Evaluation of a Genome Editing Approach in Lactic Acid Bacteria Based on the CRISPR/Cas9 System

Inês da Costa Gonçalves

Thesis to obtain the Master of Science Degree in Biotechnology

Biotechnology

Supervisor: Prof. Gabriel António Amaro Monteiro

Examination Committee

Chairperson: Prof. Arsénio do Carmo Sales Mendes Fialho

Supervisor: Prof. Gabriel António Amaro Monteiro

Member of the Committee: Prof. Leonilde de Fátima Morais Moreira

December 2017

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Gabriel Monteiro, for the guidance, availability and support during the development of this master thesis. For his ideas and perseverance has helped me see science as an everlasting improving subject in which we are always improving, not only the experiment designs, but ourselves. I would also like to thank Professor Leonilde Moreira for providing us with *E. coli* strains, and Inês Silva, for analysing and providing us the sequencing results for the desired gene regions.

I would also like to give big thanks to PhD student Sofia Duarte, not only for the laboratory guidance, patience and dedication, but also, for all the warming friendship. It was great to work in such a good environment, mostly due to the energy and positivist that she brings to the laboratory. I am also thankful to all the laboratory techniques and practises that she taught me, which, I believe, are key elements in maintaining good laboratory practises and atmospheres. Furthermore, I would also like to thank the other "Lactic Acid Bacteria Gang" member, Carolina Gonçalves, which has become a great friend, for all the help and shared work, for the discussions, and for making the long hours that we spent in the laboratory much more enjoyable.

I would also like to acknowledge my other laboratory colleagues for sharing frustrations and joys and for the enjoyable days in the laboratory. A special thanks to the PhD student Cláudia Alves, for the critical spirit, warm personality and the availability to solve problems and intelligently discuss and suggest ideas to further improve the development of this thesis and the laboratory knowledge. I also thankfully recognise the vast amount of work performed by technician Ricardo Pereira and Dona Rosa, for constantly maintaining the laboratory work homeostasis.

I am extremely grateful for the unconditional love and support of my parents and brother, who always believed in me and my capabilities and for not letting me forget that the obstacles are also great opportunities to revise our means, learn and improve ourselves. Thank you for not letting me forget to have a broad mind and reminding me that there is always a positive side to things. I also acknowledge the remaining members of my family, for pulling weigh out of my (stressful) shoulders by making me see things in a broader perspective.

A special thanks to Manuel Ramos, for the unconditional support, friendship, cheer up and worship, when I mostly needed. To the always present Sara Costa and Margarida Brotas for the true words and purest friendship. And to all my dearest Biotechnology colleagues for the incredible symbiosis that we have created, making this two years unforgettable. Special thanks to João Cotovio, André Costa, Mónica Galocha, Carolina Gonçalves, Carlota Miranda, Mariana Jerónimo, Filipe Bica, Pedro Silva and Vera Esgueira, for all the love, fun moments and working hours, for having overcome obstacles together and for letting me grow alongside incredible people.

Abstract

DNA vaccination based on the delivery of plasmid DNA is a promising tool to treat, autoimmune, infectious and genetic-related diseases. The current favourite host for plasmid production is *E. coli*. However, the majority of pDNA extraction methods co-purify lipopolysaccharides (LPS) present in *E. coli* cell wall, which generate inflammatory responses in mammal hosts. Lactic acid bacteria (LAB) are classified as GRAS and possess probiotic features, being used in human health and nutrition from ancient years. Due to its beneficial characteristics and lack of LPS production, the use of LAB as microbial cell factories for pDNA production is proposed. Hurdles associated with LAB for this mean include the lack of plasmids and features that allows them to do so. Therefore, the present work aims to genetically modify *Lactococcus lactis* LMG 19460, using the CRISPR/Cas9 system, to diminish harsh host terms for pDNA production. Applications of this improved strain include industrial production of pDNA and DNA delivery for gene therapy purposes. The previously designed nthCRISPRa plasmid, modified from pKCcas9dO, was unsuccessfully transformed by electroporation in *L. lactis* LMG 19460 to attempt *nth* knockout in a single step and a PCR reaction using 250,000 cells was developed for the assessment of both genome edition and nthCRISPRa presence. The reconstruction of nthCRISPRa into nthCRISPRe was attempted by Gibson assembly to change the selection marker from *apmR* to *eryR*. The CRISPR/Cas9 based editing plasmid containing the erythromycin resistant gene is yet to be achieved. Theoretical predictions were made for future advantageous gene knockouts. Furthermore, the OptFlux software was used to predict non-obvious knockouts that should increase DNA production.

Key Words Lactic Acid Bacteria, Gene Therapy, plasmid DNA, CRISPR/Cas9, PCR, Gibson Assembly.

Resumo

A vacinação de DNA por entrega de DNA plasmídico (pDNA) é uma ferramenta promissora no tratamento de doenças autoimunes, infeciosas e genéticas. Atualmente, *E. coli* é o hospedeiro predileto para a produção de plasmídeo. No entanto, a maioria dos métodos de extração de pDNA co-purificam lipopolissacáridos (LPS) presentes na parede celular, desencadeando respostas inflamatórias em hospedeiros. As bactérias ácido lácticas (BAL) são organismos GRAS e não possuem LPS, sendo usadas em nutrição e saúde humana há muitos anos. É então proposta a utilização das BAL como fábricas celulares microbianas para a produção de pDNA. Obstáculos associados ao uso das BAL para este fim incluem a falta de plasmídeos e características que o permitem. O presente trabalho tem por objetivo modificar geneticamente *Lactococcus lactis* LMG 19460, usando o sistema CRISPR/Cas9, de forma a diminuir as condições severas para a produção de pDNA. Aplicações desta estirpe melhorada incluem a produção industrial de pDNA e entrega de DNA no contexto da terapia génica. O plasmídeo nthCRISPRa, modificado a partir do pKCcas9dO, foi, sem sucesso, transformado por electroporação em L. lactis LMG 19460 de forma a incitar a deleção do gene *nth*. Uma reação de PCR usando 250000 células foi desenvolvida para avaliar tanto a ocorrência de edição genómica como a presença do nthCRISPRa. A reconstrução do nthCRISPRa em nthCRISPRe foi tentada por *Gibson assembly* de forma a alterar a marca de seleção apmR para eryR. Predições teóricas foram feitas para futuros knockouts. Ademais, o *software* OptFlux foi usado para prever knockouts não óbvios para o aumento da produção de DNA.

Palavras-chave Bactérias Ácido Lácticas, Terapia Génica, DNA plasmídico, CRISPR/Cas9, PCR, Gibson Assembly.

List of Contents

List of Figures

List of Tables

List of Abbreviations

BLG Beta-Lactoglobulin

bp base pairs

BPCY Biomass-Product Coupled Yield

Cas CRISPR-associated

CD4 Cluster of Differentiation 4

cfu colony forming unit

CRISPR Clustered Regularly Interspaced Short Palindromic Repeats

crRNA CRISPR RNA

CV-N Cianovirin-N

dCas9 deactivated Cas9

DNA Deoxyribonucleic Acid

DSB Double Strand Break

EFSA European Food Safety Authority

EPS Exopolysaccharide

FDA Food and Drug Administration

FnBPA Fibronectin-Binding Protein A

GFP Green Fluorescence Protein

GIT Gastrointestinal Tract

GRAS Generally Recognized as Safe

HDR Homology Directed Repair

HIV Human Immunodeficiency Virus

IBD Inflammatory Bowel Disease

IgG2a Immunoglobulin G2a

IL-6 Interleukin 6

IL1β Interleukin 1 beta

kb kilo-base pair

KO Knockout

LAB Lactic Acid Bacteria

LB Luria-Bertani

LPS Lipopolysaccharide

mRNA Messenger RNA

NGS Next Generation Sequencing

NHEJ Non-Homologous End Joining

OD Optical Density

ORF Open Reading Frame

PAM Proto-Spacer-Adjacent Motifs

P_{cmv} Cytomegalovirus Promoter

PCR Polymerase Chain Reaction

pDNA Plasmid DNA

rLAB Recombinant Lactic Acid Bacteria

RNA Ribonucleic Acid

RNase III Ribonuclease III

SCFA Short-Chain Fatty Acid

scFv single-chain variable Fragment

sgRNA Single Guide RNA

SpyCas9 *Streptococcus pyogenes* Cas9

ssODN single-stranded Oligodeoxynucleotide

TALEN Transcription Activator-Like Effector Nuclease

TNF-α Tumor Necrosis Factor alpha

tracrRNA Trans-Activating crRNA

w/ with

w/o without

ZFN Zinc-Finger Nuclease

HR Homologous Recombination

1 Introduction

1.1 Lactic Acid Bacteria

Lactic acid bacteria (LAB) are a non-taxonomic group of Gram-positive, nonsporing, low GC content, aerotolerant anaerobe and non-motile bacteria, in which are included different species such as *Lactobacillus*, *Lactococcus*, *Pediococcus*, *Streptococcus* and *Leuconostoc*.[1–3] Most of these bacteria are characterized by their capability to convert fermentable carbohydrates, mainly glucose, to lactic acid. LAB that produce lactic acid as their catabolic end-product are called homofermentative LAB. In addition to the production of lactic acid, some of these bacteria also produce ethanol/acetate and carbon dioxide in significant amounts, therefore being called heterofermentative bacteria.[3–6]

These bacteria have a long history of exploitation by humans, being used for centuries in food production and preservation and as probiotic agents to promote human health. Given their long history of food-grade applications and non-pathogenicity, LAB are classified as Generally Recognized As Safe (GRAS) by the United States Food and Drug Administration (FDA) and fulfil criteria of the Qualified Presumption of Safety (QPS) according to the European Food Safety Authority (EFSA).[7–9]

These strictly fermentative bacteria have limited biosynthetic capabilities, especially amino acid, vitamin and nitrogenous bases production, thereby requiring complex growth media. These nutritional constraint limits LAB natural habitats to those in which amino acids, peptides, fatty acids, vitamins, purines, pyrimidines and sugars are available.[6,10,11] LAB niches include foodstuffs, such as milk, wine, meat and plant surfaces, and animal mucosal cavities, such as the gastrointestinal tract, the oral cavity and the vagina. In fact, *lactobacilli* and *streptococci* are highly represented in the human and other higher vertebrates small intestine mucosa.[11] Among naturally producing molecules of economic interest are exopolysaccharides (EPSs) and short chain fatty acids (SCFAs).[10]

Some lactic acid bacteria have a fundamental role in the physiological health maintenance of their host organism, by perpetuating the intestinal homeostasis and by modulating the immune system response.[9,11] Due to their health-promoting features, LAB have been commercially sold and used as probiotics, which consist of “live microorganisms which when administered in proper amounts confer a health benefit on the host”, according to the Food and Agricultural Organization (FAO) of United Nations and the World Health Organization (WHO).[9] Different LAB strains perform their probiotic features in different host organisms using different mechanisms, meaning that LAB-mediated health benefits are strain-and-host-specific. Apart from immune system modulators, lactic acid bacteria have also been related to mucosal barrier function against pathogen invasion, carcinogen detoxification and cholesterol catabolism.[4,9]

Although LAB include several microorganisms, *Lactococcus lactis* has been the most widely used considering cloning and production of recombinant proteins and DNA.[12] It is a mesophilic bacterium, with

ideal growth temperature of 30ºC, mainly used in the dairy industry. *L. lactis* has been deeply characterized given its economic importance, being the first LAB species whose genome was fully sequenced. In addition, its genetic manipulation is relatively easy and different genetic tools, such as cloning and expression systems, have already been developed for this species.[4,7]

1.1.1 LAB Industrial Applications

Lactic acid bacteria have been used as food preservatives as early as the 8th millennium B.C., constituting one of the most ancient practices of food preservation by humans.[4,6] This process takes advantage of lactic fermentation of foods and the latter has also been increasingly used in the past decades for the manufacturing of foodstuffs, mainly dairy. Therefore, LAB are mostly used in the manufacturing of fermented foods, in which are included yogurt, cheese, milk, bread, butter, wine, sausages, fermented meats (salami), fermented vegetables (pickles, sauerkraut and silage) and sourdough bread.[3,5,9] LAB-mediated conservation and manufacturing of foodstuffs is owed to medium acidification to pH comprising 3.5 and 4.5, and to bacteriocins and organic compounds production.[4,10,11] Exceptions to the food preservation utilities in the LAB group exist in some species of the genus *Pediococcus*, which are in fact related to food spoilage. LAB are not only used in the food and feed industries for preservation, but also to modify the organoleptic characteristics of foods, including flavour and texture.[9] Formerly, the fermentation of previously mentioned products would be obtained through spontaneous fermentation, due to the naturally presence of LAB in raw materials. Nowadays, LAB are also used as starter cultures, defined as a preparation of one or more microorganisms which when introduced in raw materials accelerate and steer the fermentation process. This process came in great advantage for fermentation control and standardization of the final product, in addition to the improvement of nutritional, organoleptic and shelf-life characteristics of foods.[9]

Bacteriocins are peptides that possess antimicrobial activity against species closely related to the bacteriocin producing strains. The preservation potential of these substances have been largely studied in the food industry particularly on dairy, egg, vegetable and meat products.[9] In addition to bacteriocins, LAB also produce nisin A, which has proven to be very effective against poisoning and spoilage microbial agents. In fact nisin is used worldwide for food preservation purposes.[9,13] Besides EPS and SCFA production, LAB have also the ability to synthesize few vitamins, such as folate, vitamin B12 and riboflavin; enzymes, as α-amylase and aminopeptidase; and natural sweeteners, for example mannitol and xylitol.[9] Exopolysaccharides may be used in the industry as emulsifiers, thickeners, physical stabilizers and gelling agents and they are also used, as well as enzymes, to aromatize, flavour and texturize foodstuffs.[6,9]

Contrary to what happens in Gram-negative bacteria, these food grade organisms do not enclose endotoxins in their membranes, which are pyrogenic in humans and other mammals. Although being the first-choice microorganism for the expression of heterologous proteins, *E. coli* may present endotoxins in its derived products, limiting the final application of its products, particularly those with pharmaceutical means.[14] Given these limitations, an increased number of scientists are considering Gram-positive bacteria as safer

microorganisms for heterologous goods production, and a considerable number of homologous enzymes are already being produced by these microorganisms. While not being the safest organism, *Bacillus subtilis* is one of the most used Gram-positive bacteria used as a microbial cell factory, due to its excellent secretion system, simplifying the downstream processing, when compared to *E. coli*.[14,15] Gram-positive bacteria are in fact promising candidates for membrane and complex proteins production, proteins which majority are significant drug targets and constitute one of the major challenges in Biotechnology. The extension of heterologous proteins production in LAB would have a higher impact in Pharma and Biotech industries, which is leading to the development of key tools for goods production, such as expression vectors, optimal promoters, modified strains and improved induction and secretion systems.[14] Implementation of lactic acid bacteria as a routine cell factory expands their applications from conventional food microbiology[16] to protein and DNA production and also drug display and delivery[14,17–19], taking advantage of the GRAS features of this platform.[15]

Lactococcus lactis has been described as a promising LAB for recombinant protein production not only due to previously described characteristics, but also because it has a low batch-to-batch product variation.[14] Glenting and colleagues have shown that *L. lactis* can be used for the recombinant high yield production of genetically engineered hypoallergens for immunotherapy purposes, because this expression system overcomes problems associated with extraction, safety and variability of isolated products.[20] It has also been shown the effective production of a murine interferon-gamma cytokine and a staphylococcal nuclease. The first is also produced by *E. coli*; however, it forms inclusion bodies, hindering the extraction process.[21] Other studies prove that *L. lactis* is also a good live vector candidate for delivering antigenic or therapeutic proteins at the mucosal level. Indeed *L. lactis* has been successfully used for the delivery of bacterial, viral and protozoal antigens and also therapeutic proteins from different origins.[14]

Indeed, an exhaustive work has been done in developing different tools for the recombinant protein production using LAB as cell factories.[4,14,15,17] The development of these tools has made possible the development of LAB capable of secreting proteins of interest to the extracellular environment, becoming a key aspect when evaluating the potential of these bacteria for mucosal targeting of therapeutic molecules.[22] Alternatively, approaches based on protein display anchored to the bacteria cell wall have also been tested[23], being a system that, even not being as effective as secreted protein in terms of protein expression levels, gives a higher protection to the protein in front of degrading and denaturing agents.[7]

Other approaches for the use of these bacteria may include the production of vaccines based on plasmid DNA, which are easy to handle and rapid to construct. LAB hold great promise as they represent an alternative, efficient and safe delivery system for humans and animals.[4,24]

1.1.2 LAB in Nutrition and Human Health

The GRAS and safe status of lactic acid bacteria, the capacity to withstand the passage through the gastrointestinal tract (GIT) of mammals and the increasing availability of genetic manipulation tools has led to an increase use of these bacteria, not only in foodstuffs, but also in medical applications. The most

acknowledged application of LAB in human health is its use as probiotics, as mentioned above. Commercial probiotics are mainly constituted by bacteria of the Lactobacillus genus, and have approximately 100 recognized species, in which are included *L. acidophilus*, *L. rhamnosus*, *L. reuteri*, *L. casei*, *L. plantarum* and *L. bulgaricus*.[9] Probiotics exert their beneficial effects long term, as they can colonize the gastrointestinal tract and, thus, there is no requirement for continuous medical intervention. They have been used for over a century to treat a range of mucosal surface infections in the gut and in the vagina. To fit into the probiotic classification, bacteria must fulfil the following functional and physiological criteria: i) ability to adhere to the intestinal epithelium and to colonize the lumen of the tract; ii) capability of stabilizing intestinal microbiota; iii) counteract the action of detrimental microorganisms; vi) produce antimicrobial substances; and v) stimulate and modulate immune response.[9] Certainly, immunomodulation, enhancement of mucosal barrier function against ingested pathogens and antimicrobial activity constitute the main means by which probiotics actuate, improving mucosal defences. Different LAB have different effects on human health, but in general, they have potential health benefits when dealing with diarrheal diseases, inflammatory bowel disease, irritable bowel syndrome, colon cancer, *Helicobacter pylori* infections, lactose intolerance and blood cholesterol. Other disorders, such as bacterial vaginosis, atopic dermatitis, obesity, caries implementation and periodontal disease may also benefit with the use of LAB as probiotics.[9]

In the past years, and as previously mentioned, LAB, especially *Lactococcus lactis*, has been extensively studied for expression and secretion of heterologous proteins. For what concerns the expression of heterologous proteins, *L. lactis* is considered the model organism for these studies due to the non-invasive, non-pathogenic and GRAS profile and as much of genetic tools for LAB were developed for this organism.[8,12] Additionally, *L. lactis* secretes relatively few proteins and only one (Usp45) in measurable quantities, easing the profiling of the final product in laboratory and the downstream processing, for scalable protein production and purification purposes. In summary, the main advantages that liberate LAB as safe are: i) inexistence of LPS in the outer membrane such as endotoxins or biogenic amines; ii) suitable growth in chemically defined media; iii) bulk purified plasmid genetic integrity; and iv) availability of molecular tools for delivery and expression of heterologous antigens and therapeutic molecules.[25,26]

In the past years, the study of *lactococci* and *lactobacilli* as vehicles for medical therapies have much increased. The use of the secretion capacity of *L. lactis* for the release of recombinant anti-proteases, antioxidant enzymes and auto-antigenic peptides to treat inflammatory bowel disease (IBD) and auto-immune diseases have been recognized.[8] The LAB heterologous protein expression capacity has been used to target infectious agents, preventing infections at the mucosal surface.[11] The use of recombinant *lactococci* as vehicles to deliver DNA directly into eukaryotic cells has also been proposed.[27]

The use of *L. lactis* as a vector for the delivery of therapeutic molecules has been successfully described in the treatment of ulcerative colitis (UC) and Crohn's disease (CD), the most common forms of IBD.[11,12] *L. lactis* and *L. casei* were genetically engineered to produce and secrete elafin, an endogenous protease inhibitor found in the human gut with anti-inflammatory properties in mucosal surfaces, to

counteract the high proteolytic activity found in IBD patients. After orally administrated in IBD mice, the rLAB reached the colon and produced and excreted elafin, preventing inflammation, accelerating mucosal healing and restoring the colon homeostasis. Similar results were obtained with *in vitro* cultures of human intestinal epithelial cells. Other applications using recombinant *Lactococcus* sp. for the delivery of neutralizing molecules for the symptomatic treatment of IBD (for e.g. superoxide dismutase and trefoil factor family peptides) have also been successfully described in murine models.[7,8,11,12] Following the same reasoning as IBD treatment, *L. lactis* producing antigens to induce local and systemic antigen-specific regulatory T-cells responses have been used to rest type 1 diabetes (T1D).[12] Another very promising finding was that the oral administration of *L. lactis* can be used to induce antigen-specific OT (oral tolerance) to a secreted recombinant antigen and that the responses induced by LAB delivery were much more efficient than with purified antigen. This may avoid the need for antigen large purification and, consequentially, the costs of implementing this technique. This feature may as well be a great promise for the treatment of allergies and antigen-induced autoimmune diseases.[11] Furthermore, the LAB expression of anti-infective agents aiming the prevention of infections caused by viruses, bacteria and fungus in mucosal surfaces have been studied. Cell binding and fusion processes of HIV, for e.g., offer a wide range of possible target sets for inhibition of infection. LAB *Streptococcus gordinii*, *L. lactis* and *L. plantarum* were constructed to block the binding of the gp120 viral protein to the CD4 extracellular domain of immune cells by secreting or anchoring the high affinity microbicidal cyanovirin-N (CV-N) protein. The CV-N protein expressed by these strains was able to neutralize the infectivity of HIV-1 *in vitro*, both in laboratory and primary isolates. Another application of LAB in this matter, is its use to produce neutralizing antibodies to battle pathogens. One prominent example was the use of *S. gordinii* displaying in its cell surface and secreting the microbicidal scFv against *Candida albicans*.[11]

The use of LAB for the delivery of DNA vaccines was also described and is addressed in the above section.

