonance signals from paraffin wax and water, respectively. The founding pioneers of the subject, Edward Purcell and Felix Bloch, were awarded the Nobel Prize in physics in 1952 “for their development of new methods for nuclear magnetic precision measurements and discoveries in connection therewith”. Later on, two of the pioneers of modern NMR methods were awarded the Nobel Prize in chemistry; Richard Ernst in 1991, “for his contributions to the development of the methodology of high-resolution NMR spectroscopy”, and more recently Kurt Wüthrich in 2002, “for his development of NMR spectroscopy for determining the three-dimensional structure of biological macromolecules in solution”. In 2003, Paul Lauterbur and Peter Mansfield, who laid the foundation for snapshot MRI (image formation from data acquired in a fraction of a second), shared the Nobel Prize in Physiology or Medicine “for their discoveries concerning magnetic resonance imaging” ([www.nobelprize.org](http://www.nobelprize.org) , Pekar 2006). The field of NMR has benefited a rapid growth as an analytical technique and although having originated in the physics laboratories, nowadays, NMR is applied in chemistry, biology, medicine, materials science, food science and nutritional studies (Claridge 2008, Savorani, Rasmussen et al. 2013, Spyros, Dais et al. 2013).

In the following, some of the main theoretical aspects of NMR spectroscopy are briefly discussed.

### 3.1.1. Nuclear spin and resonances

The nuclei of all atoms, which consist of protons and neutrons, can be characterized by a nuclear spin quantum number,  $I$ , which can simply be called the ‘nuclear spin’.  $I$  may take values greater than or equal to zero which are multiples of  $1/2$ . Nuclei with no nuclear spin ( $I=0$ ) are not observed by NMR and are called ‘NMR silent’ or ‘NMR inactive’. The atomic nucleus is assumed to be spherical and rotating around its axis. Angular momentum,  $P$ , of a nucleus is calculated as:

$$P = \sqrt{I(I + 1)}\hbar \quad (3.1)$$

where  $\hbar = h/2\pi$ , and  $h$  is Planck constant ( $6.6256 \times 10^{-34}$  J.s), and as described above,  $I$  is nuclear spin. A nucleus with  $I \neq 0$ , behaves as a small magnet and its magnetic moment,  $\mu$ , is quantized:

$$\mu = \gamma P \quad (3.2)$$

where  $\gamma$  is the magnetogyric ratio and is constant for each nuclide. Both angular momentum and magnetic moment are vector quantities, and are characterized by both magnitude and direction.  $\gamma$  can be considered as a measure of how magnetic a nuclide is. If the nuclide is placed in a static magnetic field,  $B_0$ , which exists along the z-axis of a static Cartesian coordinate system, the microscopic magnetic moments (nuclear spins) align themselves relative to the field and can take up to  $2I+1$  possible orientations relative to the field, with the values  $m_I \in -I, -I+1, -I+2, \dots, I-1, I$ . Each orientation has its own energy level or state that can be calculated as:

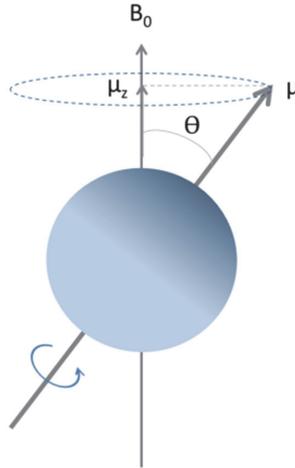
$$E = -\mu_z B_0 = -m_I \gamma \hbar B_0 \quad (3.3)$$

$\mu_z$  is the projection of the magnetic moment  $\mu$ , along the magnetic field (z-axis).

The static field causes a circular motion of the magnetic moment about the axis of the applied field, which is called ‘Larmor precession’. Each nuclide precesses with a frequency which is specific to that nuclide that is known as ‘Larmor frequency’,  $\nu_0$ , and is given by:

$$v_0 = \left(\frac{\gamma}{2\pi}\right)B_0 \quad (3.4)$$

Figure 3.1 shows a nuclear spin precessing in an external magnetic field  $B_0$ .



**Figure 3.1.** A nuclear spin precessing in an external magnetic field  $B_0$ .  
Adapted from (De Graaf 2008).

The angle between  $\mu$  and  $B_0$ ,  $\theta$ , is given by:

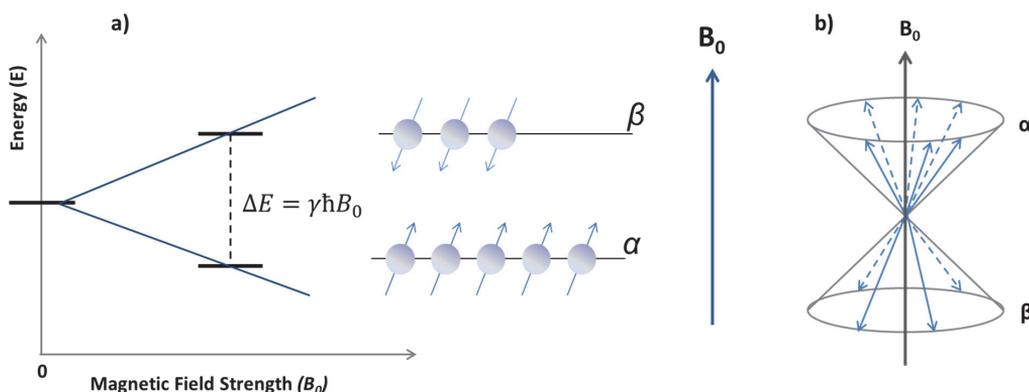
$$\cos \theta = \frac{m_I}{\sqrt{I(I+1)}} \quad (3.5)$$

So, when a nucleus with a  $1/2$  nuclear spin, such as  $^1\text{H}$ , is put in a magnetic field, the nuclei are distributed between two energy states, the parallel  $\alpha$  state ( $m_I = 1/2$ ) or antiparallel  $\beta$  state ( $m_I = -1/2$ ), according to the Boltzmann equation. The state has lower energy and a higher population. According to Boltzmann law, the relative populations of the lower ( $N_\alpha$ ) and the higher ( $N_\beta$ ) energy levels are given by:

$$\frac{N_\alpha}{N_\beta} = \exp\left(\frac{\Delta E}{K_B T}\right) = \exp\left(\frac{\gamma \hbar B_0}{K_B T}\right) \quad (3.6)$$

$\Delta E$  is the difference between the spins energy levels ( $E_\beta - E_\alpha$ ),  $K_B$  is the Boltzmann constant and  $T$  is the temperature .

Figure 3.2 presents a schematic illustration of the two mentioned energy states and a collection of the nuclei precessing around the  $B_0$  axis.



**Figure 3.2.** a) Nuclear energy states of a 1/2 nuclear spin, b) a collection of the nuclei precessing around the  $B_0$  axis. Adapted from (De Graaf 2008).

Based on the difference between the spins energy levels, it is radio waves from the electromagnetic spectrum that can cause the transitions between the different spin states.  $^1\text{H}$  is an NMR-active isotope of hydrogen which has a very high natural abundance and is the most commonly studied nuclide in NMR.

### 3.1.2. Chemical shifts and couplings

In the atom, electron cloud surrounds the nucleus. When the atom is subjected to an externally applied magnetic field  $B_0$ , a rotational motion is induced in the electron cloud and the electrons will rotate about  $B_0$  in an opposite sense relative to the proton spin precession. As electrons are charged particles, their motion will result in a magnetic moment. This gives rise to a small local magnetic field,  $B_{loc}$ , that may oppose the external field  $B_0$ , and as a result, the nucleus experiences a slightly reduced field, which can be called effective field,  $B_{eff}$ . According to the Lenz rule,  $B_{loc}$  is equal to  $\sigma B_0$ . The parameter  $\sigma$  is called the ‘shielding constant’ and depends on the density and distribution of the electron cloud that surrounds the nucleus.  $\sigma$  can take values ranging from  $10^{-6}$  for the lighter nuclides to  $10^{-3}$  for the heavier nuclides.  $B_{eff}$  is given by:

$$B_{eff} = B_0 - B_{loc} = B_0 - B_0\sigma = B_0(1 - \sigma) \quad (3.7)$$

As the resonance frequency of a nucleus depends on the (effective) magnetic field, this results in the differences in the resonance frequency of the nuclei in different molecules relative to

the Larmor frequency of the nuclide, based on the differences in their chemical environment. Taking this into account, Eq. (3.4) can be re-written as:

$$\nu = \left(\frac{\gamma}{2\pi}\right)(1 - \sigma)B_0 \quad (3.8)$$

This phenomenon is known as ‘chemical shift’ and is caused by shielding (screening) of the nuclei from the external magnetic field by the electron cloud that surrounds them. If chemical shifts are expressed in units of Hertz, this will make them dependent on the  $B_0$ . To avoid this dependency, chemical shifts are expressed in terms of (parts per million) ppm, which is independent of the field strength. By convention, chemical shift,  $\delta$ , is defined as:

$$\delta = \frac{\nu - \nu_{ref}}{\nu_{ref}} \times 10^6 \quad (3.9)$$

$\nu$  is the frequency of the compound under investigation and  $\nu_{ref}$  is that of a reference compound. The reference compound should ideally be chemically inert, should have a well-resolved singlet (a unimodal signal with no splitting), and its chemical shift should be independent of external variables (such as temperature and ionic strength). Tetramethylsilane (TMS) is widely used as a reference compound for  $^1\text{H}$  and  $^{13}\text{C}$  NMR studies in organic solvents. For aqueous solutions, 3-(trimethylsilyl) propionate (TSP) or 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) can be used as the reference compound (Friebolin 2005, De Graaf 2008). If the density of the electron cloud that surrounds the nuclei is high, the nuclei will be highly shielded and will resonate at lower frequencies or higher magnetic fields. Whereas nuclei that are surrounded by electron clouds with lower density, will be shielded less and will resonate at higher frequencies or lower magnetic fields. Shielded nuclei give rise to smaller chemical shifts relative to the internal reference than de-shielded nuclei, which appear at larger chemical shifts.

Functional groups can change chemical shifts of nuclei. If an electronegative atom or group is present in the vicinity of a nucleus, it will withdraw electrons and disturbs the electron density around the nucleus. This will cause a strong de-shielding effect and the chemical shift of the nucleus is shifted towards larger chemical shifts (higher frequencies) relative to the reference. The opposite effect can be observed with electron-donating substituent. The presence of such groups in the vicinity of a nucleus will increase the density of the electron cloud around it and the chemical shift will shift towards smaller chemical shifts (lower frequencies). In aromatic compounds, the circulation of  $p$ -electrons below and above the aromatic ring, which is referred to as the ‘ring current’, generates an anisotropic local magnetic field that has the same direction as the external field. As a result, the ring protons experience a pronounced de-shielding effect and feel a higher magnetic field, which is associated with larger chemical shift values.

Chemical shifts provide valuable information about the chemical environment of the nuclei and are a great aid in structural elucidation of compounds. However, in addition to the chemical shifts, there is an additional feature in NMR spectroscopy that can help in assigning the signals and gaining chemical information from NMR spectra. This feature is called ‘scalar coupling’, ‘J coupling’ or ‘spin-spin coupling’ and is observed as the splitting of resonances into several smaller lines. The main cause for scalar coupling is the fact that the magnetic moments of the nuclei in a molecule can influence each other, both directly through space (dipolar coupling) and also through chemical bonds (scalar coupling). Dipolar interactions average to zero because of the rapid molecular tumbling; however this is not the case for the interactions that take place through the chemical bonds which results in the scalar couplings (Friebolin 2005).

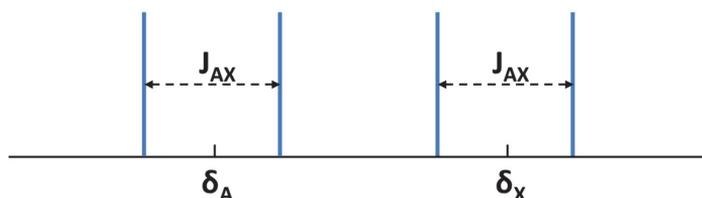
The quantum mechanical discussion related to the coupling will be avoided in this dissertation, but it can be briefly mentioned that the couplings are associated with transitions between energy levels or spin states. Multiplicity, or the number of the lines that a resonance is split into, can be calculated as:

$$M = 2nI + 1 \quad (3.10)$$

$M$  is the multiplicity,  $n$  is the number of equivalent neighboring nuclei that are coupled with the nucleus of interest, and  $I$  is the nuclear spin that was discussed earlier in this chapter. For  $I=1/2$  nuclei, Eq. (3.10) can be simplified as:

$$M = n + 1 \quad (3.11)$$

For the simplest case, a coupled two-spin system,  $AX$ , the resonance of  $A$  and  $X$  will be split into doublets split by the coupling constant,  $J_{AX}$  (Hz), centered at the Larmor frequency of  $A$  and  $X$ , as shown in Figure 3.3. Coupling constants are expressed in units of Hertz, and are independent of the magnetic field strength.

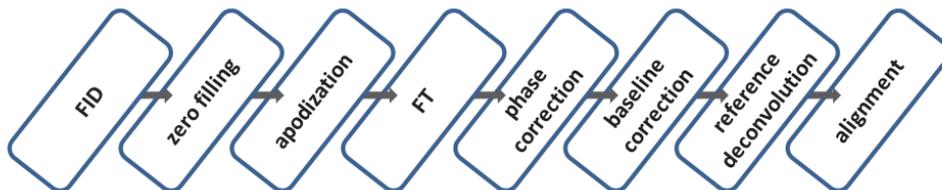


**Figure 3.3.** Simulated NMR spectra of a simple  $AX$  coupling system.

The coupling can range from one- to five-bonds. One-bond and two-bond couplings that are known as ‘germinal coupling’, and also three-bond couplings that are known as ‘vicinal coupling’ can provide important structural information. The three-bond coupling is related to the dihedral angle that is formed by the three bonds (e.g. for  $^1\text{H}$  and  $^{13}\text{C}$  nuclei in  $\text{H}-\text{C}-\text{C}-\text{H}$ ) and is the most informative one. Couplings can be heteronuclear and homonuclear, depending on the identity of the nuclei taking part in the coupling (Friebolin 2005, De Graaf 2008, Spyros, Dais et al. 2013).

## 3.2. NMR data processing

It is of high significance to optimize different acquisition parameters, such as pulse width, transmitter frequency, and receiver gain, in such a way that good quality data is ensured. Besides, appropriate processing of the acquired data can enhance the quality of the data even further, with specific emphasis on sensitivity and resolution enhancement. Data processing is performed before and after the Fourier transform of the FID. In the following subsections, different NMR data processing techniques are introduced and briefly discussed. Figure 3.4 shows a flowchart of the NMR data processing techniques that were used and are discussed briefly in the following.



**Figure 3.4.** A flowchart of the NMR data processing techniques.

### 3.2.1. Zero-filling

Zero-filling is a term that is used to describe the procedure of increasing the digital resolution of the NMR spectrum by increasing the FID data points just before performing Fourier transform, by *adding zeros to the end of the FID*. Even if zero-filling may not enhance the resolution of a spectrum, it will improve the lineshapes and the appearance of the spectrum. This can be very helpful in resolving very small couplings in multiple structures. Usually zero-filling is used to double the data points.

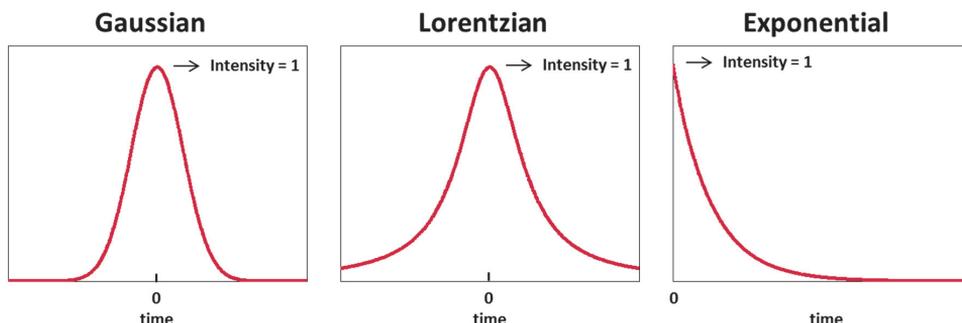
One should make sure that the FID has decayed to zero by the end of the acquisition time, because if not, baseline artifacts can occur after FT in the form of *sinc* wiggles. If the FID does not naturally fall to zero at the end of the acquisition time, window functions can be

employed to force the FID to zero, before performing zero-filling. It should be noted that after all, zero-filling cannot be a substitute for correct digitization of the recorded FID and optimizing acquisition parameters to get good quality data is of high significance (Claridge 2008, De Graaf 2008, Spyros, Dais et al. 2013).

### 3.2.2. Apodization

After collecting the data points for a reasonable number of scans and averaging the signals mathematically, the resulting time domain signal (FID) can benefit from a digital filtering type of manipulation, called ‘apodization’. Apodization is simply the practice of multiplying (convoluting) the FID with different window (weighting) functions which can be chosen to either enhance the sensitivity or resolution (or both if possible) in the final resulting spectrum. There are a variety of window functions that can be used on NMR data. Some of the most commonly used functions are demonstrated in Figure 3.5.

Exponential window functions have a decaying form:  $\exp(-t/lb)$ , where  $lb$  is the line broadening parameter in units of Hz and is chosen by the operator, and  $t$  is time. As the initial part of the FID contains most of the signal and the tail mainly the noise, multiplying the FID with an exponential window function in which  $lb > 0$ , will enhance signal-to-noise ratio (sensitivity) by digital filtration of the noise. But at the same time, truncating the FID tail will result in truncation artifacts or so called sinc wiggles. On the other hand, employing an exponential window function with  $lb < 0$ , will yield enhancement of resolution at the expense of sensitivity decrease, as it will increase both the apparent decay rate and the noise amplitude at the tail of the FID.



**Figure 3.5.** Some of the most commonly used apodization functions in NMR data processing.  $x$ -axis is time, and  $y$ -axis is intensity.

Lorentzian and Gaussian window functions (or a mixture of the two functions) are also commonly used for apodization. Equation 3.12 shows the Lorentz-Gaussian transformation, that converts a Lorentzian lineshape to a Gaussian lineshape (De Graaf 2008):

$$f_{filter}(t) = e^{-t/T_L} \times e^{-t^2/T_G^2} \quad (3.12)$$

The first part is the inverse Lorentzian function and the second part is the Gaussian function.  $f_{filter}$  is the resulted window function that will be multiplied by the original FID.  $T_L$  and  $T_G$  are the linewidth of the Lorentzian and Gaussian functions respectively. Compared to the Lorentzian function, the Gaussian function has much faster decay from the peak and narrower wings. Although the true theoretical lineshape of NMR peaks is typically Lorentzian (Ebrahimi, Nilsson et al. 2014), it can often be advantageous to use a Gaussian shape as this has a narrower base and can improve resolution, but at a moderate cost in sensitivity. A narrower linewidth for Lorentzian or Gaussian window functions will enhance resolution, whereas a wider linewidth will enhance sensitivity. If a too narrow linewidth is chosen, the sensitivity will decrease dramatically, especially for the Lorentzian function, and if the linewidth is too wide, the resolution will decrease significantly. So, choosing between the Lorentzian and Gaussian window functions, and choosing the linewidth, is in fact a trade-off between resolution and sensitivity enhancement, and naturally depends on the objective of the analysis. Comparing the results of apodization with different window functions and linewidths is often worthwhile and can help in finding the optimum condition for the data being investigated (De Graaf 2008, Spyros, Dais et al. 2013, Ebrahimi, Nilsson et al. 2014).

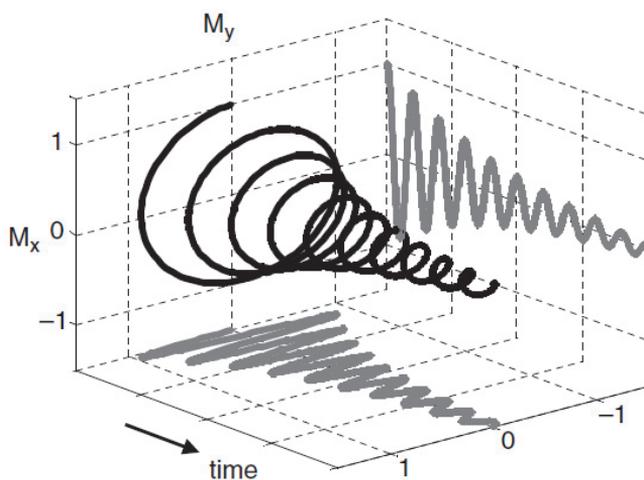
### 3.2.3. Fourier transform: *the bridge between the time and frequency domains*

After excitation of a nuclear spin in a magnetic field, by the pulses emitted from the transmitter in an NMR spectrometer, the system will start to re-establish the equilibrium condition and lose the excess energy that is imparted into the system by the applied pulse. This is done by the relaxation of the excited spins (with the rates that are specific to different nuclei). During this process the transverse magnetization (in the rotating frame) will be detected by the *rf* coil of the detection channel. This signal is called the ‘free induction decay (FID)’. The FID is a complex signal that can be decomposed into real (absorption mode) and imaginary (dispersion mode) components by projections, as shown in Figure 3.6.

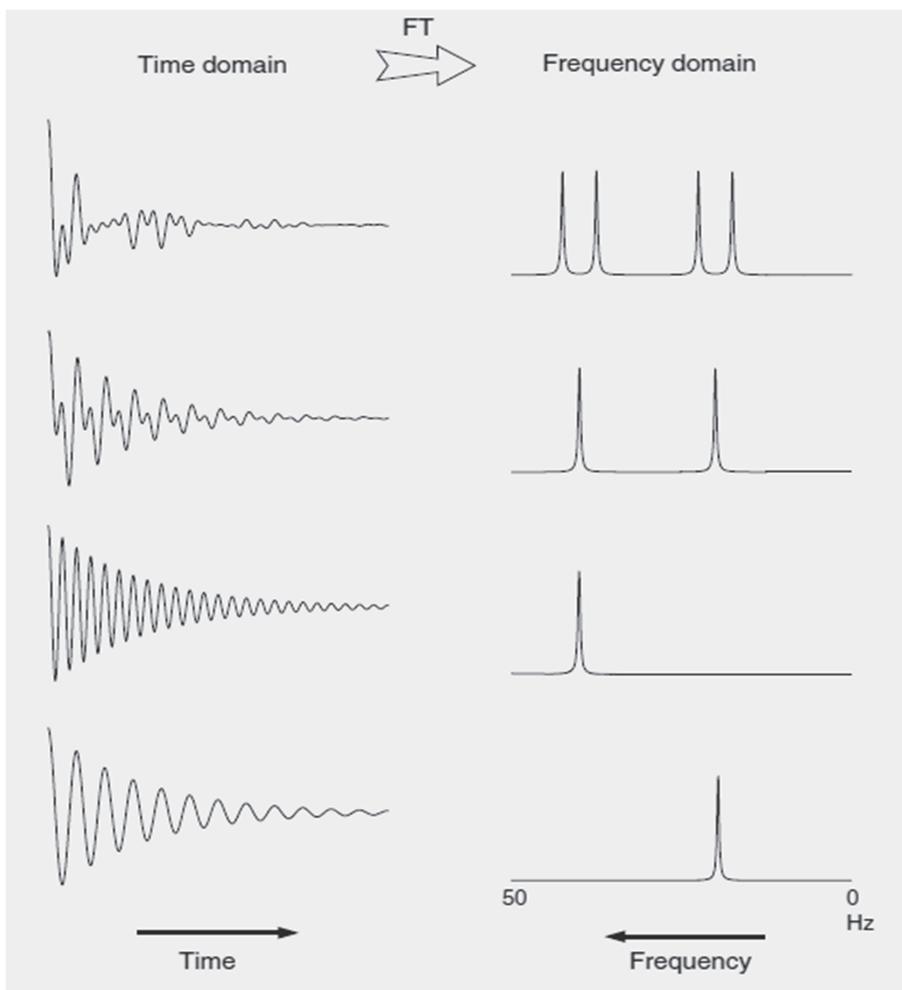
The FID is a time-domain representation of the nuclear precession frequencies within the sample and its data content is not easy to use and not that informative as it is. To make the FID interpretable, one should transform the time-domain FID into the frequency-domain NMR spectrum. This transformation is performed by a mathematical procedure called ‘Fourier transform (FT)’ and has the general form of:

$$f(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-i\omega t} dt \quad (3.13)$$

$\omega$  is equal to  $2\pi\nu$ ,  $f(\omega)$  represents the frequency domain and is a complex function, whereas  $f(t)$  represents the time-domain. It is usually only the real part of the spectrum that is displayed, however the real and imaginary parts both can represent the spectrum and it is only their phase that differs by  $90^\circ$ . FT and its applications are not limited to NMR spectroscopy, and FT can transform any function from the time-domain to the frequency-domain. In the early days of FT NMR, it was often the FT step that was the rate-limiting stage in producing a spectrum, although with today's computers and the use of the Cooley–Tukey algorithm (Cooley and Tukey 1965) for fast Fourier transform (FFT), the time requirements are not of much consequence anymore. Figure 3.7 demonstrates a number of simple FIDs and their corresponding spectra (Claridge 2008, De Graaf 2008, Spyros, Dais et al. 2013).



**Figure 3.6.** The free induction decay (FID) of nuclear magnetization precessing at the Larmor frequency of the corresponding nucleus, after an excitation pulse.  $M_x$  and  $M_y$  represent the real and imaginary components of the complex FID, respectively. *Reprinted from (De Graaf 2008).*



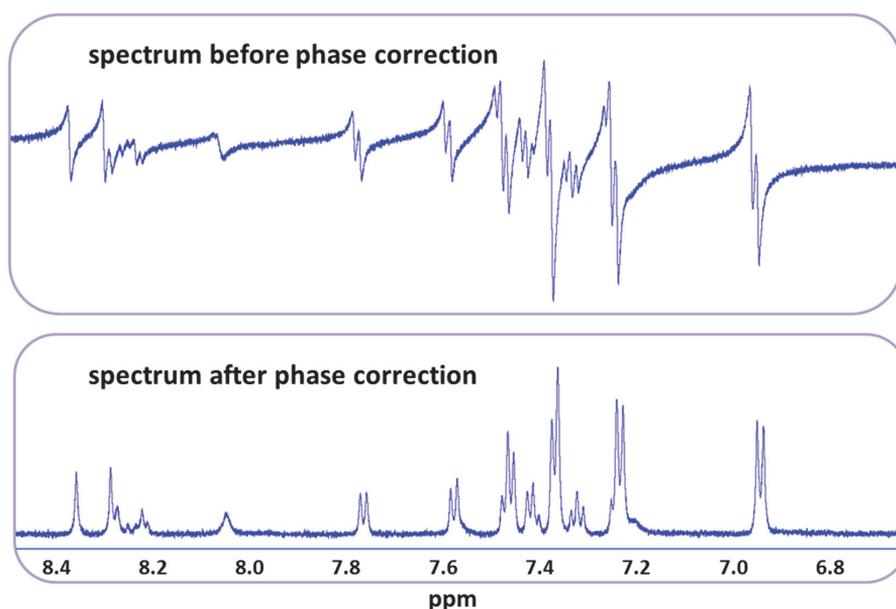
**Figure 3.7.** Examples of simple FIDs that are transformed into their frequency-domain NMR spectrum by FT. Adapted from (Claridge 2008).

### 3.2.4. Phase correction

Phase correction is the process of correcting phase errors in the spectra. Phase errors can make the signals to have a dispersive line shape or be inverted. Figure 3.8 shows an example of a spectrum with phase errors and also a spectrum that is properly phased.

There are two main reasons for the phase errors: 1) the delays in the pulse sequences, and 2) off-resonance effects. In the NMR pulse sequences, short time delays (tens of microseconds) are introduced between the closure of the transmitter after the implementation of each pulse and the opening of the receiver to acquire the FID. These delays are necessary to protect the receiver electronics from the harmful effect of the pulse. Because of these delays, the acquired

spectra will be a mixture of absorptive and dispersive signals, and the peaks will have some portion that is displaced below the baseline, not looking like their ideal Lorentzian shape. The second reason for the signals to be out of phase is off-resonance effects that originate from pulse imperfections. If the pulse power is not sufficient to excite all the nuclei in the sample equally, for the nuclei that resonate outside the excitation frequency bandwidth of the pulse, the magnetization will not flip exactly on the  $xy$ -plane after the  $90^\circ$  pulse. This will result in a phase difference between these nuclei that are namely off-resonance, and the nuclei that are on-resonance and are properly excited with the pulse. As proton nuclei have relatively narrow frequency dispersion, phase errors of this kind are usually small in  $^1\text{H}$  NMR, and being almost a linear function of frequency, are easy to correct.



**Figure 3.8.** An example of an NMR spectrum before and after phase correction.

Phasing the spectrum is a routine procedure and involves zero-order and first-order phase corrections. Zero-order correction is frequency independent and can correct phase errors that affect all the peaks equally. Whereas, the first-order correction is frequency dependent, and is necessary to correct more complicated phase errors (Claridge 2008, Spyros, Dais et al. 2013).

### 3.2.5. Baseline correction

To have accurate quantification by NMR, it is important to have a horizontal baseline. Especially in metabolomics studies where small peaks can be very important, baseline

distortions can hinder useful information. In case the first data points in FID are corrupted, because of high signal amplification or if the receiver electronics have not fully recovered from the influence of the RF pulse, this can result in a distorted baseline in the NMR spectrum. A truncated FID can also cause a distorted baseline. These can be avoided by proper adjustment of signal amplification and using a proper acquisition time to ensure full decay of the FID. Baseline correction methods try to model the baseline by fitting it to polynomial, sine or exponential functions, with polynomial function being the most common one. After modelling the baseline, it is simply subtracted from the spectrum to yield the baseline corrected spectrum (Spyros, Dais et al. 2013).

### **3.3. Enhancing quality of NMR spectra**

In NMR spectra, peaks position and area contain useful qualitative and quantitative information. Factors like temperature fluctuations and changes in the pH of the sample can lead to the peaks shape changes or shift in the position of a chemical compound between a series of similar samples. From the perspective of multivariate analysis, if position or shape of the peaks of a specific compound change between the measurements, multivariate models become unnecessarily complex and interpreting them becomes difficult. Successful use of the multivariate models will only be possible if there is good reproducibility in the data and variations in peaks shape and position between measurements do not happen or are corrected. When peaks shift or shape changes happen, data will not be bilinear which implies that methods, such as Principal Component Analysis (PCA) (Wold, Esbensen et al. 1987), and partial least squares regression (PLS) (Wold, Martens et al. 1983) will not be efficient. Therefore, it is important to correct peaks shift and shape changes prior to the multivariate analysis. Reference deconvolution (see below) and the *icoshift* alignment program (Savorani, Tomasi et al. 2010) are two of the methods that can be very useful for correcting peaks shape changes and position shift, respectively. The theory of these methods and how they can benefit data analysis of NMR data are presented in the following.

#### **3.3.1. Reference deconvolution (Paper I)**

Most of the instrumental imperfections, such as magnetic field inhomogeneity, pulse phase and amplitude errors, and field instability, affect all the signals in an NMR spectrum equally and are independent of frequency of the resonances that they affect (Morris 1988, Barjat, Morris et al. 1995). Reference deconvolution is a data processing method that is highly effective at correcting systematic and frequency independent errors in NMR data. It extracts the signal of a known reference signal from the experimental data, compares it to the theoretically expected form, and constructs the correction function needed to convert the full experimental dataset into the form that it would have had if the unwanted perturbations

experienced by the reference signal had not been present. Reference deconvolution is fast, linear (to a good approximation—the noise structure is changed slightly because the experimental noise in the reference region is convoluted onto the full spectrum), and robust. It should be noted that reference deconvolution cannot correct instrumental imperfections that are frequency dependent. Examples include errors in the spectra that originate from temperature changes and gradients, as well pH variations (Barjat, Morris et al. 1995). Reference deconvolution has been known to NMR spectroscopists for many years, but it appears to have been neglected by the NMR-based metabolomics/chemometrics community.

A number of different algorithms have been proposed for reference deconvolution (Wouters and Petersson 1977, Wouters, Petersson et al. 1977, De Graaf, Van Dijk et al. 1990), but they all share the same basis. One of the effective and simple algorithms that are used for reference deconvolution is FIDDLE (Free Induction Decay Deconvolution for Lineshape Enhancement) algorithm. The theoretical basis of FIDDLE has been discussed extensively in the literature (Morris 1988, Morris, Barjat et al. 1997, Metz, Lam et al. 2000), but a graphical illustration of the key elements is shown in Figure 3.9 and explained in the following. The NMR time-domain data, the free induction decay or FID (Figure 3.9. (a)), are zero-filled (to retain all the spectral information), Fourier-transformed (FT), and phase-corrected to yield the raw NMR spectrum (Figure 3.9. (b)). A suitable reference signal in the spectrum is then chosen, and the rest of the spectrum is set to zero. The real part (the absorption mode) of this filtered spectrum is subjected to inverse Fourier transformation to give a complex FID that contains only the reference signal (Figure 3.9. (c)). Choosing to retain only the real part of the reference spectrum excludes dispersion mode signals, making clean extraction of the reference signal much easier; no information is lost if the initial FID was zero-filled. In parallel, a synthetic FID (Figure 3.9. (e)) is calculated for the reference signal, using the known frequency (or frequencies; in the case of a reference such as TSP,  $^{29}\text{Si}$  and  $^{13}\text{C}$  satellite signals are included) and a specified lineshape. The latter is chosen by the user, according to need; while the true theoretical lineshape is typically Lorentzian, it can often be advantageous to use a Gaussian shape, because of its narrower base. This choice of target lineshape is analogous to the choice of window function (apodization) in normal FT processing, and the same considerations for resolution or sensitivity enhancement apply, as discussed in Section 3.2.1. The most conservative choice is a Lorentzian lineshape of approximately the same width as the experimental reference signal (Figure 3.9. (d)); this regularizes the lineshape (and phase and frequency) with minimum change in resolution and S/N. A complex correction function (containing both real and imaginary parts) is then constructed by dividing the ideal reference FID (Figure 3.9. (e)) by the experimental reference FID (Figure 3.9. (c)). The cumulative effect of instrumental imperfections such as field inhomogeneity, pulse phase error, and modulation sidebands is to multiply the FID that would have been recorded if the instrument had behaved ideally by a complex time-domain error function. The correction function calculated here is the inverse of that function, so when the original (full) experimental FID

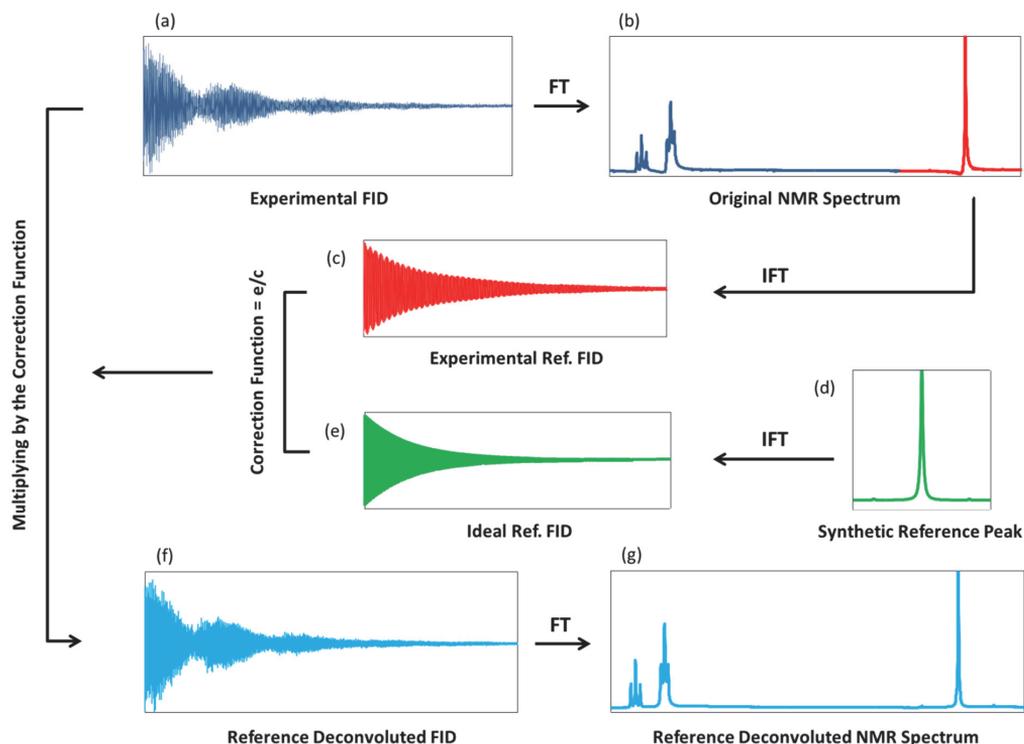
(Figure 3.9. (a)) is multiplied by the correction function, the result is a corrected FID (Figure 3.9. (f)) in which all the multiplicative errors seen in the reference FID have been corrected. The corrected FID can then be Fourier transformed to yield the reference-deconvoluted spectrum (Figure 3.9. (g)), in which such imperfections as lineshape distortions, signal amplitude errors, and signal phase changes have been corrected (Morris, Barjat et al. 1997, Metz, Lam et al. 2000, Ebrahimi, Nilsson et al. 2014)

For best results, the reference peak should be a well-resolved singlet, which is present with high signal-to-noise ratio (S/N) in all the spectra being deconvoluted. Typical examples of suitable signals are those from 3-(trimethylsilyl) propionic acid (TSP-d4), tetramethylsilane (TMS), and sodium 2,2- dimethyl-2- silapentane-5-sulfonate (DSS-d6), compounds that are commonly added to NMR samples to provide an internal standard for quantification and calibration of the chemical shift axis. The noise in the vicinity of the reference signal will be convoluted onto the entire spectrum, so if the S/N of the reference signal is too low, it can significantly degrade the quality of the data.

Multiplets are a much poorer choice for reference signals as they have FIDs that have zero amplitude at regular intervals, which results in singularity problems that are mathematically challenging. The zeroes make interpolation necessary, introducing an element of non-linearity into the algorithm. While the use of a doublet as the reference signal has been reported (Barjat, Morris et al. 1995), most software for reference deconvolution does not cater for multiplet reference signals.

The choice of the ideal peak lineshape and linewidth (the “target lineshape”) is important and warrants further discussion. The lineshape chosen for the ideal reference signal is typically Lorentzian or Gaussian or a mixture of the two, although there are many other possibilities. As noted earlier, there is a close analogy between the choice of target lineshape and the apodization procedure used in conventional Fourier transform processing. As most reference signals have a Lorentzian natural shape, and the effects of static field inhomogeneity also often approximate to a Lorentzian distribution of signal amplitude as a function of frequency, the choice of a Lorentzian target lineshape with a width close to that of the experimental reference line will produce a spectrum similar in appearance to the original but with errors in lineshape, phase, frequency, and so on corrected. However, it is often useful to change the target lineshape to aid the extraction of the features of interest from the data under analysis. If a Lorentzian target lineshape narrower than the experimental reference line is chosen, resolution will be increased, but at a severe cost in S/N; if too narrow a lineshape is used, numerical instabilities in the correction will cause severe spectral distortions. Choosing a lineshape wider than the experimental reference line will increase the S/N at a cost in resolution, with a maximum S/N improvement at twice the experimental linewidth (so-called matched filtration). The choice of a Gaussian or mixed lineshape is often a good alternative,

as the narrow base of a Gaussian improves resolution, but at a moderate cost in sensitivity. The optimum target lineshape naturally depends on the objective of the analysis, and comparison between spectra corrected with different target lineshapes is often worthwhile (Morris 1988, Ebrahimi, Nilsson et al. 2014).



**Figure 3.9.** Schematic illustration of the FIDDLE algorithm for reference deconvolution. The reference peak is extracted from the experimental spectrum (b), and its inverse Fourier transform (c) is compared to that of “perfect” FID (e) to yield a correction function (e/c). The correction is then applied in the time domain to the entire experimental FID (a) to produce the corrected FID (f).

In Paper I (Ebrahimi, Nilsson et al. 2014), the effect of reference deconvolution on the multivariate analysis results of NMR data for a triangular experimental design was investigated. The results verified that reference deconvolution can enhance the PCA and PLS models of NMR data. This can be very helpful in metabolomics studies by NMR data. In Paper II (Ebrahimi, Larsen et al. 2015), reference deconvolution was used to correct line broadening of NMR signals in *in vitro* measurements of bacterial fermentation. The results verified the great potential of reference deconvolution for improving the multivariate analysis results of *in vitro* NMR studies.

### 3.3.2. *icoshift* alignment program

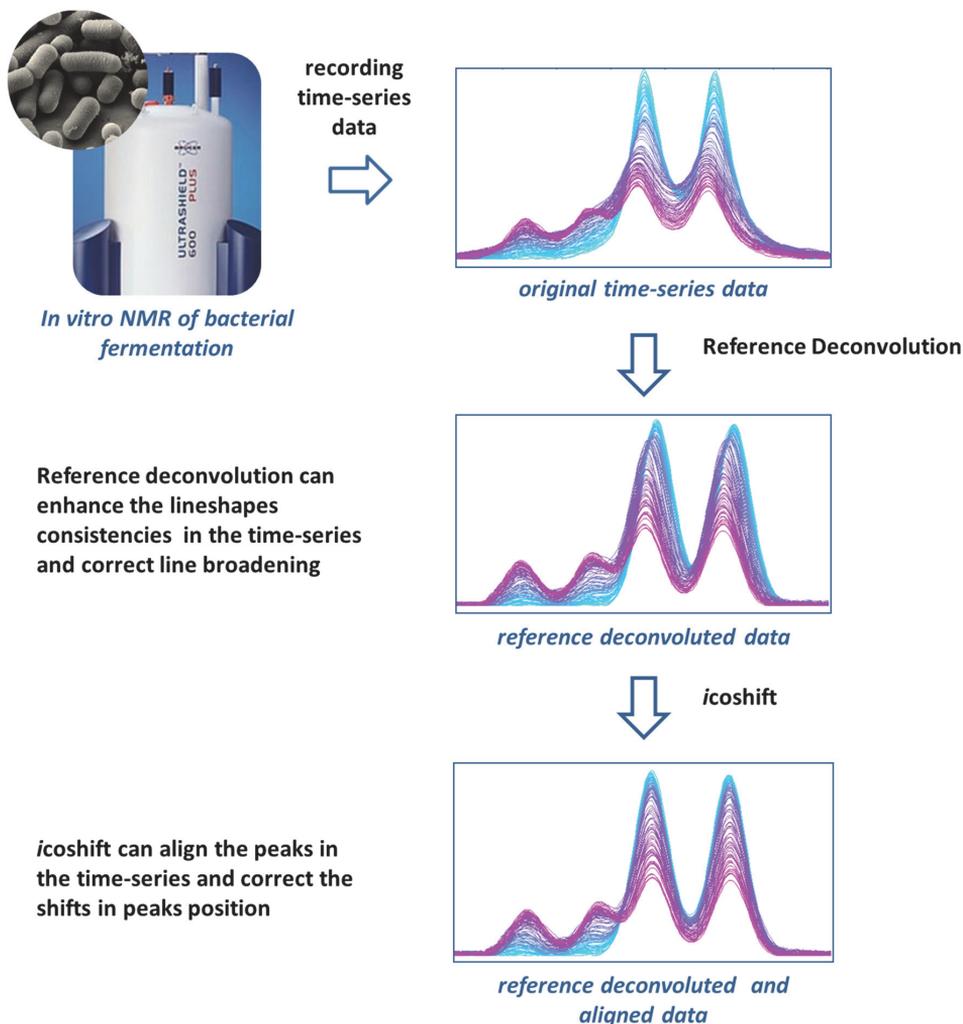
The frequencies of NMR resonances are very sensitive to the chemical environment of the nucleus, and it is in fact this property that enables structural elucidation and identification of different compounds within a sample. However, this also means that the frequencies of the resonances can be affected by small fluctuations in the temperature, pH, salt concentration, and external field during the acquisition, or between the measurements of a set of samples. This will be a source of unwanted variations in studies like metabolomics where often at least several samples are measured and the data is analyzed by multivariate analysis techniques (Pearce, Athersuch et al. 2008). From the perspective of multivariate analysis, if the resonance frequencies of a specific compound change in different measurements, multivariate models become unnecessarily complicated, and the bilinearity/ trilinearity assumptions of the different multivariate analysis techniques such as PCA, PLS, and PARAFAC, will not be fulfilled. Successful use of the multivariate models will only be possible if there is good reproducibility in the data, and if the variations in peaks position between measurements are eliminated.

Several solutions have been suggested for the alignment of 1D NMR data, from the pragmatic bucketing procedure (Spraul, Neidig et al. 1994) to more advanced procedures such as Recursive Segment-wise Peak Alignment (RSPA) (Veselkov, Lindon et al. 2009), and interval Correlation Optimized Shifting (*icoshift*) (Savorani, Tomasi et al. 2010). The *icoshift* program is a very efficient alignment program for 1D NMR data. The algorithm aligns signals in defined spectral intervals between a series of spectra, by using the maximum correlation as the criteria. Owing to the use of fast Fourier transform algorithm, *icoshift* is very fast and can align all the spectra simultaneously. The MATLAB code for *icoshift* is freely available from [www.models.life.ku.dk](http://www.models.life.ku.dk) (Savorani, Tomasi et al. 2010). In Paper II (Ebrahimi, Larsen et al. 2015), *icoshift* was successfully used for aligning the signals between the samples with different pH values.

### 3.3.3. Using reference deconvolution and *icoshift* in *in vivo*/ *in vitro* NMR studies

*In vivo/in vitro* NMR studies suffer from the inhomogeneity of the samples, which can result in the broadening of the signals in the recorded spectra. Another problem associated with investigating live organisms with NMR is that the pH of the samples can change as the result of the cellular metabolism. The pH change can subsequently cause shift in the position of NMR signals from pH sensitive compounds, and this is not desirable in the multivariate analysis of the data. Therefore, for reliable analysis of *in vivo* NMR data, it will be a great advantage to correct peaks lineshape and position inconsistencies, if any. For the *in vitro* NMR measurements of the fermentation of LAB, that is presented in Paper II (Ebrahimi,

Larsen et al. 2015) and Chapter 5, number of the cells and inhomogeneity of the samples increase during the time course of fermentation, and this leads to the line broadening of the signals. Besides, the production of acids decreases the pH and causes shift in the position of some of the signals. Fig. 3.10 shows how reference deconvolution and *icoshift* can enhance the quality of time-series *in vitro* NMR data. It is recommended that both reference deconvolution and alignment by *icoshift* are used as routine processing techniques in *in vitro* NMR studies.



**Figure 3.10.** Enhancing the quality of real time *in vitro* measurements of bacterial fermentation by reference deconvolution and *icoshift*.

### 3.4. Sensitivity in NMR spectroscopy

Despite many advantages that NMR spectroscopy offers, such as easy sample preparation, its non-invasive nature, and the wealth of the information that it can provide, its main disadvantage compared to other analytical techniques such as mass spectroscopy is its low sensitivity. Because of the small energy difference between the nuclear spins, their population differences are not high. This is the key factor that makes NMR insensitive compared to other analytical techniques such as ultraviolet (UV) and infra-red (IR) spectroscopy, for which the energy difference between the ground and excited states are much larger.

Sensitivity can define the lowest amount of a compound that can be reliably detected by an analytical technique, and therefore represents limit of detection (LOD) for a measurement. Signal-to-noise ratio can be a measure of sensitivity in a quantitative manner. Signal-to-noise ratio should be defined relative to a compound or a signal as the reference, and is calculated as:

$$S/N = \frac{P}{2N_{rms}} = \frac{2.5P}{N_{pp}} \quad (3.14)$$

where  $P$  is the reference peak intensity, and is only reliable if the lineshape is ideal and the peak does not overlap with any other signals.  $N_{rms}$  and  $N_{pp}$  are both noise levels, but estimated differently.  $N_{rms}$  is the noise calculated as the root-mean-square of the noise in a defined region of the spectrum where there are no resonances. A bandwidth of 200 Hz is commonly used.  $N_{pp}$  is the peak-to-peak estimation of the noise and is calculated by subtracting the minimum value in the noise region from the maximum value.  $N_{rms}$  is approximately one fifth of  $N_{pp}$ , and is more robust. Once comparing signal-to-noise ratios calculated for different spectra, one should know which method for noise estimation has been used, as they will give different values (Claridge 2008).

### 3.5. Quantitative NMR spectroscopy

NMR can provide a wealth of structural (qualitative) and quantitative chemical information. While chemical shifts and coupling constants give the structural information, signal intensities can be used for quantitative analysis. The quantitative nature of NMR is due to the fact that signal intensity in the spectrum is directly proportional to the number of the corresponding nuclei in the sample that give rise to that resonance. Signal intensity can be calculated by integrating the area under the peaks. It is mainly one-dimensional NMR spectra that are used in quantification studies. Proper optimization of acquisition parameters, such as the acquisition time, the relaxation delay, the pulse width, and the pulse power, is indeed very consequential to ensure a reliable quantitative measurement. Moreover, prior to the quantitative analysis, the data should be processed by the data processing techniques that were

discussed previously in this chapter (e.g. phase correction and baseline correction). Otherwise, the data will not provide accurate quantitative information. Multivariate analysis techniques enable the analysis and information recovery from complex NMR spectral data. NMR, as a quantitative analytical tool, offers a number of advantages over some of the other commonly used analytical techniques, such as chromatography and mass spectroscopy. NMR requires no or little sample pre-treatment, no prior separation of the analytes from a complex mixture is necessary, it is non-invasive and non-destructive and allows time-series studies of the sample, and finally it is not selective for any special groups of chemical compounds and makes simultaneous analysis of multiple group of compounds possible (Spyros, Dais et al. 2013).

### 3.6. Metabolomics

In 1998, the first definition of the term ‘metabolome’ was presented as the set of low-molecular-mass compounds synthesized by an organism (metabolites) (Oliver 1998). Shortly after, in 1999, the term ‘metabonomics’ was presented by Nicholson et al., and defined as the analysis of changes in the metabolic status of an organism, as a consequence of drug treatment, environmental influences, nutrition, lifestyle, genetic effects, toxic exposure, diseases, etc (Nicholson, Lindon et al. 1999). A few years later, in 2002, the first detailed definition of the term ‘metabolomics’ was presented by Fiehn, as the qualitative and quantitative study of the metabolome in a biological system (Fiehn 2002). The biological system can be cell, tissue, organ, or organism. Nowadays, the terms *metabolomics* and *metabonomics* are often used interchangeably, while the word metabolomics is accepted and used more commonly (Metabolomics Society 2010, Savorani, Rasmussen et al. 2013). However, Nicholson and coworkers have distinguished between the two terms. He defines metabolomics as “the measurement of metabolite concentrations and fluxes and secretion in cells and tissues in which there is a direct connection between the genetic activity, protein activity and the metabolic activity itself”, whereas metabonomics as “the quantitative measurement of the multivariate metabolic responses of multicellular systems to pathophysiological stimuli or genetic modification” (Nicholson, Lindon et al. 1999, Nicholson and Wilson 2003). Therefore, it can be useful to search for both terms while conducting a full literature search (Emwas, Salek et al. 2013). This is presented merely as a point to be aware of, however, throughout this thesis, only the term *metabolomics* will be used.

Similar to other omics fields, metabolomics is a dynamic research field in the effort to understand biological systems, and can provide a view over the physiological state of living systems. Although in comparison with other –omics fields like genomics, transcriptomics, and proteomics, the term metabolomics was established much later, studies that focused on the metabolome were already performed (Oldiges, Lütz et al. 2007). Metabolomics is expected to advance the fields of functional genomics, systems biology, and metabolic engineering.

Metabolomics is multifunctional and based on the topic and the goal of the study, different analytical approaches can be used (Nielsen 2001, Kell 2004, Kell, Brown et al. 2005).

Metabolomic studies can be divided into targeted analysis, metabolic profiling, metabolic fingerprinting, metabolic footprinting, and metabolomics. The definition of these terms are presented in Table 3.1 (Fiehn 2002, Kell, Brown et al. 2005). Targeted analysis may require extensive sample clean-up and selective sample preparation to avoid interference from other metabolites. In targeted analysis, by the selective sample preparation, most of the metabolome information can be ignored, however data quality can be improved and precise quantification of metabolites can be achieved. Metabolic profiling is a promising method for studying microbial metabolism in a quantitative manner. It can be used for the *in vivo* kinetic studies of the underlying metabolic networks (Buchholz, Hurlebaus et al. 2002, Chassagnole, Noisommit-Rizzi et al. 2002, Wiechert 2002), and can clarify limiting biosynthetic steps, knowledge that can be used for iterative strain optimization (Oldiges, Kunze et al. 2004, Magnus, Hollwedel et al. 2006, Oldiges, Lütz et al. 2007). Metabolic fingerprinting allows rapid classification of samples and does not require extensive sample preparation or purification steps (Fiehn 2001).

**Table 3.1.** Definition of the terms that are used to subdivide metabolomic studies based on their strategy and focus in studying the metabolome.

<b>Terms</b>	<b>Definition</b>
<i>Target analysis</i>	quantitative analysis of target groups of known metabolites
<i>Metabolic profiling</i>	quantitative analysis of a group of pre-defined metabolites, like members of a particular pathway
<i>Metabolic fingerprinting</i>	rapid classification of samples by analyzing their intra-cellular metabolites (endometabolome) of biochemical relevance
<i>Metabolic footprinting</i>	analysis of the extra-cellular metabolites (exometabolome) in a high-throughput manner
<i>Metabolomics</i>	unbiased overview of the patterns of the entire cellular metabolism

The advantage of metabolic footprinting over metabolic fingerprinting is that as metabolic footprinting focuses on the exometabolome, time consuming and sometimes rather irreproducible quenching and extraction steps are not involved. The limitation imposed by metabolic footprinting is that the full pattern of the intra-cellular metabolites cannot be

captured, as phosphorylated compounds and other highly charged metabolites are unlikely to be present in the extracellular medium. Target analysis, metabolic profiling, and metabolomics are all quantitative approaches and require exact quantification of unambiguously identified metabolites. Metabolic fingerprinting and footprinting, on the other hand, are semi-quantitative approaches and even unknown metabolites peaks can be used in these approaches to gain insight. Semi-quantitative data can originate from peak areas or heights and is error-prone. However, it can be used to good effect to compare the metabolic content between biological samples (Oldiges, Lütz et al. 2007).

The size of the metabolome varies between the different biological systems. Microorganisms and prokaryotes that are considered to have the simplest biological systems have around several hundred metabolites, whereas, in plants, more than 200,000 metabolites or phytochemicals can exist. The cellular compartmentalization in eukaryotes makes their metabolome even larger and more sophisticated, as the metabolites will also differ based on their position in the organelles. This is opposed to the prokaryotes that have much simpler structure and are not compartmentalized (Mungur, Glass et al. 2005, Oldiges, Lütz et al. 2007).

Microbiology is one of the research fields that increasingly use metabolomics. Microbial metabolomics expands from new drug discovery efforts to metabolic engineering for studying genotype-phenotype correlations and improve strain selection. Most of the applications for this purpose are targeted approaches. Metabolomics has been able to identify metabolic bottlenecks in pathways. For glycolysis pathway for instance, the conversion of phosphoenolpyruvate (PEP) to pyruvate, which is catalyzed by pyruvate kinase (PK), has been identified as key regulatory points of glycolytic flux, at least under the circumstances of limiting glucose (Theobald, Mailinger et al. 1993, Neves, Pool et al. 2005). One of the challenges facing microbial metabolomics is effective instant quenching of the metabolism of the systems to study its metabolic state. In general, in prokaryotes, the intra-cellular metabolites tend to leak more in quenching with the common cold methanol protocol than in eukaryotes, which is attributed to the less robust cell wall and membrane structure in prokaryotes (Mashego, Rumbold et al. 2007).

Different analytical tools have been used in metabolomics, to quantify metabolites and to investigate the metabolic fingerprinting or profiling. Mass spectrometry (MS) and NMR spectroscopy are the most commonly used techniques. However, other techniques such as Fourier transform infrared (FT-IR), enzymatic assays, gas chromatography, and high-performance liquid chromatography (HPLC) have also been used in metabolomics. In principal, any analytical technique can be used for different parts of a metabolomics study, but there is a trend towards highly selective and sensitive methods, which require small

sample volume to avoid any necessity for sample dilution. Unlike proteins, DNA, and RNA, the low molecular weight of metabolites eliminates the chance of up-concentration by ultrafiltration techniques. Methods like freeze-drying can be used for up-concentrating the metabolites, but loss of volatile metabolites will prevent to capture the true metabolic view of the system. (Oldiges, Lütz et al. 2007, Emwas, Salek et al. 2013). MS has high sensitivity, and if coupled with chromatographic setups, provides a powerful analytical system for studying metabolites. By coupling to two-dimensional chromatographic techniques, even higher resolution can be achieved. However, MS normally requires prior separation steps by gas or liquid chromatography. NMR has lower sensitivity relative to MS; on the other hand, it does not necessarily need prior separation, and is highly reproducible and robust which can be very useful for a study with many samples (Oldiges, Lütz et al. 2007).

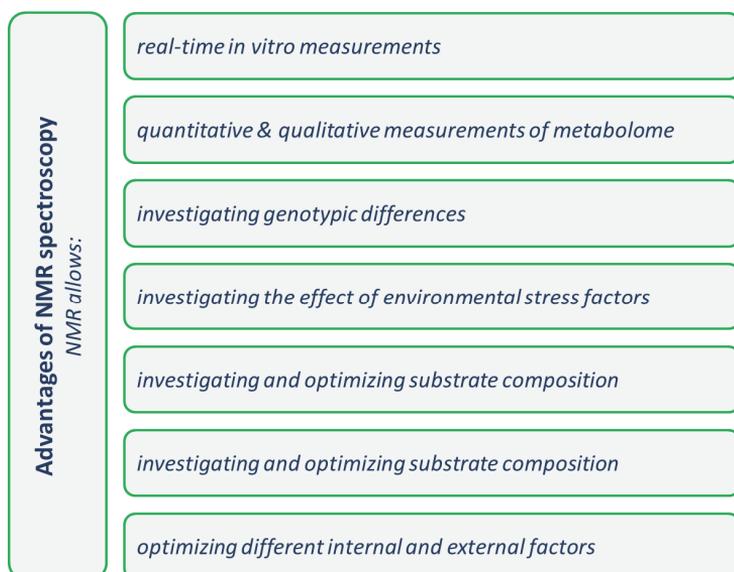
### 3.7. The role of NMR in metabolomics

NMR can offer many possibilities for metabolomics owing to its specific characteristics as an analytical technique, and can provide qualitative and quantitative information on chemical and biological samples. NMR is not selective for any special groups of chemical compounds and can be used for quantitative analysis of complex mixtures. This can be very desirable for non-targeted analysis of biological samples. Moreover, it requires no sample pretreatment, enjoys easy sample preparation, is non-destructive and leaves the sample for further analysis, and moreover is highly reproducible (Winning, Larsen et al. 2008, Emwas, Salek et al. 2013). The main limiting factor in using NMR in metabolomics is its low sensitivity. Nowadays, the significant advances in the hardware of spectrometers, like higher field strength, microprobes, cryogenically cooled probes, and also emerging NMR techniques such as the dynamic nuclear polarization (DNP), have made NMR a more efficient tool and have been able to improve sensitivity (Grivet, Delort et al. 2003, Emwas, Salek et al. 2013). Despite these, compared to some of the other analytical techniques, NMR still strives for sensitivity.

Despite the relative low sensitivity, NMR has yet a great advantage for studying live organisms, owing to its non-destructive nature, i.e. it can be used for *in vivo* and *in vitro* measurements of biological processes. *In vivo* and *in vitro* NMR measurements allow real-time investigation of biochemical processes, in a way that the metabolism is not disturbed by any means. This is a huge advantage for capturing a true image of the metabolism of the system, and observing changes that can be easily missed by any technique other than *in vivo* and *in vitro* screening. Based on these arguments, it can be concluded that the main limiting factor for using NMR in metabolomics is its sensitivity, and its unique strength which cannot be overlooked, is the feasibility of measuring live organisms and cells. This advantage puts NMR in a particular position in metabolomics.

NMR has played a central role in metabolomics. It has been used for studying metabolism since early 1970s when isotope labelling and  $^{13}\text{C}$  NMR were used to study ethanol

metabolism in rat (Wilson, Burlingame et al. 1974). The wealth of information that was obtained by the early studies, made NMR popular for studying metabolism. Since then,  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{31}\text{P}$  NMR, and other nuclides, have been used widely in metabolomics studies related to human metabolism (Govindaraju, Young et al. 2000, Rothman, Behar et al. 2003), nutrition (Gibney, Walsh et al. 2005, Wishart 2008, Cevallos-Cevallos, Reyes-De-Corcuera et al. 2009), disease diagnostics (Wang, Tso et al. 2010, Emwas, Salek et al. 2013), and biomarker discovery (Kim, Maruvada et al. 2008, Smolinska, Blanchet et al. 2012). Moreover, NMR has been used to study the metabolism of animals (Tikunov, Johnson et al. 2010), plants (Kim, Choi et al. 2010, Kim, Choi et al. 2011), and microorganisms (Grivet, Delort et al. 2003, Wishart 2008, Boroujerdi, Vizcaino et al. 2009, Sekiyama, Chikayama et al. 2011). Fig. 3.11 summarizes the advantages of NMR and the possibilities that it offers metabolomics.



**Figure 3.11.** The advantages of NMR spectroscopy and the possibilities that it offers metabolomics studies.

### 3.8. NMR for studying microbial metabolism

It was over 40 years ago, in 1972, that NMR was applied for the first time to investigate microbial metabolism. It was used to study the catabolism of  $^{13}\text{C}$ -labeled glucose by the yeast *Candida utilis* at a 1.4 T magnetic field. Although the sensitivity was quite low, using the labeled substrate allowed following the concentration profiles of several metabolites in the cells over time (Eakin, Morgan et al. 1972). Early papers in this field were actually feasibility studies rather than a source of biochemical knowledge (Grivet, Delort et al. 2003). Since then,

the advances in NMR instruments have made this technique more efficient in investigating microbial metabolism. In the following, some of the examples of the studies in the microbial metabolomics field that have benefited from the application of NMR are presented.

