Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Natural Resources and Agricultural Sciences

Inverse correlation between freeze-thaw survival and storage stability of freeze-dried Lactobacillus reuteri DSM 17938 and a microcalorimetric search for GASP

Radoslaw Sawicki

Department of Molecular Sciences
Master´s thesis • 30 hec • Second cycle, A2E
Molecular Sciences, 2017:6
Uppsala 2017

Inverse correlation between freeze-thaw survival and storage stability of freeze-dried Lactobacillus reuteri DSM 17938 and a microcalorimetric search for GAS

Radoslaw Sawicki

Supervisor:	Sebastian Håkansson, Swedish University of Agricultural Sciences, Department of Molecular Sciences
Assistant Supervisor:	Armando Hernández Garcia, Swedish University of Agricultural Sciences, Department of Molecular Sciences Stefan Roos , BioGaia AB
Examiner:	Mikael Pell, Swedish University of Agricultural Sciences, Department of Molecular Sciences

Credits: 30
Level: Second cycle, A2E
Course title: Independent project in Biology - Master's thesis
Course code: EX0565

Place of publication: Uppsala
Year of publication: 2017
Title of series: Molecular Sciences
Part number: 2017:6
Online publication: http://stud.epsilon.slu.se

Keywords: freeze-drying, lyophilization, *Lactobacillus reuteri*, probiotic, pH, temperature, freeze-thaw, storage stability, survival, morphology, microscopy, long-term stationary phase, GASP, isothermal microcalorimetry

Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Faculty of Natural Resources and Agricultural Sciences
Department of Molecular Sciences

Abstract

Lactobacillus reuteri DSM 17938 is a probiotic bacterium manufactured by the company BioGaia AB and was in focus of this study. If a probiotic product is to confer its beneficial health effect to a consumer, it must retain high viability levels during long-term storage. A common practice for conservation of living microorganisms to meet this criterion is freeze-drying. Freeze-drying of bacteria is a process dependent on many factors and needs to be carefully designed in order to be successful. Preconditioning the bacteria before freeze-drying by subjecting them to different conditions during cultivation is one of the ways toward optimizing the survivability of freeze-drying and subsequent storage-stability. In theory, finding more robust mutants of *L. reuteri* could be yet another way of reaching this goal. It has been discovered that at least in some species of bacteria, increasingly resilient generations of mutants arise in consecutive waves during prolonged batch cultivation and are referred to as "growth advantage in stationary phase" (GASP) mutants.

The overall goal of this study was to work towards a viable and storage-stable product based on *L. reuteri*. The first objective was to investigate if survival of a freeze-thaw assay, in which differently preconditioned *L. reuteri* are subjected to cycles of freezing and thawing, correlates with storage-stability. This could verify the potential use of the freeze-thaw assay instead of performing slow and laborious long-term stability studies when optimizing the production process, as well as provide insights into which cultivation conditions confer the highest storage-stability. A second objective was to investigate prolonged batch cultures of *L. reuteri* as a hypothetical source of GASP mutants. It was further hypothesized that appearance of GASPs would be detectable by isothermal microcalorimetry.

L. reuteri was cultivated at different settings of pH and temperature and freeze-dried in a 15% w/v sucrose formulation. Survival of the freeze-thaw assay and an accelerated storage assay at 37 °C were compared by viable counting and subsequent ANOVA and Tukey's test. This was complemented by microscopic morphology and Gram-stain studies. For the mutant-related experimental objective, *L. reuteri* and *E. coli* (as a control organism) were cultivated for up to 70 days while their heat production was measured in isothermal microcalorimeters (IMC).

A consistent inverse correlation was found between the freeze-thaw assay and the accelerated storage assay. High pH, and to a lesser extent, high temperature were associated with higher stability but lower freeze-thaw assay survival. Microscopic studies did not find any consistent association between stability and morphology except for the absence of cell clumps in the most storage-stable case. No GASP-revealing heat flows could be observed, neither with *L. reuteri* or *E. coli* controls. The collected data implied that *L. reuteri* did not produce GASP mutants. Future effort is suggested to be put into finding out if survival of the freeze-drying and subsequent storage-stability can both be improved independently and in such case, separate assays for both criteria should be employed. In conclusion, it is likely that the causes for differences in storage-stability and freeze-thaw survival reside in biomolecular phenomena affected by cultivation conditions. IMC does not seem to be a

feasible tool for GASP detection; the possible reason could be lack of agitation and thereby no GASP growth. However, IMC devices with stirring could still be evaluated as a tool for GASP detection.

Keywords: freeze-drying, lyophilization, Lactobacillus reuteri, storage stability, survival, freeze-thaw, morphology, microscopy, long-term stationary phase, GASP, isothermal microcalorimetry, probiotic, fermentation, preconditioning

Table of contents

Abbreviations

General abbreviations

ANOVA	Analysis of Variance
CI	Confidence Interval
CFU	Cell Functional Unit
CV	Coefficient of Variation
EMM	Estimated Marginal Mean
FT-assay	Freeze-Thaw assay
GASP	Growth Advantage In Stationary Phase (a mutant type)
IMC	Isothermal Microcalorimetry
LB	Luria-Bertani growth medium
MOPS	3-(N-morpholino)propanesulfonic acid buffer
MODDE	"Modeling and design" – a statistical software
MRS	De Man, Rogosa and Sharpe growth medium
MS_E	Mean Square of the error
OD	Optical Density
SPSS	A statistical software
T_c	Collapse temperature of an amorphous glass phase into a viscous phase
T_g'	Glass transition temperature of the bacteria-formulation precipitant left after freezing of the water. The precipitant forms the amorphous glass phase at and below this temperature.
Tukey-HSD	Tukey's Honest Significant Difference
VC	Viable Count

Experimental designs

CRD	Completely Randomized Design
RCBD	Randomized Complete Block Design
RIBD	Randomized Incomplete Block Design
BRIBD	Balanced Randomized Incomplete Block Design

Treatments

A	37 °C, pH 4.5
B	32 °C, pH 4.5
C	32 °C, pH 6.5

1 Introduction

1.1 Importance of the topic

Science is just beginning to unravel the complex and important roles of microorganisms beneficial for our health. Those microbes that have a scientifically documented beneficial health effect are called probiotics and many probiotic products are already available commercially in forms of pills and dairy products. One such probiotic bacterium is *Lactobacillus reuteri*, on which this study focused. The Swedish company BioGaia AB is specialized in *L. reuteri* and its use for healthcare. The strains currently being sold are the original DSM 17938 in oral and gut health products and ATCC PTA 5289 and ATCC PTA 6475 in oral health products. The company is research-intensive and the strains are researched at institutions all over the world, part of the research is initiated by BioGaia and part is initiated independently. This study was done in collaboration with BioGaia and SLU as part of a project involving the design of a good production process for the probiotic.

1.2 Problem formulation

An important problem that must be tackled when making a probiotic microorganism into a commercial product is its storage-stability. This means that the cells must stay alive for a long time, because only viable cells can replicate in the gut and much of the health effects lies in the actions of live microbes. By carefully drying bacteria to a so called bacterial “powder”, their metabolism may be put to a halt without killing them and the degradative chemical reactions become slowed down; these are important prerequisites for stability (Béal & Fonseca, 2015). There are a few different methods for drying bacteria, however, this study focused on the freeze-drying method which is efficient and may lead to high survival rates

and good long-term viability of the product, when properly optimized (Meng et al., 2008; Schoug & Åsa, 2009).

Freeze-drying relies on the sublimation of ice which is accomplished by very slow heating-up of frozen bacterial samples in a vacuum environment. However, freeze-drying can be a harsh process for the bacteria. That is why the freeze-drying needs to be carefully designed and controlled to protect the bacteria from its dangers, if high survival rates are to be achieved. To shield the bacteria from the dangerous conditions which arise during freeze-drying, the bacteria are "formulated" in protective agents. Besides choosing the right formulation, the freeze-drying survival can be influenced by preconditioning the bacteria (Palmfeldt & Hahn-Hägerdal, 2000; Koch et al., 2008; Meng et al., 2008; Schoug et al., 2008; Liu et al., 2014). Preconditioning may be done by for example using different settings during cultivation of the bacteria which alters the molecular profile of the cells. Therefore, it is hypothetically possible that the storage-stability could also be affected by preconditioning. Additionally, which species of microbe is chosen has a crucial effect on freeze-drying survival and storage-stability (Schoug & Åsa, 2009) and so, potentially, may be choosing the right strain or finding the right mutant. A specific type of resilient mutants called "growth advantage in stationary phase mutants" (GASPs) has been discovered to evolve in some species of bacteria, during prolonged cultivations (Zinser & Kolter, 2004; Finkel, 2006; Bačun-Družina et al., 2011). Whether GASPs mutants can appear in *L. reuteri* had not been researched before this study.

1.3 Aim and scope of study

The overall aim of this study was to contribute to BioGaia's project on finding a recipe for a viable and storage-stable product based on *L. reuteri* DSM 17938. The study built upon what was known from literature and the supervisor's previous experience, such as which regimes are useful for freeze-drying. It also partly built upon some things that others involved in the project had investigated previously, such as the survival of a freeze-thaw (FT) assay. In the FT assay, explained in detail in section 3.1.4, the cells are subjected to repeated cycles of freezing and thawing. Depending on how the cells were preconditioned during cultivation, the survival ratios of this assay differed in previous investigations (Garcia, Armando Hernández et al., unpublished).

As the first objective of this thesis, it was investigated whether survival of the FT assay could be correlated with the survival of freeze-drying and the subsequent storage-stability (depending on previous preconditioning by cultivation settings). This could potentially simplify future studies since the FT assay is much quicker

and much less work and resource intensive than direct measurement of storage-stability. Additionally, microscopy was introduced to investigate if bacterial morphology can be used as another predictor of storage-stability. Furthermore, these experiments would give some insight about which type of preconditioning gives the most storage-stable biomass. Whether preconditioning would produce differences in storage-stability was unknown; little research has been done around this subject (Yao et al., 2008; Velly et al., 2014).

The second objective involved the strategy of looking for more resilient mutants of *L. reuteri*. A hypothesis in this part of the study was that *L. reuteri* would, like some other bacteria, produce the resilient GASP mutants which could potentially prove to be more tolerant to freeze-drying and storage. The supervisors had previous experience with the use of isothermal microcalorimetry (IMC), a sensitive method which can measure the heat production of bacterial activity, elaborated in section 2.5. Therefore, it was further hypothesized that the emergence of GASP cultures could be detectable by IMC (something never investigated previously).

The first objective of this study (correlation of FT-assay), was originally supposed to be a small experiment with very simple statistics applied, while the main study would have been a small fractional factorial design using the design of experiment software ”MODDE Pro” (Umetrics™) to find a good formulation for freeze-drying. Unfortunately, since there were a lot of technical problems at the correlation experiment, it took a while to gather enough data to biologically replicate it and there was no more time for the formulation experiment. However, since statistics were supposed to be a major part of this thesis, a big effort was made to find a statistical approach applicable to the correlation experiment, as a practice in statistics. The statistical approach, which is well elaborated in section 2.6, took the path of block designs and ANOVAs (Analysis of Variance) with data transformations, largely by the aid of the statistical software IBM SPSS Statistics® (in short SPSS).

1.4 Limitations

There are other useful methods for drying bacteria: spray-drying or vacuum drying (Meng et al., 2008) and fluidized-bed drying (Albadran et al., 2015). However, these were left out of the scope. Other things left out of scope, besides investigating the effects of different formulations, were how the different cultures of *L. reuteri* tolerate gastric juices and if their beneficial interaction with host gut is retained. Furthermore, to fully optimize the cultivation conditions in order to get the

most robust and storage-stable cells, bigger factorial experiments would be needed than what was done here.

2 Background

2.1 Lactobacillus reuteri

Many species of *Lactobacillus* are beneficial for a healthy gut by complex microbe-host and microbe-microbe interactions (Hammes & Hertel, 2006). A few *Lactobacillus* species (*L. reuteri*, *L. salivarius*, *L. ruminis*, *L. gasseri* and *L. crispatus*) are indigenous in the average human gut microflora, others are transient guests whose presence can fluctuate over time, yet may have important roles (Hammes & Hertel, 2006). Some of the simplest beneficial effects of *Lactobacillus* species are the lowering of pH and the occupation of space and nutrients in the intestines, which prevents pathogenic microbes from settlement and growth (Hammes & Hertel, 2006).

Lactobacillus reuteri is a Gram positive, rod-shaped bacterium and is a common component of the microbial flora of humans and many other animals (Hammes & Hertel, 2006). It is obligately heterofermentative, meaning that instead of producing mainly one fermentation product (as in homofermentative), it produces a mixture of CO_2, lactic acid, ethanol and/or acetic acid in equimolar amounts (Hammes & Hertel, 2006). Using the strain ATCC 55730, it was shown that *L. reuteri* does not produce acetic acid when grown in glucose as the carbon source (Årsköld et al., 2008). Furthermore, *L. reuteri* produces a bacteriocin (reutericin 6) and a bacteriocin-like inhibitory substance called reuterin (Hammes & Hertel, 2006). These substances inhibit many pathogenic microorganisms (including fungi and protozoa) (Hammes & Hertel, 2006). Except for inhibiting pathogens, *L. reuteri* strains produce vitamins and essential metabolites from which the host can benefit and can also modulate the immune system of the host in different ways (Saulnier et al., 2011). There is a large genetic variability between the many different strains of *L. reuteri*: there are groups of specialized strains belonging to

specific hosts and there can be large variation between the strains (and their health effects) belonging to a single host species, such as humans (Saulnier et al., 2011).

Beneficial effects of *L. reuteri* strains against gastrointestinal disorders of children and infants has been shown in several clinical studies, it has also been proven to suppress some gastrointestinal infections and is proposed to prevent "leaky gut" (Saulnier et al., 2011; Pallin et al., 2015). There are also indications that it may be beneficial in the combat of allergic diseases (*Clinical studies and meta-analyses | BioGaia*).

2.2 Freeze-drying of Lactobacillus

Although water is essential to all known life, some forms of life can sustain at surprisingly dry conditions; this is studied in the field of anhydrobiotics. Because the anhydrobiotic condition stops the cellular metabolism and can under the right circumstances slow down surrounding chemical reactions extremely, it is a useful concept for enhancing longevity of cellular products (Béal & Fonseca, 2015).

The goal of freeze-drying bacteria is to stabilize them in a viable, easily storable form of a well-dried "powder" with minimal water content and possibly in an amorphous glass state (Fonseca et al., 2015). Formation of an amorphous glass-state matrix by a phenomenon called vitrification is beneficial for freeze-dry survival and subsequent storage-stability of cells (Schoug, 2009). Lyoprotectants such as disaccharides have a good ability to vitrify when exposed to a temperature below the glass transition temperature (T_g') during their rapid precipitation out of the aqueous solution when the water is forming ice crystals (Schoug, 2009). The water content of the precipitant affects the T_g' strongly; less water gives a higher T_g' (Schoug et al., 2010). There is also evidence of beneficial vitrification taking place within the cells themselves before the surrounding matrix vitrifies (Clarke et al., 2013).

Freeze-drying of *Lactobacilli* is composed of three main steps: First, the bacterial samples are frozen down at a low temperature below the T_g' of the formulation, typically -50 or -80 °C, the water forms ice crystals while the precipitant of bacteria and formulant forms the amorphous glass phase (Fonseca et al., 2015). Then comes the primary drying step during which the pressure is kept very low and the shelf temperature is increased, usually to -20 °C to enable sublimation of the frozen water (Fonseca et al., 2015). The sublimated water is captured at an even colder part of the freeze-drier called the "cold trap". However, the product temperature must not be increased above the collapse temperature (T_c), otherwise the amorphous glass phase collapses into a viscous phase. The pressure used does not influence T_c much directly (Fonseca et al., 2004a). However, the lowered pres-

sure enables the sublimation which cools the samples; low pressure also lowers the surrounding temperature by removing most gas molecules (Jennings, 1999). Also, as the water sublimates and the product's water content decreases, the T_g' increases and higher temperatures without collapsing become available. Lastly, comes the secondary drying step or the "step up" during which temperature is gradually increased up to between 20 to 30 °C, this makes the last residual water to evaporate. This water was never a part of the ice crystals and therefore could not sublimate during the primary drying (Fonseca et al., 2015).

Bacteria have to be properly formulated before freeze-drying, i.e. the growth medium in which they are cultivated must be exchanged to a protective aqueous solution in which the bacteria are up-concentrated. The formulation enables the formation of an amorphous glass matrix during freeze-drying. The amorphous glass phase is needed to protect the bacteria from the harsh conditions which they would otherwise experience during freeze-drying: high osmolality building up outside of the cells, cold stress, ice crystal formation damaging the membranes and the removal of water bound to lipids, cell wall and sensitive proteins leading to destabilization (Brennan et al., 1986; Meng et al., 2008; Fonseca et al., 2015). A properly chosen formulation both protects against the stresses of freeze-drying and assures the best stability during storage. *Lactobacilli* are commonly formulated in solutions of sugars and/or polymers, antioxidants may also be added but the best formulation is a rather species-specific matter (Schoug et al., 2010; Fonseca et al., 2015). "Compatible solute" cryoprotectans can protect the cells from osmotic stress when accumulated by the cells, carbohydrate/polymer cryoprotectants and trehalose can increase the T_g' which leads to a quicker inclusion of bacteria into the protective amorphous matrix (Meng et al., 2008) and prevents loss of the amorphous phase (Schoug, 2009). Other protectants such as fructo-oligosaccharides and fructans can by interacting with membrane lipids protect the membrane from damage during freeze-drying, while sugars can protect sensitive proteins and lipids by interactions as well (Schwab et al., 2007; Fonseca et al., 2015).

In order to achieve good storage-stability, the water activity in the dried powder must be low (Zayed & Roos, 2004), in other words, the samples must be well-dried but a moisture content of 2-5% seems to be the optimum (Schoug, 2009). The cooling rate is also a significant factor on cell viability (Schoug, 2009). If the freeze-dried product will be stored at temperatures above T_c, the amorphous phase will collapse into a viscous phase or crystalize (Schoug, 2009). Nevertheless, it is important that the amorphous glass state is present during the primary step of freeze-drying, as it enables high survival of the freeze-drying process and a more complete drying which will affect subsequent storage-stability dramatically.

T_g' has often been considered as the collapse threshold during primary drying, however, research provides evidence that collapsing actually takes place at the higher T_c (Fonseca et al., 2004a; b; Schoug et al., 2006). T_c is usually very close to T_g' but it can differ substantially when bacteria are present in the solution (Fonseca et al., 2004a). T_g' is independent of initial concentration of the solute and it has been determined to be about -34 °C for *Lactobacillus* suspension in sucrose at cell functional units (CFU) per mL values up to 10^{10} (Schoug et al., 2006). Fonseca et al., (2004a) conclude that T_c is usually 10 °C above T_g' for suspensions of lactic acid bacteria, which would mean that a T_c of -24 °C could be expected for the abovementioned bacterial solution. In (Fonseca et al., 2004a), the T_c of pure sucrose solution was between -33 and -32 °C. In the same study, T_g' and T_c (which were the same in this case) of a solution of 7.5 percentage by mass (wt %) sucrose with 1 wt % sodium ascorbate and 2.4 wt % ferment supernatant were -36 °C. Apparently, the additions of ascorbate and supernatant lowered the critical temperatures. Adding 8% (dry matter) of lactic acid bacteria to this solution raised T_c to -22 °C (although less detrimental collapsing at the microscopic scale began at -26 °C) while T_g' was -35 °C, which is indeed very close to the abovementioned T_g' of the *Lactobacillus* solution in sucrose determined by Schoug (-34 °C).

T_c (and T_g' to a lesser extent) are also increased more by higher cell densities, and bigger, longer cells can raise T_c by several degrees more than very short cells (Fonseca et al., 2004b). Schoug (2009) mentions that secondary drying should be done below T_c too. Since the usual temperature for secondary drying is about 20 °C (Fonseca et al., 2004b; a, 2015), it is important that the moisture level has decreased substantially during primary drying.

As it was mentioned in the introduction, freeze-thaw survival (and hypothetically storage-stability) can be affected by preconditioning of the bacteria with various stresses and with different growth conditions. Such procedures alternate the molecular profile of the cells (Meng et al., 2008; Schoug et al., 2008; Liu et al., 2014) and fermentation conditions influence cellular morphology (Palmfeldt & Hahn-Hägerdal, 2000; Koch et al., 2008). However, a potentially important fact is that the growth phase, in which the bacteria are during harvest, can be a preconditioning factor as well (Meng et al., 2008). This was also observed to affect FT assay survival in the co-supervisors' previous studies on *L. reuteri* DSM 17938 (Garcia, Armando Hernández et al., unpublished). A problem arises because the pH and temperature will affect the duration of different growth phases. Therefore, if growth phase is not intended to be a variable of the analysis, it should be controlled.

2.3 Accelerated storage

Freeze-drying bacteria is a good way of extending their shelf-life and reducing the need for costly and unpractical storage conditions such as low temperature (Fonseca et al., 2015). Nevertheless, the stability of freeze-dried bacterial powders is dependent on the temperature (Meng et al., 2008) and higher temperatures decrease the stability. This fact can, however, be used with advantage to model extended deterioration of powders within a shorter experimental timespan, by storing at a higher temperature – an accelerated storage assay (Franks, 1994). However, caution must be used in extrapolation of accelerated storage stability to standard storage stability: if the product undergoes a stepwise reaction towards deterioration or if phase transition occurs at the storage temperature, the extrapolation may be misleading and simple Arrhenius equation based extrapolations would not be appropriate (Franks, 1994).

2.4 Growth Advantage in Stationary Phase (GASP) mutants

Bacteria that grow in a batch cultivation, a type of cultivation where the amount of nutrients is predefined and limited and no new substrates are added during growth, go through four well recognized phases. The first phase after inoculation is the "lag phase" in which the growth rate is very low and the bacteria are adapting their gene expression to the new environment. After that comes the "exponential phase" also called "log phase" or "growth phase" in which bacteria grow exponentially at the maximum growth rate allowed by the conditions. Then comes the stationary phase, when the culture has suddenly run out of nutrients and waste products have accumulated. At this point, the bacteria stop dividing and their gene expression becomes adapted to starving conditions and stress tolerance (Zinser & Kolter, 2004; Bačun-Družina et al., 2011). Eventually comes the death phase, in which the bacteria begin to die. This has been explained by different theories (Finkel, 2006): either that the bacteria have too little energy to support cellular maintenance functions and become damaged by their own waste products or that the death phase is in fact a programmed collaboration strategy of the bacteria where only some mutated bacteria do not obey the program and can reinitiate growth. This new growth is also called "long-term stationary phase" (Finkel, 2006), as it replaces the death phase. In either way, this is where Growth Advantage in Stationary Phase (GASP) mutants come in. The best survivors in the death phase are the mutants which possess some advantageous traits in the new situation where nutrients come from the remains of dead cells. In *Escherichia coli* for example, it is known (Zinser & Kolter, 2004; Finkel, 2006; Bačun-Družina et al., 2011) that GASP mutations enhance amino acid catabolism which helps in surviving on cell debris, enhance

tolerance to stressful pH and osmolality and prevents the cell from going into starvation mode. The GASP phenomenon has been studied extensively in *E. coli*, summarized by (Zinser & Kolter, 2004; Finkel, 2006; Bačun-Družina et al., 2011) and some studies have been done in other Gram negatives and *Mycobacterium*, summarized by (Zinser & Kolter, 2004), which also yielded GASPs. However, few studies have been done on this subject with Gram-positive bacteria (Bruno & Freitag, 2011; Sewell et al., 2011), the group to which *Lactobacilli* belong.

A characteristic feature of GASP cultivations is that new mutant generations continue to appear in consecutive waves throughout the prolonged stationary phase, each generation having gained another mutation which makes it even more fit in the stationary phase environment. Thus, each new generation quickly takes over the culture, displacing the older generations (Figure 1).

The behavior of GASP cultures is often depicted as well defined peaks in CFU/mL over time (Finkel, 2006; Bačun-Družina et al., 2011). However, the same sources explain that the total CFU/mL remains rather constant in the long-term stationary phase, with only tiny fluctuations representing the consecutive GASPs, while the well-defined peaks represent the CFU/mL counts of specific GASP variants/"generations". It is explained that while the total CFU/mL is rather constant, there is a highly dynamic equilibrium between the dying old cells and the growing GASP cells of the newest generation (Zambrano & Kolter, 1996; Finkel, 2006) which has been experimentally proven (Zambrano et al., 1993; Finkel & Kolter, 1999). According to a review (Finkel, 2006) by the author who also participated in many of the experiments in this subject, the first generation of *E. coli* GASPs rarely appears earlier than after 8-10 days of cultivation. Furthermore, it has been demonstrated that when higher-generation GASPs are introduced as a highly-diluted minority into stationary phase cultures of earlier GASP generations or non-mutated young cell-lines, the higher generation GASPs take over the culture (Zambrano et al., 1993; Finkel & Kolter, 1999). Additionally, the viability of prolonged stationary phase GASP cultivations has been measured for up to 60 months, still maintaining significant CFU/mL values (Finkel, 2006).

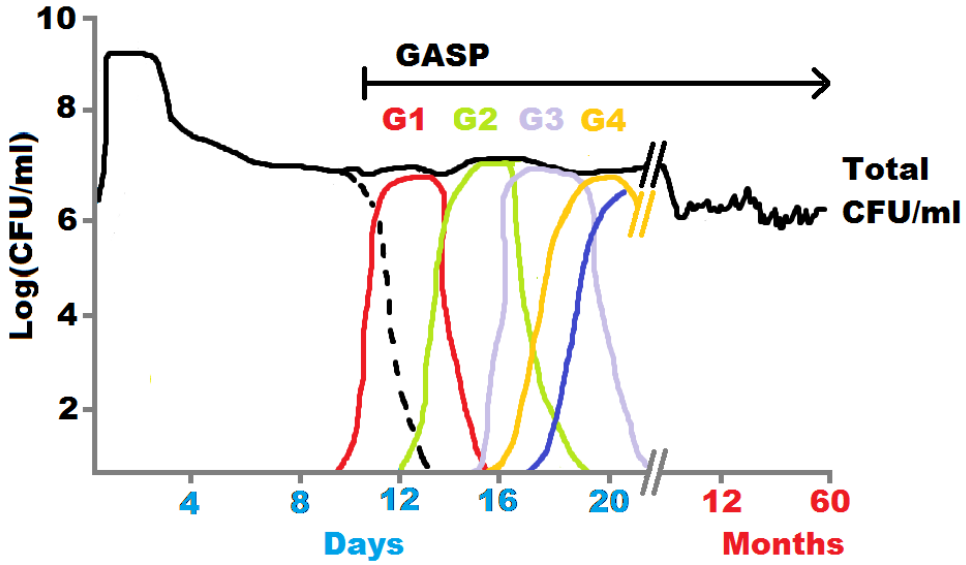

Figure 1. Representation of the growth of an *Escherichia coli* cultivation starting at inoculation and going far into the prolonged stationary phase. The black solid line represents the total CFU/mL count in the culture, it peaks at around 1×10^9 CFU/mL during early stationary phase. At around day 10 from inoculation, the original "cell line" (now represented by the dashed black line) is displaced by the first generation of GASPs (G1), and so on. The total CFU/mL stays at a significant level even after 60 months. Based on (Finkel, 2006).

The authors of a GASP study on Gram-positive bacteria (Sewell et al., 2011) claim that *Enterococcus* (a Gram positive) does exhibit the GASP phenomenon, however, that not all Gram positives display it, since it could not be observed in *Staphylococcus*. These claims are based on observed fluctuations in viable count during stationary phase of *Enterococcus* but not *Staphylococcus*, when measuring viable counts for 85 hours (3.5 days) only. That fluctuations in viable count are a sign of GASP appearance, the authors refer to the paper by (Finkel, 2006), although this source explains that the total CFU/mL in fact remains quite constant during prolonged stationary phase. The cells sampled during the viable count peaks in (Sewell et al., 2011) are shown to have enhanced pH and heat stability and hence are identified as GASPs. However, it is important to remember that even the extremely fast growing *E. coli* needs at least 8 days of cultivation before any GASPs can be found (Finkel, 2006) and also that the main GASP mutations confer enhanced amino acid scavenging, while general stress tolerance is ubiquitously exhibited by nonsporulating wild type cells shifting into starvation mode during the stationary phase (Zambrano & Kolter, 1996; Zinser & Kolter, 2004; Bačun-Družina et al., 2011). Very unfortunately, the stress assays seem not to have been done on the *Staphylococcus* to see if it also gained increased tolerance despite no cell count fluctuations! One more GASP study on Gram positives could be found, a very elegant one, done in the same manner as the classical GASP studies. (Bruno &

Freitag, 2011) could demonstrate GASPs in 12-day old cultures of *Listeria monocytogenes* which outcompeted non-GASP *Listeria* in fresh culture and the trait was stable in its heritability. This time, like in the classical GASP papers, there is no fluctuation of the total CFU/mL during the long-term stationary phase and the CFU/mL stayed at 10% of the maximum.

2.5 Isothermal microcalorimetry

An isothermal calorimeter is a device which can measure the heat production (heat flow) of various types of processes (e.g. chemical, biological and phase transitions) with great sensitivity and display it as a heat flow curve in real-time. The reason it is called "isothermal" is because the device keeps the temperature of its samples at a constant value and measures how much heat that needs to be added or removed in order to do so. Isothermal microcaloritmetry (IMC) can be used in many fields of microbiology, for example in clinical settings to quickly identify infectious agents or the efficiency of antibiotics and in environmental and food science (Braissant et al., 2015). The metabolic processes in a sample are reflected by characteristics of the heat flow curve since metabolic processes are correlated with enthalpy production. IMC is beginning to win ground in the microbiological fields as it has many unique advantages over many classical methods (Braissant et al., 2015). It is especially praised for its sensitivity and flexibility when it comes to the physical characteristics of the samples, for example both liquid and opaque solid media can be used.

2.6 Statistics

This section will introduce the statistical concepts relevant to this study and also briefly mention some alternative methodologies which were considered.

Identifying the influential factors for investigation and procedures for minimizing random and systematic errors are a crucial basis for quantitative research (Miller & Miller, 2010). However, when a study builds upon previous results, the selection of factors of interest may not require screening experiments and many standard laboratory procedures for reducing error may already be established.

A basic principle of scientific research is reproducibility. Statistical methods are used to quantify and visualize this reproducibility and so they need repeated experiments (Vaux et al., 2012). Some common terms used when referring to independently repeated experiments are "experimental repeats", "sample sizes" (Vaux et al., 2012) and "biological replicates" (Fay, 2013).