1.2 Gene Therapy

Gene therapy is roughly stated as the introduction of nucleic acids into cells, using a transfer vector, aiming to change gene expression for the treatment or prevention of a pathological process. The provision of the therapeutic vector can either be done by *in vivo* or *ex vivo* delivery, towards the target gene addition, alteration, correction and/or knockdown. The *in vivo* approach consists in the direct injection of the gene containing vector in the patient, whereas in the *ex vivo* approach the genetic modification of interest is performed *in vitro*, using autologous patient cells that are then re-introduced in the patient body.[28,29]

The idea of gene based-therapeutics arose in the early 70's, but only received serious attention after the discovery of recombinant DNA technology and the capacity to introduce and express heterologous genes in mammalian cells.[28] Indeed, the first clinical trials took place in the late 80's and gene therapy was though to overtake therapeutics for serious genetic-related diseases in a matter of years. The first reports

concerning the use of a protein-coding gene for genetic immunization were published in the beginning of the 90's and stated that RNA and DNA directly injected into mouse tissues led to the expression of antigens not naturally present in the host cells.[30,31] As such, the idea that DNA could be used to trigger humoral and cellular immune response against tumour and pathogen epitopes arose, instigating a new field study regarding DNA as a therapeutic vaccine.[31]

However, numerous of obstacles have emerged throughout the next twenty years, tempering the enthusiasm around the potentialities of this clinical approach. Reviving of gene therapeutics took place after more recent discoveries of successful examples and encouraging novel preclinical trials that promised to expand the number of possible disease targets.[28] Notwithstanding the drawbacks associated with the safety of human gene therapy, this field has been receiving more attention from the scientific community as it represents a great promise to new treatment methodologies, especially for cancer, monogenic and infectious diseases and cardiovascular disabilities.[30,32]

1.2.1 Delivery vectors

To efficiently and safely deliver genetic material into host cells, towards achieving the anticipated therapeutic result, the delivery vector is key. Regarding its nature, vectors may fall into two different categories - viral and non-viral - and must be able to deliver the gene of interest into host cells so that its expression and action is possible. Viral vectors have been the most studied and used in clinical trials, predominating the application of adenovirus and retrovirus. However, non-viral vectors have gained ground in gene therapy following the development of this vectors transfection efficiency and it appears more attractive due to safety issues.[29]

Viral vectors are widely used due to their great capacity to introduce DNA sequences of interest inside target cells, which is a key factor for gene therapeutics. When designing a delivery approach using viruses some considerations regarding their natural properties, beyond their replication cycle nature, must be crossed. They include the limiting length of genetic material and, consequentially, of the desired gene insert, the requirement of repeated sequences; gene expression extent after transduction; and the virus capability of infection of the target model. Disadvantages associated with recombinant virus delivery include possible endogenous recombination, oncogenic effects and unexpected immune responses.[29,33] Despite safety issues and reproducibility in clinical trials, these vectors continue to be the most used as gene transfer vehicles.[30,34]

Non-viral vectors constitute a much simpler, and more importantly, safer alternative for the delivery of a gene of interest, as they circumvent the disadvantages associated with viral delivery, easing large-scale production and diminishing the possibility of specific immune responses. The simplest method of gene delivery is the direct injection of naked plasmid DNA into tissues or into vessels of the host.[35] Other physical delivery methods include electroporation and gene gun.[35,36] The possible recognition of these vectors by the host immune system hurdles the delivery process. To impair this problem, the delivery mediated by chemical carriers like lipids and cationic polymers was developed.[29,36] To be safely used in human gene

therapy, the plasmid DNA vector must be prudently designed, avoiding sequences that may instigate toxic effects in the host system, and sequences that may difficult the production process, without compromising the final purpose. This system allows the delivery of large sequences of DNA; however, the transfection efficiency seems to be inversely proportional to the plasmid vector size. Larger plasmid vectors can negatively influence the expression of the gene of interest. Also, reducing the plasmid vector size may increase DNA stability and minimize integration events.[37–39] For what concerns safety of DNA delivery in mammalian cells, non-viral vectors partially or totally lacking prokaryotic elements, such as minicircles, are advantageous.[24]

DNA vaccination is a recent technique that has upraised substantial expectations since its discovery as it has been proved that it can induce humoral and cellular immune responses.[11] The vaccination process engages immunization with DNA, usually plasmid DNA (pDNA), that contain the gene or genes that code for one or more antigens of interest. Once inside the target cells, the coding sequence is expressed into the immunizing or therapeutic protein, enabling the host system response to the problem to be addressed. One great advantage of DNA delivery is the possibility to express multiple antigens using a single DNA vector.[11]

One important parameter to assess whether plasmid DNA is of high quality for transfection and will properly redeem its function in biological systems is the determination of the major pDNA conformation present after the purification process. Plasmid DNA can be present as compact supercoiled circular, relaxed open circular and linearized forms. Supercoiled fractions are preferable over open circular and linear fractions, as they are considered to be the most biologically active and as transfection procedures are facilitated.[40,41] Remaut and co-workers have shown that supercoiled topologies have higher transfection yields when transfected directly to the cytosol of Vero cells and that the transcription of pDNA is independent of its conformation, as the microinjection of all plasmid isoforms directly into the nuclei, yielded similar transcription efficiencies. Therefore, it was hypothesized that the supercoiled pDNA conformation is more suitable to reach the perinuclear region and be entrapped in the nuclei.[40] The FDA has in fact recommended the establishment of a minimum content of supercoiled plasmid, preferably higher than 80%, when manufacturing DNA vaccine products.[41]

1.2.2 Lactic Acid Bacteria as Delivery Live Vectors: An Alternative Approach

Alternatively, DNA can be delivered into host cells by live vectors. Attenuated pathogens have been used to deliver plasmids to the mucosa, but their “not-so-safe” status rise safety issues, despite the attenuation, as they can still reverse to a virulent state. Furthermore, it is difficult to construct a stable attenuated mutant strain, the presence of residual virulence may still exist and there are risks associated with assortment between the vaccine strain and the wild type.[7] The use of recombinant *E. coli* for delivery purposes was also studied.[12] However, its high inflammatory lipopolysaccharides (LPS) inherent production may cause inflammatory responses in the host, rendering it unappealing for delivery drives.

The use of *L. lactis*, particularly, seems to be attractive for plasmid delivery to epithelial membranes as the system is safe (GRAS), food-grade, transient and non-invasive.[12] It has been shown that a plasmid

containing an expression cassette with the eukaryotic cytomegalovirus promoter and a coding sequence of the bovine cow's milk allergen beta-lactoglobulin (BLG), can be transferred from *L. lactis* to mammalian Caco2 epithelial cells *in vitro* and *in vivo*.[11,42,43] Plasmid DNA and BLG were both found on the epithelial membrane of small intestine in mice orally administrated with this *L. lactis* strain. Though the BLG production was transitory, it led to a T helper 1 cells-motivated immune response characterized by a very low level of serum BLG-specific IgG2a and protection against further sensitization with BLG and cholera toxin. The *blg* gene and product were not detected when the cells were incubated with the recombinant plasmid alone, neither with *L. lactis* mixed with the recombinant plasmid.[11] Although the mechanism by which DNA is transferred from bacteria to the desired host is still not totally identified, it is conceivable that lactic acid bacteria play an important role in DNA delivering itself. It is to denote that, in addition to the need of cell-cell communication and/or vector cell internalization by epithelial cells, the need of plasmid vector relocation to the nucleus may be a limiting step for the expression of the desired gene, as well.[12]

Very recently, Mancha-Agresti *et al*. successfully corroborated that *L. lactis* is a good candidate for the delivery of genes into eukaryotic cells and for their expression, confirming previously made experiments. *L. lactis* harbouring the pExu plasmid, with an *egfp* ORF under the control of the eukaryotic P_{cmv}, was orally administrated to mice by gavage. The epifluorescence and flow-cytometry monitoring detected expression of the GFP protein between 12 and 72 hours in the duodenal portion of the GIT.[27] These results confirm the use of *L. lactis* as an effective mucosal delivery vehicle of gene expression plasmids in eukaryotic models.[27,43]

Using LAB for the delivery of DNA is very much appealing given the prospect of using mucosal routes of administration. However, this delivery method is much less effective in generating immune response as is injecting the DNA vaccines *in situ*.[42] Aiming the enhancement of DNA vaccine delivery, efforts are being made to express pathogens invasins, for instances internalin A and FnBPA, in LAB.[11,42] The receptors for these invasins are usually expressed on the basolateral membrane of epithelial cells, which means that they are not accessible throughout the entire intestinal lumen, and should only be partly expressed in some regions such in the tips of the villus. In fact, it has been shown that targeted *L. lactis* cells are only taken up by a few percentage of epithelial cells, either *in vitro* or *in vivo*.[42] Guimarães *et al*. have demonstrated that internalin-expressing *L. lactis* indeed increased the invasiveness of Caco-2 guinea pig small intestine cells by ~100-fold, compared to the wild type and was able to deliver DNA into epithelial Caco-2 cells.[44] This bacterium is not known to have an intrinsic mechanism to escape endosome or phagosome compartments. Therefore and, as previously mentioned, the mechanism by which *L. lactis* releases the pDNA material in host cells and how DNA reaches the nucleus are subjects yet to be elucidated.[11]

As here discussed, there are different successful studies of the use of rLAB for the vaccination and protection of rodent models. However, the great challenge comes in demonstrating their effectiveness in human and animal cells, and later on, systemically.[42]

Another possibility is to make use of naked DNA of LAB origin for gene therapy purposes. *E. coli* is widely used as host for production of pDNA, due to its high yield production, its robustness, its capability to fast grow with minimal nutrient requirements, its genetic knowledge, and the well-established downstream processing. However, the use of this organism as a cell factory renders some problematics, associated with the possibility of the products' co-purification with immunogenic endotoxins or lipopolysaccharides (LPS).[45] LPS is the major component of the outer membrane of gram-negative bacteria and it constitutes the primary recognition structure by immune system cells, such as monocytes and macrophages, of mammalian hosts. The innate immunological response generates the release of a range of pro-inflammatory mediators, like TNF-a, IL-6 and IL1B, which potentiate inflammation and elimination of invading organisms. In more extreme cases, the crossing of LPS to the host bloodstream may cause a systemic inflammatory reaction, which can be fatal.[46] The use of gram-positive bacteria as hosts for pDNA production, like food-grade microorganisms, that lack LPS molecules, bypasses the problematic affiliated to the use of *E. coli* as a microbial cell factory for pDNA production.[45] One suitable approach to increase pDNA production in organisms that are not characterized by high pDNA production is the improvement of those organisms using genome engineering techniques. The reasoning of such approach is to insert or delete genes that are known to improve or impair pDNA production and quality, respectively.

Either for the expression of heterologous genes or for the delivery of DNA itself from LAB, several plasmids have been developed, but their copy numbers are generally very low. This means that it is not easy to achieve high-level gene expression using those low-copy plasmids, and the limited availability of plasmids has been one of the big hurdles in using LAB as pDNA and protein production hosts.[47] As such, and again, these bacteria may benefit genome alterations in order to achieve higher plasmid production either for gene therapy purposes or increased protein production. Theoretically, knockout gene targets for enhanced pDNA yield production and quality would be genes that are responsible for DNA degradation and unspecific recombination, leading to increased pDNA production, such as endonuclease and recombinase genes. Genes whose inactivation would redirect carbon flux to the pentose phosphate pathway, leading to an increased production of nucleotides and, consequentially, pDNA, also seem to be of relevance. In *E. coli* there are some genes that have been identified as having a relation to pDNA production, whose knockout would increase DNA production and/or quality.[48] This set of genes include *endA*, *recA*, *pgi*, *pykA* and *pyk*F, coding for endonuclease I, bacterial DNA recombination protein A, phosphoglucose isomerase and pyruvate kinase, respectively.[48,49] Following this reasoning, *L. lactis* pDNA production may be assessed following the aforementioned homologous genes knockouts.

1.3 Genome Editing Engineering

Genetic engineering has emerged almost fifty years ago with the discovery of recombinant DNA technology by a group of scientists that combined the *E. coli* genome with the genes of a bacteriophage

and the SV40 virus. This technology revolutionized the molecular genetics field as it has provided a relatively easy way to manipulate DNA to construct desirable hybrid sequences, allowing the creation of transgenic microorganisms and genetically modified plants and animals. Therefore, the rapid development of this technology allowed the frenzied progress in fundamental and applied biology. However, conventional genetic engineering has several drawbacks, of which the complexity of manipulation of large DNA sequences, such as genomes.[50]

Sequencing development, especially New Generation Sequencing (NGS) technologies, has provided information on DNA nucleotide sequences of main model biological objects and, consequentially, allowed the determination of genomic functional, non-functional and regulatory elements. Although this characterization is yet to be fulfilled it has greatly aided genomic mapping, accelerating biological and medical research. In a post-genomic era, methods allowing precise genomic manipulation and regulation of gene expression are rapidly evolving.[50]

The foundation for the gene editing field was the discovery that DNA double-strand breaks (DSBs) could be used to induce endogenous DNA repair machinery. Typically, breaks in the DNA are repaired by two major DNA repair systems: homology-directed repair (HDR) or Non-Homologous End Joining (NHEJ).[51] HDR relies on strand invasion of the broken end into a homologous sequence and repair in a template-dependent manner and homologous recombination is greatly potentiated when accompanied by the introduction of a double-strand break (DSB) at the target site.[52,53] Alternatively, the error-prone NHEJ functions to repair DSB without the need of a DNA template. Instead NHEJ functions through direct re-ligation of the cleaved-ends. As no template is provided, this repair system is susceptible to insertions and/or deletions (indels) at the break site. NHEJ has been used to disrupt target genes in a wide variety of organisms, mainly eukaryotic, by taking advantage of these indels to shift the reading frame of a gene.[54,55]

Opposed to the unpredictable outcoming mutations of NHEJ, targeted DSBs can induce precise genome editing by stimulating homologous recombination with an exogenously supplied donor template for DNA repair. Thus, the co-delivery of targeted nucleases along with a targeting vector containing DNA homologous to that of the break site region enables high efficiency HDR-based gene editing. Any sequence modifications present in the donor template can thus be incorporated into a targeted endogenous locus.[54] While plasmids have been the most commonly used delivery vector, recent studies show that single stranded oligonucleotides (ssODNs), with as little as 80 base pairs of homology can be successfully delivered as donor templates for HDR.[54,56,57] For cells in which transfection may be a challenge, viral vectors, such as lentivirus or adeno-associated virus can also be used as sources to delivery ssODNs.[54]

The use of a donor template in which the desired genetic insert or modification is flanked by homology arms, enables site- specific DNA insertion through DSB-induced HDR. The use of a donor template containing the homology arms with no sequence in between them, enables the DSB-induced target deletion. Deleting large segments of DNA, like genes, can also be accomplished by flanking the sequence with two

DSBs. Indeed, it has been shown that introducing both these DSBs can give rise to genomic deletions up to several megabasepairs in size.[58,59]

Because DNA DSB-induced gene editing relies on the endogenous repair mechanisms of the cell, it is universally applicable to any cell type or organism that employs these methods of DNA repair. The critical element for implementing any gene-editing method is the precise introduction of the DSB.[54] Four major platforms currently exist for introducing these site-specific breaks: zinc-finger nucleases (ZFNs), transcription activator- like effector (TALE)-nucleases (TALENs), meganucleases and, most recently, clustered regularly interspaced short palindromic repeats (CRISPR) CRISPR-associated (/Cas) system. The present work focuses in the cutting-edge technology CRISPR/Cas9.

1.3.1 CRISPR/Cas9

The CRISPR/Cas9 system is derived from an adaptive immune system that evolved in bacteria as a defense mechanism against foreign DNA, such as the one coming from virus or other bacterial plasmids.[54] Decades of research has illuminated scientists about the mechanisms by which this system works. Short sequences of evading nucleic acids are incorporated in CRISPR loci, which consist of mysterious repeats. These repeats were first described as early as 1987 by Japanese researchers[60], but their function remained unknown for nearly 20 years. Sequencing of bacterial genomes revealed similar nucleotide sequences in genomes of many microorganisms, such as bacteria and archaea, and consisted of short regions of unique DNA, spacers, separated from each other by short palindromic repeats.[60] These CRISPR cassettes were found to be in close proximity to *cas* (CRISPR-associated) genes, the protein products of which have helicase and nuclease activities (Figure 1A).[61] In 2005, independent bioinformatics' studies reported that the spacer DNA sequence was often homologous to the DNA region of phages and plasmids.[62,63] Adding with the finding that the CRISPR locus was transcribed, scientists anticipated that CRISPR-Cas systems would be an adaptive defense mechanism that use RNAs as memory signatures of past foreign DNA invasions. Indeed, in 2007, the first evidence of the CRISPR-Cas system as a system for adaptive immunity was attained by means of lytic phage infection in lactic acid bacterium *Streptococcus thermophilus*. Cells bearing the CRISPR locus spacer that was complementary to a bacteriophage genomic DNA fragment became resistant to the phage when infected by it.[64] It became clear that the CRISPR/Cas9 system provided a unique mechanism of protecting bacterial cells against foreign DNA penetration, acting along with the restriction-modification system as a limiter of horizontal gene transfer of genetic information.[50] The aforementioned experiment with *S. thermophilus* showed that the phage was abrogated by as little as a single mismatch between the spacer sequence in the CRISPR locus and the phage sequence; however, studies revealed lower stringency of spacer-target complementary.[65] In 2008, mature CRISPR RNAs (crRNAs) were found to serve as guides to the DNA targeting activity of Cas proteins by forming a complex with them. Different Cas systems were described and, in 2011, a trans-activating crRNA (tracrRNA), a small RNA trans-encoded upstream of type-II CRISPR-cas locus, was conveyed as essential for crRNA

maturation by ribonuclease III (RNase III) and Cas9, in the sense that the crRNA maturation mediated by the tracrRNA was crucial for the sequence-specific immunity against extraneous DNA. By 2012, after the discovery of *Streptococcus pyogenes* type-II CRISPR-Cas9 as a dual-RNA-guided DNA endonuclease that directs DNA cleavage guided by the tracrRNA:crRNA duplex, this duplex structure was engineered as a single guide RNA (sgRNA) embodied by a 5' DNA sequence that determines the Watson-Crick base pairing and a 3' guide sequence that binds to Cas9. This two-component system started being extensively used in late 2012, sharing the throne with ZFNs and TALENs as the paramount technologies for genome engineering purposes and functional studies.[50]

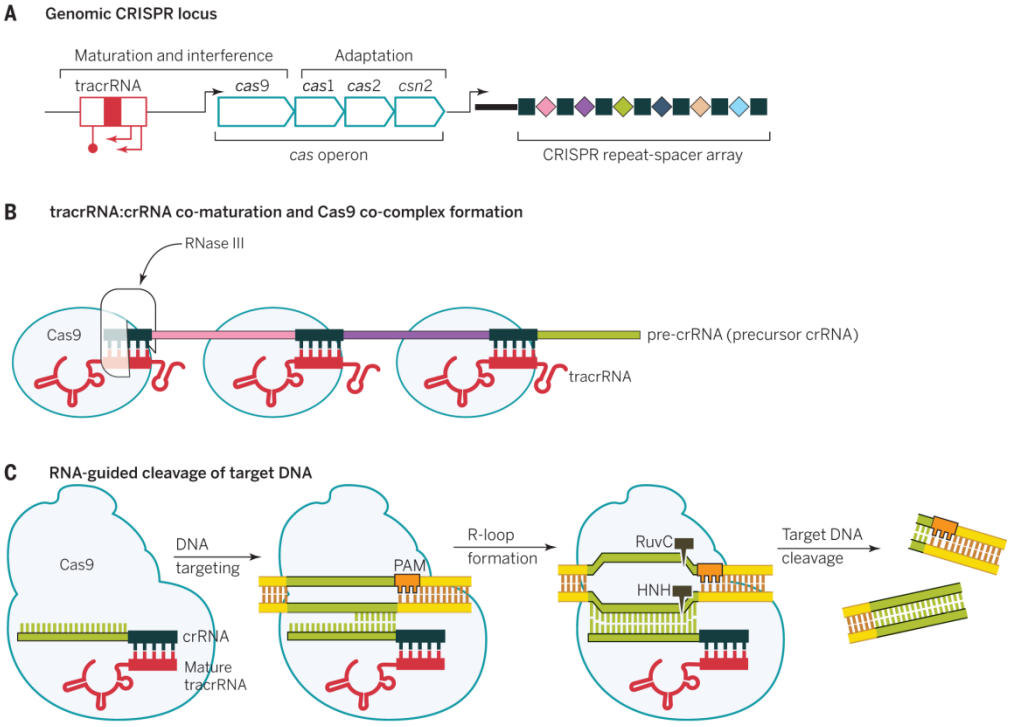

Figure 1 Representation of the type II-A CRISPR-Cas system. The type II-A system from S. pyogenes is shown as an example. (A) The cas gene operon contains the genomic sequence of tracrRNA and the CRISPR array. (B) The natural pathway of antiviral defense involve association of Cas 9 with the antirepeat-repeat RNA (tracrRNA:crRNA) duplexes, RNA co-processing by ribonuclease III, further trimming, loop formation, and target DNA cleavage. (C) Details of the natural DNA cleavage with the duplex tracrRNA:crRNA. (From Doudna et al.[66])

The naturally existing CRISPR/Cas systems mediate immunity response to invading genetic elements through three different stages: adaptation, expression and interference.[65] The adaptation step allows short pieces of DNA homologous to viral or plasmid sequences to be integrated into novel spacers in CRISPR arrays. This may also be termed as the vaccination event. Viral challenge typically triggers insertion of a single virus-derived resistance-conferring spacer, with a characteristic length of approximately 30 bp, at the leader side of a CRISPR locus. Acquisition of multiple spacers from the same foreign DNA is less frequent. Each integration event is accompanied by the duplication of a repetitive unit, thus creating a new spacer-repeat unit. The new spacer sequence is always inserted on the A-T rich side of the leader sequence, located before the existing CRISPR cassette.[50,67] The selection of spacer precursors (proto-

spacers) from the evading DNA appears to be determined by recognition of proto-spacer-adjacent motifs (PAMs). PAMs are short conserved sequences (usually 2-5 bp) and differ between variants of the CRISPR-Cas system.[68,69] During the expression stage, the CRISPR locus is transcribed into a long primary transcript (pre-crRNA) and processed into short CRISPR RNAs (crRNAs).[67] The processing step is catalyzed by endoribonucleases that either operate as a subunit of a large complex or as a single enzyme (depending on the CRISPR system) to process the long pre-crRNA into smaller crRNA molecules. The third and final step is interference, during which the foreign DNA or RNA is targeted and cleaved within the proto-spacer sequence. crRNAs guide the respective complexes of Cas nucleases to the complementary exogenous sequence that match the spacer sequence for sequence-specific cleavage.[50]