Sauer et al. used  $^1\text{H}$ - $^{13}\text{C}$  correlation spectroscopy (COSY) for investigating the central carbon metabolism in *E. coli*, by measuring the intracellular extract. The application of the 2D NMR technique in this study led to new knowledge about the tricarboxylic acid (TCA) cycle in *E. coli* (Sauer, Lasko et al. 1999). Carbon-13 NMR has been extensively used to study the diversity and extent of carbon cycling in the carbohydrate metabolism of microorganisms, and has provided new insight into the field (Portais and Delort 2002). For instance, by using labelled pyruvate and/or lactate and the labelling patterns on alanine and aspartate, the Wood–Werkman cycle in *Propionibacterium freudenreichii*, subsp. *Shermanii* was studied by  $^{13}\text{C}$  NMR analysis of the live cells, as well as the intracellular extracts (Deborde, Rolin et al. 1999, Deborde and Boyaval 2000). In another study,  $^{13}\text{C}$  NMR analysis of the cell-free supernatant of the ferments from the end-point of fermentation clearly demonstrated the occurrence of a novel glycolytic pathway in *Thermococcus zilligii* (Xavier, da Costa et al. 2000). In another study,  $^{13}\text{C}$  NMR analysis of the intracellular extracts was used for quantitative flux determinations in genetically engineered *Bacillus subtilis* producing riboflavin (Dauner, Bailey et al. 2001, Dauner and Sauer 2001).

$^1\text{H}$  NMR has also been used for studying microbial metabolism and metabolic profiling.  $^1\text{H}$  NMR has been used for studying the metabolism and substrate utilization in *E. coli* and *Pseudomonas aeruginosa*. Cell-free supernatants of the ferments that were sampled during the time-course of the fermentations were measured by  $^1\text{H}$  NMR which proved to be a great profiling tool (Behrends, Ebbels et al. 2009). In another study,  $^1\text{H}$  NMR was used for investigating the effect of temperature on the metabolome of bacterium *Vibrio coralliilyticus*, by measuring the extracted intracellular metabolites in the different samples (Boroujerdi, Vizcaino et al. 2009).  $^1\text{H}$  NMR has also been used in the field of medical microbiology. *In vitro*  $^1\text{H}$  NMR has been used to study some of the metabolic reactions in gut microflora which are assumed to protect humans against colon cancer (Combourieu, Elfoul et al. 2001). These are only examples of the studies that have used NMR successfully for gaining new insights into the metabolism of the investigated bacteria.

### **3.9. *In vivo/in vitro* NMR spectroscopy for studying microbial metabolism**

Regarding the growth of bacteria, '*in vitro*' refers to growing and studying the bacteria in a controlled environment, whereas '*in vivo*' refers to the non-controlled and native environment of the cells. In this section, the term *in vivo* will be used to refer to both *in vivo* and *in vitro*

studies, with the goal of just emphasizing the fact that the investigated microorganisms were alive.

*In vivo* NMR approaches require no sampling and quenching of the metabolism and the process is not disturbed by any means. Sampling and quenching fractions of a fermentation batch is tedious and will introduce specific errors and irreproducibilities (Mashego, Rumbold et al. 2007). Moreover, the metabolic turnover rate is high and changes can happen in the milliseconds to seconds scale (van der Werf, Takors et al. 2007), and the lag between sampling a biological process and the measurement can lead to misinterpretations. The real-time nature of *in vivo* NMR measurements prevents this problem. *In vivo* NMR is also very suitable for kinetic studies, as the time-series can be recorded on the same sample with no change in the condition. Moreover, some of the species can only be monitored under *in vivo* experiments –macromolecules such as glycogen, polyphosphates, and other intracellular polymers, as well as the proton and sodium transmembrane gradients. One of the drawbacks of *in vivo* NMR is related to the heterogeneity of the samples and accumulation of paramagnetic ions that can lead to the broadening of NMR signals (Grivet and Delort 2009). Reference deconvolution can be a good solution to solve the linebroadening problem of *in vivo* NMR, as is shown in Paper II (Ebrahimi, Larsen et al. 2015).

Among the nuclides that have been studied for *in vivo* NMR studies of biological systems,  $^{31}\text{P}$  has been the most popular nuclide, as it allows for monitoring the intracellular pH, the dynamics of intracellular phosphate pools, and obtaining information about the energetic status of the cells by measuring metabolites like ATP and ADP. Besides, the natural abundance of  $^{31}\text{P}$  is 100% and it has a relatively high magnetogyric ratio. Carbon-13 NMR has also been widely used to establish metabolic routes and to follow the fate of individual carbon atoms through different pathways, thereby enabling the determination of carbon fluxes. Because of the broad range of carbon chemical shift, if  $^{13}\text{C}$ -labelled substrates are available, carbon will be the nuclide of choice (Ramos, Neves et al. 2002). The pattern of metabolites dynamics, which is acquired from carbon and phosphate *in vivo* NMR, has provided integrated description of different metabolic pathways in bacteria. Carbon-13 NMR can provide the information regarding the intermediates in glycolysis,  $\text{NAD}^+/\text{NADH}$ , and end-products, while  $^{31}\text{P}$  NMR can provide the information regarding intracellular pH and phosphate, as well as nucleotide triphosphates that are important to understand the energetic state of the system (Ramos, Neves et al. 2002).

Lactic acid bacteria (LAB) are one group of microorganisms that have been investigated by *in vivo* NMR. In 1984, *in vivo* NMR was used for the first time to study the metabolic regulation in LAB. *In vivo*  $^{31}\text{P}$ -NMR was used to measure the intracellular inorganic phosphate ( $\text{P}_i$ ) and relative levels of glycolytic intermediates, in *L. lactis* (Thompson and Torchia 1984). Since then *in vivo* NMR of proton, carbon, and phosphate nuclides have been used to study LAB

metabolism. For instance,  $^{13}\text{C}$  *in vivo* NMR has been extensively used to investigate regulation of sugar metabolism in these microorganisms (Neves, Ramos et al. 1999, Ramos, Neves et al. 2002). *In vivo* NMR has also been used for understanding the metabolism of different carbohydrate sources by *L. Lactis* (Neves, Pool et al. 2005). *L. lactis*, due to its relative metabolic simplicity, is one of the most studied strains of LAB. Two dimensional NMR techniques have also been used for real-time *in vivo* measurements of bacterial metabolic dynamics. Sequential  $^1\text{H}$ -NMR and  $^1\text{H},^{13}\text{C}$ -HSQC (Heteronuclear Single Quantum Coherence), were successfully used to investigate the linolenic acid hydrogenation pathway of a gastrointestinal bacterium, *Butyrivibrio fibrisolvens* (Fukuda, Nakanishi et al. 2009).

One of the interesting applications of *in vivo* NMR has been to study the adaptive stress responses of microorganisms, when they are exposed to environmental stressors. NMR can be very helpful in this context, by allowing a real-time non-destructive and non-perturbing measurement of the rapid metabolic changes of microorganisms under the stress conditions. As an example of such a study, *in vivo*  $^{13}\text{C}$  NMR has been applied to investigate how the metabolism of yeast *Saccharomyces cerevisiae* is affected by the concentration of exogenous ethanol and to model the metabolic profiles (Martini, Ricci et al. 2004, Martini, Ricci et al. 2006, Ricci, Aggravi et al. 2012). Other nuclides such as  $^{23}\text{Na}$  have also been used for *in vivo* NMR studies of microorganisms. For example,  $^{23}\text{Na}$  NMR spectroscopy was used to study the sugar transport and to investigate if it is sodium dependent in live *Fibrobacter succinogenes* cells (Delort, Gaudet et al. 2002, Delort, Gaudet et al. 2004).

Different experimental setups have been devised for *in vivo* measurements, having two main types: *perfused* and *non-perfused* systems. In the perfused system, a physiological medium with controlled pH, gas concentration (e.g.  $\text{O}_2$  and  $\text{CO}_2$ ), and composition is circulated to provide the cells in the magnet with nutrients. In this case, metabolites do not accumulate in the medium, different experiments can be made on the same sample by changing the composition of the medium, and also cells can be kept alive in the magnet for a longer time. In this system, cells can be immobilized on a solid matrix to avoid their circulation with the medium (Grivet and Delort 2009). There are also other interesting setups for *in vivo* measurement, like non-perfused systems for anaerobic and aerobic cell growth, and also bioreactors. For the aerobic system, an airlift system is designed to continuously provide the sample with  $\text{O}_2$  (Santos and Turner 1986, Lemos, Serafim et al. 2003, Neves, Pool et al. 2005, Lemos, Dai et al. 2007). To increase sensitivity and decrease NMR data acquisition time in the *in vivo* studies, the group of de Graaf, designed a bioreactors that allows the in-magnet growth of the cells in high density. Despite being very interesting, this setup is not commercial and no other group has used it (Hartbrich, Schmitz et al. 1996, Gonzalez, de Graaf et al. 2000, Grivet and Delort 2009). Generally, wider NMR tubes (10-20 mm) are used for *in vivo* measurements, to increase the sample size and as a result the sensitivity (Grivet and Delort 2009). These are of course not all the examples of the studies that have applied

NMR successfully to obtain valuable knowledge and insight into microbial metabolism, and there is still a growing interest in applying NMR techniques in microbial metabolomic studies (Xu, Wang et al. 2014).

*In vivo* NMR can serve as a very efficient tool for selection and optimization of industrial strains, and to enhance their properties and nutritional value. Therefore, food industry can benefit from *in vivo* NMR for studying microorganisms like LAB.

### **3.10. The application of *in vitro* NMR in micro-PAT**

Present PhD study focused on studying the fermentation process of two strains of LAB by NMR spectroscopy, which we called ‘micro-Process Analytical Technology (micro-PAT)’. In micro-PAT concept, ‘micro-’ is derived from the subjects of the study that are microbes and PAT refers to studying the fermentation of bacteria which is a biological process. Process Analytical Technology (PAT) was first introduced by the American Food and Drug Administration (FDA) in 2004, for the quality control in the pharmaceutical industry (Food and Drug Administration 2004). However, the hypothesis and the concept are applicable to other manufacturing industries and processes, including biological processes. Multivariate data analysis and experimental design, analyzers such as spectrometers for real-time monitoring of processes, and finally continuous improvement and optimization are the main tools in PAT. In the context of biopreservation by lactic acid bacteria, PAT can help to improve and optimize the efficiency of biopreservation by real-time monitoring of the fermentation (Skibsted and Engelsen 2010).

NMR can serve micro-PAT because it is a non-destructive analytical technique that allows real-time investigation of the fermentation process. This was the main factor that made NMR the analytical method of choice for our study. No other analytical technique has the potentials that are listed in Fig. 3.11 all together and from this aspect, NMR is genuinely unique for micro-PAT. The possibility of *in vivo* measurements by NMR allows studying the effect of different internal and external factors on the metabolism of microorganisms and optimizing them to shift the metabolism towards for instance the overexpression of desirable metabolites or suppressing the undesirable ones. It also allows comparing different strains and selecting more efficient strains for specific purposes such as biopreservation.

## 4. Multivariate data analysis and chemometrics in metabolomics

Metabolomics provides an overview of the metabolic state of a biological system. The analytical techniques that are normally used in metabolomics allow for recording hundreds or even thousands of metabolites in a single measurement, which in itself is a great achievement, but at the same time creates very huge datasets that require efficient multivariate data analysis techniques to extract useful and tangible information. Multivariate data analysis and chemometrics enable quantitative modeling of multivariate and multidimensional data and can provide visual representation of the information. Nowadays, chemometrics is a well-established field, and is used for multivariate calibration, pattern recognition and multivariate statistical process monitoring and control. Three basic categories of analysis techniques can be defined for chemometrics:

- a) **Exploratory analysis:** *provides an unbiased and general overview of the data and is useful for detecting patterns and trends.*
- b) **Classification and discriminant analysis:** *identifies predetermined classes of the samples.*
- c) **Regression analysis and prediction models:** *models the quantitative relationship between two blocks of data*

In the following, a brief description is provided for the multivariate and chemometric techniques that have been used in this thesis and the included papers.

### 4.1. Principal Component Analysis (PCA)

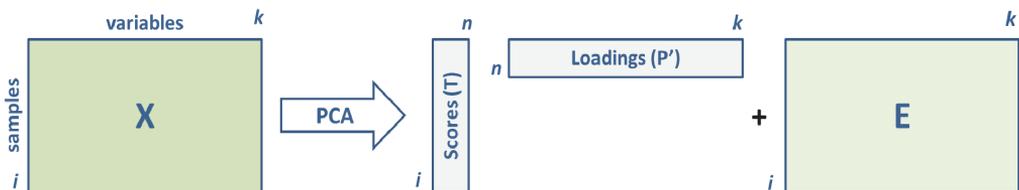
*Principal Component Analysis* (PCA) (Hotelling 1933, Wold, Esbensen et al. 1987, Jackson 2005) is an unsupervised method, which is probably the most commonly used multivariate exploratory analysis technique, and can be used to extract and visualize the systematic variations in a data matrix  $X$ . If  $X$  has  $i$  rows (samples) and  $k$  columns (variables), the data will be spread in a  $k$ -dimensional space that the variable make, in which every sample is a point. In the commonly applied analytical platforms in metabolomics, one measurement can vary from several hundreds to tens of thousands of variables. As the human vision is limited to three dimensions, it is obviously not feasible to examine such data without reducing the

dimensions. This is where PCA becomes handy by reducing dimensions of the data. PCA allows the inspection of the  $k$ -dimensional data in subspaces and also finds dimensions that are the most informative ones, based on the percentage of the variance that they represent.

Hidden latent variables or principal component's (PCs) are made by the combinations of the measurement variables. The first PC is the direction in the  $k$ -dimensional space, which includes the highest distribution of the samples or the highest percentage of variance. The second dimension with the highest percentage of variance that is orthogonal (not correlated) to PC1, is taken as PC2, and the procedure continues to find enough PCs to explain all the systematic variance in the data. So, the dimension of the data is reduced from  $k$  to the number of PCs, which is usually much smaller than  $k$ , and uninformative (non-systematic) dimensions are ignored. Every two PCs make a two-dimensional sub-space- a plane. The coordinates of the projection of samples onto each PC is called scores  $T$ , and they are weighted averages of the variables (e. g. metabolites) in the data (Trygg, Gullberg et al. 2006).

The visualization of the scores, which is called a scores plot, can be very informative as it gives an overview of all samples and how they relate to each other. Different groups of the samples, trends and outliers (deviating samples) can show up in scores plots. The other important plot that can be derived by PCA is the loadings plot, which shows which measurement variables (columns) of  $X$  make the PCs. The loading plots allow investigating which measurement variables (sometimes metabolites) correspond to the observed patterns in the scores plots (Trygg, Gullberg et al. 2006).

A PCA model can be written and shown schematically in Figure 4.1.

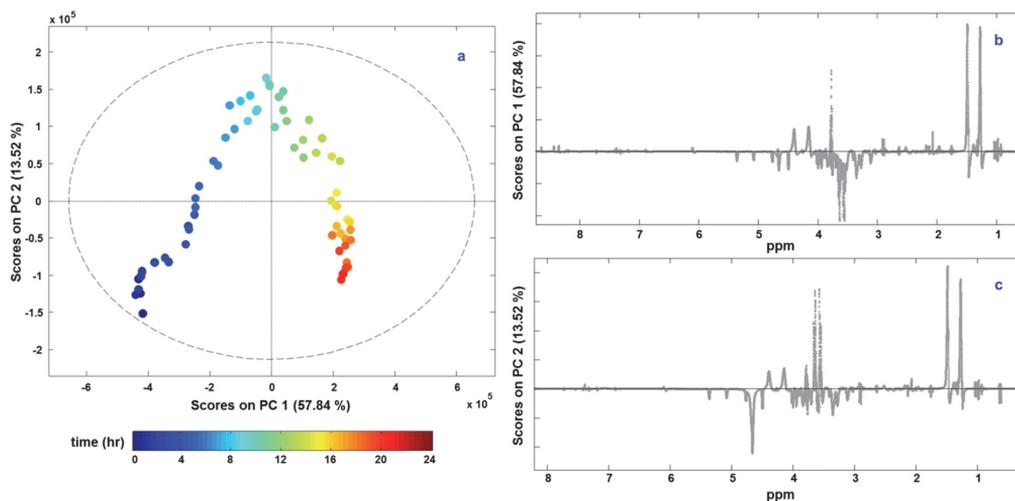


$$X = T \times P' + E$$

**Figure 4.1.** Schematic illustration of a PCA model.  $T$  and  $P$  are the scores and the loadings, and  $E$  is the residual matrix.  $i$  and  $k$  are the number of samples and variables respectively.  $n$  is number of the PCs.

Different data preprocessing techniques can be used prior to any multivariate analysis including PCA, to improve the biological information content of metabolomics data. Centering, scaling, normalization, and transformation can be used, which will not be discussed further here. For further reading about this, see (van den Berg, Hoefsloot et al. 2006).

Figure 4.2 shows an example of the PCA on a time-series *in vitro* NMR data from the fermentation of *Lactobacillus rhamnosus* DSM 20021, in which  $^1\text{H}$  spectra were recorded for 24 hrs. The scores plot, as colored by time, shows the fermentation trajectory and the metabolism of the sample during fermentation. The loadings plot shows the peaks from the metabolites or nutrients that vary during fermentation and correspond to the pattern that the scores plot shows. Therefore, they can provide valuable information of the metabolic changes that happen during the investigated biological process.



**Figure 4.2.** PCA results of a time-series *in vitro*  $^1\text{H}$  NMR data from the fermentation of *Lactobacillus rhamnosus* DSM 20021. a) the scores plot; b) the 1<sup>st</sup> loading plot; c) the 2<sup>nd</sup> loading plot.

## 4.2. Partial Least Squares regression (PLS)

Partial Least Squares Regression (PLS) method (Wold, Martens et al. 1983) is a supervised method and can be used when prior information about each sample is available. The information can be for example a quantitative value obtained from a reference method, or design information such as class membership of each sample.

The samples information is used as a response matrix  $Y$ . Then, PLS builds a quantitative relationship between the data,  $X$ , as the predictor (descriptor) matrix, and  $Y$ , as the response matrix (Wold, Ruhe et al. 1984). The  $Y$  matrix can contain both quantitative (e. g. concentration of one of the metabolites) and qualitative (e.g. class membership) information. The additional information in  $Y$  is used by the PLS method to focus only on the  $Y$ -related variance in  $X$ , e. g. separation between genotypes, rather than providing an overall view of all variation in the data in an unsupervised manner, as is performed by PCA.

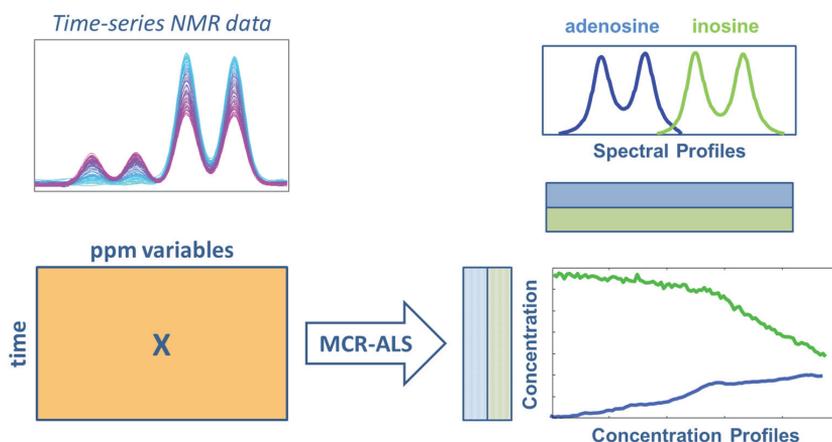
PLS is a regression method and as a result, PLS models can be used to predict the properties ( $Y$ -values) of new unknown samples, e. g. the genotype of a new sample, based on the previously calculated prediction models. The  $Y$  matrix can indicate different properties in each column. If  $Y$  contains qualitative information such as class membership of the samples, the PLS method is called PLS Discriminant Analysis (PLS-DA), in order to distinguish it from the situation where  $Y$  is quantitative (Trygg, Gullberg et al. 2006).

### **4.3. Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS)**

Multivariate Curve Resolution (MCR) (Lawton and Sylvestre 1971, de Juan, Jaumot et al. 2014) is a method that can resolve a data matrix into the pure physical/chemical profiles of the chemical mixture components. MCR was originally developed to encompass evolutionary analytical data from a process or an analytical measurement. Multivariate curve resolution with alternating least squares (MCR-ALS) works by optimizing both the concentration profile and the pure spectral profiles in an iterative cycle (Tauler 1995). The basic assumption in MCR is that the data matrix is bilinear. Based on this, MCR decomposes the data into additive bilinear models for the chemical entities. The unique advantage of MCR is that the profiles are meaningful (Engelsen, Savorani et al. 2013), and for instance for a time-series NMR data, MCR will provide the time profiles and the spectral profiles of the individual chemical compounds in the sample. This characteristic helps enormously in the interpretation of the results and ensures that the results provided by MCR methods can be easily understood, as they present physical or chemical properties. This is one of the main differences of MCR compared to other bilinear data analysis methods. For instance, methods such as PCA do not provide the true chemical or scientific models, because the real mixed contributions do not hold the orthogonal or statistical independency as natural properties. However, these methods still have a strong exploratory value, as they are extremely robust compared to the more ambiguous MCR (Engelsen, Savorani et al. 2013).

An MCR model of a time-series NMR data is built as shown in Figure 4.3. The time-series NMR data that is referred to in Figure 4.3 includes two chemical components, adenosine and

inosine. MCR resolved the data into time (concentration) and spectral profiles, which are both tangible chemical and physical properties. The iterative MCR-ALS method was applied to perform the mathematical decomposition of the data.



$$X = C \times S^T + E$$

**Figure 4.3.** An example of MCR-ALS applied to time-series NMR spectra.  $C$  and  $S$  are the time (concentration) and spectral profiles respectively, and  $E$  is the residual matrix. The time-series data shows consumption of adenosine and production of inosine. The plotted concentration profiles are not geometrically in scale with the time dimension of matrix  $X$ .

MCR-ALS is performed in an iterative manner. It starts with estimating the number of the components presented in the data matrix, which is usually done by PCA or singular value decomposition (SVD) (De Lathauwer, De Moor et al. 1994). The second step is to calculate initial estimations of spectral or concentration profiles for each one of the chemical components. There are different methods for calculating the initial estimates. One of the methods can be using evolving factor analysis (EFA) (Maeder 1987, Keller and Massart 1991) in one of the dimensions of the data in order to find selective regions of the signals, and then apply PCA on the selective regions to get the initial estimates. Methods such as SIMCA (Wold and Sjöström 1977) can also be used for this purpose. Then, an iterative alternating least squares process based on the following two equations strives to optimize the initial estimations that are introduced to the algorithm:

$$C = X(S^T)^+ \quad (4.1)$$

$$S^T = (C)^+X \quad (4.2)$$

The sign ‘+’ denotes pseudo inverse, and  $C$  and  $S$  are the concentration and spectral profiles, respectively. Different constraints including non-negativity, unimodality and closure can be applied in order to reduce the number of the possible solutions for  $C$  and  $S^T$ , during this iterative optimization process. This helps to direct the final solutions to the true profiles that make the observed experimental data. The iteration is stopped when the model is converged. It is when the relative difference between the values from two consecutive iterations reaches the defined threshold value or when number of the iterations exceeds the allowed number of iterations (Tauler 1995, de Juan, Jaumot et al. 2014).

In Paper II, MCR-ALS was used to model the time profiles of selected metabolites, from the time-series NMR data, by applying nonnegativity constraint on both dimensions.

#### **4.4. Analysis of Variance-Simultaneous Component Analysis (ASCA)**

Analysis of variance (ANOVA) can provide information about the variance between and among groups of samples in an experimental design and their significance. ANOVA partitions the variance of a variable into components that originate from different sources, and the calculated  $p$ -values indicate the significance of design factors. ANOVA is used for univariate data, when only one variable, or in the case of metabolomics data only a single metabolite is measured in samples from an experimental design (Searle 1971). Although multivariate-ANOVA (MANOVA) is a multivariate generalization of ANOVA as presented in statistics, it is not efficient for analyzing big datasets that include many variables, because the assumption of the singularity of the covariance matrix will not be satisfied (Mardia, Kent et al. 1979, St and Wold 1990); in metabolomics data, typically, the metabolites can be highly correlated.

ANOVA-simultaneous component analysis (ASCA) (Smilde, Jansen et al. 2005) is a multivariate generalization of ANOVA. ASCA can be applied to different types of data that have a balanced experimental design, or a temporal structure. Therefore, it can be applied for metabolomics, as well as other –omics datasets, especially that experimental design is being increasingly used in these fields (Smilde, Hoefsloot et al. 2008). By separating the variance into the effect matrices from the design factors, ASCA allows to investigate the importance of the design factors and also interpret the data. The current version of ASCA that was proposed by *Smilde* and co-workers in 2005, is based on the concepts of multilevel component analysis that have been proposed for time-resolved metabolic fingerprinting data (Jansen, Hoefsloot et al. 2005), and also simultaneous component models for analyzing multivariate time series (Timmerman and Kiers 2003). ASCA uses these concepts and generalizes it for data with a designed structure (Smilde, Jansen et al. 2005).

In Paper II, an experimental design was prepared with two strains of lactic acid bacteria, two pH values (pH 6.5 and 5.5), two levels of glucose concentration (2.5 and 0.25 g/l), and two batch fermentation replicates, and real-time *in vitro* NMR data were recorded of the samples fermentation. The data were then analyzed and the metabolic time profiles were extracted. ASCA was applied on the profiles, in order to investigate the significance of the experimental design factors, by investigating the resulted *p*-values, and also partition the variance of the data into the effect matrices from the design factors. Using PCA on the resulted effect matrices allowed investigating which metabolites were influenced by the different design factors. Such information is very valuable in a metabolomics study, and metabolomics can definitely benefit from ASCA.

## 4.5. Clustering by *k*-means algorithm

*K*-means clustering is an unsupervised algorithm that aims to find the best partitioning of *n* observations (or objects) into *k* clusters or groups, where *k* is a number defined by the user. The algorithm starts by randomly selecting *k* points (objects) as the initial groups' centroids. Then, the Euclidean distance between all the objects and the centroids are calculated and each object is assigned to the cluster to the centroid of which it is the closest. In the next step, for each cluster, the object which is the most similar to the average of all the objects in the *k*th cluster is defined as the new centroid and objects are clustered again, based on their distance from the new centroids. The process of finding the new centroids and re-clustering the objects is repeated iteratively until the convergence criterion is met. The convergence criterion used in our method is minimizing the Within Cluster Sum of Squares (WCSS), which is the average squared Euclidean distance between the objects and their cluster centroids. This is a measure of how well each centroid represents the group or cluster members. The algorithm has converged when WCSS does not decrease any further with iterations or decreases below a predefined threshold (MacQueen 1967, Tran, Wehrens et al. 2005). Generally, for each clustering, replicate runs/restarts are performed. Restarts of *k*-means will help to make sure that the algorithm does not converge to local minima. Each one of the replicates begins from a different randomly selected set of initial centroids and the final solution that *k*-means returns is the global minimum which has the lowest WCSS.

As presented in Paper III, *k*-mean clustering of multispectral images was used to develop a new and semi-automated approach for quantifying mold growth based on the colony size or area.



## 5. *In vitro* NMR studies of the fermentation of lactic acid bacteria (Paper II)

*In vitro* NMR can provide valuable information about the dynamic metabolism of microorganisms. As discussed earlier in Section 3.9, it can be used for studying the adaptive responses of microorganism to different stress factors. Therefore, we developed a protocol for performing real-time *in vitro* NMR measurements of the bacterial fermentation and applied it to an experimental design with the factors of interest for the two strains of LAB, *Lactobacillus rhamnosus* DSM 20021 and *Lactobacillus plantarum subsp. plantarum* DSM 20174. The experimental design included two levels of glucose concentration (2.5 and 0.25 g/l), two pH values (pH 6.5 and 5.5), and duplicate fermentation batches (Paper II, (Ebrahimi, Larsen et al. 2015)). In the following, some of the experimental and data analysis results, including discussions that are not included in the paper for the sake of conciseness, are presented.

### 5.1. Experimental design, sample preparation, and data acquisition

When including the pH value as one of the factors in an experimental design, it is important to use buffer systems that are strong enough to keep pH relatively constant during the measurement. This becomes more important in experiments like LAB fermentation where pH drops over time. Therefore, in designing the buffer system for the samples, the buffering capacity was calculated to be sufficiently high to provide a stable pH value during fermentation. Sodium phosphate buffer was used with the concentrations of 0.15M and 0.5M, in order to provide buffering systems with pH 6.5 and 5.5, and capacities of 0.09 and 0.10, respectively. Although the concentration of the pH 5.5 buffer is relatively high compared to the routine buffers that are used in the protocols for metabolomic studies by NMR, it allowed tuning and matching of the NMR probe, and the duration of the 90° pulses were between 18 and 19  $\mu$ s. However, the linewidths of the spectra for the pH 6.5 samples was slightly narrower than for the pH 5.5 samples, but as presented in Paper II (Ebrahimi, Larsen et al. 2015), this difference was corrected by using reference deconvolution. For most strains of LAB, when pH drops below 4.8, the metabolic rates are greatly decreased, and the growth is retarded. Therefore, a strong buffer system was also necessary to ensure that the pH at the end of glucose consumption, especially for pH 5.5 samples and samples with the higher glucose concentration, would not drop too low to disturb the metabolism. Table 5.1 lists the pH values

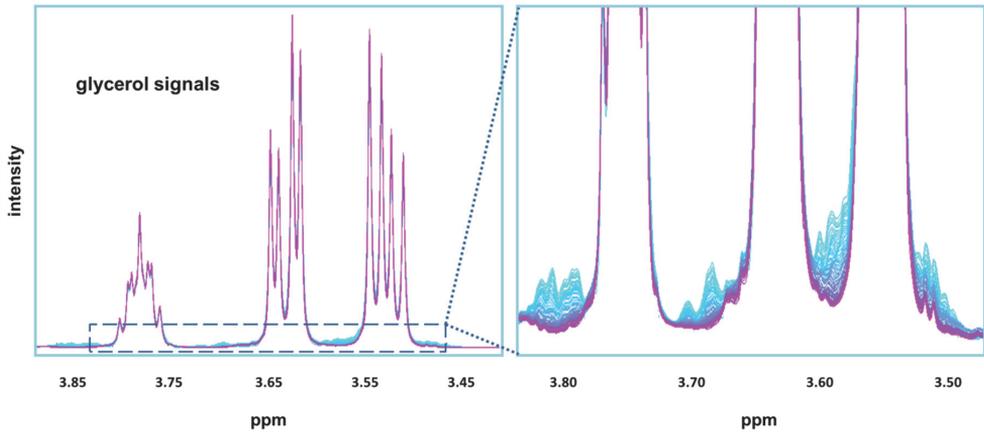
of the samples at the end of the fermentation. The pH values at the end points of fermentation show that the buffer system has worked efficiently.

**Table 5.1.** The final pH values of the samples after fermentation. ‘R’ refers to *L. rhamnosus*, ‘P’ to *L. plantarum*, ‘6.5’ to samples with pH 6.5, ‘5.5’ to samples with pH 5.5, ‘GH’ to samples with 2.5 g/l of glucose, and ‘GL’ to samples with 0.25 g/l of glucose.

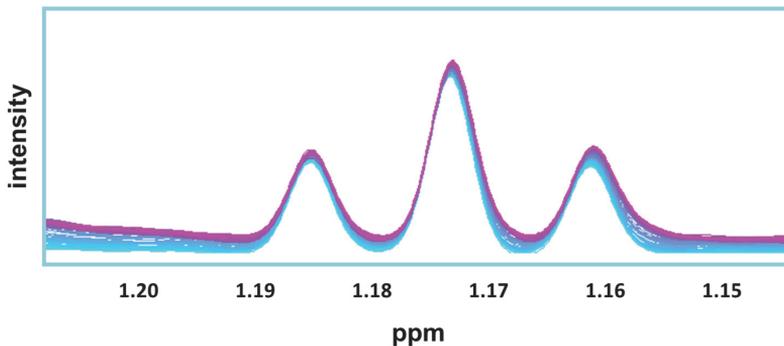
Samples	PH final	
	replicate 1	replicate 2
R 6.5 GH	6.30	6.30
P 6.5 GH	6.29	6.29
R 6.5 GL	6.52	6.56
P 6.5 GL	6.52	6.53
R 5.5 GH	5.35	5.36
P 5.5 GH	5.30	5.21
R 5.5 GL	5.59	5.44
P 5.5 GL	5.57	5.41

The bacterial cells that were used in the experiment were frozen at  $-80^{\circ}\text{C}$  in glycerol solution until use. The glycerol gave rise to strong signals in the NMR spectra which necessitated washing the cells twice with the chemically defined interaction medium (CDIM) prior to sample preparation, in order to ensure the decrease of glycerol concentration. Figure 5.1 shows an example of the spectra from a sample without the cell wash. Glycerol signals are even stronger than glucose signals, and would definitely limit the receiver gain. Therefore, it was necessary to wash the cells before inoculation, and washing the cells twice with CDIM was able to eliminate most of the glycerol signal.

In order to sterilize the NMR tubes, autoclaving proved to be more efficient than washing by ethanol. Due to the shape of NMR tubes, ethanol will not evaporate easily after wash and will appear in the samples, which is not desired. However, as ethanol was used to sterilize the bench, a very small amount can still be observed in the samples, as Figure 5.2 shows. The small residual ethanol did not change during fermentation and did not suppress cell growth.



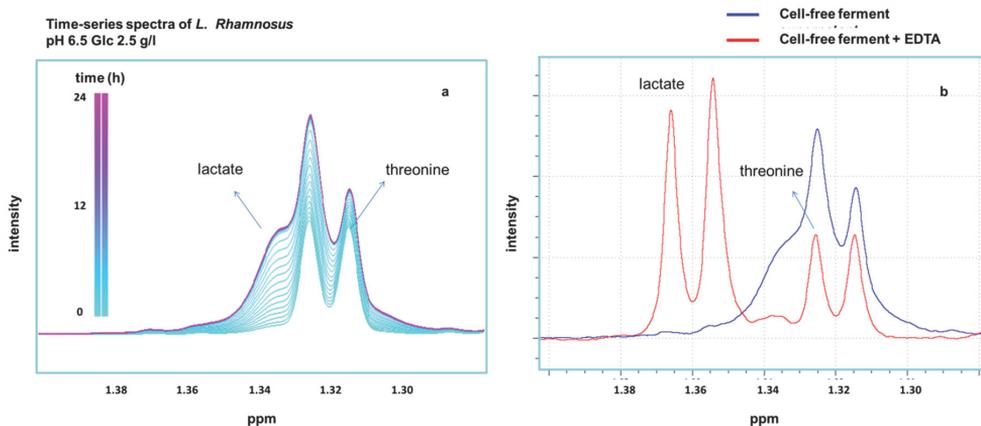
**Figure 5.1.** Strong signals from glycerol in the samples without washing the cells prior to inoculation. Glycerol signals compared to the glucose signals, that can be observed in the zoomed figure, are much stronger.



**Figure 5.2.** Ethanol triplet in one of the time-series. The signal was very weak compared to the signals from the nutrients and metabolites.

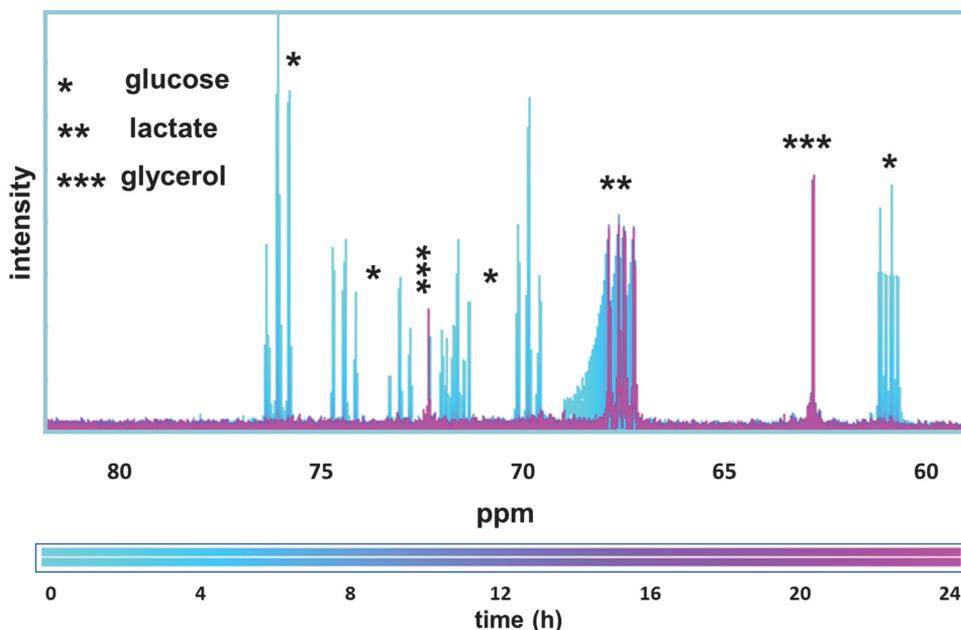
In the recorded time-series data, lactate signals were quite broad, and for instance for the signal of the methyl group in lactate at 1.32 ppm, instead of the expected doublet form, a very broad signal was building up over the time course of fermentation. Figure 5.3a shows the lactate doublet for *L. rhamnosus* sample at pH 6.5 and the glucose concentration of 0.25 g/l. In order to verify that the broadening of lactate signals were due to the formation of Na/Ca lactate, cell free supernatant of the sample at the end point of fermentation was prepared using 0.2  $\mu\text{m}$  filters, and ethylenediaminetetraacetic acid (EDTA) was added to the solution. EDTA can form chelates with  $\text{Na}^+/\text{Ca}^{2+}$  ions, and therefore release lactate. Figure 5.3b shows the spectra of the cell free supernatant with and without the added EDTA. After adding EDTA, the sharp doublet form of lactate is recovered which confirm that the line broadening is due to

the formation of Na/Ca lactate. The doublet shifts towards larger ppm values after addition of EDTA, due to the relatively high acidity of EDTA that can change the pH of the sample. This test was merely performed to explain the observed line broadening. For the purpose of the data analysis, the original samples and spectra were used. Special attention was paid to have the same time frame for preparation and measurement of all the samples. Besides, for each sample the pulse width (P1) and transmitter frequency (O1) were calibrated to enhance the spectral quality.



**Figure 5.3.** Recovery of lactate doublet by addition of EDTA. a) time-series spectra of one of the *L. rhamnosus* samples, colored by time, b) spectra of the cell-free ferment of the sample with and without added EDTA.

The use of  $^{13}\text{C}$ -labeled glucose for the time-series study was also investigated. Figure 5.4 shows the  $^{13}\text{C}$  time-series spectra that were recorded for *L. rhamnosus* at pH 6.5 and 2.5 g/l of  $^{13}\text{C}$ -labeled glucose. The goal of using labeled glucose substrate was to investigate if metabolites such as pyruvate that have key roles in the metabolic pathways of the bacteria can be observed. The only captured signals were from glucose and lactate. Glycerol signals were also observed, as the spectra were recorded before developing the protocol for washing the cells prior to inoculation, and therefore the cell suspension included considerable amount of glycerol. Considering the fact that the labeled substrate did not add more information to study, normal glucose was used for all the final measurements.



**Figure 5.4.**  $^{13}\text{C}$  time-series spectra of *L. rhamnosus* at pH 6.5 and 2.5 g/l of  $^{13}\text{C}$ -labeled glucose.

## 5.2. Data processing

Besides the routine NMR data processing techniques such as phase correction, Fourier transformation, zero-filling, and baseline correction, reference deconvolution was also used on the spectra to enhance the spectral quality and data analysis results. All the processing of the data were performed in DOSYToolbox (Nilsson 2009). Figure 5.5 shows the layout of the DOSYToolbox, and the dotted red boxes show the function of the different parts of the toolbox that were used for the data processing. For the reference target lineshape, Gaussian, Lorentzian, or the combination of the two lineshapes can be selected in the toolbox. Performing reference deconvolution by the DOSYToolbox is fast and easy, and considering the advantages that this method can offer multivariate analysis of the data, it is definitely worth trying it on different metabolomics data, and data from biological samples which are inherently inhomogeneous and the spectral quality can be an issue. The only prerequisite for being able to use reference deconvolution is the presence of a well-resolved singlet with high signal-to-noise ratio in all the processed spectra, to be used as the reference signal. The reference signal should be one of the strongest resonances in the spectrum to get the best results from reference deconvolution.

While Paper I (Ebrahimi, Nilsson et al. 2014) showed the efficiency of reference deconvolution for improving multivariate data analysis (PCA, PLS) results of an artificial

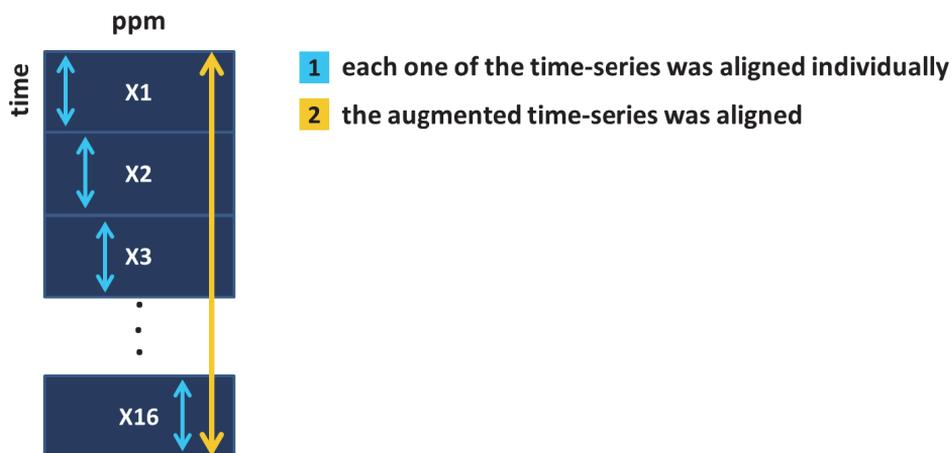
metabolomics experimental design, Paper II (Ebrahimi, Larsen et al. 2015) demonstrates for the first time the advantageous use of reference deconvolution on real metabolomics data for improving the multivariate analysis (MCR-ALS) results of the NMR data. Reference deconvolution successfully corrected the line broadening problem of the spectra in the *in vitro* measurements, which makes the analysis results of the data by methods like MCR-ALS, as well as any other method that has bilinearity of the data as its basic requirement, more reliable.



**Figure 5.5.** The layout of the DOSYToolbox. The dotted red boxes show the function of the different parts of the toolbox. The toolbox is freely available from [dosytoolbox.chemistry.manchester.ac.uk](http://dosytoolbox.chemistry.manchester.ac.uk).

Spectral shifts of NMR signals can happen as the result of the change in pH or temperature. This is not a desirable source of variation when using multivariate analysis techniques for the analysis of the data. Therefore, post-acquisition alignment techniques are necessary prior to multivariate analysis of the data. One of the methods that has proved to be very useful for this purpose is *icoshift* program (Savorani, Tomasi et al. 2010, Savorani, Tomasi et al. 2013). It allows alignment of a series of spectra, either relative to a reference signal, or in the user-defined intervals. For the time-series *in vitro* NMR data of the fermentations, the strong buffer system did not allow dramatic change of the pH and considerable shift in the signals of most of the nutrients and metabolites. However, for compounds such as acetic acid and histidine which are highly pH sensitive, even the small change in the pH resulted in large shifts in each time-series data, as well as between the time-series of the different samples. Such shifts are of course larger in samples with the higher glucose concentration that tend to produce more acidic metabolites by fermentation. In Paper II (Ebrahimi, Larsen et al. 2015), *icoshift* was

used to align each of the 16 time-series data, first relative to DSS signals, and then in defined intervals containing the selected metabolites. Intervals were defined for glucose, pyruvate, acetate, alpha-acetolactate, formate, glutamine, aspartate, adenosine, inosine, and adenine. The 16 time-series data were then augmented row-wise (along time-dimension) and *icoshift* was used again to align the selected signals. Figure 5.6 shows the alignment process schematically.



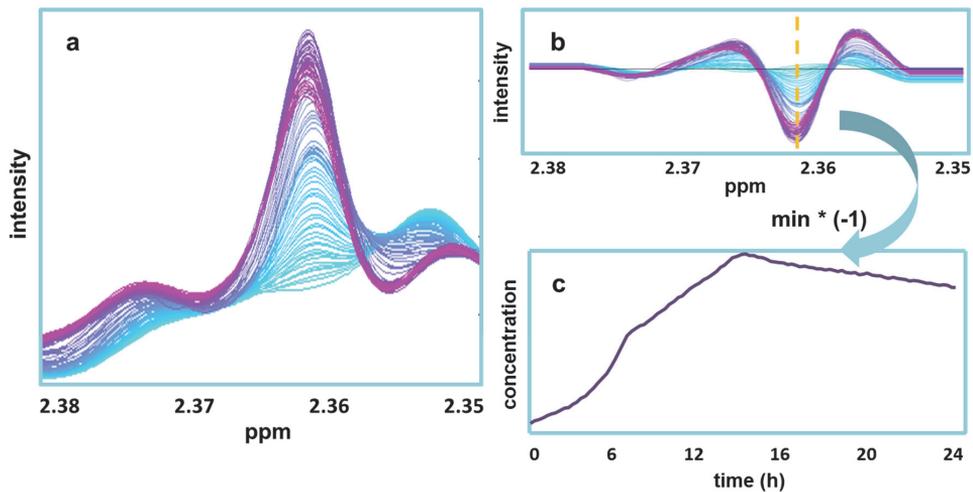
**Figure 5.6.** The schematic illustration of the alignment procedure of the data.

It is important to perform the alignment in a supervised manner, and the alignment results should be investigated before data analysis to ensure that the algorithm has not aligned non-corresponding neighboring peaks in overlapping signals, or signals that sit very close to each other.

### 5.3. Data analysis

In order to model the metabolic profiles, MCR-ALS or second derivative of the signals were used. In the data analysis, we tried to model as many signals simultaneously as possible, because it is both statistically preferred, and will also reduce the analysis time. Because of the strong buffer system, the position of most of the signals remained stable during the fermentation, and MCR-ALS was used on the augmented data to model their profile during fermentation. However, for some of the pH sensitive compounds, shift in the position of the signals was observed. For singlet signals like acetate and pyruvate, which shift considerably

between the samples, a perfect alignment was not achieved and instead the minimum of the second derivative of the signal was used to model the profile. Figure 5.7 shows the pyruvate signal in *L. rhamnosus* sample at pH 6.5 and glucose concentration of 2.5 g/l, and its second derivative that was used for the subsequent modelling of the pyruvate profile.

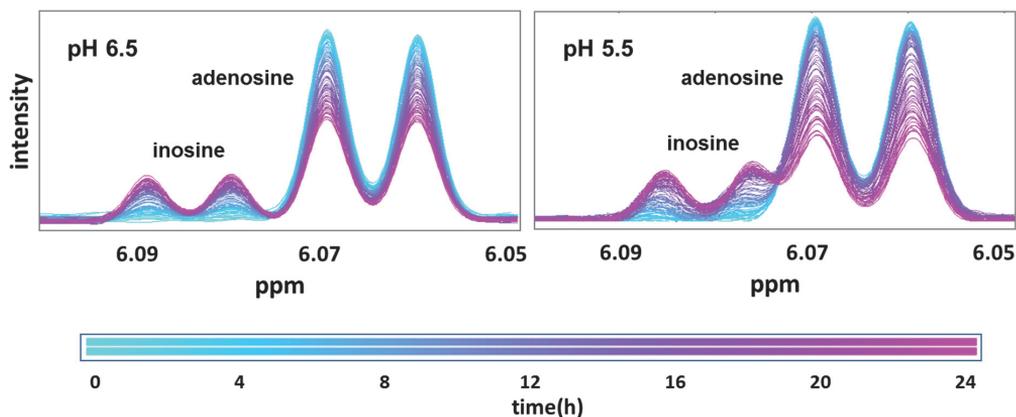


**Figure 5.7.** Calculating pyruvate signal in the *L. rhamnosus* sample, with pH 6.5, and glucose concentration of 2.5 g/l. a) pyruvate singlet, b) the second derivative of the signal, and c) calculated pyruvate profile.

The chemical shift of inosine differed for the two pH values, and at pH 5.5, the inosine doublet partly overlapped with the adenosine doublet, as shown in Figure 5.8. Because of this overlap, a suitable alignment could not be achieved between pH 6.5 and 5.5 samples. However, as the signal position was fairly stable in individual time-series and also between the samples with the same pH value, adenosine and inosine signals were modeled by two component MCR-ALS models for pH 6.5 and 5.5 separately.

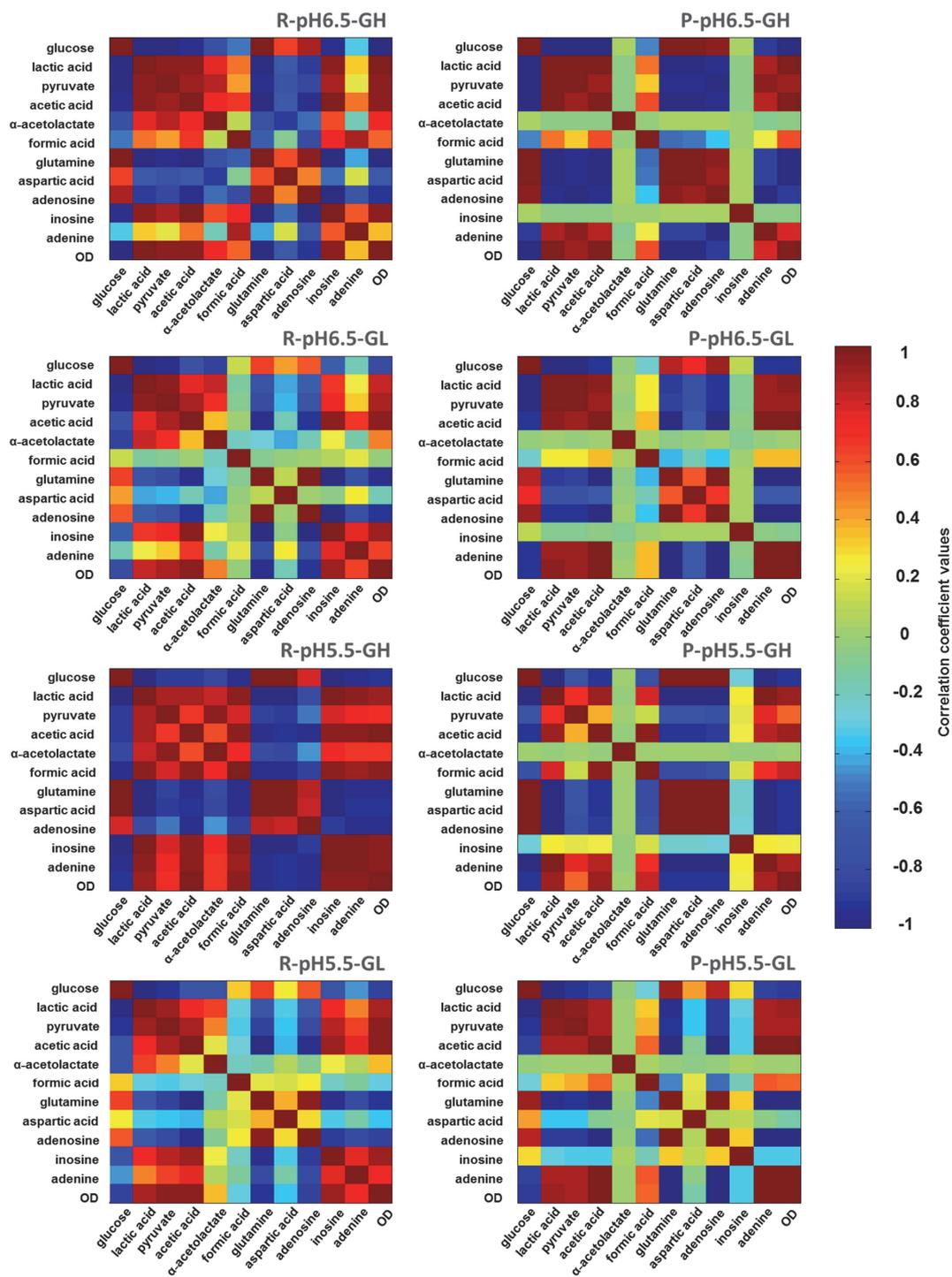
Modeling by parallel factor analysis 2 (PARAFAC2) (Kiers, Ten Berge et al. 1999) was also tested on the data, as it can be a good solution for data in which signals shift between the measurements as has been successfully performed for hyphenated chromatographic data such as LC-MS before (Khakimov, Amigo et al. 2012). The data were first structured in a tensor with the dimensions *spectra*  $\times$  *time*  $\times$  *samples*, with the NMR spectra as the flexible PARAFAC2 dimension. PARAFAC2 was applied on the intervals of selected signals, and models with different number of components were fitted to investigate if the correct metabolic profiles can be modeled. However, as the pattern of the profiles for a metabolite can vary significantly between the samples with the different strain, pH, and glucose concentration,

correct profiles could not be modeled by PARAFAC2 for these data. In these data set, both the spectral and the time dimension vary between the samples, and this cannot be handled by PARAFAC2. For signals that do not shift between the samples, using PARAFAC2 on the data with the structure  $time \times spectra \times samples$ , with time as the flexible mode can give the correct profiles. However, due to these additional complications, MCR-ALS was preferred for the final analysis results.

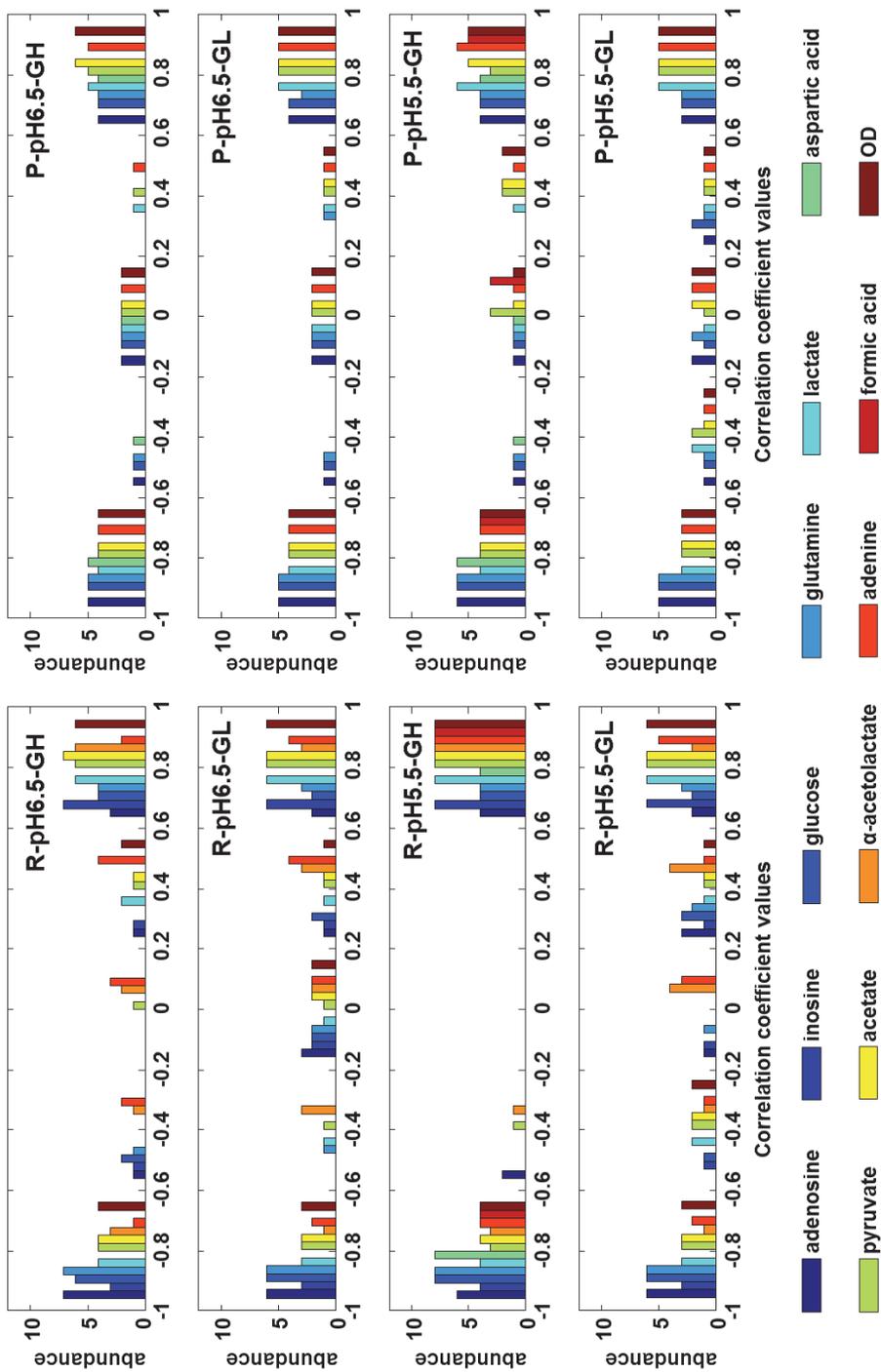


**Figure 5.8.** Adenosine and inosine signals in *L. rhamnosus* samples with pH 6.5 and 5.5, and the glucose concentration of 2.5 g/l.

Pearson correlation coefficients were calculated between the metabolic profiles that were calculated as described in detail in Paper II (Ebrahimi, Larsen et al. 2015) and partly in this chapter. Figure 5.9 shows the heat maps colored by the value of the correlations between metabolites. As the heat maps presented in the paper are limited to high correlations,  $> 0.8$  and  $< -0.8$ , this heat map is presented to provide more details on the correlations between the metabolites. In Figure 5.10, the histograms of the heat maps are presented, with the bars colored based on the metabolites. The histograms show that the correlations between the metabolites in *L. rhamnosus* samples are influenced more by the change in pH and glucose concentration than *L. plantarum* samples. In *L. rhamnosus*, by the change of the pH from 6.5 to 5.5, the correlations between metabolites considerably increase, meaning that more metabolites co-vary in the response to the acid stress (comparing the histograms ‘*R-pH6.5-GH*’ and ‘*R-pH5.5-GH*’ in Figure 5.10). Adenine and alpha-acetolactate are two of the metabolites that are influenced the most. This degree of change in the correlation coefficients is not observed in *L. plantarum* samples and the metabolism of this strain is less altered by the decrease in pH. Besides the pH, glucose concentration also affects *L. rhamnosus* more than *L. plantarum*. For *L. plantarum* at pH 6.5, the correlation levels between the samples with high and low glucose concentration almost do not change.

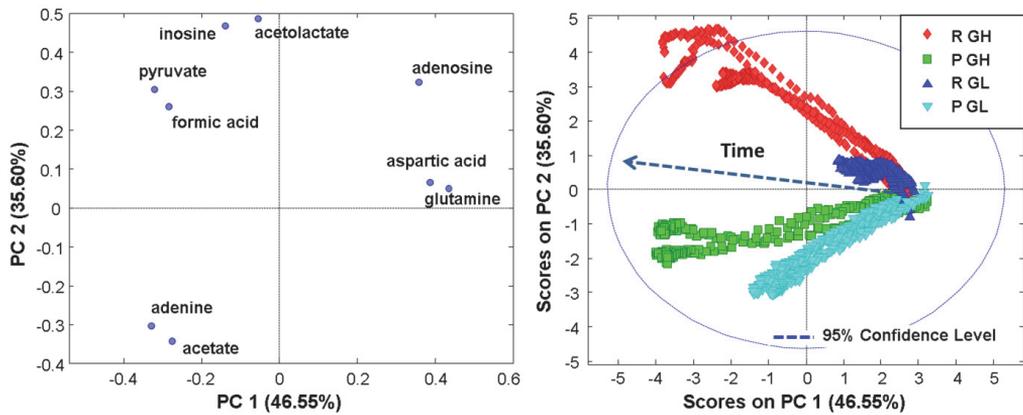


**Figure 5.9.** The heat maps colored by the Pearson correlation coefficients between the metabolites.



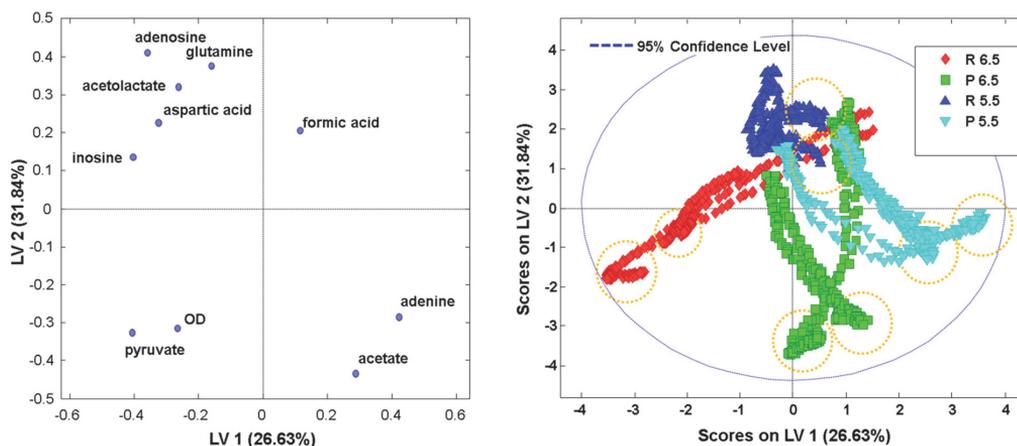
**Figure 5.10.** The histogram of the metabolites correlation coefficients heat maps.

In metabolomics data with a balanced experimental design, ASCA can be used to investigate the effects of the different design factors. In Paper II (Ebrahimi, Larsen et al. 2015), ASCA was used on the metabolic profiles to assess if the metabolic changes of the data were significantly influenced by the type of the strain, pH, and glucose concentration. Figure 5.11 shows the PCA results of the strain specific glucose concentration effect matrix. This matrix contains information on how the two different strains interact with the glucose concentration. It is obvious that the metabolism of *L. rhamnosus* is influenced more by the glucose concentration than the metabolism of *L. plantarum*; the shift in the fermentation trajectories of *L. rhamnosus* by the change in glucose concentration is highly significant.



**Figure 5.11.** PCA results of the strain specific glucose concentration effect matrix (the interaction between the strain type and the glucose concentration).

Figure 5.12 shows partial least squares discriminant analysis (PLS-DA) results of strain specific pH effect matrix. This matrix shows how the change in the pH affects the metabolism of the two strains. According to the figures, the change in the pH alters the fermentation trajectories, and pyruvate, formate, adenine, and acetate are among the main metabolites that are influenced.



**Figure 5.12.** Results from a PLS-DA of the strain specific pH effect matrix (the interaction between the strain type and pH). The dotted orange circles show the end points of the fermentations.