Before applying a statistical method, one must go through the data and reject outliers. The most basal outliers for removal are the obvious ones (once effort has been put into finding them) which have some logical explanation in the lab notes, for example if the quality of a freeze-dried sample visibly diverged from the rest or some mistake was done. After this, there is an option to do a statistical outlier test, for example Grubb's test (Miller & Miller, 2010). Grubb's test requires a normal error distribution of the data. This can also be tested but a problem arises – the outliers may impact the normality of error distribution (Fay, 2013)! Therefore, a normality test could be performed before and after Grubb's test and the removal of outliers. In general, one should be cautious in this procedure and having experience with the kind of data one is processing helps in the situation.

Transformations of data are possible and sometimes needed before the statistical analysis. Many analyses require a normal (or at least not-too-far from normal) error distribution which means that a histogram of the distribution of the errors ("residuals") should have similarity to the Gaussian curve. To achieve this, the data may have to be mathematically transformed. A popular transformation in biological sciences is the log-transformation because biological data often displays the "log-normal" distribution (Fay, 2013). However, transformed data may be hard to interpret in practical terms. Sometimes, it is feasible to back-transform the data after the statistical analysis. However, back-transforming from logarithms can only give the geometrical means and not the more commonly used arithmetical means (Montgomery, 1991; Fay, 2013). The geometrical means are similar to a median in the sense that they will show the most "typical" values of a data series; these means are not affected much by extreme values at the ends of the dataset (Fay, 2013).

A basic and very useful statistical concept is that of the confidence interval (CI). The CI tells within what interval the true value of a mean is likely to reside within a given probability (usually 95%). This information is easy to visualize so it helps to judge the results. The CI can easily be calculated with software, however it is dependent on the distribution of the data and so normally the assumption of normal distribution (Miller & Miller, 2010; Fay, 2013). Back-transforming CIs obtained in the logarithmic scale can only give the CIs of the geometric means (Montgomery, 1991).

The most basic statistical test for significant differences is perhaps the Student's t-test. In a t-test, two means are tested against each other in the light of their variances and sample sizes. A so-called t-statistic is calculated, which is compared to tabulated critical t-values. The critical values depend on the number of degrees of freedom (d.f.):

$$d.f. = sample\ size_{mean1} + sample\ size_{mean2} - 2 \qquad (1.)$$

and the predefined significance threshold (p-value) (Miller & Miller, 2010). The p-value is popularly set at 0.05 in biological research which means there is a 5% risk of accepting a random variation as a significant finding (a 5% risk of making a type I error). The most sensitive kind of t-test is the paired t-test, when applicable. It can be used when two different kinds of measurements or treatments have been done on the same material. This could for example be two treatments paired within batches and tested against each other. By comparing the differences between paired means, instead of just the overall means, the paired t-test separates the variance due to batches from the variance due to treatments which would otherwise be confounded (Miller & Miller, 2010).

When more than two means or treatments are being compared, the t-test faces a problem. Namely, the "multiple comparisons problem": the more tests that are being made with a 5% risk of having a type I error, the higher will this error inflate for the analysis as a whole. This total error is called the "family-wise error rate" (Fay, 2013). Fortunately, there are family-wise error rate corrections available such as the simple Bonferroni correction (Fay, 2013), in which the predefined p-value (say 0.05) has to be divided by the total number of comparisons made in the statistical test. Only comparisons with a p-value below this new lower limit can be accepted as significant but still at the p=0.05 level. The Bonferroni correction has low power and is acceptable for small sample sizes and there are other correction methods for available for different situations. However, there are statistical methods other than the t-test which are specially designed for dealing with multiple means as well as multiple factors (multivariate methods). Perhaps the most common one is "ANOVA" with the Tukey test. These more advanced methods are by definition family-wise corrected.

In the ANOVA method (Analysis of Variance), values called "F-statistics" are calculated from the ratio between variances attributed to each factor (or interaction of factors) and the variance which can only explained by random error (Montgomery, 1991; Langsrud, 2003; Miller & Miller, 2010). These variances are expressed as "mean of squares" for the respective variance source. Then the ANOVA performs so called "F-tests" by comparing the F-statistics with critical values, similarly to what is done in the t-test. If the F-statistic is above the critical value belonging to the pre-selected significance level (the chosen p-value, often 0.05) then a statistically significant finding is reported (with a maximum 5% risk for type I error). Using statistical software, an exact calculated p-value is reported and if it is below the p-value pre-selected by the researcher, it is a statistically significant finding. If the calculated p-values are much lower than the pre-selected ones, it indicates that the risk of a type I error may be even lower than what was judged as acceptable in the analysis.

The sums of squares (needed for the F-tests) can be calculated in different ways in ANOVA, called type I, type II and type III sums of squares. Type III is the most common standard as it is considered to cover all situations and is the usual default setting in statistical programs (Langsrud, 2003; *Help - IBM SPSS Statistics*). Some sources however, say that for an "unbalanced" design (when the sample sizes of the means are not equal) type II may be more appropriate (Langsrud, 2003).

Factors under study, e.g. "treatment", pressure or temperature, are called independent variables while the measured outcome parameters, such as yield or stability, are called dependent variables. Splitting the variance by defining more factors, such as batches, vials, bottles or dilution series can make the ANOVA more sensitive, by making the random error term smaller (Miller & Miller, 2010; Krzywinski & Altman, 2014a). When defining "nuisance variables" (variables uninteresting to us but useful for increasing sensitivity) such as batch number into the analysis it is called "blocking" as data is divided into blocks. Furthermore, variance splitting can give quantification of how great each source of variation is, something that may be useful for scientific conclusions and planning of future experiments, for example by realizing that a certain experimental step is over- or under-replicated. However, blocking and in general defining more factors inevitably reduces the number of degrees of freedom (Krzywinski & Altman, 2014a). This has the opposite effect on the sensitivity of the analysis, the less degrees of freedom there are, the higher becomes the critical value of the F-statistic to be reached in order to have a statistical significance. Therefore, blocking is only worth applying if the variance due to the blocking factor is indeed large (for example large variation between batches) so that the effect of variance separation is greater than the effect of loss of degrees of freedom (Krzywinski & Altman, 2014a). The efficiency of blocking in an ANOVA can be analyzed *de facto*. If it is uncertain whether using blocking will be beneficial when designing an experiment, it is recommended to do blocking – afterwards it can be assessed if it was beneficial and the potential losses due to the reduction of number of degrees of freedom are usually small (Montgomery, 1991). To calculate the benefit of blocking in some basic experimental designs, a simple formula can be employed (Montgomery, 1991). Furthermore, just looking at the size of the ratio of mean square of the block factor to the mean square of the error provides rough information about the benefit of blocking (Montgomery, 1991). If it turns out that blocking was not beneficial, the future experiments should be completely randomized designs.

There are two kinds of independent variables: fixed-effect factors and random-effect factors. Fixed-effect factors are those which can be controlled by the experimenter such as the treatment. Random-effect factors are those which cannot be controlled and vary randomly like the random error but can be identified and isolated as a source of variance. Batches are often treated as a random factor, because

they can be seen as a random sample from endless possible batches (Montgomery, 1991). The mathematical treatment in ANOVA differs slightly depending on which type a factor is (Montgomery, 1991; Miller & Miller, 2010).

An ANOVA with one factor is called a "one-way ANOVA" and adding more factors will make it a "two-way ANOVA" and so on. At the heart of the ANOVA lies a mathematical model created for the experimental design, an ANOVA with one fixed and one random factor (a randomized block design) with no anticipated interaction between these factors will be represented by the model in equation 2 (Montgomery, 1991):

$$y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij} \qquad (2.)$$

where y_{ij} is the (ij:th) observation, μ is the overall mean of all the treatments, τ_i is the effect of treatment i, β_j is the effect of block j and ϵ_{ij} is the random error term accompanying each measurement. This is an additive model with no square terms because it does not consider interactions.

There are different experimental designs available when doing an ANOVA. In the completely randomized design (CRD), no blocking is used even if batches are present in the experiment; the treatments are completely randomized over all batches (Montgomery, 1991). The block design comes in three main versions (Montgomery, 1991). The randomized complete block design (RCBD) has all treatment levels repeated in all blocks and these treatment levels are randomized only within blocks. The balanced randomized incomplete block design (BRIBD) does not contain all treatment levels in all blocks but each pair of treatment levels occurs together in a block the same number of times. Furthermore, each block contains the same number of treatment levels which are randomized within the block (randomized order of runs, machines used, etc.). Incomplete block designs may be necessary when it is physically impossible to fit all treatments into one block (the block is for example a batch) or as a means to save resources. Finally, there is the partially balanced randomized incomplete block design, in which the treatment level pairs do not need to occur together the same number of times. However, certain mathematical restrictions must be followed to keep the partial balance, which is out of the scope of this study.

So, an ANOVA compares the means of the dependent variable to each other and looks for significant differences with respect to levels of factors. However, the ANOVA of a randomized incomplete block design (RIBD) computes and compares something else than these simple mean values. It uses the "block adjusted means", in SPSS called the "estimated marginal means" (EMMs), which are the treatment means adjusted by the effect of the blocking factor(s). In other words, EMMs are the treatment means predicted by the statistical "model" of the ANOVA, when the effects of the other factors have been compensated for. So, if block-

ing has been used, the EMMs of the treatments are adjusted by the block means, according to which treatments were present in which blocks. This can be considered to be a more accurate representation of the treatment means than the unadjusted means. For example, if batch x always gave the highest values, the means of treatments which contained batch x in their replications will be slightly lowered. SPSS also provides the 95% confidence interval (CI) for the EMM.

The ANOVA only tells if there are *any* significant differences between the tested levels of a factor, but not between which levels. Usually, an ANOVA is coupled to a *multiple comparison* method for that purpose (Montgomery, 1991), often also called a post-hoc test (Fay, 2013), this latter term being ambiguously overused when it is in fact a special case of multiple comparisons. A popular multiple comparison method for making pairwise comparisons between all treatment means is Tukey's Honest Significant Difference test (Tukey-HSD), commonly referred to as a post-hoc test (Fay, 2013). Tukey-HSD will compare all the means to each other for individual significant differences. It uses the mean square error term from the ANOVA analysis and so it is also influenced by variance splitting. If the design is unbalanced, a modified test, "Tukey-Kramer" must be used. SPSS does this automatically when provided with unequal sample sizes (*IBM Does SPSS offer Tukey-Kramer post-hoc tests? - Sverige*, 2016).

Unfortunately, the Tukey-HSD can only be performed on the unadjusted means in SPSS, i.e. not the EMMs. This means that the effort of blocking is partly lost when it comes to Tukey-HSD, only the error mean square is reduced (due to splitting of variance into more factors in the ANOVA) but this always happens at a cost of losing degrees of freedom. Using the batch-adjusted EMMs uses the full potential of blocking. A way of using EMMs in Tukey-HSD manually and also get the resulting p-values is presented in section 3.5 of Material and methods.

A very important aspect of any statistical significance test is "power". As already mentioned there is a predefined risk of making a type I error (falsely rejecting a null hypothesis or in other words assuming a difference when there is none). Naturally, there is always also a risk of making the opposite kind of error: falsely retaining a null hypothesis or in other words missing a difference when there is one. This is the type II error and power is a measure of it according to equation 3 (Montgomery, 1991):

$$Power = 1 - type\ II\ error\ probability \qquad (3.)$$

so power gives the probability of nailing a true difference. Power depends on the size of the difference between treatments to be observed relative to the standard deviation, i.e. the "effect size" and also the number of experimental repeats and the desired significance level (p-value) (Montgomery, 1991; Krzywinski & Altman, 2014b). The greater the effect size and number of experimental repeats, the

higher power can be achieved. It is up to the experimenter how much power is desired but 80-95% seems to be the popular range (Lakens, 2013). When designing an experiment, it is important to assess the power so that it is adequate for obtaining conclusive results. One has to decide about the minimal difference between treatments that will be scientifically interesting as a basis for defining the effect size and how many experimental repeats that can afforded. When it comes to experimental repeats, it is very important not to confuse these with "technical replicates" (Fay, 2013) or just "replicates" (Vaux *et al.*, 2012). Experimental repeats are independent repetitions of an experiment while technical replicates are just repeated measurements within an experiment, such as replicated dilutions series or colony counts. Only experimental repeats can contribute to the sample size when using a statistical method to test the main hypothesis (Vaux *et al.*, 2012; Fay, 2013).

One should not rely on "observed power"/"post-hoc" power values, the power values given in the output of each SPSS ANOVA since they merely reflect the observed p-value. These observed power values show the power of the test with the assumption that the true difference between means is exactly the same as the measurements show. For power analysis to make sense, one should first decide what the difference between two means should be in minimum to still be interesting scientifically, and only from that position can the true power of testing this predefined hypothesis be calculated. This difference would be a "scientifically interesting difference" and would be expressed in terms of the *effect size*. The definition of a "scientifically interesting difference", together with the desired significance level and information (or informed guesses) about the variance (sum of squares) from previous experiments can show either the required total sample size needed to achieve a desired power level or the power level achieved with a given total sample size. There is software for assessing power, for example G*Power (*Universität Düsseldorf: G*Power*), SPSS Sample Power (*IBM SPSS SamplePower*, 2012) and many others.

2.6.1 Statistics – the assumptions

The first assumption of ANOVA is that the uncontrolled variation, the "random error" is truly random and not subject to some bias. This is achieved by using randomization and blocking in the design of the experiments (Miller & Miller, 2010).

Another assumption of ANOVA is homogeneity of variance: that the variance of the random error is not affected by treatment (Montgomery, 1991; Miller & Miller, 2010); in other words, the experimental repeats of each treatment will have an equal spread of their residuals. Some texts articulate this assumption as homogeneity of variance of the measured *data values*, since this often implies homoge-

neity of variance in the *residuals*. This is, however, unprecise; if blocking has been used, the residuals may behave differently than the variance of measured data values. If it is assumed that the model provided by ANOVA is an additive one (no interactions) according to equation (2), the block effect term β_j can vary for each individual data value, depending on which batch it came from. Therefore, since the residuals are the observed values minus the values fitted by the model, the residuals may differ in homogeneity of variance from the raw data values. The values fitted by the model in the randomized block design with one fixed and one random factor are given by equation (4) (Montgomery, 1991):

$$\hat{y}_{ij} = \bar{y}_{i.} + \bar{y}_{.j} - \bar{y}_{..} \qquad (4.)$$

where $\hat{y}_{ij}$ is the fitted value corresponding to the observation at treatment i and block j, $\bar{y}_{i.}$ is the mean of the observations at treatment i, $\bar{y}_{.j}$ is the mean of the observations at block j and $\bar{y}_{..}$ is the overall mean of the observations.

Often, Levene's test can be used to test for homogeneity of variance in a statistically formal way. However, this test is only available for one-way ANOVA. In a multifactorial ANOVA, homogeneity of variance can instead be judged "by eye" from residual plots, in which residuals are grouped by the factor of interest, e. g. by treatment. This way, the spread (the variance) of each factor level can be visualized and compared. There are some proposed rules of thumb for judging homogeneity of variance from such residual plots (*A Rule of Thumb for Unequal Variances | University of Virginia Library Research Data Services + Sciences*). The optimistic rule is that the ratio of the widest variance over the narrowest variance should not exceed 3, but this is for large sample sizes. A safer ratio, especially with small sample sizes has been suggested to be 1.5.

What to do if homogeneity of variance is not kept? If the homogeneity of variance assumption is unsatisfied, the ANOVA may be flawed. Solutions to this can be transformation of data which can improve the homogeneity (Montgomery, 1991). Otherwise, non-parametric statistical methods may have to be applied (Montgomery, 1991; Miller & Miller, 2010).

Another assumption of ANOVA is normality of distribution of the residuals (Montgomery, 1991), also called "normal error distribution". However, ANOVA is not very sensitive to departures from this assumption (Montgomery, 1991; Miller & Miller, 2010).The Shapiro-Wilk test is a statistical test of normality and keeping with the standard of 0.05 significance threshold, a lower value than 0.05 will imply that normality is violated with 95% confidence. Another way to assess the normality assumption is by examining the quantile-quantile probability plot (Q-Q plot) of the distribution of the residuals against the normal distribution (Wilk & Gnanadesikan, 1968). If the points on this plot align into a straight line, it indicates a normal distribution of residuals. Deviations from such straight line are

classified into kurtosis and skewness. Kurtosis describes the clustering of the distribution of residuals against the central point while skewness describes the asymmetry of the distribution (*IBM Knowledge Center - Display Statistics*).

Many texts are unfortunately unprecise with the normality assumption and write it out as “normality of distribution” or “samples being drawn from a normal population” (Miller & Miller, 2010) which implies that the assumption applies to the distribution of measured data. It is true that normal distribution of data will usually mean normal distribution of error, so one can test the data for normality. However, more possibilities can open up when one is testing the residuals for normality instead. Most importantly and in contrast to the data values, all the residuals from the different treatments can be tested for normality together as one group, instead of testing the normality of each treatment group separately (Montgomery, 1991). This helps to get a sensible and reliable normality test when the sample sizes are small (few measured values per treatment). If the normality assumption is grossly unsatisfied, the options are either finding a transformation for the data or using so called non-parametric and robust methods (Miller & Miller, 2010).

A final assumption when having a multifactorial design and not intending to test for interactions between factors (and to be morally correct - not expecting any interactions) is the additivity of the mathematical model behind the ANOVA (Montgomery, 1991). This means that a linear model with no square interaction terms adequately explains the observations. However, in a two-factorial block design where one factor is a random-effect factor, this assumption is unnecessary (Montgomery, 1991). Otherwise, plotting the residuals against the dependent variable values predicted by the model and looking for non-random patterns in the scatter is a way to detect unexpected interactions between the factors when using an additive model (Montgomery, 1991).

3 Material and methods

3.1 Overview of the correlation experiment

A schematic overview of the correlation experiment designed to address the first objective of the thesis is presented in Figure 2 below.

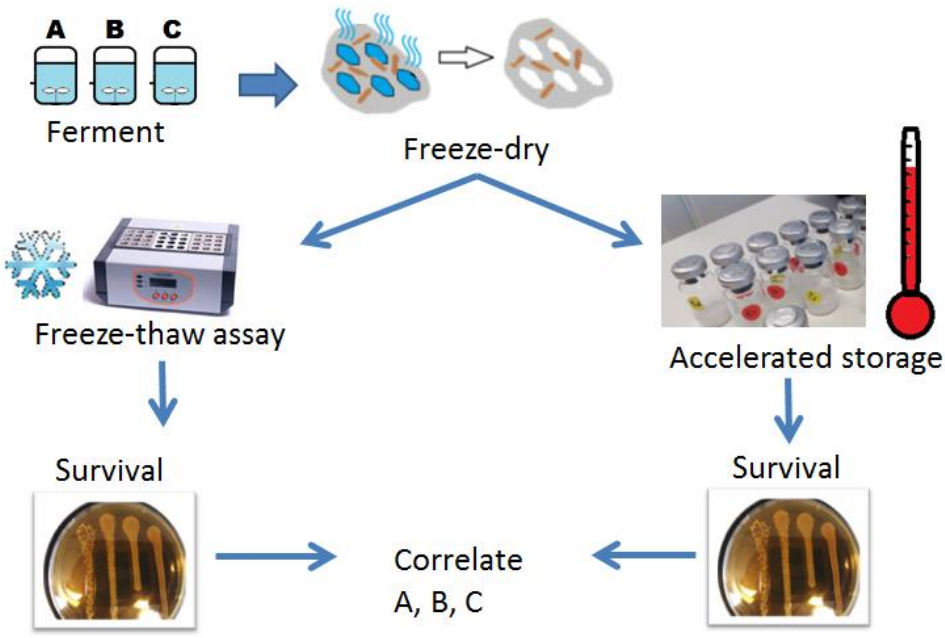

Figure 2. An overview of the correlation experiment, in which the relationship between survival of the freeze-thaw assay and survival of the accelerated storage assay was investigated, depending on cultivation settings (treatments A, B and C).

3.1.1 Fermentation and formulation

Lactobacillus reuteri DSM 17938 was cultivated in three Jenny bioreactors from Belach Bioteknik AB at three different settings of pH and temperature. These three "treatments" were "A" = 37 °C, pH 4.5; "B" = 32 °C, pH 4.5; and "C" = 32, °C pH 6.5. The three treatments were chosen on the basis of the co-supervisors' earlier studies of FT-assay stability as being the best, the midst and the worst preconditioning, respectively (Garcia, Armando Hernández et al., unpublished).

Seven batches were run in total, allowing each treatment to be replicated four times, see Table 1. In batches 1-3, three bioreactors were employed one for each treatment (A-C). In batches 4 to 7, only two bioreactors could be employed at a time with treatments A-C evenly distributed between batches. Batch 7 was a repetition of the same treatments as batch 5, because batch 5 was a collapsed batch and lost viability quickly in accelerated storage (more on this in section 4.3 of Results). These two batches were never used in the statistical analysis for the same timepoint of storage-stability measurement, leaving the sample size at four, as seen in Table 1.

Table 1. *Example of the design of experiment for storage-stability measurement after 7 days of accelerated storage. Batch 3 was never freeze-dried and so did not take part of the study, batch 7 was not investigated at the 7th-day timepoint.*

Batch number	Available treatment replicate		
1	A	B	C
2	A	B	C
4	A	-	C
5	A	B	-
6	-	B	C
7	-	-	-

All liquid cultivations, including the main fermentations as well as the preparations of inoculum were done in De Man, Rogosa and Sharpe (MRS) broth medium for *Lactobacillus* from Merck (Darmstadt, Germany). For inoculum, cell bank was thawed on ice for 20-40 minutes, a single cell bank could be re-frozen once and used for the next batch (small test experiment showed that cells from a doubly frozen cell bank grew equally well in inoculated bottles). A 10-mL bottle of MRS-medium was inoculated with 10 µL of the cell bank and put into 37 °C for six hours. After six hours, high turbidity had developed and the contents were homogenized before 3 x 500 µL was transferred to three 100 mL Duran® bottles filled with 100 mL MRS medium. These bottles were cultured at 37 °C overnight for about 15-17 hours. Large amounts of (presumably) CO_2 bubbled out of the finished cultures when these were homogenized. The starter culture was drawn into

three inoculation syringes and the syringes were knocked on during drawing of the inoculate as to minimize the volume occupied with CO_2. The optical density at 600 nm (OD_{600}) of the finished 100 mL starter cultures was also checked and it was always around 8. All of the work was done with use of thorough aseptic technique and with the aid of a sterile bench.

The reactors were filled with 1 L of MRS medium each, autoclaved and let cool to room temperature. Inoculation of the reactors was done by adding 48 mL inoculate (40 mL in batch 1) using sterile needled syringes through ethanol-flooded rubber membranes at the reactors. The mixing of the reactors was set at 100 rpm and no aeration was applied. This together with pH and temperature (specific for each treatment) was controlled by a computer connected to the bioreactor setup, including pH and temperature electrodes in the reactors. The pH was adjusted by automatic pumping of NaOH solution 20 % w/v (it is possible but untraceable that for some batches, another solution was used which was made by diluting 200 g of NaOH in 1 L of water). Acid (H_3PO_4, 3 M) was added manually with a syringe, very slowly, to adjust pH in rare cases when this was needed.

Since different cultivation conditions lead to different durations of growth phases (for example treatment A with its high, growth promoting temperature would reach the stationary phase quickest), the harvesting times had to be adjusted accordingly. Based on the co-supervisor's previous results (Garcia, Armando Hernández et al., unpublished), it was decided to divide the observed stationary phase (between the end of exponential growth phase and start of death phase) into three equal intervals: early, middle and late stationary phase. The early stationary phase starts when the growth, measured by OD, planes out. It was desirable to harvest all biomasses at an equivalent phase, to minimize the effect of this additional factor which in this case was a factor beyond the experimental scope. It was decided that it would be most practical to harvest the biomasses individually as soon as each reaches its own middle stationary phase. According to the previous unpublished studies, each of the chosen treatments would not only reach early stationary phase at different times but would also take a somewhat different time to reach middle stationary phase from that point. OD_{600} was measured (in triplicates) both to spot the early stationary phase of treatment A as a reference for when to start the harvesting, and to check if the cultures are growing well. Cultivation A usually reached early stationary phase after 8.5 hours, the other cultures were, based on the earlier unpublished studies, assumed to reach early stationary phase 4 hours later (B) and 2 hours later (C). In most batches, it took roughly the same time for cultivation A to reach early stationary phase and for a few cases, it had to be assumed that the stationary phase was reached just after the end of the measurements as it got too late to continue the readings. Harvesting was aimed at the middle stationary phase which was assumed (based on the previous unpublished results)

to be reached after 14, 13 and 13 hours from the start of the early stationary phase for treatments A, B and C, respectively. However, in practice the harvests were delayed by on average an additional hour, because of practical issues. Table 2 summarizes the time-schedule of the cultivations.

Table 2. *The approximate time-schedule of growth phases and harvesting of the three cultivation types (A-C). Based on the co-supervisors' previous results (Garcia, Armando Hernández et al., unpublished). ESP = Early Stationary Phase, MSP = Middle Stationary Phase. The rightmost column shows when the harvests were done in practice*

Treatment	Approximate time to reach ESP (h)	Approximate time to reach MSP (h)	Approximate time to harvest (h)
A (37 °C, pH 4,5)	8.5	22.5	23.5
B (32 °C, pH 4,5)	12.5	25.5	26.5
C (32 °C, pH 6,5)	10.5	23.5	24.5

Harvesting was done by pumping out the ferment from the reactor through a rubber tube and into a closed bottle, using one of the reactors own feed-pumps. The outlet of the tube was kept at the bottom of the bottle so that to avoid unnecessary oxygen shocking (although this detail was employed starting at batch 4, after some discussions of potential effects of oxygen). Pumping for sample collecting went on for about an hour to obtain 500 mL, and then the sample was gently poured down into a pre-weighted centrifuge flask. The ferment was centrifuged at 4200 rpm, 24 °C for 30 minutes. The supernatant was discarded and the bottle was again weighted to determine the weight of the bacterial pellet. The bacterial pellet was resuspended with sterilized and deionized water up to a total weight of 50 g, then, 50 g of sucrose solution (30% w/v) was added. To resuspend the pellet, vigorous flushing with a large electrical pipette (about 25 mL suction volume) was done for 10 minutes for all cases to be fair with the oxygen shock, no matter how fast a particular pellet dissolved. This resulted in a 15% w/v sucrose formulation and a 5-fold increase in the concentration of bacteria relative to the original ferment. All the numbers and graphs in this paper are based on this up-concentration as the standard, no compensatory re-calculations were done. The formulated and resuspended biomass was, while kept well-mixed, distributed with an electrical pipette into small freeze-drying vials (20mm neck 45x24/1mm), 2 mL into each vessel and usually 23 vessels per treatment in each batch. The vessels were color marked by treatment and arranged randomly in a tray, corks were put in place and the vessels were put into a -50 °C freezer for at least 2 hours to be well frozen before freeze-drying. The remaining formulated biomass was used on viable counts to give a reference starting CFU/mL value for subsequent calculations of survival.

3.1.2 Freeze-drying

Batch 1 and batch 2 were freeze-dried in a highly-automated freeze-drier (LyoStar II™, FTS Kinetics™, Stone Ridge, NY, USA). Batch 3 never got through with freeze-drying since the machine broke, so this batch was not included in the correlation experiment. All the subsequent batches were run in a simpler freeze drier, (FreeZone® 12, Labconco, Kansas City, Missouri, USA).

First, the freeze-drier was cooled down to a shelf temperature of -40 °C. The pre-frozen samples were transferred to the freeze drier as quickly as possible (within several minutes). Three randomly chosen samples had previously been given thermocouples immersed in the biomass in order to measure the actual freeze-drying temperature at different locations of the shelf. As soon as the samples were inside the freeze-drier, the vacuum pump was turned on. The temperature of the "probed" samples was watched. There was always a rise in the temperature of the probes to above -40 °C because of warming up during the transport and the opening of the machine. The pressure had to drop adequately before the shelf temperature could be increased, i.e. starting the actual freeze-drying process. It was aimed for to lower the pressure to about 5 Pa before rising the shelf temperature to -20 °C. This shelf temperature was kept for a few days before it was increased by subsequent 10 °C steps every second hour up to +10 °C for batches 1-5; the final temperature was however increased to +20 °C (in batch 6 and 7) as an attempt to improve the dryness of samples. The final temperature was kept for a final couple of freeze-drying hours.

However, the 5 Pa goal was hard to reach and various starting pressures had to be accepted because of lack of time. Batch 4 presumably reached 6 Pa but the shelf temperature was decreased to -24 °C because one of the probes measured as high as -30 °C too early in the primary drying. The temperature was adjusted back to -20 °C the following day. Batch 5 reached 16 Pa and batch 6 reached 11 Pa but the shelf temperature was decreased to -24 °C because the pressure jumped to 16 Pa when increasing the shelf temperature. It is possible that the temperature was left at -24 °C until the stepping up (there are no notes of increasing it to -20 °C). Batch 7 reached 11 Pa but the shelf temperature was only increased to -30 °C because one of the probes measured as high as -28 °C. The shelf temperature was increased to -20 °C the next day. Note that the pressure gauge was not very precise and would show only certain step-wise values.

After the final stepping up of the shelf temperature, the corks on the vials were closed by the freeze-drier, still at low pressure.

During the making of batch 2, empty freeze-drying vials were pre-weighted and then weighted again with freeze-dried biomass in order to investigate the variability and how much saline (0.9% w/v) was needed for resuspension of bacterial

cakes back into 2 mL liquid (or rather as an approximation, 2 g liquid). The variability was small, with a difference between the heaviest and the lightest sample being only 0.71%. The cakes weighted on average 0.3000 g and were therefore from there on resuspended in 1.7 mL saline for viable counts (except for batch 1 where 1.8 mL had been used at all times). This over-dilution of batch 1 was compensated for in the calculations of CFU/mL in the subsequent viable counts. Importantly, the reference formulated biomass before freeze-drying was omitted from this correction because it did not need any resuspension.

3.1.3 Accelerated storage assay

As soon as the closed freeze-drying vials were out of the freeze-drier, they were sealed with extra metal caps and stored at 37 °C for up to 29 days (accelerated storage assay). One aliquot was used directly for taking an initial survival viable count (VC) and the rest after 3±1 days, 7 days, 14 days, 22 days and 29 days. The discrepancy at around the third day of accelerated storage (±1 day) was due to practical issues. It was decided to pool these measurements into one "3 days" measurement point for the statistical analysis.

Before doing any VC, the look of the bacterial cakes was noted as well as some other characteristics. Because of large possible variability in the samples, a routine for selecting the most average-looking vials of each type was introduced starting from the 3 day VC of batch 4. At the same point, a quality scale was set-up for annotating how much each chosen sample had shrinked both in absolute terms (compared to how much possible shrinkage had been observed in batch 1) and relatively to the other samples of the same treatment (A-C) in the same batch. The scale would give one, two or three minuses in the absolute shrinkage assessment and a plus, zero or minus in the relative part. For example a(---)r(+) would mean that in absolute terms, the sample has shrinked very much but relative to the others of the same type, it is relatively big. The quality scale together with annotation of color and texture was very helpful in identifying some outliers and to see the connection between looks and survival rate.