The CRISPR system is widespread in prokaryotes, being present in 87% of archaea and 48% of eubacteria.[50,64] Long-term co-evolution of viruses and their hosts has led to the formation of viral protection mechanisms against bacterial interference mechanisms such as CRISPR, which explains the wide variety of CRISPR/Cas systems in bacteria and archaea.[70] It has been shown that CRISPR/Cas systems can either target foreign DNA or mRNA depending on the CRISPR system and organism. These findings emphasize the remarkable mechanistic and functional diversity of CRISPR-Cas systems.[65] Indeed extensive bioinformatics analysis have shown that the genome of various CRISPR-containing organisms encode more than 65 distinctive set of orthologous Cas proteins, which can be classified into 23-45 families[65], depending on the classification criteria. Currently, bioinformatics studies classify all CRISPR/Cas systems into three main types I, II and III, and at least 10 subtypes. Among these, the type II-A CRISPR/Cas system isolated from the *Streptococcus pyogenes* pathogen is presently the one used most widely in genomic engineering. This bacterium has been found to have a minimum set of *cas* genes.[65,71]

In the type II-A CRISPR/Cas system, one polyfunctional Cas9 protein performs both the processing of pre-crRNA and the interference of foreign DNA.[72] The crRNA processing also depends on small non-coding tracrRNA, which bind by complementarily to the repeat sequence in the pre-crRNA, forming a duplex. Meanwhile, the host ribonuclease III cuts the duplex in the presence of Cas9 to generate mature crRNA containing a ~20 nucleotide spacer sequence at the 5'-end and a 3' loop-like secondary structure (Figure 1B and C). In the presence of Mg^{2+} ions, the Cas9 enzyme makes a double strand break at the target locus that pairs up with the spacer sequence. Two different domains are responsible for the double nicking capacity of Cas9: the HNH nuclease domain cuts the DNA strand complementary to the crRNA and the RuvC domain cuts the non-complementary strand (Figure 1C). The target DNA for *S. pyogenes* Cas9 (SpyCas9) should necessarily contain a 5´-NGG-3' PAM sequence, three nucleotides from which the cleavage occurs. Cas9 complexed with the crRNA and tracrRNA undergoes a conformational change and associates with several PAM motifs throughout the genome, inquiring the sequence directly upstream to determine complementary with the guide RNA. The arrangement of a DNA-RNA heteroduplex at a complementary target site allows for cleavage of the target DNA by the Cas9-RNA complex, when complementarity is found.[73]

Accordingly, three main components are required for the type II CRISPR nuclease system: Cas9 protein, the mature crRNA and the tracrRNA. However, this system was engineered to reduce these requirements to two main components by fusing the crRNA and tracrRNA into a single guide RNA (sgRNA) (Figure 2).[73] Furthermore, studies showed that re-targeting of the Cas9/sgRNA complex to new target sites could be accomplished only by altering the sequence responsible for pairing up with the genome target locus. After this discovery, a series of publications have demonstrated successful strategies of CRISPR/Cas9 engineering for genetic modifications, mainly in mammalian cells. Collectively these studies have pushed the CRISPR/Cas9 technology into the spotlight of the genome-editing field.[74–76] This system can be used to disrupt, correct or add genes or features taking advantage of the error-prone or directional DNA repair mechanisms of the cell, as previously discussed.

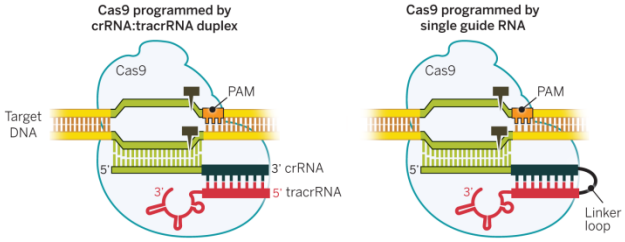

Figure 2 Evolution and structure of Cas 9. The structure of S. pyogenes Cas9 in the unliganded/non-engineered and RNA-DNA-bound/engineered forms. (From Doudna et al.[66])

The only sequence limitation of CRISPR/Cas systems derive from the necessity of a PAM sequence located immediately 3' to the target site. As previously mentioned, the PAM sequence 5´-NGG-3' is necessary for binding and cleavage by the commonly used SpyCas9.[50,77] As for other species like *S. thermophilus* and *Neisseria meningitides* Cas9, PAM sequences correspond to 5´-NGGNG-3´ and 5'-NNNNGATT-3´, respectively.[50] However, Cas9 variants with novel PAMs must be engineered by directed evolution, expanding the number of potential target sequences.[78]

Genome-wide CRISPR binding specificity, assessed using chromatin immunoprecipitation (ChIP) assays, demonstrated that RNA-guided Cas9 binding was highly specific to the target site, while off-target binding occurred at much lower intensities. As expected, regions for which the Cas9 had bond to DNA were highly enriched in NGG sites. However, the Cas9-mediated cleavage was not observed in all ChIP-bound regions. A high frequency of indels was observed at the target sites, but almost no cleavage activity was observed for off-target sites. This means that the complex may bind to different loci in the genome, driven by the existence of PAM nucleotides, yet it is the sgRNA/target ligation that ultimately induces DSBs.[79]

In order to have the most successful genome-editing tool possible, a careful selection of sites for specific introduction of DSBs and preliminary bioinformatics analysis is commended. Regions of repeated sequences and/or with high homology to other genome regions should also be avoided.[50]

One of the main drawbacks of the CRISPR/Cas9 system is a relatively high probability of off-target mutations. A number of studies carried out *in vitro* in bacteria and human cells have revealed that some single nucleotide substitutions in the ~20-nucleotide spacer region of the sgRNA may lead to a significant reduction in the activity of CRISPR/Cas9, especially if these mutations are located in the last 10-12 nucleotides of the 3´-end of this complementary sequence.[50,80] However, mutations in the 5' sgRNA region did not prove to have impact on the system's activity.[50] Although there are some exceptions to the rule, in general, the off target effect is determined by the position of the substitutions, the number of substitutions (no more than three), by features of the target site and by concentration of introduced Cas9 and sgRNA.[75] Again, these drawbacks may be overcome with the discovery of novel and more complex PAM sequences, based on the search and development of Cas9 orthologues.[65]

To increase genome editing specificity using this technology, the Cas9 can be engineered in order to lose its nuclease activity. Mutations in one of the catalytically active domains (D10A in HNH and H840 in RucC) convert the Cas9 nuclease into a DNA nickase.[73,81] In this case, the requirement of two obligate heterodimer gDNA/Cas9 complexes to bind the target DNA in a specific orientation and spacing is less probable to happen twice in the same off-target site, which in turn leads to higher DSB specificity. Additionally, it has been shown that reducing the length of complementarity between gRNA and the target site from 20 to 17 nucleotides increased the specificity of DNA cleavage by SpyCas9.[50,54]

Unlike the other nuclease systems such as ZFNs, TALENs and meganucleases, CRISPR/Cas nucleases do not require novel protein engineering for each DNA target site. The relatively ease with which new sites can be targeted, only by altering the short region of the gRNA that dictates specificity, makes this system a highly attractive method for introducing DSBs.[54] Furthermore, as the Cas9 protein is not directly coupled with the gRNA, this system is amenable to multiplexing through the introduction of multiple gRNAs to induce DSB at several loci.[75] Because the CRISPR-based gene-editing technology is relatively recent, many natural CRISPR systems are largely understudied. Thus, it is reasonable to expect new CRISPR-based technologies to emerge in a near future.[54]

Variants of native Cas9 proteins, in which both domains (HNH and RuvC) are inactivated, generating a deactivated Cas9 (dCas9), can extend the application of CRISPR/Cas9 technology applications. These DNA binding proteins can be used as transcriptional effectors, either acting as repressors or as activators blocking or enhancing transcription, respectively. This catalytically inactive dCas9 has proven to function in human, bacterial and yeast cells. The approach in which a repressor domain is coupled with dCas9 is commonly termed CRISPRi and leads to transcriptional reduction of the target RNA (Figure 3). The most regularly used Cas9-coupled repressor is the KRAB domain. Highly specific silencing and effective knockdown has been achieved in yeast. On the contrary, methods that use activator domains coupled with Cas9 to activate or enhance gene expression are typically denominated of CRISPRa (Figure 3). The most commonly used transcription activator is the viral VP64 protein, although others, such as the *E. coli* omega-subunit of RNA polymerase, may be used. Multiplex activation/repression of the

promoters of several genes has been achieved and was controlled by the target position in the gene promoter. For these purposes, it is extremely important to know the location of the transcription start sites (TSSs).[50,77,82,83]

Epigenetic modifications can similarly be achieved by coupling functional groups to dCas9. The fusion of dCas9 to enzymes that either add or remove methyl groups, responsible for epigenetic control of gene expression, allow researchers to test how such changes affect gene expression.[82]

It has also been shown that the fusion of a reporter gene, like GFP, with a deactivated *cas9* gene (EGFP-dCas9) can be used to visualize DNA loci (Figure 3). This imaging strategy provides a powerful tool for studying chromosome dynamics and structure, extending the dCas9 system beyond gene expression-based applications.[82]

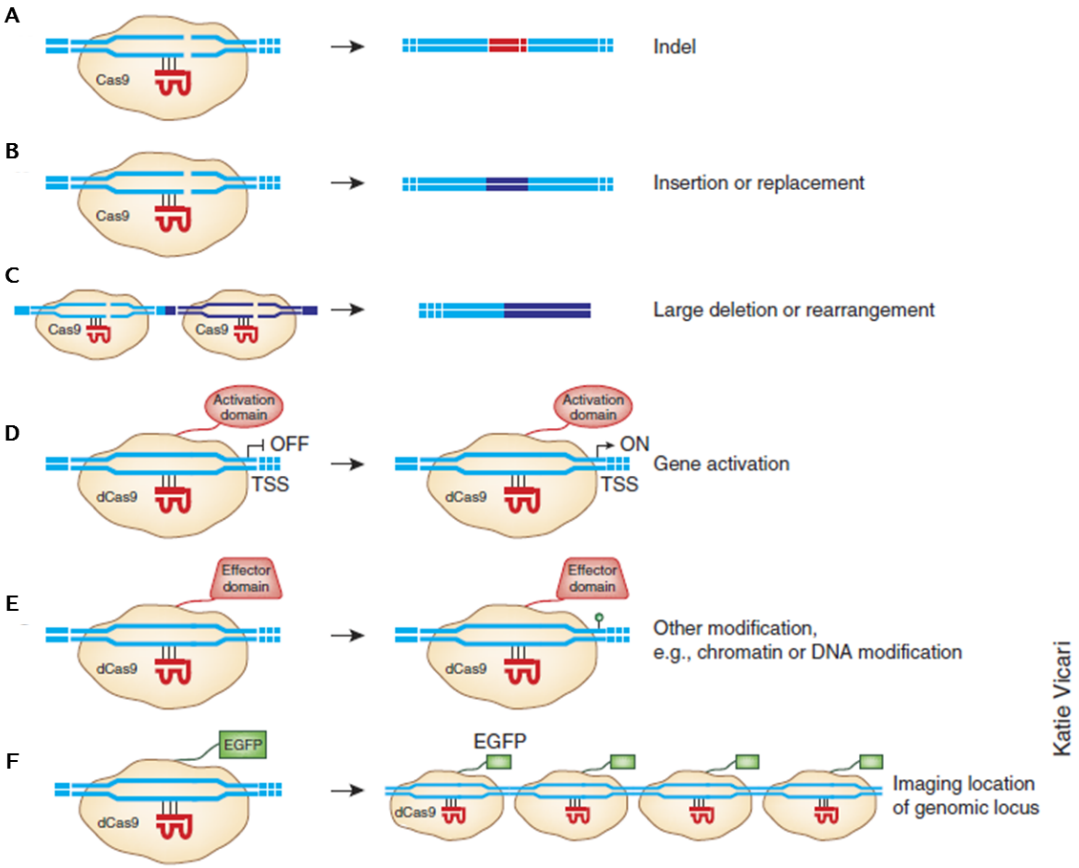

Figure 3 Overview representation of various Cas9-bassed applications. (A,B) gRNA-directed Cas9 nuclease can induce indel mutations (A) or specific sequence replacement or insertion (B). (C) Pairs of gRNA-directed Cas9 nucleases can stimulate large deletions or genomic rearrangements (e.g., inversions or translocations). (D-F) gRNA-directed dCas9 can be fused to activation domains (D) to mediate upregulation of specific endogenous genes, heterologous effector domains (E) to alter histone modifications or DNA methylation, or fluorescence proteins (F) to enable imaging of specific genomic loci. TSS stands for transcription start site. (From Sander et al.[82])

1.3.1.1 CRISPR/Cas9-mediated engineering in prokaryotes

The genome engineering studies that have been made in prokaryotes using the CRISPR/Cas9 system to engineer genomes throughout the recent years can be divided in two main groups: single- and double-stranded DNA recombineering (SSDR/DSDR). Recombineering, a phage-derived recombination system, has been used in conjugation with Cas9 counterselection, and revealed to greatly enhance the selection efficiency.[84] Cas9 counterselection addresses the CRISPR/Cas9-induced lethal DSBs, targeting non-edited native sequences. This mechanism may efficiently work in prokaryotes that lack well developed recombineering systems.[85]

The first exploitation of the heterologous expression of SpyCas9 to introduce specific mutations or to delete genes in bacterial genomes was performed in *E. coli* by Jiang *et al*. *E.* coli cells were simultaneously transformed with the pCas9 vector coding the expression system of SpyCas9 and the tracrRNA, with a linear ssDNA oligonucleotide for SSDR and with the pCRISPR vector. The SSDR was designed to introduce the desired point mutation to the target gene. The pCRISPR vector was composed of a minimal CRISPR array with a spacer that targeted the native gene, which led to the CRISPR/Cas9-based counterselection of wild-type gene cells. After this successful experiment, this tool became proof of concept for a new highly efficient genome editing tool. [84,86] Additional disruptions, insertions and deletions were successfully made in *E. coli*, either relying in the coexpression of a crRNA:tracrRNA complex[87] or in a chimeric sgRNA[88,89].

After the development of this tool in *E. coli*, a tool for *Lactobacillus reuteri* was established. The approach was based on three expression vectors containing: SpyCas9 and tracrRNA, RecT and CRISPR array. The experiment was performed on a single or dual step procedure. In the first approach, the SpyCas9 and tracrRNA expression vector was simultaneously transformed with the CRISPR array vector and the ssDNA recombineering fragments. In the dual step procedure, *L. reuteri* was transformed with the CRISPR array expression vector only after the transformation with the ssDNA recombineering fragments. Although both procedures yielded almost only recombinants, the dual step allowed a more efficient incorporation of the recombineering fragments and, consequentially, recombinant clones, as the replication cycle number was higher before the counterselection with the CRISPR array/Cas9 system. This dual step was further used to delete genomic regions up to 1kb and to perform codon saturation mutagenesis.[84,90,91]

Recently, a DSDR-CRISPR-Cas9 counterselection system was developed for *E. coli*, based on a SpyCas9 and a λ-Red expression vector and on a sgRNA expression plasmid. Using dsDNA sequences as template repairs for the recombineering process allowed for high-efficiency gene deletion. However, gene insertion was less efficient and demonstrated to be less efficient when using shorter HR template fragments. In the same study, Jiang *et al.* used a sgRNA expressing plasmid carrying a HR template as well. HR templates contained in plasmid DNA are commonly denoted plasmid-borne templates. This editing tool allowed for the simultaneous insertion of two or deletion of three genes. This system was also used to delete two genes in *Tatumella citrea*, a microorganism of biotechnological interest. As this editing procedure made

use of a plasmid-borne template and no functional λ-Red mechanism was used, the efficiency of the tool dropped, as it would be expected.[84,92]

A non-recombineering-based homologous recombination (NrHR) system was developed by Jiang *et al.* for the highly recombinogenic *Streptococcus pneumoniae*, in which a CRISPR/Cas9 targeting template and an editing template for HR were delivered to the cells.[86] This system was rapidly projected to other organisms, such as *E. coli*[93], *Clostridium beijerinckii*[94] and *Clostridium cellulolyticum*[95], recently renamed *Ruminiclostridium celluloticum*[96]. For *E. coli* and *C. cellulolyticum*, the Cas9 variant SpyCas9D^{10A}, encompassing a non-functional RuvC nuclease domain, was used to introduce single stranded-DNA breaks in the target site. This approach may help genome editing of critical genes, whose double strand breaks may be too harsh for the host, leading to cell death. Also, it paves the way for editing genomes with inefficient NHEJ and HR mechanisms.[84,93,95] For the edition of the solvent producer *C. beijerinckii*, a vector containing the spycas9 gene, the sgRNA-expressing module and the editing template was designed to be used in a single step. This approach is relevant when using organisms with low recombination efficiency and lacking functional recombineering systems. The CRISPR/Cas9 counterselection was high and the possibility for multiplexing is being studied. [84,94] Similar one-step approaches were performed in *Streptomyces* sp. by Huang and co-workers[97] using a single plasmid containing a codon-optimised *spycas9*, a HR template, a sgRNA and a temperature-sensitive replicon. The latter allowed for plasmid curing after the mutant selection.[84,97–100]

The only existing example of NHEJ repair mechanism after Cas9-induced DSDB in a bacterium was shown in *Streptomyces coelicolor*. Its incomplete NHEJ systems allowed the repair by deletions of variable sizes and the efficiency of the tool was increased by the expression of a heterologous expression of a ligase D.[100] It is however to denote that most prokaryotes contain an inefficient NHEJ repair systems or do not contain a NHEJ repair pathway at all.[84]

1.3.2 Metabolic Modulation

The dramatic increase in the number of sequenced genomes and the advance of experimental high-throughput analysis has led to the need for development of computational methods in biology, aiming the creation of data base libraries, the creation of comparative tools and the construction of models for analysis and integration of the data, in terms of systems properties. Metabolic modelling bases its functioning in metabolic networks (partial) reconstruction and simulation and analyses them within the perspective of the whole network.[101]

In silico metabolic model prediction and design using bioinformatics are emerging areas of great importance in the sense that allow researchers to predict the consequences of certain knockouts or of under- or overexpression of certain genes/reactions. Also, researchers may see it possible to predict production fluxes of interest compounds controlling nutrient availability, simulating gene/reaction deletions or creating new metabolic pathways, which comes in great advantage when designing a microbial cell factory for the

production of a given added- value product. This modelling method produces hypothesis, without the need to resort to time and resource consuming practical experimentation.

Gene or reaction deletion analysis using metabolic modelling can reveal a list of potential affected reactions in a specific model. All reactions must be analysed from a biological significance standpoint looking to obtain some insight as to how each proposed deletion modulates metabolism. This promising tool can be used to predict gene deletions or regulations by optimizing a given objective function, which could be plasmid production for gene delivery vectors, biomass production, increased production of ethanol, etc. Thus, metabolic modelling tools such as CobraToolBox[102] and Optflux[103] seem to be very advantageous platforms to predict successful genomic alterations for a given purpose. However, these tools are as limited as the metabolic acquaintance of a given organism and, therefore, fundamental studies continue to be very important for the construction of novel, more accurate metabolic models and, consequently, for prediction of more correct non-obvious genomic alterations. Moreover, *in silico* models must be viewed as a dynamic and ever improving tool that requires constant tuning and modulation attempting to describe biological metabolic phenomena as accurately as possible.

Nonetheless, metabolic modelling can be used to predict so far unknown metabolic features that may be important when planning a new genome-editing approach for a given purpose. The conjugation of such high- throughput tools renders a much faster scientific design.

2 Background and Objectives

Gene therapy through DNA vaccination has shown to be of grown interest for research studying different novel therapies to treat genetic-related disorders, to promote immunization and to impair bacterial, viral and fungal infections. Although viral delivery vectors are the most widely used for this purpose as they efficiently transduce the genetic material to be expressed in the host cells, the possibility of specific immune responses by the host immune system or integration of viral sequences in the host genome is of great concern. Aiming safe and efficient delivery protocols for human and animal therapy, the development of non-viral delivery vectors was made priority. Non-viral vectors delivery include electroporation or injection of naked DNA and the use of bacterial systems as vectors for transfer genetic material into target cells.

The use of lactic acid bacteria (LAB) as a safe platform for plasmid delivery holds great promise in scientific community, especially when considering mucosal routes of administration, as these food-grade GRAS bacteria endure the passage through the gastrointestinal tract (GIT) of mammals, adhere and colonize the GIT lumen and are LPS free. Another application of LAB can be their use as microbial cell factories for pDNA production, aiming the large-scale production and purification of plasmid DNA for gene therapy purposes. However, one big hurdle of this system is the limited availability of high copy number plasmids for these bacteria. Although changes can be done in plasmids to increase their copy number, another approach, which can be conjugated with the latter, is the editing of the host genome to allow increased pDNA production and quality.

The present work aimed at the attainment of a Lactic Acid Bacterium strain capable of producing high quantity and high quality DNA, either for large scale DNA production and purification, or for alternative gene delivery purposes *in situ*. Thus, the main objective is the knockout of *Lactococcus lactis* LMG 19460 genes that may interfere with pDNA production and quality yield, namely degradation and recombination related genes, and the assessment of increased DNA production by qRT-PCR. The first deletion to be attempted is the *nth* gene, which codes for endonuclease III, responsible for DNA degradation. The system is based on a previous work made in the gram-positive bacteria *Streptomyces coelicolor* M145, using a one-step CRISPR/Cas-mediated protocol for genome editing, in which the genome editing plasmid pKCcas9dO comprising a target-specific guide RNA, a codon-optimized cas9, and two homology-directed repair templates, was used. Therefore, the plasmid nthCRISPRa was previously constructed using pKCcas9dO as backbone with modified *nth* target-specific guide RNA and homology-directed repair templates. The introduction of the nthCRISPRa plasmid into these cells is attempted by electroporation, aiming the expression of the Cas9 system and induction of *nth* deletion by homology repair. The reconstruction of the nthCRISPRa into nthCRISPRe is attempted by Gibson assembly, aiming to change the plasmid selection marker from apramycin to the erythromycin resistance coding gene.