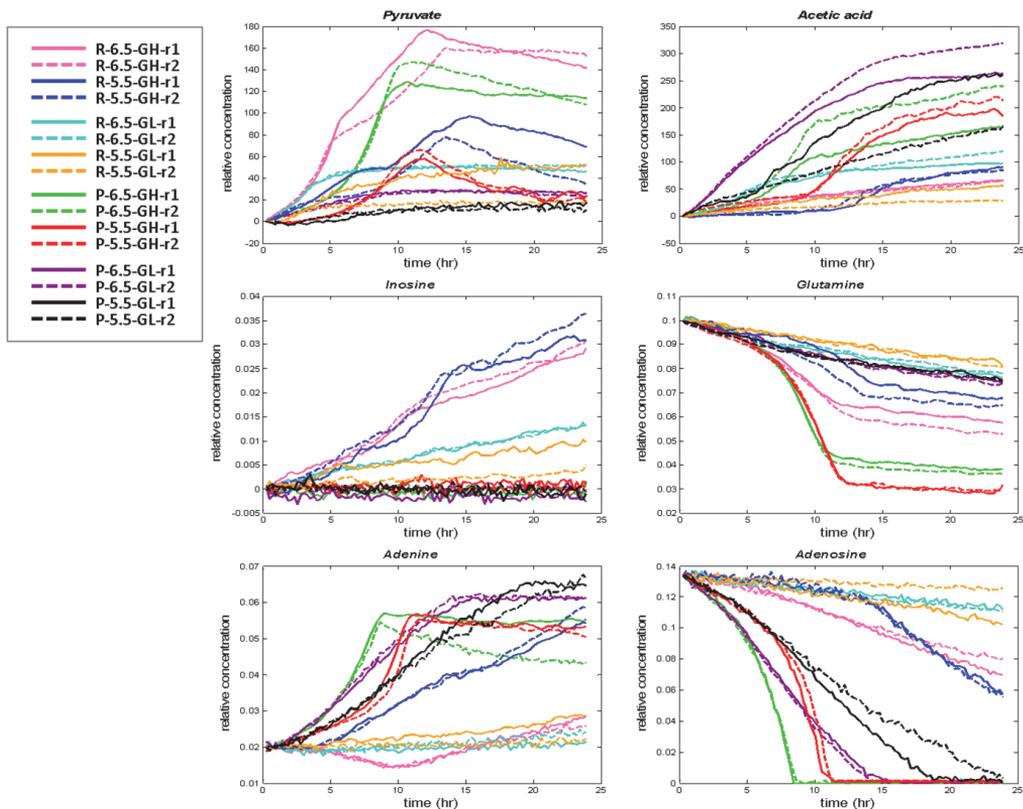
## 5.4. Outcome

This study focused on monitoring the fermentation process of two strains of LAB by *in vitro* NMR spectroscopy, which we call ‘micro-Process Analytical Technology (micro-PAT)’. In micro-PAT concept, ‘micro-’ is derived from the subjects of the study that are microbes and PAT refers to studying the fermentation of bacteria which is a biological process. NMR can serve micro-PAT due to its non-destructive nature that allows real-time investigation of the fermentation process. Therefore, an analytical protocol was developed for the *in vitro* measurements of the fermentation process of LAB and for the processing and analysis of the recorded time-series data.

The developed protocol for the *in vitro* study of bacterial fermentation and metabolism includes guidelines from the sample preparation to the kinetic modelling of metabolic changes. The application of the protocol for the described experimental design was helpful in highlighting some of the metabolic differences between the samples and investigating how the different design factors affect the metabolism. *In vitro* NMR measurements of bacterial fermentations is very useful in obtaining a metabolic overview of the investigated microorganisms, including biopreservation studies where microorganisms like lactic acid bacteria are used to preserve food. The analytical protocol has the potential to be very useful in the microbiology field, and can also be widely used by food industry for the purpose of strain optimization in biopreservation studies, as well as in the fermentation technology. The new method facilitates investigating the effect of different environmental factors on the metabolism of bacteria, in a relatively short time frame.

One of the key points in the protocol was the application of reference deconvolution for enhancing the quality of the time-series NMR spectra. As the first application of reference deconvolution to metabolomic NMR data, it proved to be a necessary and elegant solution to the problem of the inherent inhomogeneity of the samples that is encountered in *in vitro* NMR measurements of cells. When number of the cells and the inhomogeneity of the sample increase over time and give rise to the time-dependent line broadening of the resonances, application of reference deconvolution is very advantageous. Moreover, in studies that have an experimental design with factors that can affect the quality of spectra, using reference deconvolution prior to any multivariate analysis technique can improve the analysis results by eliminating the undesirable sources of variation. In the experimental design of this study for instance, two pH values, 5.5 and 6.5, were used which led to slightly different spectral quality in the samples with the two different pH values; for samples with pH 5.5 the resonances in the spectra were slightly broader than pH 6.5 samples. If the data was analyzed by multivariate analysis techniques with no prior correction or enhancement of the lineshapes, the results would be influenced by the fact that the resonances are broader in pH 5.5 samples and the metabolic differences would not be described accurately. However, after the application of reference deconvolution, the lineshapes and the linewidths in each time-series and also between the different time-series are more consistent, and the results of applying different multivariate analysis techniques, like PCA, MCR, PLS, etc., that are based on the assumption of the bilinearity of the data, will be more reliable. Therefore, we recommend that reference deconvolution should be considered as a standard tool to enhance lineshapes and improve multivariate analysis results in *in vivo* and *in vitro* NMR studies, as well as studies in which the quality of spectra can suffer from the inhomogeneity of the sample or the magnetic field.

Some of the calculated metabolic profiles from the time-series NMR spectra of the designed experiments are shown in Figure 5.13. For more detail on the experimental design and the calculation of the profiles, the reader is referred to Paper II (Ebrahimi, Larsen et al. 2015). The profiles can show the differences between the metabolism of the two investigated strains, as well as the difference in their response to the change in pH and carbohydrate source concentration. Such profiles will be the final output of the protocol.



**Figure 5.13.** Some of the calculated metabolic profiles for the real-time *in vitro* NMR study of the samples from the described experimental design.

It is expected that the developed protocol, that benefits from the advantages of *in vitro* NMR as well as the spectral quality enhancement by reference deconvolution will be used for the investigation of the metabolism of target protective cultures. Ultimately, the results of the *in vitro* experiments will have to be held against reference measurements for the antimicrobial and biopreservation performance. Then, PCLUSTER, the software that is developed for the quantification of mold growth and inhibition, can be used to record inhibition assays against pathogenic bacteria or molds at the end of the *in vitro* NMR study (see Chapter 6). This will allow for the identification of antimicrobial metabolites and quantify how different metabolites affect inhibition. Moreover, in order to get a more comprehensive illustration of the metabolism, in parallel with the *in vitro* NMR studies, similar batch fermentations can be performed, followed by regular sampling and chromatographic fractionation. The fractions can be subsequently up-concentrated by freeze-drying and then measured by NMR. Undoubtedly, the combination of the analytical NMR protocol, PCLUSTER, and adding the insight from the chromatographic fractionation, can reveal new knowledge and insight into the metabolism of bacteria.



## 6. Quantifying mold growth by multispectral imaging

In this chapter, the importance of quantifying mold growth, the approach that was developed for quantifying mold growth based on image analysis, and its application to two metabolomics studies are discussed. The new method is published in Paper III (Ebrahimi, van den Berg et al. 2015) and its applications for quantifying mold growth in two other studies are presented in Paper IV (Aunbjerg, Honoré et al. 2015), and Paper V (Honoré, Aunbjerg et al. 2015).

### 6.1. Quantifying mold growth

In food microbiology as well as in other fields that investigate mold and bacterial metabolism, it is of great relevance to quantify mold growth, as this can provide a measurement tool to investigate how different parameters influence mold growth. For instance, in studies related to biopreservation of food products, where safe and controlled microorganisms are used to inhibit growth of disease-causing molds and bacteria (Chaillou, Champomier-Vergès et al. 2005), a reliable mold quantification method can be most helpful. Such a method will allow optimizing biopreservation by investigating how mold growth is affected by different environmental factors and also different strains of protective cultures. This can be used by food industry for providing an efficient biopreservation system for food systems. However, predictive modeling of mold growth has been hindered by the lack of effective and reliable quantitative methods (Marín, Ramos et al. 2005, Marín, Cuevas et al. 2008, Ebrahimi, van den Berg et al. 2015).

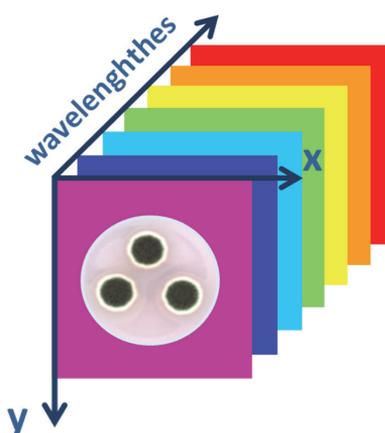
A number of approaches have been reported and used for quantifying mold growth, including colony forming units (CFU) counts, total ergosterol content and colony diameter (Marín, Ramos et al. 2005). CFU is one of the most frequently used methods for quantifying mold growth, but it reflects spore numbers rather than biomass and is, in general, a poor indicator of the extent of fungal growth. Measuring ergosterol content has also been used for mold growth quantification. Ergosterol is the dominant sterol in most fungi, and its concentration accounts for the total fungal population in a food sample (Taniwaki, Pitt et al. 2006). Although ergosterol has shown good performance as a fungal growth indicator for different species (Marín, Cuevas et al. 2008), it is not possible to determine ergosterol concentration accurately for very small colonies (e.g. a colony as small as 2 mm in diameter). By far, the simplest method to assess mold growth is measuring the colony diameter (or area). As molds often

grow in the form of surface colonies, colony diameters can be measured on Petri dishes over time and converted into growth curves (Taniwaki, Pitt et al. 2006). Colony diameter measurements show higher repeatability and sensitivity compared to ergosterol measurements (Marín, Morales et al. 2006), and diameters of very small colonies – for which ergosterol content cannot be measured accurately – can be easily determined. Although colony diameter does not take colony density and volume into account, it is the most suitable measure of the fungal biomass in solid substrates (Garcia, Ramos et al. 2010). Good correlation has been reported between ergosterol content and colony diameter (Marín, Morales et al. 2006). Colony diameter and size measurement is also non-destructive, and therefore, saves the sample for further analysis or time-series studies. In general, colony diameter is measured manually (Wang, Yan et al. 2012), and sometimes just a visual inspection of the colonies is used to estimate mold growth and grade inhibition (Magnusson and Schnürer 2001). For manual measurement of colony area, the routine practice is to measure the diameters of the mold in the two main perpendicular directions and estimate the area or to overlay tracing paper on the mold colony, trace the shape, and then overlay the tracing paper on graph paper and count the squares. These procedures obviously lack accuracy and precision and they can be even less reliable when colonies have not grown in well-shaped circular forms. In addition, they can disturb the mold, and spread the spores around which will bias the results by increasing the apparent growth. Moreover, if the investigated molds are toxic, manual measurement of the colonies can pose potential health risks to the analyst. Based on the presented arguments, the commonly used methods for quantifying mold growth are not accurate. An adequate, rapid and objective method will allow studying the effect of many different parameters and conditions on mold growth patterns, and can thus provide valuable insight and knowledge (Ebrahimi, van den Berg et al. 2015).

Paper III (Ebrahimi, van den Berg et al. 2015) presents a new and semi-automated approach for quantifying mold growth based on the colony size or area, using the unsupervised *k*-means clustering of multispectral images, recorded in the ultraviolet, visual and near-infrared regions. In order to test and demonstrate the efficiency of the new approach, three different sample sets were analyzed with the objective of quantifying white and green segments of *Penicillium* mold colonies. As white and green segments of the colonies relate to different stages of sporulation, their individual quantification can be informative. The results verified the efficiency of the proposed method for mold growth quantification. The new method and the related program, which is called 'PCLUSTER', were used in two other studies for quantifying mold growth. In the following, PCLUSTER as well as its applications are presented.

## 6.2. Multispectral imaging

Multispectral imaging combines spectroscopy and imaging, and thus provides both spectral and spatial information about the samples. Each multispectral image is a three-dimensional data structure in which two of the dimensions provide spatial information of the sample, and the third dimension represents spectral information for each picture-element (pixel). In multispectral images, each pixel is associated with a spectrum. The structure of a multispectral image is shown schematically in Figure 6.1. In multispectral images, spectral information is provided for a range of wavelengths, including ultraviolet, visible, and near-infrared which can provide much more information about the samples compared to e.g. the ordinary trichromatic (RGB) images. Multispectral images give information on the color, surface properties, water content and other important physical and chemical properties of the samples (Guo, Zeng et al. 2007, Dissing, Papadopoulou et al. 2013). This can be helpful in different fields from food quality control in industry to different biological research areas in academia (Ebrahimi, van den Berg et al. 2015).



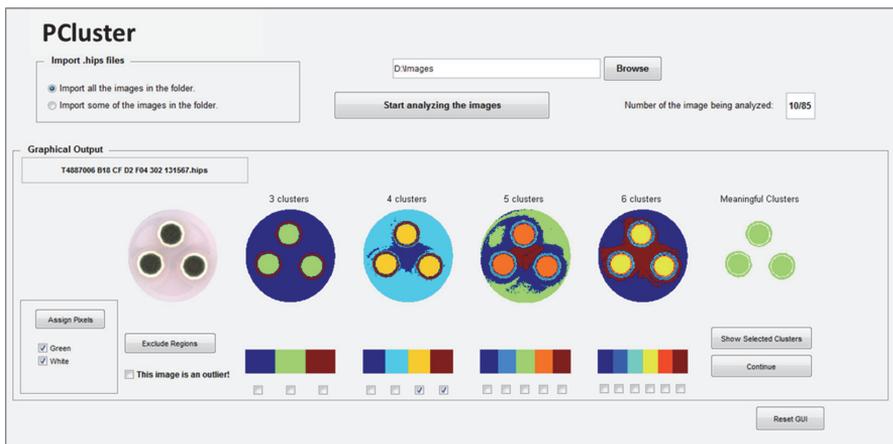
**Figure 6.1.** Schematic illustration of the structure of a multispectral image of mold colonies.  $X$  and  $y$  provide spatial information of the sample, and the wavelength dimension provides spectral information.

## 6.3. PCLUSTER software— a new approach for quantifying mold growth (Paper III)

A new approach for quantifying mold growth based on mold colonies area was developed. The method is based on clustering multispectral images by  $k$ -means, an unsupervised and simple clustering algorithm (MacQueen 1967, Tran, Wehrens et al. 2005), and provides an

accurate tool for measuring different segments of mold colonies. In order to perform the clustering and analysis of the images by the method in a semi-automated way, a Graphical User Interface (GUI) was developed using MATLAB 2012b (MathWorks, Inc., Natick, MA, USA). The outline of this software which is called ‘PCLUSTER’ is shown in Figure 6.2. PCLUSTER can be used for multispectral images that are recorded by VideometerLab2 instrument (Videometer A/S, Hørsholm, Denmark), which have ‘.hips’ format. In the following, a brief description on the instructions of the software is presented.

PCLUSTER can be started either by typing its name in MATLAB command window, or by clicking on its icon for the compiled version. Multispectral images are first imported into PCLUSTER, by selecting the import option of choice. Then, pressing the ‘Start analyzing the images’ icon will open the window for selecting a circular region of interest (ROI), which will be used for all the images in the imported set.

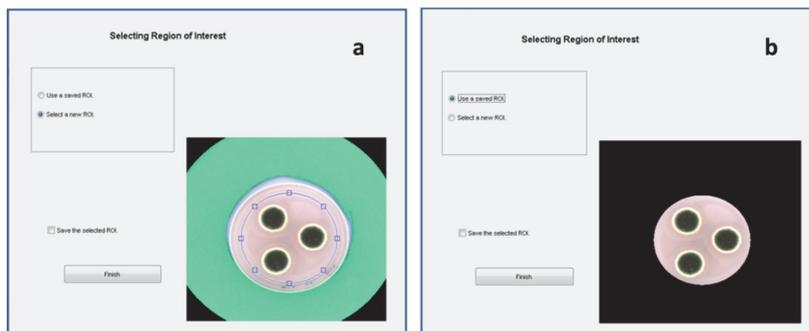


**Figure 6.2.** The layout of PCLUSTER, the software for quantifying mold colony sizes.

Choosing a ROI reduces the size of the images and accelerates the analysis. ROI selection window is shown in Figure 6.3. The software only uses the first image for this purpose, but if the images are recorded by putting the Petri dishes, or other objects of the study, in approximately the same position under the Videometer sphere during the images acquisition, the selected ROI can be safely used for all the images in a dataset. The selected ROI can also be saved to be used for other sets of images in future analysis runs.

In the  $k$ -means algorithm, the number of clusters,  $k$ , is a user defined input. In PCLUSTER, images are clustered from 3 to 6 groups ( $K=3:1:6$ ) and the results are shown as color-coded (so-called *false negative*) image objects which show membership of the pixels in the clusters (see Figure 6.2). Then, based on the graphical output, the optimum number of clustering and the meaningful clusters are selected by the user. Meaningful clusters are the ones that show

the colonies segments, and the optimum number of groups is the one that allows clustering different parts of the mold colony properly. For instance, in Figure 6.2, partitioning the pixels into 3 clusters is enough to segment the white and green parts of the mold colonies and separate them from the background. Selecting more than the optimum number of clusters will subdivide the colony segments further and can impose some errors on the quantification results, since only two colony segments can be chosen. The user can make a specific interpretation to the clusters, based on the color-coding of the pixels and select the meaningful clusters by ticking the corresponding check boxes (see Figure 6.2). PCLUSTER is specifically designed for *Penicillium* molds, for which the colonies are composed of white and green segments. However, the method and the explained concept can be applied to all types of molds.

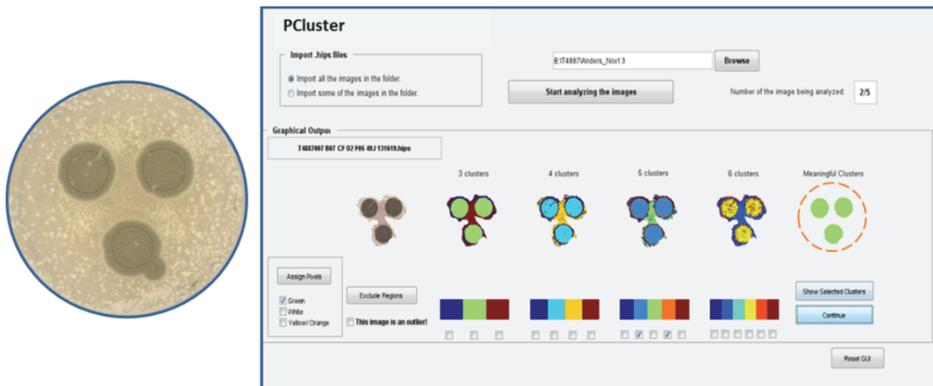


**Figure 6.3.** The window for selecting the region of interest (ROI) in PCLUSTER. a) selecting the region of interest, and b) the image reduced to the ROI.

Different built-in options are included in PCLUSTER. One of these options that can be very useful allows omitting the unwanted growth on some of the plates. An example of such a situation is shown in Figure 6.4a. The image shows a Petri dish with triple spotting of mold on top. As marked with the 'x' sign, the small growth close to one of the colonies is the result of the splashes of the mold solution while spotting on the Petri dishes. The 'Exclude regions' button in the main page of the software allows defining areas that the user does not want to be included in quantifying the image. Figure 6.4b shows the window that the user gets for excluding the regions and the different options that are implemented for this purpose. After this step, the software re-clusters the image, without the excluded regions. Figure 6.4c shows the main window of PCLUSTER for an image after re-clustering. This option allows saving some of the images while avoiding quantitative errors. Figure 6.5 shows an example of an image with considerable contamination of the Petri dish with unwanted microorganisms during incubation. Although the contamination had spread all over the dish, the defined option was able to include only the main colonies in the analysis.



**Figure 6.4.** PCLUSTER’s built-in option for excluding unwanted regions from the images. a) an example of an image with unwanted growth of mold marked with the white ‘x’ sign, b) the window that opens by ‘Exclude regions’ button, and c) re-clustering of the image after excluding unwanted regions.



**Figure 6.5.** An example of an image from a Petri dish with significant unwanted growth from contamination, and the re-clustering after excluding these regions by PCLUSTER. As shown by the red circle around the meaningful colonies, only the main mold colonies have been included in the analysis, and the growth from contamination has been excluded.

At the end of the analysis by PCLUSTER, the results are saved as Excel and Matlab files. The outputs are the size of the green and white segments of the mold colonies. So far, PCLUSTER has been used for analyzing few thousands of images from inhibition assays, and the results have verified the efficiency of the method for quantifying mold growth. Paper IV (Aunbjerg, Honoré et al. 2015) and Paper V (Honoré, Aunbjerg et al. 2015) include two applications of PCLUSTER, which are discussed in more detail in the following.

## 6.4. Applications of PCLUSTER to metabolomics studies

In this section, two applications of PCLUSTER for quantifying mold growth in metabolomics studies are presented. The first application successfully used PCLUSTER and the results of the quantitative image analysis to investigate the contribution of diacetyl to the antifungal effect of the studied strain of LAB. In the second application, PCLUSTER helped to quantify antifungal synergistic effects between six 2-hydroxy acid metabolites in three strains of LAB, by calculating correlations between the metabolic profiles of the bacteria and their inhibitory effects.

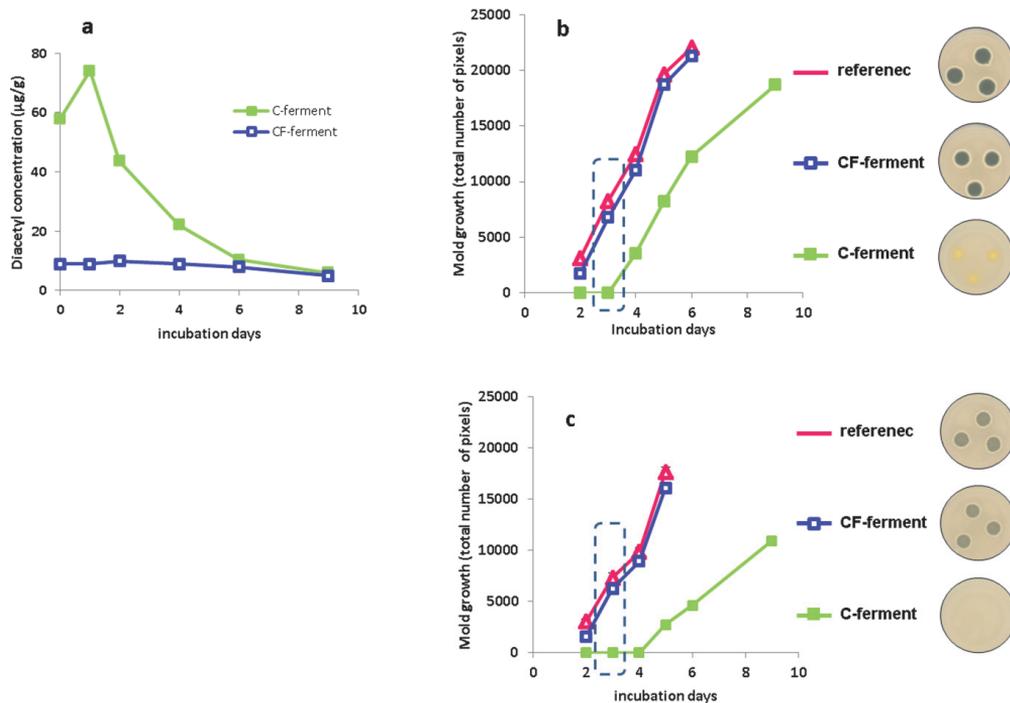
### 6.4.1. Contribution of volatiles to the antifungal effect of *Lactobacillus paracasei* in defined medium and yogurt (Paper IV)

The aim of the first study (Aunbjerg, Honoré et al. 2015) was to investigate which volatile metabolites are associated with the antifungal properties of *Lactobacillus paracasei*. Diacetyl was identified as the main volatile metabolite produced by *L. paracasei* DGCC 2132. As part of the study, inhibition assays were prepared for with-cell (C-ferments) and cell-free ferments (CF-ferments) of the strain, and reference plates with only the growth medium, using *Penicillium sp. nov.* DCS 1541, and *Penicillium solitum* DCS 302 as indicator molds. The plates were then incubated for 9 days, and multispectral images were recorded on 2, 3, 4, 5, 6 and 9 days after the incubation, and analyzed by PCLUSTER to quantify mold growth and inhibition.

In order to investigate how the inhibition of mold growth and concentration of diacetyl are correlated, solid agar plugs were taken from the plates of the prepared inhibition assay on days 0, 1, 2, 4, 6, and 9 after incubation, and diacetyl concentration was measured by headspace gas chromatography mass spectrometry (GC-MS). Figure 6.6 shows the profile for diacetyl concentration, and the total size of the mold colonies (combination of the green and white segments) for C-ferment, CF-ferment, and reference samples as resulted from PCLUSTER. The images shown beside the plots are from the Petri dishes after 4 days of incubation.

In the C-ferments, concentration of diacetyl reaches its maximum after 1 day of incubation, and decreases afterwards, whereas in the CF-ferments, diacetyl concentration stays constant in all days. The plots (Figure 6.6) of the mold colony size show how these changes affect mold growth. For CF-ferments, almost no inhibition is observed and sizes of the colonies and the shape of their plots are similar to those in reference samples, for both of the indicator molds. For C-ferments, it is observed that the growth of the colonies is inhibited by the presence of diacetyl. The high concentration of diacetyl in the first few days after incubation has delayed the growth of the indicator molds, but the colonies grow larger as the concentration of diacetyl decreases. The plots also show that diacetyl inhibits *Penicillium sp. nov.* DCS 1541 more than *Penicillium solitum* DCS 302.

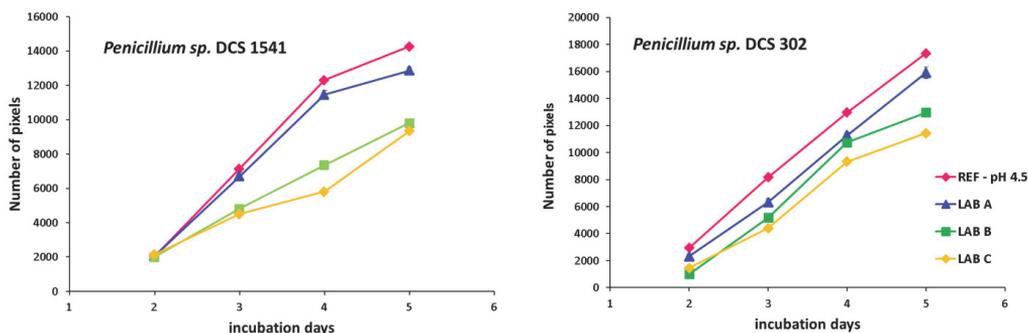
In conclusion, PCLUSTER was successfully used in this study for investigating how concentration of diacetyl affects inhibition of the two mentioned indicator molds by *L. paracasei* DGCC 2132 in the inhibition assays. The results also allow comparing the degree of the inhibition of the two molds by the different samples. Such quantitative information on mold inhibition assays cannot be reliably obtained without having an accurate tool for mold growth measurement. For instance, the common practice of the manual measurement of the size of mold colonies cannot be relied on, as manual measurements lack precision and accuracy, and are prone to subjective errors.



**Figure 6.6.** Plots and profiles for investigating the effect of diacetyl concentration on mold growth. a) concentration of diacetyl in the agar plugs from the inhibition samples, b) the growth of the *Penicillium solitum* DCS 302, and c) the growth of *Penicillium sp. nov.* DCS 1541, for C-ferments, CF-ferments, and reference samples. The images are from day 4 that is marked on the graphs by the dotted boxes.

#### 6.4.2. Metabolic footprinting for investigation of antifungal properties of *Lactobacillus paracasei* (Paper V)

The aim of the second study (Honoré, Aunbjerg et al. 2015) was to investigate the antifungal properties of three strains of *Lactobacillus paracasei*, by exometabolic profiling of the strains and find potential antifungal compounds. Multispectral imaging was used to record inhibition assays of the ferments from the three strains of LAB, *Lactobacillus paracasei* DGCC 2132, *Lactobacillus paracasei* DGCC 11287, and *Lactobacillus paracasei* DGCC 695, which were abbreviated as LAB A, B, and C respectively. Cell-free ferments of the strains were tested for the antifungal activity against two indicator molds, *Penicillium solitum* DCS 302, and *Penicillium sp. nov.* DCS 1541. The inhibition assays were incubated for five days, and multispectral images were recorded from day 2 to 5, on a daily basis. PCLUSTER was used to quantify white and green segments of the mold colonies. Figure 6.7 shows the analysis results for the three strains, and also reference samples which do not have any of the strains and are acidified to a pH value of 4.5.



**Figure 6.7.** The sizes of the mold colonies in the inhibition assays, as calculated by PCLUSTER, for the two indicator molds. a) *Penicillium sp. DCS302*, and b) *Penicillium sp. DCS1541*. REF plates are uninoculated medium acidified to pH 4.5, and LAB A, B and C are the cell free ferments of the bacteria.

Exometabolic profiles of the strains were obtained by using headspace solid-phase micro-extraction gas chromatography mass spectrometry (HS SPME-GC/MS) and Liquid chromatography–mass spectrometry (LC-MS). In order to investigate how different metabolites correlate with the inhibition properties of the strains, Partial Least Squares Regression (PLS) (Wold, Ruhe et al. 1984) was used to model the correlations between the profiles and the size of the white segments, green segments, as well as the total size of the colonies in the assays, for both of the indicator molds. Based on the PLS models, in all days there was a correlation between the growth of the indicator organism and the exometabolomic profiles. The best correlation was observed by using the green segments or the total colony size (the size of the white pixels did not change considerably after sporulation). Therefore, the following discussions are related to the PLS models that were calculated between the metabolic profiles as the X-block, and the total colony size as the Y-block.

Based on the PLS models, the consumption of nutrients was correlated to reduced mold growth. Concentration of glucose, phenylalanine, leucine, isoleucine, and adenosine, showed a high positive correlation with the mold growth. However, some of the metabolites also showed high positive correlations with reduced mold growth. Among the major metabolites that showed inhibition effect against *Penicillium sp. DCS 1541* were six 2-hydroxy acids, as listed in Table 6.1.

In conclusion, the developed PCLUSTER approach for quantifying mold growth served as an efficient tool in this study for quantifying inhibition assays, and thus helped in finding metabolites that affect the inhibition. Therefore, PCLUSTER has great potential for use in bioassay guided fractionation of protective cultures fermentation, for finding antifungal compounds and metabolites that are responsible for antimicrobial effect.

**Table 6.1.** The list of the identified 2-hydroxy acids that correlated negatively with mold growth.

<b>Compound</b>	<b>Abbreviation</b>
2- hydroxy-4- methyl pentanoic acid	OH-Me-Pe
2- hydroxy-3- phenyl propanoic acid	OH-Phe-Pr
2-hydroxy-3-(4-hydroxyphenyl)propanoic acid	OH-(OH-Phe)-Pr
2-hydroxy-3-(1H-indol-3-yl)propanoic acid	OH-Ind-Pr
2-hydroxy-3-methylbutanoic acid	OH-Me-Bu
2-hydroxy-4-(methylthio)butanoic acid	OH-MeS-Bu



## 7. Conclusions

The main aim of the present PhD project was to use NMR spectroscopy to study the metabolism of the antifungal strains of LAB. In order to achieve this, an analytical protocol was developed for real-time *in vitro* NMR measurements of bacterial fermentation and metabolism, from sample preparation to the kinetic modelling of the metabolic changes. As a non-destructive analytical technique, NMR provides a unique opportunity for studying the overall quantitative metabolism of bacteria. Bacterial fermentation can be performed in the NMR magnet and time-series spectra can thus be recorded *in vitro*. For *in vitro* NMR, no sampling and no quenching of the metabolism is required and the metabolic process is not disturbed. The developed analytical protocol was used for the measurements of an experimental design of two strains of LAB and the results, as presented in Paper II (Ebrahimi, Larsen et al. 2015), demonstrated significant metabolic differences between the strains and between the other design factors. The recommended sample preparation procedure is relatively easy and allows the investigation of different strains and experimental factors. The data analysis and kinetic modeling approaches that were used provide a good guideline for handling similar data and for extracting the correct metabolic profiles. The combined protocol shows the challenges that can be expected in the analysis of *in vitro* NMR data, and how different processing techniques can be applied to make the extracted profiles more reliable.

One of the interesting NMR data processing methods that proved to be very beneficial in metabolomics, and especially in *in vitro* studies, was reference deconvolution. Reference deconvolution is a post-measurement NMR data processing method that can correct many systematic errors in NMR spectra. For the designed artificial metabolic experiments which are presented in Paper I (Ebrahimi, Nilsson et al. 2014), it was shown that reference deconvolution can be helpful in improving the multivariate analysis results of NMR data, by enhancing the lineshapes and bilinearity of the signals. PCA and PLS models were significantly improved when applied to reference deconvoluted data compared to the data that was not reference deconvoluted. The multivariate models became simpler, and fewer latent variables were required to explain the data. This makes reference deconvolution very useful in quantitative NMR spectroscopy and in applications where quantitative pattern recognition of NMR data is of interest. In Paper II (Ebrahimi, Larsen et al. 2015), reference deconvolution was used for the first time to improve the multivariate analysis results of real metabolomics data. In this case, reference deconvolution proved to be an elegant solution to the problem of the inherent inhomogeneity of the samples that is encountered in *in vitro* NMR measurements of cells and can lead to the broadening of the signals. In this study, reference deconvolution

improved MCR-ALS models that were used for extracting the metabolic profiles, by enhancing the bilinearity of the data. Based on the results that were presented in Paper I (Ebrahimi, Nilsson et al. 2014) and Paper II (Ebrahimi, Larsen et al. 2015), it is recommended that reference deconvolution should be considered as a standard tool to enhance lineshapes and improve multivariate analysis results of NMR data.

Quantifying mold growth is of great relevance and interest in microbiology. Especially in studies related to the biopreservation of food products, the quantification of mold growth and inhibition in a reproducible and quantitative manner is highly desired, in order to compare the efficiency of the different strains of bacteria in preserving food, and to investigate the effect of different factors on inhibition. According to the literature, modeling of fungal growth has been hampered by the lack of an appropriate and accurate method for quantification (Marín, Ramos et al. 2005). Due to the importance of quantifying inhibition before actually studying the metabolism of the antifungal strains of LAB by NMR and other metabolomics platforms, one of the goals of the present PhD project was to develop an accurate method for quantifying mold growth. Therefore, a new approach was presented for quantifying mold growth, based on multispectral images and their clustering of by *k*-means algorithm. The method was further developed into a piece of software called PCLUSTER that allows measuring different segments of mold colonies in a semi-automated way. The only input that is required from the user is selecting the meaningful clusters. In Paper III (Ebrahimi, van den Berg et al. 2015), the theory of the method is presented and three different sets of samples were successfully analyzed by the method, with the objective of quantifying mold growth and the size of the colony segments of *Penicillium* mold. PCLUSTER was subsequently applied to two biopreservation related metabolomic studies of lactic acid bacteria. In Paper IV (Aunbjerg, Honoré et al. 2015) PCLUSTER was used to analyze multispectral images from the inhibition assays and was able to quantify how diacetyl affects inhibition of the indicator molds. In Paper V (Honoré, Aunbjerg et al. 2015) PCLUSTER served as an efficient tool for quantifying inhibition assays and finding metabolites that affect the inhibition. The results of the applications of PCLUSTER verified that the method has great potential for use in bioassay guided fractionation of protective cultures fermentation, for finding antifungal compounds and metabolites that contribute to the antimicrobial effects, and for strain optimization and selection to achieve enhanced biopreservation.

The project has been successful in providing analytical protocols regarding data acquisition and data analysis for *in vitro* NMR studies of bacterial metabolism and fermentation. In addition, the developed approach for quantifying inhibition and mold growth provides an efficient screening tool for the optimization of biopreservation. The combination of these new analytical techniques can be most helpful in future biopreservation studies. However, the developed tools are generic and can be used in many other contexts.

## 8. Outreach

Food spoilage is a global challenge. It causes considerable economic losses in the food industry and for the consumers. It can threaten public health, and can ruin a sustainable food production by wasting the resources and energy that have been used for its production. These concerns are very serious and warrant increased research on the preservation of food products. Biopreservation is in particular interesting, due to the advantages that it offers over the commonly used methods of food preservation, and it is expected that the research in this field will continue to increase. In the long term, biopreservation may be able to completely replace the chemical preservatives, which will be a significant improvement in the quality and healthiness of food products. Considerable effort has already been devoted to this field, but biopreservation often remains elusive and multifactorial and will require considerable research efforts in the future.

Introducing new and efficient methods for studying food protective cultures will thus be very helpful in expanding our knowledge in biopreservation. By gaining more knowledge in this field, efficient biopreservation systems can be designed and spread to commonly used food products. The analytical protocol for *in vitro* NMR measurements that was developed in this project, and the method for quantifying inhibition, PCLUSTER, are new tools which can be very beneficial for the research in biopreservation field. The *in vitro* NMR protocol will allow a reliable means for investigating the metabolism of protective cultures, the synergistic effects between bacterial strains or metabolites, and the effect of the different environmental factors on the metabolism. On the other hand, PCLUSTER provides an interactive tool for identification and quantification of the efficacy of new antifungal compounds, strain selection, and substrate optimization towards more efficient biopreservation solutions for different food systems.

In order to obtain more information about the metabolism of the investigated bacteria and to partially compensate for the relative low sensitivity of NMR in the future studies, other metabolomics analytical techniques can also be used in parallel to the suggested methods. For instance, chromatographic fractionation can be used alongside or even in hyphenation. Parallel fermentations can be performed, one in the magnet of spectrometer for the purpose of real-time *in vitro* measurements, and one in a bioreactor that allows monitoring different parameters like pH, and can be sampled in specified time intervals. The fractions should be first quenched by corresponding protocols to stop the metabolism quickly and efficiently, whereafter the metabolome can be up-concentrated by freeze drying, fractionated by a

chromatography column, and measured by NMR or mass spectrometry (MS). The fractions can also be directly introduced to hyphenated analytical platforms such as LC/GC-MS or LC-NMR. The fractionation technique obviously does not have the real-time advantage that the *in vitro* approach can offer, but can provide valuable quantitative information on the low-concentration metabolites that cannot be measured by NMR. This will be a powerful combination of analytical methods for investigating the detailed metabolism of selected strains of bacteria in a more targeted and detailed manner while still keeping the metabolic overview by *in vitro* NMR.

# References

- Annou S, M. M., Martínez-Bueno M, Valdivia E . (2007). "Biopreservation, an ecological approach to improve the safety and shelf-life of foods." *Applied Microbiology* 1: 475–486.
- Aunbjerg, S., A. Honoré, J. Marcussen, P. Ebrahimi, F. Vogensen, C. Benfeldt, T. Skov and S. Knøchel (2015). "Contribution of volatiles to the antifungal effect of *Lactobacillus paracasei* in defined medium and yogurt." *International Journal of Food Microbiology* 194: 46-53.
- Baliarda, A., H. Robert, M. Jebbar, C. Blanco, A. Deschamps and C. Le Marrec (2003). "Potential osmoprotectants for the lactic acid bacteria *Pediococcus pentosaceus* and *Tetragenococcus halophila*." *International Journal of Food Microbiology* 84: 13-20.
- Barjat, H., G. A. Morris, A. G. Swanson, S. Smart and S. C. Williams (1995). "Reference deconvolution using multiplet reference signals." *Journal of Magnetic Resonance, Series A* 116: 206-214.
- Behrends, V., T. M. Ebbels, H. D. Williams and J. G. Bundy (2009). "Time-resolved metabolic footprinting for nonlinear modeling of bacterial substrate utilization." *Applied and Environmental Microbiology* 75: 2453-2463.
- Boroujerdi, A. F. B., M. I. Vizcaino, A. Meyers, E. C. Pollock, S. L. Huynh, T. B. Schock, P. J. Morris and D. W. Bearden (2009). "NMR-Based Microbial Metabolomics and the Temperature-Dependent Coral Pathogen *Vibrio corallilyticus*." *Environmental Science & Technology* 43: 7658-7664.
- Brosnan, B., A. Coffey, E. K. Arendt and A. Furey (2014). "A comprehensive investigation into sample extraction and method validation for the identification of antifungal compounds produced by lactic acid bacteria using HPLC-UV/DAD." *Analytical Methods* 6: 5331-5344.
- Buchholz, A., J. Hurlbaeus, C. Wandrey and R. Takors (2002). "Metabolomics: quantification of intracellular metabolite dynamics." *Biomolecular Engineering* 19: 5-15.
- Calo-Mata, P., S. Arlindo, K. Boehme, T. de Miguel, A. Pascoal and J. Barros-Velazquez (2008). "Current Applications and Future Trends of Lactic Acid Bacteria and their Bacteriocins for the Biopreservation of Aquatic Food Products." *Food and Bioprocess Technology* 1: 43-63.
- Caspi, R., T. Altman, R. Billington, K. Dreher, H. Foerster, C. A. Fulcher, T. A. Holland, I. M. Keseler, A. Kothari, A. Kubo, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, D. Weerasinghe, P. Zhang and P. D. Karp (2014). "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases." *Nucleic Acids Research* 42: D459-D471.
- Castellano, P. and G. Vignolo (2006). "Inhibition of *Listeria innocua* and *Brochothrix thermosphacta* in vacuum-packaged meat by addition of bacteriocinogenic *Lactobacillus curvatus* CRL705 and its bacteriocins." *Letters in Applied Microbiology* 43: 194-199.
- Cevallos-Cevallos, J. M., J. I. Reyes-De-Corcuera, E. Etxeberria, M. D. Danyluk and G. E. Rodrick (2009). "Metabolomic analysis in food science: a review." *Trends in Food Science & Technology* 20: 557-566.
- Chaillou, S., M.-C. Champomier-Vergès, M. Cornet, A.-M. Crutz-Le Coq, A.-M. Dudez, V. Martin, S. Beaufils, E. Darbon-Rongère, R. Bossy and V. Loux (2005). "The complete genome sequence of

- the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K." *Nature biotechnology* 23: 1527-1533.
- Chassagnole, C., N. Noisommit-Rizzi, J. W. Schmid, K. Mauch and M. Reuss (2002). "Dynamic modeling of the central carbon metabolism of *Escherichia coli*." *Biotechnology and bioengineering* 79: 53-73.
- Claridge, T. D. (2008). High-resolution NMR techniques in organic chemistry, 2<sup>nd</sup> edition, *Tetrahedron Organic Chemistry Series*, Vol 27, Newnes, Elsevier.
- Combourieu, B., L. Elfoul, A.-M. Delort and S. Rabot (2001). "Identification of new derivatives of sinigrin and glucotropaeolin produced by the human digestive microflora using <sup>1</sup>H NMR spectroscopy analysis of *in vitro* incubations." *Drug metabolism and Disposition* 29: 1440-1445.
- Cooley, J. W. and J. W. Tukey (1965). "An algorithm for the machine calculation of complex Fourier series." *Mathematics of Computation* 19: 297-301.
- Dauner, M., J. E. Bailey and U. Sauer (2001). "Metabolic flux analysis with a comprehensive isotopomer model in *Bacillus subtilis*." *Biotechnology and Bioengineering* 76: 144-156.
- Dauner, M. and U. Sauer (2001). "Stoichiometric growth model for riboflavin-producing *Bacillus subtilis*." *Biotechnology and Bioengineering* 76: 132-143.
- De Angelis, M. and M. Gobetti (2004). "Environmental stress responses in *Lactobacillus*: a review." *Proteomics* 4: 106-122.
- De Graaf, A., J. Van Dijk and W. BoëE (1990). "QUALITY: quantification improvement by converting lineshapes to the Lorentzian type." *Magnetic Resonance in Medicine* 13: 343-357.
- De Graaf, R. A. (2008). *In vivo* NMR spectroscopy: principles and techniques, 2<sup>nd</sup> Edition, Wiley.
- de Juan, A., J. Jaumot and R. Tauler (2014). "Multivariate Curve Resolution (MCR). Solving the mixture analysis problem." *Analytical Methods* 6: 4964-4976.
- De Lathauwer, L., B. De Moor, J. Vandewalle and B. S. S. by Higher-Order (1994). Singular Value Decomposition. Proc. EUSIPCO-94, Edinburgh, Scotland, UK.
- Deborde, C. and P. Boyaval (2000). "Interactions between pyruvate and lactate metabolism in *Propionibacterium freudenreichii* subsp. *shermanii*: *in vivo* <sup>13</sup>C nuclear magnetic resonance studies." *Applied and Environmental microbiology* 66: 2012-2020.
- Deborde, C., D. B. Rolin and P. Boyaval (1999). "*In Vivo*<sup>13</sup>C NMR Study of the Bidirectional Reactions of the Wood–Werkman Cycle and around the Pyruvate Node in *Propionibacterium freudenreichii* subsp. *shermanii* and *Propionibacterium acidipropionici*." *Metabolic Engineering* 1: 309-319.
- Delort, A.-M., G. Gaudet and E. Forano (2002). "<sup>23</sup>Na NMR Study of *Fibrobacter succinogenes* S85: Comparison of Three Chemical Shift Reagents and Calculation of Sodium Concentration Using Ionophores." *Analytical Biochemistry* 306: 171-180.
- Delort, A.-M., G. Gaudet and E. Forano (2004). The use of chemical shift reagents and <sup>23</sup>Na NMR to study sodium gradients in microorganisms. *Environmental Microbiology*, Springer: 389-405.
- Dissing, B. S., O. S. Papadopoulou, C. Tassou, B. K. Ersbøll, J. M. Carstensen, E. Z. Panagou and G.-J. Nychas (2013). "Using multispectral imaging for spoilage detection of pork meat." *Food and Bioprocess Technology* 6: 2268-2279.

- Doyle, M. E. (2007). Microbial food spoilage, losses and control strategies; A brief review of the literature. Dissertation, University of Wisconsin, Madison.
- Eakin, R., L. Morgan, C. Gregg and N. Matwiyoff (1972). "Carbon-13 nuclear magnetic resonance spectroscopy of living cells and their metabolism of a specifically labeled  $^{13}\text{C}$  substrate." *FEBS Letters* 28: 259-264.
- Ebrahimi, P., F. H. Larsen, H. M. Jensen, F. K. Vogensen and S. B. Engelsen (2015). "Real-time metabolomic analysis of lactic acid bacteria as monitored by *in vitro* NMR and chemometrics." *Metabolomics*, submitted.
- Ebrahimi, P., M. Nilsson, G. A. Morris, H. M. Jensen and S. B. Engelsen (2014). "Cleaning up NMR spectra with reference deconvolution for improving multivariate analysis of complex mixture spectra." *Journal of Chemometrics* 28: 656-662.
- Ebrahimi, P., F. van den Berg, S. D. Aunbjerg, A. Honoré, C. Benfeldt, H. M. Jensen and S. B. Engelsen (2015). "Quantitative determination of mold growth and inhibition by multispectral imaging." *Food Control* 55: 82-89.
- Emwas, A. H. M., R. M. Salek, J. L. Griffin and J. Merzaban (2013). "NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations." *Metabolomics* 9: 1048-1072.
- Engelsen, S. B., F. Savorani and M. A. Rasmussen (2013). Chemometric Exploration of Quantitative NMR Data. *eMagRes*, Vol 2: 267-278.
- Farnworth, E. R. T. (2008). Handbook of fermented functional foods, CRC press.
- Fernández, M. and M. Zúñiga (2006). "Amino acid catabolic pathways of lactic acid bacteria." *Critical Reviews in Microbiology* 32: 155-183.
- Fiehn, O. (2001). "Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks." *Comparative and Functional Genomics* 2: 155-168.
- Fiehn, O. (2002). "Metabolomics—the link between genotypes and phenotypes." *Plant Molecular Biology* 48: 155-171.
- Food and Drug Administration (FDA) (2004). "Guidance for industry PAT—a framework for innovative pharmaceutical development, manufacturing, and quality assurance." DHHS, Rockville, Maryland.
- Friebolin, H. (2005). Basic One- and Two-Dimensional NMR Spectroscopy, Wiley.
- Fugelsang, K. C. and C. G. Edwards (2006). Wine microbiology: practical applications and procedures, Springer Science & Business Media.
- Fukuda, S., Y. Nakanishi, E. Chikayama, H. Ohno, T. Hino and J. Kikuchi (2009). "Evaluation and characterization of bacterial metabolic dynamics with a novel profiling technique, real-time metabolotyping." *PLoS one* 4: e4893.
- Galvez, A., H. Abriouel, R. L. Lopez and N. Ben Omar (2007). "Bacteriocin-based strategies for food biopreservation." *International Journal of Food Microbiology* 120: 51-70.
- Garcia, D., A. J. Ramos, V. Sanchis and S. Marín (2010). "Modelling mould growth under suboptimal environmental conditions and inoculum size." *Food Microbiology* 27: 909-917.

- Garmiene, G., J. Salomskiene, I. Jasutiene, I. Macioniene and I. Miliauskiene (2010). "Production of benzoic acid by lactic acid bacteria from *Lactobacillus*, *Lactococcus* and *Streptococcus* genera in milk." *Milchwissenschaft* 65: 295-298.
- Ghanbari, M., M. Jami, K. J. Domig and W. Kneifel (2013). "Seafood biopreservation by lactic acid bacteria - A review." *LWT-Food Science and Technology* 54: 315-324.
- Gibney, M. J., M. Walsh, L. Brennan, H. M. Roche, B. German and B. Van Ommen (2005). "Metabolomics in human nutrition: opportunities and challenges." *The American Journal of Clinical Nutrition* 82: 497-503.
- Glaasker, E., W. N. Konings and B. Poolman (1996). "Osmotic regulation of intracellular solute pools in *Lactobacillus plantarum*." *Journal of Bacteriology* 178: 575-582.
- Gonzalez, B., A. de Graaf, M. Renaud and H. Sahn (2000). "Dynamic *in vivo* <sup>31</sup>P nuclear magnetic resonance study of *Saccharomyces cerevisiae* in glucose-limited chemostat culture during the aerobic-anaerobic shift." *Yeast* 16: 483-497.
- Govindaraju, V., K. Young and A. A. Maudsley (2000). "Proton NMR chemical shifts and coupling constants for brain metabolites." *NMR in Biomedicine* 13: 129-153.
- Grivet, J. P. and A. M. Delort (2009). "NMR for microbiology: *in vivo* and *in situ* applications." *Progress in Nuclear Magnetic Resonance Spectroscopy* 54: 1-53.
- Grivet, J.-P., A.-M. Delort and J.-C. Portais (2003). "NMR and microbiology: from physiology to metabolomics." *Biochimie* 85: 823-840.
- Guo, N., L. Zeng and Q. Wu (2007). "A method based on multispectral imaging technique for white blood cell segmentation." *Computers in Biology and Medicine* 37: 70-76.
- Hartbrich, A., G. Schmitz, D. Weuster-Botz, A. De Graaf and C. Wandrey (1996). "Development and application of a membrane cyclone reactor for *in vivo* NMR spectroscopy with high microbial cell densities." *Biotechnology and Bioengineering* 51: 624-635.
- Hazan, R., A. Levine and H. Abeliovich (2004). "Benzoic acid, a weak organic acid food preservative, exerts specific effects on intracellular membrane trafficking pathways in *Saccharomyces cerevisiae*." *Applied and Environmental Microbiology* 70: 4449-4457.
- Hecker, M., W. Schumann and U. Völker (1996). "Heat-shock and general stress response in *Bacillus subtilis*." *Molecular Microbiology* 19: 417-428.
- Helinck, S., D. Le Bars, D. Moreau and M. Yvon (2004). "Ability of thermophilic lactic acid bacteria to produce aroma compounds from amino acids." *Applied and Environmental Microbiology* 70: 3855-3861.
- Hols, P., C. Defrenne, T. Ferain, S. Derzelle, B. Delplace and J. Delcour (1997). "The alanine racemase gene is essential for growth of *Lactobacillus plantarum*." *Journal of Bacteriology* 179: 3804-3807.
- Honoré, A., S. D. Aunbjerg, P. Ebrahimi, M. Thorsen, C. Benfeldt, S. Knøchel and T. Skov (2015). "Metabolic Footprinting for Investigation of antifungal properties of *Lactobacillus paracasei*." *Journal of Analytical & Bioanalytical Chemistry*, submitted.
- Hotelling, H. (1933). "Analysis of a complex of statistical variables into principal components." *Journal of Educational Psychology* 24: 417.

- Hugas, M. (1998). "Bacteriocinogenic lactic acid bacteria for the biopreservation of meat and meat products." *Meat Science* 49, *Supplement 1*: S139-S150.
- Hutkins, R. W. and N. L. Nannen (1993). "pH homeostasis in lactic acid bacteria." *Journal of Dairy Science* 76: 2354-2365.
- Jackson, J. E. (2005). *A user's guide to principal components*, Vol. 587, Wiley.
- Jansen, J. J., H. C. J. Hoefsloot, J. van der Greef, M. E. Timmerman and A. K. Smilde (2005). "Multilevel component analysis of time-resolved metabolic fingerprinting data." *Analytica Chimica Acta* 530: 173-183.
- Jewell, J. B. and E. Kashket (1991). "Osmotically regulated transport of proline by *Lactobacillus acidophilus* IFO 3532." *Applied and Environmental Microbiology* 57: 2829-2833.
- Jin, J., B. Zhang, H. Guo, J. Cui, L. Jiang, S. Song, M. Sun and F. Ren (2012). "Mechanism analysis of acid tolerance response of *Bifidobacterium longum* subsp. *longum* BBMN 68 by gene expression profile using RNA-sequencing." *PloS one* 7: e50777.
- Kandler, O. (1983). "Carbohydrate metabolism in lactic acid bacteria." *Antonie van Leeuwenhoek* 49: 209-224.
- Kell, D. B. (2004). "Metabolomics and systems biology: making sense of the soup." *Current Opinion in Microbiology* 7: 296-307.
- Kell, D. B., M. Brown, H. M. Davey, W. B. Dunn, I. Spasic and S. G. Oliver (2005). "Metabolic footprinting and systems biology: the medium is the message." *Nature Reviews Microbiology* 3: 557-565.
- Keller, H. R. and D. L. Massart (1991). "Evolving factor analysis." *Chemometrics and Intelligent Laboratory Systems* 12: 209-224.
- Kenkies, J., R. Ziehn, K. Fritsche, A. Pich and J. R. Andreessen (1999). "Proline biosynthesis from L-ornithine in *Clostridium sticklandii*: purification of  $\Delta$ 1-pyrroline-5-carboxylate reductase, and sequence and expression of the encoding gene, proC." *Microbiology* 145: 819-826.
- Khakimov, B., J. M. Amigo, S. Bak and S. B. Engelsen (2012). "Plant metabolomics: resolution and quantification of elusive peaks in liquid chromatography–mass spectrometry profiles of complex plant extracts using multi-way decomposition methods." *Journal of Chromatography A* 1266: 84-94.
- Kieronczyk, A., S. Skeie, K. Olsen and T. Langsrud (2001). "Metabolism of amino acids by resting cells of non-starter *lactobacilli* in relation to flavour development in cheese." *International Dairy Journal* 11: 217-224.
- Kiers, H. A., J. M. Ten Berge and R. Bro (1999). "PARAFAC2-Part I. A direct fitting algorithm for the PARAFAC2 model." *Journal of Chemometrics* 13: 275-294.
- Kim, H. K., Y. H. Choi and R. Verpoorte (2010). "NMR-based metabolomic analysis of plants." *Nature Protocols* 5: 536-549.
- Kim, H. K., Y. H. Choi and R. Verpoorte (2011). "NMR-based plant metabolomics: where do we stand, where do we go?" *Trends in Biotechnology* 29: 267-275.
- Kim, Y. S., P. Maruvada and J. A. Milner (2008). "Metabolomics in biomarker discovery: future uses for cancer prevention." *Future Medicine* 93-102.

- Klaenhammer, T. R. (1988). "Bacteriocins of lactic acid bacteria." *Biochimie* 70: 337-349.
- Lahtinen, S., A. C. Ouwehand, S. Salminen and A. von Wright (2011). Lactic acid bacteria: Microbiological and functional aspects, CRC Press.
- Lankadurai, B. P., E. G. Nagato and M. J. Simpson (2013). "Environmental metabolomics: an emerging approach to study organism responses to environmental stressors." *Environmental Reviews* 21: 180-205.
- Lawton, W. H. and E. A. Sylvestre (1971). "Self modeling curve resolution." *Technometrics* 13: 617-633.
- Le Bars, D. and M. Yvon (2008). "Formation of diacetyl and acetoin by *Lactococcus lactis* via aspartate catabolism." *Journal of Applied Microbiology* 104: 171-177.
- Lemos, P. C., Y. Dai, Z. Yuan, J. Keller, H. Santos and M. A. Reis (2007). "Elucidation of metabolic pathways in glycogen-accumulating organisms with *in vivo*<sup>13</sup>C nuclear magnetic resonance." *Environmental Microbiology* 9: 2694-2706.
- Lemos, P. C., L. S. Serafim, M. M. Santos, M. A. Reis and H. Santos (2003). "Metabolic pathway for propionate utilization by phosphorus-accumulating organisms in activated sludge: <sup>13</sup>C labeling and *in vivo* nuclear magnetic resonance." *Applied and Environmental Microbiology* 69: 241-251.
- Liu, S.-Q. (2003). "Practical implications of lactate and pyruvate metabolism by lactic acid bacteria in food and beverage fermentations." *International Journal of Food Microbiology* 83: 115-131.
- Liu, S.-Q., R. Holland and V. Crow (2003). "The potential of dairy lactic acid bacteria to metabolise amino acids via non-transaminating reactions and endogenous transamination." *International Journal of Food Microbiology* 86: 257-269.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Oakland, CA, USA.
- Maeder, M. (1987). "Evolving factor analysis for the resolution of overlapping chromatographic peaks." *Analytical Chemistry* 59: 527-530.
- Magnus, J. B., D. Hollwedel, M. Oldiges and R. Takors (2006). "Monitoring and modeling of the reaction dynamics in the valine/leucine synthesis pathway in *Corynebacterium glutamicum*." *Biotechnology Progress* 22: 1071-1083.
- Magnusson, J. and J. Schnürer (2001). "*Lactobacillus coryniformis* subsp. *coryniformis* strain Si3 produces a broad-spectrum proteinaceous antifungal compound." *Applied and Environmental Microbiology* 67: 1-5.
- Mardia, K. V., J. T. Kent and J. M. Bibby (1979). Multivariate analysis, Academic press.
- Marín, S., D. Cuevas, A. J. Ramos and V. Sanchis (2008). "Fitting of colony diameter and ergosterol as indicators of food borne mould growth to known growth models in solid medium." *International Journal of Food Microbiology* 121: 139-149.
- Marín, S., H. Morales, A. J. Ramos and V. Sanchis (2006). "Evaluation of growth quantification methods for modelling the growth of *Penicillium expansum* in an apple-based medium." *Journal of the Science of Food and Agriculture* 86: 1468-1474.

- Marín, S., A. J. Ramos and V. Sanchis (2005). "Comparison of methods for the assessment of growth of food spoilage moulds in solid substrates." *International Journal of Food Microbiology* 99: 329-341.
- Martini, S., M. Ricci, F. Bartolini, C. Bonechi, D. Braconi, L. Millucci, A. Santucci and C. Rossi (2006). "Metabolic response to exogenous ethanol in yeast: An *in vivo* NMR and mathematical modelling approach." *Biophysical Chemistry* 120: 135-142.
- Martini, S., M. Ricci, C. Bonechi, L. Trabalzini, A. Santucci and C. Rossi (2004). "*In vivo* <sup>13</sup>C-NMR and modelling study of metabolic yield response to ethanol stress in a wild-type strain of *Saccharomyces cerevisiae*." *FEBS Letters* 564: 63-68.
- Mashego, M. R., K. Rumbold, M. De Mey, E. Vandamme, W. Soetaert and J. J. Heijnen (2007). "Microbial metabolomics: past, present and future methodologies." *Biotechnology Letters* 29: 1-16.
- Metz, K., M. Lam and A. Webb (2000). "Reference deconvolution: a simple and effective method for resolution enhancement in nuclear magnetic resonance spectroscopy." *Concepts in Magnetic Resonance* 1: 21-42.
- Molenaar, D., A. Hagting, H. Alkema, A. Driessen and W. N. Konings (1993). "Characteristics and osmoregulatory roles of uptake systems for proline and glycine betaine in *Lactococcus lactis*." *Journal of Bacteriology* 175: 5438-5444.
- Morris, G. A. (1988). "Compensation of instrumental imperfections by deconvolution using an internal reference signal." *Journal of Magnetic Resonance* (1969) 80: 547-552.
- Morris, G. A., H. Barjat and T. J. Home (1997). "Reference deconvolution methods." *Progress in Nuclear Magnetic Resonance Spectroscopy* 31: 197-257.
- Mucchetti, G., F. Locci, P. Massara, R. Vitale and E. Neviani (2002). "Production of pyroglutamic acid by thermophilic lactic acid bacteria in hard-cooked mini-cheeses." *Journal of Dairy Science* 85: 2489-2496.
- Mungur, R., A. Glass, D. Goodenow and D. A. Lightfoot (2005). "Metabolite fingerprinting in transgenic *Nicotiana tabacum* altered by the *Escherichia coli* glutamate dehydrogenase gene." *BioMed Research International* 2005: 198-214.
- Nannen, N. L. and R. W. Hutkins (1991). "Proton-translocating adenosine triphosphatase activity in lactic acid bacterial." *Journal of Dairy Science* 74: 747-751.
- Neves, A. R., W. A. Pool, J. Kok, O. P. Kuipers and H. Santos (2005). "Overview on sugar metabolism and its control in *Lactococcus lactis*—the input from *in vivo* NMR." *FEMS Microbiology Reviews* 29: 531-554.
- Neves, A. R., A. Ramos, M. C. Nunes, M. Kleerebezem, J. Hugenholtz, W. M. de Vos, J. Almeida and H. Santos (1999). "*In vivo* nuclear magnetic resonance studies of glycolytic kinetics in *Lactococcus lactis*." *Biotechnology and Bioengineering* 64: 200-212.
- Nicholson, J. K., J. C. Lindon and E. Holmes (1999). "Metabonomics: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data." *Xenobiotica* 29: 1181-1189.
- Nicholson, J. K. and I. D. Wilson (2003). "Understanding global systems biology: metabonomics and the continuum of metabolism." *Nature Reviews Drug Discovery* 2: 668-676.
- Nielsen, J. (2001). "Metabolic engineering." *Applied Microbiology and Biotechnology* 55: 263-283.