3.1.4 Freeze-thaw assay

The three cultivation condition treatments chosen for this study were based on the co-supervisors' previous results of a FT-assay (with a different protocol than the FT-assay used for this paper) as the best, the midst and the worst survivor and had the following survival ratios: best - 32 °C, pH 4.5 -> 39%, midst - 37 °C, pH 4.5 -> 9.49%, worst - 32 °C, pH 6.5 -> 2.5% (Garcia, Armando Hernández et al., unpublished).

The protocol applied for the FT assay was adopted from previous research in the BioGaia project (persons. comm, Malin Sundelius, BioGaia AB), with some minor modifications. The freeze-dried bacterial cakes for the FT-assay were kept in the freezer at -50 °C overnight after the freeze-drying. The following day, samples of 0.1 g of the freeze-dried cakes were dissolved in 43 mL saline (0.9 % w/v), technically replicated in duplicates by using two individual freeze-drying vials from each treatment (A-C). The solutions were left to stand for 30 minutes – 60 minutes. A viable count was done on the solutions to obtain starting CFU/mL values for reference. Thereafter, 300 µL of each solution was transferred to 1.5 mL Eppendorf tubes. These Eppendorf tubes were frozen for at least 50 minutes and then thawed in a heating block at 37 °C for 10 minutes. This was repeated three more times until four freeze-thaw cycles had been done. One of these four freezes lasted overnight, which one it was could differ between batches. In the end, the freeze-thawed samples were plated, about 48 hours after the end of freeze-drying.

3.2 GASP experiment

To address the second objective of the thesis, two attempts were made to capture GASPs using IMC. In the first attempt, two sets of small ampoule cultivations were done: one with standard MRS medium (Merck) and the other with 1:2 diluted MRS medium ("half-MRS"). The idea with the half-MRS was that with less nutrient supply, the *Lactobacilli* should accumulate less acidity. This was of interest because it was suspected that a too acidic environment could inhibit the arise of GASP cultures.

The experiment was performed in a pair of isothermal microcalorimeters "TAM Air 8-channel" (TA instruments, New Castle DE, USA). Small, sterilized microcalomietry ampoules were filled with 18 mL of MRS-broth medium and half-MRS-broth medium, 8 ampoules of each. Sixteen more ampoules were filled with 18 mL of water as the baseline references for calorimetry; each reference ampoule shared a two-slot "channel" with a fermentation ampoule. The fermentation ampoules were inoculated with 10 µL of thawed cell-bank. All of the ampoules were closed with sterilized rubber/metal caps, small hooks were screwed halfway through these caps as a means for carefully sending down the ampoules into the calorimeters. To avoid CO_2 buildup, tiny sterile syringe-needles were struck through these caps. The needles were not driven down as far as to touch the liquid in order to avoid any capillary actions. The handling of the ampoules and the caps was done through vinyl gloves to avoid contamination with grease, which could interfere with the readings. The calorimeters were heated up to 37 °C, the baseline was established by the co-supervisor and then all the samples and refer-

ences were carefully submersed into the slots. The cultures were continuously cultured at 37 °C and their heat fluxes were measured for 17 days. At the end, two ampoules of the full-MRS-broth medium type were subjected to VC with single dilution series. The plaque buildup at the bottom of the ampoules had been thoroughly vortexed up into solution. The rest of the ampoules were transferred to a 37 °C incubator as the calorimeter was unavailable for a while. Five days later, two more ampoules of the whole-MRS-broth and one ampoule of the half-MRS-broth were investigated with VCs. Simultaneously, their pH were measured with a pH-electrode.

Due to unsatisfactory results of what was described above, a second attempt of the GASP experiment was done. The second attempt began by composing and testing an array of buffered versions of the half-MRS and whole-MRS media with the aim to prevent the pH from getting too acidic during cultivation by buffering action. The test-cultivations were done to assess the growth of *L. reuteri* and stability of pH in the different media. The cultivations were done in a RTS-1C Personal Bioreactor (BioSas, Riga, Latvia) which uses specialized Falcon tubes as the reaction vessels. The buffer of choice was MOPS (3-(N-morpholino)propanesulfonic acid), initially at the concentration 5.6 g/L, based on previous experiments on *Lactobacillus* for BioGaia (persons. comm, Stefan Roos, BioGaia AB). To make this medium, MOPS in powder form was weighted up and dissolved in 100 mL MRS-broth. From the resulting MOPS-MRS solution, 30 mL was sterile-filtrated into a sterile falcon tube. The tube was inoculated with *L. reuteri* from fresh colonies belonging to the other experiments, straight from a Petri-dish cultivated at 37 °C. The Falcon tube was inserted into the Personal Bioreactor and fermentation was done at 37 °C without continuous spin agitation. The cap of the Falcon tube had a membrane for CO_2 release. The OD (850 nm) was automatically measured every tenth minute (preceded by a quick spin-homogenization) and plotted on a computer connected to the bioreactor. The next day, when the culture was in death phase, its pH was measured with a pH electrode together with a temperature electrode for temperature compensation. pH measurement was also done to the fresh MRS-broth and fresh MOPS-MRS-broth. The whole procedure of cultivation in the Personal Bioreactor and measurement of pH was also repeated with half-MRS-broth medium containing 5.6 g/L MOPS buffer and finally, with half-MRS broth medium containing 11.2 g/L MOPS buffer. In the last case, the pH was measured after three days of cultivation.

At the second attempt of microcalorimetric GASP hunting, things were done in the same way as in the first attempt, except for a mistake due to a misunderstanding: the baseline was not established before inserting the samples. This eventually led to such a solution that the experiment (the collection of heat flow data) was started when samples reached stationary phase. Furthermore, in this experiment

the samples were: eight ampoules of half-MRS broth with 11.2 g/L ("twice-") MOPS buffer freshly inoculated with *L. reuteri*, four ampoules with the old death phase cultures in whole-MRS-broth (taken back from the 37 °C incubator) and four ampoules inoculated with *E. coli* strain TG1 in LB broth medium. The latter cultures were implemented as a means to verify if the GASP phases of *E. coli* cultures (Zinser & Kolter, 2004; Finkel, 2006; Bačun-Družina *et al.*, 2011) are visible at all within the heat fluxes (assuming that the GASP phenomenon would in fact take place). The *E. coli* were not freshly from a cell bank but from a previous cultivation. After eight days of cultivation, two ampoules of the half-MRS twice-MOPS cultures were taken out for VC and pH measurement. The other cultivations were kept and measured in the microcalorimeter for 19 more days.

3.3 Saline experiment

At some point in the middle of the experiments it was discovered that for some time, perhaps two weeks or so, the saline solution used for rehydrating freeze-dried samples had accidently been made at the concentration 0.09% w/v instead of the standard 0.9% w/v. To judge the impact of this mistake on the results, a small experiment was performed to see the effect of saline concentration on VC.

The biomass for this experiment was a surplus vial: after freeze-drying of batch 7 was complete, some samples were put into the freezer at -50 °C for freeze-thaw assay and there were enough of them to also use for the saline test. The next day, one such vial was sacrificed; two dilution series were done with 0.09 % w/v saline and two dilution series with 0.9% w/v saline. The dilution series were plated on Petri-dishes in triplicates.

3.4 Viable counts

The colony cultivations for VC were done in Petri-dishes with MRS-agar for *Lactobacillus* from Merck (Darmstadt, Germany). The biomass was diluted in a ten-fold dilution series (steps of 1:10 with a transfer volume of 100 µL) in 0.9% w/v saline (0.09% in some cases, see "Saline test" in Results section 6). 10 µL of each dilution of interest was dropped down on the agar and let run in one direction by tilting the plate, up to four dilutions per plate (Jett *et al.*, 1997). The dilution series was always technically duplicated. The formulated biomass before freeze-drying was drawn twice and used for two dilution series, the accelerated storage assay VC-measurements were done by using two freeze-drying vials of each treatment and likewise, all VCs in the FT-assay and most in the GASP experiment were based on technical duplicates of dilution series based on duplicates of vi-

als/ampoules. Each dilution series was in turn duplicated on two Petri-dishes, in rare cases on three. In each Petri-dish, the number of dilutions within a countable range varied and each countable dilution was included in the chain of duplications. This bottom step of duplication constitutes the lowest level "technical replicates" which were averaged to give on mean value. Values of the biological replicates (repeated experiments), which means the individual batches, were kept un-averaged for the statistical analysis, in accordance with statistical literature (Vaux *et al.*, 2012; Fay, 2013; Krzywinski & Altman, 2014a).

The inoculated Petri-dishes were incubated anaerobically in tightly closed specialized jars together with applied Anaerocult® reagent (Merck). The jars were kept at 37 °C for about two days until colonies had developed. The maximum number of colonies for a dilution drip to be considered countable (and reliable) was around 180. In a few rare cases, especially when data was sparse, values of up to 239 colonies were accepted. Usually however, the values were in a moderate range (from a few to about a hundred colonies). If some technical replicates (at the "drip" level) of a particular high-end dilution included a count of zero colonies, then this dilution was only included into the total average if there was a higher dilution available which had only zeroes. This was to guard against bias from accidental contamination of samples with the bacteria which would produce a high background noise of colony growth. Because if some samples have the value zero, it makes the non-zero ones more suspect, while the effect of this background noise is not equally drastic on higher colony counts of lower dilutions.

The CFU/mL was calculated as following:

$$\frac{CFU}{mL} = 100 \times (average\ no.\ of\ colonies) \times dilution\ factor \tag{5.}$$

The factor 100 in equation (5) is to compensate for the fact that drops of only 10 µL are used for plating while the CFU is given per mL. The "survival fraction" was calculated by dividing a CFU/mL of interest by the average CFU/mL of the formulated biomass before freeze-drying for a particular treatment (Eq. (6)).

$$survival\ fraction = \frac{\frac{CFU}{mL}\ of\ interest, treatment\ X}{\frac{CFU}{mL} formulated\ biomass\ before\ freeze\ dry, treatment\ X} \tag{6.}$$

3.5 Statistics - correlation experiment

The stability data was thoroughly examined to remove any outliers or unreliable data according to what was described in sections “Accelerated storage assay” and “Viable counts” above. Since the consistency in the remaining technical replicates seemed reasonable and the time was limited, no formal statistical test for outliers was applied.

The design of experiment (*de facto*) was an unbalanced randomized incomplete block design for a two-way ANOVA in SPSS followed by post-hoc Tukey-HSD tests. The fermentation treatment was selected as a fixed factor with three levels (treatments A, B and C) and batch number was selected as a random factor. One set of data was prepared for ANOVA by averaging the technical replicates into means for each batch/treatment combination, and one dataset was $\log_{10}$-transformed at the level of technical replicates which were then averaged into batch/treatment combination means. The $\log_{10}$-transformation was used with the aim to improve compliance of the data to the assumptions of the statistical tests used, as well as to increase their power.

Completely separate ANOVAs were done for each timepoint of accelerated storage stability measurement and for the FT-assay, as well as for the non-transformed and logarithmized versions of these analyses. Table 1 in section 3.1.1 “Fermentation and formulation” above shows an example of the experimental design for accelerated storage assay at 7 days of storage (not all timepoints had the same number of available batches, especially batch 1 was no longer viable after 14 days). The number of biological replicates at each timepoint of storage-stability measurement can be found in Tables 6 and 7 in section 4.3 “Survival in accelerated storage assay” and for the FT-assay in Table 9 in section 4.4. The reduction of batch size from three to two treatments after batch 3 was done because firstly, in batch 4, one of the cultivations failed due to a malfunctioning pH-electrode and there was no replacement for this electrode available within time for further cultivations. Secondly, because operating all three bioreactors at the harvesting day led to very stressful working load for a single person.

The smallest observable value in viable counts was 10^2 CFU/mL (one colony observed per mL when using 10 µL drops), so if no colonies were observed, the count was set to 10 CFU/mL. However, since *survival ratio* values were needed for the statistical analysis and correlation plots, the 10 CFU/mL minimal values were divided by the highest observed initial CFU/mL in the study (7.88×10^9). This was needed for the logarithmized part of analysis because a zero on the logarithmic scale would mean 100% survival. For plots of the CFU/mL over time in this paper though, the value 10 CFU/mL was used as the minimum.

Each ANOVA was accompanied by an analysis of standardized residuals in SPSS to check if its assumptions were satisfied. Normal Q-Q plots and Shapiro-Wilk tests were done to assess the normality of error distribution and additionally, kurtosis and skewness values were evaluated. These tests were done for each separate ANOVA using pooled standardized residuals from all treatment groups (A-C). The results of these tests were compared between the non-transformed and $\log_{10}$-transformed datasets at each timepoint of accelerated storage assay measurement and for the FT assay survival. The homogeneity of variance assumption was assessed by plotting standardized residuals versus treatments (A-C) and calculating ratios between the greatest and smallest within-treatment variance of residuals, in order to compare the result with the rule of thumb ratio of at most 1.5 for small sample sizes. The results were compared between the non-transformed and $\log_{10}$-transformed datasets.

After the ANOVA, Tukey-HSD tests were run to compare the individual means. To be able to use the EMMs instead of the unadjusted means, Tukey-HSD tests were done manually in Excel (since SPSS does not perform Tukey-HSD on EMMs) according to equation 7 (Montgomery, 1991):

$$T_\alpha = q_\alpha(a, f) \times \sqrt{\frac{MS_E}{n}} \tag{7.}$$

where T_α is the critical value for a pairwise difference at significance level α, $q_\alpha(a,f)$ is the statistic for the studentized range distribution at the significance level α, number of compared groups a and f degrees of freedom of the error (provided by SPSS ANOVA), MS_E is the mean square error (provided by SPSS ANOVA) and n is the sample size. Solving for $q_\alpha(a,f)$ and substituting T_α with the observed difference between means gives equation (8):

$$q = \frac{difference\ between\ means}{\sqrt{\frac{MS_E}{n}}} \tag{8.}$$

These calculated q values could be compared to tabulated critical values in order to tell which comparisons were significant. However, to do this conveniently and to also get the actual p-values out of the Tukey tests (rather than a yes/no answer), a Matlab script was employed, the "cdfTukey" (*Cumulative distribution function of the studentized range (for Tukey's HSD test) - File Exchange - MATLAB Central*). When provided with q, a and f, cdfTukey gives the "cumulative distribution function of the studentized range distribution used in Tukey's HSD test" (cdf). Then 1 - cdf = p-value of Tukey HSD.

3.6 Microscopy

Starting at batch no. 4, microscopic studies were implemented as another way of correlating measurable outcomes with storage-stability. For obtaining casual images of cell size and morphology, samples were taken just before starting the harvest or just after finishing the harvest. The samples were kept at room temperature and were observed under phase-contrast microscopes within one hour after sampling. The exception were samples from treatment B, these had to be viewed after 3 hours due to practical reasons in the time schedule of laboratory tasks. For the Gram-stain, samples were taken together with the abovementioned samples and were stained near the time when the abovementioned samples were viewed.

The Gram staining protocol was adopted from (Smith & Hussey, 2005): 1 minute of primary stain -> 2 seconds of washing -> 1 minute of mordant -> 2 seconds of washing -> 15 seconds of decolorizing -> 45 seconds of counterstain -> final wash.

The lens scales (ocular micrometers) seen in the microscopy images in the Results section do not represent µm directly but have to be compared to a µm reference (objective micrometer) for each of the two microscopes used (home lab and course lab microscopes), individually and recalculated into micrometers. This was done with the help of reference photos for each microscope (Figure 27 in Appendix).

4 Results – correlation experiment

4.1 Growth during cultivation

Batch-wise results from measurement of OD_{600} during the growth of *L. reuteri* in bioreactors at the three treatments (A-C), up to the point when the fastest culture (A) had reached stationary phase are plotted as the natural logarithm of OD_{600} over time in Figures 3-5. The entrance into stationary phase was recognized by a halt in the increase of $\ln(OD_{600})$ over time, which was generally reached for all batches of treatment A. It can be observed that there is least spread in cultivation with treatment C even when only comparing among those batches that actually contained treatment C (batches 1-4 and 6), this possibly indicates a more robust growth process. Furthermore, the $\ln(OD_{600})$ curves indicate that exponential growth, seen as a straight line when looking at ln(OD) graphs, ceased after 2-3 hours (treatment A) or 4-5 hours (treatments B and C) and became more of a linear growth, seen as a gradual flatting out of the curve.

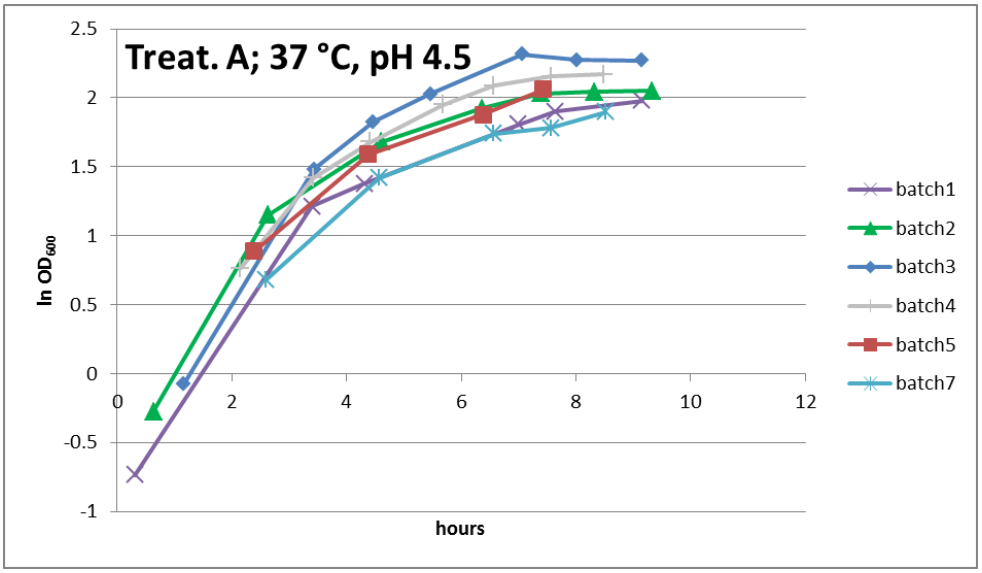

Figure 3. The growth curves during fermentation using treatment A (37 °C, pH 4.5), represented as the natural logarithm of the optical density (OD) at 600 nm.

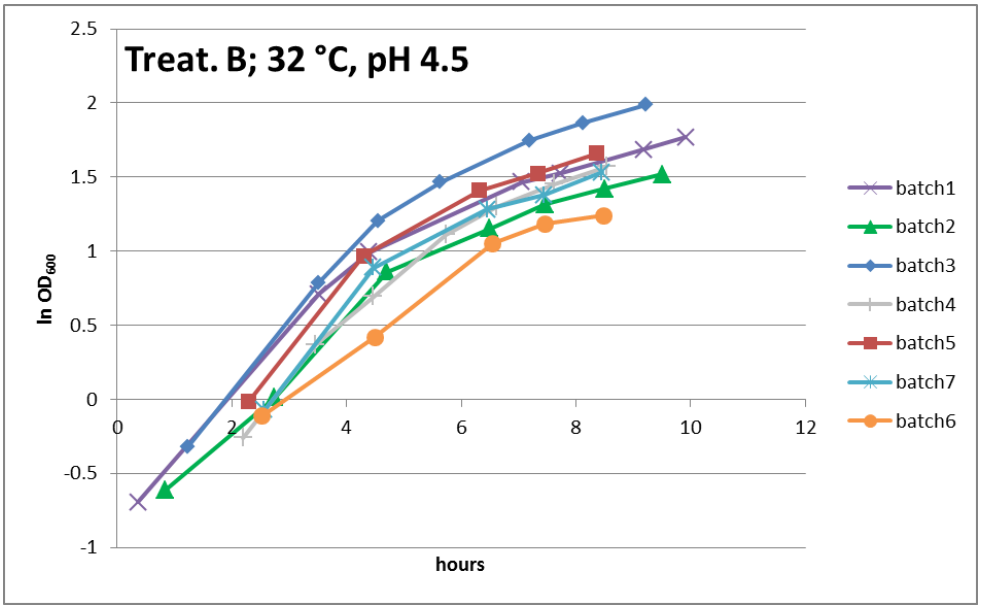

Figure 4. The growth curves during fermentation using treatment B (32 °C, pH 4.5), represented as the natural logarithm of the optical density (OD) at 600 nm.

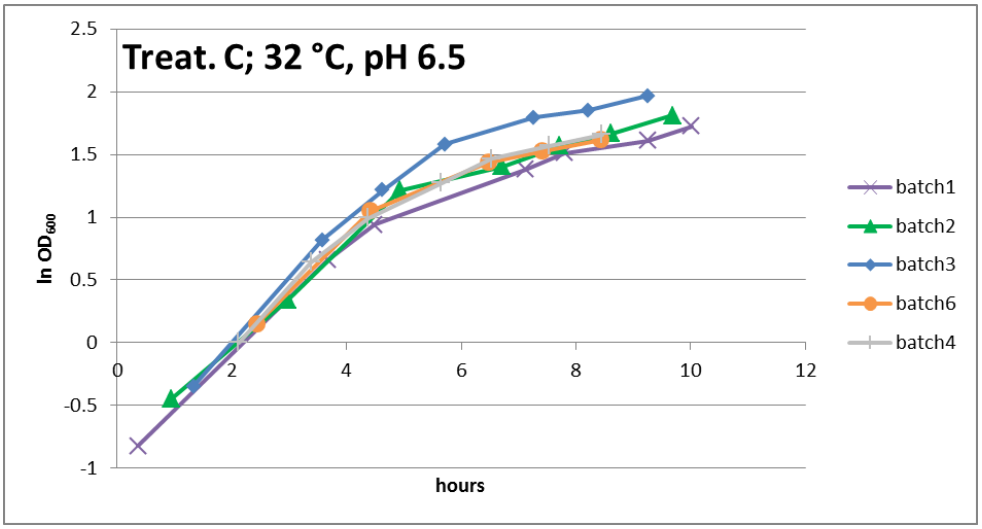

Figure 5. The growth curves during fermentation using treatment C (32 °C, pH 6.5), represented as the natural logarithm of the optical density (OD) at 600 nm.

4.2 Freeze-drying

The ferments harvested at middle stationary phase were subsequently formulated and freeze-dried and the seven batches that were run were of varying freeze-drying quality. Some batches were considered “collapsed”, by visual inspection right out of the freeze-drier; these samples were partly yellow-brownish in color and had hollow spaces with a melted look. This means that the freeze-drying was incomplete, the bacterial glass did not form to adequate extension and so more water was trapped left in the bacterial cakes. The evidently collapsed batches were batch 1, 5 and a part of batch 7. The freeze-drier LyoStar II™ seemed to have done a good job on batch 2 which was one of the, if not the most stable batch. This machine also produced the least stable collapsed batch; batch 1. However, this was no surprise since the automatic program failed during drying of batch 1 and so the freeze-drying ceased for a long while before it was restarted.

Although there was quite a lot of variation in quality between freeze-dried samples in many batches (this was mitigated by duplication and selection of the most average samples), batch 7 was exceptionally variable because it was parted in very well-dried cakes and collapsed cakes. Therefore, it was treated as two batches: “batch 7 collapsed” and “batch 7 dry”. However, only batch 7 dry was used in statistical analysis and for finding a correlation between stability in accelerated storage and freeze-thaw survival. Batch 7 collapsed was only plotted on the CFU/mL graphs (Figures 6 and 7) to see if it behaved like the other collapsed batches, which it did.

4.3 Survival in accelerated storage assay

The CFU/mL values observed throughout the accelerated storage assay are shown in Figures 6-8. Additionally, the individual CFU/ml values of the bacterial formulate before freeze-drying (which are represented graphically in Figures 6-8 at timepoint "-5") have been averaged for each respective treatment (A-C) in Table 3. The collapsed batches, labeled in Figures 6-8, showed much lower storage-stability than the non-collapsed batches. There was a large between-batch variation of storage-stability, however, treatment C clearly showed higher storage-stability than the other treatments. According to Table 3, the cell density before freeze-drying depended on treatment in the falling order: C > B > A, however, Table 3 also shows that CIs of the cell density values are too broad for making statistically confident statements about this relation, except for the relation between treatments A and C. The relation of cell density before freeze-drying of treatment B toward the other treatments remains somewhat uncertain, although it was consistent with the trend C > B > A in five out of seven possible comparisons.

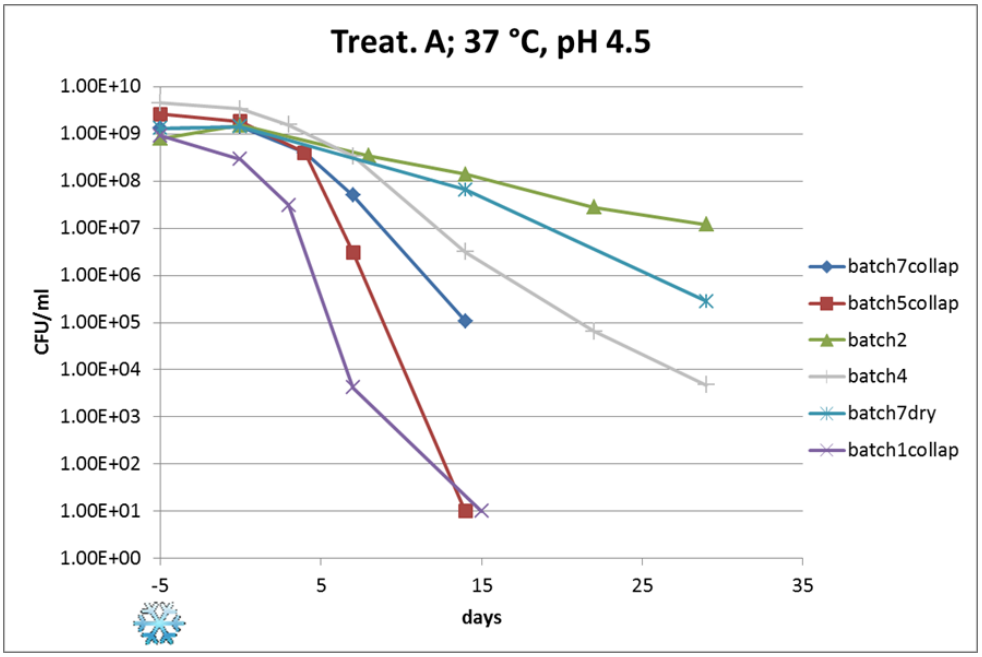

Figure 6. Storage-stability of *L. reuteri* cultivated at 37 °C, pH 4.5 represented as CFU/mL over time during accelerated storage assay at 37 °C for up to 29 days. The snowflake at "-5" days indicates the start of the freeze-drying, at day 0 accelerated storage was started. The values at day -5 are the before freeze-dry cell counts and those at day 0 are the initial freeze-drying survival cell counts.

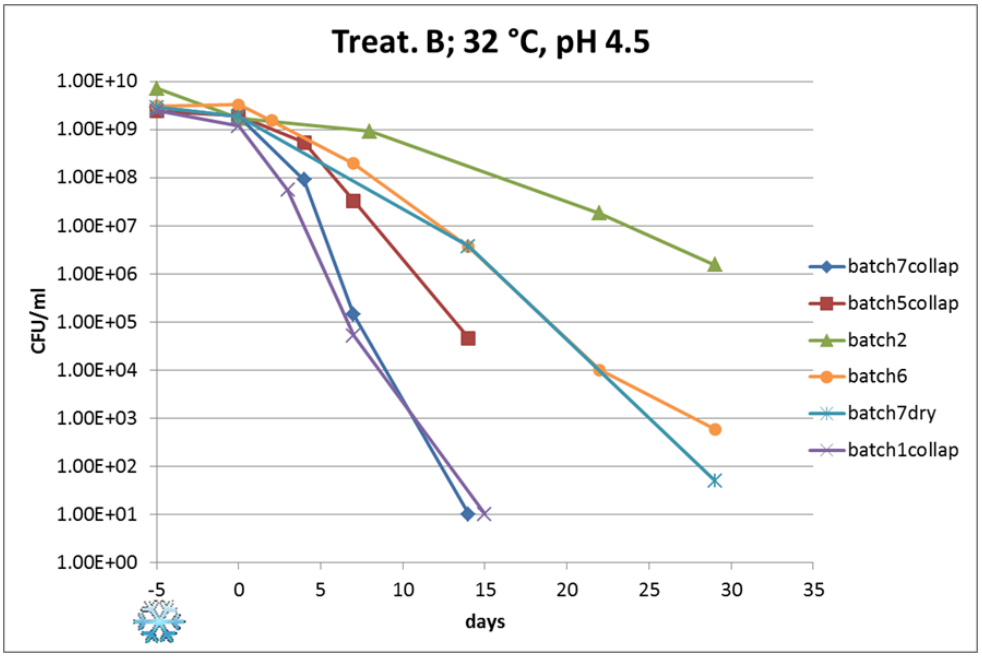

Figure 7. Storage-stability of *L. reuteri* cultivated at 32 °C, pH 4.5 represented as CFU/mL over time during accelerated storage assay at 37 °C for up to 29 days. The snowflake at "-5" days indicates the start of the freeze-drying, at day 0 accelerated storage was started. The values at day -5 are the before freeze-dry cell counts and those at day 0 are the initial freeze-drying survival cell counts.

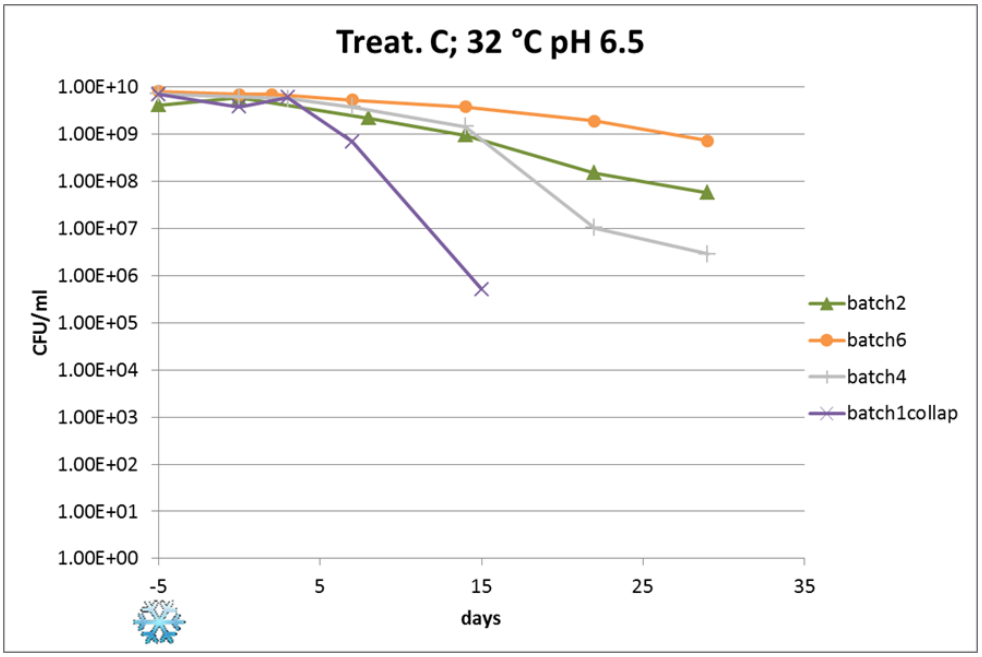

Figure 8. Storage-stability of *L. reuteri* cultivated at 32 °C, pH 6.5 represented as CFU/mL over time during accelerated storage assay at 37 °C for up to 29 days. The snowflake at "-5" days indicates the start of the freeze-drying, at day 0 accelerated storage was started. The values at day -5 are the before freeze-dry cell counts and those at day 0 are the initial freeze-drying survival cell counts.