3 Materials and Methods

3.1 Bacterial Strains and Plasmids

The designation and characteristics of bacterial strains, as well as the plasmids used throughout the present work are described in Table 1 and Table 2.

Table 1 Name and main characteristics of bacterial strains used in the present work.

Species and Strain	Relevant Features	Source
Lactococcus lactis LMG 19460	Wild-type and plasmid free.	BCCM Collection
Escherichia coli DH5α	*recA*$^-$ and *endA*$^-$. Produces methylated DNA.	Invitrogen
Escherichia coli SCS 110	*endA*$^-$ Produces unmethylated DNA (*dcm dam*).	Stratagene
Escherichia coli GM 2163	Produces unmethylated DNA (*dcm dam*).	NEB

Table 2 Name and main characteristics of plasmids used in the present work.

Plasmid	Main characteristics	Length (bp)	Source
pTRKH3	Original *E. coli*/LAB shuttle. EryR.	7,766	BCCM/LMBP Plasmid Collection
pKCcas9dO	Bacterial Expression. Thermosensitive origin of replication pSG5. ApmR. Designed for CRISPR-based genome editing.	13,382	Addgene
nthCRISPRa	*nth* flanking arms for homology repair induction. ApmR. Constructed as from pKCcas9dO.	14,146	Duarte, Sofia (unpublished)
nthCRISPRe	*nth* flanking arms for homology repair induction. EryR. Constructed as from nthCRISPRa and pTRKH3.	13,833	The present work

EryR: Erythromycin Resistant Marker. | ApmR: Apramycin Resistant Marker.

3.1.1 Construction of the nthCRISPRa plasmid

The pKCcas9dO plasmid was used by Huang and co-workers for high efficiency genome-editing in *Streptomyces coelicolor* M145 using the CRISPR/Cas9 system[97] and was previously adapted from the BERG lab previous experiments to fit the desirable *nth* knockout in *Lactococcus lactis* LMG 19460 (Duarte, S., unpublished work). The CRISPR/Cas9 genome editing system established by Huang *et al*.[97] for the one-in-one single plasmid approach derived from pKC1139 and consist of a target-specific RNA (sgRNA), a *Streptomyces* codon optimized *cas*9 (Scocas9), driven by the *tipA* promotor for expression *in vivo*, and two homology-directed repair (HDR) templates, homology harm 1 (HA1) and homology harm 2 (HA2). In the designed plasmid (Figure 4), the sequence-specific sgRNA and Scocas9, optimized with the *Streptomyces* codon bias, are expressed in a single plasmid. The *nth* recognition sequence of the sgRNA is 5'-CAGAAGCCTACGGAATTCCG -3' and it is followed by the Cas9 handle and the terminator sequences. The synthetic promotor J23119 is responsible for the transcription of the sgRNA. The homology arms have

1,028 and 955 bp and correspond to the sequence located immediately upstream and downstream of the *nth* gene in the *L. lactis* genome, respectively. The editing plasmid harbours the origin of replication of gram-negative bacteria (*ori*) and the temperature sensitive replicon pSG5, which allows further plasmid curing at 37ºC, facilitating continuous genome editing. The *acc(3)IV* gene, also known as *apmR*, is responsible for the apramycin cell resistance. In order to not interfere in the designed insert, and given the restriction sites available in the pKCcas9dO plasmid, the insertion was carried out using SpeI and ApaI restriction enzymes.

The construction of the fragment to be inserted was made using SOEing PCR. The sgRNA was designed to contain the SpeI restriction site in the 5' end and 20 bp in the 3' end that overlap the 5' region of the HA1. HA1 was amplified with primers that did not contain any overlapping sequence. HA2 primers were designed to amplify a fragment containing an overlapping sequence with the 3' end of HA1 in the 5' end, and the ApaI restriction site in the 3' end. The design of new inserts for future knockouts were made using the same rationale. The genes nucleotide sequences used to design future knockouts correspond to annotated sequences of *L. lactis* LMG 19460, provided by Silva and co-workers.[104] The *nth* knockout design was done previous to the *L. lactis* LMG 19460 genome annotation and, therefore, the closely related *L. lactis* IL1403 gene sequence was used.[105] Nonetheless, an alignment between the two strains *nth* coding sequences and upstream and downstream region (homologous harms) showed no nucleotide dissimilarities.

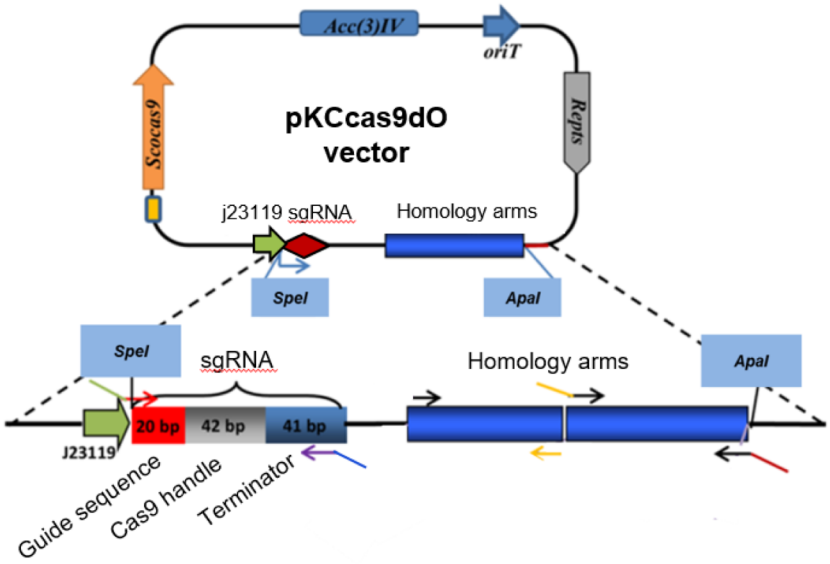

Figure 4 Design and construction of the editing plasmid for CRISPR/Cas9 in Lactococcus lactis. Adapted from Huang et al.[97]

3.1.2 Construction of the nthCRISPRe plasmid

The construction of the nthCRISPRe plasmid was carried out using a Gibson assembly protocol, described in section 3.8.2 after the amplification and purification of two fragments from the nthCRISPRa

plasmid and the erythromycin-encoding sequence (from the erm promoter region until the end of the EryR coding sequence) from the pTRKH3 plasmid. Amplification and purification protocols are described ahead (3.8. and 3.6.4) and the optimization protocol of each amplification is shown in section 4.3. Plasmids nthCRISPRa and nthCRISPRe are shown in Figure 5.

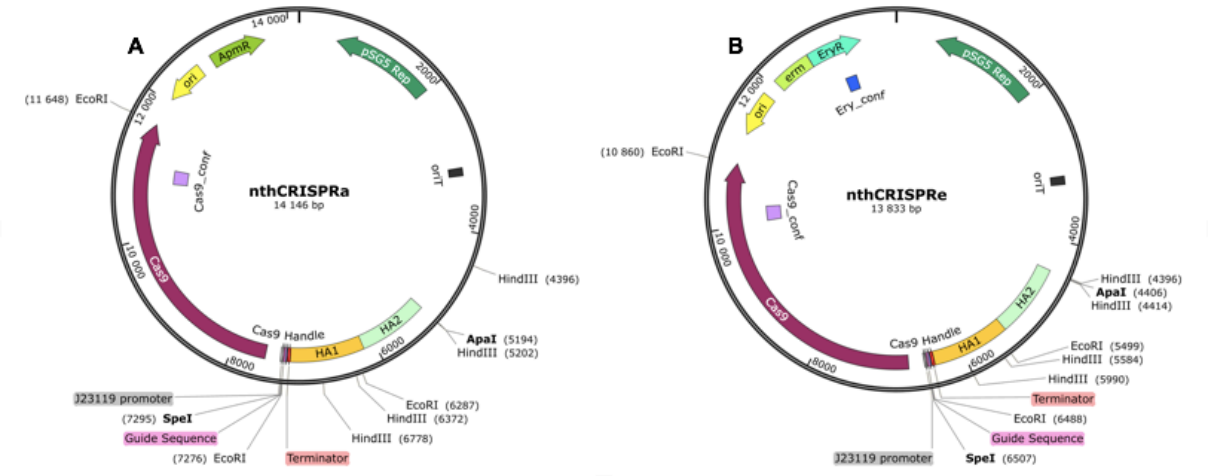

Figure 5 Schematic representation of plasmids nthCRISPRa (A) and nthCRISPRe (B) and its constituents.

3.2 Growth Conditions

3.2.1 *Escherichia coli* strains

The bacterial strains *E. coli* DH5α, *E. coli* GM 2163 and *E. coli* SCS 110 were cultured in Luria Bertani (LB) medium (25 g/L) from NZYTech, composed by tryptone 10 g/L, yeast extract 5 g/L, and NaCl 10 g/L, with pH 7. The growth of *E. coli* containing the pTRKH3 plasmid was made using LB medium supplemented with erythromycin 500 µg/mL, at 37ºC and 250 rpm agitation. *E. coli* strains containing the pKCcas9dO or nthCRISPRa plasmids were grown in LB medium supplemented with apramycin 25 µg/mL, at 30ºC and 250 rpm agitation. Incubations with controlled temperature and agitation were conducted in an ARALAB orbital shaker, model AGITORB 200.

Cells were grown in a pre-inoculum overnight, in 15 mL falcon tubes containing 5 mL of LB medium supplemented with antibiotic. After reaching the exponential growth phase cells were transferred to 15 mL falcon tubes containing 5 mL fresh LB medium supplemented with antibiotic, and incubated at the respective conditions until reaching an OD_{600nm} of approximately 1 for cell banks and late exponential for pDNA purification. Optical density (OD) measurements were performed in a U-2000 spectrophotometer, Hitachi, to assess cell growth using a visible light wave length of 600 nm (OD_{600nm}).

3.2.2 *Lactococcus lactis* LMG 19460

LAB cells were cultured in M17 broth microbiology medium (42 g/L) from SIGMA-ALDRICH supplemented with glucose 0.5% (w/v). M17 medium is composed by ascorbic acid 0.5 g/L, lactose 5 g/L, magnesium sulfate 0.25 g/L, meat extract 5 g/L, meat peptone 2.5 g/L, sodium glycerophosphate 19 g/L, soy peptone 5 g/L, tryptone 2.5 g/L, and yeast extract 2.5 g/L. LAB cells harbouring the pTRKH3 plasmid were grown using M17 broth supplemented with glucose 0.5% (w/v) and 5 μg/mL erythromycin at 37ºC, 250 rpm. Cell growth was conducted in a ARALAB orbital shaker, AGITORB 200.

Cells were grown overnight in a pre-inoculum, in 15 mL falcon tubes containing 5 mL of M17 medium supplemented with glucose and antibiotic. The next morning cells were transferred to new 15 mL falcon tubes containing 5 mL of M17 medium supplemented with glucose and antibiotic, as to start the growth with an OD_{600nm} of approximately 0.1. Cells were incubated at the respective conditions until reaching late exponential phase for pDNA extraction. Cell banks were made after cells reached an OD_{600nm} of approximately 1. OD measurements were performed in a U-2000 spectrophotometer, Hitachi.

3.2.2.1 Spectinomycin-induced pDNA production

Cell bank deriving *L. lactis* LMG 19460 cells containing the plasmid pTRKH3 or nthCRISPRa (a transformation candidate) were cultured overnight in 15 mL falcon tubes containing 5 mL of M17 medium (42 g/L) supplemented with glucose 0.5% (w/v) and erythromycin 5 μg/mL or apramycin 250 μg/mL, respectively. As they reached stationary growth, cells were transferred to new 15 mL falcon tubes containing 5 mL of M17 medium (42 g/L) supplemented with glucose 0.5% and erythromycin 5 μg/mL or apramycin 250 μg/mL, depending on the containing vector, in order to start the culture at an OD_{600nm} of 0.1. Cultures were grown until reaching mid-exponential phase, with as OD_{600nm} of approximately 2. The falcon tubes containing cultures were supplemented with spectionomycin 10 mg/mL and left for incubation for 16 hours, after which the plasmid was purified. Cells harbouring the pTRKH3 vector were incubated at 37ºC, 250 rpm, while candidate cells for nthCRISPRa transformation were grown at 30ºC, 250 rpm.

3.3 Cell banks

After growth protocols, bacterial cell banks, excluding competent cell banks (explained below), were prepared using 80 µL cell suspension solution and 20 µL glycerol 99.99%, in 1.5 mL microcentrifuge tubes. Promptly, the 100 µL cell preparations were kept on ice until storage at -80ºC.

3.4 Cell transformation

3.4.1 Chemically competent *E. coli* DH5α cells

For the preparation of chemically competent *E. coli* cells, an overnight growth in LB broth (25 g/L), at 37ºC, 250 rpm, was performed and cells were transferred to a 100 mL Erlenmeyer flask containing 20

mL LB broth (25 g/L), as to start the culture at an OD_{600nm} of 0.1. Cells were then grown as to reach an OD_{600nm} of 1 and were posteriorly centrifuged at 1,000 × g, for 10 minutes at 4ºC. Supernatants were discarded and the pellets were resuspended in 2 mL sterile (0.22 μm filter) TSS, composed by LB (25 g/L), DMSO 5% (v/v), $MgCl_2$ 50 mM and PEG 8000 10% (w/v) with pH 6.5. After the preparation of 100 μL aliquots, cell preparations were kept on ice for 10 minutes, before storage at -80ºC.

Chemically competent *E. coli* DH5α cells were transformed by heat shock. Each aliquot of competent cells was incubated with the desired DNA mass in the desired volume on ice for 30 minutes. The mixture was then placed on a 42ºC dry bath for 1 minute, and on ice for 2 minutes, immediately after. The cell mixture was then resuspended with 900 μL LB broth (25 g/L) and left for incubation at 30ºC, 250 rpm, foe 1 hour. Ultimately, these cells were plated in LB agar medium (40 g/L), from NZYTech, composed by agar 15 g/L, tryptone 10 g/L, yeast extract 5 g/L, and NaCl 10 g/L, supplemented with erythromycin 500 μg/mL and incubated at 30ºC for 48h-60h. Some exceptions regarding the antibiotic concentration are referred throughout the present work.

Transformation candidates were grown as explained in section 3.2.1 until visible growth. Cell banks were made as explained in 3.3 after cell growth reached an OD_{600nm} of ~1, if possible. Plasmid DNA was purified later after cells reached a higher OD_{600nm} value, if possible. Some transformation candidates did not exhibit an OD_{600nm}>1 and pDNA extraction was proceeded anyway.

3.4.2 Electrocompetent *E. coli* DH5α cells

E. coli DH5α cells were grown as explained in 3.4.1 until they reached an OD_{600nm} of 1. After a 3-minute centrifugation step at 6,000 × g, 4ºC, cells were washed three times with 1 mL deionized water and finally resuspended with 100 μL of a deionized water solution containing 20% glycerol. Similarly, after the preparation of 50 μL aliquots, cell preparations were kept on ice for 10 minutes, before storage at -80ºC.

Electrocompetent *E. coli* DH5α cells were transformed by electroporation. Each cell aliquot together with the desired DNA mixture was transferred to 2 mm electroporation cuvettes from BTX and incubated for 30 minutes on ice. Posteriorly, the cuvettes were dried with laboratory paper tissue and an electric field of 2,500 V (12.5 kV/cm) was applied during 5 ms using the pulse controller ECM 399 Electroporation System, BTX. The cell mixture was then resuspended with 950 μL LB broth (25 g/L) and incubated at 30ºC, 250 rpm. Ultimately, these cells were plated in LB agar medium (40 g/L) supplemented with erythromycin 250 or 500 μg/mL and incubated at 30ºC for 48h-60h.

Electroporation candidates were grown as explained in section 3.2.1 until visible growth. Cell banks were made as explained in section 3.3. after cell growth reached an OD_{600nm} value of ~1 and plasmid DNA was purified later after cells reached a higher OD_{600nm} value, if possible. Transformation candidates for which cell growth did not reach an OD_{600nm}>1 were also saved as cell banks and pDNA purification was proceeded.

3.4.3 Electrocompetent *L. lactis* LMG 19460 cells

Lactococcus lactis electrocompetent cells were obtained using a growth protocol with glycine in osmotically stabilized media. After an overnight growth in 5 mL M17 broth (42 g/L), supplemented with 0.5% (w/v) glucose, at 30ºC, 100 rpm, in a 15 mL falcon tube, cells were transferred to a 100 mL Erlenmeyer flask containing 75 mL M17 broth (42 g/L) supplemented with 0.5% (w/v) glucose (GM17 medium), as to start the culture at an OD_{600nm} of 0.1. After the culture reached an OD_{600nm} between 0.5 and 0.8, a 100-fold dilution was made into a new Erlenmeyer flask containing M17 broth (42 g/L), 0.5% (w/v) glucose, 0.5 M sucrose and glycine 2% (SGGM17 medium). For this 100-fold dilution, 750 µL of cell suspension grown in GM17 was transferred into a 100 mL Erlenmeyer flask containing 75 mL SGGM17 medium. Later one this electrocompetent protocol was also preformed using SGGM containing glycine 1%. Cells were grown in SGGM17 until reaching an OD_{600nm}>1. Afterwards, cells were centrifuged at 5,000 $\times$ g, 4ºC, for 10 minutes and then washed twice with a 1 mL water washing solution, composed by sucrose 0.5 M and glycerol 10% (v/v). Cells were finally resuspended in a particular volume of washing solution in order to have 8.89×10^{10} cells/mL in each 100 µL cell bank. After the preparation of 100 µL aliquots, cells were kept on ice until storage at -80ºC.

Transforming Lactic Acid Bacteria may be a difficult task, as different strains have different requirements and different parameters must be applied when using different DNA masses. Competent cells aliquots were diluted in a 1:3 ratio using PCR grade water and divided into 40 µl aliquots, which were mixed with the desired DNA and incubated on ice for 30 minutes in 1 mm electroporation cuvettes from BTX. The transformation process was conducted using a pulse controller BTX ECM 399 electroporator, and the conditions used for such are shown in Table 3. After the electroporation procedure, cells were immediately resuspended in 960 uL recovery medium, composed by M17 broth (42 g/L), glucose 0.5% (w/v), Sucrose 0.5 M, $MgCl_2$ 20 mM and $CaCl_2$ 2 mM, and incubated in 30ºC, without agitation, for 3 hours. Afterwards, these cells were pelleted for 3 minutes at 6000 $\times$ g, 4ºC and the pellet was transferred to 15 mL falcon tubes containing 5 mL of M17, supplemented with glucose 0.5% (w/v) and 250, 300 or 500 µg/mL apramycin, in which they grew overnight, in a Memmert incubator at 30ºC. After, cells were incubated at 30ºC on microplates containing complex agar medium constituted by tryptone 10 g/L, yeast extract 5 g/L, sucrose 200 g/L, glucose 10 g/L, gelatin 25 g/L, agar-agar 15 g/L, $MgCl_2.6H_20$ 0.5 g/L and $CaCl_2$ 0.3 g/L, supplemented with apramycin 500, 1 000 or 2 000 µg/mL, until visible growth (48-72 hours).

Throughout electroporation tests, nthCRISPRa plasmid from three different *E. coli* strains DH5α, *SCS 110 and GM 2163* were used as to see whether the transformation of methylated and non-methylated DNA had an influence in transformation efficiencies, DNA copy number or quality. As transformations were not achieved, DNA origin is not specified.

Table 3 Electroporation conditions and antibiotic concentration selection of L. lactis LMG 19460.

Number of cells in each transformation aliquot	pDNA mass (ng)	Electric pulse conditions (# of pulses × kV/cm) during 5 ms	Apramycin concentration (μg/mL)
1.185×10^{10}	10	3 × 10	500 and 1 000
	100	1 × 10	500 and 1 000
		3 × 10	
		3 × 20	1 000 and 2 000
	250	3 ×10	500
	500	1 × 10	500 and 1000
		3 × 10	

3.5 Relation between optical density and cell number

To ascertain the number of cells in a given growth broth, the relation of OD_{600} = 0.1 or 7×10^7 cells, used by Jones *et al.*[106] for *L. reuteri* cell counting, was also used in the present work. This relation was confirmed for *L. lactis* LMG 19460 in the laboratory by Martins, M.[107]. In the present work, it was used to determine the resuspension volume of washing solution necessary for 100 mL cell banks of *L. lactis* LMG19460 electrocompetent cells to have 8.89×10^{10} cells/mL (3.4.3) and to determine the necessary pellet volume and dilutions for cell PCR reactions (3.7.3).

3.6 DNA purification, quantification and quality assessment

3.6.1 Purification of plasmid DNA from *E. coli* DH5α

Following cell growth, either from cell banks or from transformant growth protocol, plasmid DNA was purified using the NZYMiniprep kit from NZYTech following the manufacturer's instructions. Some manufacture recommendations were performed: using double the A1, A2 and A3 buffer volumes and preheat the elution solution until 70ºC. For DNA purification procedures, the elution buffer used was PCR grade water (filtered and autoclaved). The use of PCR grade water, instead of buffer AE in the elution step avoids inhibitory effects that may happen in PCR reactions and circumvents electric arcs during electroporation due to the elution buffers' salt composition. Purified pDNA mixtures were stored at 4ºC until use.