- Niku-Paavola, M. L., A. Laitila, T. Mattila-Sandholm and A. Haikara (1999). "New types of antimicrobial compounds produced by *Lactobacillus plantarum*." *Journal of Applied Microbiology* 86: 29-35.
- Nilsson, M. (2009). "The DOSY Toolbox: A new tool for processing PFG NMR diffusion data." *Journal of Magnetic Resonance* 200: 296-302.
- Oda, Y., K. Saito, H. Yamauchi and M. Mori (2002). "Lactic acid fermentation of potato pulp by the fungus *Rhizopus oryzae*." *Current Microbiology* 45: 1-4.
- Oldiges, M., M. Kunze, D. Degenring, G. Sprenger and R. Takors (2004). "Stimulation, monitoring, and analysis of pathway dynamics by metabolic profiling in the aromatic amino acid pathway." *Biotechnology Progress* 20: 1623-1633.
- Oldiges, M., S. Lütz, S. Pflug, K. Schroer, N. Stein and C. Wiendahl (2007). "Metabolomics: current state and evolving methodologies and tools." *Applied Microbiology and Biotechnology* 76: 495-511.
- Oliver, S. G., Winson, M.K., Kell, D.B. and Baganz, R. (1998). "Systematic functional analysis of the yeast genome." *Trends in biotechnology* 16: 373-378.
- Palumbo, E., C. F. Favier, M. Deghorain, P. S. Cocconcelli, C. Grangette, A. Mercenier, E. E. Vaughan and P. Hols (2004). "Knockout of the alanine racemase gene in *Lactobacillus plantarum* results in septation defects and cell wall perforation." *FEMS Microbiology Letters* 233: 131-138.
- Parke, D. and D. Lewis (1992). "Safety aspects of food preservatives." *Food Additives & Contaminants* 9: 561-577.
- Pearce, J. T., T. J. Athersuch, T. M. Ebbels, J. C. Lindon, J. K. Nicholson and H. C. Keun (2008). "Robust algorithms for automated chemical shift calibration of 1D <sup>1</sup>H NMR spectra of blood serum." *Analytical Chemistry* 80: 7158-7162.
- Pekar, J. J. (2006). "A brief introduction to functional MRI." *Engineering in Medicine and Biology Magazine, IEEE* 25: 24-26.
- Piper, P. W., C. Ortiz-Calderon, C. Holyoak, P. Coote and M. Cole (1997). "Hsp30, the integral plasma membrane heat shock protein of *Saccharmyces cerevisiae*, is a stress-inducible regulator of plasma membrane H<sup>+</sup>-ATPase." *Cell Stress & Chaperones* 2: 12.
- Poolman, B. and E. Glaasker (1998). "Regulation of compatible solute accumulation in bacteria." *Molecular Microbiology* 29: 397-407.
- Portais, J.-C. and A.-M. Delort (2002). "Carbohydrate cycling in micro-organisms: what can <sup>13</sup>C-NMR tell us?" *FEMS Microbiology Reviews* 26: 375-402.
- Ramos, A., A. R. Neves and H. Santos (2002). Metabolism of lactic acid bacteria studied by nuclear magnetic resonance. *Lactic Acid Bacteria: Genetics, Metabolism and Applications*, Springer: 249-261.
- Reis, J., A. Paula, S. Casarotti and A. Penna (2012). "Lactic acid bacteria antimicrobial compounds: characteristics and applications." *Food Engineering Reviews* 4: 124-140.
- Ricci, M., M. Aggravi, C. Bonechi, S. Martini, A. M. Aloisi and C. Rossi (2012). "Metabolic response to exogenous ethanol in yeast: an *in vivo* statistical total correlation NMR spectroscopy approach." *Journal of Biosciences* 37: 749-755.

- Rijnen, L., P. Courtin, J.-C. Gripon and M. Yvon (2000). "Expression of a heterologous glutamate dehydrogenase gene in *Lactococcus lactis* highly improves the conversion of amino acids to aroma compounds." *Applied and Environmental Microbiology* 66: 1354-1359.
- Rothman, D. L., K. L. Behar, F. Hyder and R. G. Shulman (2003). "In vivo NMR studies of the glutamate neurotransmitter flux and neuroenergetics: Implications for brain function." *Annual Review of Physiology* 65: 401-427.
- Santos, H. and D. L. Turner (1986). "Characterization of the improved sensitivity obtained using a flow method for oxygenating and mixing cell suspensions in NMR." *Journal of Magnetic Resonance* 68: 345-349.
- Sauer, U., D. R. Lasko, J. Fiaux, M. Hochuli, R. Glaser, T. Szyperski, K. Wüthrich and J. E. Bailey (1999). "Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism." *Journal of Bacteriology* 181: 6679-6688.
- Savorani, F., M. A. Rasmussen, M. S. Mikkelsen and S. B. Engelsen (2013). "A primer to nutritional metabolomics by NMR spectroscopy and chemometrics." *Food Research International* 54: 1131-1145.
- Savorani, F., G. Tomasi, and S. B. Engelsen. (2013) "Alignment of 1D NMR data using the *icoshift* tool: a tutorial." *Magnetic Resonance in Food Science*, Royal Society of Chemistry 343: 14-24.
- Savorani, F., G. Tomasi and S. B. Engelsen (2010). "*icoshift*: A versatile tool for the rapid alignment of 1D NMR spectra." *Journal of Magnetic Resonance* 202: 190-202.
- Searle, S. (1971). *Linear Models*, Wiley.
- Sekiyama, Y., E. Chikayama and J. Kikuchi (2011). "Evaluation of a semipolar solvent system as a step toward heteronuclear multidimensional NMR-based metabolomics for <sup>13</sup>C-labeled bacteria, plants, and animals." *Analytical Chemistry* 83: 719-726.
- Serrazanetti, D. I., M. E. Guerzoni, A. Corsetti and R. Vogel (2009). "Metabolic impact and potential exploitation of the stress reactions in *lactobacilli*." *Food Microbiology* 26: 700-711.
- Serrazanetti, D. I., M. Ndagijimana, S. L. Sado-Kamdem, A. Corsetti, R. F. Vogel, M. Ehrmann and M. E. Guerzoni (2011). "Acid stress-mediated metabolic shift in *Lactobacillus sanfranciscensis* LSCE1." *Applied and Environmental Microbiology* 77: 2656-2666.
- Skibsted, E. and S. B. Engelsen (2010). "Fundamentals of Spectroscopy: Spectroscopy for process analytical technology (PAT)." In *Encyclopedia of Spectroscopy and Spectrometry* (Eds. J. Lindon, G. Tranter and D. Koppenaal), 2<sup>nd</sup> edition, Elsevier, Vol 3: 2651-2661.
- Smart, J. B. and T. D. Thomas (1987). "Effect of oxygen on lactose metabolism in lactic streptococci." *Applied and Environmental Microbiology* 53: 533-541.
- Smilde, A. K., H. Hoefsloot and J. Westerhuis (2008). "The geometry of ASCA." *Journal of Chemometrics* 22: 464-471.
- Smilde, A. K., J. J. Jansen, H. C. Hoefsloot, R.-J. A. Lamers, J. Van Der Greef and M. E. Timmerman (2005). "ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data." *Bioinformatics* 21: 3043-3048.
- Smolinska, A., L. Blanchet, L. M. Buydens and S. S. Wijmenga (2012). "NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review." *Analytica Chimica Acta* 750: 82-97.

- Spraul, M., P. Neidig, U. Klauck, P. Kessler, E. Holmes, J. K. Nicholson, B. C. Sweatman, S. R. Salman, R. D. Farrant, E. Rahr, C. R. Beddell and J. C. Lindon (1994). "Automatic reduction of NMR spectroscopic data for statistical and pattern recognition classification of samples." *Journal of Pharmaceutical and Biomedical Analysis* 12: 1215-1225.
- Spyros, A., P. Dais and P. Belton (2013). NMR spectroscopy in food analysis, Royal Society of Chemistry.
- St, L. and S. Wold (1990). "Multivariate analysis of variance (MANOVA)." *Chemometrics and Intelligent Laboratory Systems* 9: 127-141.
- Steen, A., E. Palumbo, M. Deghorain, P. S. Cocconcelli, J. Delcour, O. P. Kuipers, J. Kok, G. Buist and P. Hols (2005). "Autolysis of *Lactococcus lactis* is increased upon D-alanine depletion of peptidoglycan and lipoteichoic acids." *Journal of Bacteriology* 187: 114-124.
- Stiles, M. E. (1996). "Biopreservation by lactic acid bacteria." *Antonie Van Leeuwenhoek* 70: 331-345.
- Ström, K., J. Sjögren, A. Broberg and J. Schnürer (2002). "*Lactobacillus plantarum* MiLAB 393 produces the antifungal cyclic dipeptides cyclo (L-Phe-L-Pro) and cyclo (L-Phe-trans-4-OH-L-Pro) and 3-phenyllactic acid." *Applied and Environmental Microbiology* 68: 4322-4327.
- Tammam, J., A. Williams, J. Noble and D. Lloyd (2000). "Amino acid fermentation in non-starter *Lactobacillus* spp. isolated from Cheddar cheese." *Letters in Applied Microbiology* 30: 370-374.
- Taniwaki, M. H., J. I. Pitt, A. D. Hocking and G. H. Fleet (2006). Comparison of hyphal length, ergosterol, mycelium dry weight, and colony diameter for quantifying growth of fungi from foods. *Advances in Food Mycology*, Springer: 49-67.
- Tanous, C., A. Kieronczyk, S. Helinck, E. Chambellon and M. Yvon (2002). Glutamate dehydrogenase activity: a major criterion for the selection of flavour-producing lactic acid bacteria strains. *Lactic Acid Bacteria: Genetics, Metabolism and Applications*, Springer: 271-278.
- Tauler, R. (1995). "Multivariate curve resolution applied to second order data." *Chemometrics and Intelligent Laboratory Systems* 30: 133-146.
- Theobald, U., W. Mailinger, M. Reuss and M. Rizzi (1993). "*In vivo* analysis of glucose-induced fast changes in yeast adenine nucleotide pool applying a rapid sampling technique." *Analytical Biochemistry* 214: 31-37.
- Thomas, E. L., T. W. Milligan, R. E. Joyner and M. M. Jefferson (1994). "Antibacterial activity of hydrogen peroxide and the lactoperoxidase-hydrogen peroxide-thiocyanate system against oral streptococci." *Infection and Immunity* 62: 529-535.
- Thomas, T. D., D. C. Ellwood and V. M. C. Longyear (1979). "Change from homo- to hetero-lactic fermentation by *Streptococcus lactis* resulting from glucose limitation in anaerobic chemostat cultures." *Journal of Bacteriology* 138: 109-117.
- Thompson, A., H. Griffin and M. J. Gasson (2002). "Characterization of an alanine racemase gene from *Lactobacillus reuteri*." *Current Microbiology* 44: 246-250.
- Thompson, J. and D. Torchia (1984). "Use of <sup>31</sup>P nuclear magnetic resonance spectroscopy and <sup>14</sup>C fluorography in studies of glycolysis and regulation of pyruvate kinase in *Streptococcus lactis*." *Journal of Bacteriology* 158: 791-800.
- Tikunov, A. P., C. B. Johnson, H. Lee, M. K. Stoskopf and J. M. Macdonald (2010). "Metabolomic investigations of American oysters Using <sup>1</sup>H-NMR spectroscopy." *Marine Drugs* 8: 2578-2596.

- Timmerman, M. and H. L. Kiers (2003). "Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences." *Psychometrika* 68: 105-121.
- Tran, T. N., R. Wehrens and L. M. Buydens (2005). "Clustering multispectral images: a tutorial." *Chemometrics and Intelligent Laboratory Systems* 77: 3-17.
- Trygg, J., J. Gullberg, A. Johansson, P. Jonsson and T. Moritz (2006). Chemometrics in metabolomics—an introduction. Plant metabolomics, Springer: 117-128.
- van den Berg, R. A., H. C. Hoefsloot, J. A. Westerhuis, A. K. Smilde and M. J. van der Werf (2006). "Centering, scaling, and transformations: improving the biological information content of metabolomics data." *BMC Genomics* 7: 142.
- van der Werf, M. J., R. Takors, J. Smedsgaard, J. Nielsen, T. Ferenci, J. C. Portais, C. Wittmann, M. Hooks, A. Tomassini and M. Oldiges (2007). "Standard reporting requirements for biological samples in metabolomics experiments: microbial and *in vitro* biology experiments." *Metabolomics* 3: 189-194.
- Vermeiren, L., F. Devlieghere and J. Debevere (2004). "Evaluation of meat born lactic acid bacteria as protective cultures for the biopreservation of cooked meat products." *International Journal of Food Microbiology* 96: 149-164.
- Veselkov, K. A., J. C. Lindon, T. M. D. Ebbels, D. Crockford, V. V. Volynkin, E. Holmes, D. B. Davies and J. K. Nicholson (2009). "Recursive segment-wise peak alignment of biological  $^1\text{H}$  NMR spectra for improved metabolic biomarker recovery." *Analytical Chemistry* 81: 56-66.
- Von Wright, A. and L. Axelsson (2011). "Lactic acid bacteria: an introduction." *Lactic Acid Bacteria: Microbiological and Functional Aspects*, CRC Press.
- Wang, H., V. K. Tso, C. M. Slupsky and R. N. Fedorak (2010). "Metabolomics and detection of colorectal cancer in humans: a systematic review." *Future Oncology* 6: 1395-1406.
- Wang, H., Y. Yan, J. Wang, H. Zhang and W. Qi (2012). "Production and characterization of antifungal compounds produced by *Lactobacillus plantarum* IMAU10014." *PLoS one* 7: e29452.
- Weber, U. and H. Thiele (2008). *NMR-spectroscopy: modern spectral analysis*, Wiley.
- Wiechert, W. (2002). "Modeling and simulation: tools for metabolic engineering." *Journal of biotechnology* 94: 37-63.
- Williams, A., J. Noble and J. Banks (2001). "Catabolism of amino acids by lactic acid bacteria isolated from Cheddar cheese." *International Dairy Journal* 11: 203-215.
- Williams, A. G., J. Noble and J. M. Banks (2001). "Catabolism of amino acids by lactic acid bacteria isolated from Cheddar cheese." *International Dairy Journal* 11: 203-215.
- Williams, A. G., S. E. Withers and J. M. Banks (2000). "Energy sources of non-starter lactic acid bacteria isolated from Cheddar cheese." *International Dairy Journal* 10: 17-23.
- Wilson, D., A. Burlingame, T. Cronholm and J. Sjövall (1974). "Deuterium and carbon-13 tracer studies of ethanol metabolism in the rat by  $^2\text{H}$ ,  $^1\text{H}$ -decoupled  $^{13}\text{C}$  nuclear magnetic resonance." *Biochemical and Biophysical Research Communications* 56: 828-835.
- Winning, H., F. H. Larsen, R. Bro and S. B. Engelsen (2008). "Quantitative analysis of NMR spectra with chemometrics." *Journal of Magnetic Resonance* 190: 26-32.

- Wishart, D. S. (2008). "Metabolomics: applications to food science and nutrition research." *Trends in Food Science & Technology* 19: 482-493.
- Wishart, D. S. (2008). "Quantitative metabolomics using NMR." *TrAC Trends in Analytical Chemistry* 27: 228-237.
- Wold, S., K. Esbensen and P. Geladi (1987). "Principal component analysis." *Chemometrics and Intelligent Laboratory Systems* 2: 37-52.
- Wold, S., H. Martens and H. Wold (1983). The multivariate calibration problem in chemistry solved by the PLS method. *Matrix pencils*, Springer: 286-293.
- Wold, S., A. Ruhe, H. Wold and I. Dunn, WJ (1984). "The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses." *SIAM Journal on Scientific and Statistical Computing* 5: 735-743.
- Wold, S. and M. Sjöström (1977). "SIMCA: a method for analyzing chemical data in terms of similarity and analogy." *Chemometrics: Theory and Applications*, ACS Ser., Vol. 52, 243-282.
- Wouters, J. M. and G. Petersson (1977). "Reference lineshape adjusted difference NMR spectroscopy. I. Theory." *Journal of Magnetic Resonance* 28: 81-91.
- Wouters, J. M., G. Petersson, W. C. Agosta, F. Field, W. Gibbons, H. Wyssbrod and D. Cowburn (1977). "Reference lineshape adjusted difference NMR spectroscopy. II. Experimental verification." *Journal of Magnetic Resonance* 28: 93-104.
- [www.nobelprize.org](http://www.nobelprize.org) [http://www.nobelprize.org/nobel\\_prizes/medicine/laureates/2003/](http://www.nobelprize.org/nobel_prizes/medicine/laureates/2003/).
- Xavier, K. B., M. S. da Costa and H. Santos (2000). "Demonstration of a novel glycolytic pathway in the hyperthermophilic archaeon *thermococcus zilligii* by <sup>13</sup>C-Labeling experiments and nuclear magnetic resonance analysis." *Journal of Bacteriology* 182: 4632-4636.
- Xu, Y.-J., C. Wang, W. E. Ho and C. N. Ong (2014). "Recent developments and applications of metabolomics in microbiological investigations." *TrAC Trends in Analytical Chemistry* 56: 37-48.
- Yang, E. J. and H. C. Chang (2010). "Purification of a new antifungal compound produced by *Lactobacillus plantarum* AF1 isolated from kimchi." *International Journal of Food Microbiology* 139: 56-63.

# Paper I

---

**Cleaning up NMR spectra with reference deconvolution for improving multivariate analysis of complex mixture spectra**

Parvaneh Ebrahimi, Mathias Nilsson, Gareth A. Morris,  
Henrik M. Jensen, and Søren B. Engelsen

*Journal of Chemometrics, Vol. 28(8), Pages 656-662, 2014*

---



# Cleaning up NMR spectra with reference deconvolution for improving multivariate analysis of complex mixture spectra

Parvaneh Ebrahimi<sup>a,\*</sup>, Mathias Nilsson<sup>a,b</sup>, Gareth A. Morris<sup>b</sup>, Henrik M. Jensen<sup>c</sup> and Søren B. Engelsen<sup>a</sup>

NMR spectroscopy provides valuable data for metabolomics, but the information sought can be partly obscured by errors from hardware imperfection, causing frequency, phase, and spectral lineshape to change significantly between measurements. Clearly, this is a highly undesirable source of variation in multivariate quantitative studies such as metabolomics. Fortunately, many hardware imperfections affect all resonances in the same way. They can therefore be corrected for by comparing an experimental reference peak with the known correct peak shape, in a procedure known as reference deconvolution. This post-measurement processing method can correct many systematic errors in data. The aim of this study is to investigate how reference deconvolution can improve the results obtained by multivariate analysis of NMR data. For this purpose, <sup>1</sup>H NMR data were recorded for a set of 136 mixture samples. Spectra were then produced with and without reference deconvolution and analyzed by principal component analysis and partial least squares methods. The results showed that reference deconvolution resulted in simpler and improved models, requiring fewer latent variables to explain the same or higher percentage of the variance. It was also evident that the recovery of the design concentrations was significantly enhanced. This confirms that reference deconvolution can significantly improve multivariate data analysis and should be considered as a standard tool in high throughput quantitative NMR spectroscopy. Copyright © 2014 John Wiley & Sons, Ltd.

**Keywords:** reference deconvolution; NMR; multivariate data analysis; PCA; PLS

## 1. INTRODUCTION

The quality of NMR spectra has improved substantially with recent improvements in spectrometer design and manufacture. Despite these improvements, significant instrumental imperfections remain, and these are often the limiting factor in determining the amount and quality of information obtainable from NMR experiments. This is particularly true for experiments involving multiple data acquisitions, such as multidimensional NMR methods and chemometric studies, for which instrumental reproducibility is vital. Most instrumental imperfections affect all the signals in a spectrum in the same way [1]. For example, magnetic field inhomogeneity broadens all lines to the same extent, radiofrequency pulse phase error imposes the same phase shift on all signals, and receiver gain variation changes all signal amplitudes equally. Other imperfections in NMR data vary from signal to signal, for example, shifts in peak position due to changes in temperature, concentration, and pH. All of these types of imperfection in NMR data represent sources of unwanted, non-relevant variance and can blur the picture obtained from the biological variance in a metabolomics study, or any investigation relying on quantitative pattern recognition. Any method that can reduce or suppress such unwanted spectral distortion is a welcome addition to the arsenal of methods available to the data analyst.

One data processing method that is highly effective at correcting systematic errors in NMR data is reference deconvolution [1–5]. This extracts the signal of a known reference material from the experimental data, compares it to the theoretically expected form, and

constructs the correction function needed to convert the full experimental dataset into the form that it would have had if the unwanted perturbations experienced by the reference signal had not been present. The reference signal should ideally be a well-resolved singlet of high signal-to-noise ratio (S/N), for which the theoretical lineshape is known [1]. Typical examples of suitable signals are those from 3-(trimethylsilyl)propionic acid (TSP) and tetramethylsilane (TMS), compounds that are commonly added to NMR samples to provide an internal standard for quantification and calibration of the chemical shift axis. Reference deconvolution is fast, linear (to a good approximation—the noise structure is changed slightly because the experimental noise in the reference region is convoluted onto the full spectrum), and robust; it has been known to NMR spectroscopists for many years, but it appears to have been neglected by the NMR-based metabolomics/

\* Correspondence to: Parvaneh Ebrahimi, Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Copenhagen, Denmark.  
E-mail: parvaneh@food.ku.dk

a P. Ebrahimi, M. Nilsson, S. B. Engelsen  
Department of Food Science, Copenhagen University, Rolighedsvej 30, DK-1958 Frederiksberg C, Copenhagen, Denmark

b M. Nilsson, G. A. Morris  
School of Chemistry, University of Manchester, Oxford Road, Manchester M13 9PL, UK

c H. M. Jensen  
DuPont Nutrition Biosciences ApS, Edwin Rahrs Vej 38, Brabrand, Denmark

chemometrics community. The algorithm used in this article, Free Induction Decay Deconvolution for Lineshape Enhancement (FIDDLE), has been used in a wide variety of different contexts [1–9] but has yet to be applied to chemometrics.

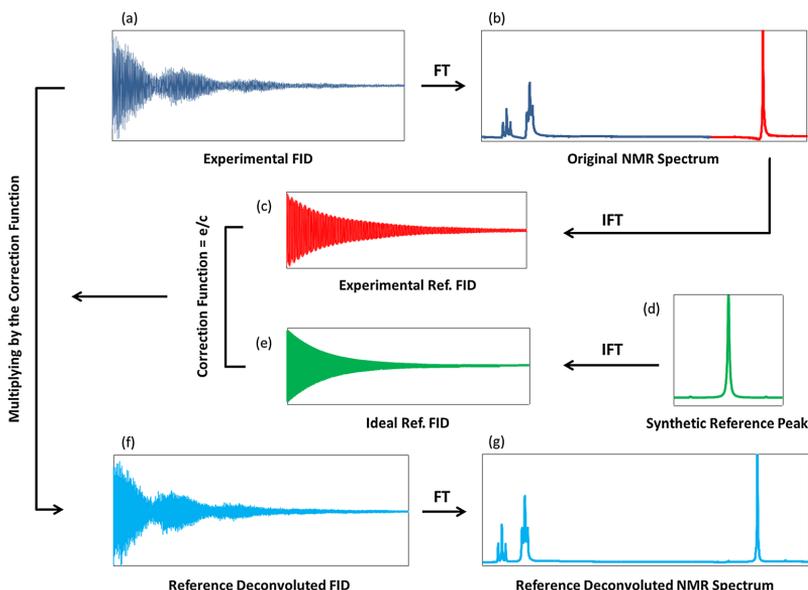
This study investigates how reference deconvolution can help in multivariate data analysis of NMR data. For this purpose, a ternary experimental design was prepared of 136 mixture samples with different concentrations of lactic acid, propionic acid, and lactose, and a constant artificial “metabolic” background consisting of eight different amino acids and carbohydrates.  $^1\text{H}$  NMR spectra were acquired using a standard metabolomics protocol [10], except that a higher-than-usual concentration of the reference material (TSP) was used. The effect of reference deconvolution was then investigated by subjecting the corrected and uncorrected experimental data to two of the most common data mining methods, principal component analysis (PCA) and partial least squares (PLS) methods. PCA and PLS models from the corrected data were superior to those from the uncorrected spectra, demonstrating the ability of reference deconvolution to reduce systematic imperfections in NMR data and, in turn, improve the consistency of a spectral dataset.

## 2. THEORY

A number of different algorithms have been proposed for reference deconvolution [2–4], but they are all based on the same foundations. The FIDDLE algorithm is effective and simple; the theoretical basis has been discussed extensively in the literature [1,5,7,8], but a graphical illustration of the key elements is shown in Figure 1 and explained in the following.

The NMR time-domain data, the free induction decay or FID (Figure 1(a)), are zero-filled (to retain all the spectral information),

Fourier-transformed (FT), and phase-corrected to yield the raw NMR spectrum (Figure 1(b)). A suitable reference signal in the spectrum is then chosen, and the rest of the spectrum is set to zero. The real part (the absorption mode) of this filtered spectrum is subjected to inverse Fourier transformation to give a complex FID that contains only the reference signal (Figure 1(c)). Choosing to retain only the real part of the reference spectrum excludes dispersion mode signals, making clean extraction of the reference signal much easier; no information is lost if the initial FID was zero-filled [6]. In parallel, a synthetic FID (Figure 1(e)) is calculated for the reference signal, using the known frequency (or frequencies; in the case of a reference such as TSP,  $^{29}\text{Si}$  and  $^{13}\text{C}$  satellite signals are included) and a specified lineshape. The latter is chosen by the user, according to need; while the true theoretical lineshape is typically Lorentzian, it can often be advantageous to use a Gaussian shape as this has a narrower base. This choice of target lineshape is analogous to the choice of window function (apodization) in normal FT processing, and the same considerations for resolution or sensitivity enhancement apply [8]. The most conservative choice is a Lorentzian lineshape of approximately the same width as the experimental reference signal (Figure 1(d)); this regularizes the lineshape (and phase and frequency) with minimum change in resolution and  $S/N$ . A complex correction function is then constructed by dividing the ideal reference FID (Figure 1(e)) by the experimental reference FID (Figure 1(c)). The cumulative effect of instrumental imperfections such as field inhomogeneity, pulse phase error, and modulation sidebands is to multiply the FID that would have been recorded if the instrument had behaved ideally by a complex time-domain error function. The correction function calculated here is the inverse of that function, so when the original (full) experimental



**Figure 1.** Schematic illustration of the FIDDLE algorithm for reference deconvolution. The reference peak is extracted from the experimental spectrum (b), and its inverse Fourier transform (c) is compared to that of “perfect” FID (e) to yield a correction function (e/c). The correction is then applied in the time domain to the entire experimental FID (a) to produce the corrected FID (f).

FID (Figure 1(a)) is multiplied by it, the result is a corrected FID (Figure 1(f)) in which all the multiplicative errors seen in the reference FID have been corrected. The corrected FID can then be Fourier transformed to yield the reference-deconvoluted spectrum (Figure 1(g)), in which such imperfections as lineshape distortions, signal amplitude errors, and signal phase changes have been corrected [5,7–9].

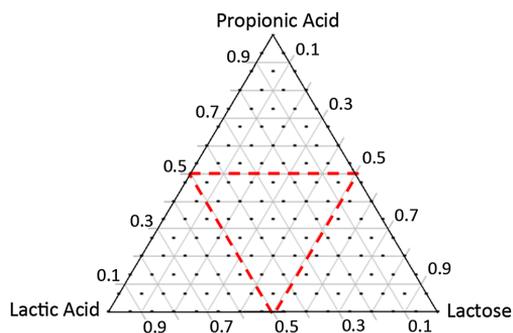
For best results, the reference peak should be a well-resolved singlet, which is present with high amplitude in all the spectra being deconvoluted [1]. The noise in the vicinity of the reference signal will be convoluted onto the entire spectrum, so if the S/N of the reference signal is too low, it can significantly degrade the quality of the data [5]. Multiplets are a much poorer choice for reference signals as they have FIDs that have zero amplitude at regular intervals, which results in singularity problems that are mathematically challenging [9]. The zeroes make interpolation necessary, introducing an element of non-linearity into the algorithm. While the use of a doublet as the reference signal has been reported [11], most software for reference deconvolution does not cater for multiplet reference signals. In this study, only singlet reference signals have been used.

The choice of the ideal peak lineshape and linewidth (the “target lineshape”) is important and warrants further discussion. The lineshape chosen for the ideal reference signal is typically Lorentzian or Gaussian or a mixture of the two [1], although there are many other possibilities. As noted earlier, there is a close analogy between the choice of target lineshape and the apodization procedure used in conventional Fourier transform processing. As most reference signals have a Lorentzian natural shape, and the effects of static field inhomogeneity also often approximate to a Lorentzian distribution of signal amplitude as a function of frequency, the choice of a Lorentzian target lineshape with a width close to that of the experimental reference line will produce a spectrum similar in appearance to the original but with errors in lineshape, phase, frequency, and so on corrected. However, it is often useful to change the target lineshape to aid the extraction of the features of interest from the data under analysis. If a Lorentzian target lineshape narrower than the experimental reference line is chosen, resolution will be increased, but at a severe cost in S/N; if too narrow a lineshape is used, numerical instabilities in the correction will cause severe spectral distortions. Choosing a target lineshape wider than the experimental reference line will increase the S/N at a cost in resolution, with a maximum S/N improvement at twice the experimental linewidth (so-called matched filtration). The choice of a Gaussian or mixed lineshape is often a good alternative, as the narrow base of a Gaussian improves resolution, but at a moderate cost in sensitivity. The optimum target lineshape naturally depends on the objective of the analysis, and comparison between spectra corrected with different target lineshapes is often worthwhile.

### 3. MATERIALS AND METHODS

#### 3.1. Experimental design

A ternary mixture of lactic acid, propionic acid, and lactose was designed using JMP software, version 9 (SAS Institute Inc., Cary, NC, USA) with 16 increments from 0 to 15 mM for each component, which yielded a total of 136 mixtures (see experimental design in Figure 2). Each ternary mixture was prepared in distilled water and added to a metabolic background consisting of a mixture of amino acids and carbohydrates (L-alanine, L-asparagine, L-glutamate,



**Figure 2.** A schematic illustration of the ternary experimental design. A total of 136 mixture samples of lactic acid, propionic acid, and lactose were designed by JMP software. To validate the PLS models (Section 4), mixtures in the center of the design (shown by the dashed triangle) were used as the calibration set, and the remainder of the samples as the test set.

L-leucine, L-phenylalanine, sucrose, glucose, and galactose) at 15 mM each in distilled water. Sodium azide was added to prevent the growth of bacteria and fungi (20 mg per 100 mL of the metabolic background solution). Phosphate buffer with pH 7.4 was also prepared with deuterated water according to a protocol for biological samples [10], which includes TSP as a chemical shift reference. However, concentration of TSP was increased by a factor of 10, relative to the concentration in the original protocol, to 10 mM, in order to ensure high S/N for the TSP singlet to be used as the reference signal in reference deconvolution. The 10-fold increase in the concentration of TSP did not affect the pH of the buffer. To prepare samples for NMR measurement, 200  $\mu$ L of the artificial metabolic background and 200  $\mu$ L of the phosphate buffer were added to 200  $\mu$ L of each ternary design mixture. In the final samples, the concentrations of the ternary design components varied between 0 and 5 mM.

#### 3.2. NMR data acquisition and processing methods

<sup>1</sup>H NMR spectra of the samples were recorded on a Bruker DRX 500 spectrometer (Bruker Biospin GmbH, Rheinstetten, Germany) operating at a proton frequency of 500.13 MHz. For each spectrum, 32,768 complex points were acquired in 64 scans with a recycle delay of 2 s at a nominal temperature of 298 K. The spectrometer was equipped with a 5-mm broadband inverse (BBI) probe, and spectra were recorded using the one-dimensional nuclear Overhauser effect sequence for suppression of the solvent (water) signal. All processing of the data, including phase correction, apodization, Fourier transformation, baseline correction, referencing to TSP signal, and reference deconvolution, was performed using the DOSY Toolbox [12]. Spectra were processed with and without reference deconvolution. Linewidths are expressed as full widths at half-height throughout this paper. Reference deconvolution was performed using the TSP methyl signal as reference, using Gaussian or Lorentzian lineshapes with linewidths ranging from 1 to 5 Hz in 0.25 Hz increments. In order to ensure comparability, FIDs that were not reference-deconvoluted were weighted with Gaussian and Lorentzian apodization functions adjusted to give reference linewidths corresponding as closely as possible to those obtained using reference deconvolution. For example, to make models of

conventional data and reference-deconvoluted with 3 Hz Lorentzian data comparable, line broadening was added to the FID in conventional data to make the width of the reference peak equal to that of the reference signal. The resultant spectra from the DOSY Toolbox were imported into MATLAB 2012b (MathWorks, Inc., Natick, MA, USA) and further processed by normalizing the spectra with respect to the TSP signal area. The MATLAB code for the DOSY Toolbox is freely available from dosytoolbox.chemistry.manchester.ac.uk.

### 3.3. Multivariate analysis

Prior to the multivariate analysis, spectral regions containing only noise, water, or TSP signals were removed from the data. The PLS Toolbox, version 7.0 (Eigenvector Research, Inc., WA, USA), was used for the multivariate analysis. PCA models [13] were calculated for mean-centered datasets. PLS models [14] between the mean-centered data and concentration of lactic acid in the samples were also calculated and cross-validated by the leave-one-out method. Two of the samples were in all cases identified as score outliers (outside the limit of confidence in the primary scores plot) and removed from the datasets.

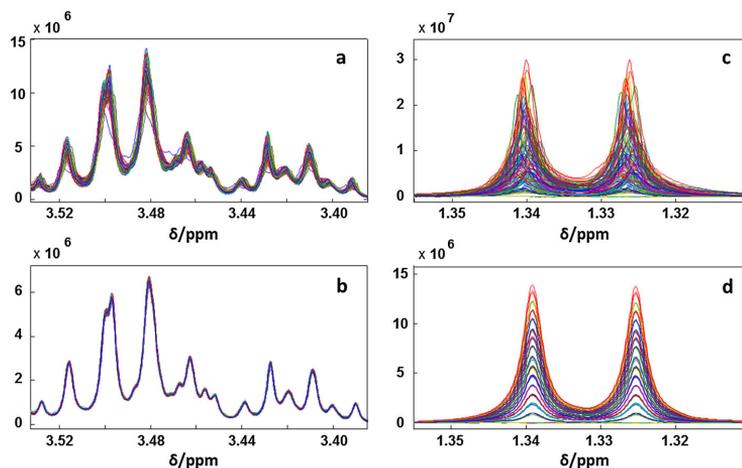
## 4. RESULTS AND DISCUSSION

Selected regions of the conventional and reference-deconvoluted spectra from the 136 samples are shown in Figure 3. The spectra were reference-deconvoluted with a 1.5 Hz Lorentzian target lineshape. The experimental linewidths for the reference (TSP) signal in the spectra measured were around 1.5 Hz; the aim here was to correct spectral errors while minimizing any change in linewidth between uncorrected and corrected spectra, in order to facilitate comparison.

Comparing the conventional and reference-deconvoluted spectra in Figure 3, it can be seen that the signals from the constant metabolic background in the samples are much more consistent in the reference-deconvoluted spectra. For these

signals, reference deconvolution has significantly reduced the effects of experimental and instrumental irreproducibilities—which do not have a chemical/biological source—between the spectra. Inspecting the lactic acid doublet, it is also clear that in the reference-deconvoluted spectra, the lineshapes are much more consistent, and the 16 increments in concentration in the design can be easily observed. Depending on the nature and extent of the lineshape errors in the experimental data, reference deconvolution with a target linewidth equal to the experimental width can increase or decrease S/N. The effect on S/N here was, as expected for good quality data, marginal, with the S/N of the lactic acid doublet for the average spectrum in Figure 3 decreasing from  $3.0 \times 10^4$  in the normal spectrum to  $2.9 \times 10^4$  in the reference-deconvoluted spectrum. Just as in conventional processing of NMR data, the target lineshape in reference deconvolution can be chosen to enhance either the sensitivity or the resolution of the spectrum. Typical choices are a Lorentzian target lineshape broader than the experimental reference line for the former and a Gaussian lineshape narrower than the experimental reference line for the latter. If necessary, the target lineshape can be varied between datasets to maintain the desired balance between resolution and S/N. Where the spectral lines of interest are naturally broader than those of the reference material, resolution enhancement is best achieved by choosing a target lineshape for the reference that contains a negative Lorentzian width contribution and a positive Gaussian (i.e., the corresponding time-domain function corresponds to a rising exponential multiplied by a decaying Gaussian). The negative Lorentzian contribution should correspond to the difference in natural linewidth between the signals of interest and the reference.

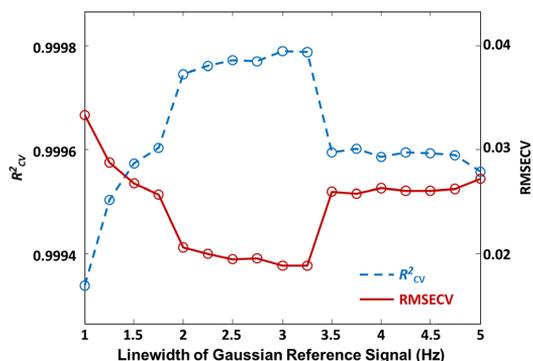
In order to optimize the target linewidth, reference deconvolution with a Gaussian linewidth varying from 1 to 5 Hz in 0.25 Hz increments was performed on all the spectra. The Gaussian lineshape was chosen because it represents a good compromise between resolution and S/N. For each increment, a PLS model was calculated between the spectral data and the



**Figure 3.** NMR spectra with and without reference deconvolution with a 1.5 Hz Lorentzian target lineshape: (a) signals from the constant metabolic background without reference deconvolution; (b) signals from the constant background with reference deconvolution; (c) the doublet originating from lactic acid without reference deconvolution; and (d) the doublet originating from lactic acid with reference deconvolution.

concentration of lactic acid as the response variable. The resultant root mean square error of cross-validation (RMSECV) and the squared Pearson correlation coefficient of cross-validation ( $R_{CV}^2$ ) values as a function of target linewidth are plotted in Figure 4. It can be seen that linewidths between 2 and 3 Hz resulted in the lowest RMSECVs and the highest  $R_{CV}^2$  values. As the optimum region forms a plateau, a linewidth value of 2.5 Hz can safely be chosen as the optimum. The optimum value will depend strongly on the data: where peaks in the raw data are well resolved, an increase in  $S/N$  is beneficial, while for crowded spectra, resolution enhancement may be the better option.

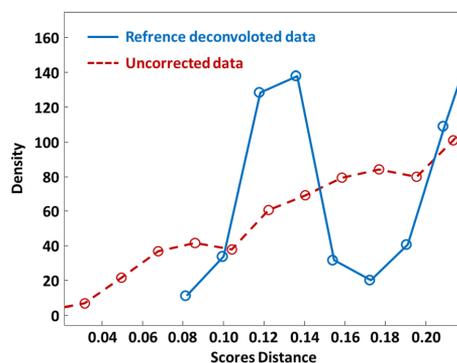
In order to investigate further the spectral variance in the ternary design, and demonstrate how reference deconvolution can improve component modeling of the data, a PCA model was calculated [15,16]. Figure 5 shows the PCA scores plot of the normal spectra and that of the reference-deconvoluted spectra. In this case, the reference-deconvoluted spectra were calculated using the optimal 2.5 Hz Gaussian lineshape, and the normal spectra were weighted with a  $-1.5$  Hz Lorentzian and a  $+2.5$  Hz Gaussian apodization function in order to achieve similar lineshapes and facilitate comparison. From Figure 5, it is clear that the triangular design is much better recovered in the



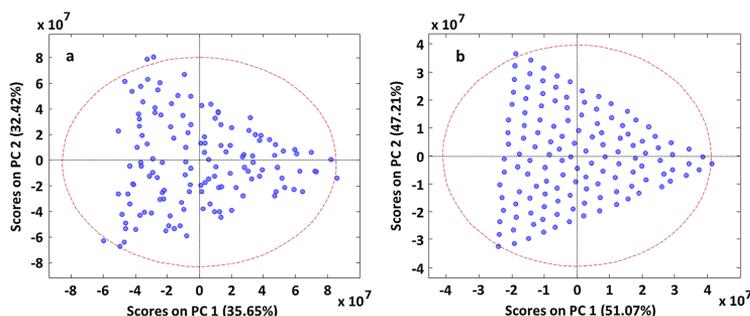
**Figure 4.** Root mean square error of cross-validation (RMSECV) and the squared Pearson correlation coefficient of cross-validation ( $R_{CV}^2$ ) of the PLS models calculated for the experimental NMR data using reference deconvolution with different Gaussian linewidths. Lactic acid concentration was used as the response variable. All the samples were included in the models with the exception of the two outliers.

scores plot from reference-deconvoluted data. Moreover, the percentage of the explained variance for the first two principal components is higher for the reference-deconvoluted data. These are both strong and credible indicators that systematic irregularities have been removed from the data by reference deconvolution and that as a result, simpler PCA models are required to explain the data.

In order to obtain a quantitative measure of the regularity of the PCA scores plots shown in Figure 5, the distances between the scores in normalized scores plots were calculated. This allows numerical confirmation of the higher regularity observed for reference-deconvoluted data. The density plot of the resulted distance distributions (in PC1 and PC2 scores) is shown in Figure 6. The average distance between the sample scores in a normalized plot, considering the span of normalized plots and the 16 increments in the ternary design, should be approximately 0.13 (dividing 2 by the 15 gaps between the scores in the base of the triangle). As evidenced by the plot, for the reference-deconvoluted data, a clear and well-defined peak is observed around 0.13, as compared to the uncorrected data, which only shows a broad shoulder. This implies that in the scores space, the samples appear closer to the correct positions expected for the ternary design. In addition, the distribution is



**Figure 6.** Score distance density plot showing the regularity of the PCA scores. Plots show the density of the distances between the scores for the uncorrected spectral data (red line) and reference-deconvoluted spectral data generated using an optimal 2.5 Hz Gaussian linewidth (blue line).



**Figure 5.** PCA scores plots of (a) raw data weighted with  $-1.5$  Hz Lorentzian and  $+2.5$  Hz Gaussian apodization functions and (b) data reference-deconvoluted using a 2.5 Hz Gaussian target lineshape.

more regular for reference-deconvoluted data, and the density of distances below 0.08 is zero.

Subsequently, PLS models were calculated between the spectral data and the lactic acid concentration of the samples. Models were calculated for a number of different sets, including uncorrected data, reference-deconvoluted data with 1.5 Hz Lorentzian target linewidth, uncorrected data with 1 Hz Lorentzian apodization, reference-deconvoluted data with 2.5 Hz Lorentzian linewidth, uncorrected data with  $-1.5$  Hz Lorentzian and 2.5 Hz Gaussian apodization, and reference-deconvoluted data with 2.5 Hz Gaussian target linewidth. To test the predictive ability of the PLS models, the central part of the triangular design—shown with dashed lines in Figure 2—was used as the calibration set (28 samples) and all the other samples in the design as the test set (106 samples). The statistics for all the PLS models are summarized in Table I. The most noticeable result is that for the PLS models built on the reference-deconvoluted data, each latent variable explains more variance compared to the uncorrected data, and fewer latent variables are needed to describe the data adequately. For the uncorrected data, both with and without apodization, PLS models composed of four latent variables are appropriate, whereas for reference-deconvoluted data, only three latent variables are needed. This is mainly because in reference-deconvoluted data, variations in peaks shape and amplitude due to instrumental inconsistencies—as were observed for lactic acid doublet in Figure 3—have

been corrected. As a result, the data become more bilinear, and simpler multivariate models can be constructed to explain the data and focus on the interesting variance. The root mean square error of calibration (RMSEC) and root mean square error of prediction (RMSEP) values decrease when window functions are applied; this is attributable to the smoothing effect of apodization and broadening of the lines. However, both RMSEC and RMSEP values are further improved in the reference-deconvoluted data when compared to uncorrected data with corresponding apodization (linewidth); this improvement is not attributable to the smoothing effect. Consistent with the prediction errors, the squared Pearson correlation coefficients of the calibration ( $R_{cal}^2$ ) and the prediction ( $R_{pred}^2$ ) are higher in the PLS models of the reference-deconvoluted data.

The percentages of the cumulative variances captured for the **X** and **y** blocks are given in Table I. The cumulative variance captured for the **X** block shows an increase with apodization of the raw data and increases by approximately 10% when reference deconvolution is used. Plots of the variance captured for the **X** and **y** blocks versus number of latent variables are shown in Figure 7; uncorrected data, uncorrected data with 2.5 Hz Gaussian apodization, and reference-deconvoluted data with 2.5 Hz Gaussian target lineshape are included. Inspection of the **X** variance captured for each latent variable in Figure 7(a) shows that for the models built on the reference-deconvoluted data, the latent variables explain more of the variance in the **X** block,

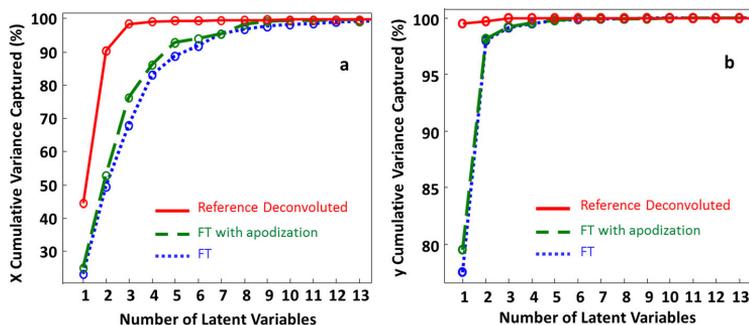
**Table I.** Statistics of the PLS models between the spectra and lactic acid concentration as the response variable

Datasets	Number of LVs	RMSEC	$R_{cal}^2$	RMSEP	$R_{pred}^2$	<b>X</b> Cum. Var. (%)	<b>y</b> Cum. Var. (%)
FT	4	0.0080	0.995	0.0213	0.994	83.06	99.52
RD 1.5 Hz Lorentzian	3	0.0022	0.999	0.0096	0.999	97.72	99.96
FT 1.5 Hz Lorentzian Apodization	4	0.0055	0.998	0.0132	0.997	83.77	99.78
RD 2.5 Hz Lorentzian	3	0.0034	0.999	0.0055	0.999	95.79	99.91
FT 2.5 Hz Gaussian Apodization <sup>a</sup>	4	0.0073	0.996	0.0220	0.994	86.16	99.60
RD 2.5 Hz Gaussian	3	0.0022	0.999	0.0058	0.999	98.40	99.96

Samples from the central part of the triangular design were used as the calibration set and all the other samples as the test set (the two outliers were removed—see Section 3.3).

FT, uncorrected spectral data; RD, reference-deconvoluted spectral data; LVs, latent variables.

<sup>a</sup>Besides +2.5 Hz Gaussian apodization,  $-1.5$  Hz Lorentzian apodization was also used to eliminate the natural linewidth.



**Figure 7.** Captured variance in the **X** and **y** blocks versus number of latent variables. (a) Variance in the **X** block and (b) variance in the **y** block; plots from normal Fourier-transformed data, Fourier-transformed data with 2.5 Hz Gaussian apodization, and reference-deconvoluted data with a 2.5 Hz Gaussian target lineshape are shown in blue, green, and red, respectively.

and that with only three latent variables, almost all the  $\mathbf{X}$  variance is explained. In contrast, for the uncorrected datasets, lower  $\mathbf{X}$  variance is explained for each latent variable. The cumulative variance captured for the response variable ( $\mathbf{y}$  block in Table I) is also higher in the reference-deconvoluted data than for uncorrected data. Figure 7(b) shows the  $\mathbf{y}$  variance captured for each latent variable; for the reference-deconvoluted data, only two latent variables explain almost 100% of the variance, whereas for the uncorrected spectra, at least four latent variables are required to explain a comparable amount of variance in the  $\mathbf{y}$  block.

## 5. CONCLUSIONS

For a designed set of 136 samples,  $^1\text{H}$  NMR spectra were recorded and processed with and without reference deconvolution. Then, PCA and PLS models were calculated, and a comparison was made between the models of the data with and without reference deconvolution. The results clearly demonstrate that reference deconvolution substantially improves PCA and PLS models of the NMR data. This is mainly because reference deconvolution corrects systematic artifacts such as lineshape errors, and as a result, data become more bilinear. The resultant multivariate models become simpler, as they can capture more of the relevant variance, and fewer latent variables are needed to explain the data. Reference deconvolution can be particularly helpful in quantitative NMR spectroscopy, and where quantitative pattern recognition of NMR data is of interest, for example, in NMR-based metabolomics. Investigations are in progress to study the extent to which multivariate analysis of data from real NMR metabolomics studies can benefit from reference deconvolution.

## Acknowledgements

The Danish Council for Strategic Research is acknowledged for generous support to the MicroPAT project under the inSPIRe (Danish Industry–Science Partnership for Innovation and Research in Food Science) consortium. The Faculty of Science

and The Ministry of Science and Technology are acknowledged for a grant to the NMR metabolomics infrastructure.

## REFERENCES

1. Morris GA. Compensation of instrumental imperfections by deconvolution using an internal reference signal. *J. Magn. Reson.* 1988; **80**: 547–552.
2. De Graaf AA, Van Dijk JE, Bovée WMMJ. QUALITY: quantification improvement by converting lineshapes to the Lorentzian type. *Magn. Reson. Med.* 1990; **13**: 343–357.
3. Wouters JM, Petersson GA. Reference lineshape adjusted difference NMR spectroscopy. I. Theory. *J. Magn. Reson.* 1977; **28**: 81–91.
4. Wouters JM, Petersson GA, Agosta WC, Field FH, Gibbons WA, Wyssbrod H, Cowburn D. Reference lineshape adjusted difference NMR spectroscopy. II. Experimental verification. *J. Magn. Reson.* 1977; **28**: 93–104.
5. Morris GA. Reference deconvolution. *eMagRes* 2007; DOI: 10.1002/9780470034590emrstm0449.
6. Bartholdi E, Ernst RR. Fourier spectroscopy and the causality principle. *J. Magn. Reson.* 1973; **11**: 9–19.
7. Metz KR, Lam MM, Webb AG. Reference deconvolution: a simple and effective method for resolution enhancement in nuclear magnetic resonance spectroscopy. *Concepts Magn. Reson.* 2000; **12**: 21–42.
8. Morris GA, Barjat H, Horne TJ. Reference deconvolution methods. *Prog. Nucl. Magn. Reson. Spectrosc.* 1997; **31**: 197–257.
9. Morris GA. Data handling in science and technology (Chapter 16: Reference Deconvolution in NMR), vol 18 Elsevier, 1996; 346–361.
10. Beckonert O, Keun HC, Ebbels TMD, Bundy J, Holmes E, Lindon JC, Nicholson JK. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protocols* 2007; **2**: 2692–2703.
11. Barjat H, Morris GA, Swanson AG, Smart S, Williams SC. Reference deconvolution using multiplet reference signals. *J. Magn. Reson. Ser A* 1995; **116**: 206–214.
12. Nilsson M. The DOSY Toolbox: a new tool for processing PFG NMR diffusion data. *J. Magn. Reson.* 2009; **200**: 296–302.
13. Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemometr. Intell. Lab.* 1987; **2**: 37–52.
14. Wold S, Martens H, Wold H. The multivariate calibration problem in chemistry solved by the PLS method, *Lecture Notes in Mathematics* 1983; **973**: 286–293.
15. Winning H, Larsen FH, Bro R, Engelsen SB. Quantitative analysis of NMR spectra with chemometrics. *J. Magn. Reson.* 2008; **190**: 26–32.
16. Engelsen SB, Savorani F, Rasmussen MA. Chemometric Exploration of Quantitative NMR Data. *eMagRes* 2013; DOI: 10.1002/9780470034590emrstm1304.



# Paper II

---

**Real-time metabolomic analysis of lactic acid bacteria as  
monitored by *in vitro* NMR and chemometrics**

Parvaneh Ebrahimi, Flemming H. Larsen, Henrik M. Jensen,  
Finn K. Vogensen, and Søren B. Engelsen

*Metabolomics, submitted*

---



## **Real-time metabolomic analysis of lactic acid bacteria as monitored by *in vitro* NMR and chemometrics**

<sup>a</sup>Parvaneh Ebrahimi, <sup>a</sup>Flemming Hofmann Larsen, Henrik Max Jensen, <sup>a</sup>Finn Kvist Vogensen, <sup>a</sup>Søren Balling Engelsen

<sup>a</sup> Spectroscopy & Chemometrics, Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 26, 1958 Frederiksberg, Denmark

<sup>b</sup> DuPont Nutrition Biosciences ApS, Edwin Rahrs vej 38, 8220 Brabrand, Denmark

Corresponding author: Parvaneh Ebrahimi ([parvaneh@food.ku.dk](mailto:parvaneh@food.ku.dk)), Tel: +45 3533 2971

## Abstract

Lactic acid bacteria (LAB) play an important role in the food industry as starter cultures to manufacture fermented food, and as probiotics. In recent years, there has been an increasing interest in using LAB cultures for biopreservation of food products. It is therefore of great interest to study the detailed metabolism of these bacteria. This study presents an efficient analytical protocol for real-time *in vitro* NMR measurements of LAB fermentation, from sample preparation over data acquisition and preprocessing to extracting the metabolic kinetic profiles. The analytical protocol is applied to an experimental design with two LAB strains (*Lactobacillus rhamnosus* DSM 20021 and *Lactobacillus plantarum subsp. plantarum* DSM 20174), two pH levels (pH: 6.5 and 5.5), two levels of glucose concentration (2.5 and 0.25 g/l), and two batch fermentation replicates. All the design factors proved to be strongly significant and led to interesting biological information. The protocol allowed for detailed real-time metabolic analysis of 11 major metabolites that are involved in the glycolysis, pyruvate catabolism, amino acid catabolism and cell energy metabolism. Among the obtained biological knowledge are the different patterns of glutamine and aspartic acid consumption by the two strains. It was observed that *L. plantarum* consumes more glutamine at low pH (pH5.5) whereas the opposite applies to *L. rhamnosus*. Regarding aspartic acid, both of the strains consume it higher at low pH, and overall *L. plantarum* consumes it more. *L. rhamnosus* did not consume aspartic acid at pH 6.5. The developed analytical protocol allows relatively easy investigation of different fermentation factors, such as new strains, temperature, and pH.

## Keywords

*in vitro* NMR, lactic acid bacteria, fermentation, modelling metabolic profiles, multivariate curve resolution, reference deconvolution

## 1. Introduction

Lactic acid bacteria (LAB) play a very important role in food industry, and are used as starter cultures to manufacture dairy products, sourdough bread, fermented sausages, and etc. (van de Guchte, Serror et al. 2002). Some of the strains of lactobacilli are used as probiotics, and different health benefits have been claimed for them when used as food supplements (Ljungh and Wadstrom 2006, Kumari, Catanzaro et al. 2011). In recent years, LABs are increasingly being used by food industry for biopreservation, which is regarded as an ecological solution to the problem of food spoilage and can be defined as the controlled use of antifungal and/or antibacterial microorganisms or their metabolites for the purpose of food preservation (Annou S 2007, Delavenne, Ismail et al. 2013). There are many health and consumer political issues associated with the chemical preservatives that are commonly added to food products, whereas the use of biopreservation does not raise such same concerns. However, in order to provide an efficient biopreservation system, detailed knowledge about the metabolism of the protective cultures and their responses to different environmental factors is of great importance.

Environmental metabolomics, that investigates the interactions of organisms with their environment and how they respond to different environmental stressors (Lankadurai, Nagato et al. 2013), can be very useful in the biopreservation studies, as well as other studies that are related to the application of LAB for enhancing the sensory properties of food products. Growth and survival of bacteria that are used as biopreservatives depend upon their stress response, i.e. how they can adapt to the environmental changes such as pH, temperature, carbohydrate source concentration, other exogenous metabolites and cell population density. Sometimes, the adaptive metabolic responses of the cells can lead to the secretion or increased production of desired antimicrobial metabolites. In fermented food products, the reroute of the bacterial metabolism under the stress condition can result in the production of more diverse metabolites, and such stress induced metabolites can in certain cases significantly enhance the sensory quality and structural properties of the food system (Serrazanetti, Guerzoni et al. 2009). Therefore, stress conditions can be designed to augment the quality of fermented food products. In order to achieve this, a detailed understanding of the metabolism of the target microorganism and the mechanisms of stress resistance is a prerequisite. The choice of a suitable analytical platform and carefully designed experimental plans can provide this knowledge.

Nuclear magnetic resonance (NMR) spectroscopy is a powerful analytical technique that can provide qualitative and quantitative information on chemical and biological samples. Proton NMR has been increasingly used in systems biology and metabolomics (Nicholson and Lindon 2008), but it was already applied to study microbial metabolism about forty years ago (Eakin, Morgan et al. 1972). NMR has a great potential for studying living organisms, owing to its non-destructive nature, i.e. it can be used for *in vivo* and *in vitro* measurements of biological processes with no quenching of the metabolism required. Sampling and quenching fractions of a fermentation batch is tedious and likely to introduce specific errors and irreproducibilities (Mashego, Rumbold et al. 2007). Moreover, the metabolic turnover rate is high and changes can

happen in the milliseconds to seconds scale (van der Werf, Takors et al. 2007), and the lag between the sampling a biological process and the measurement can lead to misinterpretations. Present study presents an efficient analytical protocol for real-time *in vitro*  $^1\text{H}$  NMR analysis of bacterial fermentations. The protocol is applied to an experimental design with two strains of LAB, *Lactobacillus rhamnosus* DSM 20021 and *Lactobacillus plantarum* subsp. *plantarum* DSM 20174 (hereafter, *L. plantarum*), two initial pH ( $\text{pH}_i$ ) values and two levels of glucose concentration of a chemically defined medium. The kinetic profiles of selected metabolites are modeled by multivariate curve resolution alternating least squares (MCR-ALS) analysis (Lawton and Sylvestre 1971, de Juan, Jaumot et al. 2014), or by using the second order derivative of the signal, after preprocessing the data with reference deconvolution (Morris, Barjat et al. 1997, Morris 2007) for lineshape enhancements between the time-series, and *icoshift* (Savorani, Tomasi et al. 2010) for spectral alignment. The results of the experimental design are further analyzed by ANOVA-simultaneous component analysis (ASCA) (Smilde, Jansen et al. 2005), and metabolite-metabolite correlations presented in heat maps.

## 2. Materials and methods

In this section, the experimental design and procedure, as well as the data acquisition, processing and analysis are presented. In the last subsection, some theory and background on the relevant metabolism of lactic acid bacteria are presented.

### 2.1. Experimental design and sample preparation

The experimental design of the study includes two strains of LAB, *L. rhamnosus* DSM 20021 and *L. plantarum* DSM 20174, two levels of glucose concentration, 2.5 and 0.25 g/l, and two initial pH ( $\text{pH}_i$ ) values, 6.5 and 5.5 (See Fig. 1S, the Online Resource). The goal was to investigate how the fermentation and the metabolism of the bacteria are influenced by the design factors, and observe some of their biological differences. In order to avoid the full repetition of the names of the samples in the paper, the following abbreviations were used: 'R' for *L. rhamnosus*, 'P' for *L. plantarum*, 'GH' for samples with high glucose concentration, and 'GL' for samples with low glucose concentration. As an example, 'R6.5GH' refers to the *L. rhamnosus* samples with  $\text{pH}_i$  6.5 and high glucose concentration. The abbreviations for the samples names are listed in Fig. 1S (the Online Resource). The strains were obtained from Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (DSMZ, Braunschweig, Germany). Chemically defined interaction medium (CDIM) was prepared as described previously (Aunsbjerg, Honoré et al. 2015), but glucose and lactate were excluded. This CDIM was used as the growth medium for the bacteria. All the chemicals that were used to prepare the CDIM and the samples were obtained from Sigma-Aldrich (Schnelldorf, Germany). Water that was used for preparing the CDIM was freshly produced Milli-Q quality (Merck Millipore, Billerica, MA, USA) water. Phosphate buffer was prepared with pH 6.5 (0.15M and the buffer capacity of 0.09), and pH 5.5 (0.5M and the buffer capacity of 0.10). The buffers were used to adjust the pH of the CDIM to 6.5 and 5.5, and glucose was added to make the two glucose levels of the design. The prepared CDIM were filter-sterilized by 0.2  $\mu\text{m}$  pore size filter (Nalgene®, Thermo Fisher Scientific Inc., Waltham, MA, USA), before storage in the freezer and also prior to use in the experiments. The

strains were stored in Ringer solution with 10% v/v glycerol at -80 °C until use. Before inoculating the CDIM, the cells were centrifuged for 15 min at 6000xg and 4 °C and washed with CDIM twice, to wash out glycerol and avoid its signals in the <sup>1</sup>H NMR spectra, as they can overlap with the signals of interest. Then, the cell pellets were used to inoculate CDIM with 10<sup>7</sup> CFU/ml. Deuterated water (D<sub>2</sub>O) containing 0.1% 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) was sterilized by 0.2 μm pore size syringe filters (Minisart®, Sartorius, Goettingen, Germany) prior to sample preparation. DSS signal serves as an internal reference for the calibration of chemical shift axis. The samples were prepared by adding 1200 μl of CDIM and 300 μl of D<sub>2</sub>O/DSS solution to the cell pellets after the wash. One ml of the samples was put in autoclave sterilized NMR tubes for *in vitro* measurements. To follow the cell growth in parallel, 200 μl of the sample was put in 96-well plates, besides 200 μl of the pure cell-free CDIM as reference, both in duplicates, to measure optical density (OD) during the fermentation. For all of the 8 samples in the design, duplicate samples were prepared and analyzed by this procedure.

## 2.2. Data acquisition and data processing

The sample in the tube was fermented inside the magnet of the NMR spectrometer at 37 °C (310 K) for 24 hours, and <sup>1</sup>H NMR spectra of the sample were recorded every 14 minutes, resulting in 102 spectra for each experiment. The spectrometer was a Bruker Avance-III 600 spectrometer (Bruker Biospin GmbH, Rheinstetten, Germany) operating at a proton frequency of 600.13 MHz (14.1 T), using a double-tuned TCI probe (cryoprobe) equipped for 5 mm sample tubes. All the spectra were recorded using the 'noesygppr1d' pulse sequence, employing a spectral width of 20 ppm, an acquisition time of 843 s, a recycle delay of 10 s, and 64 scans. Taking the duplicates into account, the final acquired NMR data consisted of 16 data matrices, having 102 spectra in the rows and the ppm variables in the columns. The processing of the data, including phase correction, apodization, Fourier transformation, baseline correction, referencing to DSS signal, and reference deconvolution, was performed using the DOSY Toolbox (Nilsson 2009). For reference deconvolution, DSS singlet and a 5 Hz Gaussian target lineshape were used.