Table 3. *The average cell densities of the bacterial formulate before freeze-drying for each of the treatments with the respective confidence interval and sample sizes*

treatment	average cell density of formulate (CFU/mL)	95% confidence interval	sample size
A; (37 °C, pH 4.5)	2.03×10^9	$\pm 1.74\times10^9$	5
B; (32 °C, pH 4.5)	3.64×10^9	$\pm 1.74\times10^9$	5
C; (32 °C, pH 6.5)	6.45×10^9	$\pm 1.95\times10^9$	4

Rough predictions about the storage-stability could often be made from the look of the freeze-dried bacterial cakes. Especially the collapsed batches, exhibited dramatically lower storage-stability. When the samples were fresh out of the freeze dryer, in the case of well-dried samples, the cake was white and homogenous and without any hollow cracks, in contrary to collapsed samples (unfortunately no picture was taken of freshly dried collapsed samples). The best well-dried samples kept their original look for a long time in accelerated storage as exemplified on the right side of Figure 9 E. In contrary, collapsed samples quickly shrinked and became discolored, or changed into a foam during accelerated storage as seen in Figure 9. Medium-well-dried samples (of non-collapsed batches, picture not shown) would shrink to a moderate extent but kept their dry look and the white color for many weeks and had a relatively good storage-stability.

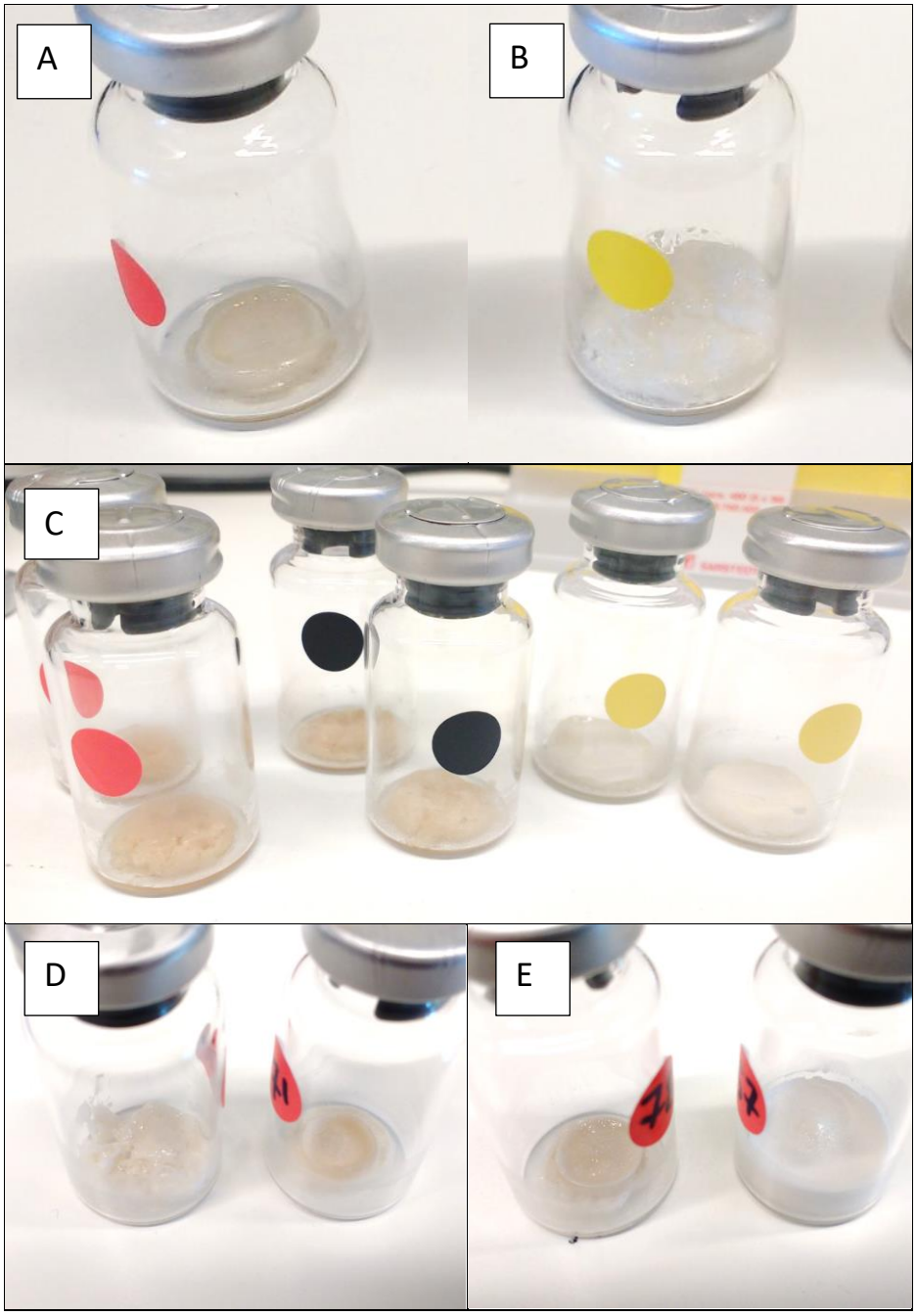

Figure 9. Visual inspection of the different looks of collapsing of freeze-dried cakes of *Lactobacillus reuteri* with a well-dried sample represented in picture E (right side). (A) “batch 1 collapsed” after 3 days of accelerated storage; (B) “batch 1 collapsed” after 3 days of accelerated storage; (C) “batch 1 collapsed” after 2 weeks of accelerated storage; (D) “batch 7 collapsed” after 4 days of accelerated storage; (E) “batch 7 collapsed” (left) and “non-collapsed” (right) after 4 days of accelerated storage.

The collapsed samples were observed to develop in a few different ways during accelerated storage. Two different types of collapses in batch 1 can be seen in Figure 9 A and B. In Figure 9 A, the cake has shrinked a lot and become brownish in color and moist. It corresponds to treatment B (32 °C pH 4.5). In Figure 9 B, the sample is white, voluminous and looks like a foam. It corresponds to treatment C (32 °C pH 6.5). Both of these types of collapses led to very low stability – the survival rates were close to zero after 1-2 weeks of accelerated storage. Figure 9 C shows samples corresponding to all three treatments (A-C) from the collapsed batch 1 after 14 days of accelerated storage; interestingly it shows that biomass from treatment C (32 °C pH 6.5) collapsed in the white, foamy way while samples from the other two treatments have gained a "crumbled", dry and hollowed-out look. The latter type corresponds to a later stage of deterioration of the shrinked, darkened and moist samples from Figure 9 A.

Collapsed samples from other batches showed similar looks during storage to what was observed in batch 1, as exemplified in Figure 9 D where collapsed samples of batch 7 belonging to treatment A exhibit a foamy (left) and a shrinked and discolored but dry appearance (right). Another collapsed sample of batch 7 belonging to treatment A that exhibits a shrinked, discolored and moist appearance is seen in the left part of Figure 9 E. The shrinked and discolored but dry sample type had much better storage-stability and never went into the "crumbled" stage as did the shrinked, darkened and moist type. It also had much better storage-stability than the foamy type.

It was observed in many batches that different treatments could lead to different developments of collapsed samples and even different tendencies towards collapsing at all during freeze-drying. This is summarized in Table 4 below. Treatment A seemed least robust – it showed the greatest variability in appearance when collapsed, this tendency was best represented in batch 5 and 7 (Table 4). As it already has been demonstrated, treatment C led to the highest storage-stability. Furthermore, as can be deduced from Table 4, the superiority of treatment C was preserved even in collapsed conditions. The relation in storage-stability between treatments A and B was more subtle. To clarify the relation in storage-stability between treatments A and B, Figure 10 shows within-batch comparisons of survival fractions (CFU/mL at a given time divided by initial CFU/mL) after treatment A (represented by rhombus marks) and treatment B (represented by square marks). Each batch is represented by a unique color. Treatment A was normally more stable than treatment B (see non-collapsed batches, Figure 10) but in collapsed batches it became equivalent to treatment B (batch 1, Figure 10) or worse than treatment B (batch 5, Figure 10). It is most likely that treatment A would have been worse or equivalent to treatment B also in "batch 7 collapsed" (see Figure 10), if the foamy samples had not been excluded from measurements. These were

excluded because of experience from batch 5, where such samples showed extremely low storage-stability, while it was meant to use the most “average” samples from each batch.

Table 4. *Summary of the connections between freeze-drying success, appearance of samples during accelerated storage and the storage-stability*

Batch	Treatment A	Treatment B	Treatment C	Stability
Batch 1 (collapsed)	Shrinked and darkened, hard to dissolve	Shrinked and darkened, hard to dissolve	Foamy	C > A/B
Batch 2	Looks good	Looks good (slightly nicer – but stability worst)	Looks good	C > A > B
Batch 4	Looks good, shrinked more, harder to dissolve	-	Looks good, shrinked less	C > A
Batch 5 (collapsed)	Looks much worse, parted 50/50 into foamy and shrinked, darkened	Looks better, never foamy, shrinked less	-	B > A (half of measured A-samples were foamy!)
Batch 6	-	Looks very nice	Looks very nice	C > B
Batch 7 (collapsed)	All types of collapsing represented, some are foamy	Never foamy	-	A > B (foamy A-samples excluded from viable counting!)
Batch 7 (dry)	Looks very nice	Looks very nice (fewer uncollapsed samples than A)	-	A > B

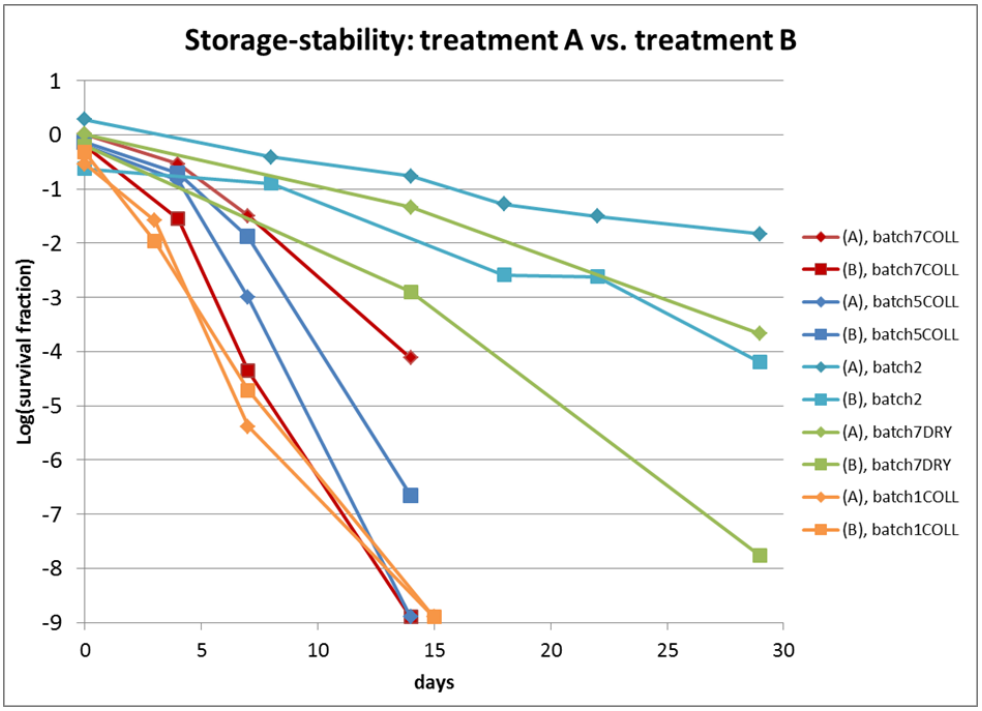

Figure 10. Survival of *Lactobacillus reuteri* subjected to accelerated storage assay at 37 °C for up to 29 days; within-batch pairwise comparisons between cultivation preconditioning: treatment A (37 °C, pH 4.5) versus treatment B (32 °C, pH 4.5). Batch 7 contained both well-dried and collapsed assemblies of samples.

So far, the storage-stability of samples from the different fermentation treatments (A-C) has been presented as raw data and related to quality of the freeze-drying process and the appearance of freeze-dried samples. The storage-stability data was further processed statistically in the form of survival fractions through ANOVA followed by post-hoc Tukey HSD tests. The results, obtained from both non- and $\log_{10}$-transformed data, are represented graphically in Figure 11 as the batch-adjusted means with 95% CIs. The exact sizes of differences between means along with p-values of the statistical tests are presented in Table 5, while the exact numbers behind what was plotted in Figure 11 can be found in Tables 6 and 7, along with other useful descriptive statistics. p-values from Tukey-HSD which were just slightly above the 0.05 threshold were also judged as significant since it was tested that cdfTukey (see Materials and methods section 3.5) always gives slightly higher p-values than SPSS would.

A clear and rather consistent trend of falling survival rates is seen in Figure 11 and the relation between treatments is quite clear. Biomass of fermentation treatment C (32 °C pH 6.5) was easily the most stable, in all cases. The relation between treatments A (37 °C pH 4.5) and B (32 °C pH 4.5) was, as mentioned, more subtle. The plots have been parted in a “reliable” and “unreliable” region. This was

done to symbolize some facts: firstly, it indicates where the homogeneity of variance assumption was kept best. Secondly, it is where the coefficient of variation (CV) was lowest (see Tables 6 and 7 below).

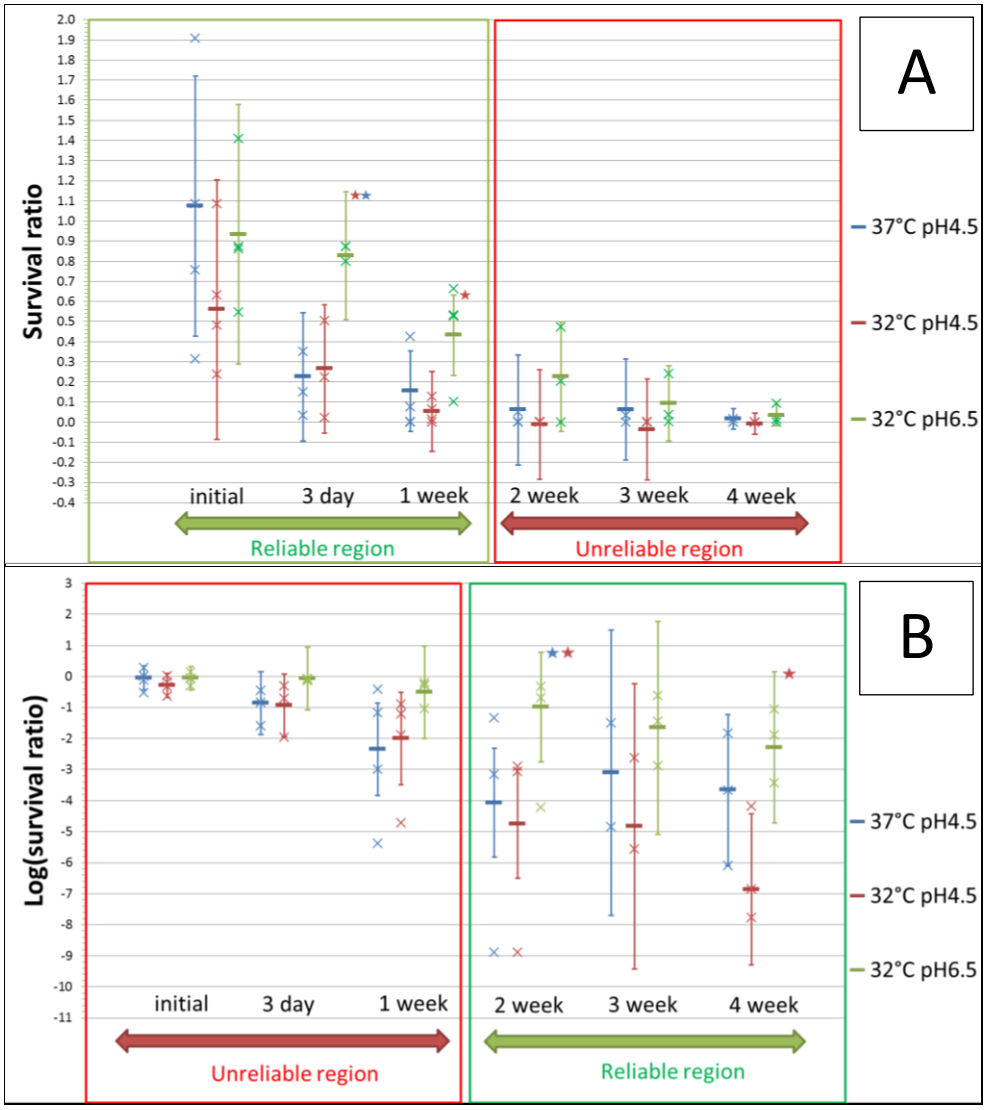

Figure 11. Storage-stability analysis of *Lactobacillus reuteri*, fermented at three different cultivation conditions (rightmost labels). Analyzed by ANOVA and Tukey-HSD, using non-transformed (graph A) and $\log_{10}$-transformed (graph B) data of survival in accelerated storage assay during four weeks, respectively. Rectangle markers indicate treatment means (batch-adjusted). The cross markers indicate non-adjusted values of individual biological replicates i.e. of individual batches (see Tables 6 and 7). Bars indicate 95% CI from pooled standard deviation of the adjusted means. Colored stars indicate statistically significant differences ($p \leq 0.05$) from means of that color. Reliable/unreliable region indicates the range of compliance with statistical assumptions and reasonable batch-adjusted mean-values. Timepoints at the x-axis are approximate (±1 day).

Probably as the consequence of lower CV, the statistically significant differences were found in the “Reliable regions”. The CV is a relative, unitless measure of how much variance there is in what is measured - relative to the mean (Fay, 2013), CV = standard deviation/mean. By looking at Tables 6 and 7, it can be seen that the CV starts low but increases with time for non-transformed data and the opposite is true for $\log_{10}$-transformed data. Thirdly, at two weeks and onward is where slightly negative adjusted means for non-transformed data appear in the analysis. This is unrealistic, yet affects the significance.

Table 5. *Statistical evaluation of survival in accelerated storage assay for up to four weeks of Lactobacillus reuteri fermented at three different treatments. Only data belonging to "reliable regions" is shown (see Figure 11). Differences between block-adjusted means of "raw" survival fractions and log10-transformed survival fractions are shown with their corresponding significance p-values from the ANOVA and Tukey HSD test. Highlighted in green are significant findings ($p \leq 0.05$) and these p-values are marked with a "*". Tukey HSD p-values slightly above the 0.05 threshold are also judged significant and marked with a "~*" as well as highlighted in green. Timepoints are approximate (±1 day)*

Timepoint and ANOVA sig.	Treatment (I)	Treatment(J)	Difference between adjusted means (I - J)	Tukey HSD sig.
Reliable region using non-transformed data				
"ini"raw	32°C pH 6.5	32°C pH 4.5	0.3735	p= n/a
ANOVA:	32°C pH 6.5	37°C pH 4.5	-0.1412	p= n/a
p=0.405	37°C pH 4.5	32°C pH 4.5	0.5147	p= n/a
"3day"raw	32°C pH 6.5	32°C pH 4.5	0.5628	p= 0.051 ~*
ANOVA:	32°C pH 6.5	37°C pH 4.5	0.6035	p= 0.043 *
p= 0.042 *	37°C pH 4.5	32°C pH 4.5	-0.0407	p= 0.952
"1week"raw	32°C pH 6.5	32°C pH 4.5	0.3782	p= 0.056 ~*
ANOVA:	32°C pH 6.5	37°C pH 4.5	0.2765	p= 0.151
p= 0.045 *	37°C pH 4.5	32°C pH 4.5	0.1017	p= 0.772
Reliable region using $\log_{10}$-transformed data				
"2week"$\log_{10}$	32°C pH 6.5	32°C pH 4.5	3.7652	p= 0.032 *
ANOVA:	32°C pH 6.5	37°C pH 4.5	3.0854	p= 0.052 ~*
p= 0.035 *	37°C pH 4.5	32°C pH 4.5	0.6798	p= 0.663
"3week" $\log_{10}$	32°C pH 6.5	32°C pH 4.5	3.1772	p= n/a
ANOVA:	32°C pH 6.5	37°C pH 4.5	1.4471	p= n/a
p= 0.254	37°C pH 4.5	32°C pH 4.5	1.7301	p= n/a
"4week" $\log_{10}$	32°C pH 6.5	32°C pH 4.5	4.5819	p= 0.044 *
ANOVA:	32°C pH 6.5	37°C pH 4.5	1.3724	p= 0.464
p= 0.054	37°C pH 4.5	32°C pH 4.5	3.2095	p= 0.102

Table 6. *Quantitative statistics from survival of Lactobacillus reuteri during accelerated storage assay for up to four weeks, after fermentation at three different treatments. Means represent survival fraction, adjusted means have been batch-adjusted through a block design in SPSS. 95% CI are the confidence intervals. Abs(Average CV) is the absolute of the coefficient of variation, averaged for all treatments. Statistically significant and non-significant results of the omnibus ANOVA test (p=0.05) are indicated by "sig" and "n/s", respectively. Timepoints are approximate (±1 day)*

Time-point	Treatment	Biol. repl.	Adjusted means (EMMs)	Unadjusted means	95% CI (adjusted means)		Pooled std. dev. (adjusted means)	Average CV
					lower	upper		
Ini	37°C pH4.5	4	1.074	1.016	0.429	1.720		
(n/s)	32°C pH4.5	4	0.559	0.608	-0.086	1.205	0.502	0.587
	32°C pH6.5	4	0.933	0.923	0.288	1.579		
3day	37°C pH4.5	3	0.225	0.178	-0.094	0.543		
(sig)	32°C pH4.5	3	0.265	0.250	-0.053	0.584	0.173	0.394
	32°C pH6.5	3	0.828	0.848	0.510	1.146		
1week	37°C pH4.5	4	0.155	0.125	-0.045	0.355		
(sig)	32°C pH4.5	4	0.053	0.051	-0.146	0.253	0.156	0.732
	32°C pH6.5	4	0.432	0.456	0.232	0.632		
2week	37°C pH4.5	3	0.060	0.0170	-0.212	0.333		
(n/s)	32°C pH4.5	3	-0.013	0.0008	-0.285	0.260	0.149	1.637
	32°C pH6.5	3	0.226	0.2251	-0.047	0.498		
3week	37°C pH4.5	2	0.063	0.017	-0.188	0.315	0.082	
(n/s)	32°C pH4.5	2	-0.037	0.001	-0.289	0.214	0.082	2.000
	32°C pH6.5	3	0.093	0.093	-0.095	0.280	0.076	
4week	37°C pH4.5	3	0.016	0.00510	-0.036	0.068		
(n/s)	32°C pH4.5	3	-0.008	0.00007	-0.060	0.044	0.028	2.000
	32°C pH6.5	3	0.034	0.03532	-0.018	0.086		

Table 7. *Quantitative statistics from survival of Lactobacillus reuteri during accelerated storage assay for up to four weeks, after fermentation at three different treatments. Means represent log_{10}(survival fraction), and were batch-adjusted through a block design in SPSS. 95% CI are the confidence intervals. Abs(Average CV) is the absolute of the coefficient of variation, averaged for all treatments. Statistically significant and non-significant results of the omnibus ANOVA test (p=0.05) are indicated by "sig" and "n/s", respectively. Timepoints are approximate (±1 day)*

Time-point	Treatment	Biol. repl.	Adjusted means (EMMs)	95% CI (adjusted means)		Pooled std. dev. (adjusted means)	Abs(Average CV)
				lower	upper		
Initial	37°C pH4.5	4	-0.054	-0.439	0.331		
(n/s)	32°C pH4.5	4	-0.294	-0.679	0.091	0.300	2.206
	32°C pH6.5	4	-0.059	-0.444	0.326		
3day	37°C pH4.5	3	-0.859	-1.873	0.155		
(n/s)	32°C pH4.5	3	-0.936	-1.950	0.078	0.553	0.892
	32°C pH6.5	3	-0.064	-1.078	0.950		
1week	37°C pH4.5	4	-2.349	-3.835	-0.863		
(n/s)	32°C pH4.5	4	-1.999	-3.485	-0.513	1.156	0.713
	32°C pH6.5	4	-0.514	-2.000	0.972		
2week	37°C pH4.5	3	-4.072	-5.829	-2.314		
(sig)	32°C pH4.5	3	-4.752	-6.509	-2.994	0.956	0.292
	32°C pH6.5	3	-0.986	-2.744	0.771		
3week	37°C pH4.5	2	-3.099	-7.696	1.498	1.510	
(n/s)	32°C pH4.5	2	-4.829	-9.426	-0.232	1.510	0.459
	32°C pH6.5	3	-1.652	-5.078	1.775	1.379	
4week	37°C pH4.5	3	-3.658	-6.092	-1.224		
(border	32°C pH4.5	3	-6.867	-9.301	-4.434	1.325	0.310
case sig.)	32°C pH6.5	3	-2.285	-4.719	0.148		

4.4 Freeze-thaw assay survival

Survival fractions of the FT assay, according to fermentation treatment used, are shown below in Figure 12. The best survivor was after fermentation at treatment B (32 °C, pH 4.5) followed by treatment A (37 °C, pH 4.5). The drastically worst survivor was after treatment C (32 °C pH 6.5).

Statistical analysis of the FT-assay survival data with ANOVA followed by Tukey-HSD revealed that the mean survival of FT assay after treatment C differed significantly (at p-value below 0.05) from the mean survival of the other two treatments, when using $\log_{10}$-transformed data but not non-transformed data, see Table 8. Additional quantitative statistics data is provided in Table 9, where it can be seen that the CV of $\log_{10}$-transformed data is much lower than that of non-transformed data, and this explains the greater power of the statistical tests for the $\log_{10}$-transformed dataset.

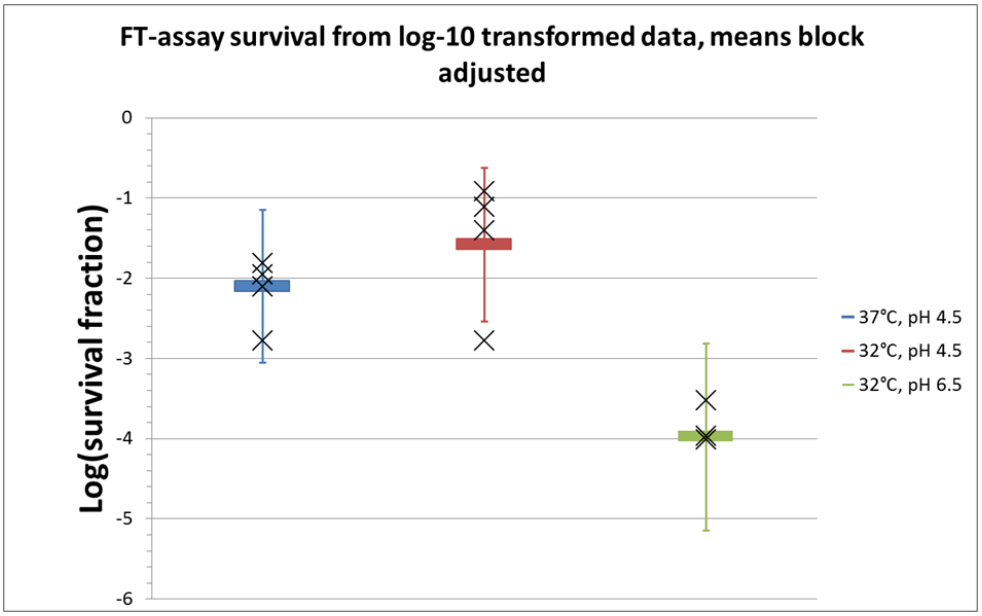

Figure 12. Freeze-thaw assay survival of *Lactobacillus reuteri* fermented at three different treatments: (A) 37 °C pH 4.5; (B) 32 °C pH 4.5 and (C) 32 °C pH 6.5. Results based on $\log_{10}$-transformed data. The rectangular markers represent the treatment means (batch adjusted), the bars show 95% CIs and the black cross markers represent the individual biological replicates (batches). The colored stars next to the CI bar indicate a statistical difference between the mean and the means of the stars colors.

Table 8. *Statistical evaluation of freeze-thaw assay survival of Lactobacillus reuteri fermented at three different treatments. Differences between block-adjusted means of survival fractions are shown together with their corresponding significance p-values from the ANOVA and Tukey HSD multiple-comparisons test. Highlighted in green are significant findings ($p \leq 0.05$) and these p-values are marked with a "*". Tukey HSD p-values slightly above the 0.05 threshold are also judged significant and marked with a "~*" as well as highlighted in green*

Type of data (FT assay survival)	ANOVA omnibus test	Treatment (I)	Treatment (J)	Survival mean diff. (I - J)	Tukey HSD significance:
Non-transformed	p=0.100	32°C pH 6.5	32°C pH 4.5	-0.0744	p=0.090
		32°C pH 6.5	37°C pH 4.5	-0.0272	p=0.616
		37°C pH 4.5	32°C pH 4.5	-0.0472	p=0.273
log_{10}-transformed	p=0.030 *	32°C pH 6.5	32°C pH 4.5	2.3984	p=0.026 *
		32°C pH 6.5	37°C pH 4.5	1.8768	p=0.054 ~*
		37°C pH 4.5	32°C pH 4.5	0.5216	p=0.674

Table 9. *Quantitative statistics for FT assay survival of Lactobacillus reuteri fermented at three different treatments. Means represent survival fraction/log_{10}(survival fraction), according to type of data. 95% CI are the confidence intervals. Abs(Average CV) is the absolute of the coefficient of variation, averaged for all treatments. Statistically significant and non-significant results of the omnibus ANOVA test (p=0.05) are indicated by "sig" and "n/s", respectively*

Type of data	Treatment	No. of biol. repl.	Unadjusted means	Adjusted means (EMMs)	95% CI (adjusted means)		Pooled std. dev. (adjusted means)	Average CV
					lower	upper		
Non-transf. (n/s)	37°C pH4.5	4	0.0090	0.013	-0.032	0.058	0.032	1.678
	32°C pH4.5	4	0.0598	0.060	0.015	0.105	0.032	
	32°C pH6.5	3	0.0002	-0.014	-0.069	0.041	0.035	
Log_{10} transf. (sig)	37°C pH4.5	4	-2.162	-2.102	-3.058	-1.146	0.688	0.275
	32°C pH4.5	4	-1.552	-1.580	-2.536	-0.624	0.688	
	32°C pH6.5	3	-3.833	-3.979	-5.146	-2.811	0.727	

4.5 Correlating survival of freeze-thaw and accelerated storage assays

The survival of *L. reuteri* after one and two weeks in the accelerated storage assay was plotted against its survival of the FT-assay in Figures 13 and 14, respectively.