3.6.2 Purification of plasmid DNA from *L. lactis* LMG19460

Following cell growth, either from cell banks or from transformant growth protocol, plasmid DNA was purified based on the Plasmid DNA Purification user manual of the Nucleo Spin® Plasmid, DNA, RNA and protein purification kit, from Macherey Nagel Bioanalysis. The first two steps of purification were based on

"Isolation of plasmids from Gram-positive bacteria", which includes a lysis step with lysozyme 10 mg/mL to aid cell wall lysis. For the remaining purification steps the protocol "Isolation of low-copy plasmids" was considered.

3.6.2.1 Alternative lysis step: boiling

The harvesting of cells from growth cultures was made by centrifugation at 4 000 × g and 4ºC for 15 minutes. The pelleted cells were resuspended in 1 mL of "ice cold" STE solution, containing 0.1 M NaCl, 10 mM Tris.Cl (pH 8) and 1 mM EDTA (pH 8). The resuspended pellet was then re-centrifuged using the same conditions. The lysis by boiling was then performed as an adapted protocol from Holmes and Quigley.[108] The pellet was resuspended in 100 μL STET containing 0.1 M NaCl, 10 mM Tris.Cl (pH 8) and 1 mM EDTA (pH 8) and 5% (v/v) Triton X-100. The resuspended pellet was transferred to an Erlenmeyer and 10 μL of lysozyme 10 mg/mL, prepared with 10 mM Tris.Cl (pH 8), was added. The samples were boiled under a Bunsen burner under continuous manual stirring. After the samples started to boil, erlenmeyers were immerged in a 99ºC wet bath for 40 seconds. The samples were then cooled by placing the Erlenmeyers in "ice-cold water" for 5 minutes. The viscous content was transferred to centrifuge flasks and centrifuged at 13,000 rpm, 4ºC, for 30 minutes. The remaining purification steps were made as in 3.6.2, excluding the lysis step.

3.6.3 Purification of genomic DNA from *L. lactis* LMG 19460

For *L. lactis* genome extraction, the protocol "Isolation of Genomic DNA from Gram Positive and Gram Negative Bacteria" from Wizard® Genomic DNA Purification Kit, Promega, using the Gram-positive bacteria specifications, was used. However, in the last purification step, the rehydration of the DNA pellet was done using PCR grade water, instead of Rehydration Solution, as to minimize the inhibitory effects that these solutions may have in PCR enzymes in following PCR procedures.

3.6.4 PCR clean up and purification of DNA from TAE agarose gels

In order to purify DNA amplified from PCR reactions, the NZYGelpure kit from NZYTech was used based on the "Protocol for PCR clean-up or DNA purification from enzymatic reactions". DNA originated from PCR reactions that could not be purified using the PCR clean up protocol, due to unspecified amplifications, had to be purified using the same kit, but using the "Protocol for plasmid DNA purification from Agarose Gels". Once more, the elution step of both protocols was made using PCR grade water rather than Elution Buffer.

3.6.5 Quantification of purified DNA products

Following the DNA purification protocol, DNA concentration was assessed using the NanoVue Plus from GE Lifesciences. The purification degree was determined considering the absorbance values (A) ratio between DNA and protein and DNA and salts and solvents, measured in the wavelengths for which these components have maximum absorption (A(260/280) and A(260/230), respectively). To confirm the quality

of the extracted DNA and to make sure that the DNA was present in the purified mixture, 500 ng to 1 μg of DNA was run in agarose gels for visualization.

3.6.6 Digestion of nthCRISPRa

To access DNA quality and integrity of the purified plasmid nthCRISPRa, a digestion using the Promega restriction enzyme EcoRI, with restriction site G^AATTC, was conducted. This enzymes's action leads to the digestion of the plasmid into three different fragments with 8,785 bp, 4,372 bp and 989 bp. The digestion was performed using 500 ng of nthCRISPRa, 0.3 μL (12 U/μL) EcoRI, 1.5 μL 10x buffer H and PCR grade water, making a total volume of 15 μL. The mixture was incubated at 37ºC for 1 hour and 30 minutes. Ultimately, the digested samples were run in 1% agarose gels to confirm if the fragments size were the ones expected.

3.7 PCR amplifications

3.7.1 *E. coli* DH5α colony PCR

E. coli colony PCR was conducted to confirm the presence of nthCRISPRe plasmid after cell transformation. To do so, the primers Ery_conf and Cas9_conf (Table 4) were either used to amplify a region inside the erythromycin resistance gene or the Scocas9 gene, respectively. None of these genes make part of the *E. coli* DH5α genome. The transformation cfu candidates, which have grown in the solid medium supplemented with erythromycin, were picked to a 1.5 mL Eppendorf containing 70 μL PCR grade water. After careful resuspension, 50 μL of the cell mixture was incubated in 99ºC for 5 minutes, to promote cell rupture. The remaining 20 μL were kept under sterile conditions in 4ºC to be either grown in liquid medium or solid medium supplemented with erythromycin, if PCR amplifications were positive and had the expected base pair length. After the heat treatment, cells were centrifuged at 12,000 × g, 4ºC for 1 minute. 10 μL of supernatant were used for the PCR reaction, which was carried out with the NovaTaq™ Hot Start Master Mix kit, from Merck Millipore. Each 25 μL reaction contained 10 μL of supernatant (containing the pDNA, if present), 12.5 μL master mix (final concentration of 1×), 1 μL of each forward and reverse primers (0.4 μM final concentration) and 0.5 μL PCR grade water. The PCR conditions are shown in Table 5.

Table 4 Primer designation, sequence, melting temperature (Tm) and product size used in colony PCR for transformation confirmation purposes.

Primer designation	Primer sequence (5'-3')	Primer Tm (ºC)	PCR product size (bp)
Ery_conf_forward	CCATGCGTCTGACATCTATCT	55.2	190
Ery_conf_reverse	CTGTGGTATGGCGGGTAAGT	55.2	
Cas9_conf_forward	CCCAGGTCAACATCGTCAAG	57	241
Cas9_conf_reverse	TCCATGATGGTGATGCCCAG	58	

Table 5 PCR conditions for amplifications using Ery_conf and Cas9_conf primers.

PCR step	Ery_conf		Cas9_conf	
Activation	95ºC, 2 minutes		95ºC, 2 minutes	
Melting	35 cycles	95ºC, 30 seconds	40 cycles	95ºC, 30 seconds
Anneling		55ºC, 30 seconds		56ºC, 30 seconds
Extension		72ºC, 30 seconds		70ºC, 3 minutes
Final Extension	72ºC, 10 minutes		-	

3.7.2 PCR amplification for confirmation purposes using purified DNA as template

To ascertain the *nth* genomic knockout in *L. lactis* LMG 19460, the primers Nth_conf (Table 6) that prime the upstream and downstream flanking regions of the homology harms sequences of the *nth* gene in the LAB genome were used. If the knockout had happened, the amplicon would be 657 bp smaller than the amplicon containing the intact *nth* gene (2,790 bp). Figure 6 shows a schematic representation of the genomic region amplified by Nth_conf, in negative and positive *nth* knockout scenarios (w/o KO and w/ KO, respectively). This PCR procedure was done using purified genomic DNA, before testing the cell PCR protocol. The amplification programme is shown in Table 7.

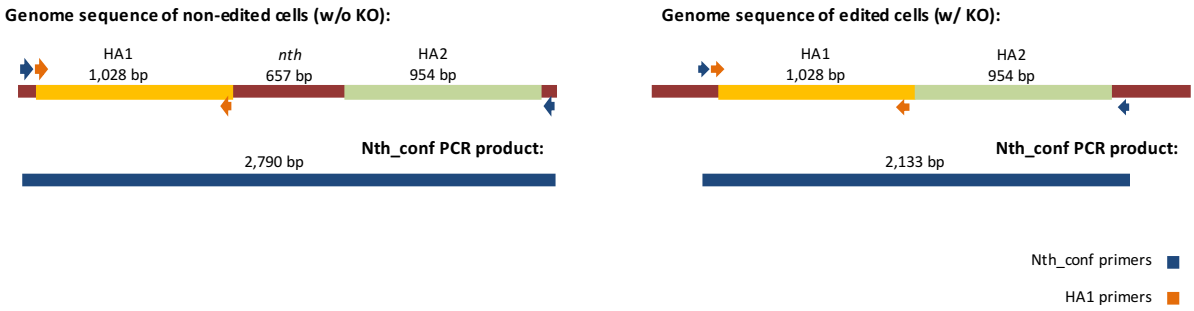

Figure 6 Schematic representation of L. lactis LMG 19460 genomic region of nth gene and Nth_conf amplicon, comprising negative and positive nth gene knockout events. Primers are represented by arrows.

Early in the present work, the presence of the nthCRISPRa plasmid in transformed and grown *L. lactis* LMG 19460 cells was assessed using the HA1 primers (Table 6), which amplify the homology arm 1, downstream of the *nth* sgRNA in the nthCRISPRa plasmid (Figure 5A). As the HA1 sequence was also present in the *L. lactis* LMG 19460 genome, upstream of the *nth* gene (Figure 6), a purified mixture of pDNA was used as template.

Both reactions were performed using KOD Hot Start DNA Polymerase kit, from Merck Millipore and each 25 μL reaction mixture contained 2.5 μL buffer 10× (final concentration 1×), 2.5 μL dNTP mixture (final concentration 0.2 mM each), 2.3 μL $MgSO_4$ (final concentration 2.25 mM), 0.8 μL of each forward and reverse primers (final concentration 0.32 mM each), 0.5 μL KOD Hot Start Polymerase (final concentration

0.02 U/µL), DNA and water. For genomic (Nth_conf primers) and plasmid (HA1 primers) amplifications, 100-200 ng and 10-100 ng DNA mass were used, respectively. The PCR programmes are shown in Table 7.

Table 6 Primer designation, sequence, melting temperature (Tm) and product size used in confirmation PCRs.

Primer designation	Primer sequence (5'-3')	Primer Tm (ºC)	PCR product size (bp)	
Nth_conf_forward	CAAGGGCAAGTCAAATATAC	49	w/o KO	2,790
Nth_conf_reverse	GTATGGCTGGTATAGACAGCA	54	w/ KO	2,133
HA1_forward	GAACCGTTGTAGAGTAAGTC	40	1,028	
HA1_reverse	CGTGCACGTCCATTATCTCT	50		

Table 7 PCR conditions for amplifications using Nth_conf and HA1 primers using KOD Hot Start Polymerase.

PCR step	Nth_conf		HA1	
Activation	95ºC, 2 minutes		95ºC, 2 minutes	
Denaturation	35 cycles	95ºC, 30 seconds	35 cycles	95ºC, 30 seconds
Annealing		53ºC, 30 seconds		65ºC, 30 seconds
Extension		70ºC, 3 minutes		70ºC, 1.5 minutes
Final Extension	-		-	

3.7.3 *L. lactis* LMG 19460 cells PCR

Extracting pDNA and gDNA to perform PCR confirmations can be time consuming, especially when there is a need to analyse a large set of transformant candidates. Here a protocol using pelleted cells was used and the number of cells for a clear amplification was optimized. Contrary to the colony PCR used for *E. coli* DH5α, this cell PCR does not require the 5 minutes lysis step. The activation step seems to be sufficient to disrupt the cell wall. Purified nthCRISPRa (extracted from *E. coli*) and *L. lactis* genome DNA were used as controls. The same PCR protocol amplified both Scocas9 internal sequence and the sequence amplified by Nth_conf primers, in the control sequences. Control samples of purified pDNA (extracted from *E. coli*) and LAB gDNA used were 1 ng and 200 ng, respectively. Three different annealing temperatures were tested (50ºC, 53ºC and 56ºC), and 56ºC was chosen for further amplifications. The number of cells chosen to be the best for this PCR protocol was 250 000, the one with a cleaner amplification profile (with a clear expected PCR product and less smear). The number of cells tested for the PCR were 1,000, 10,000, 25,000, 50,000 cells and 250,000. The PCR was carried out using the NovaTaq™ Hot Start Master Mix kit, from Merck Millipore. Each 25 µL reaction contained 12.5 µL master mix (final concentration 1×), 1 µL of each forward and reverse primers (0.4 µM final concentration), the desired volume of diluted pelleted cells and the remaining volume of PCR grade water. Cas9_conf and Nth_conf primers are shown in Table 4 and Table 6, respectively. The PCR programme is shown in Table 8.

A confirmation PCR test was also performed in *L. lactis* LMG 19460 cells using Ery_conf primers and the conditions shown in Table 5, with the exception that all three steps of the cycling programme lasted for 1 minute, instead of 30 seconds.

Table 8 PCR conditions for amplifications using Cas9_conf and Nth_conf primers in L. lactis LMG 19460 cells.

PCR step	Cas9_conf and Nth_conf	
Activation	95ºC, 2 minutes	
Denaturation	40 cycles	95ºC, 30 seconds
Annealing		56ºC, 30 seconds
Extension		70ºC, 3 minutes
Final Extension	-	

3.8 Reconstruction of nthCRISPRa plasmid into the nthCRISPRe plasmid

3.8.1 PCR amplifications for reconstruction of the nthCRISPRa plasmid into nthCRISPRe

Aiming to change the selection marker sequence, the reconstruction of nthCRISPRa into nthCRISPRe was performed. Two fragments (Apa_Ery and Ery_Apa) were amplified by PCR from nthCRISPRa and the fragment containing the erythromycin resistance gene and promotor sequence (erm cassette) was amplified by PCR from pTRKH3. To amplify the desired products, a long optimization procedure described in 4.3 was performed. Primers and respective PCR products are shown in Table 9. Primers were designed using the available online software NEB Builder Assembly tool v1.12.16.[109] Highlighted nucleotides correspond to overlapping sequences for posterior Gibson ligation protocol. Ery_Apa2 primers were designed to have ten additional nucleotides in the seeding region (non-overlapping region), which correspond to the lowercase nucleotide representative letters in Table 9.

PCR reactions were performed until proper amplification patterns were obtained. The optimization rationale and experiments are shown in 4.3 and the final PCR conditions are shown in Table 10. After the optimization process, 5 ng of template were used for the Ery, Apa_Ery and Ery_Apa2 amplifications. The Ery_Apa NovaTaq Hot Start Mastermix amplification was performed using 20 ng of DNA template. The Ery and Apa_Ery fragments were amplified using KOD Hot Start Kit, the Ery_Apa using the Nova Taq Hot Start Master Mix and the Ery_Apa2 using the KOD Hot Start Master Mix. The first three PCR reaction mixtures were prepared as described before for each polymerase kit. The 25 μL mixture of the KOD Hot Start Master Mix contained 10 μL master mix, 0.6 μL of each forward and reverse primers (0.24 μM final concentration), the desired volume of DNA mixture and the remaining volume of PCR grade water.

The first three PCR reaction mixtures were prepared as described before for each polymerase kit. The 25 μL mixture of the KOD Hot Start Master Mix contained 10 μL master mix, 0.6 μL of each forward and

reverse primers (0.24 μM final concentration), the desired volume of DNA mixture and the remaining volume of PCR grade water.

Table 9 Primer designation, sequence, melting temperature (Tm) and PCR product length for amlification of sequences for Gibson assembly.

Primer designation	Primer sequence (5'-3')	Primer Tm (ºC)	PCR product size (bp)
Ery_forward	TGATCGACTGAGTCTAGAATCGATACGATTTTG	57.9	1,299
Ery_reverse	GCTCATGAGCTTATTTCCTCCCGTTAAATAATAG	57.9	
Apa_Ery_forward	TGCCAAGCTTGGGCCCATATATAAGCTTC	60.3	7,686
Apa_Ery_reverse	ATTCTAGACTCAGTCGATCATAGCACGATC	60.3	
Ery_Apa_forward	GAGGAAATAAGCTCATGAGCGGAGAACG	67.1	4,908
Ery_Apa_reverse	ATATGGGCCCAAGCTTGGCACTGGCCGT	67.1	
Ery_Apa2_forward	GAGGAAATAAGCTCATGAGCGGAGAACGagatgacgtt	65.1	4,908
Ery_Apa2_reverse	ATATGGGCCCAAGCTTGGCACTGGCCGTcgttttacaa	69.6	

Table 10 PCR conditions for amplifications of the nthCRISPRe constituent fragments.

Fragment	DNA Polymerase	PCR step	Respective primers	
Ery	KOD Hot Start	Activation	40 cycles	95ºC, 2 minutes
		Denaturation		95ºC, 1 minute
		Annealing		59ºC, 1 minute
		Extension		70ºC, 2 minutes
Apa_Ery	KOD Hot Start	Activation	40 cycles	95ºC, 2 minutes
		Denaturation		95ºC, 1 minute
		Annealing		62ºC, 1 minute
		Extension		70ºC, 8 minutes
Ery_Apa	Nova Taq Hot Start Master Mix	Activation	40 cycles	94ºC, 2 minutes
		Denaturation		94ºC, 30 seconds
		Annealing		62ºC, 30 seconds
		Extension		72ºC, 5 minutes
Ery_Apa2	KOD Hot Start Master Mix	Activation	40 cycles	95ºC, 2 minutes
		Denaturation		95ºC, 30 seconds
		Annealing		54ºC, 30 seconds
		Extension		70ºC, 2 minutes and 15 seconds

3.8.2 Gibson Assembly

Prior to DNA ligation protocol, the DpnI enzyme from Promega was used to digest methylated template DNA, leaving PCR amplified products intact. This digestion aimed the degradation of template DNA, either pTRKH3 or nthCRISPRa, in order to not interfere in subsequent transformations. Digestion was done by adding 1 μL (10 U/μL) DpnI to 10 μL of each PCR reaction product mixture and incubating the mixture at 37ºC for 1 hour and 30 minutes. DpnI inactivation was accomplished by submitting the samples to 80ºC for 20 minutes.

Aiming the ligation of the amplified and purified PCR products Ery, Apa_Ery and Ery_Apa or Ery_Apa2, the protocol of NEBuilder HiFi DNA Assembly Master Mix/NEBuilder HiFi DNA Assembly Cloning Kit Instruction Manual was used. Ligations using different molarities and masses were performed and are shown in Table 11. In each 20 μL reaction volume, DNA fragments volume accounted for a maximum of 10 μL and the other 10 μL for NEBuilder HiFi DNA Assembly Master Mix. PCR grade water was added to the fragments mixture, until a total volume of 10 μL was reached.

Gibson reactions #1 to #4 were performed using the Apa_Ery fragment mixture, purified from TAE agarose gel and the Ery_Apa fragment amplified using the low-fidelity NovaTaq Polymerase, and purified from TAE agarose gel. Gibson reactions #5 and #6 were performed using the Apa_Ery fragment mixture obtained by PCR cleanup, and the Ery_Apa fragment was amplified using the high-fidelity KOD Hot Start Master Mix, and purified from TAE agarose gel.

Table 11 DNA masses (in ng and pmol) used for different attempts of Gibson assembly-based ligation of the cited fragments.

Fragments	Size (bp)	Gibson #1		Gibson #2-#5		Gibson #6	
		Mass (ng)	Mass (pmol)	Mass (ng)	Mass (pmol)	Mass (ng)	Mass (pmol)
Ery	1,299	120.00	0.142	120.00	0.142	84.44	0.100
Apa_Ery	7,686	60.00	0.012	80.00	0.016	250.00	0.050
Ery_Apa	4,908	60.00	0.019	40.00	0.013	159.10	0.050
Total Mass			0.173		0.171		0.200

Gibson #1 was performed under the recommendations of the protocol, using the DNA mass ratio of vector:insert=1:2, assuming the larger fragments (Apa_Ery and Ery_Apa) as a vector fragments and the shorter fragment (Ery) as the insert. Gibson #2, #3, #4 and #5 were made under the assumption of Apa_Ery and Ery_Apa as a unique vector fragment, hence the approximate equimolarity (pmol per reaction volume). The double of the larger fragments mass mean was used to determine the Ery fragment mass. Gibson #6 was made using the DNA molarity ratio of vector:insert=1:2. Every Gibson protocol took into account the maximum molarity of 0.2 pmol for each ligation protocol.

Samples were incubated at 50ºC and taken and stored after 15 minutes, 1 hour, 2 hours and 4 hours of incubation, aimed at further *E. coli* DH5α transformation. Either 2 μL or 1 μL were used for chemical transformation or electroporation, respectively.

3.9 Gel Electrophoresis

Gel electrophoresis was performed to visualize DNA after purification or from PCR origins, by separating the products according to their size. Agarose gels were prepared using 1 × concentrated TAE buffer and Seakem LE agarose (Lonza) to obtain 1% concentrated gels. The molecular weight marker NZYDNA ladder III was used to infer samples' base pair number. Electrophoresis was performed at 100 V for 1 hour, when using small gels and at 120 V for 1 hour and 30 minutes when using bigger gels. The DNA was stained with ethidium bromide for 15 to 40 minutes, depending on the agents' stability. Visualization was made on the Eagle Eye II image acquisition system (Stratagene).

3.10 OptFlux Simulation

A briefly metabolic modelling approach was conducted using the OptFlux 3.3.3 software, aiming the discovery of new knockout targets for pDNA optimized production. To this mean, the iAP358 metabolic model from *Lactococcus lactis*, subspecies lactis Il1403 was imported to the modelling software. This model was the one available with more similarity to what should be the metabolic map of *L. lactis* LMG19460. The critical genes/reactions were determined and the "Optimization-Evolutionary" tool was used to predict new possible knockouts for DNA production optimization. For the simulations, the maximum number of solutions evaluations was limited to 5,000 and the maximum number modifications were chosen to be 6. Critical and drain reactions were excluded from the simulation. The used method was "SPEA2 reaction knockout" and the objective function setup was determined as "BPCY: Biomass-Product Coupled Yield". The chosen product whose production would be optimized was nrdD_1, involved in the pyrimidine and purine synthesis pathways. A series of possible genes to be knocked out in a near future in *L. lactis* was obtained.