Reference deconvolution is a post-measurement NMR data processing method that can correct many systematic errors in NMR spectra. It is previously shown by designed artificial 'metabolic' experiments that reference deconvolution can be helpful in improving the multivariate analysis results of NMR data, by enhancing the line-shapes and bi-linearity of the signals (Ebrahimi, Nilsson et al. 2014). In this work, reference deconvolution was used to solve the line broadening problem of the *in vitro* fermentations, and thus to preserve the bi-linearity of the data. The latter is a pre-requisite for application of methods such as multivariate curve resolution alternating least squares (MCR-ALS) (Engelsen, Savorani et al. 2013).

The data was then imported into MATLAB 2013b (MathWorks, Inc., Natick, MA, USA), and further processed by normalizing the spectra relative to the DSS signal area. Then, the *icoshift* program (Savorani, Tomasi et al. 2010) was used to align the signals, first relative to the DSS signals and then in the defined intervals, which included the signals of interest. For a few signals such as acetate perfect alignment could not be achieved due to large shifts between the samples with the different pH<sub>i</sub> values. As the last preprocessing step, and prior to multivariate

data analysis, spectral regions containing only noise, water, or DSS signals were excluded from the data.

The OD of the samples in the 96-well plates were measured at 600 nm, automatically every 14 minutes, in parallel with the *in vitro* NMR measurements, by a Multiskan™ FC Microplate photometer (Thermo Scientific Inc., Waltham, MA, USA). The plates were shaken by the instrument for 2 s before each reading. Only 100 readings were possible for each experiment, as it was the maximum allowed number of readings for each time-series measurement on the photometer. The OD curves were plotted by averaging the duplicate wells of the sample and subtracting the mean OD value of the CDIM.

2D NMR spectra of some of the ferments at the endpoint of the fermentation were also recorded primarily to assist the assignments of the signals; these data was not used in the data analysis.

### 2.3. Data analysis

#### 2.3.1. Extracting the metabolic profiles

Kinetic metabolic profiles were extracted by using multivariate curve resolution-alternating least squares (MCR-ALS) algorithm (de Juan, Jaumot et al. 2014) on spectral intervals containing the signals that were well-aligned between the spectra, or by modeling the minimum of the peaks second derivative for (singlet) resonances that could not be aligned by *icoshift* (Savorani, Tomasi et al. 2010), because of the considerable shift or concentration difference between the samples. For MCR-ALS modeling, MCR-ALS graphical user interface (GUI) was used (Jaumot, Gargallo et al. 2005). Non-negativity constraints were applied on both the spectral and concentration (time) profiles. As the initial concentrations of the nutrients in the CDIM were known, their profiles were scaled accordingly.

#### 2.3.2. Heat maps and ASCA

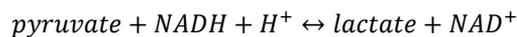
Pearson correlation coefficients were calculated between the kinetic profiles of the metabolites in each sample (duplicates were averaged). In order to focus on the significant correlations, the coefficients less than -0.8 were set to -1, and the coefficients greater than +0.8 were set to +1. The coefficients that were in-between were all set to zero. Correlation coefficients are shown in heat maps that are colored by the value of the coefficients.

Analysis of variance-simultaneous component analysis (ASCA) (Smilde, Jansen et al. 2005) was performed on the extracted metabolic profiles and the OD profiles, excluding lactate and glucose profiles, by using an in-house MATLAB code. The null hypothesis was tested for the main and interaction ASCA effect matrices by a permutation test with 50,000 permutations, in order to evaluate the statistical significance as expressed in the *p*-values. Some of the resulting effect matrices were analyzed further by principal component analysis (PCA). Prior to the PCA, the residual matrix was added to the effect matrices and autoscaling was used.

#### 2.4. Metabolism of lactic acid bacteria

Fig. 1 shows homolactic fermentation (glycolysis pathway) in which pyruvate and subsequently lactate are produced from glucose. The transfer of a phosphate group from phosphoenolpyruvate (PEP) to ADP, which is catalyzed by pyruvate kinase (PK), yields one molecule of pyruvate and one molecule of ATP. This reaction is identified as one of the bottlenecks in glycolysis, at least under the circumstances of limiting glucose concentration. It is suggested that after the addition of glucose or close to its depletion, the high concentration of intracellular inorganic phosphate,  $P_i$ , leads to the reduced activity of PK and as a result the accumulation of PEP and the formation of a metabolic bottleneck at this level (Neves, Pool et al. 2005, Kowalczyk and Bardowski 2007). However, it seems that it is not only a single enzyme, but rather many different enzymes that control the glycolytic flux (Brian J. Koebmann 2002).

Pyruvate can be reduced to lactic acid, through the reaction:

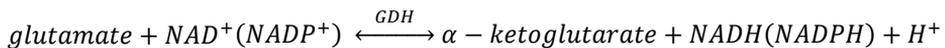


catalyzed by lactate dehydrogenase (LDH) (Hung and Yellen 2014). NAD is an important cofactor for some of the reduction-oxidation reactions in biochemical processes such as the glycolysis. The cytosolic free ratio of NADH/NAD<sup>+</sup> determines the redox state of the cell and is important to keep the redox balance in the cell during glycolysis. It is suggested that NADH/NAD<sup>+</sup> can regulate glycolysis by affecting the activity of involved enzymes such as LDH (Neves, Pool et al. 2005). Accordingly, the ratio of lactate/pyruvate is regulated by the redox state of the cell according to the above biochemical reaction, and NADH/NAD<sup>+</sup> ratio is the main factor that controls how much of the metabolized pyruvate is reduced to lactate during glycolysis (Hung, Albeck et al. 2011, Sun, Dai et al. 2012).

Depending on the type of the bacteria and the physiological conditions, pyruvate can be converted into other metabolites than lactate, through other biochemical pathways. In this metabolism that is known as mixed-acid fermentation, acetic acid, formic acid, succinic acid, and also ethanol can be produced in addition to lactic acid. A higher NADH/NAD<sup>+</sup> ratio stimulates higher LDH activity and consequently promotes homolactic metabolism and higher lactate formation from pyruvate. On the other hand, a lower NADH/NAD<sup>+</sup> ratio reduces the activity of LDH and shifts the metabolism towards mixed-acid fermentation (Kowalczyk and Bardowski 2007). Some of the alternative pathways for pyruvate consumption through mixed-acid fermentation are shown in Fig. 1. Acetic acid is one of the important metabolites that pyruvate can be converted to. Pyruvate dehydrogenase can convert pyruvate to acetyl-CoA and subsequently acetate kinase can catalyze the production of acetate from acetyl-p. The activity of pyruvate oxidase can also lead to the production of acetate from pyruvate. Pyruvate can also be converted to alpha-acetolactate by acetolactate synthase (Lahtinen, Ouwehand et al. 2011), which according to the pathways shown in Fig. 1, can be subsequently converted to diacetyl and acetoin. Acetolactate is converted to diacetyl through a slow chemical, and non-enzymatic reaction and diacetyl can be subsequently reduced to acetoin by diacetyl reductase. Acetolactate can also be directly converted into acetoin by alpha-acetolactate decarboxylase (Caspj, Altman et al. 2014). Synthesis of alpha-acetolactate is active only under the condition of

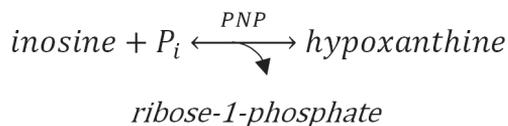
surplus pyruvate, and will be enhanced at low pH values (Le Bars and Yvon 2008). By mixed-acid fermentation pyruvate can also be converted into formic acid. This pathway is catalyzed by pyruvate–formate lyase, which is highly sensitive to oxygen.

Metabolism of amino acids by LAB have important physiological roles including intracellular pH control, controlling the redox state of the cells, and being involved in metabolic stress responses. As LAB are widely used for the production of fermented food products by the industry, studying the catabolism of amino acids by LAB is relevant for the safety and the quality of fermented products. Studying the metabolism of amino acids by LAB has received special attention, because the products from amino acids catabolism can significantly enhance the sensory properties of food products. It is previously reported that for many of LAB strains, catabolic pathways of amino acids start with transamination, which requires presence of alpha-ketoglutarate (Amarita, Requena et al. 2001, Helinck, Le Bars et al. 2004). LAB strains that have active glutamate dehydrogenase (GDH) can produce alpha-ketoglutarate from deamination of glutamic acid, by the biochemical reaction (Tanous, Kieronczyk et al. 2002):



As catabolism of amino acids can produce aromatic compounds, especially from the aromatic, branched-chain and, sulfur-containing amino acids, strains that have higher activity of GDH can be used for flavor enhancement in food products (Kieronczyk, Skeie et al. 2003, Helinck, Le Bars et al. 2004, Liu, Nauta et al. 2008). Amino acids can also be used as energy sources, play a role in controlling intracellular pH, or be involved in the pathways regarding cells stress responses (Fernández and Zúñiga 2006). There is not detailed knowledge on the catabolism of glutamine by LAB, but it has been observed that several strains of LAB can metabolize it (Kieronczyk, Skeie et al. 2001, Williams, Noble et al. 2001). For aspartic acid, three catabolic pathways have been reported that are catalyzed by three different enzymes: aspartate aminotransferase, aspartase, and aspartate decarboxylase. The pathway catalyzed by the aspartate aminotransferase produces oxaloacetate and pyruvate, the one catalyzed by aspartase, depending on the conditions can produce fumaric acid, succinic acid or malic acid, and finally the pathway catalyzed by aspartate decarboxylase can synthesize alanine (Fernández and Zúñiga 2006).

Nucleosides are very important endogenous metabolites in LAB, as they are used as the building blocks of DNA and RNA after phosphorylation and forming of nucleotides. They also serve a very important role in cells energy metabolism. Some bacteria can use nucleosides as carbon source by degrading their pentose ring. Moreover, it is reported that adenosine may be utilized as nitrogen source without degradation of purine ring (Kilstrup, Hammer et al. 2005). To metabolize adenosine as a purine source, it should be deaminated by adenosine deaminase (ADD) to yield inosine. Inosine can then undergo phosphorolytic cleavage, generating ribose-1-P and hypoxanthine, catalyzed by purine nucleoside phosphorylase (PNP), according to the following reaction:



PNP can also convert adenosine into adenine (Kilstrup, Hammer et al. 2005).

### 3. Results and Discussion

#### 3.1. The time-series NMR spectra and the extraction of the metabolic profiles

When the number of cells increases during fermentation, the samples can become inhomogeneous because of the cells coagulation. The inhomogeneity in the sample will result in the line broadening of the signals in the NMR spectra towards the end of the fermentations. Therefore, for a time-series measurement of fermentation, signals will get broader over the time-course of fermentation, as long as the cells are propagating. Lineshape changes by time will definitely lead to errors in quantitative multivariate analysis of the metabolic changes of such data, by methods such as principal component analysis (PCA), partial least squares (PLS) regression, and multivariate curve resolution-alternating least squares (MCR-ALS) that have bilinearity of the data as their main principal. Reference deconvolution was used on the data, in order to enhance the quality of the spectra and to solve the line broadening problem.

Fig. 2 shows some of the signals from one of the time-series spectra processed with and without reference deconvolution (RD) with a 5 Hz Gaussian target lineshape, and subsequently aligned by *icoshift* program. As can be observed from the figure, RD significantly improves the quality of the spectra, and the lineshapes become more consistent between the spectra after reference deconvolution, which in turn makes the data more suitable for bilinear modelling. Fig. 2 also shows that the signals from the amino acids like lysine, arginine, and iso-leucine which do not change significantly during the fermentation become more consistent between the reference-deconvoluted spectra, which again leads to simpler multivariate models and facilitates the visual inspection of the data. Moreover, Fig. 2 shows that using *icoshift* on reference-deconvoluted data corrects the shifts in the position of the resonances between the spectra. This will also lead to the improved multivariate data analysis results. To our knowledge, this is the first application of reference deconvolution to improve the multivariate analysis results of real metabolomics data. In any further analysis of the data and modeling the metabolic profiles, the reference deconvoluted and *icoshifted* data was used.

Table 1 lists the metabolites that were observed to change during fermentation. It also shows the specific signals of each metabolite that was used for extracting the kinetic profiles, and the method that was applied for this purpose (see Section 2.3). The NMR signals of the modeled metabolites in one of the 'R5.5GH' time-series are shown in Fig 2S (the Online Resource). The signals were assigned by using human metabolome database (HMDB) (Wishart, Tzur et al. 2007,

Wishart, Knox et al. 2009, Wishart, Jewison et al. 2012), and biological magnetic resonance bank (BMRB) (Ulrich, Akutsu et al. 2008) databases. As adenosine and inosine signals overlap slightly, one MCR-ALS model with two components was used to model the profiles of both signals simultaneously. Besides, as inosine signal shifts between the samples with pH<sub>i</sub> 6.5 and 5.5, models were calculated for the two pH values separately.

### 3.2. Cell growth and substrate depletion

The extracted metabolic profiles, as well as the OD curves are presented in Fig. 3. To avoid repetition of the samples names, abbreviations that are defined in Section 2.1 and Fig. 1S (the Online Resource) will be used for discussing the profiles. The OD curves (Fig. 3(a)) show that cells grow more rapidly in samples with pH<sub>i</sub> 6.5, and in samples with the high glucose concentration. Cell growth (OD values) is lower in all samples with pH<sub>i</sub> 5.5, and this effect is more pronounced in *L. plantarum* samples. According to the glucose profiles (Fig. 3(b)), and Fig. 3(c)), *L. plantarum* depletes glucose faster in samples with high glucose concentration, whereas in samples with low glucose concentration, *L. rhamnosus* depletes glucose faster. In the beginning of the fermentation *L. rhamnosus* consumes glucose faster, but since *L. plantarum* continuously increases its glucose consumption rate, it overhauls *L. rhamnosus* towards the end of the fermentation process. The rate of glucose consumption decreases in the low pH<sub>i</sub> value and this affects *L. rhamnosus* more than *L. plantarum*. Both of the studied strains can obtain energy from glucose through the glycolysis pathway, which is shown in Fig. 1.

### 3.3. The homofermentative pathway and mixed-acid fermentation

Pyruvate is a key metabolite in the homofermentative pathway where it is converted to lactate as the main metabolic end-product. Fig. 3(d) and Fig. 3(e) show the kinetic profiles of lactate, for high and low glucose concentration, respectively. For both strains, a higher amount of lactate is produced and more pyruvate is accumulated at pH<sub>i</sub> 6.5 samples compared to pH<sub>i</sub> 5.5 samples. This difference can be explained by the difference in the intracellular pH. At pH 6.5, the intracellular pH is close to neutral (~7.5), whereas at pH 5.5, the intracellular pH is close to 6.5 (Siegumfeldt, Rechinger et al. 2000). This leads to the higher activities of PK and LDH (the two enzymes that catalyze the production of pyruvate and lactate respectively) in samples with pH<sub>i</sub> 6.5, in which the intracellular pH is closer to the optimum pH of the enzymes activity. In addition, the activity of LDH seems to be higher in *L. plantarum* samples, as higher amount of lactate is produced in these samples (Fig. 3(d) and Fig. 3(e)). On the other hand, as Fig. 3(f) shows, more pyruvate is accumulated in *L. rhamnosus* samples, which can result from the lower activity of LDH that converts pyruvate to lactate. According to Fig. 3(e), for 'P6.5GL' and 'P5.5GL' samples, the patterns of the lactate profiles are different from the other samples. In 'P5.5GL', the concentration of lactate increases slightly even after glucose depletion, whereas it remains constant in the other samples. According to the extracted metabolic profiles, after glucose depletion (see Fig. 3(b)), pyruvate concentration decreases in samples with the higher glucose concentration (see Fig. 3(f)). However, this decrease is not observed in samples with low glucose concentration. Samples with the lower glucose concentration have a lower metabolic flux and NADH/NAD<sup>+</sup> ratio, which as discussed earlier in Section 2.4 can lead to the reduced activity of

enzymes that are involved in the conversion of pyruvate to other metabolites. (Brian J. Koebmann 2002).

Fig. 3(g) shows the kinetic profiles of acetic acid, which is one of the metabolites that pyruvate can be converted to through mixed-acid fermentation. As the profiles show all the design factors affect the formation of acetic acid. Overall, *L. plantarum* samples produce more acetic acid than *L. rhamnosus* samples, and its highest yield is observed in 'P6.5GL' samples (Fig. 3(g)). Amongst the *L. rhamnosus* samples, the highest acetic acid was achieved at pH<sub>i</sub> 6.5 and low glucose concentration. The higher production of acetic acid in samples with the low glucose concentration can be explained by the smaller NADH/NAD<sup>+</sup> ratio in these samples that shifts the fermentation towards mixed-acid fermentation. Besides, among the samples with the low glucose concentration, the activity of acetate kinase is higher in samples with pH<sub>i</sub> 6.5 that leads to a higher production of acetate (Fig. 3(f)). In a similar way, the lower production of acetate in samples with high glucose concentration is related to the higher NADH/NAD<sup>+</sup> ratio in these samples that stimulates homolactic fermentation. In samples with high glucose concentration, formation of acetic acid is increased close to or after glucose depletion, whereas in samples with low glucose concentration, acetic acid is formed already from the beginning of the fermentation. This trend is observed by comparing the acetic acid profiles of 'P6.5GL' and 'P6.5GH' samples (Fig. 3(g)), for instance. In samples with high glucose concentration, close to the depletion of glucose, the reduced NADH/NAD<sup>+</sup> ratio stimulates mixed-acid fermentation, whereas in the samples with low glucose concentration the small NADH/NAD<sup>+</sup> ratio from the beginning of the fermentation shifts the metabolism towards mixed-acid fermentation.

Fig. 3(h) shows the kinetic profiles of alpha-acetolactate. Alpha-acetolactate can be produced from pyruvate in mixed-acid fermentation, catalyzed by acetolactate synthase (ALS). According to the alpha-acetolactate profiles (Fig. 3(h)), alpha-acetolactate is accumulated only in *L. rhamnosus* samples, at least to a concentration that is detectable by NMR. However, according to Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000) both of the strains have ALS genes. The reason that alpha-acetolactate is accumulated and observed in only one of the strains may be attributed to the difference in the activity of ALS or diacetyl synthase between the strains. The highest concentration of this metabolite is produced in 'R5.5GH' samples, which have high amount of accumulated pyruvate and also a lower pH value—the two factors that enhance the production of alpha-acetolactate (Le Bars and Yvon 2008). 'R6.5GL' samples produce the lowest amount of alpha-acetolactate, which can be explained by the same reasoning. A considerable decline in the concentration of alpha-acetolactate is observed in 'R5.5GH' samples after glucose depletion (Fig. 3(h)), which implies that diacetyl or acetoin are produced. However, signals from acetoin and diacetyl could not be identified, presumably due to their low concentration. It is previously reported that for *Lactococcus lactis*, under the condition of surplus pyruvate, at pH 5.0 alpha-acetolactate was mainly converted into diacetyl, whereas between pH 5.5 and 7.0, it was mainly converted into acetoin (Le Bars and Yvon 2008). Diacetyl and acetoin are redox pairs, and depending on the free NADH/NAD<sup>+</sup> ratio in the cells, and also activity of diacetyl reductase, the ratio of the diacetyl that is reduced to acetoin can vary. Activity of alpha-acetolactate decarboxylase also affects the rate of the direct conversion of alpha-acetolactate into acetoin. The extracted alpha-acetolactate profiles are not as

reproducible as the other profiles for the duplicates. Alpha-acetolactate is highly unstable and in the presence of oxygen will be converted to diacetyl through the related chemical reaction, and under anaerobic condition it can be converted to acetoin.

Fig. 3(i) shows the formic acid profiles. Pyruvate can be converted to formic acid by pyruvate-formate lyase. Only 'R5.5GH' and 'P5.5GH' samples produce formic acid according to the profiles, with the highest amount being produced by 'R5.5GH'. The other samples do not produce detectable amount of formic acid. Accordingly, it can be concluded that pyruvate-formate lyase is active in the low pH value, pH<sub>i</sub> 5.5, and when there is higher amount of carbohydrate source present, and under these conditions, it is more active in *L. rhamnosus* samples.

#### 3.4. The catabolic pathways of amino acids

The kinetic profiles of glutamine catabolism are shown in Fig. 3(j). According to KEGG, both of the studied strains have GDH (Kanehisa and Goto 2000). When examining the extracted metabolic profiles for glutamine (Fig. 3(j)), *L. plantarum* samples consume higher amount of glutamine compared to the *L. rhamnosus* samples. Besides, in *L. plantarum* samples, the consumption of glutamine is higher at pH<sub>i</sub> 5.5 than at pH<sub>i</sub> 6.5 (Fig 3(j)). For *L. rhamnosus* samples, 'R6.5GH' and subsequently 'R5.5GH' consume the highest amount of glutamine. For both *L. plantarum* and *L. rhamnosus*, more glutamine is consumed in samples with the higher glucose concentration. For samples with the lower glucose concentration, glutamine is consumed to a much lower extent, with 'R5.5GL' samples being the lowest (Fig. 3(j)). As glutamine, after being converted to alpha-ketoglutarate, plays a role in the catabolism of amino acids by starting the transamination step as the amino group acceptor, augmenting its consumption can benefit the catabolism of amino acids. This can be of high interest for enhancing the sensory properties of food products, as the catabolism of some amino acids can lead to the formation of desirable aroma compounds. For this purpose, the application of the strains that have high GDH activity and also optimizing influential parameters like pH for increasing glutamine consumption will be helpful. The extracted kinetic profiles of glutamine provide interesting information about the efficiency and the differences of the two investigated strains in metabolizing glutamine. It is observed that *L. plantarum* consumes more glutamine than *L. rhamnosus* (Fig. 3(j)), which can be attributed to the higher activity of GDH in *L. plantarum*. Another interesting observation was the contradictory effect of pH on the consumption of glutamine by the two strains; for *L. plantarum*, the low pH value (pH<sub>i</sub> 5.5) enhances glutamine consumption, whereas for *L. rhamnosus* samples, consumption of glutamine is higher at pH<sub>i</sub> 6.5 (Fig. 3(j)). These information can be very helpful when these strains are used in food products. Signals from the catabolic products of glutamine were not strong enough to be identified reliably by NMR.

The kinetic profiles of aspartic acid catabolism are shown in Fig. 3(k). According to KEGG (Kanehisa and Goto 2000) both of the studied strains have aspartase, which suggests that they can consume aspartic acid and produce fumaric acid, succinic acid or malic acid (Kanehisa and Goto 2000). According to Fig 3(k), *L. plantarum* samples consume more aspartic acid than *L. rhamnosus*, and aspartic acid consumption increases at the high glucose concentration and the

low  $pH_i$  value. For *L. rhamnosus* samples, only 'R5.5GH' samples consume considerable amount of aspartic acid, and at  $pH_i$  6.5 samples, consumption of aspartic acid is very low.

### 3.5. Nucleosides and cell energy metabolism

Fig 3(l) shows the extracted kinetic profiles for adenosine. In our experiment, *L. plantarum* samples consume much more adenosine compared to *L. rhamnosus* samples. In all *L. plantarum* samples adenosine is depleted, which is not the case in *L. rhamnosus* samples (Fig. 3(l)). For *L. plantarum*, the rate of the adenosine consumption is higher in samples with high glucose concentration and the higher  $pH_i$  value, with glucose concentration having the main effect (Fig. 3(l)). The same glucose concentration dependence was observed for the *L. rhamnosus*. In 'R6.5GH' samples, adenosine is consumed at the same rate throughout the fermentation, but in 'R5.5GH' the rate of the adenosine consumption increases considerably after glucose depletion (Fig. 3(l)). *L. rhamnosus* samples with low glucose concentration consume the least amount of adenosine among all the other samples. One of the reasons for the higher consumption of adenosine at the higher glucose concentration is the higher rate of cell propagation and nucleic acid consumption for the new cells. Fig. 3(m) shows the kinetic profiles of inosine. Adenosine can be converted to inosine by ADD or to adenine by PNP. As Fig. 3(m) shows, inosine is only produced in *L. rhamnosus* samples, and analysis of the genome sequences from *L. plantarum* indicates that ADD is missing in these species (Kanehisa and Goto 2000). The maximum amount of inosine is formed in 'R5.5GH' samples (Fig. 3(m)). PNP based on KEGG (Kanehisa and Goto 2000) is present in both of the strains. Fig. 3(n) shows the kinetic profiles of adenine. Adenine can be synthesized from adenosine by purine nucleoside phosphorylase (PNP). As Fig. 3(n) shows, adenine is synthesized in the samples, and the patterns of its profiles are similar but inversed relative to the shape of the corresponding adenosine profiles. Overall, *L. plantarum* samples have a higher rate of adenine synthesis, and the highest amount of adenine is accumulated in 'P6.5GL' and 'P5.5GL' samples (Fig. 3(n)). In these samples, the concentration of adenine increases till the end of glucose depletion, but declines afterwards, especially in one of the 'P6.5GH' samples. Amongst the *L. rhamnosus* samples, the highest amount of adenine is synthesized in 'R5.5GH'. In 'R6.5GH', the concentration of adenine decreases till the end of glucose depletion, but starts to rise afterwards (Fig. 3(n)).

### 3.6. Experimental design and significant factors

In order to investigate the significance of the experimental design factors, and to analyze some of the effect matrices, ASCA was applied on the extracted metabolic profiles including the OD profiles. Fig. 4 shows the workflow for the data analysis and the ASCA results. As glucose was one of the experimental design factors and lactate production is highly correlated to the glucose consumption, these two metabolites were excluded prior to ASCA, in order to focus on the variation of the other metabolites. Based on the  $p$ -values of the effect matrices in Fig. 4, all the design factors, strains,  $pH_i$ , and glucose concentrations are statistically highly significant. The preliminary ASCA results showed that the  $p$ -values for the variance that is related to the replicates were not significant, as can be expected for a correctly performed experiment. Therefore, in the final ASCA, replicates were not included as the design factors. The  $p$ -values verify the experimental design and the experimental work.

The matrices that represent the strain related and pH related variances of the dataset were constructed by adding the residual matrix to the corresponding effect matrix that is given by ASCA, and were subsequently autoscaled and analyzed by PCA. Fig. 5(a) shows the scores and the loadings biplot of the PCA analysis of the ASCA separated strains effect matrix. The first loading mainly describes consumption of adenosine, glutamine, and aspartic acid, and the concomitant production of acetic acid and adenine. The second loading mainly describes production of inosine, pyruvate, alpha-acetolactate, and formic acid. The PCA biplot shows that both *L. rhamnosus* and *L. plantarum* can metabolize adenosine, glutamine, and aspartic acid, but *L. plantarum* consumes them more. Furthermore, *L. plantarum* can produce more acetate and adenine, whereas, *L. rhamnosus* can produce more inosine, pyruvate, acetolactate, and formic acid. This information was presented in a more segmented, but detailed manner by the extracted profiles, but the PCA results of the ASCA separated strain effect matrix, provides a more general and comprehensive overview of the strains differences in terms of the productions and consumption of the metabolites.

Fig. 5(b) shows the PCA results of the pH separated effect matrix. For samples with pH<sub>i</sub> 6.5, the end points of the fermentation mainly lead towards pyruvate and acetate, and for the samples with pH<sub>i</sub> 5.5, towards alpha-acetolactate, adenine, formate, and inosine. As observed from this plot, aspartate consumption increases at pH<sub>i</sub> 5.5 than at pH<sub>i</sub> 6.5, whereas consumption of adenosine and glutamine decrease.

The strain specific pH effect matrix (the interaction between the strains and pH<sub>i</sub>) was also analyzed by PCA. The resulting scores and loadings plots are presented in Fig. 3S (the Online Resource). In the scores plot, only samples with the high glucose concentration are shown, because the samples with the lower glucose concentration are not influenced by the pH change as much as the samples with the higher glucose concentrations. The combination of the first and the third PCs show the fermentation trajectories of the two strains and how they are affected by the change in pH<sub>i</sub>. The strains follow quite different trajectories at the same pH<sub>i</sub> value, and also the metabolic shift in their metabolism due to the change in the pH<sub>i</sub> is different. The perturbation in *L. rhamnosus* fermentation by pH<sub>i</sub> change is more pronounced than in the *L. plantarum*, as the metabolic shift from 'R6.5GH' to 'R5.5GH' is stronger than the shift from 'P6.5GH' to 'P5.5GH'.

### 3.7. Pearson correlations and metabolite-metabolite heat maps

Fig. 4S (the Online Resource) shows the correlation between the metabolites as heat maps, calculated as described in Section 2.3.2. These maps show which metabolites are highly correlated during the fermentation process. Table 1S shows number of the positive correlations that are bigger than 0.8 and the negative correlations that are smaller than -0.8, for all the samples. For *L. rhamnosus* samples with the higher glucose concentration, the numbers of the positive and negative correlations are increased considerably by the change of the pH<sub>i</sub> from 6.5 to 5.5. This means that the change in the pH can significantly alter the metabolic state of the *L. rhamnosus* samples with the higher glucose concentration and by the decrease in pH, more metabolites will decrease or increase concordantly, to reach a new metabolic equilibrium. This probably will help the cells to deal with the exerted pH stress. For *L. plantarum* samples with the

high glucose concentration, the numbers of the positive and negative correlations decrease slightly at  $\text{pH}_i$  5.5 compared to  $\text{pH}_i$  6.5. However, this level of perturbation in the metabolism is not comparable to the change that is observed in the corresponding *L. rhamnosus* samples. Therefore, it can be concluded that the metabolism of *L. rhamnosus* is more sensitive to pH decrease than the metabolism of *L. plantarum*. For *L. rhamnosus* and *L. plantarum* samples with the lower glucose concentration, the numbers of the positive and negative correlations between the metabolites are smaller compared to the samples with the higher glucose concentration. The numbers do not change between the two  $\text{pH}_i$  values. This may be due to the fact that the concentration of glucose in these samples is too low for the metabolism to reach equilibrium, and therefore the cells, being under the shortage of the carbohydrate source, cannot respond to the pH stress by shifting to a new metabolic state.

#### **4. Concluding remarks**

An efficient analytical protocol was developed for *in vitro* NMR studies of bacterial metabolism, from sample preparation to kinetic modelling of metabolic changes. The protocol was successfully applied to an experimental design involving two LAB strains, two pH values and two initial glucose levels, where all the design factors proved to be strongly significant and led to interesting biological information. One of the interesting biological findings was the different patterns of glutamine consumption by the two strains. As glutamine plays a role in the catabolism of amino acids, augmenting its consumption can benefit the catabolism of amino acids, and consequently enhance the sensory properties of food products. Using the developed protocol, NMR proved to be an excellent analytical platform for studying the gross real-time details of bacterial metabolism. As for the data processing, reference deconvolution proved to be a necessary and elegant solution to the inherent inhomogeneity of *in vitro* NMR measurements of cells. It is recommended that reference deconvolution should be considered as a standard tool to enhance lineshapes and improve multivariate analysis results. The protocol can be used for studying different aspects of bacterial metabolism, and allows relatively easy investigation of different fermentation factors such as new strains, cohabitations, new substrates and deleterious metabolites, as well as temperature and pH.

## References

- Amarita, F., T. Requena, G. Taborda, L. Amigo and C. Pelaez (2001). "*Lactobacillus casei* and *Lactobacillus plantarum* initiate catabolism of methionine by transamination." *Journal of applied microbiology* **90**: 971-978.
- Annou S, M. M., Martínez-Bueno M, Valdivia E . (2007). "Biopreservation, an ecological approach to improve the safety and shelf-life of foods." *Applied Microbiology* **1**: 475-486.
- Aunbjerg, S., A. Honoré, F. Vogensen and S. Knøchel (2015). "Development of a chemically defined medium for studying foodborne bacterial-fungal interactions." *International Dairy Journal* **45**: 48-55.
- Brian J. Koebmann, H. W. A., Christian Solem & Peter R. Jensen (2002). "Experimental determination of control of glycolysis in *Lactococcus lactis*." *Antonie van Leeuwenhoek* **82**: 237-248.
- Caspi, R., T. Altman, R. Billington, K. Dreher, H. Foerster, C. A. Fulcher, T. A. Holland, I. M. Keseler, A. Kothari, A. Kubo, M. Krummenacker, M. Latendresse, L. A. Mueller, Q. Ong, S. Paley, P. Subhraveti, D. S. Weaver, D. Weerasinghe, P. Zhang and P. D. Karp (2014). "The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases." *Nucleic Acids Research* **42**: D459-D471.
- de Juan, A., J. Jaumot and R. Tauler (2014). "Multivariate Curve Resolution (MCR). Solving the mixture analysis problem." *Analytical Methods* **6**: 4964-4976.
- Delavenne, E., R. Ismail, A. Pawtowski, J. Mounier, G. Barbier and G. Le Blay (2013). "Assessment of lactobacilli strains as yogurt bioprotective cultures." *Food Control* **30**: 206-213.
- Eakin, R., L. Morgan, C. Gregg and N. Matwiyoff (1972). "Carbon-13 nuclear magnetic resonance spectroscopy of living cells and their metabolism of a specifically labeled <sup>13</sup>C substrate." *FEBS Letters* **28**: 259-264.
- Ebrahimi, P., M. Nilsson, G. A. Morris, H. M. Jensen and S. B. Engelsen (2014). "Cleaning up NMR spectra with reference deconvolution for improving multivariate analysis of complex mixture spectra." *Journal of Chemometrics* **28**: 656-662.
- Engelsen, S. B., F. Savorani and M. A. Rasmussen (2013). Chemometric exploration of quantitative NMR data. eMagRes, DOI: 10.1002/9780470034590.emrstm1304; John Wiley & Sons, Ltd.
- Fernández, M. and M. Zúñiga (2006). "Amino acid catabolic pathways of lactic acid bacteria." *Critical Reviews in Microbiology* **32**: 155-183.
- Helinck, S., D. Le Bars, D. Moreau and M. Yvon (2004). "Ability of thermophilic lactic acid bacteria to produce aroma compounds from amino acids." *Applied and Environmental Microbiology* **70**: 3855-3861.
- Hung, Y. P., J. G. Albeck, M. Tantama and G. Yellen (2011). "Imaging Cytosolic NADH-NAD<sup>+</sup> Redox State with a Genetically Encoded Fluorescent Biosensor." *Cell Metabolism* **14**: 545-554.

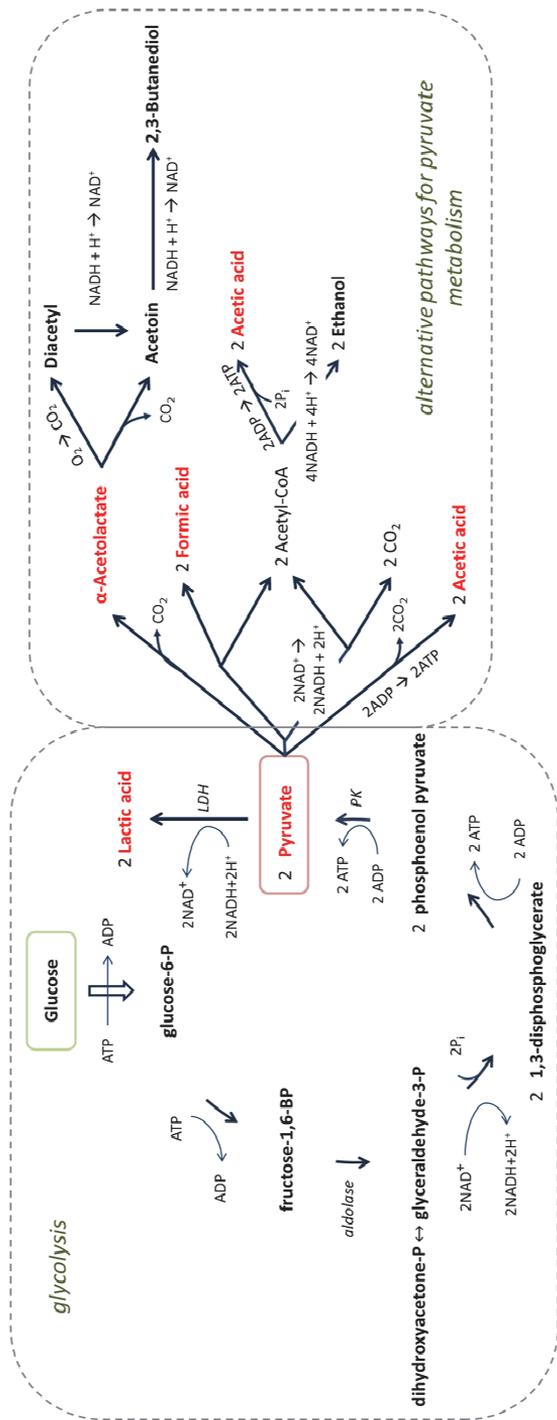
- Hung, Y. P. and G. Yellen (2014). Live-cell imaging of cytosolic NADH–NAD<sup>+</sup> redox state using a genetically encoded fluorescent biosensor. *Fluorescent Protein-Based Biosensors*, Springer: 83-95.
- Jaumot, J., R. Gargallo, A. de Juan and R. Tauler (2005). "A graphical user-friendly interface for MCR-ALS: a new tool for multivariate curve resolution in MATLAB." *Chemometrics and Intelligent Laboratory Systems* **76**: 101-110.
- Kanehisa, M. and S. Goto (2000). "KEGG: kyoto encyclopedia of genes and genomes." *Nucleic Acid Research* **28**: 27-30.
- Kieronczyk, A., S. Skeie, T. Langsrud and M. Yvon (2003). "Cooperation between *Lactococcus lactis* and nonstarter lactobacilli in the formation of cheese aroma from amino acids." *Applied and Environmental Microbiology* **69**: 734-739.
- Kieronczyk, A., S. Skeie, K. Olsen and T. Langsrud (2001). "Metabolism of amino acids by resting cells of non-starter lactobacilli in relation to flavour development in cheese." *International Dairy Journal* **11**: 217-224.
- Kilstrup, M., K. Hammer, P. R. Jensen and J. Martinussen (2005). "Nucleotide metabolism and its control in lactic acid bacteria." *FEMS Microbiology Reviews* **29**: 555-590.
- Kowalczyk, M. and J. Bardowski (2007). "Regulation of sugar catabolism in *Lactococcus lactis*." *Critical Reviews in Microbiology* **33**: 1-13.
- Kumari, A., R. Catanzaro and F. Marotta (2011). "Clinical importance of lactic acid bacteria: a short review." *Acta bio-medica: Atenei Parmensis* **82**: 177-180.
- Lahtinen, S., A. C. Ouwehand, S. Salminen and A. von Wright (2011). *Lactic acid bacteria: Microbiological and functional aspects*, CRC Press.
- Lankadurai, B. P., E. G. Nagato and M. J. Simpson (2013). "Environmental metabolomics: an emerging approach to study organism responses to environmental stressors." *Environmental Reviews* **21**: 180-205.
- Lawton, W. H. and E. A. Sylvestre (1971). "Self modeling curve resolution." *Technometrics* **13**(3): 617-633.
- Le Bars, D. and M. Yvon (2008). "Formation of diacetyl and acetoin by *Lactococcus lactis* via aspartate catabolism." *Journal of Applied Microbiology* **104**: 171-177.
- Liu, M., A. Nauta, C. Francke and R. J. Siezen (2008). "Comparative genomics of enzymes in flavor-forming pathways from amino acids in lactic acid bacteria." *Applied and Environmental Microbiology* **74**: 4590-4600.
- Ljungh, A. and T. Wadstrom (2006). "Lactic acid bacteria as probiotics." *Current Issues in Intestinal Microbiology* **7**: 73-90.
- Mashego, M. R., K. Rumbold, M. De Mey, E. Vandamme, W. Soetaert and J. J. Heijnen (2007). "Microbial metabolomics: past, present and future methodologies." *Biotechnology Letters* **29**: 1-16.

- Morris, G. A. (2007). Reference Deconvolution. *eMagRes*, DOI: 10.1002/9780470034590.emrstm0449; John Wiley & Sons, Ltd.
- Morris, G. A., H. Barjat and T. J. Home (1997). "Reference deconvolution methods." *Progress in Nuclear Magnetic Resonance Spectroscopy* **31**: 197-257.
- Neves, A. R., W. A. Pool, J. Kok, O. P. Kuipers and H. Santos (2005). "Overview on sugar metabolism and its control in *Lactococcus lactis*—the input from *in vivo* NMR." *FEMS Microbiology Reviews* **29**: 531-554.
- Nicholson, J. K. and J. C. Lindon (2008). "Systems biology: Metabonomics." *Nature* **455**: 1054-1056.
- Nilsson, M. (2009). "The DOSY Toolbox: A new tool for processing PFG NMR diffusion data." *Journal of Magnetic Resonance* **200**: 296-302.
- Savorani, F., G. Tomasi and S. B. Engelsen (2010). "icoshift: A versatile tool for the rapid alignment of 1D NMR spectra." *Journal of Magnetic Resonance* **202**: 190-202.
- Serrazanetti, D. I., M. E. Guerzoni, A. Corsetti and R. Vogel (2009). "Metabolic impact and potential exploitation of the stress reactions in lactobacilli." *Food Microbiology* **26**: 700-711.
- Siegmundfeldt, H., K. B. Rechanger and M. Jakobsen (2000). "Dynamic changes of intracellular pH in individual lactic acid bacterium cells in response to a rapid drop in extracellular pH." *Applied and Environmental Microbiology* **66**: 2330-2335.
- Smilde, A. K., J. J. Jansen, H. C. Hoefsloot, R.-J. A. Lamers, J. Van Der Greef and M. E. Timmerman (2005). "ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data." *Bioinformatics* **21**: 3043-3048.
- Sun, F., C. Dai, J. Xie and X. Hu (2012). "Biochemical issues in estimation of cytosolic free NAD<sup>+</sup>/NADH ratio." *PloS one* **7**: e34525.
- Tanous, C., A. Kieronczyk, S. Helinck, E. Chambellon and M. Yvon (2002). Glutamate dehydrogenase activity: a major criterion for the selection of flavour-producing lactic acid bacteria strains. *Lactic Acid Bacteria: Genetics, Metabolism and Applications*, Springer: 271-278.
- Ulrich, E. L., H. Akutsu, J. F. Doreleijers, Y. Harano, Y. E. Ioannidis, J. Lin, M. Livny, S. Mading, D. Maziuk and Z. Miller (2008). "BioMagResBank." *Nucleic Acids Research* **36**: D402-D408.
- van de Guchte, M., P. Serror, C. Chervaux, T. Smokvina, S. D. Ehrlich and E. Maguin (2002). "Stress responses in lactic acid bacteria." *Antonie Van Leeuwenhoek* **82**: 187-216.
- van der Werf, M. J., R. Takors, J. Smedsgaard, J. Nielsen, T. Ferenci, J. C. Portais, C. Wittmann, M. Hooks, A. Tomassini and M. Oldiges (2007). "Standard reporting requirements for biological samples in metabolomics experiments: microbial and *in vitro* biology experiments." *Metabolomics* **3**: 189-194.
- Williams, A. G., J. Noble and J. M. Banks (2001). "Catabolism of amino acids by lactic acid bacteria isolated from Cheddar cheese." *International Dairy Journal* **11**: 203-215.

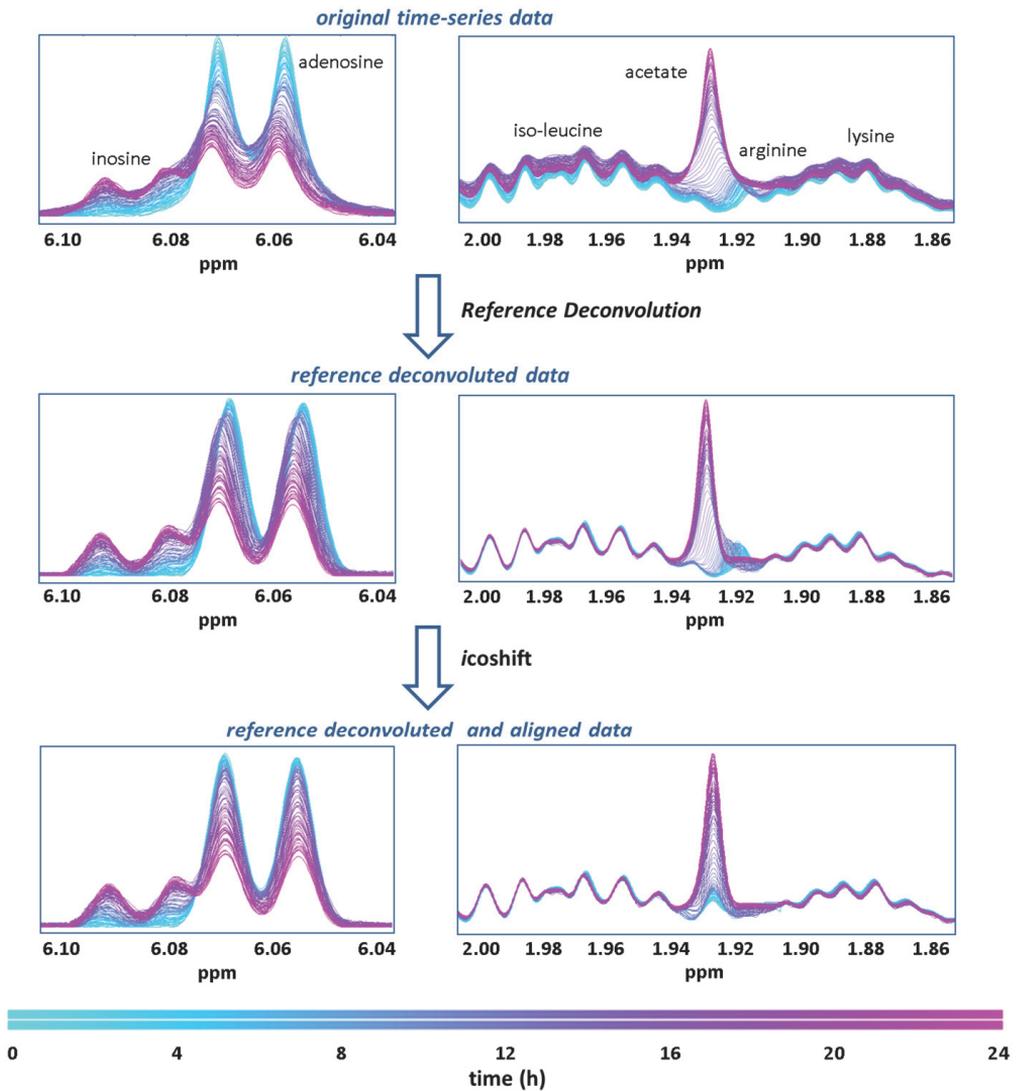
Wishart, D. S., T. Jewison, A. C. Guo, M. Wilson, C. Knox, Y. Liu, Y. Djoumbou, R. Mandal, F. Aziat and E. Dong (2012). "HMDB 3.0—the human metabolome database in 2013." *Nucleic Acids Research*: gks1065.

Wishart, D. S., C. Knox, A. C. Guo, R. Eisner, N. Young, B. Gautam, D. D. Hau, N. Psychogios, E. Dong and S. Bouatra (2009). "HMDB: a knowledgebase for the human metabolome." *Nucleic Acids Research* **37**: D603-D610.

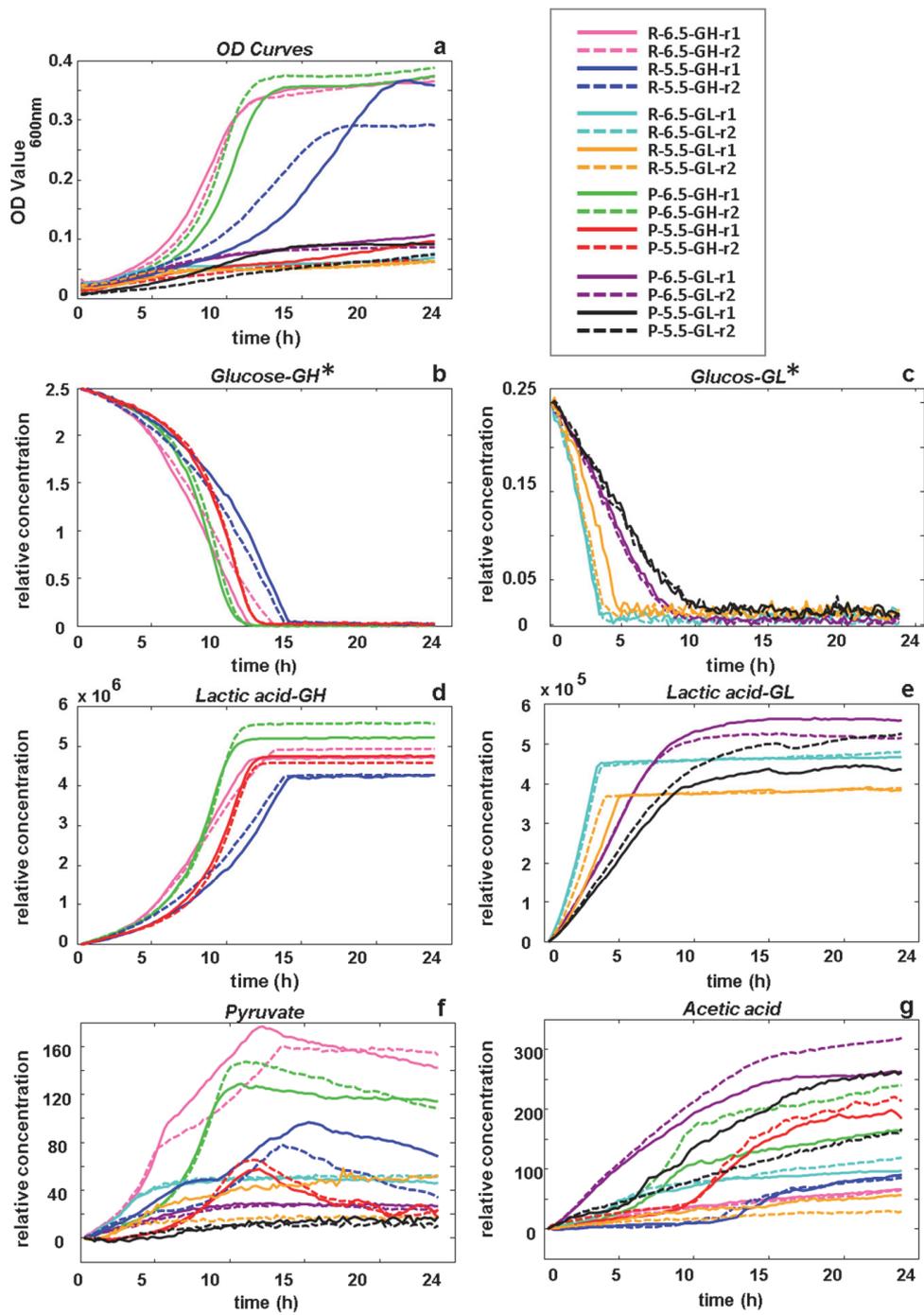
Wishart, D. S., D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt and S. Sawhney (2007). "HMDB: the human metabolome database." *Nucleic Acids Research* **35**: D521-D526.

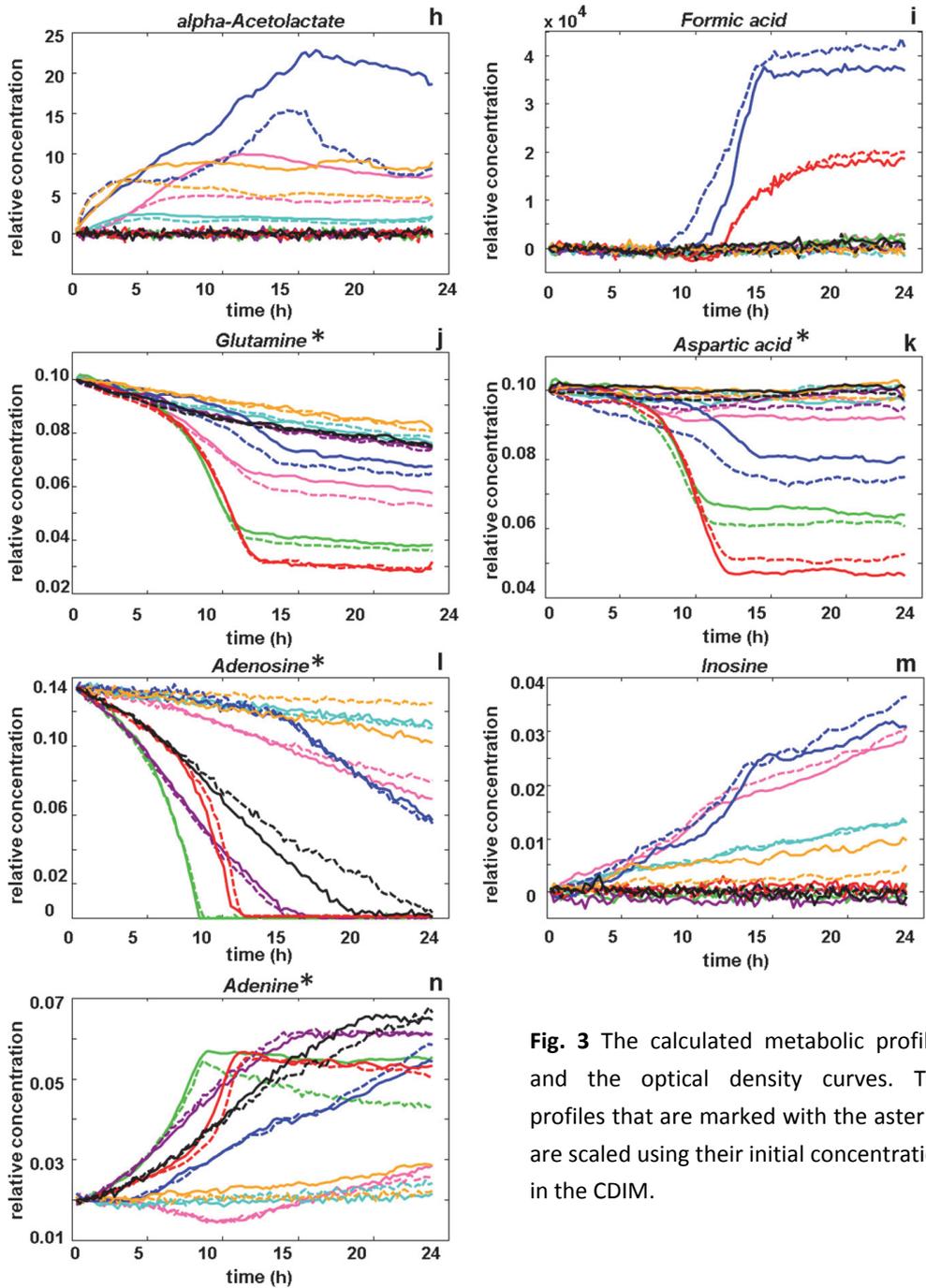


**Fig. 1** Glycolysis (Embden–Meyerhof–Parnas pathway) and the alternative routes for pyruvate catabolism. The metabolites that are written in red are among the metabolites that were identified by the *in vitro* NMR measurements of the experimental design samples.

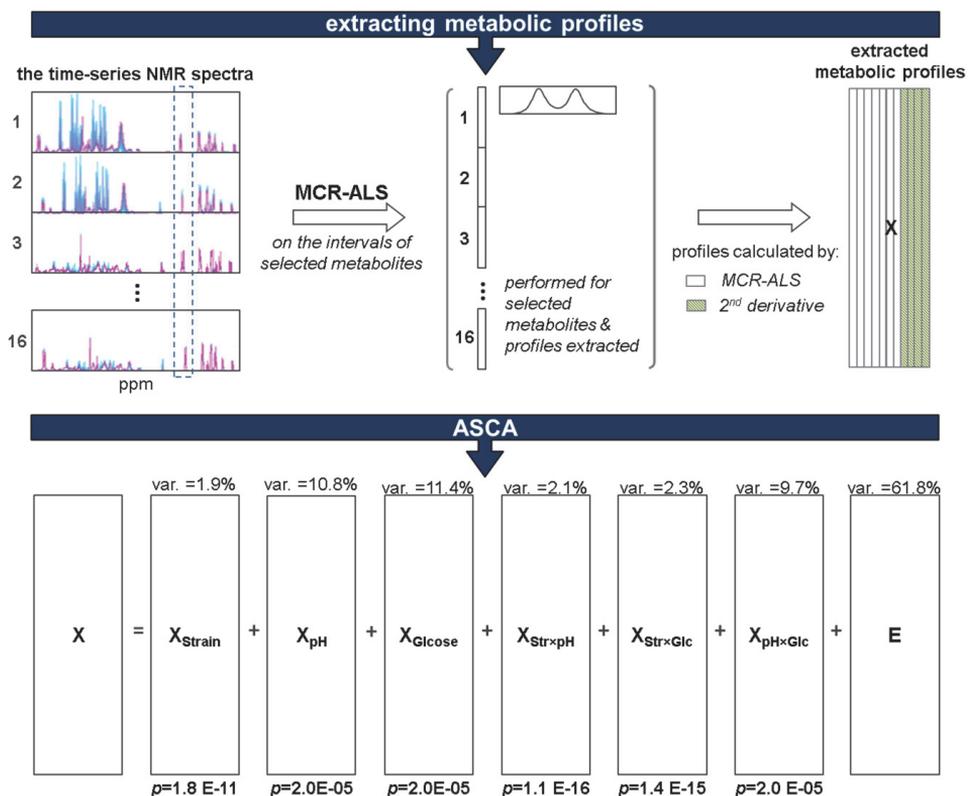


**Fig. 2** Enhancing the quality of real-time *in vitro* measurements of bacterial fermentation by reference deconvolution and *icoshift*. Reference deconvolution can enhance the lineshapes consistencies and correct line broadening, and *icoshift* can align the resonances and correct the shifts in peaks position.

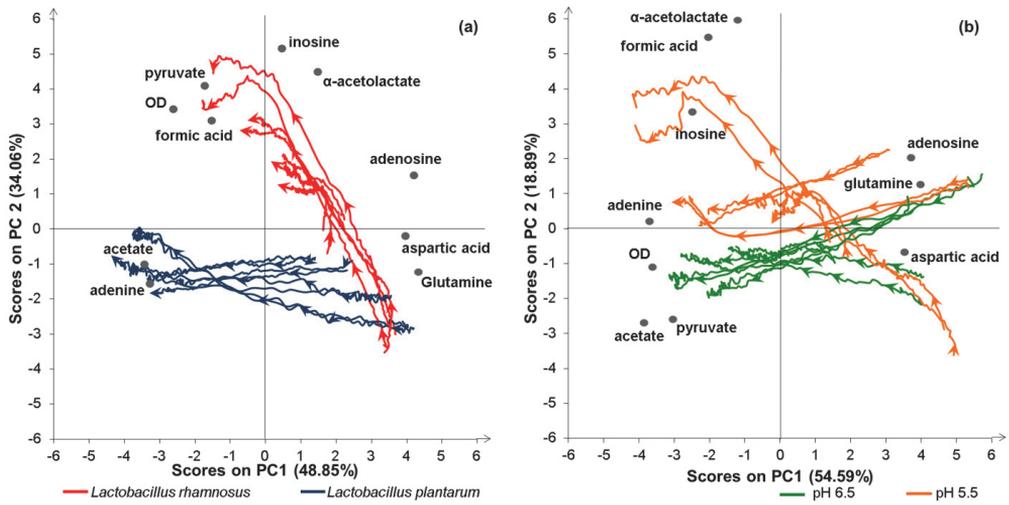




**Fig. 3** The calculated metabolic profiles and the optical density curves. The profiles that are marked with the asterisk are scaled using their initial concentration in the CDIM.



**Fig. 4** The workflow for extracting the metabolic profiles by MCR-ALS and subsequent ASCA analysis. Some of the profiles were extracted by 2<sup>nd</sup> derivative approach. The explained variance and the  $p$ -values of the ASCA effect matrices are shown. ‘ $X_{\text{Strain}}$ ’, ‘ $X_{\text{pH}}$ ’, and ‘ $X_{\text{Glucose}}$ ’ are the main effect matrices of the corresponding design factors. ‘ $X_{\text{StrxpH}}$ ’ and ‘ $X_{\text{StrxGlc}}$ ’ are the strain-pH and strain-glucose interaction effect matrices, and ‘ $X_{\text{pHxGlc}}$ ’ is the pH-glucose interaction effect matrix.

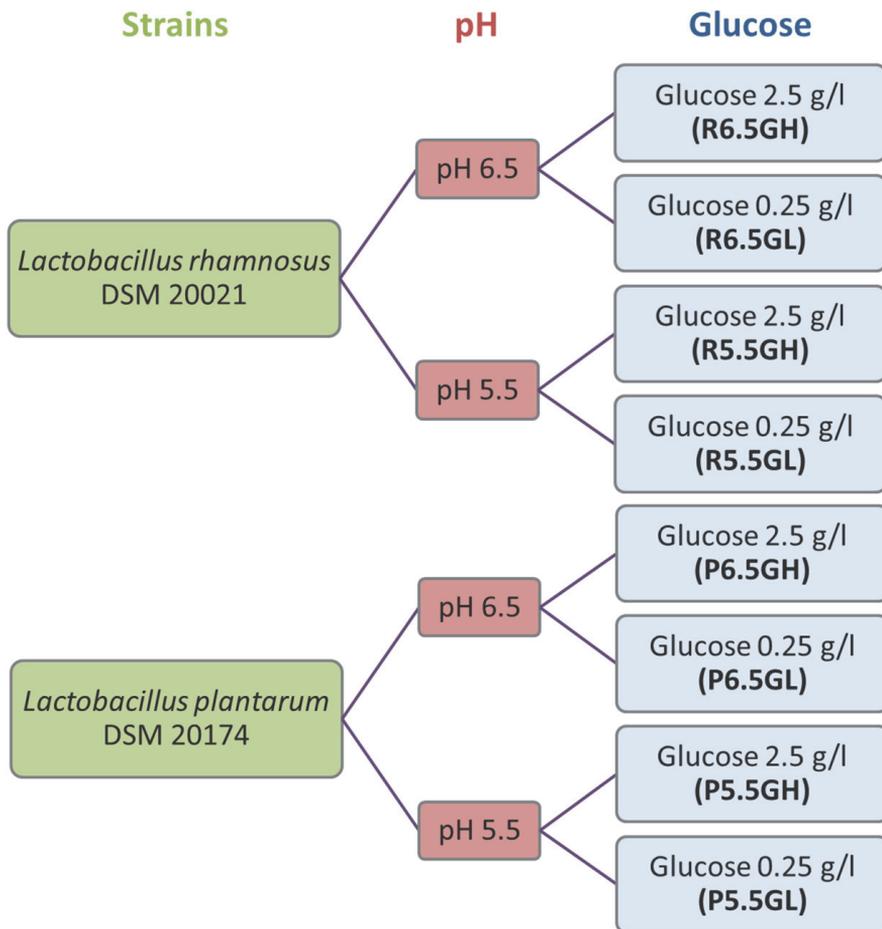


**Fig. 5** PCA results of the ASCA effect matrices: a) strains effect matrix, b) pH effect matrix. The arrows on the scores trajectories show the time progression.