In both cases, a negative trend was seen; higher FT-assay survival meant lower storage-stability. The negative slope was steeper after two weeks because naturally, more cells had died after an additional week of accelerated storage. Batch 5 was included in these plots but specifically labeled, since it seemed to be quite an outlier. It is the only collapsed batch in this part of the analysis because for each of the plots, only those batches were used for which there was data for both storage-stability and FT-assay survival (and FT-survival had not been measured for batch 1 with the appropriate assay procedure). Therefore, the correlated storage-stability/FT-assay survival means belonging to each of the three fermentation treatments (A-C), plotted in Figures 13 and 14, do not include batch 1 (and Figure 13 does not include batch 7 due to the experimental design). The correlated means were not batch-adjusted. See Tables 6 and 9 in the previous sections to compare with both batch-adjusted and unadjusted means when including all available batches.

The mean stability after 1 week was plotted versus the negative 10-logarithm of mean FT-assay survival for each treatment, in order to fit a line and get a predictive equation (Figure 15). Only batches for which both storage-stability and FT-assay survival data was available were included, except for batch 5 which was excluded in this case. The formula behind the fitted line, $y = 0.0901x^{1.336}$, is a power equation. The FT-assay survival was transformed by ($-\log_{10}$) to avoid negative signs in the equation. For additional information, including batch 5 gave the equation: $y = 0.0224x^{2.43}$. This could perhaps be a useful, more robust but less precise prediction if one anticipates some batches to be collapsed. At week 2 and not including batch 5, the equation became $y = 0.0004x^{5.1119}$ (steeper). See Discussion section regarding conclusions about using FT-assay as a predictor of stability. Note that to use these equations, the "x" must be provided as the negative $\log_{10}$ of FT-assay surviving fraction.

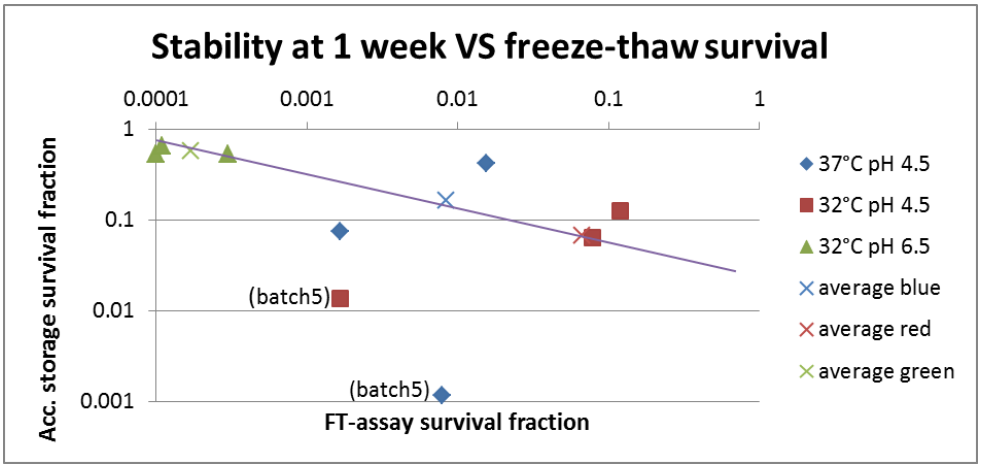

Figure 13. Correlation between survival of *Lactobacillus reuteri* after 1 week in accelerated storage assay and its survival of freeze-thaw assay. Using unadjusted means, data from a combination of batches 2, 4, 5 and 6. Cross-markers indicate the correlated average for each treatment and the other markers indicate measurements from individual batches.

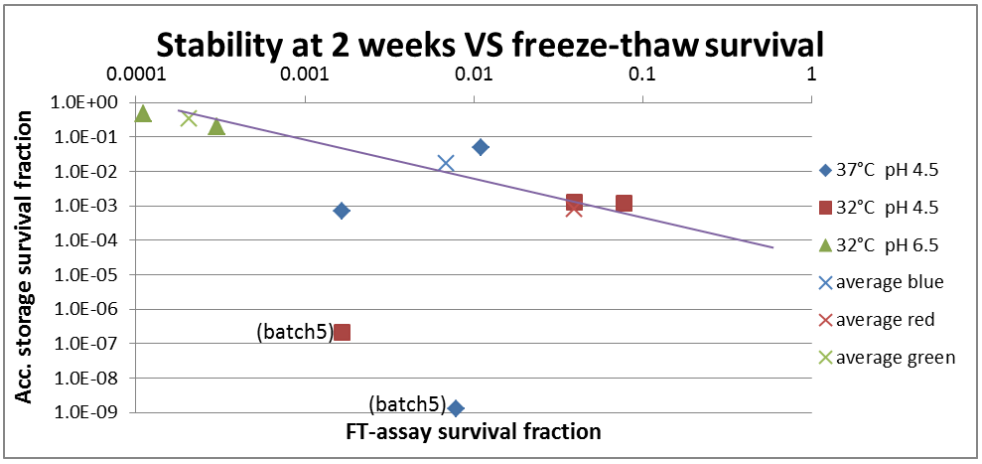

Figure 14. Correlation between survival of *Lactobacillus reuteri* after 1 week in accelerated storage assay and its survival of freeze-thaw assay. Using unadjusted means, data from a combination of batches 4, 5, 6 and 7(dry). Cross-markers indicate the correlated average for each treatment and the other markers indicate measurements from individual batches.

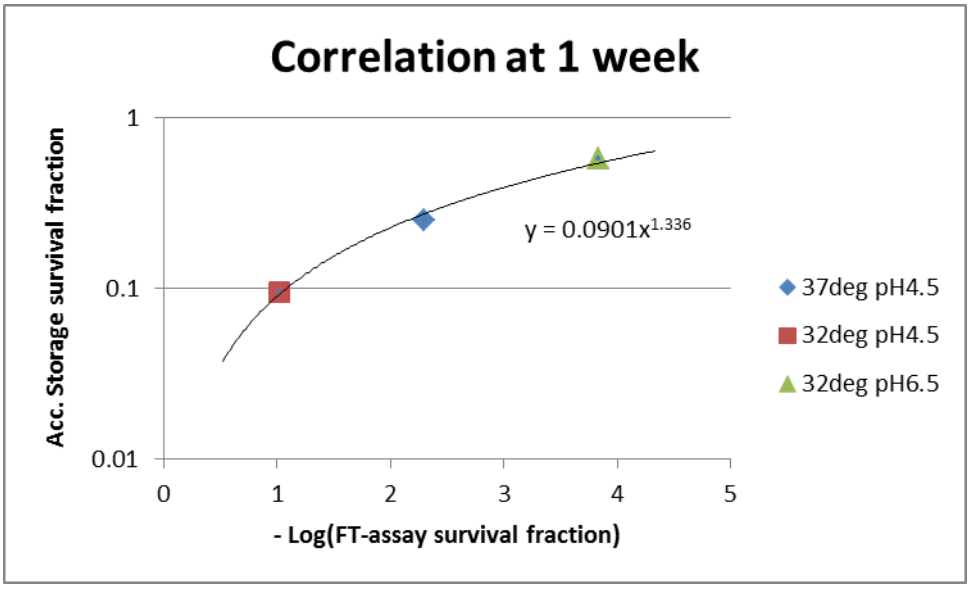

Figure 15. Correlation between survival of *Lactobacillus reuteri* after 1 week of accelerated storage and the negative logarithm (base 10) of its freeze-thaw assay survival. Unadjusted values, combination of batches 2, 4 and 6 (batch 5 not included).

4.6 Statistics

The ANOVAs and Tukey-HSD tests that were used to statistically evaluate the difference between survival means in the FT and accelerated storage assays were based on a statistical model with some assumptions (see Background section 2.6.1). The following subsections describe how well these statistical assumptions were fulfilled.

4.6.1 Assumptions of statistical tests – normality

An example of the normal Q-Q plots of standardized residuals which were plotted for each ANOVA and compared between non-transformed and $\log_{10}$-transformed data is shown in Figure 16. The rest of the plots can be found in Figures 23 and 24 in Appendix. The normal Q-Q plots allow the reader to judge by eye and experience how well the residuals fit the normality assumption of ANOVA. The closer to the line the dots are, the better. Also, it is better if the dots are randomly scattered around the line instead of forming curves. The graphical information is represented in Table 10 in the quantitative form of skewness, kurtosis and the results of Shapiro-Wilk test; which is easier to judge objectively.

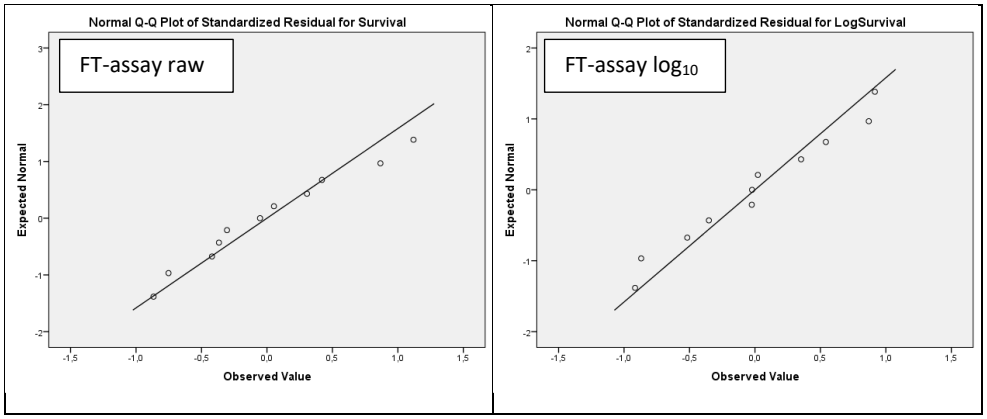

Figure 16. Normal Q-Q plots of expected residuals for normal distribution versus observed residuals, for evaluation of the normality assumption of ANOVA in the freeze-thaw (FT) assay. Plots based on non-transformed (raw) and log_{10}-transformed data are compared side by side. Each plot contains the combined standardized residuals of all the biological replicates for all treatments and batches.

Table 10. *Statistical significances (sig.) of Shapiro-Wilk test and the skewness and kurtosis compared to their respective standard errors. The underlying data is for each ANOVA (non-transformed ="raw" or log_{10}-transformed) the combined standardized residuals for all treatments and batches*

	Accelerated storage assay Timepoints (±1day)						Freeze-thaw assay
Type of evaluation	Ini raw	3day raw	1week raw	2week raw	3week raw	4week raw	FT-assay raw
Shapiro-Wilk, sig.	0.970	0.505	0.197	0.468	0.394	0.802	0.776
Skewness / Std error	-0.501 / 0.637	0.603 / 0.717	-0.005 / 0.637	-0.339 / 0.717	0.267 / 0.794	-0.075 / 0.717	0.448 / 0.661
Kurtosis / Std error	0.210 / 1.232	-0.789 / 1.400	-1.669 / 1.232	-0.419 / 1.400	-1.427 / 1.587	-1.082 / 1.400	-0.598 / 1.279
Type of evaluation	Ini Log_{10}	3day Log_{10}	1week Log_{10}	2week Log_{10}	3week Log_{10}	4week Log_{10}	FT-assay Log_{10}
Shapiro-Wilk, sig.	0.700	0.839	0.680	0.657	0.393	0.606	0.600
Skewness / Std error	-0.647 / 0.637	0.405 / 0.717	0.738 / 0.637	0.141 / 0.717	-0.361 / 0.794	-0.281 / 0.717	0.010 / 0.661
Kurtosis / Std error	*0.293 / 1.232*	*-0.816 / 1.400*	*0.244 / 1.232*	*-1.447 / 1.400*	*-1.382 / 1.587*	*-1.150 / 1.400*	*-1.029 / 1.279*

From the Shapiro-Wilk test of normality (Table 10), there were no cases close to significantly violating normality at 95% probability ($p \leq 0.05$) which can be seen

also from the fairly good looking Q-Q plots in which only pattern-formation sometimes looked a little suspect but not the general trend. There is no clear difference in how well the residuals of $\log_{10}$-transformed or non-transformed data follow normal distribution which is also seen in the values of skewness and kurtosis (Table 10). Sometimes, transformed data is better and sometimes non-transformed data is better. A rule of thumb for acceptable skewness and kurtosis for the normal distribution (*IBM Knowledge Center - Display Statistics*) says that the absolute value of skewness should be less than twice its standard error (compared in Table 10) and the same goes for the absolute value of kurtosis. There were no cases close to these borders. Therefore, adequate normality of error distribution was assumed.

4.6.2 Assumptions of statistical tests - Homogeneity of variance

The standardized residuals of each ANOVA of the accelerated storage assay and FT-assay were plotted versus treatments (Figure 17 and Figures 25 and 26 in Appendix) to assess the second assumption of ANOVA, homogeneity of variance. It is the relative homogeneity of residual variance (similarity of spread) between treatments and not absolute values that is judged, so the scale on the y-axis in these figures is not important. The numerical values in textboxes in the plots show the ratio between the widest and narrowest variances. By looking at Figures 25-26 in Appendix it can be seen that for the "ini" timepoint (initial survival), there is no notable difference in this ratio between $\log_{10}$-transformed and non-transformed data. After that, the non-transformed data gives more homogenous variances until 2 weeks. Starting at two weeks and forward, the $\log_{10}$-transformed data gives most homogenous variances. For the freeze-thaw assay, in which survival rates were similar to the 2 weeks timepoint in accelerated storage assay, the $\log_{10}$-transformed data gives most homogenous variance of residuals (Figure 17). This could mean that an ANOVA will be more reliable when performed on non-transformed data until 1 week of accelerated storage is reached and after that (and for the FT-assay), $\log_{10}$-transformed data will yield the most reliable ANOVA. Following this conclusion and using transformed and non-transformed data for the ANOVAs accordingly, the homogeneity of variance assumption seems largely fulfilled, according to the rule of thumb (see Background section 2.6.1) with the exception of the timepoint at 3 weeks. This case, however, had very small sample size (and the sample size was unequal between fermentation treatments) due to missing data.

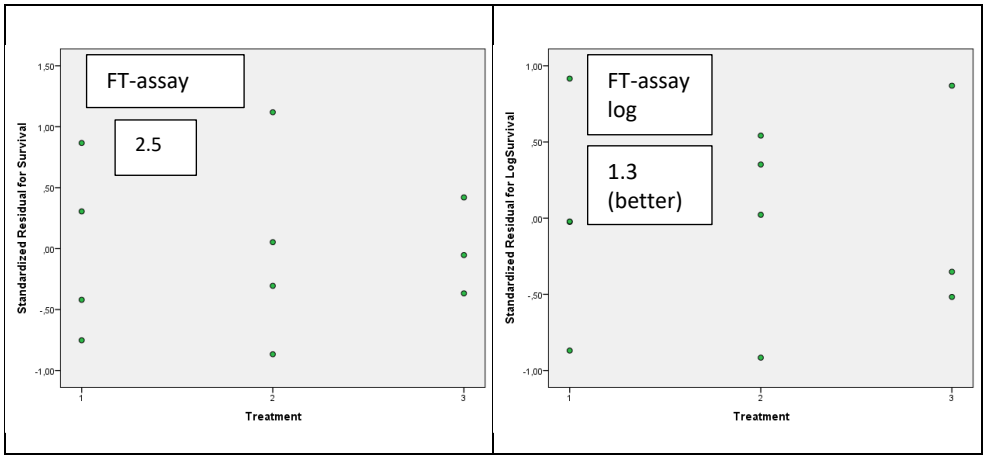

Figure 17. Standardized residuals versus treatments, where treatment no. 1 = 37 °C, pH 4.5; 2 = 32 °C, pH 4.5; 3 = 32 °C, pH 6.5. Freeze-thaw assay, comparison of non-transformed (raw) and $\log_{10}$-transformed data, respectively. Ratio of widest variance over narrowest variance is shown in the textboxes.

4.7 Microscopy

The microscopic studies revealed many differences in morphology of *L. reuteri* resulting from the different fermentation conditions (treatments A-C). Treatment A led to short cells, treatment B to medium-length cells and treatment C tended to lead to long cells. The variance between the shortest and longest cell observed seemed to follow the same increasing trend with few exceptions. Treatment A and B gave cell chains and cell clumps while treatment C gave fewer chains and no clumps at all. These findings are represented by three example photographs from the respective treatments (A-C) in Figure 18, while photos from the rest of the batches can be found in Figure 28 in Appendix. Furthermore, Table 11 summarizes the morphology observations. The cell sizes in Table 11 have been recalculated into µm using the objective micrometer scales seen at the photographs, with the help of Figure 27 in Appendix.

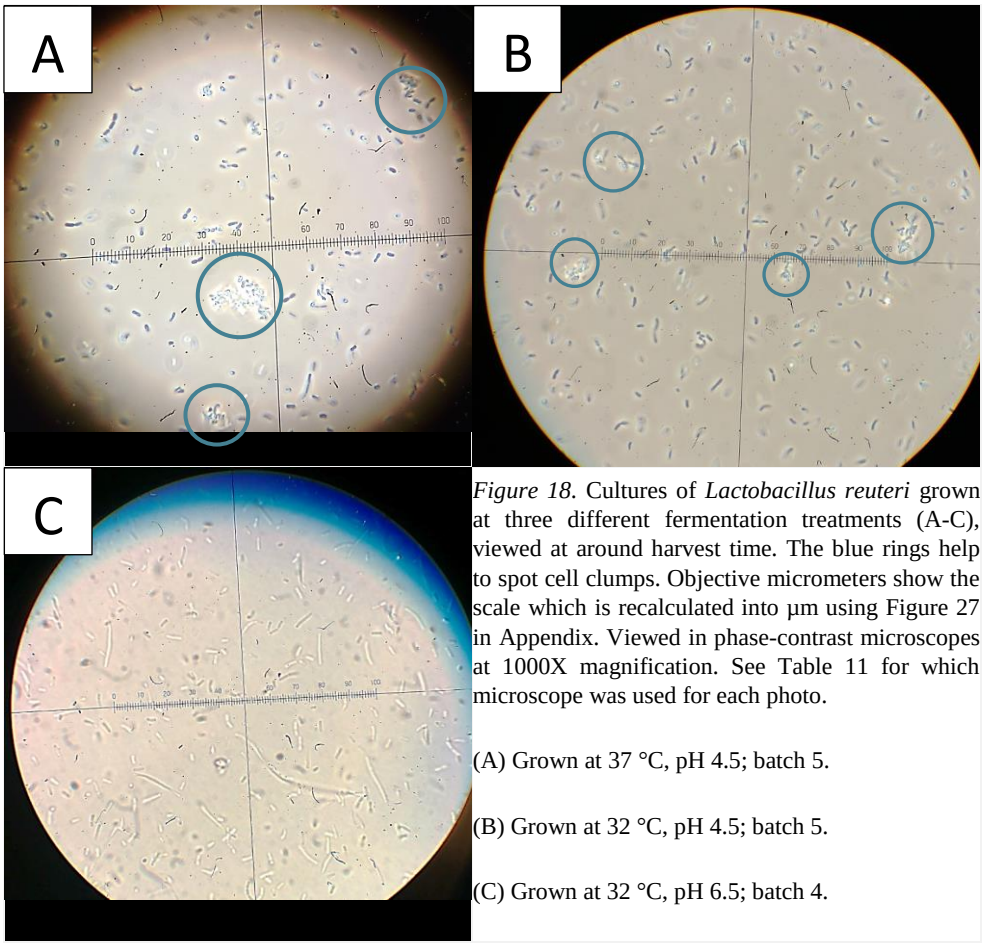

Figure 18. Cultures of *Lactobacillus reuteri* grown at three different fermentation treatments (A-C), viewed at around harvest time. The blue rings help to spot cell clumps. Objective micrometers show the scale which is recalculated into µm using Figure 27 in Appendix. Viewed in phase-contrast microscopes at 1000X magnification. See Table 11 for which microscope was used for each photo.

(A) Grown at 37 °C, pH 4.5; batch 5.

(B) Grown at 32 °C, pH 4.5; batch 5.

(C) Grown at 32 °C, pH 6.5; batch 4.

Table 11. *Summarized microscopic morphology observations of Lactobacillus reuteri grown at three different fermentation treatments (A-C); A = 37 °C, pH 4.5; B = 32 °C, pH 4.5; C = 32 °C, pH 6.5*

Treatment	Batch	Average cell size (µm)	Variation in cell size (µm)	Chains	Clumps	Microscope
A	4	2.1	1.2-2.4	Some dual	Yes	Local lab
A	5	2.1	1.3-2.7	Some dual & longer	Yes (a lot)	Course lab
A	7	2.1[1]	1.9-4.3[1]	Some dual	Yes (a lot)	Course lab
B	5	2.6	1.9-6.7	Some dual	Yes	Course lab
B	6	2.6	1.6-2.8	A lot of different chains[2]	Yes	Course lab
B	7	3.0	1.7-6.2	Most dual, few longer	Yes, small	Course lab

Treatment	Batch	Average cell size (µm)	Variation in cell size (µm)	Chains	Clumps	Microscope
C	4	7.1	2.0-15.8	Few dual	No	Local lab
C	6	5.0	1.8-11.2	Few dual	No	Course lab

1. Scale is lacking at this photo but it was compared with a later photo of formulated biomass from same batch which had a scale and looked similar.

2. Microscopy was very delayed after sampling (3h) and it was observed from later microscopy studies that chains evolve while keeping the samples on hold.

When it comes to microscopy observations of Gram-stained samples, there was no consistent difference in staining between fermentation treatments 37 °C, pH 4.5 and 32 °C, pH 4.5. Staining these resulted in anything between all cells red to 70% blue cells (Table 12). The fermentation treatment 32 °C, pH 6.5 did stand out against the other two dramatically in being stained completely blue; however, staining of this treatment was not replicated (since Gram-staining was implemented at batch 5, leaving only one opportunity to stain cells from this fermentation treatment).

There was on the other hand an association of Gram-stain results with batch number; the batches replicated in Gram staining (although for different treatments) were batch 5, 6 and 7. It seems that batch 5 stained 50/50 red/blue, batch 6 stained strongly blue and batch 7 stained strongly red irrespectively of treatment (Table 12). The actual microscopy photos of all Gram-stained samples are found in Figure 29 in Appendix.

The Gram-stain results were also compared with the relative storage-stabilities within treatments (to eliminate the effect of treatment on storage-stability) and the batch quality (collapsed or non-collapsed) and this is summarized in Table 12 below. The relative within-treatment storage-stabilities were judged by comparing the CFU/mL curves during accelerated storage (Figures 6, 7 and 8). The percentage of blue/red cells after Gram-staining did not correlate with the relative storage-stability, even when excluding collapsed batches. To illustrate this by examining Table 12, consider that "(A) batch 7" consists of all red cells and corresponds to 2.8×10^5 CFU/mL after 4 weeks of accelerated storage while "(B) batch 6" consists of 70% blue cells but corresponds to only 5.95×10^2 CFU/mL after 4 weeks. At the same time, the 100% blue "(C) batch 6" is extremely storage-stable (7.21×10^8 CFU/mL after 4 weeks).

Table 12. *Comparing Gram-stain of Lactobacillus retueri cultivated at different conditions with relative within-treatment storage-stability and with batch quality. Corresponding microscopy photos are found in Figure 29 in Appendix*

Photo in Fig. 29 (Appendix)	Red/blue cell ratio	Relative stability within treatment (Figs. 6-8)	Batch quality
”(A) batch 5”	50/50	very low	Collapsed
“(A) batch 7”	all red	above average	Good (batch 7 dry)
“(B) batch 5”	50/50	below average	Collapsed
“(B) batch 6”	70% blue	above average	Good
“(B) batch 7”	almost all red	medium	Good (batch 7 dry)
“(C) batch 6”	all blue	above average	Good

Trying to gain more understanding about the Gram-stain results, these were compared with the ln(OD) growth-curves (Figures 3-5), fermentation durations, FT-assay survival and specific events during the fermentation and during the microscopy itself (summarized in Tables 13). The Gram-stain did not correlate whatsoever to the height of the ln(OD) curves, the cultivation time nor the FT-assay survival. No trend could be found with how soon after sampling the Gram-stain was done nor with the delay of harvest against the cultivation length plan (Material and methods, section 3.1.1 “Fermentation and formulation”).The most delayed cases were strongly blue, the moderately delayed were strongly red and the least delayed were colored 50/50). The only factor that is possible to link with the result of the Gram-stain is the batch number.

Table 13. *The relations between results of observations of Gram-stained Lactobacillus reuteri from different cultivations and heights of optical density (OD) curves, harvest time, time from anticipated entry into middle stationary phase (MSP) to harvest, delay between sampling to staining and survival of freeze-thaw (FT) assay*

Photo in Fig. 29 (Appendix)	Red/blue cell ratio	growth curve height (Figs. 3-5)	Inoculum to harvest (h)	Delay from MSP to harvest (h:min)	Microscopy: from sampling to Gram-stain	FT assay surviving fraction (and ranking)
”(A) batch 5”	50/50	Medium	23:08	0:42	quick	0.0079 (4)
“(A) batch 7”	all red	Very low	23:39	1:14	medium	0.0111 (3)
“(B) batch 5”	50/50	Above average	26:15	0:49	slow	0.0017 (5)
“(B) batch 6”	70% blue	Very low	26:21	1:34	slow	0.0778 (1)
“(B) batch 7”	almost all red	Medium	26:38	1:11	slow	0.0395 (2)
“(C) batch 6”	all blue	Medium	24:50	2:03	quick	0.0001 (6)

5 Results – GASP experiment

In the first attempt to capture GASP mutants by IMC, it was observed that during the initial growth phase, whole-MRS medium gave rise to a higher heat flow (peaking at about 8.25 mW, Figure 19 A) than half-MRS medium (peaking at about 6.15 mW, Figure 19 C) but the shapes of the curves are very similar. Figures 19 B and D show the development of the heat flows starting from the stationary phase and going into the long-term stationary phase, in a very zoomed-in heat flow scale since the heat activity was small during this time. The heat flow became slightly negative in the stationary phase (more so with whole-MRS, Figure 19 B) implying a small endothermic activity; or this was due to positioning of the baseline. After this dip, the heat flow slowly climbed up during the long-term stationary phase, to values near zero heat flow (whole-MRS) or around 5 µW (half-MRS). No drastic peaks that would evidently imply GASP activity were observed. The erratic fluctuations in the long-term stationary phase heat flows usually follow exactly the same pattern for all ampoules and therefore must be noise, probably due to small external disturbances.

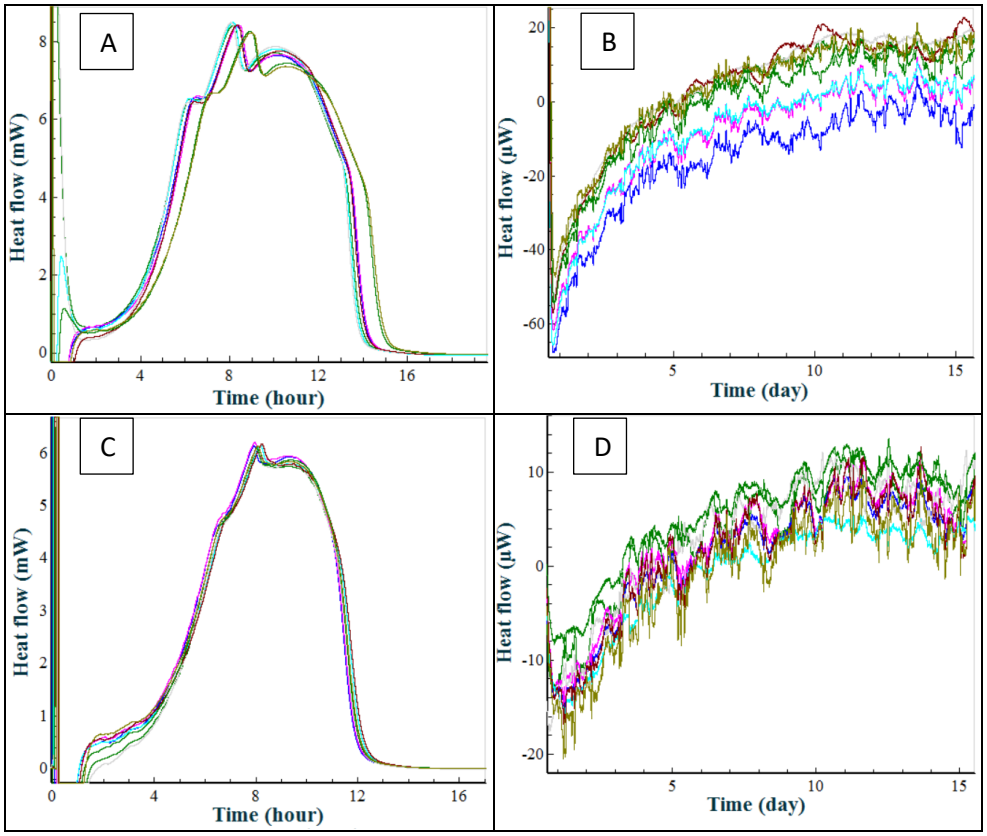

Figure 19. First attempt of GASP cultivation, each colored line indicates an individual ampoule. A = Heat flow curve of whole-MRS-broth *L. reuteri* during growth phase. B = Heat flow curve of whole-MRS-broth *L. reuteri* during long-term stationary phase. C= Heat flow curve of half-MRS-broth *L. reuteri* during growth phase. D = Heat flow curve of half-MRS-broth *L. reuteri* during long-term stationary phase. Note that A and C are in mW scale while B and D are in µW scale. Note also that A and C are in hour scale while B and D are in day scale.