4 Results and Discussion

4.1 Potential gene knockouts for increased plasmid production and quality

4.1.1 OptFlux *Lactococcus lactis* Metabolic Optimization using the iAP358 Metabolic Model

The 2005 metabolic model iAP358 of the strain species *L. lactis* IL1403[110] was used to predict potential knockouts considering a given objective function in the OptFlux software. Although not being the same strain, *L. lactis* IL1403 is the most phylogenetically closed organism to *L. lactis* LMG 19460, among other *L. lactis* species.[111]

The simulation of critical reactions is extremely helpful in the sense that it allows the user to exclude gene/reaction deletions that would be lethal to the cell, decreasing the vast amount of generated data that the human eye has to critically process. The "Optimization- Evolutionary" tool allows the optimization of a determined selected feature by different methods, which include reaction knockout and reaction under/over expression. This software allows the user to choose the number of possible combinations of alterations for the given objective function. Using all the data in metabolic models would be an almost impossible thing to solve, mathematically speaking, even for high processing computers. In order to solve this problem OptFlux uses an algorithm that makes all possible combinations of gene/reaction alterations within a group of limited number. The number of solutions in that group is defined by the user before the simulation. Therefore, different simulations using exactly the same parameters can follow different outputs and solutions, as the system "randomly" changes the group composition. It seems that different simulations using the same parameters can be advantageous in the sense that solutions that are more likely to promote a given function could appear more times and in different combinations. In addition to critical reactions, the drain and transport reactions were also excluded from the simulation, because transport reactions usually play important roles in the cell stress response, nutrient uptake, signalling pathways, etc., and, although these "reactions" may be virtually possible, in practice, could have nefarious consequences for cell growth and manipulation. The chosen objective function setup was the BPCY and the chosen "reaction" to be optimized (product) was the *nrdD* gene product. The *nrdD* gene codes for an anaerobic ribonucleoside-triphosphate reductase (EC number 1.17.4.2)[112] and it is involved in the pyrimidine and purine metabolic pathways, responsible for the dCTP, dGTP, dATP and dGTP production, which leads to DNA synthesis. This was the closest product to the DNA synthesis pathway that one could find in the software and have results, which does not mean that other genes/reactions could not offer better results.

After a couple simulations using the same parameters, the solutions that were more common between knockout predictions were chosen and inferred about their function (Table 12). The biological significance standpoint of each proposed deletion was analysed and their function information was taken from NCBI data base[111] and Bolotin *et al*[105].

Table 12 List of chosen genes/reactions outputted for knockout using Biomass-nrdD Coupled Yield in OptFlux Software. Gene functions were assessed using the NCBI ID number.

Gene/Reaction	Name and Function	NCBI ID number
araT	Aromatic amino acid aminotransferase: constitutes the first step in aromatic amino acid degradation in lactococci.[113]	1113660
nucA	Nucleotidase: nucleotide and nucleoside interconversions.	1114734
yjhF	Phosphoglycerate mutase (glycolysis pathway).	1114580
gltB	Glutamate synthase large subunit: glutamate family; nitrogen metabolism; brings together nitrogen and carbon metabolism.	1114935
nrdF	Ribonucleoside-diphosphate reductase beta chain (pyrimidine pathway): catalyzes the rate-limiting step in dNTP synthesis; converts nucleotides to deoxynucleotides.	1114605
adhE	Bifunctional acetaldehyde-CoA/alcohol dehydrogenase (energy metabolism; fermentation).	1115832
argH	Argininosuccinate lyase: catalyzes the formation of arginine from (N-L-arginino) succinate (arginine biosynthesis).	1113731
argG	Argininosuccinate synthase: catalyzes the formation of 2-N(omega)-(L-arginino) succinate from L- citrulline and L-aspartate in arginine biosynthesis, AMP-forming (arginine biosynthesis).	1113730
purH	Bifunctional purine biosynthesis protein PurH: *de novo* purine biosynthesis.	1115175

Most of these genes/reactions are involved in two main pathways: glycolysis/gluconeogenesis (Figure S 1 (Spplemental Data)) and purine and pyrimidine metabolism (Figure S 2 (Supplemental Data)). These knockout proposals can be explained by the need to channel carbon pools to DNA synthesis, instead of fermentative pathways, for example. The knockout proposal of *purH* gene/reaction does not seem realistic as this gene/reaction is directly involved in *de novo* purine biosynthesis by directing carbon coming from de pentose phosphate pathway to other pathways that lead to DNA production, including the ones in which nrdD is involved.[114] In principle, it does not seem a good approach to knockout the *purH* gene in a strain aiming the increase of DNA production. This particular result evidences the importance of human eye analysis and judgment when working with simulation programs.

It is also important to denote that most of the genes/reactions here proposed are involved in other metabolic pathways and that their knockout could result in pleiotropic effects. As the objective function of OptFlux in the present work is the couple yield of biomass and nrdD reaction product, the simulator does not consider other pathways that may be affected. Another point to be taken into consideration is that these kinds of metabolic modulations oversee genes as reactions and does not know non-translated elements and, therefore, do not have genic regulation issues into account, which can be extremely important in real life application.

The obtained proposed gene deletions could be potentially used in further strain optimization aiming to obtain a highly DNA productive bacteria, suitable for gene therapy applications. It is to consider that the metabolic model here used is from 2005 and may not be fully updated. These tools are as limited as their

metabolic acquaintance and the detail that it has been put when designing and programing them. Therefore, some simulation outputs may not correspond to what would happen in reality. Nonetheless, OptFlux constitutes a very promising tool for knockout prediction for a given purpose and the development of new metabolic models considering other issues than reactions are much awaited in the scientific community.

From a theoretical perspective, the target genes for knockouts, aiming the increase of pDNA yield production and quality, should include genes that are associated with DNA degradation and unspecific recombination, such as endonuclease and recombinase genes. Genes whose inactivation redirect the carbon flux to the pentose phosphate pathway, leading to an increased production of nucleotides also seem to be pertinent. In *E. coli* the genes *endA*, *recA*, *pgi*, *pykA* and *pyk*F, have been related to pDNA production, whose knockout increase DNA production and/or quality.[48,49] It seems reasonable that the deletion of these genes' homologs in *L. lactis* could have the same impact on pDNA production, if orthology is confirmed for this feature.

In fact, the first plasmid construction for CRISPR/Cas9-based knockout in *L. lactis* LMG 19460 here presented is targeted at the endonuclease III gene, *nth*, which may have a similar function to the *E. coli*'s *endA* gene. In *E. coli*, the endonuclease I coding gene, *endA,* knockout did not prove to increase DNA concentration or quality (supercoiling fraction) for high density, batch mode cultivations.[115] However, it has been earlier shown that the quality of plasmid DNA prepared from *endA*$^+$ strains appeared inconstant, and after extraction and long-term storage, after quality assessed by gel electrophoresis and restriction digestion, the DNA appeared degraded.[116] Therefore, it seems plausible to extrapolate the same effect of the *nth* product in *L. lactis*. The *nth* knockout should eliminate the activity of the non-specific endonuclease that affects DNA stability and shelf-life.[117]

Another approach to increase DNA production and quality yields would be the knockout of the *recA* gene. The product of *recA* is responsible for recombination, which may lead to multimerization of the plasmid. Therefore, its knockout should prevent unwanted recombination events and thus, stabilization of the plasmid vectors sizes.[117] Indeed, it has been proved that *ΔrecA* mutants have a 2.5-fold increase of DNA production, compared to the respective wild-type *E. coli* strain, as well as an increase in pDNA supercoiled fractions.[115] As these gene is ubiquitous among bacteria[118], one could expect the same effect in *L. lactis* at least concerning DNA quality and production. However, it is important to stress out that deletion of the *recA* gene in *L. lactis* strain MG1363 has been shown to have a pleiotropic effect. Duwat *et al.* have shown that RecA not only is involved in homologous recombination and DNA repair, but also in responses to oxygen and thermal stresses.[119] *L. lactis ΔrecA* cells cultured in aeration showed 10^2-10^3-fold lower viability compared to mutant non-aerated cells, in the stationary phase. Furthermore, the exponential phase was characterized by a two-fold increase in doubling time (depending on the culture medium used). More surprising, were the results observed when *L. lactis* was submitted to thermal variations. After shifting the cell growth of *ΔrecA* cells from 30ºC to 37ºC, cell growth stopped for 8-10 hours after a three generation-

growth. Immunological data have shown that the *recA* mutant is deficient in the heat-shock response proteins DnaK, GroEL and GrpE, indicating that *recA* plays a role in regulating heat-shock in *L. lactis*.[119] Hereupon, when deleting *L. lactis recA* gene, one must take into account that the growth conditions must be controlled in order to avoid cell growth arrest, namely, controlling the availability of oxygen. The heat-shock response proteins deficiency could posture a problem when using temperature sensitive replicons, which is the case of the editing plasmid nthCRISPRa. Tests would have to be made using the lowest instability temperature possible. RecA plays a fundamental role in homologous recombination and DNA damage repair.[118] Furthermore, the deletion of this gene should be the last to be made when designing a multi-gene-knockout strain using the present CRISPR/Cas9 system approach, which relies in homologous recombination.

The *pgi* gene codes for the enzyme phosphoglucose isomerase, which catalyses the conversion of glucose-6-phosphate into fructose-6-phosphate. The knockout of this gene in *E. coli* leads to the redirection of the carbon flux to the pentose phosphate pathway[120], which, in turn, leads to an increase nucleotide production and, consequentially, pDNA synthesis[121] (Figure S 3 (Supplemental Data)). Gonçalves *et al*. described this phenomenon for the first time, showing that *E. coli* GALG20 (MG1655*ΔendAΔrecAΔpgi*) produced 25-fold more plasmid DNA than its parental strain (MG1655*ΔendAΔrecA*), having glucose as the primary carbon source (20 g/L).[121] It should be expected that the same redirection of carbon flux would happen in *L. lactis ∆pgi*, hopefully resulting in increased DNA production.

The knockout of two pyruvate kinase-coding genes in *E. coli*, *pykA* and *pykF*, were shown to increase DNA production and reduced acetate production, when grown in glucose containing media (20 g/L), compared to the parental strain.[121,122] Thus, these knockout strategy is not only advantageous when creating an added-value plasmid DNA production strain, but also allow for easier cultivation of high cell concentration when scaling up for commercial purposes[122]. Again, the effects of the unique *pyk* gene knockout in *L. lactis* can also be extrapolated from the effects caused in *E. coli*, as it should also be expected to occur a similar redirection of carbon flux to generate key precursors for plasmid DNA synthesis[122], due to similarities in central carbohydrate metabolism[114]. However, one must take into account that this gene is also responsible for dGTP conversion in purine metabolism, which in turn may not be advantageous, if the genes involved in the same reaction are not properly functioning (Figure S 2 (Supplemental Data)).

Given these knockout possibilities in *L. lactis* LMG 19460, the primers for new sgRNAs and homologous arms for recombination, as well as primers for KO confirmation, were designed and are shown in Table S 1 (Supplemental Data). The reasoning explained in the end of section 3.1.1 was followed. As the SpeI and ApaI restriction sites are not found in the fragment constructions of *recA*, *pgi* and *pyk*, these restriction sites could be again used for fragment design, and these enzymes can be used for restriction digestion of the fragments and vector for posterior ligation. One and two nucleotides were added in the 5' end of the ApaI and SpeI restriction sequences, respectively, in order to increase higher efficiency of docking and digestion, as proposed by NEB.[123] Figure 4 also applies as schematic representation of the complete

fragment and primers here described. The sgRNA PCR amplification does not require template, as the primer forward and the primer reverse overlap in the Cas9 handle sequence. The PCR amplification of HA1 and HA2 require *L. lactis* LMG 19460 genomic DNA as template. A SOEing PCR of the HA1 and HA2 fragments is made before ligating the whole fragment together (sgRNA and homology arms). If SpeI and ApaI restriction sites were present in the designed fragments (recognition sequence or homology arms), the fragment could be inserted in the vector by Gibson assembly, after the vector SpeI and ApaI digestion.

4.2 *Lactococcus lactis* LMG 19460 transformation and *nth* deletion assessments

L. lactis LMG 19460 first transformations were made using 100 ng and 500 ng of plasmid nthCRISPRa, and electroporation pulses of 1 × 1,000 V and 3 × 1,000 V. Surprisingly, the transformation protocol yielded more than 1,000 colonies using 500 ng of DNA and a 1 × 1,000 V pulse and using 100 ng of DNA and 3 × 1,000 V pulses, These electroporation procedures yielded highly packed cell growth, compared with transformation of LMG cells with pTRKH3, which demonstrated a transformation efficiency of 20 cfu/μg, using 500 ng DNA and with no overnight recovery step. Although the number of colonies obtained is higher than 400, it must be noted that the packed cell growth may secret other outcomes and interfere with the proper growth of cells. Even so, four colonies from the first mentioned transformation protocol were picked to liquid medium and after the incubation process, the gDNA and the pDNA were purified. To ascertain whether the knockout had occurred and whether the plasmid was present in electrotransformation-submitted cells, a confirmation PCR was performed using the purified DNA. The PCR to genomic DNA using the Nth_conf primers yielded a fragment between 2,500 bp and 3,000 bp, indicating that no knockout had occurred. As the use of primers HA1 in a PCR reaction using purified pDNA yielded the expected product (1,028bp), one assumed that the plasmid was present in transformation-submitted cells, however cells were not edited. The same did happen for another transformation experiment in which 100 ng nthCRISPRa were pulsed inside cells with a three 1,000 V pulse appliance. Transformation generated 200 colonies and four were grown in liquid medium and analysed. Again, no knockout was observed and the HA1 fragment was positively amplified. Additionally, an electrophoresis analysis of the samples did not ensued in plasmid DNA visualization, which disagreed with the HA1 positive amplification. As cells would grow in antibiotic medium and render positive results for the presence of the homologous arm 1 (HA1), one pondered the possibility that the plasmid would not be seen in agarose gels due to its low copy number nature.

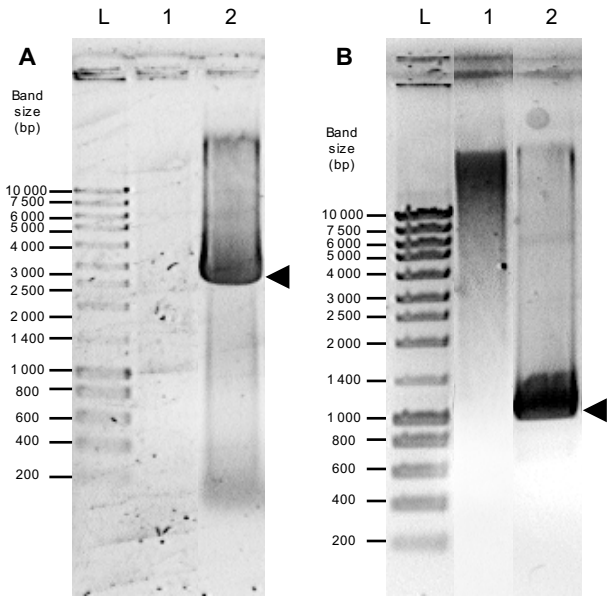

Figure 7 Analysis of agarose gel electrophoresis of (A) gDNA and respective Nth_conf PCR product and (B) nthCRSIPRa and respective HA1 PCR product, purified from a L. lactis LMG 19460 transformation candidate. Lane L- Molecular-weight marker (NYZ DNA Ladder III,NYZTech). (A) Lane 1- 1 μg nthCRISPRa. Lane 2- HA1 amplicon using 100 ng nthCRISPRa template. (B) Lane 1- 1 μg genomic DNA. Lane 2- Nth_conf amplicon using 200 ng gDNA as template.

Figure 7 demonstrates typical results of *L. lactis* LMG 19460 edition and transformation assessement. As it can be observed in Figure 7A, genomic DNA is in good condition, as no intense smear is observed throughout lane 1. In lane 2, the Nth_conf PCR amplicon evidences that no edition event had ocurred, as the molecular wheigh of the product seems to be ~2,800 bp. To assessed whether plasmid DNA was present inside the electroporation candidates, pDNA was extracted and run in an electrophoresis gel. In Figure 7B one can observe that clear pDNA isoforms are not oberved (lane 1). Instead, it is possible to observe a smear, which, at the time, seemed to be indicating that pDNA was not present in the sample, at least in a non-degrated or visible state. Lane 2 observations led to the conclusion that, as the HA1 ~1000 bp amplicon was amplified from purified pDNA sample, nthCRISPRa was present, although not visible, which could be related to its low-copy number nature.

Aiming to increase the plasmid DNA copy number, a protocol of cell growth using spectinomycin 10 mg/mL to induce the increase of pDNA production, by inhibiting protein synthesis, was used.[124,125] pTRKH3-containing cells were used as controls, as the presence of the plasmid had already been confirmed. Either for the control or transformation candidate cells, four growth protocols were performed. Two of each were grown standardly and the growth was arrested after cells reached the mid exponential phase. The pDNA purification was done using the standard method (3.6.2) or using an initial alternative boiling step (3.6.2.1). The other two were grown until mid-exponential phase, after which spectinomycin was added (3.2.2.1). Plasmid samples were purified using the standard method (3.6.2) or using an alternative boiling step (3.6.2.1). Resulting DNA concentrations in purified mixtures are shown above (Table 13). For visualization in agarose gel, 700 ng of each sample was observed after electrophoresis (Figure 8).

Table 13 DNA concentration values (based on Abs260nm quantification) of DNA purified mixtures after different growth protocols and purification coupled methods.

Growth Protocol	**Standard**	**+ Spectinomycin**	**Standard**	**+ Spectinomycin**
pDNA Extraction	**Standard**	**Standard**	**+ Boiling Step**	**+ Boiling step**
pTRKH3	262.5 ng/μL	503 ng/μL	13.6 ng/μL	19 ng/μL
nthCRISPRa	15.9 ng/μL	496.5 ng/μL	15.1 ng/μL	1.9 ng/μL

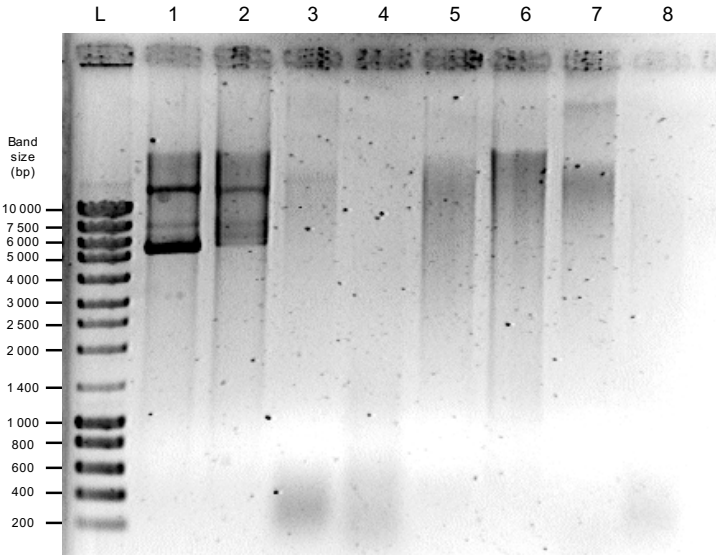

Figure 8 Analysis of agarose gel electrophoresis of samples containg 700 ng of each DNA sample. Lane L- Molecular-weight marker (NYZ DNA Ladder III,NYZTech). pTRKH3 (lanes 1-4) extracted from cells under standard growth (lanes 1, 3) or submitted to spectinomycin exposure (lanes 2, 4) and standard extraction conditions (lanes 1, 2) and using an initial boiling step (lanes 3, 4). nthCRISPRa (lanes 5-8) extracted from cells under standard growth (lanes 5, 7) or submitted to spectinomycin exposure (lanes 6, 8) and standard extraction conditions (lanes 5, 6) and using an initial boiling step (lanes 7, 8).

As one can infer by the observation of the results shown above, the measured DNA concentration does not imply that the extracted DNA is present in the purified mixture. Indeed, nthCRISPRa purified with the standard method using the spectinomycin protocol revealed high concentration values but it did not result in a clear observation in the gel. In fact, as an intense smear is observed, the high DNA concentration value may be due to the presence of high contents of degraded DNA. Furthermore, in this case, DNA concentration measurements do not seem to be good indicators for the presence of DNA. As one can observe, the sample in lane 1 (Figure 8) shows higher contents of DNA when compared to the lane 2 but DNA concentration measurements were higher in the sample loaded in lane 2 (Figure 8 and Table 13). Therefore, it is extremely important to run samples in agarose gels to assess DNA presence and quality. Most importantly, the spectinomycin growth protocol together with standard purification procedures did not improve pTRKH3 copy number (lane 2 in Figure 8) and the standard methods of cell growth and pDNA extraction seem to be more advantageous. nthCRISPRa did not appear in the gel (lanes 5-8 in Figure 8) and although the candidate submitted to the protocol using the spectinomycin seemed to have more (degraded) DNA, this protocol did not serve its purpose. The alternative lysis step did not increase pDNA

concentration nor quality for both samples (lanes 7 and 8 in Figure 8). It is to denote that running these samples in agarose gels is not an accurate test to assess plasmid copy number; it serves only as a fast comparative assay to understand which DNA samples have higher and lower concentrations and to see whether they have the expected characteristics.

Although the protocol did not improve the apparent plasmid copy number in the control *L. lactis* LMG 19460 harbouring the pTRKH3 plasmid, the absence of plasmid DNA isoforms in samples extracted from a nthCRISPRa *L. lactis* LMG 19460 transformation candidate may indicate that, in fact, the plasmid DNA was not present inside the cells, and that this candidate was a false positive.

The use of two independently purified nthCRISPRa samples has led the conclusion that pDNA samples were contaminated with genomic DNA: a fragment of approximately 2,800 bp was amplified when using primers Nth_conf and purified pDNA as template (data not shown). The Nth_conf primers prime to regions upstream and downstream of the homologous arms in the *L. lactis* genome, which means that would not be possible to amplify a product of 2,790 bp using pure pDNA as template. Therefore, the conclusions made until now would not be applicable. One cannot confirm whether the plasmid has entered cells using these set of primers. Therefore, a new pair of primers that amplify a small region (241 bp) of the internal sequence of Scocas9, only present in the plasmid, was designed. An optimization protocol allowed for the determination of a PCR programme suitable for the amplification of both Cas9_conf and Nth_conf, using a suspension containing a known number of cells as template. This PCR protocol would allow a less time-consuming protocol to test a larger set of transformation candidates.