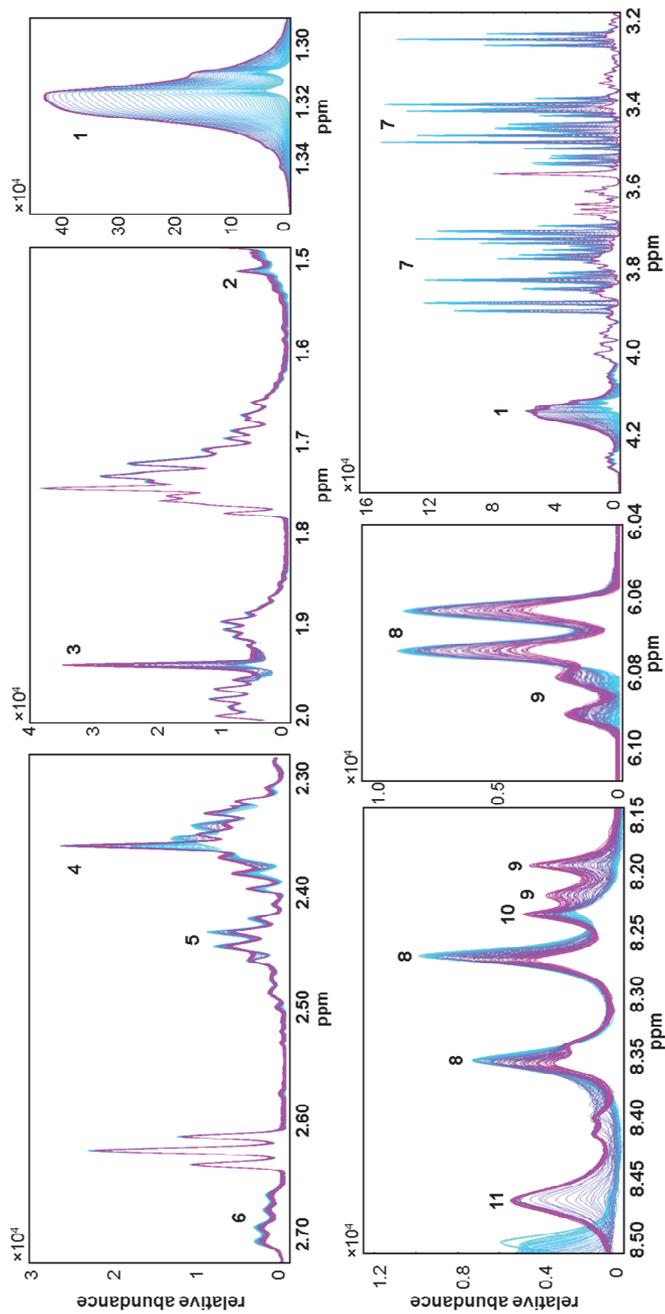
**Table 1** List of the modeled metabolites, the specific signal that was used for extracting the profiles, and the modelling approach that was applied.

<b>Metabolites</b>	<b>NMR signal</b>	<b>Signal type</b>	<b>Profile extracting approach</b>
<b>glucose</b>	5.23 ppm	doublet	MCR-ALS on all the data
<b>lactate</b>	1.32 ppm	doublet	MCR-ALS on all the data
<b>pyruvate</b>	2.36 ppm	singlet	2 <sup>nd</sup> derivative approach
<b>acetic acid</b>	1.91 ppm (pH 6.5)- 1.94 ppm (pH 5.5)	singlet	2 <sup>nd</sup> derivative approach
<b><math>\alpha</math>-acetolactate</b>	1.50 ppm (pH 6.5)- 1.51 ppm (pH 5.5)	singlet	2 <sup>nd</sup> derivative approach
<b>formic acid</b>	8.46 ppm	singlet	MCR-ALS on all the data
<b>glutamine</b>	2.45 ppm	quartet	MCR-ALS on all the data
<b>aspartic acid</b>	2.70 ppm	doublet of doublet	MCR-ALS on all the data
<b>adenosine</b>	6.06 ppm	doublet	MCR-ALS on pH 6.5 and 5.5 datasets separately
<b>inosine</b>	6.08 ppm (pH 6.5)-6.09 ppm (pH 5.5)	doublet	MCR-ALS on pH 6.5 and 5.5 datasets separately
<b>adenine</b>	8.20 ppm (pH 6.5)- 8.22 ppm (pH 5.5)	singlet	2 <sup>nd</sup> derivative approach

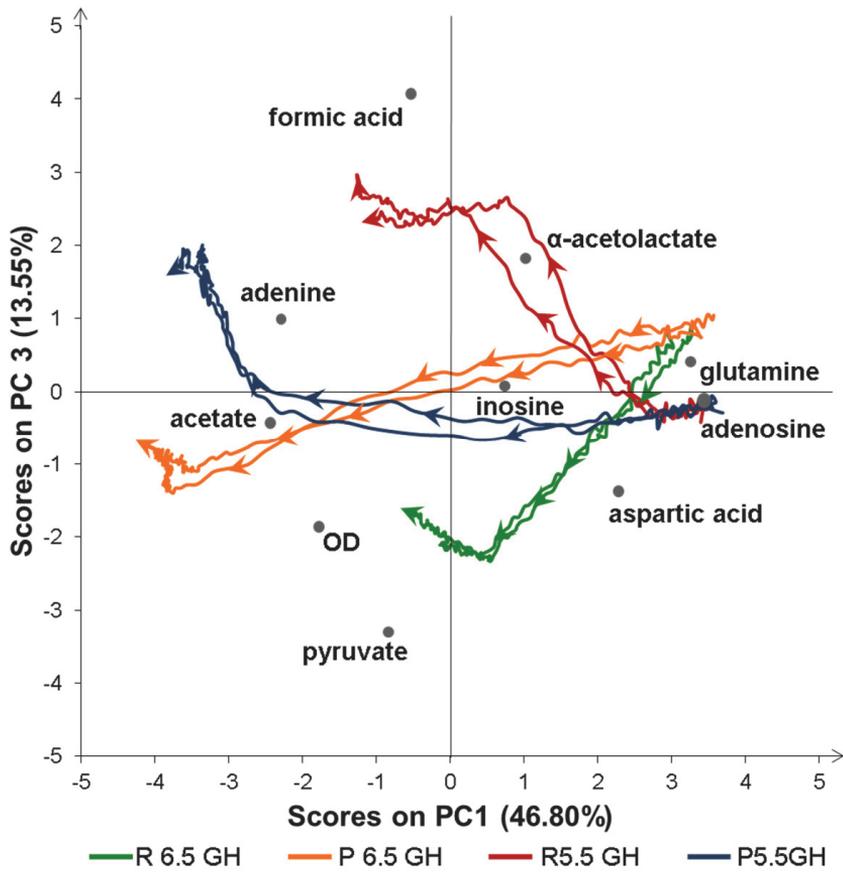
## Supplementary Materials



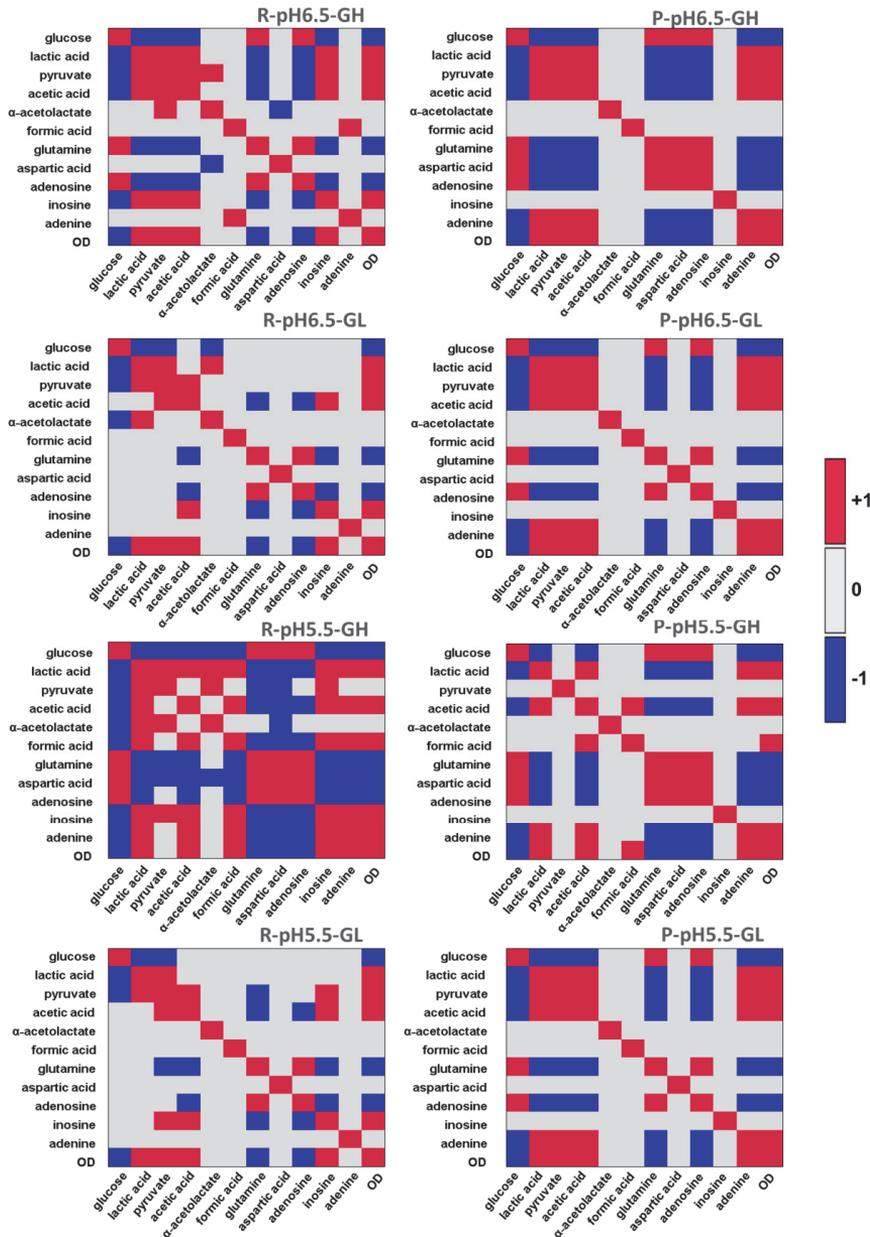
**Fig. 15** The experimental design. The presented abbreviations for the two strains and the samples are used throughout the article in order to avoid the full repetition of the names.



**Fig. 25** Some of the NMR signals from the 'R5.5GH' time-series spectra, including the signals from the modeled metabolites. The assignmet numbers are as follows: 1) lactic acid; 2)  $\alpha$ -acetolactate; 3) acetic acid; 4) pyruvate; 5) glutamine; 6) aspartic acid; 7) glucose; 8) adenosine; 9) inosine; 10) adenine; and 11) formic acid. Lactate signal appears as a broad signal due to the formation of Na/Ca lactate in the sample.



**Fig. 35** PCA results of the ASCA separated strain specific pH effect matrix (the interaction between strain and pH). The arrows on the scores trajectories show the time progression.



**Fig. 4S** The heat maps of the Pearson correlation coefficients calculated between the metabolic profiles of the samples. Only correlation coefficients bigger than +0.8 (red) and smaller than -0.8 (blue) are colored.

## Table-Online Resource

**Table 1S** Number of the significant correlations between the metabolic profiles in the calculated heat maps.

	R 6.5 GH	P 6.5 GH	R 6.5 GL	P 6.5 GL	R 5.5 GH	P 5.5 GH	R 5.5 GL	P 5.5 GL
# of $ corrl  > 0.8$	42	44	30	38	62	40	30	38
# of $ corrl  < -0.8$	32	40	20	30	58	32	20	30



# Paper III

---

## **Quantitative determination of mold growth and inhibition by multispectral imaging**

Parvaneh Ebrahimi, Frans van den Berg, Stina D. Aunbjerg, Anders  
Honoré, Connie Benfeldt, Henrik M. Jensen, and Søren B. Engelsen

*Food Control, Vol. 55, Pages 82-89, 2015*

---





## Quantitative determination of mold growth and inhibition by multispectral imaging



Parvaneh Ebrahimi <sup>a,\*</sup>, Frans van den Berg <sup>a</sup>, Stina D. Aunbjerg <sup>a</sup>, Anders Honoré <sup>a,b</sup>,  
Connie Benfeldt <sup>b</sup>, Henrik M. Jensen <sup>b</sup>, Søren B. Engelsen <sup>a</sup>

<sup>a</sup> Spectroscopy & Chemometrics Group, Department of Food Science, Faculty of Science, Copenhagen University, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

<sup>b</sup> DuPont Nutrition Biosciences ApS, Edwin Rahrs Vej 38, Brabrand, Denmark

### ARTICLE INFO

#### Article history:

Received 24 September 2014

Received in revised form

7 January 2015

Accepted 13 January 2015

Available online 17 February 2015

#### Keywords:

Quantification

Inhibition

Biopreservation

Mold growth

Mold colony size

Multispectral imaging

### ABSTRACT

Quantifying mold growth is of great relevance and interest in microbiology. However, predictive modeling of filamentous fungal growth has been hampered by the lack of an appropriate and accurate method for quantification. An adequate, rapid and objective method will allow studying the effect of many different parameters and conditions on mold growth patterns and can thus provide valuable insight and knowledge. This study outlines a new approach for quantifying mold growth by providing an accurate tool for measuring different segments of mold colonies. The method is based on clustering multispectral images by *k*-means, an unsupervised and simple clustering algorithm. In order to demonstrate the efficiency of the new approach, three different sample sets were analyzed by the developed method, with the objective of quantifying mold growth and size of the colony segments of *Penicillium* mold. The results verify the ability of the proposed method to quantify mold growth and colony composition (relative size of the white and green segments) accurately. This provides a robust measure for interpreting inhibition activity against mold in different samples and makes a quantitative comparison possible. Among the virtues of the method are: 1) the ability to quantify very small differences in the size of colonies which cannot be easily discriminated by visual inspection, 2) the ability to quantify mold growth on transparent as well as on opaque media (e.g. milk), and 3) no prior assumptions for the shape and multiplicity of colonies. The accuracy and non-destructive characteristic of the method allow dynamic quantification of mold growth which can be very valuable in predictive microbiology and in studies related to biopreservation of food products.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Introduction

In food microbiology, it is of great relevance and interest to quantify mold growth and investigate how different parameters influence the process – for instance, in studies related to biopreservation of food products, where safe bacteria are used to inhibit growth of disease-causing microorganisms (Chaillou et al., 2005). However, predictive modeling of filamentous fungal growth has been hindered by the lack of adequate quantitative

methods (Marín, Cuevas, Ramos, & Sanchis, 2008; Marín, Ramos, & Sanchis, 2005). A number of approaches have been reported and used for this purpose, including colony forming units (CFU) counts, total ergosterol content and colony diameter (Marín et al., 2005). CFU is one of the most frequently used methods for quantifying mold growth, but it suffers from serious drawbacks. It usually reflects spore numbers rather than biomass and is, in general, a poor indicator of the extent of fungal growth. CFU also appears to correlate poorly with other parameters such as ergosterol content (Taniwaki, Pitt, Hocking, & Fleet, 2006) which is the second most commonly used method to quantify mold. Ergosterol is the dominant sterol in most fungi, and its concentration accounts for the total fungal population in a food sample (Taniwaki et al., 2006). Although ergosterol has shown good performance as a fungal growth indicator for different species (Marín et al., 2008), it is not possible to determine ergosterol concentration accurately for very

**Abbreviations:** CDIM, Chemically Defined Interaction Medium; GUI, Graphical User Interface.

\* Corresponding author. Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark. Tel.: +45 353 32971.

E-mail address: [parvaneh@food.ku.dk](mailto:parvaneh@food.ku.dk) (P. Ebrahimi).

<http://dx.doi.org/10.1016/j.foodcont.2015.01.050>

0956-7135/© 2015 Elsevier Ltd. All rights reserved.

small colonies (e.g. a colony as small as 2 mm in diameter). By far, the simplest method to assess mold growth is measuring the colony diameter (or area). As molds often grow in the form of surface colonies, colony diameters can be measured on petri dishes over time and converted into growth curves (Taniwaki et al., 2006). Colony diameter measurements show higher repeatability and sensitivity compared to ergosterol measurements (Marín, Morales, Ramos, & Sanchis, 2006), and diameters of very small colonies – for which ergosterol content cannot be measured accurately – can be easily determined. Although colony diameter does not take colony density and volume into account, it is the most suitable measure of the fungal biomass in solid substrates (García, Ramos, Sanchis, & Marín, 2010). Good correlation has been reported between ergosterol content and colony diameter (Marín et al., 2006). Colony diameter and size measurement is also non-destructive, and therefore, saves the sample for further analysis or time-series studies. In General, colony diameter is measured manually (Wang, Yan, Wang, Zhang, & Qi, 2012), and sometimes just a visual inspection of the colonies is used to estimate mold growth and grade inhibition (Magnusson & Schnurer, 2001). For manual measurement of colony area, the routine practice is to measure the diameters of the mold in the two main perpendicular directions and estimate the area or to overlay tracing paper on the mold colony, trace the shape, and then overlay the tracing paper on graph paper and count the squares. These procedures obviously lack accuracy and precision and they can be even less reliable when colonies have not grown in well-shaped circular forms. In addition, they can disturb the mold, and spread the spores around which will bias the results by increasing the apparent growth. Moreover, if the investigated molds are toxic, manual measurement of the colonies can pose potential health risks to the analyst.

This study proposes a new and semi-automated approach for quantifying mold growth based on the colony size or area, using the unsupervised *k*-means clustering of multispectral images, recorded in the ultraviolet, visible and near-infrared regions. Multispectral imaging combines spectroscopy and imaging, and thus provides both spectral and spatial information on the samples, which in this study are mold colonies grown on different media in standard petri dishes. The *k*-means clustering algorithm is a simple procedure which was employed to subdivide multispectral images of the mold colonies on petri dishes, and quantify different segments of the mold colonies.

In order to test and demonstrate the efficiency of the new approach, three different sample sets were analyzed with the objective of quantifying white and green segments of *Penicillium* mold colonies. As white and green segments of the colonies relate to different stages of sporulation, their individual quantification can be informative. The first set of the analyzed samples was inhibition assays for indicator fungi spotted on the cell-free ferments of antifungal bacterial cultures in a chemically defined interaction medium (CDIM), which had a transparent background. The second set of samples was inhibition assays for indicator fungi spotted on the acidified un-inoculated CDIM samples with different pH values. The third set of samples consisted of inhibition assays for indicator fungi spotted on the ferments of the microbial strains on a milk-based medium, which had an opaque background. The new method, called 'PCluster', performs the analysis of the images in a semi-automatic way and is distributed as a freely available MATLAB Graphical User Interface (GUI). The results verify the ability of the proposed strategy to quantify mold growth and colony composition in response to an inhibitor challenge, both on transparent and opaque media. By colony composition, we refer to the relative area (size) of the white and green segments of a colony. For example, if there are two colonies having the same total area, it can be deduced that the one with the bigger size of the green segment, is

more advanced in growth. This is one of the arguments which make separate quantification of the different segments of the mold colony advantageous.

## 2. Theory and methods

### 2.1. *K*-means clustering

*K*-means clustering is an unsupervised algorithm that aims to find the best partitioning of *n* observations (or objects) into *k* clusters or groups, where *k* is a number defined by the user. The algorithm starts by randomly selecting *k* points (objects) as the initial groups' centroids. Then, the Euclidean distance between all the objects and the centroids are calculated and each object is assigned to the cluster to the centroid of which it is the closest. In the next step, for each cluster, the object which is the most similar to the average of all the objects in the *k*th cluster is defined as the new centroid and objects are clustered again, based on their distance from the new centroids. The process of finding the new centroids and re-clustering the objects is repeated iteratively until the convergence criterion is met. The convergence criterion used in our method is minimizing the Within Cluster Sum of Squares (WCSS), which is the average squared Euclidean distance between the objects and their cluster centroids. This is a measure of how well each centroid represents the group or cluster members. The algorithm has converged when WCSS does not decrease any further with iterations or decreases below a predefined threshold (MacQueen, 1976; Mohd, Beg, Herawan, & Rabbi, 2012; Tran, Wehrens, & Buydens, 2005). Generally, for each clustering, replicate runs/restarts are performed. Restarts of *k*-means will help to make sure that the algorithm does not converge to local minima. Each one of the replicates begins from a different randomly selected set of initial centroids and the final solution that *k*-means returns is the global minimum which has the lowest WCSS.

### 2.2. Multispectral images

Each multispectral image is a three-dimensional data structure in which two of the dimensions provide spatial information of the sample, and the third dimension represents spectral information for each picture-element (pixel; see Fig. 1). Spectral information is provided for a range of wavelengths, including ultraviolet, visible, and near-infrared which can provide much more information about the samples compared to e.g. the ordinary trichromatic (RGB) images. Multispectral images give information on the color, surface properties, water content and other important physical and chemical properties of the samples (Dissing et al., 2013; Guo, Zeng, & Wu, 2007). This can be helpful in different fields from food quality control in industry to different biological research areas in academia.

### 2.3. Developed approach for quantifying mold growth

The *k*-means clustering algorithm was used to subdivide multispectral images of the mold colonies on petri dishes and quantify different segments of the mold colonies. A graphical illustration of the procedure is presented in Fig. 1. Images were acquired using the VideometerLab 2 instrument (Videometer A/S, Hørsholm, Denmark). In these images, each pixel is associated with a spectrum and can be considered as an object in a *w*-dimensional space, where *w* is number of the wavelengths. Each image data cube is unfolded into a matrix, where the *n* rows are the pixels and the *w* columns are the wavelengths ( $n \gg w$ ; see Fig. 1). Then, the *k*-means algorithm is used to cluster the pixels in this unfolded matrix, using the information from all the spectral bands. The concept is the same as

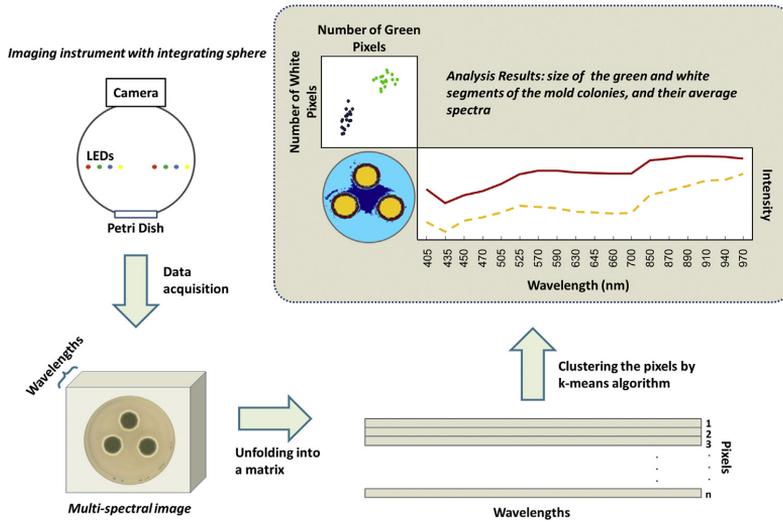


Fig. 1. The workflow for quantifying mold growth by multispectral images. The outputs of the analysis allow different samples to be quantified and discriminated.

clustering objects in e.g. a three-dimensional Cartesian coordinates, just expanded to a  $w$ -dimensional space. The final outputs of the analysis are quantification of different segments of the mold colonies in pixel-counts and the corresponding average spectrum for each segment. For instance, for the *Penicillium* molds that were used in this study, mold colonies consist of white and green segments. Results of the analysis are thus number of the white and green pixels in the mold colony and their average spectra (see Fig. 1). This allows quantifying the mold growth as well as the composition of the colony.

To perform the clustering and analysis of the images in a semi-automated way, a Graphical User Interface (GUI) has been developed using MATLAB 2012b (MathWorks, Inc., Natick, MA, USA). The layout of the GUI, which is called 'PCluster', is shown in Fig. 2. PCluster is specifically designed for *Penicillium* molds, for which the colonies are composed of white and green segments. However, the method and the explained concept can be used for all types of

molds. Multispectral images are first imported into PCluster, and then the user selects a circular region of interest which will be used for all the images in the imported set. In the  $k$ -means algorithm, the number of clusters,  $k$ , is a user defined input. In PCluster, images are clustered from 3 to 6 groups and the results are shown as color-coded (so-called *false negative*) image objects which show membership of the pixels in the clusters (see Fig. 2). Then, based on the graphical output, the optimum number of clustering and the meaningful clusters,  $k$ , are selected by the user. Meaningful clusters are the ones which show the colonies segments, and the optimum number of groups is the one which allows clustering different parts of the mold colony properly. For instance, in Fig. 2, partitioning the pixels into 4 clusters is enough to segment the white and green parts of the mold colony and separate them from the background. Selecting more than the optimum number of clusters will subdivide the colony segments further and will impose some errors on the quantification results, since only two colony segments can be

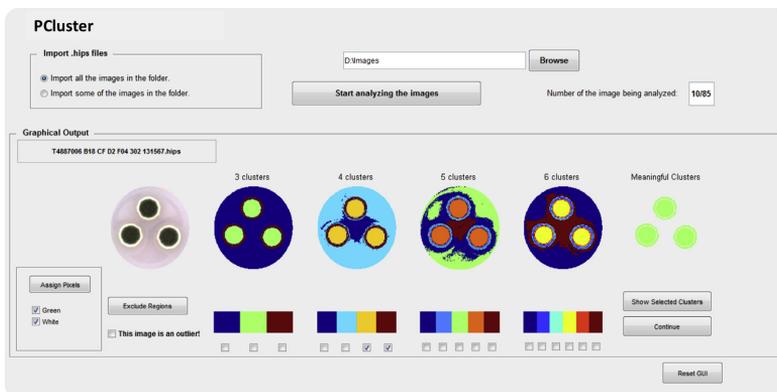


Fig. 2. The layout of the designed Graphical User Interface (PCluster).

chosen. The user can make a specific interpretation to the clusters, based on the color-coding of the pixels and select the meaningful clusters by ticking the corresponding check boxes (see Fig. 2). Although it may not seem so at the first sight, the procedure is quite simple and will come natural to the user in a short while.

This method is more accurate than the manual measurement of mold colony size. Fig. 3 shows two images of mold colonies and also the clustered images, which allow quantification of the different segments of the colonies as different clusters of pixels. It is obvious that manual measurement of these colonies in perpendicular directions and calculating the area would not provide the true size of the colonies, as the colonies do not grow quite symmetrically.

### 3. Materials and methods

#### 3.1. Microorganisms and culture conditions

The bacteria and molds which were used in this study are listed in Table 1. All bacterial strains were freeze dried and kept at  $-18^{\circ}\text{C}$  until use. Indicator molds were stored at  $-80^{\circ}\text{C}$ , in 20% glycerol and water, containing 0.1% Tween 80 (Merck).

#### 3.2. Ferments of chemically defined medium

Chemically defined interaction medium (CDIM) was prepared as described by (Aunsbjerg, Honoré, Vogensen, & Knöchel, 2015). In addition, an enhanced medium (CDIM<sup>+</sup>) was prepared by adding an extra agent to the standard medium (proprietary information, DuPont Nutrition Bioscience ApS, Brabrand, Denmark). Both media were inoculated with  $10^7$  CFU/mL of *Lactobacillus paracasei* DGCC 2132 (LAB A) or *L. paracasei* DGCC 11287 (LAB B) and incubated for 22 h at  $37^{\circ}\text{C}$ . All batches were made in triplicate. An overview of the batches and the average pH values are presented in Table 2. After fermentation, all the batches were centrifuged (5000 g, 15 min at  $5^{\circ}\text{C}$ ), followed by a filtration step through a  $0.45\ \mu\text{m}$  filter (Frisenette ApS, Knebel, Denmark). Two plates of cell-free ferments were prepared for each replicate, by mixing the extracts with agar (1%) and pouring into petri dishes.

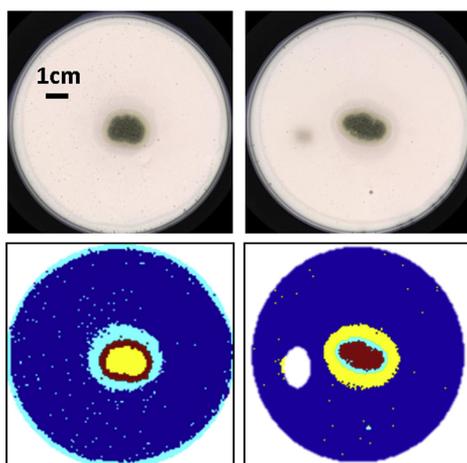


Fig. 3. Two examples of mold colonies (top) and the corresponding clustered images (bottom; clustered images are slightly magnified by the software).

**Table 1**  
Bacteria, commercial cultures and fungi used in the study and their incubation temperatures.

Microbial strain	Incubation temperature ( $^{\circ}\text{C}$ )	
	CDIM	Milk
<b>Antifungal bacteria</b>		
<i>Lactobacillus paracasei</i> DGCC 2132	37	43
<i>Lactobacillus paracasei</i> DGCC 11287	37	43
<b>Commercial cultures</b>		
YO-MIX <sup>®</sup> 410	–	43
HOLDBAC <sup>®</sup> YM-C	–	43
<b>Indicator fungi</b>		
<i>Penicillium</i> sp. DCS 1541	25	25
<i>Penicillium solitum</i> DCS 302	25	25
<i>Penicillium glabrum</i> DCS 305	25	25

**Table 2**  
An overview of the batches from the ferments of the chemically defined medium (CDIM).

Batches	Description	Average pH values
LAB A	<i>Lb. paracasei</i> DGCC 2132 in CDIM	4.54
LAB A <sup>+</sup>	<i>Lb. paracasei</i> DGCC 2132 in CDIM <sup>+</sup>	4.33
LAB B	<i>Lb. paracasei</i> DGCC 11287 in CDIM	5.72
LAB B <sup>+</sup>	<i>Lb. paracasei</i> DGCC 11287 in CDIM <sup>+</sup>	4.46

'Lb.' is the abbreviation for '*Lactobacillus*'.

#### 3.3. Acidified un-inoculated chemically defined medium samples

To investigate the influence of acidification with lactic acid on mold growth inhibition, un-inoculated samples were prepared by titrating CDIM with 80% DL-Lactic acid to pH values of 4.0, 4.5, 5.0, 5.5, 6.0 and 6.5 in the 6 samples. Subsequently, two plates of each sample were prepared by mixing with agar (1%) and pouring into petri dishes.

#### 3.4. Ferments of milk-based medium

Skim milk powder (Lactalis, Laval, France) was dissolved in distilled water (10% solution) and heat-treated at  $90^{\circ}\text{C}$  for 10 min. To prepare reference (control) batches, milk solutions were inoculated with 20 DCU (Internal dosing unit) of YO-MIX<sup>®</sup> 410 starter culture. For HOLDBAC<sup>®</sup> YM-C (HB) batches, in addition to 20 DCU of the starter culture, milk solutions were inoculated with 20 DCU of HOLDBAC<sup>®</sup> YM-C, which is an antifungal culture. Control batches were included to provide a reference to compare the HB batches with and investigate how the antifungal culture can enhance inhibition. Prior to inoculation, the pH of the heat-treated milk was approximately 6.5. After inoculation, both the control and HB batches were fermented to a pH value of approximately 4.6. Four biological replicates of reference and HB were prepared and incubated at  $43^{\circ}\text{C}$  for 8 h. Then, for each replicate, six plates of cell-containing ferments were prepared by mixing with agar (1%) and pouring into petri dishes.

#### 3.5. Mold inhibition test

Inhibition tests were performed using *Penicillium* sp. DCS 1541 and *Penicillium solitum* DCS 302 indicator fungi for CDIM batches, and *Penicillium* sp. DCS 1541 and *Penicillium glabrum* DCS 305 for the ferments of milk-based medium (see Table 1). In all the three sample sets, half of the prepared sample plates were used for each of the tested indicator fungi. Plates of CDIM and milk ferments were spotted with  $20\ \mu\text{L}$  of  $10^5$  spores/mL indicator molds. CDIM plates,

both inoculated and un-inoculated batches, were spotted in triplicate, whereas milk plates were spotted with a single spot. Ferments of CDIM, acidified un-inoculated samples and milk plates were incubated at 25 °C for 4, 5 and 6 days, respectively.

### 3.6. Multispectral image acquisition

A VideometerLab 2 spectral imaging instrument (Videometer A/S, Hørsholm, Denmark) was used to acquire the multispectral images. To record images by this instrument, the sample is placed inside the sphere of the instrument – so-called Ulbricht sphere – where diffused light from light emitting diodes (LEDs) is provided at 18 different wavelengths, ranging from 405 to 970 nm, and a single-channel image is recorded for each wavelength. The lid of the petri dishes was removed prior to image acquisition to avoid reflection. The size of all the acquired images was  $2056 \times 2056 \times 18$  and the size of each pixel, for the present configuration and instrumental settings, was  $45.8 \mu\text{m} \times 45.8 \mu\text{m}$ . For the ferments of CDIM, multispectral images were recorded after 2, 3 and 4 days of incubation of the indicator molds which were used to perform the inhibition tests. For the acidified samples, multispectral images were recorded after 3, 4 and 5 days of incubation of the indicator molds. The samples were taken out of the incubator, the images were recorded and then the samples were returned back to the incubator. Recording the images is quick, taking less than a minute per sample. For the milk ferments, multispectral images were recorded only once, after 6 days of incubation of the indicator molds. Some representative images (shown here as conventional RGB) of the CDIM and milk samples after incubation of the indicator molds are shown in Fig. 4.

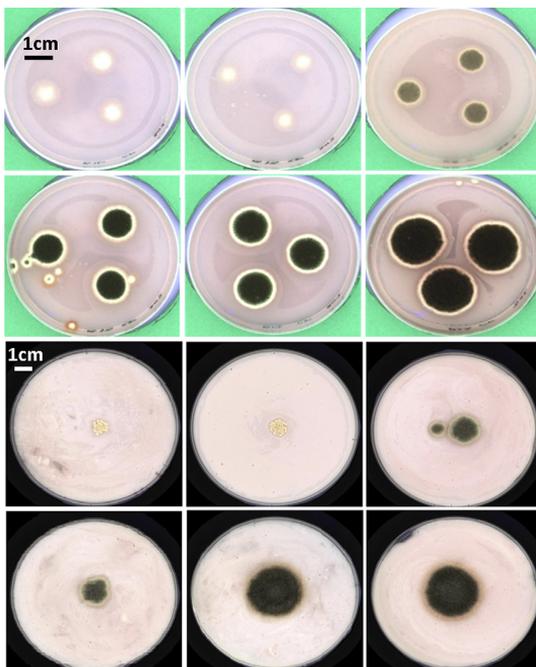


Fig. 4. Sample RGB images of the ferments from the chemically defined medium, spotted with *Penicillium* sp. DCS 1541 (top) and milk medium, spotted with *Penicillium glabrum* DCS 305 (bottom).

### 3.7. Analysis of multispectral images

The images were analyzed by PCluster, as detailed in Section 2.1. For all the acquired images, the number of the white and green pixels in the mold colony and their average spectra were calculated. The analysis time depends on the performance of central processing unit (CPU) of the computer and also the size of the images. For this study, the required time was on average a few minutes per image, using a standard office computer. Unwanted growth on some of the plates, caused by splashes of the mold solution while spotting (see Figs. 3 and 4), was removed from the images before quantification, using one of PCluster's built-in options. This allows saving some of the images while avoiding quantitative errors. PCluster, including the full MATLAB source code, is freely available from [www.models.life.ku.dk](http://www.models.life.ku.dk).

## 4. Results

In this section, the results of analyzing the images from the three designed experiments are presented and interpreted. The goal was to demonstrate the capability and potentials of the method in providing a quantitative comparison between different samples, in terms of mold growth and inhibition. The difference between the inhibition properties of the samples, although briefly discussed, is not the main interest of the article and is merely included to verify the performance of the method and provide examples of the analysis results.

### 4.1. Ferments of chemically defined medium

For the ferments of the CDIM, the images were analyzed with the aim of quantifying the size of the green and white segments of the mold colonies as a function of incubation days. The calculated sizes of the segments (in pixels unit) are shown in Fig. 5a. For all the batches and both tested molds, after 2 days of incubation, mold colonies are quite similar in size. Results show that at this stage, the colonies are only composed of white segments; green segments, which indicate a more advanced stage of the mold growth, have not appeared yet. After 3 days of incubation, mold colonies grow larger and, as the spores mature, they turn from white to green. According to the colony sizes, the inhibition of *Penicillium* sp. DCS 1541 is higher in comparison with *P. solitum* DCS 302. Moreover, the mold colonies grew less on the batches with the enhanced medium (LAB A<sup>+</sup> and LAB B<sup>+</sup>) compared to the batches with the standard medium (LAB A and LAB B). Among all the batches and for both of the indicator molds, the lowest inhibition is observed for LAB B. After 4 days of incubation, the growth patterns between the molds are discriminated better and the batches spotted with these molds are separated as two distinct groups in Fig. 5a. The average spectra of the green and white segments of the colonies, as shown in Fig. 5b, have different patterns and their intensities are clearly different in all the spectral channels, as expected.

### 4.2. Acidified un-inoculated chemically defined medium samples

For the acidified un-inoculated CDIM, the images were analyzed and the green and white segments of the mold colonies were quantified. The aim was to investigate and compare the growth patterns of the two indicator molds in response to pH variations. For both of the indicator molds, the correlation between the size of the green and white segments (and also the sum of the two segments) with pH of the samples and the growth of the colonies were examined. The best correlation was observed for the green segments and data from the white segments will not be shown or discussed further. Sizes of the green segments of the colonies, as a function of the pH of

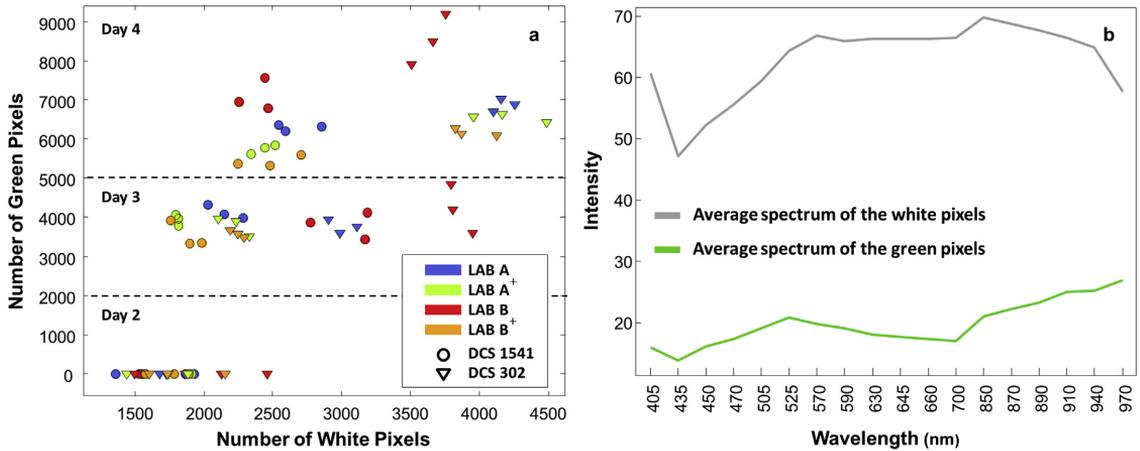


Fig. 5. Results of analyzing the images from the ferments of the chemically defined medium. a) size of the green and white segments of the colonies (in pixels unit); b) average spectra of the green and white segments. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the samples at different days of the incubation are shown in Fig. 6. For both of the indicator molds, size of the green segments in the colonies naturally increases by days of incubation. For *P. solitum* DCS 302, growth seems to be inhibited at lower pH values and this trend is observed for all days. By increasing the pH from 4.0 to 5.5, growth increases, whereas by increasing the pH further, the growth does not seem to be influenced. Contrary to *P. solitum* DCS 302, growth in *Penicillium* sp. DCS 1541 is not significantly affected by pH. For *Penicillium* sp. DCS 1541, the image of the sample with pH 6 in day 5 of incubation was not recorded properly and could not be analyzed.

4.3. Ferments of milk-based medium

For the ferments of milk-based medium, images were analyzed with the aim of quantifying green and white segments of the mold colonies. Results are shown in Fig. 7.

For control (reference) batches, which were only inoculated with the starter culture, mold colonies have grown much larger compared to HB batches, that in addition to the starter culture were inoculated with a culture known to have antifungal properties. On average, considering the total colony size regardless of the composition, *Penicillium* sp. DCS 1541 colonies on the control batches are 3 times larger than on the HB batches. For *P. glabrum* DCS 305, colonies are approximately 9 times larger on the control compared to the HB batches. In the control batches, *P. glabrum* DCS 305 has a significantly higher growth rate compared to *Penicillium* sp. DCS 1541. Sizes of the green and white segments of *P. glabrum* DCS 305 colonies are respectively around 5 and 3 times larger, when compared to *Penicillium* sp. DCS 1541. For HB batches, the colonies are only composed of white segments, and the more mature green spores have not appeared, even after 6 days of incubation. Although size of the colonies for the two types of the molds are quite close in HB batches,

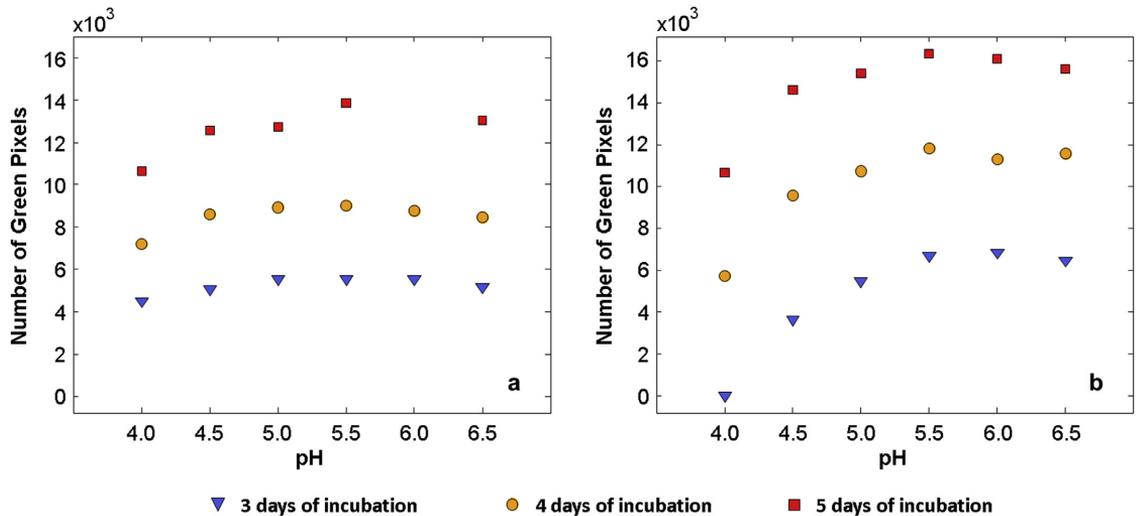


Fig. 6. Sizes of the green segments of the mold colonies (in pixels unit) for the acidified un-inoculated chemically defined medium samples. a) *Penicillium* sp. DCS 1541; b) *Penicillium solitum* DCS 302. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

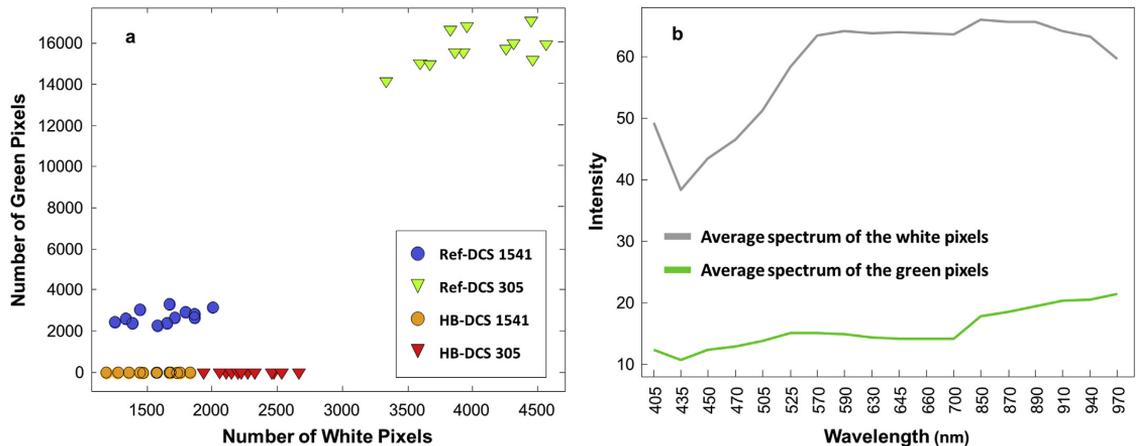


Fig. 7. Results of analyzing the images of the milk-based medium ferments. a) size of the green and white segments of the colonies (in pixels unit); b) average spectra of the green and white segments. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

they still appear as two separate groups in Fig. 7a. The average spectra of the green and white segments of the colonies are shown in Fig. 7b. As expected, white and green segments of the mold colonies have different spectral patterns and their intensities are also different in all the spectral channels.

## 5. Discussion

For the ferments of the CDIM, the developed approach – Pcluster – could accurately quantify mold growth, based on the area of the white and green segments of the mold colonies. The areas are presented in pixels unit. Pixel is the building unit of an image and for all of the images in the current study, the size of each pixel was  $45.8 \mu\text{m} \times 45.8 \mu\text{m}$ . As the number of the pixels and their size was constant between the images, pixel could easily be used as a measure of area for a quantitative comparison of mold growth. Converting number of the pixels to the more commonly used units of area, is just a matter of multiplying them by the area of each pixel. Differences between the growth patterns of the two indicator molds, at different days of incubation, could be discriminated. Moreover, the results could show how mold growth was influenced by the change in the medium.

For the acidified CDIM samples, influence of the pH on the growth of the indicator molds could be quantified. The method detected the differences between the two indicator molds in terms of their response to pH variation and their growth patterns. The variation which was created between the samples could be converted into objective measures. This can be very helpful in investigating how different parameters influence mold growth patterns.

The method could also quantify mold growth and size of the white and green segments in the ferments with the milk-based medium. Looking at the RGB images of some of the milk plates, the white edge of the mold colonies is not easily distinguishable on the white milk background (see Figs. 3 and 4), whereas the proposed method, benefitting from the advantages of multispectral imaging in combination with *k*-means clustering, could quantify the white sporulating segments reliably. For the HB batches, the differences in the size of the colonies for the two types of the indicator molds were small, and visual inspection of the plates would

not allow deciding which mold was inhibited more efficiently. The results also demonstrated how the mold growth and colony composition (relative size of the segments) differs between the HB and the control batches. The average spectra of the white and green segments have very different spectral patterns in both the CDIM and milk-based samples, and it is because of these different spectral patterns of the colony's segments that they can be reliably clustered as different groups by the *k*-means algorithm.

The method does not make any assumptions for the shape of the colonies, and can also quantify mold growth on transparent media, like the CDIM used in the first and second sets, as well as on opaque media, like the milk-based medium in the third set. Furthermore, the semi-automated analysis, using PCluster, can discriminate between the molds in a less labor-intensive and a more objective way. Following mold growth with appropriate imaging systems that allow comparative studies over time, and using the developed methodology for quantifying the colonies can be very useful in predictive microbiology and in studying how different parameters affect biopreservation of food products by antifungal bacterial cultures.

## Acknowledgments

The Danish Council for Strategic Research (Grant no. 10-095397) is acknowledged for generous financial support to the project entitled “microPAT” under the inSPIRe (Danish Industry–Science Partnership for Innovation and Research in Food Science) consortium (Copenhagen, Denmark). Furthermore, Dupont ApS (Brabrand, Denmark), formerly Danisco A/S, is acknowledged for financial support.

## References

- Aunshjerg, S. D., Honoré, A. H., Vogensen, F. K., & Knöchel, S. (2015). Development of a chemically defined medium for studying foodborne bacterial–fungal interactions. *International Dairy Journal*, 45, 48–55.
- Chaillou, S., Champomier-Verges, M.-C., Cornet, M., Crutz-Le Coq, A.-M., Dudez, A.-M., Martin, V., et al. (2005). The complete genome sequence of the meat-borne lactic acid bacterium *Lactobacillus sakei* 23K. *Nature Biotechnology*, 23, 1527–1533.

- Dissing, B., Papadopoulou, O., Tassou, C., Ersbøll, B., Carstensen, J., Panagou, E., et al. (2013). Using multispectral imaging for spoilage detection of pork meat. *Food and Bioprocess Technology*, 6, 2268–2279.
- García, D., Ramos, A. J., Sanchis, V., & Marín, S. (2010). Modelling mould growth under suboptimal environmental conditions and inoculum size. *Food Microbiology*, 27, 909–917.
- Guo, N., Zeng, L., & Wu, Q. (2007). A method based on multispectral imaging technique for white blood cell segmentation. *Computers in Biology and Medicine*, 37, 70–76.
- Mac Queen, J. (1976). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 281–297). University of California Press. No. 14.
- Magnusson, J., & Schnurer, J. (2001). *Lactobacillus coryniformis* subsp. *Coryniformis* strain Si3 produces a broad-spectrum proteinaceous antifungal compound. *Applied and Environmental Microbiology*, 67, 1–5.
- Marín, S., Cuevas, D., Ramos, A. J., & Sanchis, V. (2008). Fitting of colony diameter and ergosterol as indicators of food borne mould growth to known growth models in solid medium. *International Journal of Food Microbiology*, 121, 139–149.
- Marín, S., Morales, H., Ramos, A. J., & Sanchis, V. (2006). Evaluation of growth quantification methods for modelling the growth of *Penicillium expansum* in an apple-based medium. *Journal of the Science of Food and Agriculture*, 86, 1468–1474.
- Marín, S., Ramos, A. J., & Sanchis, V. (2005). Comparison of methods for the assessment of growth of food spoilage moulds in solid substrates. *International Journal of Food Microbiology*, 99, 329–341.
- Mohd, W. M. B. W., Beg, A. H., Herawan, T., & Rabbi, K. F. (2012). An improved parameter less data clustering technique based on maximum distance of data and Lloyd k-means algorithm. *Procedia Technology*, 1, 367–371.
- Taniwaki, M., Pitt, J., Hocking, A., & Fleet, G. (2006). Comparison of hyphal length, ergosterol, mycelium dry weight, and colony diameter for quantifying growth of fungi from foods. In A. D. Hocking, J. I. Pitt, R. Samson, & U. Thrane (Eds.), *Advances in food mycology* (Vol. 571, pp. 49–67). Springer US.
- Tran, T. N., Wehrens, R., & Buydens, L. M. C. (2005). Clustering multispectral images: a tutorial. *Chemometrics and Intelligent Laboratory Systems*, 77, 3–17.
- Wang, H. K., Yan, Y. H., Wang, J. M., Zhang, H. P., & Qi, W. (2012). Production and characterization of antifungal compounds produced by *Lactobacillus plantarum* IMAU10014. *PLoS One*, 7, e29452.

# Paper IV

---

## **Contribution of volatiles to the antifungal effect of *Lactobacillus paracasei* in defined medium and yogurt**

Stina D. Aunbjerg, Anders Honoré, Jorn Marcussen, Parvaneh Ebrahimi,  
Finn K. Vogensen, Connie Benfeldt, Thomas Skov, and Susanne Knøchel

*International Journal of Food Microbiology. Vol. 194, Pages 46-53, 2015*

---





## Contribution of volatiles to the antifungal effect of *Lactobacillus paracasei* in defined medium and yogurt



S.D. Aunbjerg<sup>a,1</sup>, A.H. Honoré<sup>a,b,1</sup>, J. Marcussen<sup>b</sup>, P. Ebrahimi<sup>a</sup>, F.K. Vogensen<sup>a</sup>, C. Benfeldt<sup>b</sup>, T. Skov<sup>a</sup>, S. Knøchel<sup>a,\*</sup>

<sup>a</sup> Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 26, 1958 Frederiksberg C, Denmark

<sup>b</sup> DuPont Nutrition Biosciences ApS, Edwin Rahrs Vej 38, 8220 Brabrand, Denmark

### ARTICLE INFO

#### Article history:

Received 13 June 2014

Received in revised form 17 October 2014

Accepted 5 November 2014

Available online 11 November 2014

#### Keywords:

Antifungal

Diacetyl

GC–MS

*Lactobacillus*

Molds

Volatiles

### ABSTRACT

Lactic acid bacteria with antifungal properties can be used to control spoilage of food and feed. Previously, most of the identified metabolites have been isolated from cell-free fermentate of lactic acid bacteria with methods suboptimal for detecting possible contribution from volatiles to the antifungal activity. The role of volatile compounds in the antifungal activity of *Lactobacillus paracasei* DGCC 2132 in a chemically defined interaction medium (CDIM) and yogurt was therefore investigated with a sampling technique minimizing volatile loss. Diacetyl was identified as the major volatile produced by *L. paracasei* DGCC 2132 in CDIM. When the strain was added to a yogurt medium diacetyl as well as other volatiles also increased but the metabolome was more complex. Removal of *L. paracasei* DGCC 2132 cells from CDIM fermentate resulted in loss of both volatiles, including diacetyl, and the antifungal activity towards two strains of *Penicillium* spp. When adding diacetyl to CDIM or yogurt without *L. paracasei* DGCC 2132, marked inhibition was observed. Besides diacetyl, the antifungal properties of acetoin were examined, but no antifungal activity was observed. Overall, the results demonstrate the contribution of diacetyl in the antifungal effect of *L. paracasei* DGCC 2132 and indicate that the importance of volatiles may have been previously underestimated.

© 2014 Elsevier B.V. All rights reserved.

### 1. Introduction

Antifungal lactic acid bacteria (LAB) have been studied in a range of foods and feed like sourdough (Black et al., 2013; Lavermicocca et al., 2000), dairy products (Schwenninger and Meile, 2004) fermented vegetables (Yang and Chang, 2010), and silage (Ström et al., 2002). While most of the efforts have been directed towards finding new potent strains, there is an increasing interest in understanding the antifungal mechanism including the identification and quantification of bioactive compounds produced by these strains. Several recent reviews exist on antifungal compounds produced by LAB (Crowley et al., 2013a; Dalié et al., 2010; Schnürer and Magnusson, 2005; Schwenninger et al., 2011). Phenyllactic acid (PLA) has been

reported as an antifungal compound of LAB in several publications. However, Ndagano et al. (2011) found the MIC value of PLA against *Aspergillus* spp. and *Penicillium* spp. to be far higher (180 mM) than the concentrations produced by LAB (0.1–0.5 mM). The same tendency was observed for lactic acid and acetic acid which were produced in concentrations much lower (44.8–76.8 and 1.2–7.5 mM, respectively) than the observed MIC values (>500 mM and 83–125 mM, respectively). This suggests that the antifungal effect is due to synergistic or additive effects of several compounds.

Schwenninger and Meile (2004) described the antifungal properties of a co-culture of *Lactobacillus paracasei* subsp. *paracasei* and *Propionibacterium jensenii* in fermented milk and cheese. They concluded that the inhibition was not solely based on the organic acids produced since acetic and propionic acids did not fully explain the antifungal effect (Schwenninger et al., 2008). Furthermore, some observations indicate a loss of antifungal activity upon cell removal (Schwenninger and Meile, 2004). This could indicate that some of the antifungal compounds disappear in cell-free fermentates, e.g. by being degraded, volatile or being consumed. Bacteria can produce a wide range of volatile organic compounds (Kai et al., 2009; Schulz and Dickschat, 2007). Several studies have shown the antifungal potential of some of these bacterial volatiles. *Pseudomonas* spp. isolated from canola and soybean plants produced the antifungal volatiles cyclohexanol, decanal, 2-Ethyl-1-hexanol, nonanal, benzothiazole and dimethyl

**Abbreviations:** AL,  $\alpha$ -Acetolactate; ALDC,  $\alpha$ -Acetolactate decarboxylase; AR, Acetoin reductase; CDIM, Chemically defined interaction medium; C-fermentate, Cell-containing fermentate; CF-fermentate, Cell-free fermentate; GC, Gas chromatography; GUI, Graphical User Interface; LAB, Lactic acid bacteria; LEDs, Light emitting diodes; MEA, Malt extract agar; MIC, Minimal inhibitory concentration; MS, Mass spectrometry; REF, Reference, acidified non-inoculated CDIM; SHS, Static headspace analysis; SPME, Solid phase microextraction; UHT, Ultra-high-temperature.

\* Corresponding author at: Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark. Tel.: +45 353 333258.

E-mail address: [skn@food.ku.dk](mailto:skn@food.ku.dk) (S. Knøchel).

<sup>1</sup> Stina D. Aunbjerg and Anders H. Honoré contributed equally to the manuscript.

trisulfide (Fernando et al., 2005). In another study, unidentified soil bacteria produced the antifungal volatiles trimethylamine, 3-Methyl-2-pentanone, dimethyl disulfide, benzaldehyde and *N,N*-Dimethyloctylamine (Chuankun et al., 2004). In spite of this, most of the studies on antifungal compounds produced by LAB strains focus on the non-volatile, liquid cell-free fermentate, often using bioassay-guided fractionation (Ström et al., 2002) and/or pre-concentration, e.g. extraction and drying (Schwenninger et al., 2008). These concentration techniques are well suited for concentrating compounds less volatile than water (or the solvent). However, compounds more volatile than water will be lost in the process. The aim of the current study was to examine the role of the major volatile compounds produced by the antifungal *L. paracasei* DGCC 2132. We screened fermentations in both chemically defined interaction medium (CDIM) and milk using qualitative and quantitative methods and minimal sample processing in order to elucidate the volatile profile. Diacetyl was identified as a major volatile compound. The antifungal activity of cell-containing as well as cell-free fermentates was tested towards selected fungal spoilers. The activity of cell-containing fermentate was also investigated after diacetyl formation had been inhibited by converting the precursor enzymatically. The antifungal effect of added diacetyl and acetoin was furthermore examined at concentrations comparable to those produced by *L. paracasei* DGCC 2132. The results help to explain the role of diacetyl in the antifungal effect of *L. paracasei* DGCC 2132 and highlight the likely importance of volatiles.

## 2. Materials and methods

### 2.1. Chemicals and materials

2,3-Butanedione (diacetyl) with purity 97%; 3-hydroxy-2-butanone (acetoin) with purity  $\geq 96\%$ ; Hydrochloric acid (37%) and 2-hydroxypropionic acid (DL-lactic acid)  $\geq 85\%$  were purchased from Sigma-Aldrich (Schneidorf, Germany).  $\alpha$ -Acetolactate decarboxylase, ALDC (3000 ADU/g) was supplied by DuPont Nutrition Biosciences ApS (Brabrand, Denmark). Tween 80 was from Merck, and UHT Milk was from MILSANI®.

### 2.2. Microbial strains, media and growth conditions

*L. paracasei* DGCC 2132 isolated from a dairy matrix was used in these studies. The strain had previously been identified by 16S rRNA sequencing. YO-MIX™ 410 starter culture (DuPont Nutrition Biosciences ApS, Denmark) was used for the production of yogurt. Freeze dried bacteria were stored at  $-18\text{ }^{\circ}\text{C}$  until use.

*Penicillium* sp. nov. DCS 1541 (tentative name *Penicillium salamii*, closely related to *Penicillium olsonii*, Centraalbureau voor Schimmelcultures, Fungal Biodiversity Centre) and *Penicillium solitum* DCS 302 were used as indicator molds since they had previously been isolated from spoiled fermented dairy products and initial results had shown different sensitivities towards antifungal LAB (unpublished results). Molds were grown on malt extract agar (Galloway and Burgess, 1952) (MEA, 30 g/L malt extract, 5 g/L peptone, 15 g/L agar) for 5–7 days and spores were harvested by adding water containing 0.01% Tween 80 (Merck). The harvested spores were supplemented with 20% glycerol (v/v) and stored at  $-80\text{ }^{\circ}\text{C}$  until use. All bacteria and mold strains were supplied by DuPont Nutrition Biosciences ApS, Brabrand, Denmark.

### 2.3. Culture conditions

To test the antifungal activity of *L. paracasei* DGCC 2132 in yogurt, UHT-milk (MILSANI®) was inoculated with 10 DCU/100 L YO-MIX™ 410 starter culture (DuPont Nutrition Biosciences ApS, Denmark) and  $10^7$  CFU/mL of *L. paracasei* DGCC 2132 followed by fermentation at

$43\text{ }^{\circ}\text{C}$  for 7 h. Yogurt without added *L. paracasei* DGCC 2132 was used as control.

A chemically defined interaction medium (CDIM) used for growth of *L. paracasei* DGCC 2132 and antifungal activity tests was prepared based on defined media previously reported for growth of fungi (Andersen et al., 2003; Bockelmann et al., 1999; Emeh and Marth, 1976; Hobot and Jennings, 1981; Meyers and Knight, 1958), LAB (Morishita et al., 1974; Møretro et al., 1998; Saguir and de Nadra, 2007; Savijoki et al., 2006) and other potential antifungal species such as propionic acid bacteria (Dherbécourt et al., 2008; Glatz and Anderson, 1988).

CDIM (200 mL) was inoculated with *L. paracasei* DGCC 2132 ( $10^7$  CFU/mL) in 250 mL blue cap flasks and fermented at  $37\text{ }^{\circ}\text{C}$  for 22 h to obtain a cell-containing fermentate (C-fermentate). pH was measured continuously in batches every 15 min during fermentation (Cinac, Alliance Instruments, Frepillon, France). All batches were made in triplicate. Cell-free fermentates (CF-fermentates) were prepared by centrifugation of C-fermentate ( $5000 \times g$ , 15 min at  $5\text{ }^{\circ}\text{C}$ ) followed by filtration of the supernatant through a  $0.45\text{ }\mu\text{m}$  filter (Frisenette, ApS). Un-inoculated CDIM kept at  $37\text{ }^{\circ}\text{C}$  for 22 h and acidified with lactic acid to pH 4.5 was used as reference (REF).