In the second attempt to capture a GASP by IMC, the baseline was not established in the best way (see Materials and methods), so the graphs should be viewed with some caution. Figure 20 A shows the heat flow during initial growth phase of the *E. coli* in LB broth controls. The graph is from a screenshot, this data was not collected into the "experiment" file in the TAM Air Assistant program because the "experiment" was actually started at the beginning of the lag phase (18 hours after insertion of samples into the calorimeter). Therefore, the absolute values may be biased in this graph. The heat flows during long-term stationary phase of *E. coli* controls (colored lines Figure 20 B) show a dip into negative values at the very beginning of the experiment. This dip is very likely explained by the late baseline establishment – there must have been some residual heat activity from initial growth of the cultures when the baseline was established and this heat flow then diminished, giving a negative heat flow reading. Otherwise, not much happened

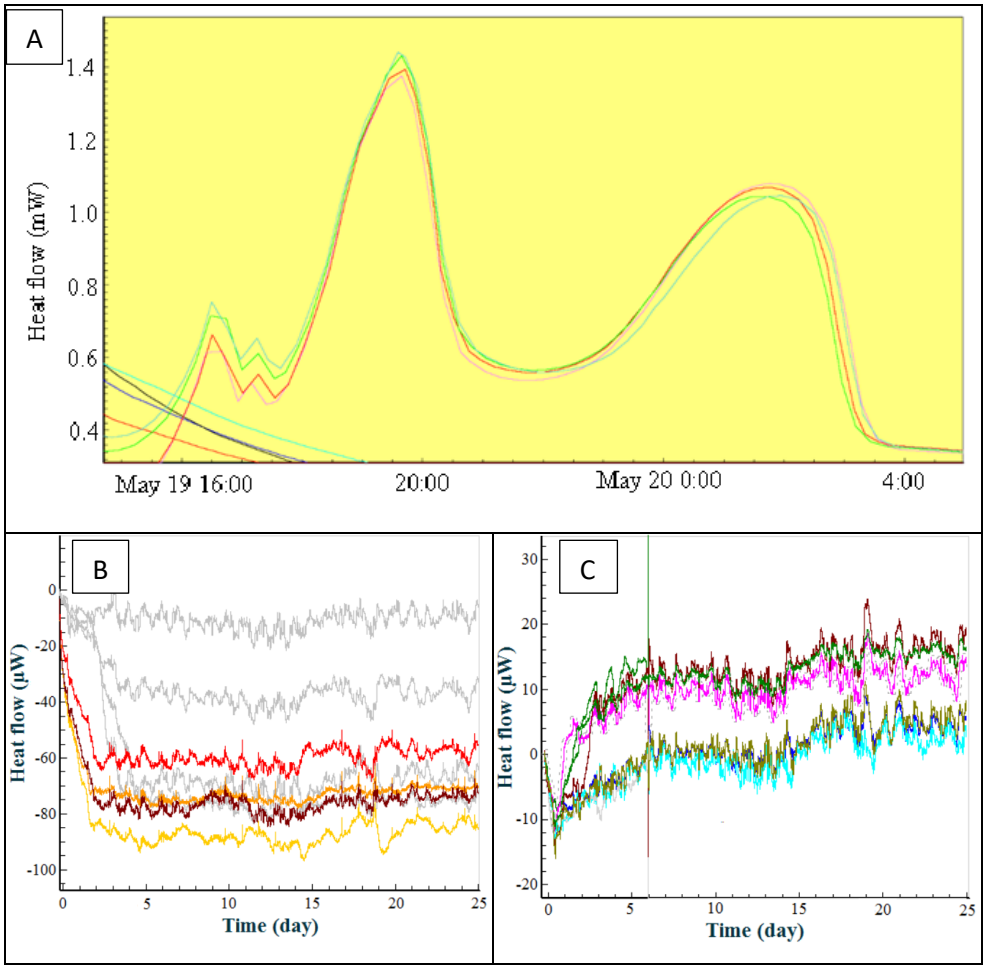

Figure 20. Second attempt of GASP cultivation, each colored line indicates an individual ampoule. A = Heat flow curve of *E.coli* in LB broth during growth phase. B = Heat flow curve during long-term stationary phase of *E. coli* in LB broth control (colored lines) and old whole-MRS *Lactobacillus* (gray lines). C = Heat flow curve of *L. reuteri* in half-MRS-twice-MOPS broth during long-term stationary phase. Note that in B and C, day zero is the start of experiment which in this case is at the beginning of the stationary phase. The completely vertical heat flow lines seen in picture C are the heat flows resulting from retrieval of two ampoules for viable counting.

with the heat flow of *E. coli* controls during the long-term stationary phase. As a remark however, the patterns of fluctuations were not always identical between the different *E. coli* ampoules. Figure 20 B also shows the heat flows of the very old whole-MRS cultures from the previous GASP experiment (gray lines) which had been kept incubated at 37 °C for 30 days after the first IMC experiment and brought back to IMC for the second experiment. These samples were quite intriguing; most of them show a rather strong dip into negative heat flow, some as low as -70 mW, after a few days in IMC. Why would these cultures, which had stabi-

lized around 0 heat flow during the previous experiment, suddenly drop so low a few days after the next experiment was started?

The behavior of the half-MRS-twice-MOPS *Lactobacillus* samples during long-term stationary phase is shown in Figure 20 C. The heat flow pattern was quite similar to that of half-MRS *Lactobacillus* in the first attempt (Figure 19 D): the negative dip reached about -15 µW, then climbed up to a peak of 8 µW after 11 days (in second attempt, half of the samples peak at 0 µW while the other half peak at about 11 µW, after 7 days). After the peak, half-MRS samples in the previous experiment (Figure 19 D) showed a small depression and then the beginning of a new, mild climb at day 15. The half-MRS-twice-MOPS samples (Figure 20 C) also showed the small depression followed by a new, mild climb, also starting at day 15. This climb seems to continue very vaguely throughout the experiment.

The results of the VCs done on ampoules retrieved from the calorimeters are presented in Table 14 below. After 21 days, the non-buffered samples had lost all detectable viability. VC of the MOPS-buffered samples was unfortunately only measured after 7 days and at that time exhibited a GASP-like level.

Table 14. *Viable count results of L. reuteri cultivated in different media during the GASP experiments, in which the bacteria were cultivated for a prolonged time in isothermal calorimeters*

Samples – first experiment	CFU/mL
whole-MRS, day 16	2.3×10^3
(two ampoules)	5.7×10^3
whole-MRS, day 21	$<10^2$
(two ampoules)	$<10^2$
half-MRS, day 21	$<10^2$
Samples – second experiment	**CFU/mL**
half-MRS-twice-MOPS,	1.5×10^6
day 7	1.9×10^6
(two ampoules)	

Additionally, a summary of the pHs in stationary phase of *L. reuteri* cultivations in various media is found in Table 15 below.

Table 15. *Summary of pH measurements of different media tested with L. reuteri in the "personal bioreactor" before the second attempt of GASP cultivation*

Medium	Cultivation time	pH
whole-MRS	fresh medium	5.53
whole-MRS	22 days	4.29
half-MRS	22 days	4.24
whole-MRS-MOPS	fresh medium	6.34
whole-MRS-MOPS	~24 hours	4.58
half-MRS-MOPS	fresh medium	6.82
half-MRS-MOPS	~24 hours	4.75
half-MRS-twice-MOPS	fresh medium	7.25
half-MRS-twice-MOPS	3 days	5.46

Finally, Figure 21 below shows the growth curve in half-MRS-twice-MOPS medium, measured by the Personal Bioreactor. Growth seemed exponential to perhaps the third hour, after which it became rather linear. Similar results were seen in (Årsköld *et al.*, 2008) when a different strain of *L. reuteri* grew on glucose.

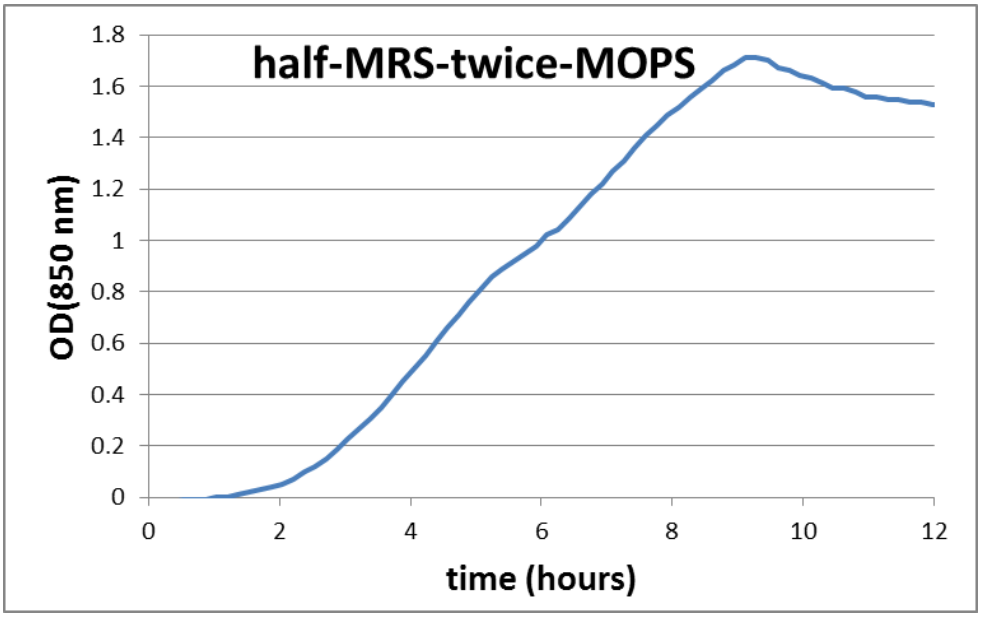

Figure 21. Growth curve of a test cultivation in the "personal bioreactor". The personal bioreactor has an OD_{850} measurment range of 0-8 at 30 mL cultivation, equivalent to OD_{600} 0-19 (*RTS-1C, Personal bioreactor — Biosan*).

6 Results – Saline test

There was little difference in the VCs when equivalent biomass was resolved in either 0.9% or 0.09% saline solution (Figure 22).

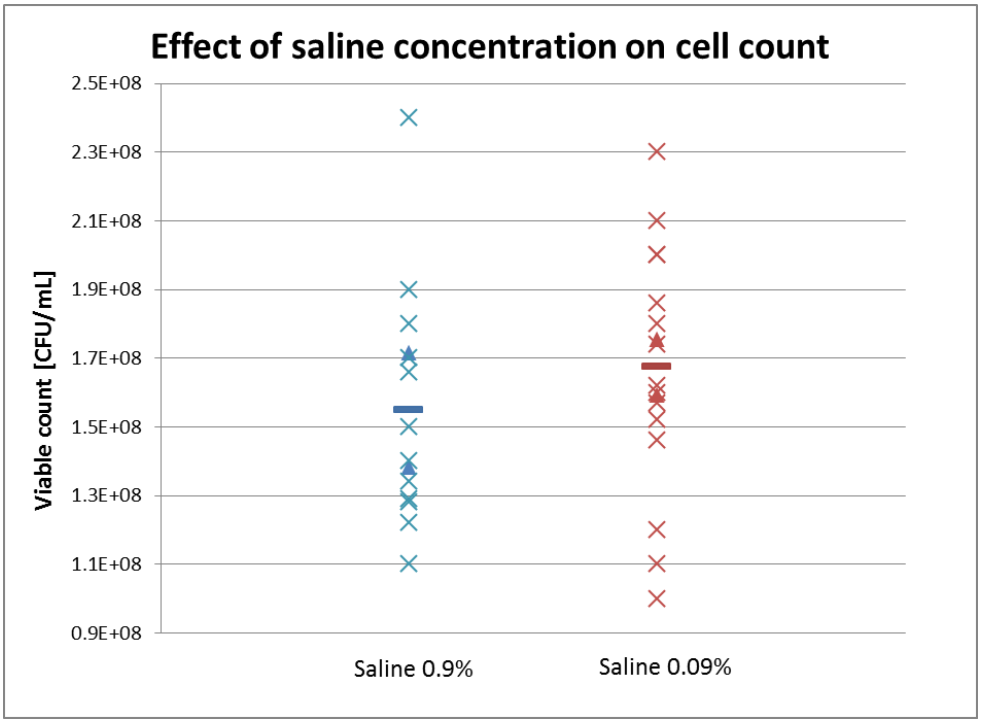

Figure 22. Viable counts of equivalent biomass when resuspended with different concentrations of saline. The rectangular markers are the grand means of the counts. The triangles are the means of the individual dilution series. The cross markers are the individual counts.

Interestingly, the 0.09% solution gave slightly higher viable counts than the standard solution of 0.9% but the difference was so small that it was far from statistical significance, with this amount of measurements. It seems reasonable to say that the accidental use of the 0.09% solution for a period of time in the experiments did

not affect the results or conclusions noticeably. Furthermore, samples of all treatments were treated with the same saline at a given viable count, so the relation between treatments should not have changed.

7 Discussion

7.1 Growth during cultivation

The observation that exponential growth changed to a linear growth after 2-5 hours of cultivation (Figures 3-5), time being dependent on fermentation conditions**,** is similar to what was observed with the strain ATCC 55730 when grown on glucose (Årsköld *et al.*, 2008) due to a redox imbalance. Adding sucrose to the cultivation media could enhance the growth rate and the yield because it provides an external electron acceptor (Årsköld *et al.*, 2008). Perhaps, as in that study, supplementing the medium with sucrose to improve the redox balance would give faster growth and better yield. In addition, it could perhaps prepare the cells for the subsequent sucrose formulation. It could even precondition the cells for the high osmolality during drying, as for example, shown by (Tymczyszyn *et al.*, 2007) where sucrose-supplemented MRS medium for cultivation of a different *Lactobacillus* species improved the drying survival, although the drying method was vacuum-drying. As a remark, using fructooligosacharides as the formulant was better than sucrose for increasing freeze-dry survival, long-term storage after freeze-dry survival and membrane integrity of *L. reuteri* TMW1.106 (Schwab *et al.*, 2007).

Finally, it is important to remember that the resistance to gastric juices and beneficial health effects of the product can be affected by the whole process starting from cultivation conditions and ending at resuspension of the product. This must also be assessed early in the studies, to avoid spending research resources towards a potential dead-end. Therefore, it is important to find good and not to laborious assays for these qualities and couple them with upcoming optimization experiments.

7.2 Freeze-drying and collapsing

Regarding the freeze-drying procedures in this study, apparently collapsing and variable batch quality were major issues. When it comes to the vacuum level, according to (Fonseca *et al.*, 2015) the recommended pressure to reach before the primary drying step is 20 Pa. Then, the pressures used in this study seem to have been sufficiently low, as the highest cases were 16 Pa (although the pressure gauge had low resolution). Unfortunately, the Labconco freeze-drier did not have both a Pirani gauge and a capacitance manometer to identify the point of a finished sublimation nor was it possible to record the probe (thermocouple) temperature. The temperature was tracked during the vacuum pumping and the beginning of the primary drying to look for collapse risks. When the samples had been inserted into the freeze-drier at a shelf temperature of -40 °C and the vacuum pumping was initiated, the temperature of the samples (according to thermocouples) would keep rising for as long as 20-30 minutes. In batch 4, the worst-case thermocouple showed -32 °C during this phase. However, when the primary drying was initiated, the temperatures of the probed samples were below the shelf temperature due to sublimation (Fonseca *et al.*, 2015) (-30 to -35 °C versus a shelf temperature of -20 °C) and increased to above the shelf temperature during the next day, as it is supposed to be when sublimation is finished (Fonseca *et al.*, 2015). In batch 5, the temperature of the probes slowly increased (over the course of 28 minutes) from -40 °C to between -31 and -34 °C during the vacuum pumping. In batch 6, the probes stayed at -38 to -41 °C during vacuum pumping. In batch 7, the temperature increased to between -28 and -33 °C during vacuum pumping. Unfortunately, the notes of the temperature behavior during primary drying are lacking in batches other than no. 4. Batches no. 1 and 2 were not recorded. In compliance to what was written in the Introduction, section 1.2, it can be anticipated that the bacterial formulation used in this work had a T_c no less than -24 °C. Perhaps the shelf temperature during primary drying should have been lower to avoid collapse. However, the major issue turned out to be that the Labconco machine was in need of service and therefore struggled with vacuum pumping. This made the sublimation of ice less efficient and therefore increased the collapse risk, since it results in higher humidity and temperature around the samples.

As it is mentioned in (Fonseca *et al.*, 2015), the presence of a thermocouple probe in a sample will catalyze its sublimation process. Indeed, in batch 7 in which some of the samples were well-dried and the others were collapsed, those that had a probe were the best-looking ones (maybe because a higher water content decreases T_g' (Higl *et al.*, 2007), the probed samples dried faster and avoided collapsing). This could be seen as an error source toward the other samples because it is an inconsistency in the freeze-drying process and since these samples were most

probably the driest, they would have a lower tendency of deteriorating during storage as well. Nevertheless, non-collapsed biomass was scarce in batch 7 and so these samples had to be used for the measurements too. These samples also lacked vacuum because the thermocouples had to be taken out before closing the corks. However, in this study there were a few other cases in which a non-probed sample lacked vacuum and could be compared to a duplicate with vacuum. There was no consistent trend in the influence of vacuum on storage-stability.

7.3 The effect of fermentation conditions on storage-stability

In this study, it was demonstrated that a neutral pH (6.5) during cultivation of *L. reuteri* led to higher storage-stability after freeze-drying than when using lower pH, clearly seen in Figure 11. Microscopy revealed that this fermentation treatment also gave rise to an elongated cellular morphology and absence of cell clumps. When looking at the more subtle relation between treatments A (37 °C, pH 4.5) and B (32 °C, pH 4.5), usually, treatment A gave better storage-stability than B but after 3 days (and 1 week in log-transformed data), it was the other way around. The reason for the fact that the relation between A and B could change between non-transformed and $\log_{10}$-transformed data at timepoints 3 days and 1 week lies in the procedure of transforming the data; first logarithmizing the technical replicates, taking their average for each biological replicate and finally letting SPSS calculating the treatment means. In cases where the difference is small, this mathematical procedure could turn the relation. Regarding how to interpret this in practice, it is suggested to simply trust the "reliable region" on each graph (in Figure 11 A and B). However, consultation with a statistician regarding this issue would be beneficial. For plotting the correlation of FT assay survival with stability in accelerated storage assay, however, the data used was neither batch-adjusted nor transformed. In conclusion, there was a consistent difference between treatments but the low number of replicates and large variance in stability made the 95% CIs broad. Still, the treatment of 32 °C, pH 6.5 stood out from the others, even significantly. The difference between treatments A and B, although not statistically significant, is supported by a consistent relation in all cases except one early timepoint (3 days).

In a similar study to the present one (Palmfeldt & Hahn-Hägerdal, 2000), *L. reuteri* ATTC 55730 also exhibited elongated cells at pH 6 versus pH 5. According to that study, the relative initial survival after freeze-drying was better at pH 5 than at pH 6 (at 37 °C in both cases). Unfortunately, the storage-stability of the biomass was not investigated. In another study (Koch *et al.*, 2008), the same scenario is repeated with *Lactobacillus delbrueckii*: lower pH – shorter cells – higher

freeze-dry survival but no storage-stability follow up. In that study, best freeze-drying survival was after fermentation with uncontrolled pH which ended at pH 3.4 and at the second place came the lowest controlled pH of 5. Interestingly, in the present study, the initial freeze-dry survival was likewise better at 37 °C, pH 4.5 versus 32 °C, pH 6.5 (and overall best but not statistically significant, see Figure 11). In the present study, however, when it comes to storage-stability, the tables became turned and the higher pH treatment was best. Longer cells may be more susceptible to damage during freeze-drying (as well as simple freezing) as suggested in (Fonseca *et al.*, 2000) and by the results of the present study's FT-assay and the abovementioned references, but the present study also showed that these longer cells seemed to have a big advantage in long-term storage (and perhaps robustness), for reasons yet to be discussed. However, treatment B (32 °C pH 4.5) which gave medium long cells had, in general, the worst storage-stability (see Table 16 below and Figure 11 in section 4.3), so there might be something else than cell size involved (although the difference in storage-stability between treatments A (37 °C, pH 4.5) and B (32 °C, pH 4.5) was not statistically significant). It would be very interesting to compare the current treatment C (32 °C, pH 46.5) with a treatment of 37 °C, pH 6.5 which combines a higher temperature with higher pH.

In a third study (Schoug *et al.*, 2008), *Lactobacillus coryniformis* was subjected to many different temperatures and pHs before freeze-drying. Best survival was observed when treated with optimal growth conditions (pH 5.5 and 34 °C) or by warm treatment (42 °C). Higher and lower pH gave lower survival. However, in that study, the fermentations were done at the optimal pH and temperature for cell yield and the "treatment" pH and temperature was applied first upon late logarithmic growth phase. Highest survival was linked to less unsaturation of fatty acids in cell membrane. Finally, a fourth study (Liu *et al.*, 2014) performed in a similar fashion to the last one, *L. reuteri* I5007 was subjected to different temperatures and pHs before freeze-drying. This time, high temperatures (47 °C) or high pH (6.7) gave the best freeze-drying survival but there was no storage-stability investigation. In that study, the fermentations were also done at the optimal pH and temperature for cell yield and the "treatment" pH and temperature was applied first upon the stationary phase. In contrary to (Schoug *et al.*, 2008), the higher survival was linked to higher levels of unsaturated fatty acids in the cell membranes. Perhaps another local optimum was found, or it is a matter of difference between species. In the present study, the cell density before freeze-drying was not controlled as in (Schoug *et al.*, 2008) and (Liu *et al.*, 2014) because the special growth conditions were fixed throughout the cultivation. Because of this, the cell density factor was uncontrolled for and this factor has been shown to be significant for survival of *Lactobacillus coryniformis* in (Schoug *et al.*, 2006) where higher

cell densities gave higher survival of freeze-drying. In the present study, treatment C (32 °C pH 6.5) coincided with highest cell densities out of the treatments tested (see Table 16) and as mentioned, correlated with better storage-stability but worse freezing survival. Because there was no storage assay in (Schoug *et al.*, 2006), it is hard to draw a direct parallel to the first finding. At least a rational reason for the worse freezing survival of the elongated cells might lie in the large cell surface of elongated cells, leading to more damage from extracellular ice formation; also the small cell surface : cell volume ratio of elongated *Lactobacilli* affects the water flux from the cell and may promote intracellular ice formation (Fonseca *et al.*, 2000). One possible reason for the better storage-stability of cells from treatment C (32 °C pH 6.5) is that longer cells and higher cell densities increase T_c substantially (Fonseca *et al.*, 2004b) and since there were problems with collapsing (perhaps even the non-collapsed batches were on the verge of collapse at the microscopic scale), this could have been an advantage for these cells. However, this is contradicted by the fact that treatment B (32 °C pH 4.5) with medium long cells and medium cell density were the least storage-stable case, although more statistical power would be needed to confirm this relation. Perhaps the absence of cell clumps in treatment C (32 °C pH 6.5) was the main factor, or there are some deeper molecular causes for this result.

Table 16. *Summary of fermentation treatments and survival of freeze-drying (initial survival), freeze-thaw (FT) assay and 1 week of accelerated storage assay, in relation to observed cell length and cell density of formulated L. reuteri before freeze-drying. For simplicity, the shown average values are not transformed or batch-adjusted*

Treatment	Average initial survival	Average FT-assay survival	Average survival after 1 week	Average estimated cell length (µm)	Average cell density of formulate (CFU/mL)	95% confidence interval (cell density)	Sample size (cell density)
37 °C pH 4.5	1.074	0.00903	0.125	2.1	2.03×10^9	$\pm 1.74\times10^9$	5
32 °C pH 4.5	0.559	0.05980	0.051	2.7	3.64×10^9	$\pm 1.74\times10^9$	5
32 °C pH 6.5	0.933	0.00017	0.456	6.1	6.45×10^9	$\pm 1.95\times10^9$	4

The differences between initial survival ratios in the present study were far from statistically significant as the spread of individual measurements was large and this should be kept in mind. It is worrisome that while the grand adjusted means are in an almost reasonable range, one individual initial survival ratio is as high as 1.9

and another is 1.4. No ratios should be higher than one, anything else indicates cell growth, which would be unlikely. However, one theory is that this could be due to dissolving of cell clumps after freeze-drying (which was actually observed to some extent in microscopy, data not shown). However, the 1.4 initial survival ratio belongs to treatment C (32 °C, pH 6.5) which never had any cell clumps. Unfortunately, this particular measurement belonged to batch 2 which was before microscopic studies were applied so it cannot be confirmed whether there were no cell clumps. In future experiments, some anti-clumping agent could be considered for the before-freeze-drying VC. Perhaps a more likely explanation is that the procedure for doing the before-freeze-drying VC was suboptimal and gave underestimated starting cell counts. This can be suspected because the formulated biomass had to stand and wait during which all the freeze-drying vials were filled and put into the freezer and other laboratory tasks had to be done. The bacteria could have been damaged by floating around in a 15% sucrose-in-distilled-water solution for a long time (something that could be experimentally investigated). In such case, the relative difference between the initial survivals should still be valid, as all samples were subjected to similar procedures.

7.4 Correlation experiment analysis

The correlation between FT assay survival and stability in accelerated storage assay was mainly explored after 1 week of accelerated storage (Figures 13 and 15). This choice of timepoint was based on several reasons; firstly, there were many biological replicates at that point (see Table 6). Secondly, there was a statistically significant finding in the non-transformed, more easily interpreted dataset (Figure 11 A) while at the same time, it was a timepoint at which the samples had been exposed to accelerated storage for a substantial period. Additionally, both of the timepoints 1 week and 2 weeks have rather low coefficients of variation (CVs) in non-transformed values (see Table 6). It is important to remember that this correlation may not apply to other formulations than 15% sucrose, will not apply to storage assays at different temperatures and that better avoidance of collapsing may change it too.

The FT-assay survival correlates negatively with storage-stability, or more precisely, with the total CFU/mL after freeze-drying and accelerated storage. This includes the survival of a freeze-drying and a given time of storage. To really isolate the storage-stability itself, the reference CFU/mL value for calculating storage survival would have to be the CFU/mL after the freeze-drying and not (as done in this paper) the CFU/mL of the formulated biomass before freeze-drying. However, this was omitted due to lack of time. Since when it comes to surviving one freeze-

drying (a freezing process involved), treatment A (37 °C, pH 4.5) was actually the best and treatment C (32 °C, pH 6.5) was the midst (although this was not significant) which breaks the correlation seen in FT-survival (another freezing process) where treatment B (32 °C, pH 4.5) is the best and C (32 °C, pH 6.5) is the worst. This could, however, simply be due to random variation since the initial survival differences were not statistically significant.

Although the correlation between the FT-assay and the accelerated storage stability assay is inverse, it does not exclude the use of FT-assay for storage-stability prediction; using the correlation equation will give a prediction assuming similar experimental conditions. However, it is important to make sure that the CFU/mL is acceptable before the FT-assay, otherwise a low FT-survival could simply mean that the bacteria were dead from the beginning.

The inverse correlation lead the thoughts to the possibility that some (possibly cellular) factor makes the cells more resistant to freezing but at the same time less resistant to storage, and vice versa. This *could* be one factor but it could also depend on two (or more) factors which could then perhaps be optimized independently of each other to some extent, but happened to be confounded in this study by varying only pH and temperature. If there are many factors involved, a goal should be to identify these and design two assays, one which positively correlates with high survival of the freeze-drying process and one which positively correlates with storage-stability.

Further optimization of the freeze-drying of *L. reuteri* (by using factorial designs) may hopefully lead to a recipe that assures both a relatively high freeze survival and storage-stability. In such case, the correlation equation found here will not be of much practical use but may at this point contribute to a wider understanding of the phenomena affecting freeze-dried bacteria.

7.5 Statistics – experimental design

As mentioned in the introduction, the original experimental plan looked differently than what was eventually done. The learning goal within the field of statistics was shifted from the cancelled formulation-screening experiment to the FT-assay/accelerated storage assay correlation experiment and a big literature study was done to find a statistical strategy fit for the situation. This turned out to be a much more winding and strenuous path than expected, especially considering the collapses of some batches and the changing batch size. Ideally, a researcher should not look for statistical methods that fit to the collected data but instead plan the experimental design knowing that feasible statistical methods will be applicable to the data generated. In this case, it had to be the other way around to get hands on

with a little bit more advanced statistics and to actually make better use of the batch-wise scattered data.

Paired t-tests were also considered when choosing the statistical method. However, using three paired t-tests to test the differences between the three treatments was not feasible because not all treatments were represented in each batch. Therefore, in a batch-paired set of t-tests, the mean of treatment A, for example, would be different depending on if it was being compared with the mean of treatment B or C which would give inconsistent results and would be very awkward to present.

The finally applied unbalanced RIBD was a “design of necessity”. All treatments were paired (A paired with B, A paired with C and B paired with C) in blocks an equal number of times (except for timepoint 22 days which was lacking some data) and the sample size was equal for all treatments (again, except for timepoint 22 days). These are two out of three criteria for a balanced RIBD, however, the criterion of equal amount of treatments in each block (the third criterion) was not fulfilled since some of the early batches contained all three treatments while the rest contained only two treatments each. Excluding these larger batches would have given a balanced design but the sample size would become worryingly low; the homogeneity of variance would not be meaningful to assess and a lot of data would be wasted. Alternatively, after assessing the assumptions of ANOVA, the large batches could be excluded to get a balanced design while assuming that since the data is taken from the same populations, compliance with assumptions of ANOVA can still be assumed. This is however, left for the interested reader. There are ways to analyze block designs with unequal block sizes which are still called balanced (Gupta & Jones, 1983; John, 1998) but this was left out of the scope of this study.

The variance was not split more than into treatments (as a fixed factor) and batches (as a random factor) for two reasons. Firstly, to split it further into individual freeze-dry vials as a factor (or ”source of variation”) would require the use of a so called “nested/hierarchical” design (Miller & Miller, 2010) because the vial identity would never be replicated, each vial is unique and only used once. A nested design does not require all combinations of factors (such as the impossible scenario of having vial no. x in all of batches no. a, b and c), instead they build on a hierarchic structure of factors and factors within factors. However, attempting a nested design did not seem worthwhile because of the dominating batch variation; separating the smaller sources of variation would not help. Furthermore, it would perhaps even be impossible since the design was already an incomplete unbalanced one and deeper studies of the nested design were omitted. It is obvious that batch number was the dominating nuisance variable and efforts should be done in future experiments to reduce this variance. Making two replications of the dilution

series and the plates may seem futile in the light of this; however, it was an action against outliers and possible gross mistakes.