Several transformation tests using 10 ng, 100 ng, 250 ng and 500 ng of nthCRISPRa followed and rendered the same results: approximately 30-500 colonies grew after respective apramycin selection and tested candidates did not underwent genome edition. Information about the transformation tests, KO and pDNA presence assessment are shown in Table S 2 (Supplemental Data).

4.2.1 Optimizing the cell number for *Nth_conf* and *Cas9_conf* PCR amplification

Aiming the development of a fast approach to test *L. lactis* LMG 19460 candidates for CRISPR modification and plasmid presence, a PCR protocol using cells as templates was developed. The elaboration of such protocol would enable to test more candidates at a time, without the need to resort time and resources at purifying plasmid and genomic DNA. In this assay, the volume of cell pellets was calculated and the number of cells and needed dilutions were made as explained in 3.5.

Initially, the amplification of the fragment amplified using the Nth_conf primers was performed using 50,000 and 10,000 cells as PCR templates and the conditions described in Table 7, with KOD Hot Start Polymerase. Results rendered tenuous amplification of the 2,790 bp fragment with higher intense profile in PCR reactions using 10,000 cells (data not shown), which may indicate that, although containing more genetic material, higher number of cells may cause inhibition or efficiency loss of PCR components,

probably due to the presence of high cellular debris content, at least when using KOD Hot Start DNA Polymerase.

A PCR amplification, encompassing 40 cycles, was then tested using KOD Hot Start Polymerase and Nova Taq Hot Start Master Mix, for comparison purposes. PCR reactions, using Nova Taq Hot Start Master Mix not only improved the 2,790 bp fragment amplification intensity in agarose gel, but also decreased the smear that was characteristic of the amplifications using KOD Hot Start Polymerase.

As NovaTaq DNA polymerase exhibited higher amplification efficacy and less smear (data not shown), a PCR using 250,000 cells and a 40-cycle programme was tested using three different annealing temperatures: 50ºC, 53ºC and 56ºC. The number of cells was increased using the new DNA polymerase, as to increase the acuteness of the genomic fragment amplification and to increase the chance of amplification of the fragment of this low copy number plasmid, if present. The procedure provided good amplifications, specially using 56ºC of annealing temperature. The use of 250,000 cells did not have a negative impact in the amplification of the fragment using NovaTaq polymerase and agreed with a more intense amplification of the genomic fragment. Therefore, we may conclude that cell debris did not interfere with the present PCR reaction. Although functioning with less cells, 250,000 cells were chosen as being ideal for, at least the Nth_conf amplified fragment.

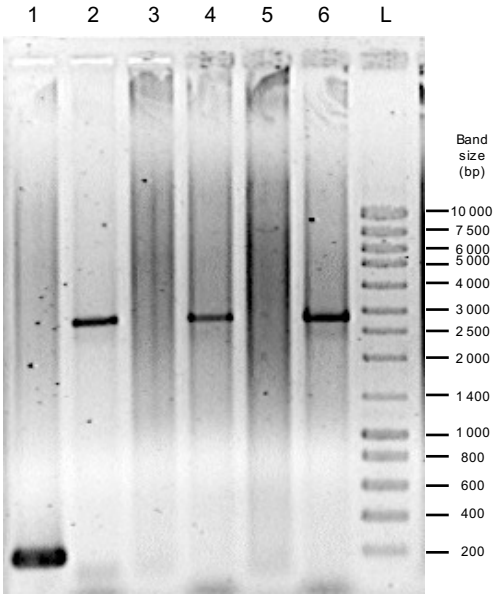

Figure 9 Agarose gel analysis of the electrophoresis of the PCR using L. lactis LMG 19460 transformation candidates. Lane 1- PCR reaction product using Cas9_conf primers and 1 ng of E. coli-purified nthCRISPRa template; lane 2- PCR reaction product using Nth_conf primers and 200 ng L. lactis LMG 19460 genomic DNA template; lanes 3 and 5- PCR reaction productusing Cas9_conf primers and 250,000 L. lactis LMG 19460 electroporated cells; lanes 4 and 6- PCR reaction product using Nth_conf primers and 250,000 L. lactis LMG 19460 electroprated cells; lane L- Molecular-weight marker (NYZ DNA Ladder III,NYZTech). Lanes 3 and 4 correspond to one candidate test, as well as lanes 5 and 6.

The chosen PCR conditions are represented in Table 8 and an example of such PCR result is shown in Figure 9. The present figure shows two of the 21 PCR tests in transformation candidates, which rendered similar results. As one can observe in lane 1, the internal fragment of the Scocas9 fragment of the purified

plasmid is highly and clearly amplified. The PCR products of the Nth_conf fragment using cell as templates (lanes 4 and 6) are as good amplified as the control sample (lane 2), using purified DNA from a previously tested candidate. As the difference between the fragments of Nth_conf amplifications from cells without and with the expected edited genome is as small as 657 bp, one can expect that edited cells have similar amplification efficiencies using the programme mentioned above. However, it is only possible to conclude such supposition when an edited sample is selected.

Simultaneously, different pDNA masses regarding the control sample were tested. After the shift to NovaTaq DNA master mix, 10 ng and 1 ng were used for the PCR reaction. The use of 1 ng proved to be enough to have a good amplification (Figure 9). Regarding the genomic DNA extracted from a *L. lactis* 19460 transformation candidate, 200 ng were used in every test.

Conclusion remarks about the amplification of the internal fragment of Scocas9 were not possible, as no tested candidates encompassed the nthCRISPRa plasmid. As soon as the internal fragment of the nthCRISPR is detected, the number of cells can be restudied, if necessary, to assure a clear amplification of both Nth_conf and Cas9_conf resulting fragments.

So far, the only amplification accomplished using cells as templates was the amplification of the 2,790 bp Nth_conf resulting fragment, amplified from genomic DNA. To assure that plasmid sequences would be amplified using cells as templates, a PCR test using *L. lactis* LMG 19460 containing the high-copy number pTRKH3 plasmid was performed using 10,000 and 1,000 cells as templates. For the present PCR reaction, the polymerase NovaTaq Master Mix and Ery_conf primers were used and the PCR programme is mentioned in 3.7.3. All the three samples tested rendered positive for the 190 bp fragment amplification and 10,000 cells showed to be better for confirmation purposes, as a clearer amplification was obtained. Although PCR amplifications may depend on DNA template sequence, size, structure and quality, it is possible to conclude that PCR protocols to amplify pDNA sequences using cells are possible and that, when present in *L. lactis* cells, the nthCRISPRa should be able to be amplified and detected.

Electroporation of this lactic acid bacterium with nthCRISPRa proved to be a difficult task and the surviving colonies did not prove to be true transformants. Certain conclusions cannot be made, as one is not positive about apramycin sensitivity in *L. lactis* LMG 19460. This uncertainty relies in the observation of bacterial growth using supposedly high concentrations of antibiotic, which did not bear the nthCRISPRa plasmid. However, transformation of lactic acid bacteria can be a difficult task, as their thick cell walls are highly resistant to mechanical disruptions.[126] Furthermore, the existence of unspecific endonucleases may not ease the abiding of the foreign pDNA inside the cells. Future work should include a larger set of electroporation conditions, for example higher range voltage values, number of pulses and DNA concentrations. Electroporation of lactic acid bacteria is not a straightforward method and a single protocol for a single species is yet to be achieved. Indeed, this might be a difficult task as strains within the same species show huge variations in transformability and growth conditions. Furthermore, transformability is

influenced by a variety of different parameters, such as growth conditions and growth arrest phase, final cell concentration, plasmids to be transformed, composition of media and solutions used throughout the electrocompetence and transformation protocols, field strength and duration and number of pulses.[127] Despite these drawbacks, the transformation of *L. lactis* LMG 19460 with the pTRKH3 plasmid was successful. Although the electroporation of LMG cells with nthCRISPRa may need some optimization, given the existence of pTRKH3 true transformants, most probably, the surviving colonies in apramycin after nthCRISPRa electroporation are not true transformants, and it was hypothesized that apramycin was not a suitable selection marker for *L. lactis*.

4.3 PCR optimization for the amplification of the nthCRISPRe constituent fragments

The transformation of *L. lactis* LMG 19460 cells with the plasmid nthCRISPRa was not achieved, even using higher apramycin concentrations and different electrotransformation conditions. The lack of detection of nthCRISPRa plasmid in *L. lactis* cells led to the debate whether apramycin was a good selection marker for this system, although previous minimum inhibitory concentration studies had been made for solid and liquid medium growth (unpublished). It was decided to change the selection marker apramycin to a selection marker already established for lactic acid bacteria, namely erythromycin, which is the pTRKH3 selective marker. Erythromycin has been used in the laboratory for a long time and it embodies in the list of ten antimicrobials that the EFSA has set as a basic requirement for the LAB group.[128] However, the plasmid sequence did not contain restriction sites that would allow the removal of the *apmR* cassette and the insertion of erythromycin gene and its promoter (erm) using restriction enzymes and ligation protocols, respectively. Therefore, the assembly of desired fragments would have to be obtained using a Gibson assembly protocol. Due to the large size of nthCRISPRa, the amplification of the plasmid was decided to be made by the amplification of two different fragments. One fragment narrowed from the only restriction site of the ApaI restriction enzyme until the beginning of the *apmR* cassette, where the sequence of the erm cassette would start (Apa_Ery with 7,666 bp); and the other from the end of the *apmR* cassette, where the sequence of the *eryR* would finish until the ApaI restriction site (Ery_Apa with 4,888 bp) (Figure 5). The 1,279 bp erm and *eryR* sequence would be placed in the same location as the previous resistance marker. To shorten the resulting plasmid sequence, the deletion of the region between the ApaI and the HindIII restriction sites (Figure 5) in the Huang and colleagues[97] pKCcas9dO vector, was removed by designing primers to do as such: the Apa_Ery primers have an addition of 10 nucleotides that overlap the HindIII restriction region and the Ery_Apa primers have an addition of nucleotides that overlap the ApaI restriction region (Figure 5). The primer sequences were designed according to the protocol and tools provided by the technology holder (3.8.2) and are seen in Table 10.

4.3.1 Ery Amplification

The Ery (erm and *eryR*) amplification from pTRKH3 was promptly amplified using a single gradient test for three independent annealing temperatures, using the Ery primers shown in Table 9. The high fidelity KOD Hot Start DNA Polymerase was used to amplify the Ery fragment from 0.5 ng of template nthCRISPRa. A first attempt using 0.05 ng template was also successful (data not shown), but using 5 ng instead has permitted clearer visualization of bands in the agarose gel. All the annealing temperatures used, 59ºC, 60ºC, and 61ºC, allowed the desired amplification of the Ery fragment. However, the annealing temperature of 59ºC managed a clearer and more intense amplification of the 1,299 bp fragment. A brief analysis is shown in Figure 10.

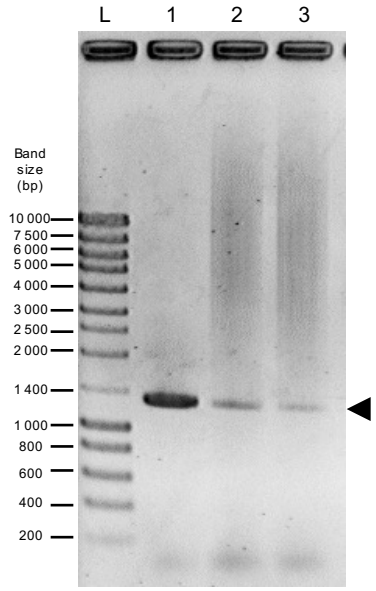

Figure 10 Agarose gel analysis of the electrophoresis of PCR products of pTRKH3 amplification using the Ery primers. Lane L- Molecular-weight marker (NYZ DNA Ladder III,NYZTech); lanes 1, 2 and 3- PCR products of the amplification using an annealing temperature of 59ºC, 60ºC and 61ºC, respectively.

The chosen PCR conditions are described in Table 10.

4.3.2 Apa_Ery Amplification

The amplification of the Apa_Ery fragment was firstly attempted using the PCR programme shown in Table 10 and three independent annealing temperatures of 58ºC, 60.5ºC and 63ºC. The high fidelity KOD Hot Start DNA Polymerase was used to amplify the fragment from 0.05 ng of template nthCRISPRa, using the primers shown in Table 9. The amplification of the desired 7,686 bp product was achieved using 60.5ºC as the annealing temperature. However, the PCR reaction also yielded other unspecific amplifications, some of which having higher intensity in the agarose gel comparatively to the desired product. More problematic was the amplification of fragments a few base pairs larger and a few base pairs smaller, rendering it difficult to posteriorly purify the DNA band from the TAE agarose gel (Figure 11A).

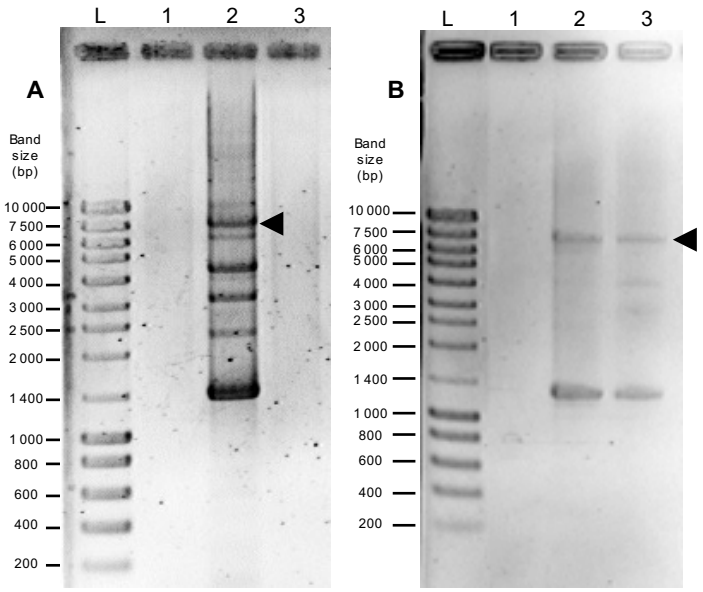

Figure 11 (A) Agarose gel analysis of the electrophoresis of PCR products of nthCRISPRa amplification using the Apa_Ery primers. Lane L- Molecular-weight marker (NYZ DNA Ladder III,NYZTech); lanes 1, 2 and 3- PCR products of the amplification using 0.05 ng DNA template and annealing temperatures of 58ºC, 60.5ºC and 63ºC, respectively. (B) Agarose gel analysis of the electrophoresis of PCR products of nthCRISPRa amplification using the Apa_Ery primers. Lane L- Molecular-weight marker (NYZ DNA Ladder III,NYZTech); lanes 1, 2 and 3- PCR products of the amplification using 5 ng DNA template and annealing temperatures of 61ºC, 62ºC and 63ºC, respectively.

It is known that magnesium salt Mg^{2+} acts as a cofactor during the PCR reaction process and it is required for the proper function of thermostable DNA polymerases. Therefore, one can easily manipulate the magnesium concentration to control the stringency of the PCR reaction. In general, increasing the Mg^{2+} concentrations lead to increased PCR product yields, as it increases the stability of primer-template duplexes. However, the presence of surplus magnesium may lead to unspecific amplifications or may prevent the complete denaturation of the DNA template, resulting in no PCR products.[129–131] Taking this into account, the magnesium ($MgSO_4$) concentration was increased and decreased 0.5 mM, comparatively to the final concentration of 2.25 mM, usually used. The annealing temperature of 60.5ºC was used and the extension time was prolonged to 10 minutes. The remaining amplification parameters were maintained. The PCR procedure yielded no products, either specific or unspecific (data not shown). Thus, changing the magnesium concentration was not a good strategy to specifically improve this amplification.

Several PCR reaction tests followed. The previously successfully amplified sample was used as template for a new PCR reaction using the same conditions, aiming to decrease unspecific amplifications. It was only possible to observe a generalized smear with no amplification products (data not shown). One can presume that residual PCR components from the previous amplification may have interfered with the proper functioning of the consecutive PCR reaction. A 10-fold increase in the nthCRISPRa template mass was made and 0.05 ng of template was also used. The annealing temperatures of 59.5 ºC, 60.5ºC and 61.5ºC were tested. Increasing the template mass proved to be advantageous for the amplification, regardless the annealing temperature used (data not shown). This PCR reaction revealed an apparent

decrease of the amplification of the products upstream and downstream of the desired 7,686 bp product, which would be advantageous for future purification of the DNA band from the TAE agarose gel, if a better optimization could not be achieved. The product with the most intense amplification was the one amplified with the highest annealing temperature. It is difficult to predict why the PCR reaction did not amplify samples from 0.05 ng of template DNA, especially, using 60.5ºC annealing temperature, as it corresponded to the conditions that previously amplified the desired fragment (Figure 11A). The absence of results could be due to low DNA concentration and, consequentially, low DNA uptake. Trying to subvert the unspecific amplifications, the retrieval of a small cylinder of the desired amplified product from the TAE agarose gel was made using a 100 μL pipette tip, to be used as template for a PCR reaction. The annealing temperature of 61.5 ºC was used. This PCR procedure poorly and only amplified the unspecific fragment of approximately 1,4 00 bp, and was characterized by a generalized smear (data not shown). It can be concluded that this PCR protocol is not adequate for increasing the amplification specificity and it can be postulated that the presence of agarose may negatively influence the PCR reaction.

After several amplification tentative optimizations, some of which described above, a new gradient PCR was performed using 61ºC, 62ºC and 63ºC as annealing temperatures, 5 ng of DNA template and the remaining parameters as show in Table 10. The results are shown in Figure 11B. PCR amplifications using 62ºC and 63ºC positively amplified the desired fragments as well as unspecific products. Nonetheless, it was decided that no further optimizations would be attempted and that the desired Apa_Ery product would be amplified using an annealing temperature of 62ºC and purified from the TAE agarose gel. Final amplification conditions are shown in Table 10. Later in the work, this fragment was purified using a PCR clean up protocol, although having one unspecific product.

4.3.3 Ery_Apa Amplification

The amplification of the Ery_Apa fragment was very laborious, as several tests using different annealing temperatures (from 59.5ºC to 70.5ºC), different DNA template masses (0.05, 0.5 and 5 ng) and MgSO4 concentrations (0.75, 1, 1.25, 1.75, 2.25, 2.75 and 3.25 mM) were tested. The magnesium concentrations were varied due to same reasons explained above. The magnesium concentration variation was made to test whether it would have an effect in the polymerase specificity and fidelity. The increment in DNA template mass was used to increase the DNA template molecules available to template the reaction. Independently purified templates were used as to test if the problem would have been related to the DNA integrity. The reduction of the annealing temperature was made to try to reduce the stringency of the reaction. The increment of the annealing temperature was made in order to test whether an increase in the stringency of the primer:DNA complex would improve the amplification. None of the tests described above and performed using the KOD Hot Start Polymerase followed amplifications.

Due to the lack of results using the KOD Hot Start Polymerase, although not ideal due to its non-proofreading and lower fidelity nature, a PCR protocol using NovaTaq Hot Start Master Mix was tested, using 0.5 ng of nthCRISPRa template. PCR steps were performed as in Table 10, using a gradient test with

independent annealing temperatures 50ºC, 54ªC, 58ºC, 62ºC, 66ºC and 70ºC. The PCR reaction amplified the desired 4,908 bp product with 58ºC and 62ºC annealing temperatures, although characterized by an intense generalized smear (data not shown).

Aiming to increase the binding specificity of the primers to the DNA template, the Ery_Apa2 pair of primers was designed as to have 10 additional seeding nucleotides in its sequence (Table 9). The KOD Hot Start DNA Polymerase was used to attempt the amplification from 5 ng of template using the primers Ery_Apa2, and annealing temperatures of 56ºC, 58ºC, 60.5ºC, 62ºC and 64ºC. This PCR reaction did not amplify the desired fragment (data not shown). As PCR reactions using KOD Hot Start polymerase were not handing results, it was decided that the amplification of the Ery_Apa fragment would be attempted using the NovaTaq DNA Polymerase, carrying in mind the disadvantages of doing so, which include the possibility of undesirable mutants after Gibson assembly and transformation, given the higher error-rate of this polymerase. Nova Taq Master Mix DNA polymerase was used to amplify the desired fragment from 20 ng DNA template using annealing temperatures of 54ºC, 58ºC, 62ºC and 66ºC and primers Ery_Apa. In addition to nthCRISPRa, the pKCcas9dO plasmid was also used as template, as the desired fragment is the same in both plasmids. The PCR reaction yielded specific amplifications using 58ºC and 62ºC annealing temperatures, using both plasmids as templates, as well as one unspecific product with size ranging between 800 bp and 1,000 bp, more pronounced in samples amplified from pKCcas9dO (Figure 12A).

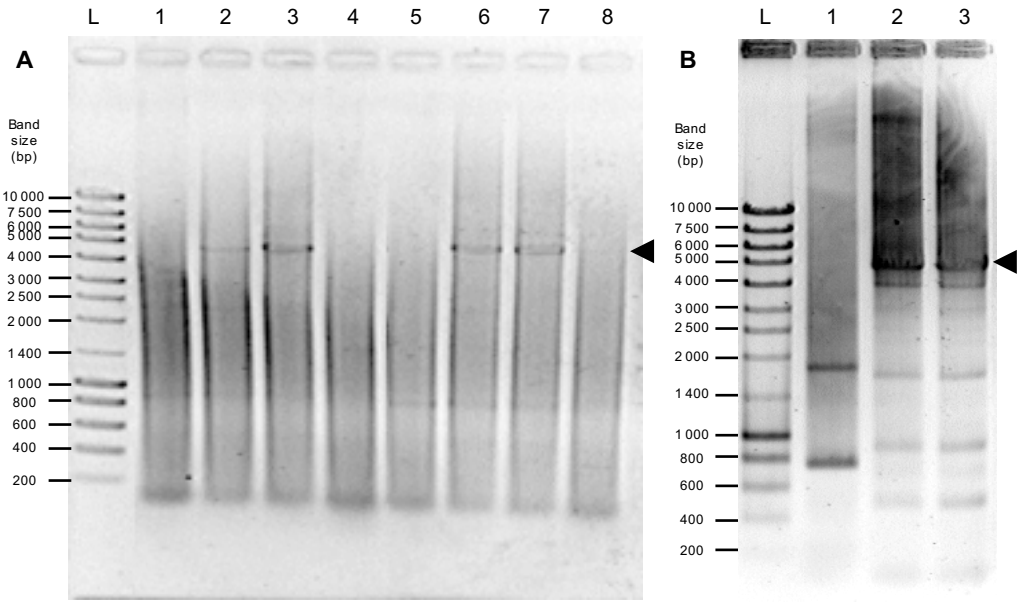

Figure 12 (A) Agarose gel analysis of the electrophoresis of PCR products of nthCRISPRa or pKCcas9dO amplifications using the Ery_Apa primers and NovaTaq Master Mix. Lane L- Molecular-weight marker (NYZ DNA Ladder III,NYZTech); lanes 1, 2, 3 and 4- PCR products of the amplifications using annealing temperature of 54ºC, 58ºC, 62ºC and 66ºC, respectively, and nthCRISPRa as template; lanes 5, 6, 7 and 8- PCR products of the amplifications using annealing temperature of 54ºC, 58ºC, 62ºC and 66ºC, respectively, and pKCcas9dO as template. (B) Agarose gel analysis of the electrophoresis of PCR products of nthCRISPRa amplification using the Ery_Apa2 primers and KOD Hot Start Master Mix. Lane L- Molecular-weight marker (NYZ DNA Ladder III,NYZTech); lanes 1, 2 and 3- PCR products of the amplifications using annealing temperature of 50ºC, 54ºC and 58ºC, respectively.