REF, C-fermentate, and CF-fermentate were tempered in a  $48\text{ }^{\circ}\text{C}$  water bath. After mixing with melted, tempered agar (1%), the media were poured into petri dishes. The plates were used after solidification and a short drying period ( $<3$  h).

### 2.4. Antifungal activity test

Antifungal activity of *L. paracasei* DGCC 2132 was tested by spotting  $20\text{ }\mu\text{L}$  of spore dilution ( $10^5$  spores/mL) of each mold in triplicate on plates of yogurt and plates of REF, C-fermentate and CF-fermentate and incubating at  $25\text{ }^{\circ}\text{C}$  for 9 days. Mold growth was documented by recording and analyzing multispectral images with the objective of quantifying area of the mold colonies.

Contribution of volatiles to antifungal activity was assessed in a “plate-on-plate” test system without direct contact between molds and C-fermentate. A REF plate was spotted with  $20\text{ }\mu\text{L}$  of a mold spore dilution ( $10^5$  spores/mL). On top of the REF plate a C-fermentate plate or a REF plate (control) was placed upside down and sealed with Parafilm® M. The inhibitory activity was assessed by growth on REF plates after 4 days of incubation at  $25\text{ }^{\circ}\text{C}$ .

### 2.5. Acquisition and analysis of multispectral images

A VideometerLab 2 spectral imaging instrument (Videometer A/S, Hørsholm, Denmark) was used to record objective and reproducible images of the petri dishes with spotted mold. Images were recorded after 2, 3, 4, 5, 6 and 9 days of incubation. To record multispectral images by the VideometerLab, the sample was placed inside the sphere of the instrument (Ulbricht sphere) where diffused light from light emitting diodes (LEDs) was provided at 18 different wavelengths, ranging from 375 to 970 nm. A single-channel image was recorded for each wavelength. The size of all the acquired images was  $2056 \times 2056 \times 18$  and the lid of the petri dishes was removed prior to image acquisition to avoid reflection.

The images were subsequently analyzed using PCluster, an in-house MATLAB Graphical User Interface (GUI) developed by Ebrahimi et al. (unpublished results). PCluster is specifically designed for *Penicillium* molds, for which the colonies are often composed of white and green segments; however, its main idea and the concept can be used for quantifying all types of molds. PCluster clusters the pixels in the multispectral images with the objective of quantifying mold growth. The outputs of PCluster are the size (in pixels unit) of the green and white segments of the mold colonies and their average spectra. In the current study, quantification of mold growth was based on the total size (area)

of the colonies by summing up the number of pixels in the white and green segments of the colonies.

## 2.6. Analysis of volatiles by Gas Chromatography Mass Spectrometry

Volatiles were measured by headspace Gas Chromatography Mass Spectrometry (GC–MS) configured for two types of headspace sampling. For qualitative screening and semi-quantitative work, headspace solid-phase microextraction (HS–SPME) was employed. For quantitative purposes, static headspace sampling (SHS) was used.

Using the two methods, three types of samples were analyzed: Liquid fermentations in CDIM (SPME + SHS), liquid fermentations in milk (SPME) and solid CDIM agar plugs (SPME + SHS).

### 2.6.1. Liquid calibration and sampling

Calibration solutions were prepared by acidifying CDIM with DL-lactic acid to pH 4.5 and adding diacetyl in the concentrations 0, 10, 25, 50, 100 and 200 µg/mL. The solutions were stored for a maximum of 24 h at 5 °C. Aliquots of 2 mL of both calibration solutions and CDIM inoculated with *L. paracasei* DGCC 2132 were transferred to 20 mL headspace vials and sealed. Real-time development of volatiles during fermentation at 37 °C and 43 °C in liquid samples of milk and at 37 °C in liquid samples of CDIM, was monitored by having a set of subsamples of the inoculated medium placed in the GC autosampler, and the headspace was sampled alternately from the vials. Samples were maintained at 37 °C (for CDIM fermentations) or at 37 °C and 43 °C (for milk fermentations) during headspace sampling and in between analyses by SHS GC/MS.

### 2.6.2. Solid calibration and sampling

Calibration plates with diacetyl were prepared by acidifying CDIM with DL-lactate to pH 4.5 tempering at 48 °C in a water bath and adding diacetyl in the concentrations 0, 10, 30 and 90 µg/g, respectively, followed by mixing with 48 °C agar (1%). The calibration range was selected according to the observed production during fermentations. The development of volatiles in plates of C-fermentate or volatiles present in plates of CF-fermentate was monitored by daily sampling of agar plugs from plates placed in a 25 °C incubator. Plugs were taken from a sample plate (C- or CF-fermentate) or a calibration plate by using an inverted sterile 1000 µL pipette tip (BioHit, Sartorius, Helsinki, Finland) with vacuum suction. The plug was transferred to a 20 mL headspace vial, sealed and frozen at –18 °C. Collected agar plugs were stored at –18 °C for a maximum 14 days prior to analysis by SHS GC/MS.

Headspace analyses were performed using a CTC PAL autosampler (CTC Analytics, Zwingen, Switzerland) mounted on an Agilent 6890 GC coupled to an Agilent 5973N single quadrupole MSD (Agilent Technologies, Waldbronn, Germany). The CTC PAL could be configured for both solid phase microextraction headspace analysis and static headspace analysis. Both the CTC PAL and Agilent GC/MS were controlled by Agilent MS Chemstation E. 02.00.49. Details on SPME and SHS instrumental conditions are listed in Table A.1.

## 2.7. Antifungal activity of diacetyl and acetoin

Antifungal activity of different concentrations of diacetyl and acetoin in CDIM (0, 10, 45, 60, 75 and 200 µg/mL) was tested at high (6.5) and low (4.5) pH levels, respectively. DL-Lactic acid or HCl was used to acidify CDIM to pH 4.5. Plates were prepared by mixing with tempered agar (1%) after the addition of diacetyl and acids. Antifungal activity of different concentrations of diacetyl (0, 45, 75 and 200 µg/mL) was tested in yogurt and milk acidified with DL-lactic acid to pH 4.5. Plates of CDIM, milk and yogurt were spotted with 20 µL of spore dilution (105 spores/mL) of each mold in triplicate and incubated at 25 °C for up to 20 days.

## 2.8. Reduction of inhibition by addition of $\alpha$ -acetolactate decarboxylase

Diacetyl is formed from a spontaneous chemical oxidation and decarboxylation of the metabolic intermediate  $\alpha$ -acetolactate (AL) formed from pyruvate (Fig. 1). Alternatively, AL can be converted to acetoin by  $\alpha$ -acetolactate decarboxylase (ALDC) (Kleerebezem et al., 2000; von Wright and Axelsson, 2011). To promote the formation of acetoin from AL rather than diacetyl to reduce the inhibitory effect, 400 µg/mL ALDC was added to CDIM prior to inoculation of *L. paracasei* DGCC 2132 ( $10^7$  CFU/mL). Plates of C-fermentate with and without added ALDC were made by mixing with agar (1%). Antifungal activity was tested against *P. solitum* DCS 302 as described in Subsection 2.4.

## 2.9. Production of volatiles in yogurt with and without *L. paracasei* DGCC 2132

UHT-milk at 5 °C was inoculated with 10 DCU/100 L YO-MIX™ 410 starter culture (DuPont Nutrition Biosciences ApS, Denmark) and with 10 DCU of both starter culture and  $5 \times 10^6$ – $10^7$  CFU/mL *L. paracasei* DGCC 2132. After mixing,  $6 \times 5$  mL aliquots were pipetted to 20 mL headspace vials and capped. The vials were placed in an autosampler thermostated at 37 °C. The headspace of the vials was sampled alternately with a cycle time of 20 min and analyzed by HS–SPME–GC/MS as described in Subsection 2.6.

## 3. Results

### 3.1. Antifungal activity of bacterial fermentate with and without cells

*L. paracasei* DGCC 2132 showed antifungal properties against *P. solitum* DCS 302 and *Penicillium* sp. nov. DCS 1541 in both CDIM and yogurt (Figs. 2 and 5). The cell-containing CDIM fermentate (C-fermentate) showed antifungal activity towards both molds, with *Penicillium* sp. nov. DCS 1541 being more sensitive. The cell-free CDIM fermentate (CF-fermentate) had little effect on the indicator mold growth (Fig. 5b, c).

In the plate-on-plate test system, with no direct contact between C-fermentate and indicator molds, both molds were completely inhibited for 4 days, results of *P. solitum* DCS 302 are shown as an example (Fig. 3),

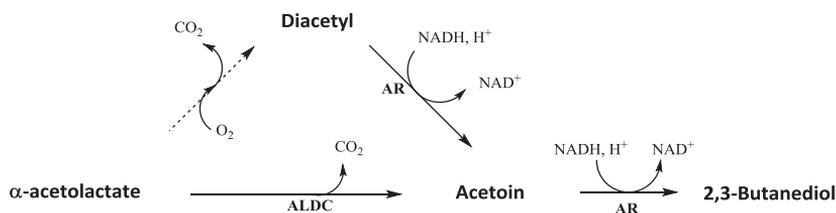
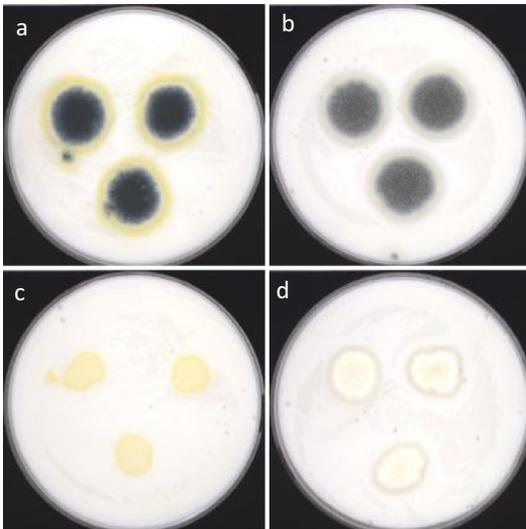


Fig. 1. Schematic biochemical pathway for diacetyl and acetoin in lactic acid bacteria modified from Von Wright and Axelsson (2011). Enzymes marked in bold, ALDC:  $\alpha$ -acetolactate decarboxylase, AR: acetoin reductase.



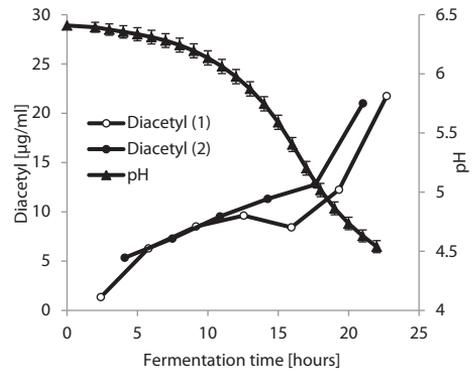
**Fig. 2.** Antifungal activity of *L. paracasei* DGCC 2132 in yogurt fermented at 43 °C for 7 h. Indicator molds: *Penicillium solitum* DCS 302 (a, c) and *Penicillium* sp. nov. DCS 1541 (b, d). Yogurt without (a, b) and with added *L. paracasei* DGCC 2132 (c, d). Images were recorded after 4 days incubation at 25 °C.

strongly indicating the inhibitory role of some volatiles. Methylene blue was added to C-fermentate as a redox indicator to test if suppressed mold growth was due to oxygen exhaustion. The color of the C-fermentate remained blue throughout incubation thereby indicating that lack of oxygen was unlikely to be the reason for mold growth inhibition in the plate-on-plate test system (results not shown).

### 3.2. Diacetyl production and inhibition by C-/CF-fermentate

Chemical analysis of the headspace of CDIM during fermentation with *L. paracasei* DGCC 2132 identified diacetyl as the primary volatile compound measured. Production of diacetyl was detected within the first few hours of fermentation with an increased production rate after 20 h of fermentation (Fig. 4). Prolongation of fermentation time increased diacetyl concentration further (data not shown). Besides diacetyl, traces of acetoin were observed in the CDIM fermentation.

To test if decreased antifungal activity of CF-fermentate compared to C-fermentate was due to difference in the volatile content, the diacetyl was measured in plugs from C-fermentate as well as CF-fermentate plates after 0, 1, 2, 4, 6, and 9 days of incubation at 25 °C. Diacetyl concentration increased in C-fermentate from 58 to 74 µg/g after 1 day of



**Fig. 4.** Diacetyl formation and pH change during initial fermentation of *L. paracasei* DGCC 2132 in CDIM at 37 °C. pH was measured continuously in 5 replicates while diacetyl was measured during fermentation in two replicates.

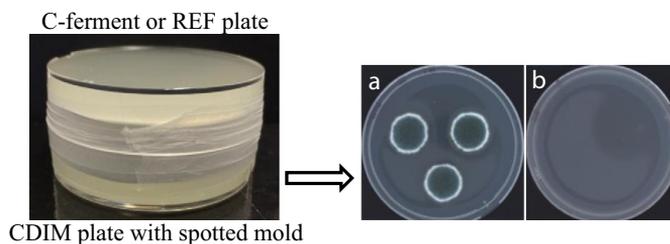
incubation and then decreased upon further incubation. Diacetyl concentration in the CF-fermentate was 9 µg/g at all days (Fig. 5a).

Growth of the two molds was quantified based on the size of the colonies calculated from the multispectral images of REF, CF-fermentate and C-fermentate plates during incubation (Fig. 5b and c). Growth of both molds was similar on CF-fermentate and un-inoculated REF plates. Plates with CF-fermentate were overgrown after 5 and 6 days for *Penicillium* sp. nov. DCS 1541 and *P. solitum* DCS 302, respectively. In contrast, the presence of LAB cells in C-fermentate caused a delay in the onset of growth until day 3 for *P. solitum* DCS 302 and until day 4 for *Penicillium* sp. nov. DCS 1541, as well as a markedly reduced growth rate of *Penicillium* sp. nov. DCS 1541.

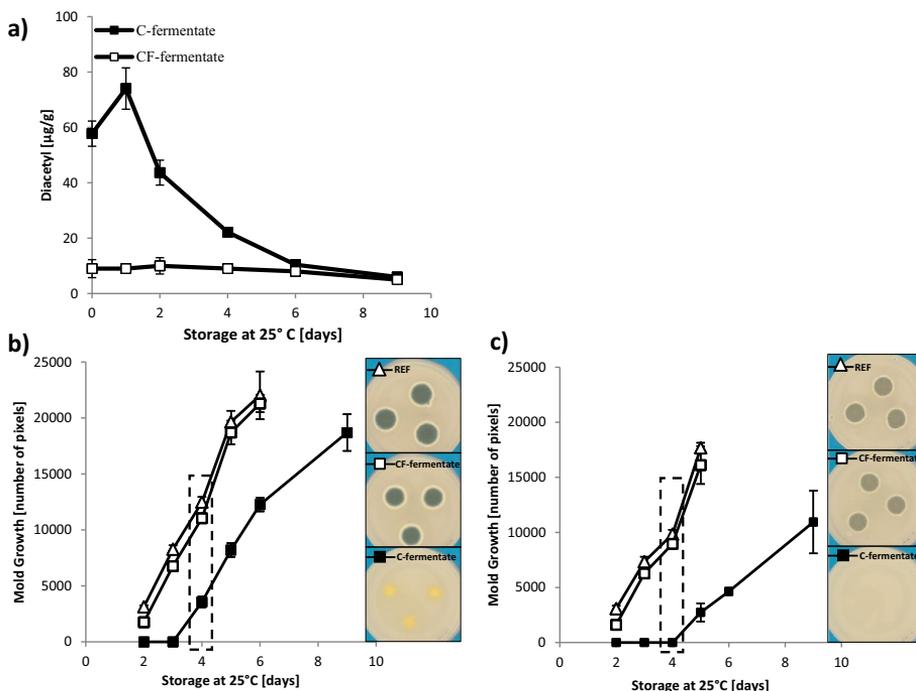
### 3.3. Antifungal activity of diacetyl and acetoin

Diacetyl was added to CDIM in concentrations corresponding to those produced by *L. paracasei* DGCC 2132 as well as in higher concentrations as used in previous mold inhibition studies (Jay, 1982a). Increasing diacetyl concentration correlated with increased inhibition of indicator molds, with *Penicillium* sp. nov. DCS 1541 being the most sensitive target organism. Mold growth was markedly inhibited on un-inoculated CDIM when diacetyl was added in the concentration corresponding to the amount present in C-fermentate at days 0 and 2 (~60 and 45 µg/mL, respectively) (Fig. 6). At 75 µg/mL diacetyl mold growth was suppressed for up to 5 days (data not shown).

*Penicillium* sp. nov. DCS 1541 was more sensitive to diacetyl at pH 4.5, whereas *P. solitum* DCS 302 was more sensitive at pH 6.5. The



**Fig. 3.** Influence of volatiles in *L. paracasei* DGCC 2132 C-fermentate (plate on top) on growth of *Penicillium solitum* DCS 302 spotted in triplicates on a REF plate (bottom). Control: The C-fermentate plate was replaced with a REF plate. Growth of mold on REF plate with a C-fermentate plate on top (b) and a REF plate on top (a), respectively, was assessed visually after incubation at 25 °C/4 days.



**Fig. 5.** a) Diacetyl content in agar plugs of C-fermentate and CF-fermentate, b) growth of *P. solitum* DCS 302, c) and *Penicillium* sp. nov. DCS 1541 (average  $\pm$  standard deviation,  $n = 3$ ) in REF (i.e. CDIM at pH 4.5), C-fermentate and CF-fermentate plates over 9 days. Next to graphs: Images of mold growth on day 4. Day 4 is marked on the graphs by dotted rectangles. Plates were incubated at 25 °C.

use of HCl to acidify CDIM showed same results as lactic acid (data not shown). If high amounts (200 µg/mL) of diacetyl were added to either CDIM, milk or yogurt, *P. solitum* DCS 302 did not grow until day 14, and no growth of *Penicillium* sp. nov. DCS 1541 was observed after 20 days (data not shown).

Acetoin did not show antifungal activity towards the two molds at any of the concentrations tested (data not shown).

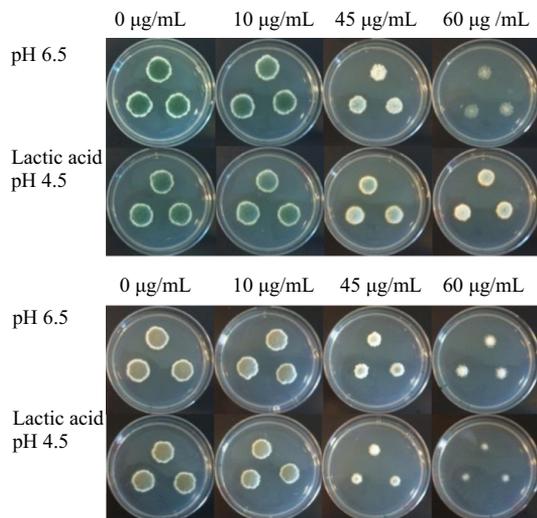
#### 3.4. Reduction of inhibition by addition of $\alpha$ -acetolactate decarboxylase

ALDC has been used in the brewing industry to remove diacetyl which is considered an off flavor in beer (Yamano et al., 1995). ALDC was added to CDIM prior to inoculation with *L. paracasei* DGCC 2132 in order to promote the formation of acetoin from AL and thereby decreasing the formation of diacetyl during fermentation. The addition of ALDC did not influence growth of molds on un-inoculated plates of CDIM (data not shown). Upon the addition of ALDC to the C-fermentate, the inhibitory effect of the fermentate decreased (Fig. 7).

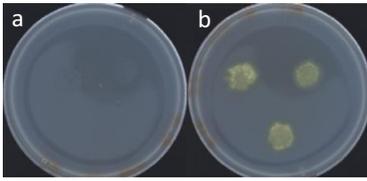
#### 3.5. Production of volatiles in yogurt with and without *L. paracasei* DGCC 2132

The volatiles in the headspace of yogurt with and without added *L. paracasei* DGCC 2132 were measured by SPME-GC/MS in order to follow their formation when the antifungal *L. paracasei* DGCC 2132 was added to yogurt. Two temperatures were used for fermentation, 37 °C and 43 °C. A number of volatiles were detected in both types of yogurt. When *L. paracasei* DGCC 2132 was added together with the starter culture at 37 °C, some of the easily identifiable changes were increased amounts of diacetyl and acetoin but also some relative increases in

potential antifungal compounds like 2,3-pentadione, acetic acid, and butanoic acid were observed (results not shown). This reflects a more complex volatile profile due to the milk medium and the additional



**Fig. 6.** Influence of lactic acid and diacetyl (0, 10, 45 and 60 µg/mL) on growth of *P. solitum* DCS 302 (top) and *Penicillium* sp. nov. DCS 1541 (bottom) on CDIM after 4 days of incubation at 25 °C.



**Fig. 7.** 400  $\mu\text{g/mL}$  ALDC was added to CDIM prior to inoculation with *L. paracasei* DGCC 2132 to decrease formation of diacetyl. C-fermentate (a) and C-fermentate with added ALDC (b). The influence of added ALDC on inhibition of *P. solitum* DCS 302 was tested after fermentation at 37 °C/24 h. Plates were incubated at 25 °C until mold growth appeared.

effect of the starter culture. In yogurts fermented at 43 °C the addition of *L. paracasei* DGCC 2132 markedly increased the formation of diacetyl (Fig. 9), whereas no marked difference was seen for other volatile compounds (Fig. 8). Storage of yogurts at 5 °C for 9 days increased diacetyl production markedly to  $13.5 \pm 0.54 \mu\text{g/mL}$  in yogurts with added *L. paracasei* DGCC 2132 compared to lower levels of  $3.1 \pm 0.20 \mu\text{g/mL}$  in yogurt without the antifungal strain.

### 3.6. Antifungal activity of diacetyl in yogurt and acidified milk

The antifungal activity of yogurt with added diacetyl against *P. solitum* DCS 302 and *Penicillium* sp. nov. DCS 1541 supported the results obtained in CDIM (Fig. 10). Similar results were observed for milk acidified with lactic acid and added diacetyl (results not shown).

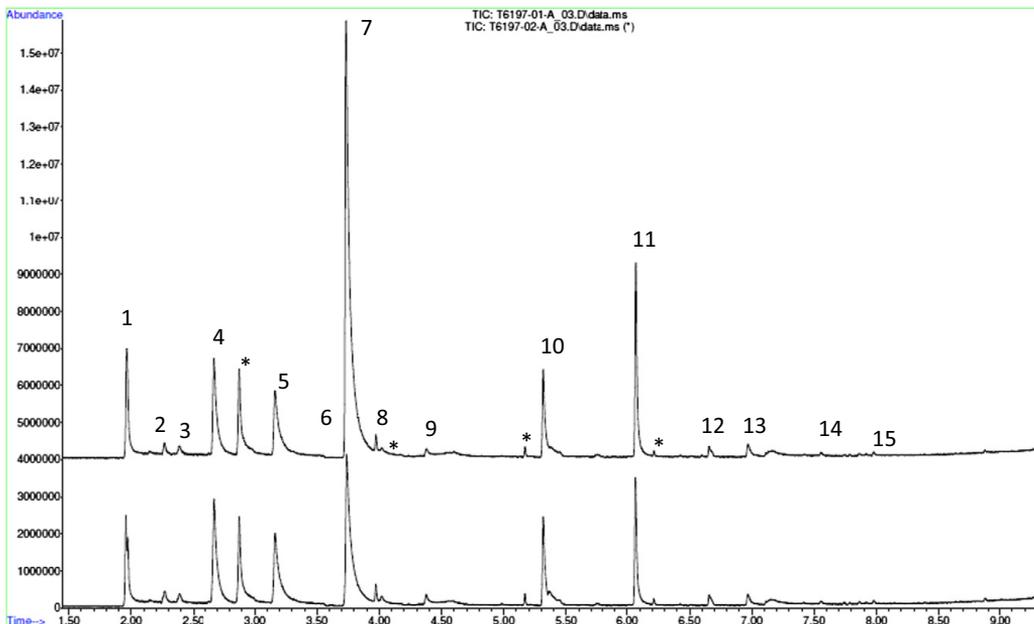
## 4. Discussion

Many earlier studies of bacterial cultures with antifungal effect have focused on identification of potent antifungal compounds in cell-free

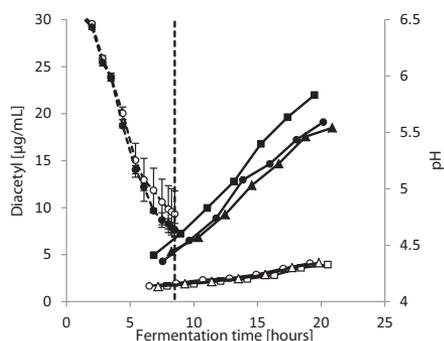
fermentates (Crowley et al., 2013b; Magnusson and Schnürer, 2001; Rouse et al., 2008; Schwenninger et al., 2008). However, we observed loss of antifungal activity of culture fermentates after cell removal, an observation also made by Schwenninger and Meile (2004). Since we observed antifungal activity in a plate-on-plate test with no direct contact between C-fermentate and mold, volatiles seemed to be involved in the antifungal activity. We moreover found that most of the volatile diacetyl produced during fermentation disappeared when preparing cell-free fermentates. A reason for this could be that volatiles are easily lost in the steps involved in cell removal e.g. centrifugation, filtration and heating of fermentate prior to mixing with agar. If this is not taken into account, the contribution of volatiles to the antifungal activity could be overlooked or underestimated.

In CDIM, the main volatile compound measured in the headspace of C-fermentate during fermentation was diacetyl. Diacetyl production increased rapidly at pH below 5 and increasing fermentation time was found to further increase diacetyl (data not shown). This is in agreement with a study on *Lactobacillus casei* by Branan and Keenan (1971) who observed a rapid increase in diacetyl content at pH below 5.5 with the highest production measured between pH 4.5 and 5.5. In CDIM, the concentration of diacetyl was 22  $\mu\text{g/g}$  after the initial 22 h of fermentation and it rose to 58  $\mu\text{g/g}$  during the preparation of C-fermentate plates indicating continuous production of diacetyl by *L. paracasei* DGCC 2132.

When molds were spotted on CDIM with levels of diacetyl corresponding to those found in the plates of C-fermentate (~60  $\mu\text{g/g}$ ) inhibition was observed for up to 3 days. Diacetyl levels corresponding to those found in CF-fermentate (~10  $\mu\text{g/g}$ ) showed no antifungal activity. This indicated that the high diacetyl concentration in C-fermentate was the main cause of the inhibitory activity (Fig. 6). The addition of diacetyl as a single compound to yogurt and acidified milk showed inhibitory activity against the two molds with results similar to CDIM. The results from yogurt are shown as an example (Fig. 10).



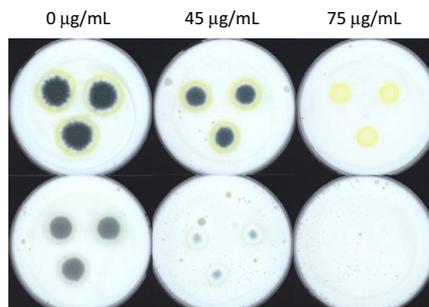
**Fig. 8.** Total ion chromatogram of headspace of yogurt fermentations after 7.2 h at 43 °C with added *L. paracasei* DGCC 2132 (top) and without *L. paracasei* DGCC 2132 (bottom). Peak annotations are 1:  $\text{CO}_2$ , 2: Acetaldehyde, 3: Dimethylsulfide, 4: Acetone, 5: 2-Butanone, 6: 2-Pentanone, 7: Diacetyl, 8: Ethanol, 9: 2,3-Pentadione, 10: 2-Heptanone, 11: Acetoin, 12: 2-Nonanone, 13: Acetic acid, 14: Propanoic acid, and 15: Butanoic acid, \*: System peaks.



**Fig. 9.** Formation of diacetyl (solid lines) and pH change (dotted lines) during fermentation of yogurts at 43 °C with (closed markers) and without (open markers) *L. paracasei* DGCC 2132. The markers, □, ○ and Δ, represent three replicates. The vertical line shows the end of fermentation.

Diacetyl is a common volatile metabolite in fermented milk products (Ott et al., 2000) with reported concentrations between 0.045 and 27 µg/g depending on the product (Rincon-Delgado et al., 2012). The presence of diacetyl and acetoin contributes positively to the buttery or creamy flavor perception but the range of sensory acceptable levels varies between different dairy products. In environments, where workers are constantly inhaling very high peak levels of diacetyl, such as in microwave popcorn production, exposure to high concentrations of diacetyl has been associated with serious respiratory illness (SCOEL, 2014). However, consumption of fermented milk products is not expected to present any risk since both the concentration of naturally produced diacetyl in the dairy products and the vapor pressure at consumption temperature are very low and the exposure is infrequent. Yogurt is a complex medium which, in itself, contains numerous volatile compounds apart from diacetyl after fermentation. When *L. paracasei* DGCC 2132 was added together with the starter culture at 37 °C a marked increase in diacetyl and acetoin was observed. In addition, potential antifungal compounds like 2,3-pentadione, acetic acid, and butanoic acid slightly increased. Co-fermentation of the *L. paracasei* DGCC 2132 strain with a yogurt starter culture at 43 °C demonstrated a marked increase in diacetyl as compared to yogurt without *L. paracasei* DGCC 2132 (Fig. 9), whereas other volatile compounds did not increase (Fig. 8). The diacetyl production was, however, lower in yogurt than in CDIM due to the shorter fermentation time of yogurt. Diacetyl levels continued to increase with longer fermentation time (Fig. 9) as well as during storage at 5 °C. The addition of the antifungal *L. paracasei* DGCC 2132 greatly increased the concentration of diacetyl in yogurt. The measured levels may still be too low to explain all of the inhibition, indicating involvement of synergistic or additive interactions with other compounds. Another aspect could be that the employed sampling techniques do not measure diacetyl associated with the milk/yogurt matrix and therefore unavailable for measurement in the headspace although it may still contribute to the antifungal effect. Further studies could include analysis of the liquid phase after dilution with acetone (De Leonardi et al., 2013; Macciola et al., 2008).

The antifungal properties of diacetyl as an isolated compound were assessed as early as 1941 towards several molds including *Penicillium* sp. and *Fusarium* sp. with inhibition observed at concentrations above 86 µg/mL (Lagoni, 1941). Other studies showed that molds and yeast were sensitive to diacetyl concentrations of 100 and 200 µg/mL, respectively (Jay, 1982a, 1982b), with increased activity at low pH for some molds (Jay, 1982a). The mechanism behind the inhibition of mold growth has not been elucidated and little is known of interactions with other compounds. Here, we found that the *Penicillium* strains were totally inhibited for up to 5 days in CDIM with added diacetyl



**Fig. 10.** Influence of 0, 45 and 75 µg/mL diacetyl added to yogurt on growth of *P. solitum* DCS 302 (top) and *Penicillium* sp. nov. DCS 1541 (bottom). Images were recorded after 4 days of incubation at 25 °C.

concentrations above 75 µg/mL and, furthermore, that the influence of the pH was strain dependent.

Suomalainen and Mäyrä-Mäkinen (1999) found that levels of diacetyl and acetic acid increased in a bacterial culture mix with activity against yeast and *Bacillus* spp. While the amount of acetic acid doubled, the diacetyl levels increased dramatically in yogurt and in quark from <0.5 to 24 and from 0.6 to 49 µg/mL, respectively. The addition of the protective culture increased the sensory properties of the products which the authors explained by the increased diacetyl production. The inhibitory effect was not investigated, but it was assumed, based on the data from Jay (1982b), that the diacetyl levels would be insufficient to account for the antifungal activity. Similar statements are found in several reviews on antifungal LAB (Caplice and Fitzgerald, 1999; Schnürer and Magnusson, 2005) and a recent review on antifungal compounds produced by LAB does not mention diacetyl at all (Crowley et al., 2013a). This may reflect that previous approaches have underestimated the potential contribution of diacetyl to the antifungal activity of LAB strains. Apart from the loss of volatiles in many sampling operations, some caution in the interpretation of volatile production from fermentations should be exerted since the equilibrium between the headspace and liquid phase is influenced by temperature, pH and other factors (Pawliszyn, 1997). In this study, two headspace sampling techniques were employed to follow the dynamics of volatiles independent of the matrix. For initial profiling of the headspace, SPME was used as an unbiased, sensitive screening technique and SHS was used for absolute quantification of e.g. diacetyl. There may, however, still be changes in compounds which are not readily detected.

The differences in diacetyl production between LAB strains may be the absence or decreased activity of the acetoin reductase (AR), which reduces diacetyl to acetoin or could be due to the absence or suppression of ALDC, resulting in the formation of diacetyl instead of acetoin from accumulated AL (Kleerebezem et al., 2000). ALDC was added to CDIM prior to inoculation with *L. paracasei* DGCC 2132 in order to test if the low antifungal activity observed in CF-fermentate was caused by a decrease in diacetyl content after cell removal. The addition of ALDC markedly decreased the antifungal activity of the C-fermentate, although the activity was not completely abolished. The expected higher amounts of acetoin caused by the ALDC addition could not explain the remaining antifungal activity of C-fermentate when ALDC was added, since acetoin showed no antifungal activity at the relevant concentrations. Jay et al. (1983) also found low antifungal activity of acetoin. We observed remaining activity after the ALDC addition which could be a result of antifungal metabolites other than diacetyl or inefficient removal of all diacetyl by the added ALDC.

Lactic acid bacteria known as diacetyl producers have been associated with antifungal properties, but no causal link between the antifungal effect of these cultures and their diacetyl production has previously been documented. Here, we link the production of volatile diacetyl by

*L. paracasei* DGCC 2132 with inhibition of *Penicillium* strains thereby pointing at a previously overlooked contribution to the anti-mold effect of LAB. A multitude of bacterial metabolites, both volatile and non-volatile, may influence potential mold growth and the interactions can be complex and depending on many factors. The use of cultures with continuous production of diacetyl and other synergistically active compounds may have a strong potential as clean label ingredients in products prone to fungal spoilage but further knowledge is needed to optimize cultures for specific products and spoilage organisms.

## Acknowledgment

This work was financed by the University of Copenhagen (692886) as a scholarship for Stina Dissing Aunbjerg and by the Danish Ministry of Science, Innovation and Higher Education (FI 10-084173). We would also like to thank Heike Morgenstern, Senior Scientist, DuPont Nutrition Biosciences ApS for providing both cultures and laboratory support.

## References

- Andersen, D., Renshaw, J.C., Wiebe, M.G., 2003. Rhodotorulic acid production by *Rhodotorula mucilaginosa*. *Mycol. Res.* 107, 949–956.
- Black, B.A., Zannini, E., Curtis, J.M., Gänzle, M.G., 2013. Antifungal hydroxy fatty acids produced during sourdough fermentation: microbial and enzymatic pathways, and antifungal activity in bread. *Appl. Environ. Microbiol.* 79, 1866–1873.
- Bockelmann, W., Portius, S., Lick, S., Heller, K.J., 1999. Sporulation of *Penicillium camemberti* in submerged batch culture. *Syst. Appl. Microbiol.* 22, 479–485.
- Branen, A.L., Keenan, T.W., 1971. Diacetyl and acetoin production by *Lactobacillus casei*. *Appl. Microbiol.* 22, 517–521.
- Caplice, E., Fitzgerald, G.F., 1999. Food fermentations: role of microorganisms in food production and preservation. *Int. J. Food Microbiol.* 50, 131–149.
- Chuankun, X., Minghe, M., Leming, Z., Kejin, Z., 2004. Soil volatile fungistasis and volatile fungistatic compounds. *Soil Biol. Biochem.* 36, 1997–2004.
- Crowley, S., Mahony, J., van Sinderen, D., 2013a. Current perspectives on antifungal lactic acid bacteria as natural bio-preservatives. *Trends Food Sci. Technol.* 33, 93–109.
- Crowley, S., Mahony, J., van Sinderen, D., 2013b. Broad-spectrum antifungal-producing lactic acid bacteria and their application in fruit models. *Folia Microbiol.* 58, 291–299.
- Dalié, D.K.D., Deschamps, A.M., Richard-Forget, F., 2010. Lactic acid bacteria – potential for control of mould growth and mycotoxins: a review. *Food Control* 21, 370–380.
- De Leonardis, A., Lopez, F., Nag, A., Macciola, V., 2013. Occurrence and persistence of diacetyl in unfermented and fermented milks. *Eur. Food Res. Technol.* 236, 691–697.
- Dherbécourt, J., Maillard, M.-B., Catheline, D., Thierry, A., 2008. Production of branched-chain aroma compounds by *Propionibacterium freudenreichii*: links with the biosynthesis of membrane fatty acids. *J. Appl. Microbiol.* 105, 977–985.
- Ebrahimi, P., van der Berg, F., Aunbjerg, S.D., Honoré, A.H., Benfeldt, C., Jensen, H.M., Engelsens, S.B., n.d. unpublished results. Quantitative determination of mold growth and inhibition by multispectral imaging.
- Emeh, C.O., Marth, E.H., 1976. Growth and synthesis of rubratoxin by *Penicillium rubrum* in a chemically defined medium fortified with organic acids and intermediates of the tricarboxylic acid cycle. *Mycopathologia* 59, 137–142.
- Fernando, W.G.D., Ramarathnam, R., Krishnamoorthy, A.S., Savchuk, S.C., 2005. Identification and use of potential bacterial organic antifungal volatiles in biocontrol. *Soil Biol. Biochem.* 37, 955–964.
- Galloway, L.D., Burgess, R., 1952. *Applied Mycology and Bacteriology*, 3rd ed. Leonard Hill, London.
- Glatz, B.A., Anderson, K.I., 1988. Isolation and characterization of mutants of *Propionibacterium* strains. *J. Dairy Sci.* 71, 1769–1776.
- Hobot, J.A., Jennings, D.H., 1981. Growth of *Debaryomyces hansenii* and *Saccharomyces cerevisiae* in relation to pH and salinity. *Exp. Mycol.* 5, 217–228.
- Jay, J.M., 1982a. Effect of diacetyl on foodborne microorganisms. *J. Food Sci.* 47, 1829–1831.
- Jay, J.M., 1982b. Antimicrobial properties of diacetyl. *Appl. Environ. Microbiol.* 44, 525–532.
- Jay, J.M., Rivers, G.M., Biosvert, W.E., 1983. Antimicrobial properties of  $\alpha$ -dicarbonyl and related compounds. *J. Food Prot.* 46, 325–329.
- Kai, M., Hausteiner, M., Molina, F., Petri, A., Scholz, B., Piechulla, B., 2009. Bacterial volatiles and their action potential. *Appl. Microbiol. Biotechnol.* 81, 1001–1012.
- Kleerebezem, M., Hols, P., Hugenholtz, J., 2000. Lactic acid bacteria as a cell factory: rerouting of carbon metabolism in *Lactococcus lactis* by metabolic engineering. *Enzym. Microb. Technol.* 26, 840–848.
- Lagoni, H., 1941. Vergleichende Untersuchungen über den mikrobiciden Effekt verschiedener Konservierungsmittel. *Zentralblatt fuer Bakteriologie, Parasitenkunde und Hygiene, Abteilung 2* (103), 225–231.
- Lavermicocca, P., Valerio, F., Evidente, A., Lazzaroni, S., Corsetti, A., Gobetti, M., 2000. Purification and characterization of novel antifungal compounds from the sourdough *Lactobacillus plantarum* strain 21B. *Appl. Environ. Microbiol.* 66, 4084–4090.
- Macciola, V., Candela, G., De Leonardis, A., 2008. Rapid gas-chromatographic method for the determination of diacetyl in milk, fermented milk and butter. *Food Control* 19, 873–878.
- Magnusson, J., Schnürer, J., 2001. *Lactobacillus coryniformis* subsp. *coryniformis* strain SI3 produces a broad-spectrum proteinaceous antifungal compound. *Appl. Environ. Microbiol.* 67, 1–5.
- Meyers, E., Knight, S.G., 1958. Studies on the nutrition of *Penicillium roqueforti*. *Appl. Microbiol.* 6, 174–178.
- Møretø, T., Hagen, B.F., Axelsson, L., 1998. A new, completely defined medium for meat lactobacilli. *J. Appl. Microbiol.* 85, 715–722.
- Morishita, T., Fukada, T., Shirota, M., Yura, T., 1974. Genetic basis of nutritional requirements in *Lactobacillus casei*. *J. Bacteriol.* 120, 1078–1084.
- Ndagano, D., Lamoureux, T., Dortu, C., Vandemoten, S., Thonart, P., 2011. Antifungal activity of 2 lactic acid bacteria of the *Weissella* genus isolated from food. *J. Food Sci.* 76, M305–M311.
- Ott, A., Germond, J.-E., Chaintreau, A., 2000. Vicinal diketone formation in yogurt: branched-chain amino acids  $^{13}\text{C}$  precursors and effect of branched-chain amino acids. *J. Agric. Food Chem.* 48, 724–731.
- Pawliszyn, J., 1997. SPME method development. *Solid Phase Microextraction: Theory and Practice*. Wiley-VCH, New York, pp. 97–140.
- Rincon-Delgado, M.I., Lopez-Hernandez, A., Wijaya, I., Rankin, S.A., 2012. Diacetyl levels and volatile profiles of commercial starter distillates and selected dairy foods. *J. Dairy Sci.* 95, 1128–1139.
- Rouse, S., Harnett, D., Vaughan, A., van Sinderen, D., 2008. Lactic acid bacteria with potential to eliminate fungal spoilage in foods. *J. Appl. Microbiol.* 104, 915–923.
- Saguir, F.M., de Nadra, C.M., 2007. Improvement of a chemically defined medium for the sustained growth of *Lactobacillus plantarum*: nutritional requirements. *Curr. Microbiol.* 54, 414–418.
- Savijoki, K., Suokko, A., Palva, A., Varmanen, P., 2006. New convenient defined media for [ $^{35}\text{S}$ ]methionine labelling and proteomic analyses of probiotic lactobacilli. *Lett. Appl. Microbiol.* 42, 202–209.
- Schnürer, J., Magnusson, J., 2005. Antifungal lactic acid bacteria as biopreservatives. *Trends Food Sci. Technol.* 16, 70–78.
- Schulz, S., Dickschat, J.S., 2007. Bacterial volatiles: the smell of small organisms. *Nat. Prod. Rep.* 24, 814–842.
- Schwenninger, S.M., Meile, L., 2004. A mixed culture of *Propionibacterium jensenii* and *Lactobacillus paracasei* subsp. *paracasei* inhibits food spoilage yeasts. *Syst. Appl. Microbiol.* 27, 229–237.
- Schwenninger, S.M., Lacroix, C., Truttmann, S., Jans, C., Spörndli, C., Bigler, L., Meile, L., 2008. Characterization of low-molecular-weight antiyeast metabolites produced by a food-protective *Lactobacillus-Propionibacterium* coculture. *J. Food Prot.* 71, 2481–2487.
- Schwenninger, S.M., Meile, L., Lacroix, C., 2011. Antifungal lactic acid bacteria and propionibacteria for food biopreservation. In: Lacroix, C. (Ed.), *Protective Cultures, Antimicrobial Metabolites and Bacteriophages for Food and Beverage Biopreservation*. Woodhead Publishing, Oxford, pp. 27–62.
- SCOEL, 2014. Recommendation from the Scientific Committee on Occupational Exposure Limits for Diacetyl. *SCOEL/SUM/149*, June 2014. *Eur. Comm.* 1–31.
- Ström, K., Sjögren, J., Broberg, A., Schnürer, J., 2002. *Lactobacillus plantarum* MiLAB 393 produces the antifungal cyclic dipeptides cyclo(*l*-Phe-*l*-Pro) and cyclo(*l*-Phe-trans-4-OH-*l*-Pro) and 3-phenyllactic acid. *Appl. Environ. Microbiol.* 68, 4322–4327.
- Suomalainen, T.H., Mäyrä-Mäkinen, A.M., 1999. Propionic acid bacteria as protective cultures in fermented milks and breads. *Lait* 79, 165–174.
- Von Wright, A., Axelsson, L., 2011. Lactic acid bacteria: an introduction. In: Lahtinen, S., Ouwehand, A.C., Salminen, S., von Wright, A. (Eds.), *Lactic Acid Bacteria – Microbiological and Functional Aspects*. CRC Press, London, pp. 1–16.
- Yamano, S., Tomizuka, K., Sone, H., Imura, M., Takeuchi, T., Tanaka, J., Inoue, T., 1995. Brewing performance of a brewer's yeast having  $\alpha$ -acetylacetyl decarboxylase from *Acetobacter aceti* subsp. *xylinum*. *J. Biotechnol.* 39, 21–26.
- Yang, E.J., Chang, H.C., 2010. Purification of a new antifungal compound produced by *Lactobacillus plantarum* AF1 isolated from kimchi. *Int. J. Food Microbiol.* 139, 56–63.

# Paper V

---

## **Metabolic footprinting for Investigation of antifungal properties of *Lactobacillus paracasei***

Anders Honoré, Stina D. Aunsbjerg, Parvaneh Ebrahimi, Michael Thorsen,  
Connie Benfeldt, Susanne Knøchel, and Thomas Skov

*Analytical and Bioanalytical Chemistry, submitted*

---



# Metabolic Footprinting for Investigation of antifungal properties of *Lactobacillus paracasei*

Anders H. Honoré<sup>1, 2, #</sup>, Stina D. Aunsbjerg<sup>2</sup>, Parvaneh Ebrahimi<sup>2</sup>, Michael Thorsen<sup>1</sup>, Connie Benfeldt<sup>1</sup>, Susanne Knøchel<sup>2</sup>, Thomas Skov<sup>2</sup>

1: DuPont Nutrition Biosciences Aps, Denmark

2: Department of Food Science, Faculty of Science, University of Copenhagen

## #: Corresponding author:

Anders H. Honoré, DuPont Nutrition Biosciences Aps, Denmark, Edwin Rahrs vej 38, DK-8220 Brabrand, Denmark, E-mail: anders.honore@dupont.com, Phone: +45 89435348

## Running Title:

Metabolic footprinting for Investigation of Antifungal Properties

## ABSTRACT

Some lactic acid bacteria with antifungal properties are applied for bio-preservation of food. For understanding their antifungal mechanism, there is an ongoing search for the bioactive molecules. With focus on formed metabolites, bioassay guided fractionation and comprehensive screening have identified compounds as antifungal. Although active, the compounds have been found in concentrations too low to account for the effect. It has been hypothesized that the formation of metabolites and consumption of nutrients during bacterial fermentations form the basis for the antifungal effect, *i.e.* the composition of the exometabolome. To build a more comprehensive view of the chemical changes induced by bacterial fermentation and the effects on mold growth, a strategy for correlating the exometabolomic profiles to mold growth was applied. The antifungal properties were assessed by measuring mold growth of two *Penicillium* strains on cell free ferments of three strains of *Lactobacillus paracasei* pre-fermented in a chemically defined medium. Exometabolomic profiling was performed by reversed phase liquid chromatography in combination with mass spectrometry in electrospray positive and negative modes. By multivariate data analysis, the three strains of *Lb. paracasei* were readily distinguished by the relative difference of their exometabolomes. The relative differences correlated to the relative growth of the two *Penicillium* strains. Metabolic footprinting proved as a supplement to bioassay guided fractionation for investigation of antifungal properties of bacterial ferments. Additionally, three previously identified and three novel antifungal metabolites from *Lb. paracasei* and their potential precursors were detected and assigned using the strategy.

## Key words:

Untargeted profiling, LC/MS, multivariate, minimal inhibitory concentration, antifungal

## LIST OF ABBREVIATIONS

ArAT	Aromatic aminotransferase
BCAA	Branched chain amino acid
BcAT	Branched chain aminotransferase
BPC	Base peak chromatogram
CDIM	Chemically defined interaction medium
CF	Cell free ferment
Da	Dalton
ESI	Electrospray ionization
FID	Flame ionization detector
GC	Gas chromatography
id	Internal diameter
ID	Inhibition degree
IS	Internal standard
LC	Liquid chromatography
<i>m/z</i>	mass-to-charge ratio
MIC	Minimal inhibitory concentration
MS	Mass spectrometry
Neg	Negative, as for negative electrospray mode
OD	Optical density
PC	Principal Component
PCA	Principal Component analysis
PLSR	Partial least squares regression
Pos	Positive, as for positive electrospray mode
ppm	Parts per million
REF	Reference, un-inoculated substrate
UPLC	Ultra performance liquid chromatography
TIC	Total ion chromatogram
VIP	Variable importance in Projection

## INTRODUCTION

The use of lactic acid bacteria as an additional safety is widely used in food products to help controlling fungal spoilage. Much work has been invested in identifying the antifungal metabolites produced by lactic acid bacteria. Focus has been on bioassay guided fractionation [1], i.e. with a hypothesis of finding one or several compounds responsible for the antagonistic effect e.g. [2–5]. Over time, numerous compounds have been assigned as antifungal as compiled in recent comprehensive reviews [6, 7]. Although the activity of compounds have been shown, the level produced in biological system has been below minimal inhibitory concentration (MIC) [8–10]. Some studies have shown that combinations of metabolites increase the antifungal activity of ferments which could indicate synergistic effects between metabolites [10–12]. The strategy of bioassay guided fractionation is especially suited for detecting one or few potent antagonistic compounds, but has shortcomings when the effect is composed.

Recent studies have employed more comprehensive analytical screening strategies with a minimum of sample preparation or fractionation [13, 14]. The workflow excludes sample preparation steps like solid phase extraction (SPE), where loss of bioactive compounds have been observed [13, 15]. Instead, untreated or pre-concentrated cell free ferments (CFs) are studied on hyphenated high performance chromatography – high-resolution mass spectrometry analytical platforms. This enables recognition and quantification of multiple known compounds [13].

Within the potential of state-of-the-art analytical platforms, the analytical coverage can be expanded to go beyond the recognized compounds to include all available signals. Then, the study becomes an untargeted or data-driven study of the CF or exometabolome of lactic acid bacteria. Studies of an extensive range of extracellular metabolites including residual nutrients have been defined as exometabolomic footprinting [16]. Exometabolomic studies hold potential for new learning related to the antifungal properties of lactic acid bacteria. The mode of action of a new antibiotic compound was elucidated using this strategy [17]. Paczia et al. (2012) demonstrated the dynamics of *extended* overflow metabolism showing passive and active transportation of central metabolic intermediates to the extracellular environment during batch fermentations [18]. The change in metabolism of branched chain amino acids (BCAA) as stress induced by acidic conditions was examined for *Lactobacillus sanfranciscensis* [19].

By adopting the workflow of untargeted metabolomics, the present study aimed at providing a supplementary tool for investigating the antifungal properties of three strains of *Lactobacillus paracasei*. The three *Lb. paracasei* strains were fermented in a chemically defined interaction medium (CDIM). The CFs were subjected to metabolic footprinting and tested for effect on growth of two *Penicillium* indicator strains. With this approach, the dynamics for both nutrient availability and formation of metabolites was taken into account. The data driven approach applied multivariate methods for identifying biomarkers from correlations between compounds and biological effect. The correlations were converted into hypotheses for targeted studies and tested in model systems to test for causal links between compound concentrations and biological activity.

## MATERIALS AND METHODS

### Chemicals

Solvents and chemicals, including formic acid, acetic acid, propionic acid, butanoic acid, 2-ethylbutanoic acid, lactic acid (80%), 2-hydroxy-3-methylbutanoic acid (= '2-hydroxy-isovaleric acid '), 2-hydroxy-3-phenylpropanoic acid (= 'phenyllactic acid'), 2-hydroxy-3-(4-hydroxyphenyl)propanoic acid (= '(4-hydroxyphenyl)lactic acid'), 2-hydroxy-3-(1H-indol-3-yl)propanoic acid (= 'indolelactic acid'), 2-hydroxy-4-methylpentanoic acid (= '2-hydroxyisocaproic acid'), 2-hydroxy-(4-methylthio)butanoate Calcium salt, inosine, hypoxanthine, L-phenyl-D<sub>5</sub>-alanine, sodium chloride, sulfuric acid, diethyl ether, acetonitrile and ethanol were purchased from Sigma-Aldrich (Schneelldorf, Germany) with purity in excess of 95% or as *pro analysis* quality unless otherwise specified. All water employed was of freshly prepared Milli-Q quality (Merck Millipore, Billerica, MA, USA).

### Microbial strains

All strains used in these studies were supplied by DuPont Nutrition Biosciences ApS. The strains and their growth conditions are listed in Table 1.

**Table 1 Overview of microbial strains and their growth conditions.**

Microbial strain	Short name	Growth medium	Incubation temperature
<i>Lactobacillus paracasei</i> DGCC 2132	LAB A	MRS	37°C
<i>Lactobacillus paracasei</i> DGCC 11287	LAB B	MRS	37°C
<i>Lactobacillus paracasei</i> DGCC 695	LAB C	MRS	37°C
<i>Penicillium solitum</i> DCS 302	DCS 302	MEA	25°C
<i>Penicillium</i> sp. nov. DCS 1541*	DCS 1541	MEA	25°C

**MRS**, Man Rogosa Sharpe (Oxoid)

**MEA**, Malt extract agar [20] (30 g/L malt extract (Becton Dickinson), 5 g/L peptone (Becton Dickinson), 15 g/L agar (Becton Dickinson))

\*Tentative name *Penicillium salamii*, closely related to *Penicillium olsonii*, Centraalbureau voor Schimmelcultures, Fungal Biodiversity Centre

### Preparation of *Lactobacillus paracasei* cell free ferments (CFs)

The three strains of *Lb. paracasei* were pre-inoculated in MRS and incubated at 37°C overnight. The cells were harvested by centrifugation followed by washing of cells twice in 0.9% NaCl. A chemically defined interaction medium (CDIM) was prepared as described by Aunbjerg et al. [21]. Blue cap bottles (250 ml) with CDIM were inoculated with washed cells of the three strains to an optical density at 600 nm (OD<sub>600</sub>) of 0.05. Five biological replicates were made of each batch. Batches were placed in a 37°C water bath with CINAC pH electrodes monitoring pH every 15 minutes throughout fermentation. Aliquots were taken from each batch after 65 hours of fermentation and placed on ice. The biomass of a 45 ml sample aliquot was determined as the dry mass after filtration through Advantec GB140 (Toyo Roshi Kaisha, Ltd, Japan). CFU/ml and OD<sub>600</sub> was in addition measured after 65 hours of fermentation.

CFs were prepared by centrifuging at 4500 g for 10 min at 0°C (Heraeus XR3 Multifuge, Thermo Fisher Scientific, Waltham, MA, USA) and the supernatant was filtered through a sterile 0.45 µm pore size filter (cellulose acetate membrane, Q-Max, Frisenette Aps, Knebel, Denmark). Aliquots (1000 µl) of the filtrate were distributed into eppendorf tubes for liquid chromatography/mass spectrometry (LC/MS) analysis, frozen on dry ice and stored at -80°C.

### **Headspace Solid-Phase Microextraction Gas chromatography Mass Spectrometry for Monitoring Fermentation Volatiles**

The ferments were analyzed for the volatile profile by headspace solid phase microextraction gas chromatography with mass spectrometric detection (HS SPME-GC/MS). The method was as described in Aunbjerg et al. [22]. In brief, the analysis was performed using a CTC PAL autosampler (CTC Analytics, Zwingen, Switzerland) mounted on an Agilent 6870 GC coupled to Agilent 5973N single quadrupole MSD (Agilent Technologies, Waldbronn, Germany). Both the CTC pal and Agilent GC/MS were controlled by Agilent MSD Chemstation E. 02.00.49. Immediately after inoculation, an aliquot of 2 ml were transferred to 20 ml headspace vials. The samples were placed on the autosampler thermostated to 37°C and the headspaces were alternately sampled by SPME with an extraction time of 900 seconds. The GC was mounted with a 60m x 0.25mm internal diameter (id) Agilent J&W (Santa Clara, CA) GC column DB-1701 (cross-linked and surface bonded 14% cyanopropyl-phenyl/86% dimethylpolysiloxane) with a film thickness of 1 µm. The carrier gas was helium flowing at 2.0 ml/minutes. The SPME fiber (85 µm Carboxen/polydimethylsiloxane (Supelco, Bellefonte, PA) was desorbed for 30 seconds with a split ratio of 1:5 split/split less injector kept at 260°C. The oven was initially held at 60°C for 2 min then the temperature was increased to 110°C with a rate of 10°C/minute and then to 240°C with 20°C/minute. The transfer line and MS was held at 260°C and 230°C, respectively. The ionization voltage was 70 eV, and MS scan range was mass-to-charge ( $m/z$ ) 29-300.

### **Mold growth assay**

Mold growth on *Lb. paracasei* CF: CFs and un-inoculated CDIM acidified to pH 4.5 with DL-lactic acid (REF) were mixed with agar (1%) and poured into petri dishes. Antifungal activity of 3 of 5 biological replicates of the three *Lb. paracasei* CFs and REF was tested against *P. solitum* DCS 302 and *Penicillium* sp. nov. DCS 1541. The molds were spotted (20 µL of 10<sup>5</sup> spores/ml) on plates in triplicates and incubated at 25°C for 2-5 days.

Multispectral images of *Lb. paracasei* ferments were recorded on day 2-5 on a daily basis using a VideometerLab 2 spectral imaging instrument (Videometer A/S, Hørsholm, Denmark). Diffused light from light emitting diodes (LEDs) were provided at 18 different wavelengths, ranging from 375 to 970 nm. Single-channel images were recorded for each of the 18 wavelengths (from 375 to 970 nm). The instrument was mounted with a JAI (JAI A/S, Valby Denmark) model BM-500GE 32 bit camera with a resolution of 2056 x 2056 resulting in a pixel size of 45.8 µm. The lid of the petri dishes was removed prior to image acquisition to avoid reflection. The multispectral images were subsequently analyzed using a modified version of PCluster, a GUI developed under MATLAB environment (Version 2012b, MathWorks, Inc., Natick, MA, USA) by Ebrahimi et al. [23]. In the modified version of PCluster, the pixels in the multispectral images were clustered with the objective of quantifying different segments of the mold colonies, and as a result, quantifying mold

growth and mold color of colonies. The outputs of the analysis were number of the white and green segments in the colonies and their average spectra. Mold growth was calculated as the sum of white and green pixels.

### **Minimal inhibitory concentrations of identified compounds**

The six identified 2-hydroxy acids (Table 4) were tested for their minimal inhibitory concentration (MIC) against the two *Penicillium* strains. Solutions of CDIM acidified with DL-lactic acid (pH 4.5) were prepared with 0 (REF), 0.1, 1, 5 and 10 mg/ml of all six compounds, respectively. The test solutions were distributed into a sterile flat-bottomed 96 well microplate (Fisher Scientific) and  $10^4$  spores/ml of each *Penicillium* strain were added to the wells. Triplicate determinations were made. Microplates were incubated at 25°C for up to 48 hours. OD<sub>600</sub> at start ( $t_0$ ) and after 48 hours ( $t_f$ ) was recorded at 600 nm using a Varioskan Flash (Thermo Fisher Scientific Oy, Finland). The inhibition degree (ID) was calculated and evaluated as [10]:

$$ID = [OD_{600}(\text{test solution}, t_f) - OD_{600}(\text{test solution}, t_0)] / [OD_{600}(\text{Reference}, t_f) - OD_{600}(\text{Reference}, t_0)]$$

The minimal inhibition concentration (MIC 50) was defined as the concentration providing ID < 0.5 [10].

### **Metabolic footprinting by Liquid Chromatography Mass Spectrometry (LC/MS)**

Frozen CFs were thawed in an Eppendorf Thermomixer Comfort (Eppendorf Nordic Aps, Horsholm, Denmark) for 10 minutes at 20°C with 750 rpm. A pooled sample (MIX) was prepared by sampling and mixing 200 µl of each CF included in the study. Aliquots of 200 µl of the CF, the MIX sample or Milli-Q water (solvent blank) were diluted with 800 µl 0.1%v/v formic acid in water containing 0.06 mmol/l L-phenyl-D<sub>5</sub>-alanine as internal standard (IS). The diluted solution was centrifuged at 13300 x g for 10 minutes prior to analysis (Spectrafuge Labnet International Inc. Edison, NJ, USA). The LC/MS system was equilibrated with a minimum of ten replicate injections of the MIX sample prior to analyzing samples. Sample injections were performed in duplicate. Each set of replicates were placed in randomized brackets containing ten samples (representing two replicates of the five biological treatments) and one MIX sample. Each injection bracket was initiated by a solvent blank (Milli-Q water) treated as sample. The LC/MS analysis was performed using an Agilent (Agilent Technologies, Waldbronn, Germany) modular 1290 ultra performance liquid chromatography (UPLC) instrument coupled to a Bruker (Bruker Daltonics, Billerica, MA) maXis 4G single quadrupole time-of-flight mass spectrometer (MS) via an electrospray interface. The UPLC was mounted with a Waters (Waters Corporation, Milford, Ma, USA) HSS T3, 2.1 x 150 mm id column packed with 1.8 µm particles. Mobile phases were A) water/formic acid 1000/1 v/v and B) acetonitrile/formic acid 1000/1 v/v. Vials were kept at 5°C in the autosampler prior to injection of 10 µl. Elution was performed with a flow of 400 µl/min and a gradient starting at 0% B at t=0 and kept for 1.0 minute, then to 100% B at 15 minutes and kept for 0.5 minutes. Then back to 0% B over 0.1 minutes and maintained for 4.4 minutes. The electrospray interface with nebulizer at 2.5 bar and dry gas at 9.0 L/min at 200°C was operated in both positive and negative mode (capillary voltage at 4000 V and 3200 V, respectively). Mass spectra in the range  $m/z$  60 – 1250 were acquired with a

frequency of 3 Hz. Spectra were saved as centroided. The  $m/z$  axis was calibrated with sodium formate clusters (solution of water/2-propanol/1 mol/l sodium hydroxide/formic acid 250/250/2.5/0.5 v/v/v/v) infused prior to each chromatographic run via a divert-valve-loop setup. The instrument was controlled using Bruker Daltonics micrOTOFcontrol version 3.1 and acquired data was handled with Data Analysis version 4.0 SP4.

### **Identification and quantification by LC/MS**

Stock solutions of the individual standards (Table 4) of 1 mg/ml were prepared. Dilution series of combinations of the stock solutions were prepared with concentrations of ca. 3, 10, 30, 100  $\mu\text{g/ml}$  of each of the six 2-hydroxy acids. The dilution series were used for preparing calibration solutions and for spiking CFs of LAB C. Aliquots of 400  $\mu\text{l}$  of the dilution series were added to 200  $\mu\text{l}$  of either 0.1% formic acid (calibration solutions) or the CF (spiked ferments) and additional 400  $\mu\text{l}$  0.1%v/v formic acid in water containing 0.12 mmol/l L-phenyl- $D_5$ -alanine as IS. The samples were analyzed by the LC/MS conditions listed previously. Verification of identity by retention time assignment and mass spectrum of compounds was made by analyzing solutions of 0.1 mg/ml.