When it comes to the benefit of having used a block design, no formula was found for assessing this in RIBDs. The mean squares of the block factor were almost always greater than the mean square error, only the initial survival was an exception. However, to judge whether these ratios were high enough to consider blocking worthwhile would require more literature research. Compared to the mean square of the treatment factor, the block mean square could be much smaller, similar or larger in the many ANOVAs without any completely certain trend. When it comes to how the significance of the treatment effect was influenced by blocking, one-way ANOVAs without the block factor were run (results not shown). The initial survival non-transformed and the 2, 3 and 4 week $\log_{10}$-transformed ANOVAs yielded increased significance of the treatment effect thanks to blocking while in the other cases the significance got lower. Perhaps the benefit of blocking was a border case in this experimental design. However, the ANOVAs run without also resulted in greater heterogeneity of residuals (especially residuals of treatment C had lower variance). Therefore, blocking could in this case have improved compliance with the statistical assumptions.

Another aspect important to remember is that the $\log_{10}$-transformed analyses are not as straightforward in the practical interpretation as the non-transformed. The reason for this is not only that the geometric means must be used as the real-world interpretation; it is also about the fulfillment of assumptions and the reliability of the analysis. According to (Montgomery, 1991), when transformation is needed to get homogeneity of variance, the ANOVA will only be reliable in the transformed scale. Also, the coefficients of variation (CVs) examined in Table 7 for $\log_{10}$-transformed data are valid only for the means of $\log_{10}$ transforms.

As a final remark, regression methods (Altman & Krzywinski, 2015) for statistical evaluation of the correlation between FT assay and accelerated storage assay itself, were not used due to lack of time for researching the methodology.

7.6 Statistics – power

No power analysis had been done before the study since this statistic approach was not planned. However, a power analysis can be done after the experiments too (in order to at least see what the power was, as long as the scientifically interesting difference has been predefined). It was attempted to make such a power analysis in G*Power 3 (Faul *et al.*, 2007; *Universität Düsseldorf: G*Power*). However, it seemed like in G*Power 3, for a multifactorial ANOVA power analysis, each *combination of factors* must consist of at least two experimental repeats because

the “total sample size” must be larger than “number of groups”. Since each batch could only contain one of each treatment (at most), the requirement from the software was not fulfilled. More literature research would be needed about how to analyze the power of block designs with only one replica of block-treatment combinations. Fortunately, having only one replicate per treatment in a batch seems to be a standard situation (Krzywinski & Altman, 2014a) so there must be plenty of software available for this situation.

How big should effect size (see section 2.6 “Statistics”) be set for power analysis prior to potential future experiments, if one would like to build upon the present study? This will only be discussed in general terms, as it depends on many choices in the experimental plan. Firstly, it depends at what timepoint of accelerated storage one wants to do the measurements. It will also depend on whether $\log_{10}$-transformation will be used. A sensible difference to detect could be one $\log_{10}$-exponent. If no transformation will be applied, the *difference* between two non-transformed means corresponding to one $\log_{10}$-exponent is going to become smaller and smaller as the means get lower while the differences between transformed means would always be equal to one. For example, the difference between logarithmized means -2 and -1 is 1 while 0.1 – 0.01 is 0.09 and the difference between -3 and -2 is 1 while 0.01 – 0.001 is 0.009. The choice of timepoint will also make the standard deviations of these means to change but not necessarily keep their proportions to the means unchanged. Therefore, the effect size corresponding to a difference of one $\log_{10}$-exponent will depend on chosen timepoint and transformation. Furthermore, it will depend on the general design of experiment such as blocking because the standard deviations may be estimated from a block adjusted or completely randomized design. Nevertheless, once these choices have been made, standard deviation data from this study may facilitate a proper choice of sample size. The effect size can be calculated and used in a power analysis software to calculate the sample size needed for a specified statistical method, or the information can be used with tabulated “operating characteristic curves” (Montgomery, 1991) to find the required sample size. As an example, consider that one is interested in measurements after 1 week of accelerated storage, with non-transformed data and that an interesting difference between treatment group means would be one of 50%. Treatment C could be kept as a standard because it turned out to be so storage-stable; according to the present study it would have a mean survival ratio of (roughly approximating) 0.5 and a standard deviation of 0.24. However, this was the standard deviation of the unadjusted mean. The pooled standard deviation of the block adjusted means was 0.156 in this case (see Table 6) so this value could be used if it would be intended to do a block design. Since new fermentation treatments would probably be chosen in the successful high-pH area, a reasonable (and simple) assumption would be that these treatments would have a

similar standard deviation (otherwise the anticipated standard deviations must be pooled).

For a multilevel ANOVA, an appropriate measure of effect size is Cohen's f (Maher *et al.*, 2013) according to equation (9):

$$f = \frac{\sqrt{\frac{\sum(m_i - \bar{m})^2}{k}}}{\sigma} \quad (9.)$$

where k is the number of groups compared, m_i is the mean of group i, $\bar{m}$ is the mean of group means and σ is the pooled standard deviation. The mean survival of treatment group C can from the results of the present study be entered as 0.5. If the difference of interest is 50%, another mean can be entered as 0.25. The third mean (if it is assumed that k=3 in this planned experiment too) can have an arbitrary but not more extreme value, for example 0.5 again. Calculating the mean of these means will yield $\bar{m}$ = (0.5 + 0.5 + 0.25) / 3 = 0.417 and σ can be guessed to stay at 0.156. Entering these values will give f = 0.76. This is the effect size that can be used in a power analysis software once the statistical method/model has been chosen (remembering that it must employ blocking because σ has been estimated in a block design).

7.7 Microscopy

Gram-staining unfortunately did not yield any conclusive results, since the results only correlated with batch number. Before implementing Gram-staining for *L. reuteri* in this study, a test session of Gram-staining was done with different bacteria on the same objective and the results looked appropriately. However, there should have been reference bacteria sharing the objective with every subsequent *L. reuteri* sample to see if the stains were successful, as pointed out in (Smith & Hussey, 2005). This source also pointed out that the samples for staining should better be thin but this was overlooked, however. Following this advice by diluting the samples could perhaps have given more consistent results.

The suspected correlation of Gram-stain result with batch number could be explained by the medium quality and other batch-biased procedures. More speculatively, there could have been a procedural difference in how the Gram-stain was performed each time, although a strict protocol was followed. However, there could have been a difference in how effectively the microscopy plates were washed in running tap water for each batch during the steps of Gram-staining; it is important to do it delicately so that the sample is not washed away yet quickly to follow the protocol times.

As seen in panel "(C) batch 6" in Figure 29 in Appendix, there are quite many short cells present in this case, although at close inspection plenty of long cells can still be found. The large amount of short cells in this sample was surprising, since it belongs to treatment C (32 °C, pH 6.5) which in all other microscopy photos showed mostly medium to very long cells. An unstained sample from the same biomass can be seen in Figure 28, panel "(C) batch 6" in Appendix and it exhibits much more elongated cells. Perhaps the short cells were somehow up-concentrated in the Gram-stained sample during washing steps of the staining procedure. Anyhow, it would be best to repeat Gram-staining of a sample from treatment C. A small remark is that the cell size in photos of Gram-stained samples seemed overall smaller than in the non-stained microscopy images from Figure 18 in section 4.7 and Figure 28 in Appendix. Either the staining procedure did shrink the cells or it improved the contrast allowing to see their true sizes.

7.8 GASP

What could be the reason behind that no obvious signs of GASPs were observed? Could the two mild "hills" in the heat flow during the long-term stationary phase (Figure 19 D and Figure 20 C) represent the activity of two subsequent GASP cultures? Why then, was this not observed in *E. coli*? At the time of doing the experiment, it was anticipated that the appearance of GASPs would be seen as obvious growth peaks reaching perhaps hundreds of µWs. Therefore, the experiment seemed to be fruitless and not much VCs were taken. Filling in the missing VCs would greatly help to answer the questions and fortunately, this can be easily ameliorated by simple complementary experiments - a repetition of the cultivations with MOPS buffer and those with *E. coli* in the same conditions as previously but without any need of microcalorimetry.

Nevertheless, many conclusions can already be drawn: since the heat-flow pattern seemed very similar between non-buffered cultures (Figure 19 D) and MOPS-buffered cultures (Figure 20 C), it did not look promising for the GASPs, since the cultures in non-buffered medium had died out (see Table 14).

Furthermore, examining the literature more carefully provided some new insights. Firstly, all previous GASP papers, which have a sufficiently detailed methods section to tell, describe cultivations with shaking and/or aeration (Zambrano *et al.*, 1993; Finkel & Kolter, 1999; Smeulders *et al.*, 1999; Vasi & Lenski, 1999; Bruno & Freitag, 2011). Supposedly there was an anaerobical cultivation too (with unknown shaking), in a PhD thesis which could not be retreated, referred to from (Zinser & Kolter, 2004). Neither shaking or efficient aeration was possible in the IMC, could this explain lack of GASPs in the *E. coli* control cultures? Assuming,

of course, that there was in fact a lack of GASPs, it is something to be checked with VCs (to begin with). Without shaking, the biomass formed a hard-to-dissolve plaque at the bottom of the ampoules. Perhaps all the nutrients needed for GASP growth were "locked" within this plaque? Without aeration, *E. coli* shifts to the heterofermentative "mixed-acid" pathway and produces acetic acid, lactic acid, succinic acid, ethanol, CO_2 and H_2 (Madigan & Brock, 2009). However, this is the case when fermenting glucose. The LB medium (also used in some of the successful GASP studies) contains only small amounts of various sugars; instead amino acids are the main carbon source (Sezonov *et al.*, 2007). As mentioned in the introduction, amino acids seem to be of greatest importance also during the GASP growth. Evidently, *E. coli* does grow anaerobically in LB medium (Colón-González *et al.*, 2004). According to (Losen *et al.*, 2004), when *E. coli* grows in unmodified LB-medium aerobically, it uses very low amounts of oxygen. The bacteria mainly use the amino acids as an energy source and therefore, the pH of such culture rises up to 8.7 (Losen *et al.*, 2004). However, this rise in pH neither influences final OD or respiration. Adding buffer does not change anything except the pH and the growth limitation is therefore a limitation of carbon sources (Losen *et al.*, 2004). All of these arguments point toward the hypothesis that the anaerobicity of the cultivations should not have influenced GASP evolution because if GASPs live off amino acids, they should cope with anaerobic conditions equally well as the original LB-broth cultivation, unless the anaerobic situation makes the pH climb even higher. What metabolic pathways are used when *E. coli* grows anaerobically on LB broth (amino acids) would need deeper studies. Furthermore, it is unfortunate that the *E. coli* inoculate used for the controls was not freshly out of a cell bank, but taken from the neighboring laboratories cultures. This could potentially mean that these *E. coli* had already gone into GASP formation before the start of the experiment. Also, it would have been best to make sure that the same *E. coli* strain was being used as in the classical GASP experiments.

As a final remark, the theoretical detectability of GASP growth in IMC will be discussed. As already mentioned before, the total CFU/mL stays quite constant in the long-term stationary phase but with a dynamic interchange of living bacteria if GASPs appear. Looking at the CFU/mL values in the GASP literature, with *E. coli*, the highest cell count during a cultivation was about 5×10^9 CFU/mL and in long-term stationary phase it fell to about 5×10^7 CFU/mL (Finkel & Kolter, 1999). The values were roughly the same presented in (Finkel, 2006) and it was commented that 99% of the cells died (two logs) in the passage from stationary phase to long-term stationary phase. With the Gram positive *Listeria*, there was roughly 5×10^9 CFU/mL at the peak while it went down to 5×10^8 during long-term stationary phase, which means that the net decline in CFU/mL was only 90% (Bruno & Freitag, 2011). Even if the latter scenario is assumed, that 10% of the original cell

count must re-grow in the form of a GASP, this should roughly give off as much heat as 10% of the heat given off by the original growth phase. Furthermore, when the GASPs grow, the old cells must lyse and many of their components such as peptides and DNA must be hydrolyzed. The hydrolysis of polymers such as peptides and nucleic acids is itself an exothermic process (Sturtevant, 1953; Armstrong *et al.*, 1977) and would add up to the heat. Could some other endothermic process have cancelled out the heat-producing reactions of polymer hydrolysis and GASP growth?

8 Conclusions

8.1 Correlation experiment conclusions

FT-assay survival correlated negatively with accelerated storage stability after freeze-drying (more precisely, with the total CFU/mL after freeze-drying and accelerated storage). Highest pH during cultivation, highest cell density and longest cell length was associated with best storage-stability even among collapsed batches (but lowest FT-assay and medium freeze-dry survival). However, the treatment which gave medium length cells and medium cell density was least stable (although this was not statistically significant), so cell length and cell density were probably not the direct (or not the only) causes. Directly, high pH during cultivation seemed to be very beneficial as well as high temperature to a lesser extent. Furthermore, the most storage-stable treatment was also the only one lacking cell clumps, a factor worth further investigation. A combination of high pH and high temperature would be interesting for future investigations. Gram-stain appearance was inconsistent and correlated only with batch number. Better control procedures should have been applied during the Gram-staining.

There are two possible future scenarios: if the initial survival and storage-stability depend on exactly the same molecular factors and are confounded, then the equation proposed in the present study could be used as a predictor of product stability under the same experimental conditions. In the second scenario, if the tolerance to freezing can be improved independently of the accelerated storage-stability, for example by changing formulation, controlling cell density or changing the freeze-drying process, then the correlation formula would become of low practical usefulness, except for shedding some light on the complex dependencies. If the second scenario is true, the FT-assay could be further investigated for testing the freeze-drying tolerance while another assay must be found which correlates independently with the storage-stability. Furthermore, future experiments could

benefit from supplementing the MRS-medium with sucrose and fructooligosaccharides could be worth trying as the formulant.

8.2 GASP experiment conclusions

In conclusion, it seems unlikely that IMC is a feasible option for GASP detection, whether it is due to no possibility of agitation or to the cancelling out of endothermic and exothermic heat flows. To make sure however, a repetition of the anaerobic, non-agitated *E. coli* cultivations with VCs during long-term stationary phase could be done. If these cultivations maintain a GASP-like level of CFU/mL, then GASP detection with IMC can probably be ruled out. However, there are IMC systems with the possibility of stirring (Maskow *et al.*, 2014).

References

Main references

A Rule of Thumb for Unequal Variances | University of Virginia Library Research Data Services + Sciences. [online]. Available from: http://data.library.virginia.edu/a-rule-of-thumb-for-unequal-variances/. [Accessed 2017-03-18].

Albadran, H. A., Chatzifragkou, A., Khutoryanskiy, V. V. & Charalampopoulos, D. (2015). Stability of probiotic Lactobacillus plantarum in dry microcapsules under accelerated storage conditions. *Food Research International*, 74, pp 208–216.

Altman, N. & Krzywinski, M. (2015). Points of Significance: Simple linear regression. *Nature Methods*, 12(11), pp 999–1000.

Armstrong, D. W., Nome, F., Fendler, J. H. & Nagyvary, J. (1977). Novel prebiotic systems: Nucleotide oligomerization in surfactant entrapped water pools. *Journal of Molecular Evolution*, 9(3), pp 213–223.

Bačun-Družina, V., Butorac, A., Mrvčić, J., Landeka Dragičević, T. & Stehlik-Tomas, V. (2011). Bacterial Stationary-Phase Evolution. *Faculty of Food Technology and Biotechnology*, 49(1), pp 13–23.

Béal, C. & Fonseca, F. (2015). Freezing of Probiotic Bacteria. In: Foerst, P. & Santivarangkna, C. (Eds) *Advances in Probiotic Technology*. pp 179–212. CRC Press. ISBN 978-1-4987-3453-0.

Braissant, O., Bachmann, A. & Bonkat, G. (2015). Microcalorimetric assays for measuring cell growth and metabolic activity: methodology and applications. *Methods (San Diego, Calif.)*, 76, pp 27–34.

Brennan, M., Wanismail, B., Johnson, M. C. & Ray, B. (1986). Cellular Damage in Dried Lactobacillus acidophilus. *Journal of Food Protection*, 49(1), pp 47–53.

Bruno, J. C. & Freitag, N. E. (2011). Listeria monocytogenes adapts to long term stationary phase survival without compromising bacterial virulence. *FEMS microbiology letters*, 323(2), pp 171–179.

Clarke, A., Morris, G. J., Fonseca, F., Murray, B. J., Acton, E. & Price, H. C. (2013). A Low Temperature Limit for Life on Earth. *PLOS ONE*, 8(6), p e66207.

Clinical studies and meta-analyses | BioGaia. Available from: https://www.biogaia.com/research/clinical-studies/. [Accessed 2017-04-07].

Colón-González, M., Méndez-Ortiz, M. M. & Membrillo-Hernández, J. (2004). Anaerobic growth does not support biofilm formation in Escherichia coli K-12. *Research in Microbiology*, 155(7), pp 514–521.

Cumulative distribution function of the studentized range (for Tukey's HSD test) - File Exchange - MATLAB Central. [online]. Available from: http://se.mathworks.com/matlabcentral/fileexchange/37450-cumulative-distribution-function-of-the-studentized-range--for-tukey-s-hsd-test-. [Accessed 2017-03-18].

Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), pp 175–191.

Fay, D. S. (2013). A biologist's guide to statistical thinking and analysis. *WormBook*, pp 1–54.

Finkel, S. E. (2006). Long-term survival during stationary phase: evolution and the GASP phenotype. *Nature Reviews Microbiology*, 4(2), pp 113–120.

Finkel, S. E. & Kolter, R. (1999). Evolution of Microbial Diversity during Prolonged Starvation. *Proceedings of the National Academy of Sciences of the United States of America*, 96(7), pp 4023–4027.

Fonseca, F., Béal, C. & Corrieu, G. (2000). Method of quantifying the loss of acidification activity of lactic acid starters during freezing and frozen storage. *Journal of Dairy Research*, 67(1), pp 83–90.

Fonseca, F., Cenard, S. & Passot, S. (2015). Freeze-Drying of Lactic Acid Bacteria. In: Wolkers, W. F. & Oldenhof, H. (Eds) *Cryopreservation and Freeze-Drying Protocols*. pp 477–488. New York, NY: Springer New York. ISBN 978-1-4939-2192-8.

Fonseca, F., Passot, S., Cunin, O. & Marin, M. (2004a). Collapse temperature of freeze-dried Lactobacillus bulgaricus suspensions and protective media. *Biotechnology Progress*, 20(1), pp 229–238.

Fonseca, F., Passot, S., Lieben, P. & Marin, M. (2004b). Collapse temperature of bacterial suspensions: the effect of cell type and concentration. *Cryoletters*, 25(6), pp 425–434.

Franks, F. (1994). Accelerated stability testing of bioproducts: Attractions and pitfalls. *Trends in Biotechnology*, 12(4), pp 114–117.

Gupta, S. C. & Jones, B. (1983). Equireplicate balanced block designs with unequal block sizes. *Biometrika*, 70(2), pp 433–440.

Hammes, W. P. & Hertel, C. (2006). The Genera Lactobacillus and Carnobacterium. In: Dr, M. D. P., Falkow, S., Rosenberg, E., Schleifer, K.-H., & Stackebrandt, E. (Eds) *The Prokaryotes*. pp 320–403. Springer US. ISBN 978-0-387-25494-4.

Help - IBM SPSS Statistics. [online]. Available from: http://127.0.0.1:55902/help/index.jsp?topic=%2Fcom.ibm.spss.statistics.help%2Fspss%2Fbase%2Fovervw_auto_0.htm. [Accessed 2017-03-27].

Higl, B., Kurtmann, L., Carlsen, C. U., Ratjen, J., Först, P., Skibsted, L. H., Kulozik, U. & Risbo, J. (2007). Impact of Water Activity, Temperature, and Physical State on the Storage Stability of Lactobacillus paracasei ssp. paracasei Freeze-Dried in a Lactose Matrix. *Biotechnology Progress*, 23(4), pp 794–800.

IBM Does SPSS offer Tukey-Kramer post-hoc tests? - Sverige. [online] (2016-09-07). Available from: http://www.ibm.com/support, https://www-304.ibm.com/support/docview.wss?uid=swg21478616. [Accessed 2017-03-19].

IBM Knowledge Center - Display Statistics. [online]. Available from: https://www.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.modeler.help/dataaudit_displaystatistics.htm. [Accessed 2017-03-18].

IBM SPSS SamplePower - Get the right sample size the first time. [online] (2012-08-14). Available from: https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=YTD03015GBEN&appname=skmwww. [Accessed 2017-03-19].

Jennings, T. A. (1999). *Lyophilization: introduction and basic principles*. Englewood. CO: Interpharm Press. ISBN 978-1-57491-081-0.

Jett, B. D., Hatter, K. L., Huycke, M. M. & Gilmore, M. S. (1997). Simplified agar plate method for quantifying viable bacteria. *BioTechniques*, 23(4), pp 648–650.

John, P. W. M. (1998). *Statistical design and analysis of experiments*. Philadelphia: Society for Industrial and Applied Mathematics. (Classics in applied mathematics; 22). ISBN 978-0-89871-427-2.

Koch, S., Eugster-Meier, E., Oberson, G., Meile, L. & Lacroix, C. (2008). Effects of strains and growth conditions on autolytic activity and survival to freezing and lyophilization of Lactobacillus delbrueckii ssp. lactis isolated from cheese. *International Dairy Journal*, 18(2), pp 187–196.

Krzywinski, M. & Altman, N. (2014a). Points of significance: Analysis of variance and blocking. *Nature Methods*, 11(7), pp 699–700.

Krzywinski, M. & Altman, N. (2014b). Points of significance: Designing comparative experiments. *Nature Methods*, 11(6), pp 597–598.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology* [online], 4. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3840331/. [Accessed 2017-03-25].

Langsrud, Ø. (2003). ANOVA for unbalanced data: Use Type II instead of Type III sums of squares. *Statistics and Computing*, 13(2), pp 163–167.

Liu, X. T., Hou, C. L., Zhang, J., Zeng, X. F. & Qiao, S. Y. (2014). Fermentation conditions influence the fatty acid composition of the membranes of *Lactobacillus reuteri* I5007 and its survival following freeze-drying. *Letters in Applied Microbiology*, 59(4), pp 398–403.

Losen, M., Frölich, B., Pohl, M. & Büchs, J. (2004). Effect of oxygen limitation and medium composition on Escherichia coli fermentation in shake-flask cultures. *Biotechnology Progress*, 20(4), pp 1062–1068.

Madigan, M. T. & Brock, T. D. (Eds) (2009). *Brock biology of microorganisms*. 12. ed., internat. ed. San Francisco, Calif.: Pearson/Benjamin Cummings. ISBN 978-0-321-53615-0.

Maher, J. M., Markey, J. C. & Ebert-May, D. (2013). The Other Half of the Story: Effect Size Analysis in Quantitative Research. *CBE Life Sciences Education*, 12(3), pp 345–351.

Maskow, T., Morais, F. M., Rosa, L. F. M., Qian, Y. G. & Harnisch, F. (2014). Insufficient oxygen diffusion leads to distortions of microbial growth parameters assessed by isothermal microcalorimetry. 4(62), pp 32730–32737.

Meng, X. C., Stanton, C., Fitzgerald, G. F., Daly, C. & Ross, R. P. (2008). Anhydrobiotics: The challenges of drying probiotic cultures. *Food Chemistry*, 106(4), pp 1406–1416.

Miller, J. N. & Miller, J. C. (2010). *Statistics and chemometrics for analytical chemistry*. 6. ed. Harlow: Prentice Hall. ISBN 978-0-273-73042-2.

Montgomery, D. C. (1991). *Design and analysis of experiments*. 3rd ed. New York: Wiley. ISBN 978-0-471-52000-9.

Pallin, A., Sveriges lantbruksuniversitet & Institutionen för mikrobiologi (2015). *Fermentation of barley flour with Lactobacillus reuteri: a source of bioactive compounds against a leaky gut?*. Diss. Uppsala: Department of Microbiology, Swedish University of Agricultural Sciences.

Palmfeldt, J. & Hahn-Hägerdal, B. (2000). Influence of culture pH on survival of Lactobacillus reuteri subjected to freeze-drying. *International Journal of Food Microbiology*, 55(1–3), pp 235–238.

RTS-1C, Personal bioreactor — Biosan. [online]. Available from: http://www.biosan.lv/en/products/bioprocessing-technologies/personal-bioreactors/rts-1c-real-time-cell-growth-logger. [Accessed 2017-03-09].

Saulnier, D. M., Santos, F., Roos, S., Mistretta, T.-A., Spinler, J. K., Molenaar, D., Teusink, B. & Versalovic, J. (2011). Exploring Metabolic Pathway Reconstruction and Genome-Wide Expression Profiling in Lactobacillus reuteri to Define Functional Probiotic Features. *PLOS ONE*, 6(4), p e18783.

Schoug, å., Mahlin, D., Jonson, M. & Håkansson, S. (2010). Differential effects of polymers PVP90 and Ficoll400 on storage stability and viability of *Lactobacillus coryniformis* Si3 freeze-dried in sucrose. *Journal of Applied Microbiology*, 108(3), pp 1032–1040.

Schoug, Å., Fischer, J., Heipieper, H. J., Schnürer, J. & Håkansson, S. (2008). Impact of fermentation pH and temperature on freeze-drying survival and membrane lipid composition of Lactobacillus coryniformis Si3. *Journal of Industrial Microbiology & Biotechnology*, 35(3), pp 175–181.

Schoug, Å., Olsson, J., Carlfors, J., Schnürer, J. & Håkansson, S. (2006). Freeze-drying of Lactobacillus coryniformis Si3—effects of sucrose concentration, cell density, and freezing rate on cell survival and thermophysical properties. *Cryobiology*, 53(1), pp 119–127.

Schoug & Åsa (2009). *A dry phase of life: freeze-drying and storage stability of Lactobacillus coryniformis Si3 in sucrose-based formulations*. Diss. Uppsala: Swedish University of Agricultural Sciences.

Schwab, C., Vogel, R. & Gänzle, M. G. (2007). Influence of oligosaccharides on the viability and membrane properties of Lactobacillus reuteri TMW1.106 during freeze-drying. *Cryobiology*, 55(2), pp 108–114.

Sewell, D., Laird, K. & Phillips, C. (2011). Growth advantage in stationary phase phenomenon in Gram-positive bacteria. *Journal of Hospital Infection*, 78(1), pp 73–75.

Sezonov, G., Joseleau-Petit, D. & D'Ari, R. (2007). Escherichia coli Physiology in Luria-Bertani Broth. *Journal of Bacteriology*, 189(23), pp 8746–8749.

Smeulders, M. J., Keer, J., Speight, R. A. & Williams, H. D. (1999). Adaptation of Mycobacterium smegmatis to Stationary Phase. *Journal of Bacteriology*, 181(1), pp 270–283.

Smith, A. C. & Hussey, M. A. *Gram Stain Protocols*. [online] (2005-09-30). Available from: http://asmscience.org/content/education/protocol/protocol.2886. [Accessed 2017-02-15].

Sturtevant, J. M. (1953). Heats of Hydrolysis of Amide and Peptide Bonds. *Journal of the American Chemical Society*, 75(8), pp 2016–2017.

Tymczyszyn, E. e., Gómez-Zavaglia, A. & Disalvo, E. a. (2007). Effect of sugars and growth media on the dehydration of Lactobacillus delbrueckii ssp. bulgaricus. *Journal of Applied Microbiology*, 102(3), pp 845–851.

*Universität Düsseldorf: G*Power*. [online]. Available from: http://www.gpower.hhu.de/en.html. [Accessed 2017-03-16].

Vasi, F. K. & Lenski, R. E. (1999). Ecological strategies and fitness tradeoffs inEscherichia coli mutants adapted to prolonged starvation. *Journal of Genetics*, 78(1), pp 43–49.

Vaux, D. L., Fidler, F. & Cumming, G. (2012). Replicates and repeats—what is the difference and is it significant? *EMBO Reports*, 13(4), pp 291–296.

Velly, H., Fonseca, F., Passot, S., Delacroix-Buchet, A. & Bouix, M. (2014). Cell growth and resistance of Lactococcus lactis subsp. lactisTOMSC161 following freezing, drying and freeze-dried storage are differentially affected by fermentation conditions. *Journal of Applied Microbiology*, 117(3), pp 729–740.

Wilk, M. B. & Gnanadesikan, R. (1968). Probability Plotting Methods for the Analysis of Data. *Biometrika*, 55(1), pp 1–17.

Yao, A. A., Coulibaly, I., Lognay, G., Fauconnier, M.-L. & Thonart, P. (2008). Impact of polyunsaturated fatty acid degradation on survival and acidification activity of freeze-dried Weissella paramesenteroides LC11 during storage. *Applied Microbiology and Biotechnology*, 79(6), pp 1045–1052.

Zambrano, M. M. & Kolter, R. (1996). GASPing for Life in Stationary Phase. *Cell*, 86(2), pp 181–184.

Zambrano, M. M., Siegele, D. A., Almirón, M., Tormo, A. & Kolter, R. (1993). Microbial competition: Escherichia coli mutants that take over stationary phase cultures. *Science*, 259(5102), pp 1757–1760.

Zayed, G. & Roos, Y. H. (2004). Influence of trehalose and moisture content on survival of Lactobacillus salivarius subjected to freeze-drying and storage. *Process Biochemistry*, 39(9), pp 1081–1086.

Zinser, E. R. & Kolter, R. (2004). Escherichia coli evolution during stationary phase. *Research in Microbiology*, 155(5), pp 328–336 (Genome plasticity and the evolution of microbial genomes).

Årsköld, E., Lohmeier-Vogel, E., Cao, R., Roos, S., Rådström, P. & Niel, E. W. J. van (2008). Phosphoketolase Pathway Dominates in Lactobacillus reuteri ATCC 55730 Containing Dual Pathways for Glycolysis. *Journal of Bacteriology*, 190(1), pp 206–212.

Images used

By Lilly_M - Own work(in cooperation with the school of photography - Fotoedukacja), CC BY-SA 3.0, *Laboratory heating block Techne-03.jpg*
https://commons.wikimedia.org/w/index.php?curid=13200077. [Accessed 2017-06-22].

snowflake-clipart-outline-5673-snowflake-design
http://www.clipartpanda.com/categories/snowflake-clipart-transparent-background. [Accessed 2017-06-22].

Popularized summary

Just as the ecosystems at land and sea are built upon complex interactions of a vast array of species, so are the hidden ecosystems of our own bodies. Science is barely beginning to scratch the surface of the interplay between species in our so called microbiota, but one thing is becoming clear – it has huge effects on our health. It is not surprising that probiotics, supplements of good bacteria, are becoming popular. *Lactobacillus reuteri* is one such bacterium available at the probiotic market and was the subject of this study. For probiotics to have any effects, however, they must stay alive for a long time in the form of tablets and powders. This is where the engineering comes in. One of the methods for preserving bacteria is freeze-drying; an efficient method which allows further storage at ambient temperatures. Paradoxically, freeze-drying can become a lethal process for the bacteria if done arbitrarily. In fact, the whole manufacturing process, starting at cultivating the bacteria in bioreactors and ending at the dry bacterial powder, has to be properly optimized for each individual bacterial species if high survival rates and good storage-stability are to be achieved. There are different ways to optimize the freeze-drying survival and storage-stability of bacteria. One strategy, in focus of this study, is "preconditioning" done by subjecting the bacteria to different environmental conditions during or after the cultivation, which alters their molecular expression and morphology. Another strategy can be to search for more robust mutants of the bacterium. A discovery made by other researchers related to this concern is that increasingly resilient mutants can be obtained from some bacterial species by over-fermenting the cultures for a long time. These mutants, referred to as "growth advantage in stationary phase mutants" (GASPs), take over the culture by scavenging the remains of their ancestors while themselves tolerating the scarce and polluted environment of an overgrown cultivation. Therefore, it was of interest whether such mutants could also be observed in *L. reuteri*.