### **Quantification of organic acids by gas chromatography with flame ionization detection (GC/FID)**

Aqueous stock solutions of standards were prepared containing the six 2-hydroxy acids of interest (Table 4) and the IS, 2-ethylbutanoic acid. Aliquots of standards (0.5-400  $\mu\text{l}$ ) or CF samples (100, 200, 300  $\mu\text{l}$ ) mixed with IS solution (50  $\mu\text{l}$ ) were derivatized with 2 ml ethanol in sulfuric acid (concentrated, p.a.) 150/50 v/v for 2 hours at 80°C. The solution was cooled to ambient temperature and 10 ml of a 10 w/w% sodium chloride in water was added. The combined solution was extracted with 2 ml diethyl ether for 30 minutes by vigorous shaking. The diethyl ether extracts were analyzed on an Agilent (Agilent Technologies, Waldbronn, Germany) 6890N GC with flame ionization detector (FID). The Agilent GC was controlled by Agilent GC Chemstation rev. B.04.01 SP1. The GC was mounted with a 10m x 0.10mm id Quadrex (Quadrex Corporation, Bethany, CT) 007-FFAP (nitroterephthalic acid modified polyethylene glycol polymer) column with a film thickness of 0.1  $\mu\text{m}$ . The carrier gas was He flowing at 0.8 ml/min. The sample injection was 1  $\mu\text{l}$  with split ratio of 1:100 in a split/split less injector kept at 250°C. The FID was held at 240°C. Two different temperature gradients were employed. For low boiling derivatives, the oven initially was held at 40°C for 2 minutes then temperature was increased to 240°C with a rate of 20°C/minute. For higher boiling derivatives, the oven start temperature held at 150°C for 2 minutes, and then was increased to 240° with a rate of 20°C/min to 240°C held for 10 minutes.

### **Data processing**

Metadata like pH (from CINAC) and biomass as well as randomizing injection sequences was handled in Microsoft Excel 2007 (Microsoft). Prior to feature extraction, chromatographic data of MIX samples (total ion chromatograms (TICs) and base peak chromatograms (BPCs)) were inspected visually for irregularities like drift in intensities and retention time. Bruker raw data files were converted into mzXML files by Bruker CompassXport v.3.0.5 (Bruker Daltonics). Feature extraction made with MZmine2 [24]. Peak detection was based on an  $m/z$  tolerance of 0.001 Dalton (Da) or 5 parts per million (ppm) and a peak duration time

range of 0.025 – 0.35 minutes. Chromatograms were deconvoluted using the ‘local minimum search’ algorithm, de-isotoped and peaks were aligned using the Join aligner algorithm. Peak lists were filtered using a criterion of minimum a feature being detected in five chromatograms (the data set held 5 biological replicates analyzed twice; technical duplicates). The peak list was gap filled and filtered for duplicate peaks with a retention time tolerance of 0.1 minutes and 0.001 Da or 5 ppm. Multivariate data analysis was performed in MATLAB R2013a version 8.1.0.604, 64-bit (MathWorks, Natick, MA, USA) and PLS Toolbox version 7.3.1 (Eigenvector Research, Inc., Wenatchee, WA, USA).

## RESULTS

### Characteristics of *Lb. paracasei* Fermentations and Cell Free Ferments

The three *Lb. paracasei* strains were fermented in parallel for 65 hours (end of fermentation). All three strains were diacetyl producing (data not shown). From an initial similar inoculation concentration (in terms of OD), two of the strains LAB B and LAB C fermented to a lower pH than LAB A. Corresponding to the larger decrease in pH, the LAB B and LAB C strains generated significantly higher amount of biomass and higher cell counts than LAB A and LAB C (Table 2).

**Table 2 Characteristics of *Lb. paracasei* fermentations (average  $\pm$  standard deviation)**

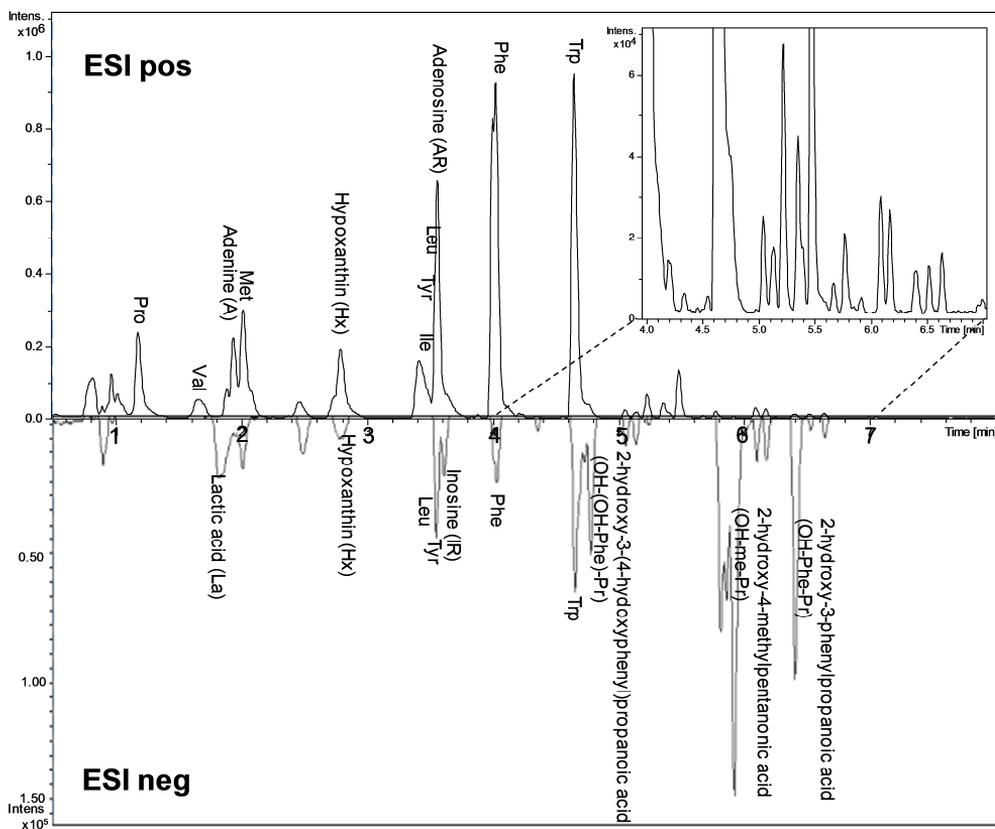
	<i>Lb. paracasei</i> strain (Short name)			Reference (REF)
	DGCC2132 (LAB A)	DGCC1128 7 (LAB B)	DGCC695 (LAB C)	
Cell count (log CFU/mL, n=3) Start of fermentation End of fermentation	6.8 $\pm$ 0.1 7.8 $\pm$ 0.1	7.6 $\pm$ 0.1 9.4 $\pm$ 0.2	6.9 $\pm$ 0.2 9.4 $\pm$ 0.2	
OD <sub>600</sub> (abs, n=5), Start of fermentation End of fermentation	0.052 $\pm$ 0.002 0.50 $\pm$ 0.04	0.048 $\pm$ 0.002 1.42 $\pm$ 0.02	0.050 $\pm$ 0.004 1.50 $\pm$ 0.05	
Diacetyl production during fermentation	Yes	Yes	Yes	
pH (n=5) Start of fermentation End of fermentation	6.55 $\pm$ 0.03 4.53 $\pm$ 0.08	6.55 $\pm$ 0.03 3.84 $\pm$ 0.06	6.56 $\pm$ 0.05 3.77 $\pm$ 0.07	
Biomass (g/L, n=5) End of fermentation	0.15 $\pm$ 0.01	0.56 $\pm$ 0.02	0.65 $\pm$ 0.01	
Mold growth after three days at 25 °C (10 <sup>3</sup> pixels, n=3) DCS 302 DCS 1541	6.31 $\pm$ 0.25 6.70 $\pm$ 0.22	5.19 $\pm$ 0.27 4.82 $\pm$ 0.01	4.40 $\pm$ 0.18 4.51 $\pm$ 0.10	7.38 $\pm$ 0.07 8.01 $\pm$ 0.25

The growth of the two indicator molds, *P. solitum* DCS 302 and *Penicillium* sp. nov. DCS 1541 was tested on the agar plates of un-inoculated media at pH 4.5 and CFs of the three *Lb. paracasei* strains. On a daily basis, mold growth was monitored by multispectral imaging which was processed into a measure of the sum of white and green pixels. Mold growth was characterized by only white pixels on day two and from day three and onwards by both white and green pixels as outlined in supplementary material S1. From day three and onwards, the three *Lb. paracasei* CFs could be ranked according to the mold growth: REF > LAB A > LAB B > LAB C (Table 2).

### Metabolic footprinting by Liquid Chromatography Mass Spectrometry (LC/MS)

The chemical profiles of the *Lb. paracasei* exometabolomes (i.e. CFs) were generated by reversed phase-LC/MS via both electrospray ionization (ESI) positive (pos) and negative (neg)

mode. The two ESI polarities were complementary in detecting both substrate components and metabolites as shown in Fig. 1. Compounds with acidic protons, like carboxylic acids, were more prone to detection in ESI neg by formation of M-H ions. The most intense peaks of the ESI pos were amino acids, adenosine and adenine from the CDIM. The most polar amino acids (like arginine, asparagine, aspartic acid, glutamine, glutamic acid) were only slightly retained on the column and eluted within or just after the void volume.



**Fig. 1** Base peak chromatogram of CF of LAB C showing electrospray positive (with insert of enlargement of retention time 4.0 – 7.0 min) and negative (inverted) mode with annotation of selected peaks.

Feature extraction using MZmine2 resulted in 977 features in ESI pos and 142 features in ESI neg mode. The metabolites for both polarities were normalized with the corresponding feature representing IS. Especially in ESI pos mode, compounds were represented with multiple features due to different adducts and in-source fragmentation. As an example glucose was observed as  $[M+Na]^+$ ,  $[M+NH_4]^+$ ,  $[M+K]^+$ , as  $[2M+Na]^+$  cluster and as corresponding in-source fragments after water loss. Glucose adducts and clusters were designated "Glc" and in-source fragments "Glc\*". At least two sets of features represent the sum of positional isomers closely eluting, namely 'Leu+Ile' covering leucine and isoleucine and 'OH-Me-Pe' covering 2-hydroxy-Y-methylpentanoic acid with Y being 3 or 4.

Principal component analysis (PCA) after mean centering and Pareto scaling grouped the samples according to their origin, i.e. un-inoculated CDIM (REF), the ferments of the three *Lb. paracasei* strains and a centered group of the pooled quality control sample (Supplementary material S2).

The technical replicates (n=2) were averaged for each biological replicate (n=5). For each of the two ESI modes, scores and loadings of PCAs of the averaged data are shown in Fig. 2. The first principal component (PC) for both ESI pos and neg mode data was correlated with both acidification and biomass production, i.e. grouping of LAB A versus LAB B and LAB C. Inspection of loadings of pos mode data showed that components from the CDIM such as glucose, amino acids (leucine + isoleucine, (Leu+Ile); phenylalanine (Phe), tryptophane (Trp), tyrosine (Tyr)), adenosine (AR) and adenine (A) were positively correlated with principal component 1 (PC1). Accordingly, PC1 correlated negatively with nutrient consumption.

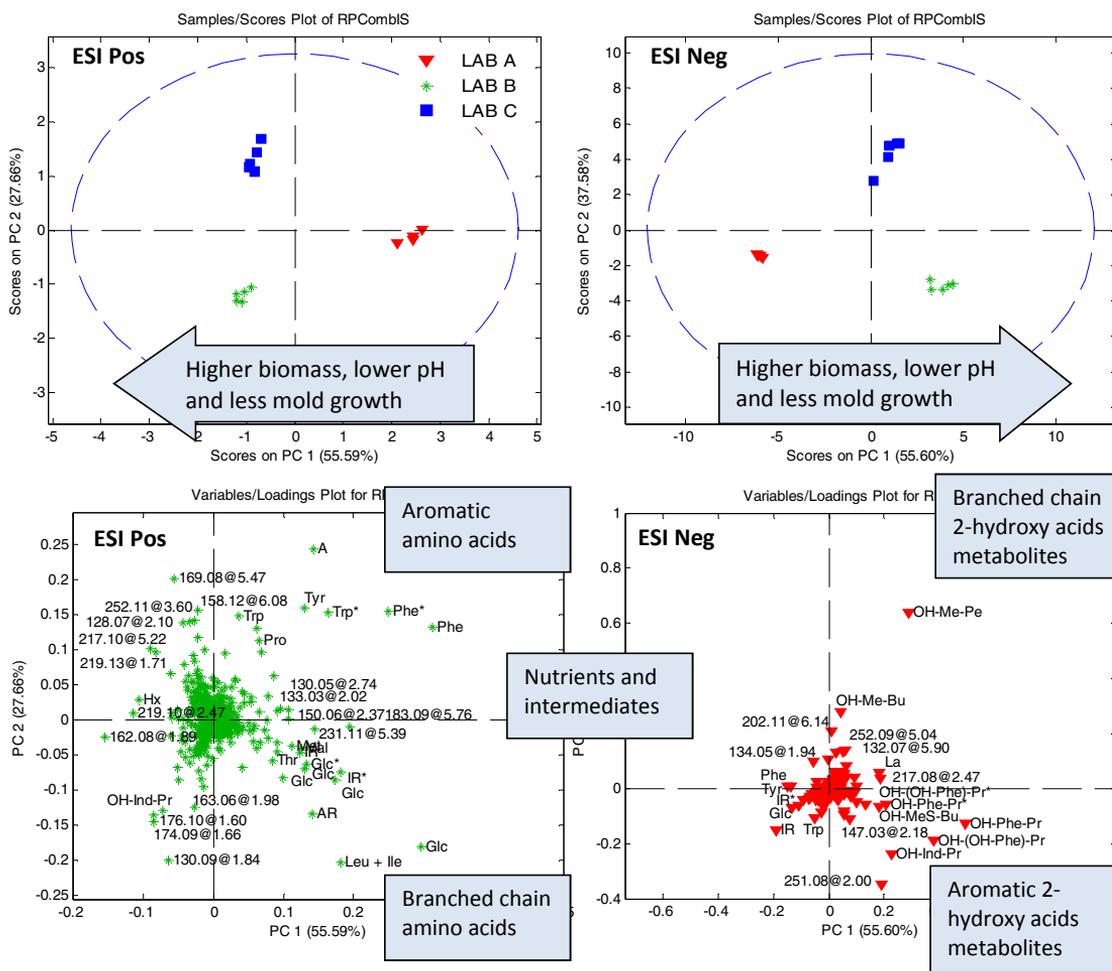
For ESI neg data, PC1 loadings correlated positively with formation of 2-hydroxy acids, e.g. lactic acid (La), 2-hydroxy-3-phenylpropanoic acid (OH-Phe-Pr), 2-hydroxy (4-hydroxyphenyl)propanoic acid (OH-(OH-Phe)-Pr) and 2-hydroxy-4-methylpropanoic acid (OH-Me-Pr). Consumption of glucose (Glc), amino acids (Phe,Tyr) and the intermediate inosine (IR) correlated negatively with PC1.

According to the highest loadings in PC2 in positive mode, the three LAB strains differed by a preference for metabolizing aromatic (Phe, Tyr, Trp) or branched chain (Leu + Ile) amino acids. Corresponding to this, LAB C produced relatively higher amounts of branched 2-hydroxy acids (seen in negative mode), whereas the LAB B had relatively higher amounts of the corresponding aromatic catabolism 2-hydroxy acids products.

### **Exploring the Correlation between Exometabolite Profiles and Mold Growth**

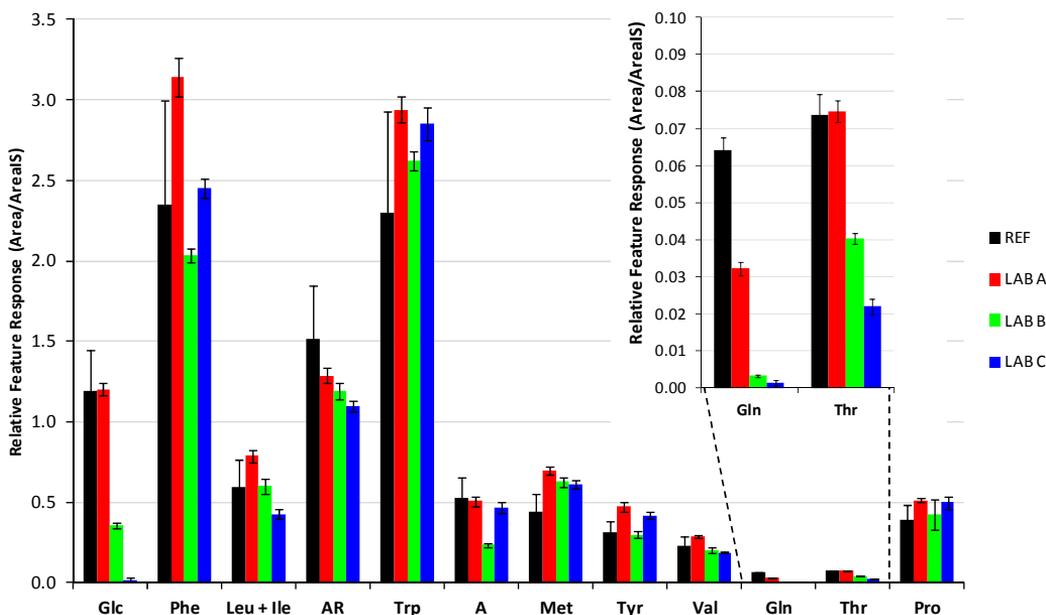
Correlation between the exometabolomes and mold growth were explored by partial least squares regression (PLSR) with the exometabolome profile as independent **X** data block and mold growth as dependent **Y** data block. In the previous section, exploratory multivariate data analysis of the two analytical modes demonstrated that the exometabolome contained information to differentiate the *Lb. paracasei* CFs, both in terms of nutrient consumption and formation of metabolites. The two sets of data (ESI pos and neg) were merged into one **X** data block to obtain a more comprehensive description of the differentiation. Low molecular weight organic acids like formic, acetic, propionic, butanoic acid, were analyzed by targeted GC/FID analysis. All these analytes were either below limit of quantification or not detected (Table 4) and were not included in the **X** data block.

The merged **X** data block was group-scaled (i.e. variance scaled to equal sum-of-square weighing) and mean centered. The mold growth **Y** data was mean centered. Partial least squares regression (PLSR) was performed with the mold growth for day 3 for each of the two indicator organisms versus the merged **X** data block. The characteristics of the resulting PLSR models as well as loading and scores plots are listed in supplementary material S3. For both organisms, the PLSR models of all days showed correlation between growth of the indicator organism and exometabolomic profile. All following observations below are based on PLSR models for day 3 growth of each of the two indicator molds (DCS 302 and DCS 1541).



**Fig. 2** PCA plot of biological replicates (normalized with IS, pareto scaled) of the three *Lb. paracasei* strain ferments showing PC1 and PC2 scores (top) and loadings (below) for ESI pos (left) and ESI neg data (right). Identified features designated with abbreviations and unidentified as mass to charge at retention time are listed in Table 3.

According to the loadings for latent variable 1 in supplementary material, increased mold growth, correlated to relatively high amounts of primarily glucose (as several adducts), phenylalanine, leucine + isoleucine and adenosine. Reduced mold growth was strongly correlated to depletion of these nutrients. The residual amount of each of the highest ranking nutrients was estimated as the sum of relative responses of all the compounds' adducts and in-source fragment features (Fig. 3).



**Fig. 3** Relative responses of nutrients in the CDIM (REF) and three *Lb. paracasei* CFs showing almost complete depletion of glucose and glutamine in LAB C but only limited consumption of other nutrients. (Sum of responses from adducts and in-source fragments; average  $\pm$  standard deviation, n=5)

None of the observed nutrients were completely depleted. However, glucose and glutamine were almost depleted in LAB C ferments. For all ferments all other measured nutrients remained at more than 50% of the content observed in the CDIM. Besides consumption of nutrients, formation of metabolites correlated positively with reduced mold growth. Among, the major metabolites were 2-hydroxy-4-methylpentanoic acid (OH-Me-Pe), 2-hydroxy-3-phenylpropanoic acid (OH-Phe-Pr) and 2-hydroxy-3-(4-hydroxyphenyl)propanoic acid (OH-(OH-Phe)-Pr) (Table 3). The variables (or features) mostly influencing the overall model were identified by calculation and inspection of variable importance in projection, or VIP scores [25]. For both mold models, the variables were grouped according to a positive or negative correlation to mold growth and within the groups ranked according to VIP scores and listed with the univariate Pearson correlation coefficient towards mold growth (Table 3).

**Table 3 Features in *Lb. paracasei* CF which correlated positively (top) and negatively (below) in latent variable 1 with mold growth and sorted according to VIP score (decreasing from top to bottom)**

Observed m/z	Ret. time [min]	ESI Polarity	Ion/adduct	Assignment <sup>a</sup>	Abbreviation	Exact m/z	m/z Error [ppm]	Pearson correlation coefficient (R) <sup>b</sup>
<b>Features positively correlated with mold growth</b>								
203.0526	1.03	Pos	[M+Na] <sup>+</sup>	Glucose	Glc	203.0526	0.2	0.98
166.0864	4.02	Pos	[M+H] <sup>+</sup>	Phenylalanine	Phe	166.0863	0.6	0.72
132.1020	3.57	Pos	[M+H] <sup>+</sup>	Leucine + Isoleucine	Leu + Ile	132.1019	0.5	0.95
120.0808	4.02	Pos	[M-H-COOH+H] <sup>+</sup>	Phe in-source fragment	Phe*	120.0808	0.5	0.55
268.1041	3.56	Pos	[M+H] <sup>+</sup>	Adenosine	AR	268.1040	0.2	0.87
183.0917	5.76	Pos	[M+H] <sup>+</sup>	Unk-42; C12H10N2	183.09@5.76	-	-	0.87
137.0459	3.61	Pos	[M-C <sub>5</sub> H <sub>8</sub> O <sub>4</sub> +H] <sup>+</sup>	Inosine in-source fragment	IR*	137.0458	0.6	0.94
198.0972	1.03	Pos	[M+NH4] <sup>+</sup>	Glucose	Glc	198.0972	0.1	0.96
188.0707	4.64	Pos	[M-NH <sub>3</sub> +H] <sup>+</sup>	Trp in-source fragment	Trp*	188.0706	0.4	0.36
205.0973	4.64	Pos	[M+H] <sup>+</sup>	Tryptophan	Trp	205.0972	0.5	0.09
267.0736	3.61	Neg	[M-H] <sup>-</sup>	Inosine	IR	267.0735	0.3	0.97
150.0584	2.02	Pos	[M+H] <sup>+</sup>	Methionine	Met	150.0583	0.5	0.80
136.0618	1.94	Pos	[M+H] <sup>+</sup>	Adenine	A	136.0618	0.4	0.21
180.0867	1.04	Pos	[M-H <sub>2</sub> O+NH4] <sup>+</sup>	Glucose in-source fragment	Glc*	180.0867	0.2	0.95
383.1159	1.03	Pos	[2M+Na] <sup>+</sup>	Glucose	Glc	383.1160	-0.2	0.97
231.1128	5.39	Pos	-	Unk-47	231.11@5.39	231.1128	0.0	0.88
269.0882	3.61	Pos	[M+H] <sup>+</sup>	Inosine	IR	269.0880	0.6	0.94
182.0812	3.55	Pos	[M+H] <sup>+</sup>	Tyrosine	Tyr	182.0812	0.4	0.32
118.0863	1.66	Pos	[M+H] <sup>+</sup>	Valine	Val	118.0863	0.3	0.91
<b>Features negatively correlated with mold growth</b>								
131.0714	5.90	Neg	[M-H] <sup>-</sup>	2-hydroxy-γ <sup>c</sup> -methylpentanoic acid	OH-Me-Pe	131.0714	0.3	-0.92
165.0558	6.39	Neg	[M-H] <sup>-</sup>	2-hydroxy-3-phenylpropanoic acid	OH-Phe-Pr	165.0557	0.5	-0.62
181.0507	4.77	Neg	[M-H] <sup>-</sup>	2-hydroxy-3-(4-hydroxyphenyl)propanoic acid	OH-(OH-Phe)-Pr	181.0506	0.6	-0.48
169.0761	5.47	Pos	[M+H] <sup>+</sup>	Unk-28; C11H8N2	169.08@5.47	-	-	-0.77
								-0.58

251.0773	2.00	Neg	[M-H] <sup>-</sup>	Unk-981	251.08@2.00	-	0.09	-0.21
117.0557	4.67	Neg	[M-H] <sup>-</sup>	2-hydroxy-3-methylbutanoic acid	OH-Me-Bu	-0.4	-0.79	-0.60
162.0761	1.89	Pos	[M+H] <sup>+</sup>	Unk-43; C6H11NO4	162.08@1.89	-	-0.81	-0.95
89.0243	1.83	Neg	[M-H] <sup>-</sup>	Lactic acid	La	-1.0	-0.78	-0.91
204.0667	6.62	Neg	[M-H] <sup>-</sup>	2-hydroxy-3-(1H-indol-3-yl)propanoic acid	OH-Ind-Pr	0.4	-0.17	-0.44
217.0830	2.47	Neg	[M-H] <sup>-</sup>	Unk-991; C8H14N2O5	217.08@2.47	-	-0.83	-0.95
217.0972	5.22	Pos	[M+H] <sup>+</sup>	Unk-66; C12H12N2O2	217.10@5.22	-	-0.97	-0.91
147.0452	6.39	Neg	[M-H2O-H] <sup>-</sup>	In-source fragment of PLA	OH-Phe-Pr*	0.3	-0.61	-0.78
137.0458	2.79	Pos	[M+H] <sup>+</sup>	Hypoxanthine	Hx	0.4	-0.84	-0.96
130.0863	1.84	Pos	-	Unk-97	130.09@1.84	-	0.04	-0.24
219.0976	2.47	Pos	[M+H] <sup>+</sup>	Unk-40 = Unk-991	219.10@2.47	-	-0.88	-0.98
219.1339	1.71	Pos	[M+H] <sup>+</sup>	Unk-85; C9H18N2O4	219.13@1.71	-	-0.96	-0.89
202.1085	6.14	Neg	-	Unk-1004	202.11@6.14	-	-0.66	-0.46
252.1091	3.60	Pos	-	Unk-67	252.11@3.60	-	-0.69	-0.45
158.1176	6.08	Pos	-	Unk-126	158.12@6.08	-	-0.66	-0.41
149.0279	4.72	Neg	[M-H] <sup>-</sup>	2-hydroxy-4-(methylthio)butanoic acid	OH-MeS-Bu	0.6	-0.55	-0.73

- a) Unk = Unknown identified with tentative elementary composition where possible  
b) Correlation between mold growth and feature response  
c) Y = 4 or 3, i.e. 2-hydroxy-4-methylpentanoic acid or 2-hydroxy-3-methylpentanoic acid

### Minimal inhibition concentrations

The contents of selected carboxylic acids were quantified by targeted LC/MS and GC/FID resulted in concentrations listed in Table 4. Chemical inhibition studies were performed with the 2-hydroxy acids (number 1-6 in Table 4 below) individually as presented in supplementary material S4. Minimal inhibition concentration for 50% inhibition (MIC 50) of the individual compounds, estimated as the lowest concentration where the inhibition degree was below 0.5 [10], was in the range of 5 to 10 mg/ml for all six compounds. This MIC 50 value was more than 75 times higher than the concentration produced of the most abundant of the six acids (OH-Me-Pe at 65 mg/l).

**Table 4 Quantification of produced and identified metabolites with highest VIP scores in mold growth models (average  $\pm$  stdev, n=5) and minimal inhibitory concentration for 50% inhibition (MIC 50)**

<i>L. paracasei</i> strain Compound (abbreviation; number)	LAB A	LAB B	LAB C	Limit of detection mg/l	MIC 50 (g/l)	
	mg/l	mg/l	mg/l		DCS 302	DCS 1541
2-hydroxy- $\gamma^3$ -methyl pentanoic acid (OH-Me-Pe; #1)	12.3 $\pm$ 0.2	24.5 $\pm$ 0.5	65 $\pm$ 6	0.3	5	5
2-hydroxy-3-phenylpropanoic acid (OH-Phe-Pr; #2)	<2	21.1 $\pm$ 0.9	11.6 $\pm$ 1.0	0.8	5	5
2-hydroxy-3-(4-hydroxyphenyl) propanoic acid (OH-(OH-Phe)-Pr; #3 )	<1	15.8 $\pm$ 0.7	6.5 $\pm$ 0.4	0.4	10	10
2-hydroxy-3-methylbutanoic acid (OH-Me-Bu; #5)	5.6 $\pm$ 0.1	5.3 $\pm$ 0.3	23.4 $\pm$ 2.2	0.7	5	5
Lactic acid <sup>b</sup> (La)	9.6 $\times 10^3 \pm$ 0.6 $\times 10^3$	15.6 $\times 10^3 \pm$ 0.5 $\times 10^3$	16.8 $\times 10^3 \pm$ $\pm 0.4 \times 10^3$	1 $\times 10^3$	NA	NA
2-hydroxy-3-(1H-indol-3- yl)propanoic acid (OH-Ind-Pr; #4)	<3	4.4 $\pm$ 0.2	<3	0.9	5	5
2-hydroxy-4-(methylthio)butanoic acid (OH-MeS-Bu; #6)	1.2 $\pm$ 0.03	4.4 $\pm$ 0.2	2.7 $\pm$ 0.2	0.2	10	10
Formic acid <sup>b</sup> (Fo)	ND	ND	ND	10	NA	NA
Acetic acid <sup>b</sup> (Ac)	<10	<10	<10	3	NA	NA
Propionic acid <sup>b</sup> (Pr)	ND	ND	ND	1	NA	NA
Butanoic acid <sup>b</sup> (Bu)	ND	ND	ND	1	NA	NA

a) Y = 4 or 3, i.e. 2-hydroxy-4-methylpentanoic acid or 2-hydroxy-3-methylpentanoic acid

b) Quantified by GC/FID

ND: Not detected, i.e. below limit of detection

<xx: Detected, but below limit of quantification of xx mg/ml

NA: Not analyzed

### DISCUSSION

Recent studies showed that a major contributor to the antifungal properties of LAB A (*Lb. paracasei* DGCC 2132) owed to release of diacetyl during the metabolism of live cells. Removal of the cells from the LAB A ferment effectively reduced the diacetyl concentration

and consequently reduced the inhibitory effect. After removal of cells, some residual antifungal effect remained in the cell free ferment, indicating other factors contributing to antifungal activity [22].

The purpose of this study was to demonstrate the use of an untargeted metabolomics approach for investigation of the factors responsible for the antifungal effect observed in CFs of selected *Lb. paracasei* strains. Fermentation of CDIM by the selected *Lb. paracasei* strains induced relative changes in the exometabolomes in terms of a) consumption of nutrients and b) formation of metabolites. By removal of cells and hence the main contribution from diacetyl, the relative growth of molds on the cell free ferments (CF) would provide the biological response to the combined change in the exometabolomes. Hereby the mold growth assists to assign the factors that contribute to antifungal effect in the CF. Potentially, this approach could serve as a supplement to bio-assay guided fractionation for identifying antifungal metabolites.

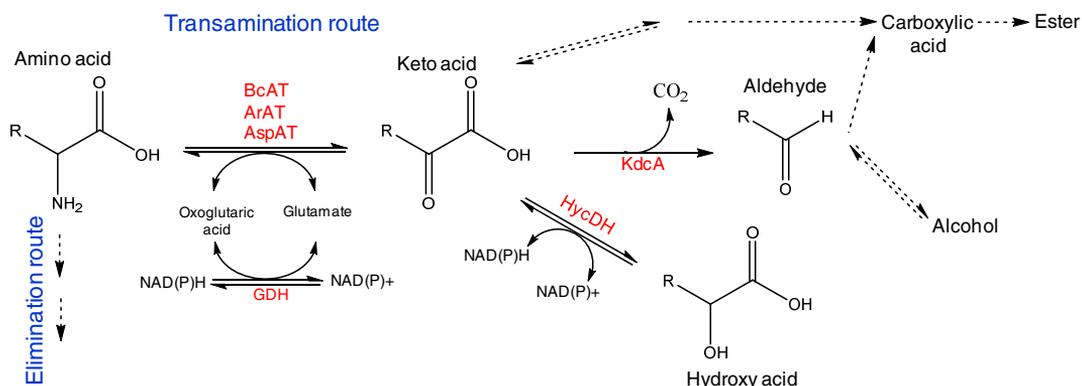
The three *Lb. paracasei* strains were selected based on their variation in effect on mold growth, i.e. inhibitory properties. However, LAB B and C produced more biomass and acidified to lower pHs than LAB A. All three strains were diacetyl producers, but the inhibition factor from diacetyl was considerably reduced by removing cells by centrifugation and filtration as previously observed [22]. Mold growth was assessed by recording and analyzing multispectral images. Spores of *Penicillium* are normally white in the initial growth phase. As the mold matures the color of the spores often change into green-grey or blue-green, and the development is dependent e.g. on the composition of the medium [26]. By the use of multispectral images, the approach allows for precise and objective description of the growth of *Penicillium* spp. by measuring the white, green segments and the total area of the mold growth (sum of white and green segments) [23]. Visually observed differences in mold growth during incubation were best described by the development of total (or green) amount of pixels as the number of white pixels remained constant.

The LC/MS-based metabolic footprinting of the CF was designed to include a wide range of nutrients and metabolites. However, the use of the reversed phase protocol excluded compounds more hydrophilic than e.g. glucose and lactic acid. Furthermore, the LC/MS footprint did not include low molecular short chain fatty acids like formic, acetic, propionic and butanoic acids. According to literature, these compounds contribute to the antifungal properties of lactic acid bacteria (e.g. [10, 11]) and were included in the study via GC/FID analysis. All four acids were either not detected or below a limit of quantification, which was more than 300 times below MIC values (e.g. MIC of acetic acid at pH 4.0 [10]: 3 g/L). Hence, these were not included in the multivariate data analyses. In spite of these limitations, the three *Lb. paracasei* strains could readily be classified by PCA. This classification was based on both the relative consumption of nutrients and formation of metabolites.

The sole purpose of doing PLSR was to identify compounds in the exometabolome which correlate with mold growth. The very limited number of *Lb. paracasei* strains did not allow the model to be used for any prediction of mold growth. Although the two indicator molds responded slightly different to the three *Lb. paracasei* CFs, their respective PLSR models showed practically similar ranking of the variables.

In terms of the consumption of nutrients, glucose ranked highest and was together with glutamine the only medium components almost fully consumed (by LAB C). Other nutrients remained at 50% or more of the initial content. The metabolites which had the highest correlation with decreased mold growth was lactic acid (primarily fermentation product of glucose), together with additional six 2-hydroxy acids and several unidentified compounds. Two metabolites, probably intermediates from metabolism of nucleotides [27], namely inosine and hypoxanthine also correlated with mold growth. However, the two compounds were not considered relevant factors for reduced mold growth based on the absence of anti-mold effect previously reported for other nucleosides [2]. Based on the PLSR model and GC/FID data it was hypothesized, that the minimal mold growth should be due to primarily three factors, namely (complete) consumption of glucose, acidification by lactic acid and formation of six known metabolites.

To our knowledge, indolelactic acid, 2-hydroxy-(4-methylthio)butanoic acid and 2-hydroxy-3-methylbutanoic acid have not previously been reported as antifungal compounds produced by *Lb. paracasei* strains. MIC 50 values obtained during this study were in accordance with the previously reported 2-hydroxy acids from LAB (2-hydroxy-4-methyl-pentanoic acid - MIC 10 g/L [28], phenyllactic acid – MIC 4 g/L [29] , 4-hydroxyphenyllactic acid - MIC 5 g/L [2]). The six 2-hydroxy acids were recognized as metabolites from lactic acid bacteria catabolism of amino acids [19, 30]. Their structural similarity with amino acids (Leu, Phe, Tyr, Trp, Met, Val) suggested that the metabolites were formed via the transamination route via a keto acid, which is reduced by a hydroxyacid dehydrogenase (Fig. 4). This is supported by the loadings observed where e.g. 2-hydroxy-3-phenylpropanoic acid was inversely correlated with the precursor Phe ( ).



**Fig. 4** Catabolism of amino acids (AA) with the focus on transamination of amino acids for formation of hydroxy acids in LAB. Enzymes are marked in red: **BcAT**, **ArAT**, **AspAT**, branched-chain-, aromatic- and aspartat aminotransferase; **GDH**, glutamate dehydrognase, **HycDH**, hydroxy acid dehydrogenase, **KdcA**: alpha-keto decarboxylase. Adapted from [31]

Even though the screening method favored these compounds, it was striking that transaminated and reduced metabolites were so abundant in ESI neg mode. The most obvious explanation could be a significantly higher MS response of these components compared to metabolites from the alternative route in the transamination, where the keto-

acids are decarboxylated and reduced/oxidized. Products from this pathway would be aldehydes, alcohols and carboxylic acids with a  $C_{n-1}$  carbon chain. Aldehydes and alcohols are expected to have lower responses in ESI neg mode than corresponding carboxylic acids. These metabolites could have been detected by a more sensitive method like headspace sampling followed by GC/MS [19]. The headspaces of the fermentations were analyzed during fermentation by SPME-GC/MS to test for decarboxylated metabolites. For LAB A, B and C the main observed components were diacetyl and acetoin and the method was not able to detect any components related to the decarboxylation metabolic pathway outlined in Fig. 4. However, detecting the 2-hydroxy acids is in accordance with presence of the genes for both D- and L-hydroxy acid dehydrogenases (HycDH) in *Lactobacilli* (and especially for *Lb. casei*) as described by Liu [31]. It was interesting to notice that none of the corresponding keto acids were detected being the precursors for the 2-hydroxy acids. Although the used exometabolomic approach is simplistic, it demonstrates the potential embedded in including the combinations of compounds. Hereby, it holds promise being a valuable supplement to e.g. bioassay guided fractionation for finding novel compounds and factors responsible for antimicrobial effect.

## **CONCLUSION**

Exometabolomic footprints of *Lb. paracasei* strains were correlated to mold growth to assist in identifying the factors responsible for the antifungal properties of the bacteria. For the selected mold growth test system, both nutrient consumption and metabolite formation correlated with the inhibition observed. The study demonstrated metabolic footprinting as a valuable supplement to bioassay guided fractionation for investigation of antifungal properties of bacterial cultures. Additionally, the strategy enabled the detection and assignment of three previously identified and three novel antifungal compounds and indicated potential precursors for the metabolites.

## **ACKNOWLEDGEMENT**

This work was partially financed by the Danish Ministry of Science, Innovation and Higher Education and by the University of Copenhagen as a scholarship for Stina Dissing Aunsbjerg. The analytical support for GC work by Research Associate Marianne Termansen and Technician Lasse Hørup, DuPont Nutritional BioSciences Aps was highly appreciated.

## REFERENCES

1. Ström K, Sjögren J, Broberg A, Schnürer J (2002) *Lactobacillus plantarum* MiLAB 393 Produces the Antifungal Cyclic Dipeptides Cyclo ( L -Phe – L -Pro ) and and 3-Phenyllactic Acid. *Appl Environ Microbiol* 68:4322–4327.
2. Ryan LAM, Zannini E, Dal Bello F, et al. (2011) *Lactobacillus amylovorus* DSM 19280 as a novel food-grade antifungal agent for bakery products. *Int J Food Microbiol* 146:276–83.
3. Lind H, Jonsson H, Schnürer J (2005) Antifungal effect of dairy propionibacteria - contribution of organic acids. *Int J Food Microbiol* 98:157–65.
4. Magnusson J, Ström K, Roos S, et al. (2003) Broad and complex antifungal activity among environmental isolates of lactic acid bacteria. *FEMS Microbiol Lett* 219:129–135.
5. Sjögren J, Magnusson J, Broberg A, et al. (2003) Antifungal 3-Hydroxy Fatty Acids from *Lactobacillus plantarum* MiLAB 14 Antifungal 3-Hydroxy Fatty Acids from *Lactobacillus plantarum* MiLAB 14. Society. doi: 10.1128/AEM.69.12.7554
6. Schwenninger SM, Meile L, Lacroix C (2011) Antifungal lactic acid bacteria and propionibacteria for food biopreservation. In: Lacroix C (ed) *Prot. Cult. Antimicrob. Metab. Bacteriophages Food Beverage Biopreservation*, 1st ed. Woodward, Oxford, pp 27–57
7. Crowley S, Mahony J, van Sinderen D (2013) Current perspectives on antifungal lactic acid bacteria as natural bio-preservatives. *Trends Food Sci Technol* 33:93–109.
8. Ryan LAM, Dal Bello F, Czerny M, et al. (2009) Quantification of phenyllactic acid in wheat sourdough using high resolution gas chromatography-mass spectrometry. *J Agric Food Chem* 57:1060–4.
9. Ryan LAM, Dal Bello F, Arendt EK, Koehler P (2009) Detection and quantitation of 2,5-diketopiperazines in wheat sourdough and bread. *J Agric Food Chem* 57:9563–8.
10. Schwenninger SM, Lacroix C, Truttmann S, et al. (2008) Characterization of low-molecular-weight antiyeast metabolites produced by a food-protective *Lactobacillus-Propionibacterium* coculture. *J Food Prot* 71:2481–7.
11. Corsetti A, Gobbetti M, Rossi J, Damiani P (1998) Antimould activity of sourdough lactic acid bacteria : identification of a mixture of organic acids produced by *Lactobacillus sanfrancisco* CB1. *Appl Microbiol Biotechnol* 50:253–256.
12. Niku-Paavola ML, Laitila A, Mattila-Sandholm T, Haikara A (1999) New types of antimicrobial compounds produced by *Lactobacillus plantarum*. *J Appl Microbiol* 86:29–35.

13. Brosnan B, Coffey A, Arendt EK, Furey A (2012) Rapid identification, by use of the LTQ Orbitrap hybrid FT mass spectrometer, of antifungal compounds produced by lactic acid bacteria. *Anal Bioanal Chem* 403:2983–95.
14. Guo J, Brosnan B, Furey A, et al. (2012) Antifungal activity of *Lactobacillus* against *Microsporium canis*, *Microsporium gypseum* and *Epidermophyton floccosum*. *Bioeng Bugs* 3:104–113.
15. Armaforte E, Carri S, Ferri G, Caboni MF (2006) High-performance liquid chromatography determination of phenyllactic acid in MRS broth. *J Chromatogr A* 1131:281–4. doi: 10.1016/j.chroma.2006.07.095
16. Mashego MR, Rumbold K, De Mey M, et al. (2007) Microbial metabolomics: past, present and future methodologies. *Biotechnol Lett* 29:1–16.
17. Birkenstock T, Liebeke M, Winstel V, et al. (2012) Exometabolome analysis identifies pyruvate dehydrogenase as a target for the antibiotic triphenylbismuthdichloride in multiresistant bacterial pathogens. *J Biol Chem* 287:2887–95.
18. Paczia N, Nilgen A, Lehmann T, et al. (2012) Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms. *Microb Cell Fact* 11:122.
19. Serrazanetti DI, Ndagijimana M, Sado-Kamdem SL, et al. (2011) Acid stress-mediated metabolic shift in *Lactobacillus sanfranciscensis* LSCE1. *Appl Environ Microbiol* 77:2656–66.
20. Galloway LD, Burgess R (1952) *Applied Mycology and Bacteriology*. *Appl. Mycol. Bacteriol.*, 3rd ed. Leonard Hill, London, pp 54–57
21. Aunbjerg SD, Honoré AH, Vogensen FK, Knøchel S (2015) Development of a chemically defined medium for studying foodborne bacterial–fungal interactions. *Int Dairy J* 45:48–55.
22. Aunbjerg SD, Honoré a H, Marcussen J, et al. (2014) Contribution of volatiles to the antifungal effect of *Lactobacillus paracasei* in defined medium and yogurt. *Int J Food Microbiol* 194C:46–53.
23. Ebrahimi P, Berg F van den, Aunbjerg SD (2015) Quantitative determination of mold growth and inhibition by multispectral imaging. *Food Control* 55:82–89.
24. Pluskal T, Castillo S, Villar-Briones A, Oresic M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* 11:395.
25. Chong I-G, Jun C-H (2005) Performance of some variable selection methods when multicollinearity is present. *Chemom Intell Lab Syst* 78:103–112.

26. Hanson JR (2008) Pigments and Odours of Fungi. Chem. Fungi. RSC Publishing, Cambridge, England, pp 127–146
27. Kilstrup M, Hammer K, Ruhdal Jensen P, Martinussen J (2005) Nucleotide metabolism and its control in lactic acid bacteria. FEMS Microbiol Rev 29:555–90.
28. Broberg A, Jacobsson K, Ström K, Schnürer J (2007) Metabolite profiles of lactic acid bacteria in grass silage. Appl Environ Microbiol 73:5547–52.
29. Lavermicocca P, Valerio F, Evidente A, et al. (2000) Purification and characterization of novel antifungal compounds from the sourdough *Lactobacillus plantarum* strain 21B. Appl Environ Microbiol 66:4084–90.
30. Kieronczyk A, Skeie S, Langsrud T, Yvon M (2003) Cooperation between *Lactococcus lactis* and Nonstarter *Lactobacilli* in the Formation of Cheese Aroma from Amino Acids. 69:734–739.
31. Liu M, Nauta A, Francke C, Siezen RJ (2008) Comparative genomics of enzymes in flavor-forming pathways from amino acids in lactic acid bacteria. Appl Environ Microbiol 74:4590–600.

## Supplementary material S1:

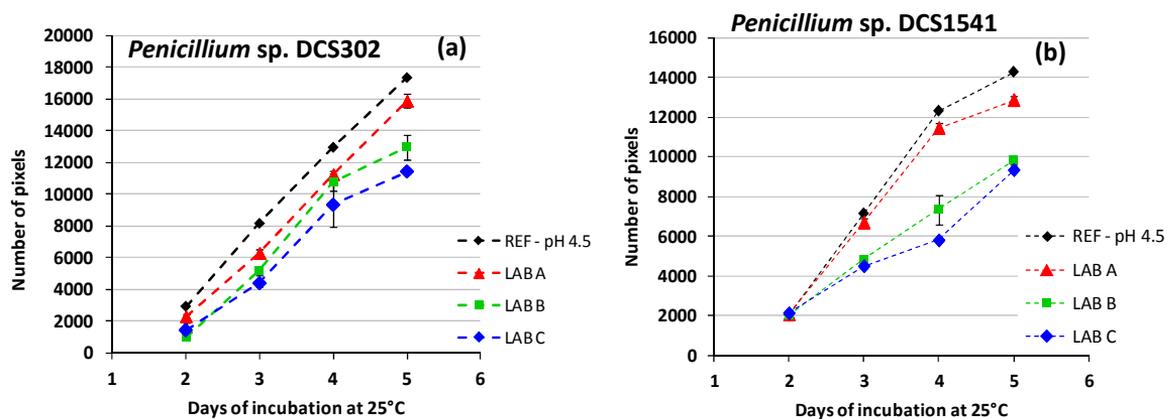
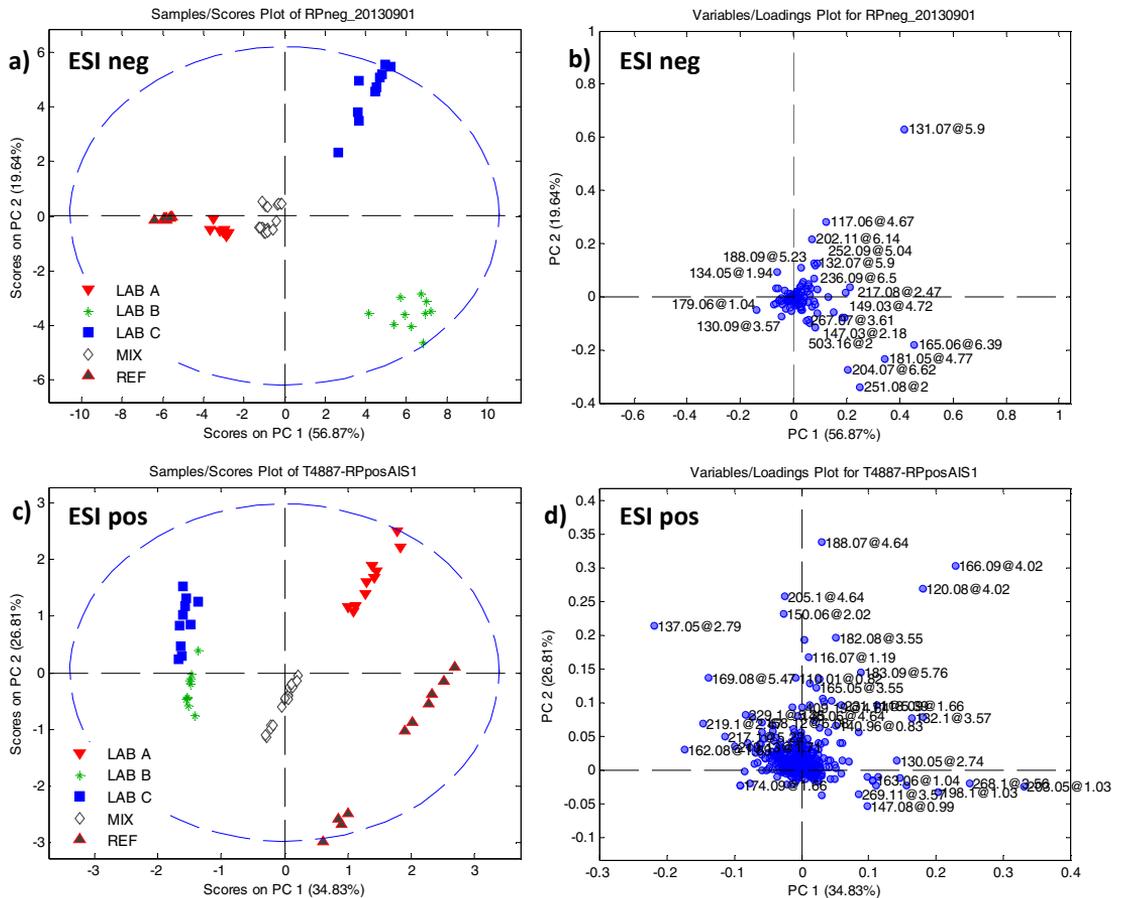


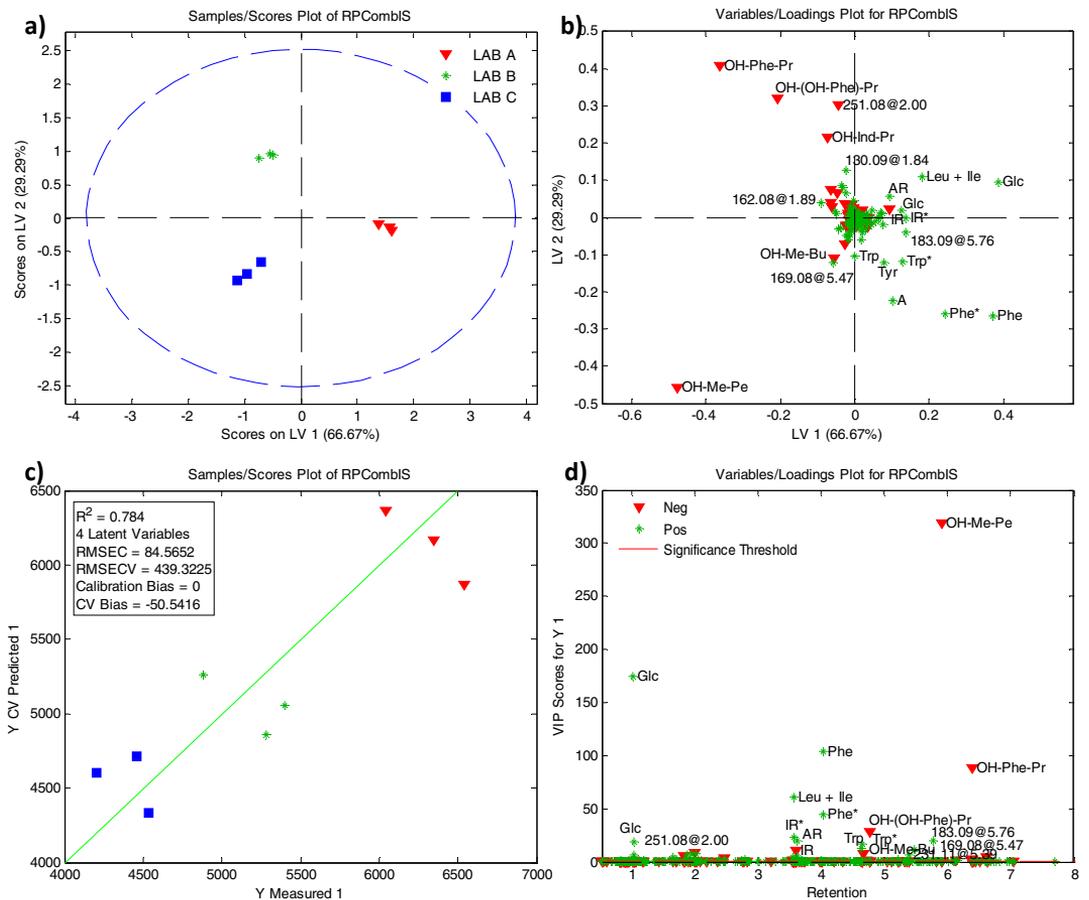
Fig. 1 Growth at 25°C of indicator mold strains a) *Penicillium* sp. DCS302 and b) *Penicillium* sp. DCS1541 spotted on plates of REF (un-inoculated CDIM acidified to pH 4.5) and cell free ferments of LAB A, B and C. (average  $\pm$  standard deviation, n=3)

Supplementary material S2:

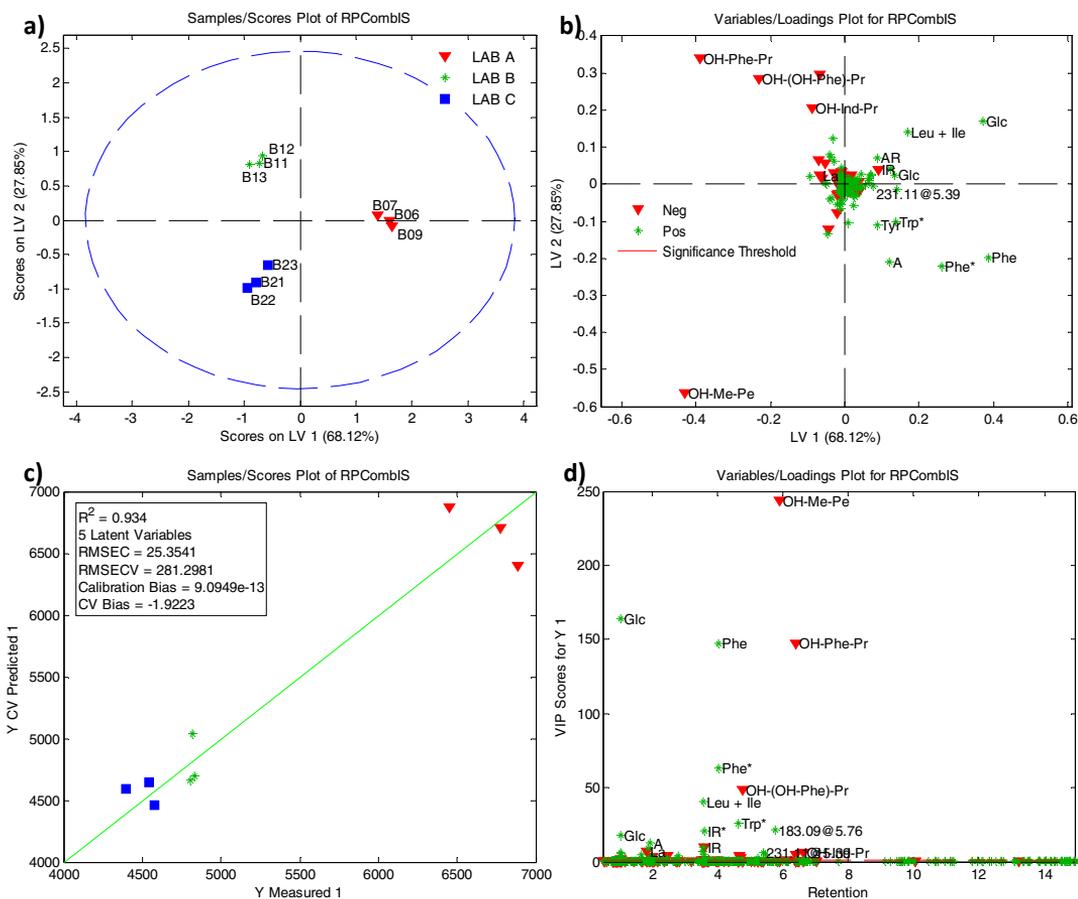


**Fig. 1** Pareto scaled PCA scores and loadings of PC1 versus PC2 for ESI neg data (top, a + b)) and ESI pos data (top, c + d)) for un-inoculated medium (REF), *Lb. paracasei* ferments (LAB A, B and C) and pooled control samples (MIX).

Supplementary material S3:



**Fig. 1** PLSR Model for *Penicillium sp.* DCS 302, day 3 scores plot a), latent variable 2 versus 2 loadings b), predicted versus observed mold growth c) and VIP scores d). Loadings from Pos variables are marked with green stars, and Neg variables are marked with red triangles.



**Fig.2** PLSR Model for *Penicillium sp.* DCS 1541, day 3 scores plot a), latent variable 2 versus 2 loadings b), predicted versus observed mold growth c) and VIP scores d). Loadings from Pos variables are marked with green stars, and Neg variables are marked with red triangles.

Supplementary information S4:

Table 1 Inhibition degree of OH-Me-Pe, OH-Phe-Pr, OH-(OH-Phe)Pr, OH-Ind, OH-Me-Bu and OH-MeS-Bu towards *P. solitum* DCS 302 *Penicillium* spp. DCS 1541 in CDIM acidified with DL lactic acid to pH 4.5 (Average  $\pm$  standard deviation, n=3). Compounds are added in the concentrations 0.1, 1, 5 and 10 mg/mL. The lowest concentration needed to give MIC 50 is highlighted in bold

Compound (abbreviation)	Concentration mg/ml	Inhibition Degree (48 hours)					
		DCS 302			DCS 1541		
OH-Me-Pe	0.1	1.07	$\pm$	0.20	1.13	$\pm$	0.13
	1	0.75	$\pm$	0.03	0.87	$\pm$	0.02
	<b>5</b>	<b>0.36</b>	$\pm$	<b>0.04</b>	<b>0.32</b>	$\pm$	<b>0.05</b>
	10	0.06	$\pm$	0.01	0.07	$\pm$	0.01
OH-Phe-Pr	0.1	0.89	$\pm$	0.03	1.01	$\pm$	0.03
	1	0.79	$\pm$	0.10	1.01	$\pm$	0.14
	<b>5</b>	<b>0.33</b>	$\pm$	<b>0.02</b>	<b>0.45</b>	$\pm$	<b>0.03</b>
	10	0.01	$\pm$	0.05	0.05	$\pm$	0.07
OH-(OH-Phe)Pr	0.1	0.98	$\pm$	0.13	1.13	$\pm$	0.08
	1	0.80	$\pm$	0.03	1.01	$\pm$	0.02
	5	0.71	$\pm$	0.03	0.98	$\pm$	0.08
	<b>10</b>	<b>0.43</b>	$\pm$	<b>0.04</b>	<b>0.40</b>	$\pm$	<b>0.49</b>
OH-Ind-Pr	0.1	0.81	$\pm$	0.03	1.06	$\pm$	0.04
	1	0.77	$\pm$	0.07	1.09	$\pm$	0.06
	<b>5</b>	<b>0.49</b>	$\pm$	<b>0.02</b>	<b>0.26</b>	$\pm$	<b>0.03</b>
	10	0.02	$\pm$	0.05	0.01	$\pm$	0.04
OH-Me-Bu	0.1	0.87	$\pm$	0.03	1.31	$\pm$	0.04
	1	0.69	$\pm$	0.03	1.15	$\pm$	0.03
	<b>5</b>	<b>0.41</b>	$\pm$	<b>0.02</b>	<b>0.53</b>	$\pm$	<b>0.01</b>
	10	0.06	$\pm$	0.02	0.00	$\pm$	0.00
OH-MeS-Bu	0.1	0.78	$\pm$	0.03	1.24	$\pm$	0.05
	1	0.76	$\pm$	0.14	1.33	$\pm$	0.06
	5	0.60	$\pm$	0.02	1.19	$\pm$	0.02
	<b>10</b>	<b>0.53</b>	$\pm$	<b>0.06</b>	<b>0.32</b>	$\pm$	<b>0.40</b>

PARVANEH EBRAHIMI

**Metabolic profiling of food protective cultures by *in vitro* NMR spectroscopy**



Food spoilage is of major concern to the food industry, because it leads to considerable economic losses and to possible public health hazards. There has been considerable effort and research to preserve food which have mainly relied on the application of chemical preservatives or drastic physical treatments. However, chemical preservatives are becoming increasingly unpopular by the consumers, and some have even proven to be toxic and linked to different health problems. Physical treatments of the products, on the other hand, can deteriorate the sensory properties of the products, and may even destroy some of the nutrients and vitamins. In this context, 'Biopreservation', which is defined as the use of safe antibacterial/antifungal microorganism (so-called protective cultures) has unexploited potential to inhibit the growth of pathogenic microorganisms and enhance the shelf life of the final food product. In order to apply biopreservation in food products effectively, detailed knowledge on the metabolism of protective cultures is required.

This PhD study mainly focused on the application of *in vitro* NMR spectroscopy for studying the metabolism of protective cultures. An analytical protocol was developed for real-time *in vitro* NMR measurements of bacterial fermentation, which includes guidelines from the sample preparation to the data processing and the modelling of the metabolic profiles. As a part of this work, an NMR data preprocessing technique, called 'Reference Deconvolution', was employed for the first time to improve the multivariate analysis of the *in vitro* real-time metabolomics data and proved a necessary and elegant solution to the inherent inhomogeneity problem of the samples in the *in vitro* NMR measurements of cells. As the second part of the project, an accurate approach for quantifying mold growth and inhibition, based on multispectral images and k-means clustering was developed. The method was developed into a software package called 'PCluster', and was demonstrated to be very helpful in two other biopreservation related metabolomic studies. The developed analytical tools are expected to be very beneficial in the studies related to the biopreservation, and will be used in the future investigations of the protective cultures.