Since the overall goal of this study was to work towards finding the recipe for a more freeze-dry tolerant and storage-stable *L. reuteri*, both of the abovementioned strategies were applied. Firstly, it was investigated whether a quick assay could be

used to predict the freeze-drying survival and storage-stability instead of measuring this directly by laborious viability measurements over many weeks, and the candidate of choice was a freeze-thaw assay. In the freeze-thaw assay, bacteria are subjected to several cycles of freezing and thawing after which their survival rates are measured. The survival of the freeze-thaw assay was then compared to the survival of conventional storage-stability measurements. The bacteria for this experiment were cultivated at three different combinations of pH and temperature; therefore the experiment would simultaneously shed some light onto which cultivation conditions lead to the highest freeze-drying survival and storage-stability in general. Secondly, microscopic studies were applied to the abovementioned experiment to see if the morphology of the cells could be linked with the freeze-drying survival and storage-stability. Finally, as a strategy towards finding more robust *L. reuteri* GASP mutants, batch cultivations were grown for a prolonged time in isothermal microcalorimeters. These very sensitive pieces of equipment measure the heat given off by various processes, down to a few microwatts. It was hypothesized that GASPs would be detectable by the heat released through their metabolic activity, since consecutive generations of GASPs mutants appear in a characteristic wave-like manner.

The results of this study showed that higher pH, and to lesser extent higher temperature, gave the best storage-stability. The differences in survival of the freeze-drying process itself were not statistically significant because there was a large spread in the data at this point of measurement. A correlation between storage-stability and the freeze-thaw assay was found and it was inverse; a high storage-stability meant low survival of the freeze-thaw assay and vice versa. There was no consistent link between morphology and storage-stability, although interestingly, at highest pH the cells did not form any clumps in contrast to the other treatments. These results imply that the observed differences in freeze-thaw survival and storage-stability are caused by molecular changes in the cells due to the different cultivation conditions; something worth deeper studies in the future. When it comes to the quest for GASP mutants, no activity of interest could be observed; this either means that isothermal microcalorimetry is not a feasible tool for detecting GASPs or that *L. reuteri* (and *E. coli* controls) did not produce any GASP mutants. Overall, this study has hopefully lit up a path towards understanding the phenomena which govern the fitness for freeze-drying preservation of *L. reuteri*.

Acknowledgements

To my supervisor Sebastian Håkansson and my examiner Mikael Pell, thank you both for accepting me for doing this thesis. It was always fun brainstorming and making plans with you Sebastian and learning the ways of lyophilization!

Huge thanks to Armando Hernández Garcia, my co-supervisor for teaching me how the equipment and things in general work in the labs, spending time for discussing ideas and sharing helpful insights from your ongoing research. Your companionship in the labs where we spent many hours together would always lift the mood. Not to mention the times where you changed your own research plans to let me have the equipment first so I could get my replicates!

Thanks to Hans Jonsson for arranging the "break in" to the course labs with me where I could do the Gram-staining! Couldn't have done it without you. You spend much of your time for preparing everything I needed down to the protocol with such engagement, it was fun.

Thanks to Stefan Roos from BioGaia AB who participated in outlining the focus of our experiments and kindly introduced me to the *L. reuteri* species and BioGaia's work.

Appendix

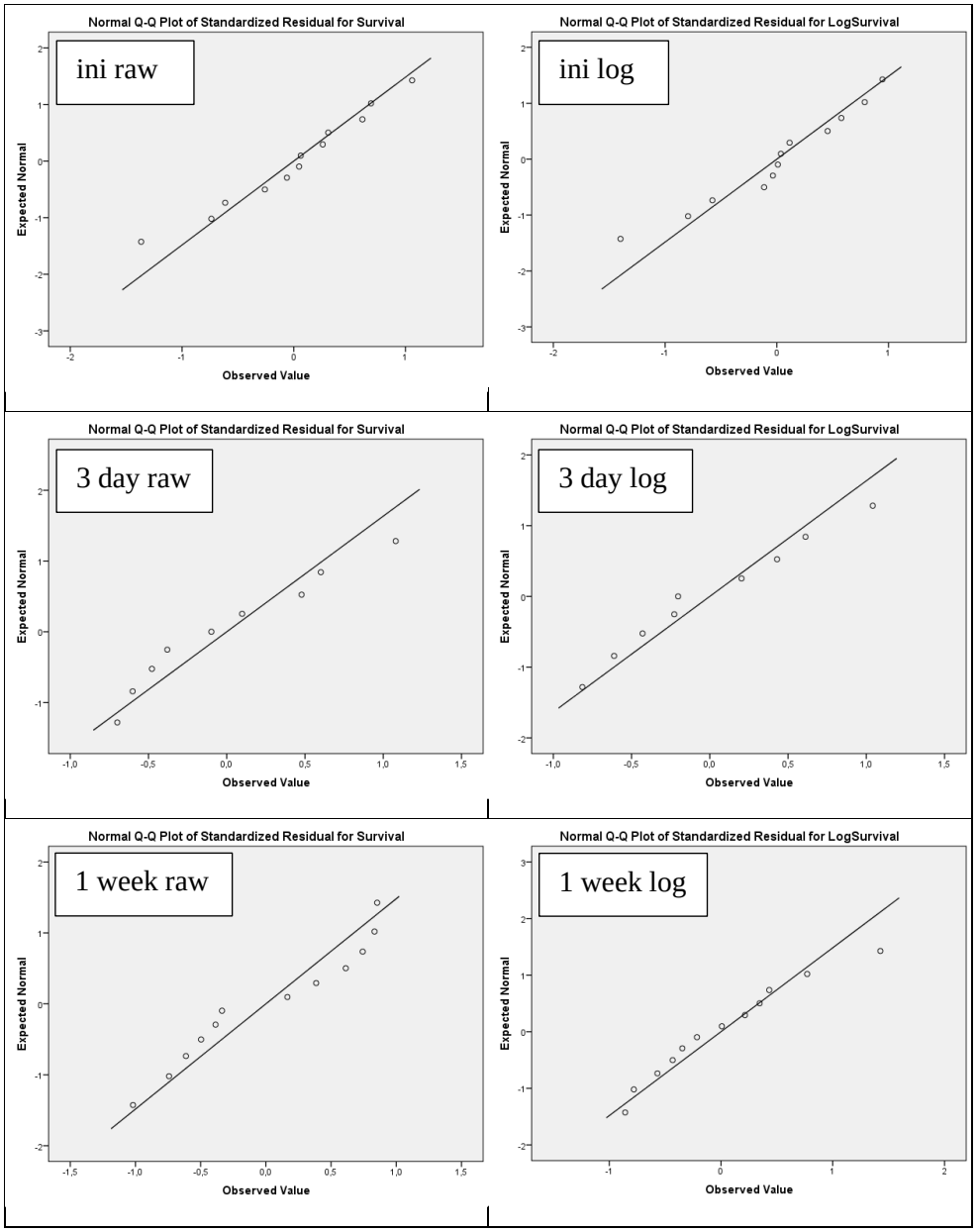

Figure 23. Normal Q-Q plots of expected residuals for normal distribution versus observed standardized residuals, for evaluation of the normality assumption of ANOVAs in the accelerated storage assay. Plots based on non-transformed (raw) and $\log_{10}$-transformed data are compared side by side. Each plot contains the combined residuals of all the biological replicates for all treatments and batches at the given timepoint of accelerated storage (±1day), starting at the initial survival after freeze-drying ("ini") and ending at 1 week of storage.

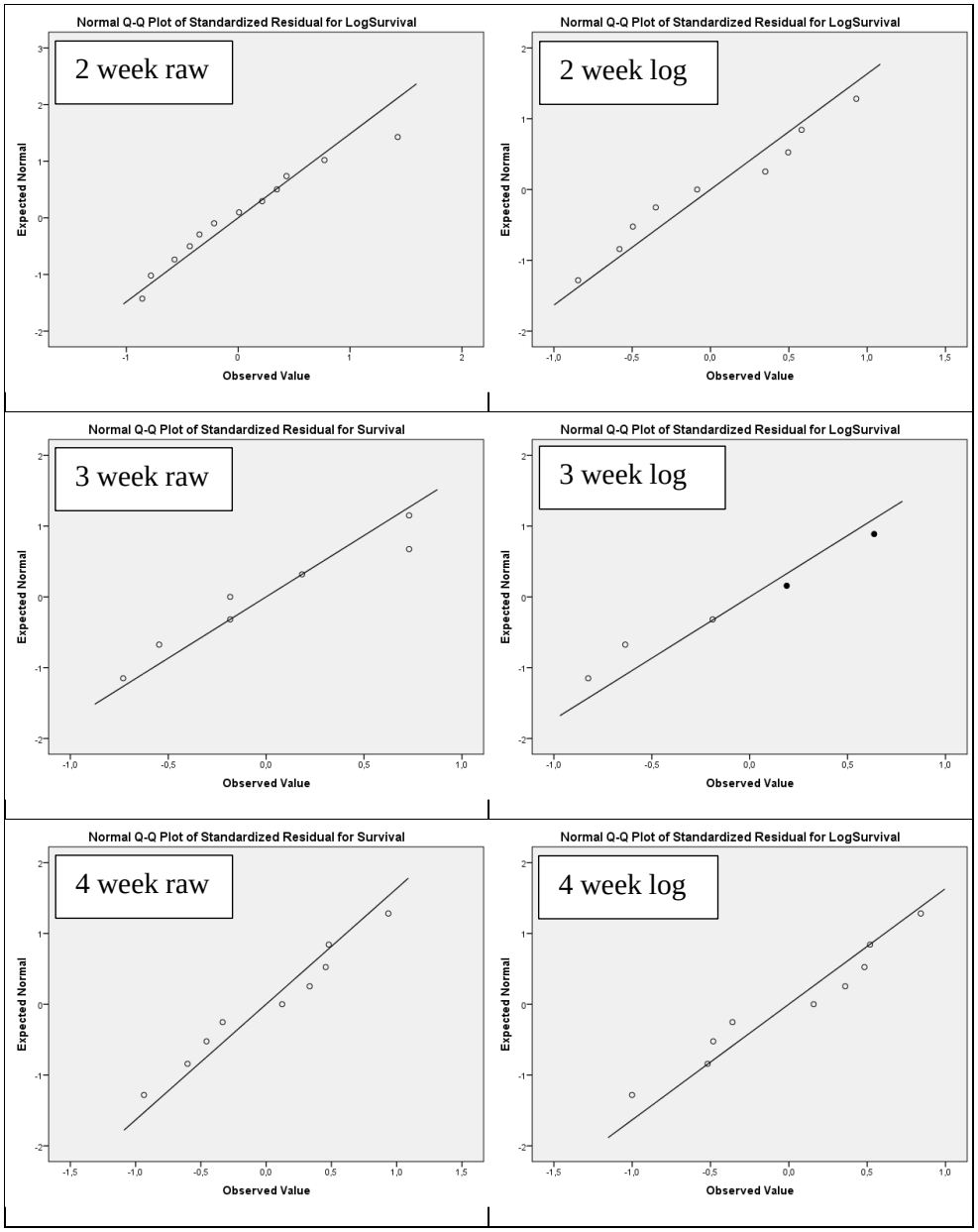

Figure 24. Normal Q-Q plots of expected residuals for normal distribution versus observed standardized residuals, for evaluation of the normality assumption of ANOVAs in the accelerated storage assay. Plots based on non-transformed (raw) and $\log_{10}$-transformed data are compared side by side. Each plot contains the combined residuals of all the biological replicates for all treatments and batches at the given timepoint of accelerated storage (±1day), in this figure starting at 2 weeks of; and ending at 4 week of storage.

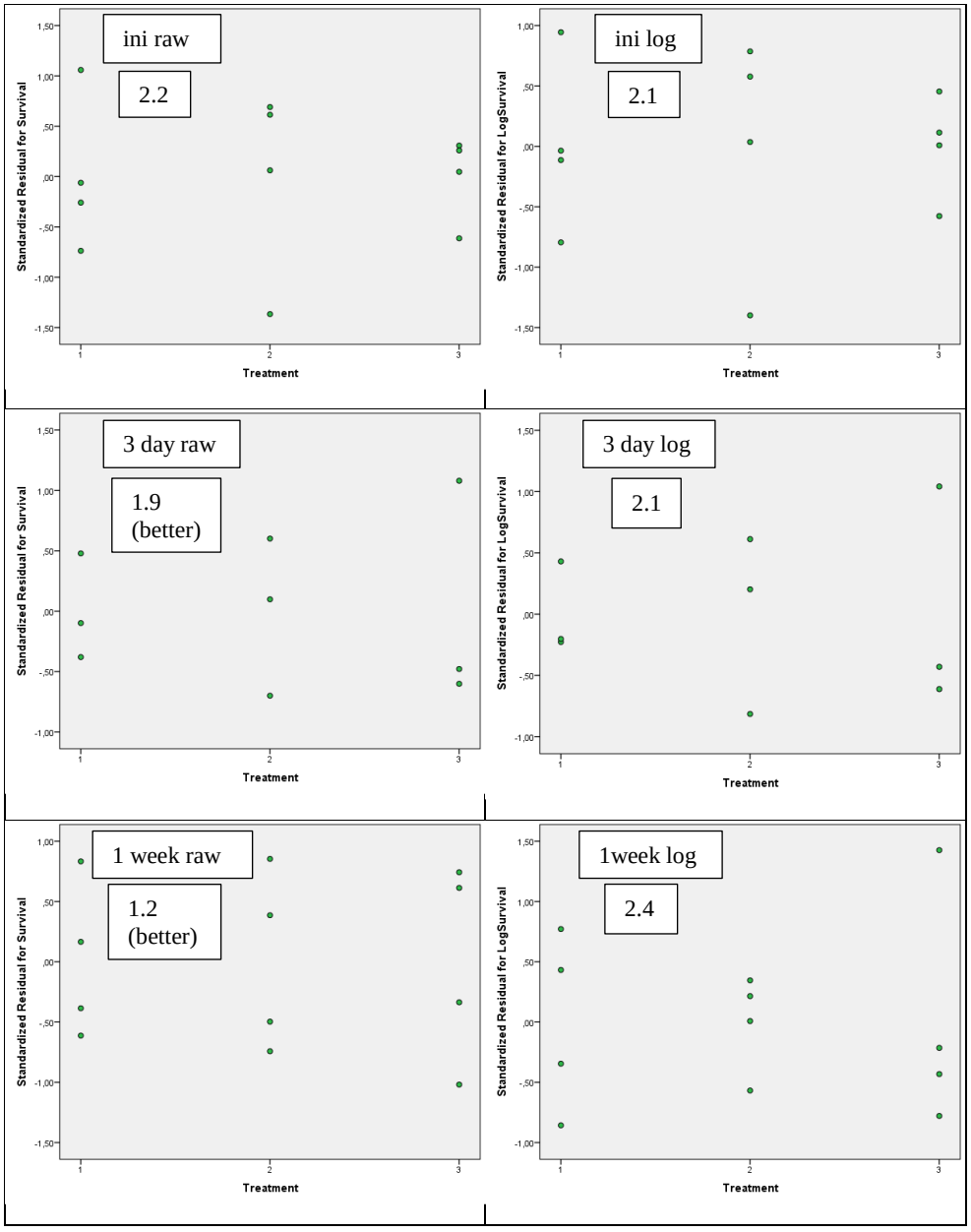

Figure 25. Standardized residuals versus treatments, where treatment nr. 1 = 37 °C, pH 4.5; 2 = 32 °C, pH 4.5; 3 = 32 °C, pH 6.5. From accelerated storage assay; timepoints (±1day) of storage starting at initial survival after freeze-dry ("ini") to 1 week of storage. Comparison of non-transformed (raw) and log_{10}-transformed data, respectively. Ratio of widest variance over narrowest variance is shown in the textboxes.

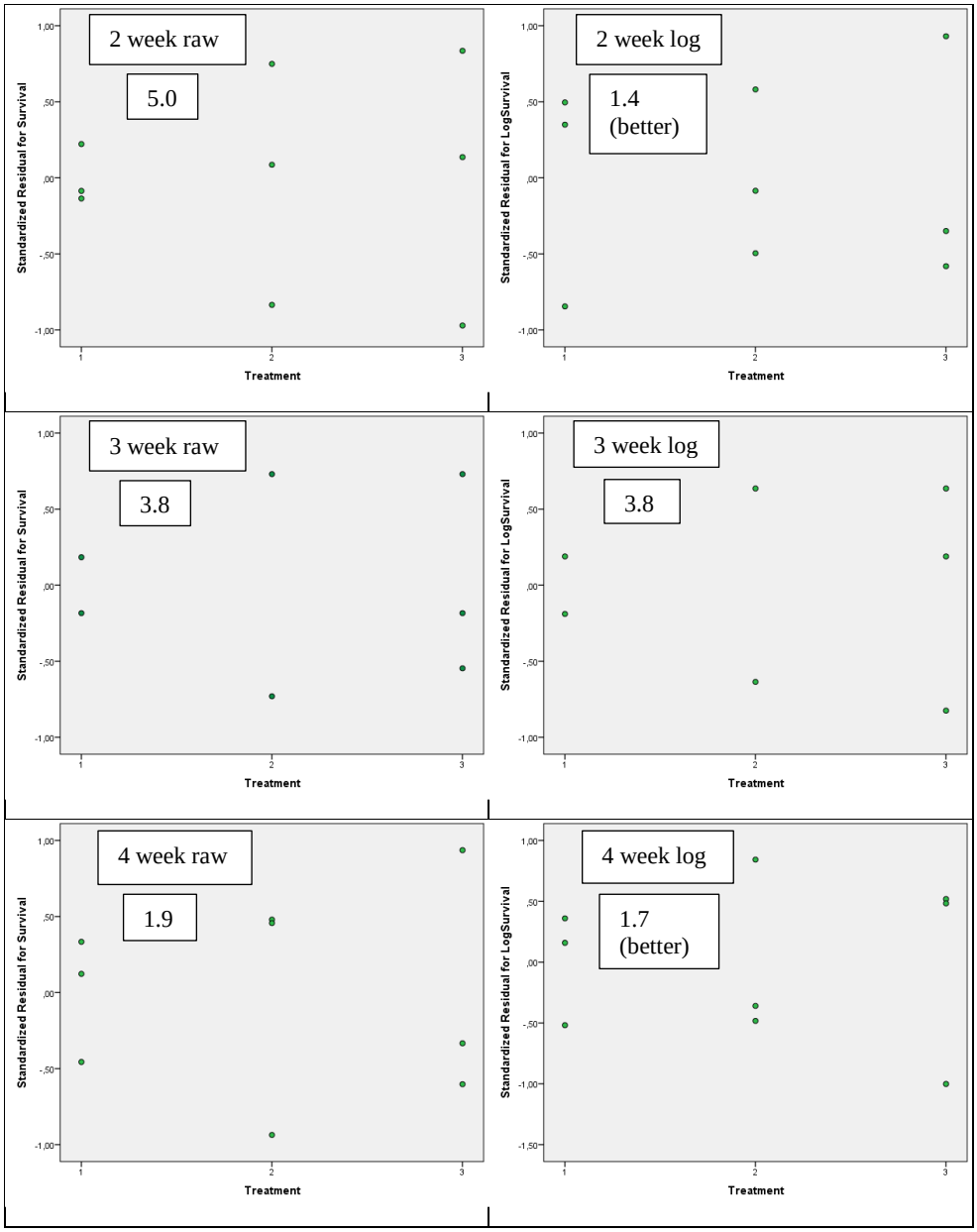

Figure 26. Standardized residuals versus treatments, where treatment nr. 1 = 37 °C, pH 4.5; 2 = 32 °C, pH 4.5; 3 = 32 °C, pH 6.5. From accelerated storage assay; timepoints (±1day) starting at 2 weeks of storage and ending at 4 weeks. Comparison of non-transformed (raw) and log_{10}-transformed data, respectively. Ratio of widest variance over narrowest variance is shown in the textboxes.

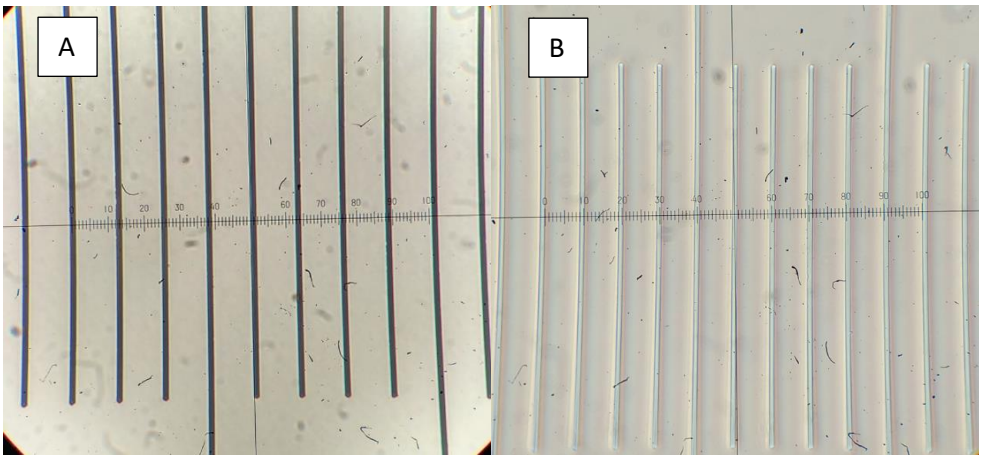

Figure 27. A = home lab microscope reference. B = course lab microscope reference. The big scale is the objective micrometer (10 µm between ticks) and the small scale is the ocular micrometer. Ocular micrometer was calculated to represent 0.79 µm between ticks at the home lab microscope and 0.99 µm between ticks at the course lab microscope.

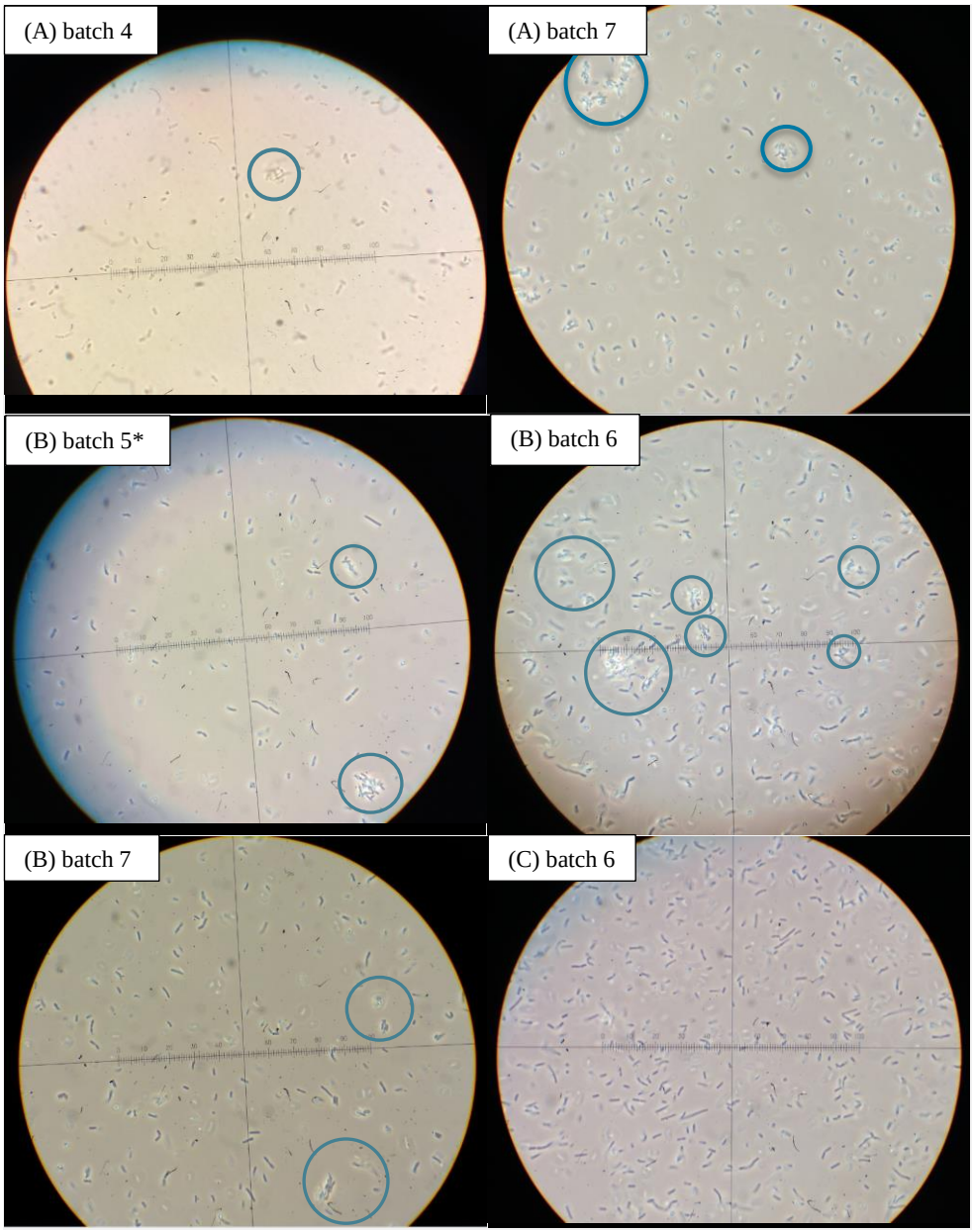

Figure 28. Cultures of *L. reuteri* grown at three different fermentation treatments (A-C), examined at around harvest time. The blue rings help to spot cell clumps. Objective micrometers show the scale which is recalculated into µm using Figure 27 in Appendix. Viewed in phase-contrast microscopes at 1000X magnification. See Table 11 for which microscope was used for each photo. Picture index: (A) 37 °C, pH 4.5; (B) 32 °C, pH 4.5; (C) 32 °C, pH 6.5. *Another photo of the sample in Figure 18 B, here showing that very long cells were, albeit very rarely, found in this sample.

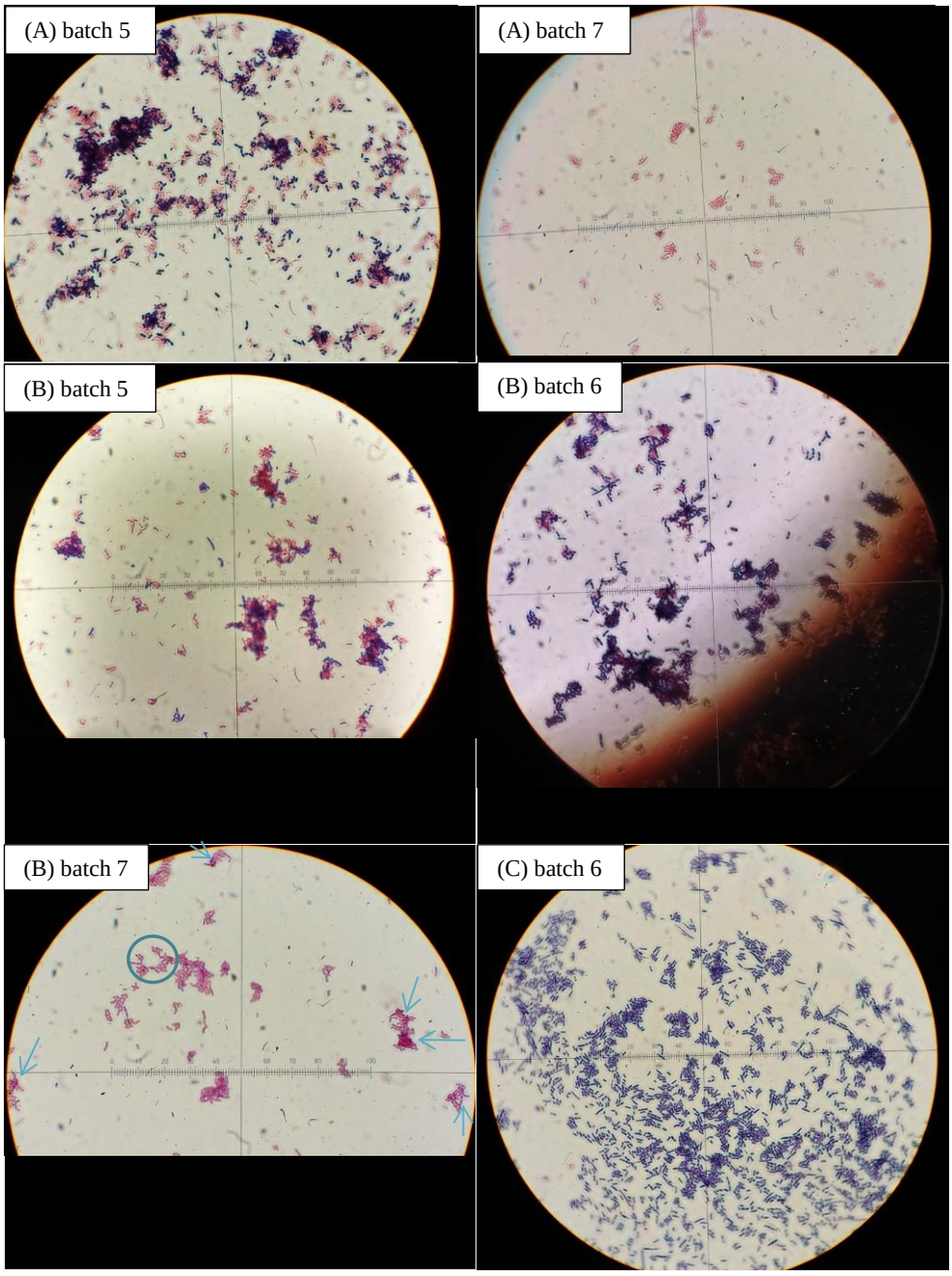

Figure 29. Gram-stain of *L. reuteri* grown at three different fermentation treatments (A-C). Viewed in phase-contrast microscope at 1000X magnification. Objective micrometers show the scale which is recalculated into µm using Figure 27 B in Appendix (all photos are from the course lab microscope). The blue arrows and a ring in the bottom left photo spot small groupings of lightly blue-tinted cells (only some cells in the ring). Picture index: (A) 37 °C, pH 4.5; (B) 32 °C, pH 4.5; (C) 32 °C, pH 6.5.