The sample with the clearest amplification was the one amplified from nthCRISPRa, using an annealing temperature of 62ºC. Although yielding other unspecific products, as shown in Figure 12A, the annealing temperature of 62ºC was chosen for further amplifications which would be followed by the Ery_Apa fragment TAE agarose gel purification. Final PCR conditions are shown in Table 10.

Later, another set of tests was made using different DNA polymerases, employing the manufacturer's recommended conditions for each enzyme, using 5 ng of template and a gradient programme with independent annealing temperatures of 58ºC, 60ºC and 62ºC. Platinum Supermix HiFi DNA Polymerase had an additional PCR reaction with an annealing temperature of 55ºC, the recommended temperature for annealing step. The DNA polymerases used in this study were: KOD Hot Start DNA Polymerase (Merck Millipore), KOD Hot Start Master Mix (Merck Millipore), BIOTAQ™ (Bioline), Platinum Supermix HiFi DNA Polymerase (Thermo Fisher Scientific), Pfu Turbo (Agilent), Nova Taq Master Mix (Merck Millipore), Nzylong (NZYTech) and Supreme NzyLong (NZYTech). Only Nova Taq Master Mix and KOD Hot Start Master Mix yielded an amplification of the desired product. The first enzyme amplified the ~4,900 bp product with an annealing temperature of 62ºC, together with three unspecific products of approximately 2,500 bp, 900 bp and 800 bp (data not shown). KOD Hot Start Master Mix yielded the desired product using annealing temperatures of 58ºC and 60ºC, together with seven unspecific products varying from approximately 2,500 bp to 400 bp (data not shown). As a proofreading high fidelity enzyme, a KOD Hot Star Master Mix amplification would be preferable over an amplification using the error prone Nova Taq Polymerase, when projecting a cloning protocol.

Using both pair of primers available for the amplification of the Ery_Apa fragment (Ery_Apa and Ery_Apa2), KOD Hot Start Master Mix and 50ºC and 60ºC annealing temperatures, it was possible to assure that, the use of the pair of primers Ery_Apa2 was preferable, as a more defined product was visualized in a 1% TAE agarose gel (data not shown). Finally, an annealing temperature gradient test was performed using 50ºC, 54ºC and 58ºC and the results are shown in Figure 12. Unexpectedly, the amplification using 50ºC yielded no specific product, which disagrees with the previous result. However, the programme using the other two annealing temperatures amplified the desired product together with one major unspecific product of ~4,000 bp. It was decided that further amplifications for purification would be done using 54ºC annealing temperature. The PCR conditions used for the amplification of Ery_Apa using the longest set of primers (Ery_Apa2) are shown in Table 10. The 4,908 bp product would be purified from the TAE agarose gel.

4.4 Gibson Assembly of the nthCRISPRe constituent fragments and Transformation of *Escherichia coli* DH5α

Six different Gibson assembly reactions were independently performed and transformed in *E. coli* DH5α cells. The transformation events are described above and summarized in Table S 3 (Supplemental Data).

Gibson Assembly #1

The reaction mixtures of Gibson Assembly #1 were used in three different transformation events, two heat shock chemical transformations and an electroporation protocol. After selection in erythromycin 500 μg/mL, transformation candidates accounted a total of 4, 18 and 3 cfu, respectively. The first 4 did not exhibited growth in liquid medium (DO600nm<0.04). Of the 18 cfu, 4 were picked to liquid medium and 3 exhibited very little growth. Nevertheless, the plasmid DNA of these three candidates was extracted and its concentration was measured, varying from 12.4 to 16.6 ng/L. To test whether nthCRISPRe pDNA was present in the purified mixture, 500 ng were run in 1% TAE agarose gel. It was not possible to observe any pDNA products in the electrophoresis analysis (data not shown). The 3 electroporation candidates were submitted to colony PCR using the Cas9_conf primer pair. The colony PCR procedure did not amplify the internal cas9 fragment present in the plasmid DNA to be constructed (data not shown), which suggested that the plasmid DNA was not present inside the cells.

Gibson Assembly #2

Gibson #2 products were used to perform three different transformation events: one electroporation and two heat shock protocols. After erythromycin 500 ug/mL selection, the electroporation protocol yielded 3 colonies, 2 of which were analysed for the Scocas9 fragment presence by colony PCR. Again, the 241 bp fragment was not detected (data not shown), suggesting that the nthCRISPRe plasmid was not present inside cells. One chemical transformation procedure was made using 250 g/mL erythromycin to select transformants. The reasoning behind the antibiotic concentration decrease was the possibility of the 13,833 bp resulting nthCRISPRe plasmid to be too large to produce sufficient number of copies to produce enough erythromycin resistance protein to cope with the high concentration of antibiotic in the medium. This transformation protocol yielded a total 742 candidates (from three different transformations: 142, 313 and 287 cfu), 12 of which were submitted to colony PCR to assess the presence of the cas9 containing fragment. After electrophoresis of PCR products, the presence of such fragment was confirmed in 11 of the total 12 transformation candidates (Figure 13).

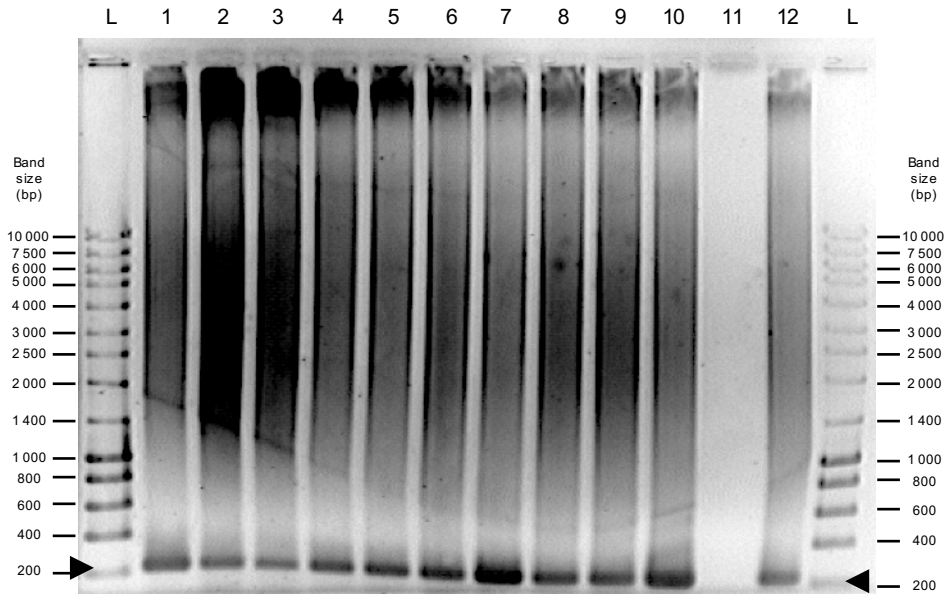

Figure 13 Agarose gel analysis of the electrophoresis of E. coli DH5α colony PCR using Cas9_conf primers for the assessment of pDNA presence. Lanes L - Molecular-weight marker (NYZ DNA Ladder III,NYZTech); lanes 1 to 4 –PCR products of transformation candidates with the 15 minute Gibson #2 reaction mixture; lanes 5 to 8 – PCR products of transformation candidates with the 1 hour Gibson #2 reaction mixture; lanes 9 to 12 – PCR products of transformation candidates with the 4 hour Gibson #2 reaction mixture.

The 11 candidates that were positive for the presence of the Scocas9 fragment were picked to liquid medium supplemented with Ery 250 g/mL. Six candidates exhibited growth ($0.57 \leq OD_{600nm} \leq 1.57$) and plasmid DNA was purified from four of these candidates. Its concentration in purified mixture was evaluated, varying from 0.1 to 13.4 ng/μL, with the higher concentration belonging to the candidate which exhibited higher growth. The concentration of DNA demonstrated to be too low, which was not expected as some candidates had higher OD_{600nm} measurements, for which pDNA production is characteristic, if present. The total DNA mixture was loaded into the agarose gel for visualization. However, plasmid was not observed and the samples appeared degraded, with a generalized smear starting from a tenuous band above 10,000 bp (data not shown). Additionally, 21 cfu were directly picked to liquid medium containing Ery 250 μg/mL. After incubation, none of the 21 candidates unveiled growth in liquid medium ($OD_{600nm} < 0.01$). Furthermore, 30 cfu were picked to new solid medium with the same antibiotic concentration as to form close lines to gain critical mass. 8 candidates exhibited growth and were transferred to liquid medium. 3 of the 8 candidates ($0.3 \leq OD_{600nm} \leq 1.66$) with higher apparent growth ($1.27 \leq OD_{600nm} \leq 1.66$) were used to extract pDNA. After its concentration measurement, and because DNA concentration was low, all the mixture was loaded into agarose gel for visualization after electrophoresis. None of the candidates exhibited a clear pDNA structure, one of the candidates exhibited high DNA content above 10,000 bp, although with degraded appearance (data not shown). The pDNA of this candidate was extracted again and a PCR using the Cas9_conf and Ery_conf primers was performed to assess, not only the presence of the Scocas9, but also the presence of the selection marker. As the results rendered positive for both fragments (data not shown), pDNA was sent for sequencing in STAB VIDA. However, results came out inconclusive and no sequencing was achieved.

These results should be more straightforward, with a clearer appearance of DNA in the agarose gel. Although the electrophoresis analysis did not show a quality plasmid DNA, the PCR confirmation using Cas9_conf and Ery_conf did give positive results. This may happen as cells could have taken up the linear DNA fragments, instead of the pDNA to be constructed. This would explain the growth of cells in erythromycin containing media.

Gibson #2 reaction products were, once more, used to transform *E. coli* DH5α cells by heat shock. Solid medium growth in Ery 350 μg/mL yielded 42 cfu, of which 8 were picked to liquid medium. Only one candidate exhibited growth; however, after pDNA extraction and electrophoresis analysis, no product was found.

Gibson Assembly #3

Gibson #3 products were used to transform *E. coli* either by chemical means (heat shock) or by electroporation. The heat shock protocol was slightly altered: it lasted 90 seconds, instead of one minute. After transformation and recovery, cells were divided and plated in solid medium either containing Ery 250 μg/mL or Ery 500 μg/mL and incubated accordingly. Chemical transformation yielded 1 cfu after selection in Ery 500 μg/mL, and electroporation yield 60 cfu after selection in 250 μg/mL and 1 cfu after selection in 500 μg/mL. Although the 6 analysed cfu rendered positive results for the presence of the Scocas9 241 bp fragment (data not shown), none exhibited growth in liquid medium.

Gibson Assembly #4

After electroporation of *E. coli* with Gibson #4 reaction products, and selection in Ery 250 μg/mL containing medium, 253 cfu were perceived. Nine candidates were analysed for the presence of the erythromycin selection mark, operating with Ery_conf primers. All nine candidates gave positive results concerning the presence of the internal region of the Ery resistant gene (Figure 14).

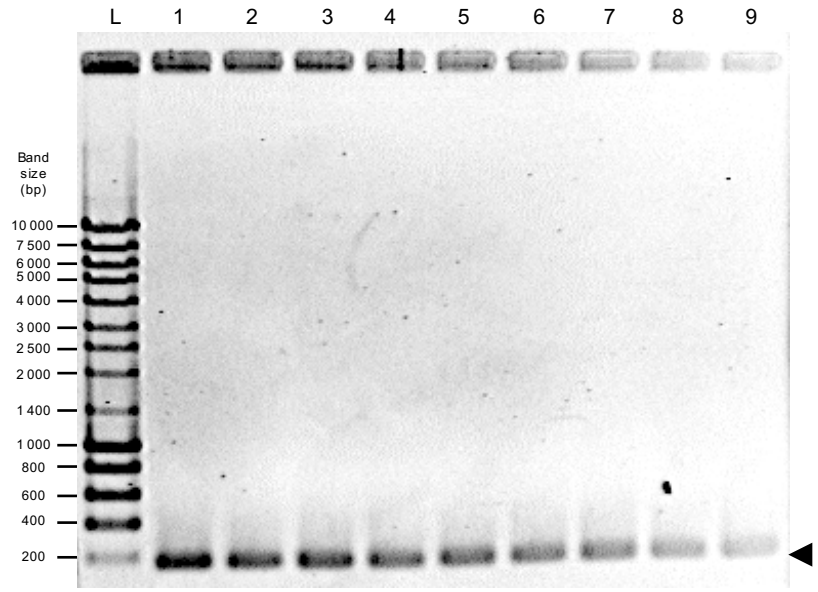

Figure 14 Agarose gel analysis of the electrophoresis of E. coli DH5α colony PCR using ery_conf primers for the assessment of pDNA presence. Lane L - Molecular-weight marker (NYZ DNA Ladder III,NYZTech); lanes 1 to 3- PCR products of Gibson #4 with 15 minutes reaction time; lanes 4 to 6- PCR products of Gibson #4 with 1 hour reaction time; lanes 7 to 9- PCR products of Gibson #4 with 2 hour reaction time.

This time, instead of using 20 μL for liquid medium growth, the suspension of cells was plated onto new solid medium containing Ery 250 μg/mL, to gain critical mass for potential growth in liquid medium. All candidates exhibited growth and were picked into liquid medium. Only one candidate, transformed with reaction mixture of one hour, showed growth (OD_{600nm}=1.5). The respective purified DNA concentration measurements were 9.5 ng/μL. However, no DNA was observed after gel electrophoresis and observation. Once again, it seems that the fragments are present in cells that undergo transformation, but the plasmid is not forming itself during the Gibson Assembly Protocol.

Furthermore, the Gibson #4 of 4 hours-reaction was used to electropore cells with three different reaction volumes: 2, 3 and 6 μL. Cells were plated onto solid medium containing Ery 500 μg/mL. No colonies unveiled growth in neither of the three transformations.

Gibson Assembly #5 and #6

Gibson #5 and #6 assembly reactions were made using all three DNA fragments amplified by high fidelity, proofreading DNA polymerases. Also, after purification, DNA concentrations rendered higher yields, giving one a more confident accreditation in future results. Also, the cooling of each fragment mixture after DpnI inactivation was made in a controlled manner, with a decrease of 2ºC per minute. Theoretically, this should allow a more controlled and organized annealing of the constituting strands of the fragments. For each Gibson reaction, a chemically transformed cell bank was plated in LB plates containing Ery 250 μg/mL and another in Ery 500 μg/mL. Furthermore, a control test using the amount of pTRKH3 present in each

transformation aliquot was used to transform the same cells and selected them in both Ery concentrations tested. The results are shown in Table 14.

Table 14 Number of colonies and analysis obtained in Gibson #5 and #6 of transformant candidates in selection medium containing 500 or 250 μg/mL erythromycine (Ery).

Gibson #	Ery 500 μg/mL	Ery 250 μg/mL
5	7	18
6	16	8
pTRKH3	0	3

The results regarding the transformation of cells with the DpnI digested pTRKH3 grown in Ery 500 μg/mL, suggested that the presence of colonies without the plasmid DNA is not due to the presence of undigested pTRKH3. Although the plating of pTRKH3 transformed cells in Ery 250 μg/mL containing plates resulted in the formation of three colonies, one does not know whether that is due to the presence of undigested pTRKH3 or due to the presence of false positives, as the antibiotic concentration is half the concentration determined to be adequate for this organism. Therefore, candidates grown in Ery 500 μg/mL supplemented medium should be more reliable than the others. All 23 candidates grown in Ery 500 were picked to liquid medium. One candidate exhibited growth (OD_{600nm}=0.69); however, the 10.2 ng/μL measured purified pDNA was not observed in the agarose gel after electrophoresis (data not shown).

With all the above-mentioned results one can conclude that something appears to allow cell growth without the presence of the plasmid, and that could be the presence of non-ligated fragments. It is difficult to counteract this issue using the present system, as one cannot make sure that the ligation is happening before the transformation process. Furthermore, the concentration of DNA used in the present protocol of Gibson Assembly is so low that the ligation product cannot be visualized in an agarose gel electrophoresis. One must trust in ligation probability events until the cloning of the desired product is obtained. Moreover, the ligation itself might be working, but transformation efficiency may not be high enough to see it. To increase Gibson Assembly efficiencies one can increase the number of overlapping nucleotides in each fragment. Although not making use of one of the major advantages of Gibson Assembly (ligation of multiple fragments in a single ligation reaction), the assembly protocol may be done using two fragments at a time, instead of the ligation of three fragments at once.

4.5 nthCRISPRa and nthCRISPRe considerations

The nthCRISPRa plasmid was constructed prior to the present work using pKCcas9dO as backbone. A modified target-specific guide RNA and two homology-directed repair templates (HA1 and HA2), to be used for the *nth* knockout in *L. lactis* LMG 19460, were incorporated in the pKCcas9dO vector after digestion with SpeI and ApaI restriction enzymes of both the vector and the fragment to be inserted. However, it is

important to highlight that the pKCcas9dO plasmid was used to induce genome editing in *Streptomyces coelicolor* M145, and although being a Gram-positive bacterium as well, the system is yet to be proved to work in *L. lactis*.

The naturally temperature sensitive replicon pSG5, from *Streptomyces ghanaensis* DSM2932, is stably inherited under growth conditions below 34ºC. Above this temperature, the plasmid does not replicate and is lost after some replication cycles. Therefore, the use of this plasmid as a mutational platform seems extremely useful, as these vectors can be eliminated from the host cells very efficiently by increasing the incubation temperature, after edition.[132] In the present work, *L. lactis* cells that would render positive results for the *nth* deletion would be then cultivated in 37ºC to lost the nthCRISPRa/nthCRISPRe plasmid, to proceed with the other transformations for the study of plasmid copy number or to proceed with other knockout experiments. The pSG5 replicon is of wide host range within *Sptreptomyces* species; however, it has not been shown to work in lactic acid bacteria. If such origin is not applicable to LAB, the present system would resemble a non-replicative editing system, which is much less effective. The increased efficiency of the CRISPR/Cas editing system relies in proper induction of DSB; therefore, using a non-replicative plasmid, the probability of enough expression of Scocas9 and sgRNA for gene targeting and cleavage is decreased and, consequentially, editing events are decreased. Much more transformation candidates would have to be tested for editing features. Alternatives to this replication origins are replicons known to function in LAB, such as pAMβ1 and pWV01 replicons.[133,134] The pvW01-derivative pVE6002 thermosensitive replicon, non-functional from 35ºC, is also a possible alternative.[135] In fact, thermosensitive replicons may be advantageous in this kind of systems as the plasmid curing efficacy is increased.[90]

The insertion of the fragment containing the sgRNA and homology arms was made by the SpeI restriction site, present in the J23119 promoter. Hence, the distance between the end of the promoter and the beginning of the guide RNA sequence is 2 bp, whereas the distance between the promoter and the sgRNA of Huang *et al*. design is 19 bp[97]. This may not be problematic as the objective is to obtain a non-translatable guide transcript. In fact, the base pair reduction between these components may reduce some instability effects that larger intermediated fragments may cause.

The tipA promoter should be responsible, in the nthCRISPRa plasmid, for the expression of Scocas9 *in vivo*. The transcription of this promoter is known to be highly induced by thiostrepton and other thiostrepton-like antibiotics, even when using extremely low concentrations.[136–138] This system was already tested in thiostrepton non-producing *Streptomyces* organisms, such as *Streptomyces lividans*.[139] The construction of a system with a tipA promoter upstream of a secretion Tendamistat coding gene from *Streptomyces tendae* showed little secretion in the absence of thiostrepton and almost an 80-fold secretion increase in presence of the antibiotic (5 μg/mL), in *S. lividans*. Therefore, in this organism, tipA promoter reveals expression with and without thiostrepton, but is highly inducible by the latter.[139] In the present work, the thiostrepton antibiotic was not used. As no *L. lactis* transformation candidates were obtained, one cannot conjecture about the activity of the tipA promoter in this organism. If colonies did contain nthCRISPRa and

were not edited, it could be due to low expression of the Scocas9 and lack of DSB events. In this case, the addition of thiostrepton to the medium could be tested to see whether edited cells would grow. However, one cannot make sure that the presence of thiostrepton would not impair *L. lactis* growth profiles. In that case, the induction of expression using extremely low concentrations of thiostrepton would be an advantage, as such low concentration could be sufficient to promote expression to the desired extent and not enough to impair *L. lactis* growth. The best-case scenario would be that tipA promoter would have a basal expression, enough to produce the sufficient amounts of Scocas9 to induce the desired DSB. As this system has not been described in *L. lactis* yet, one cannot predict whether it will work. An alternative would be to change this promoter with native promoters, such as the inducible nisin A promoter or the constitutive promoters P45 or P32.[140] The use of conditional or weak promoter seems advantageous when considering that the overexpression of Cas9 proteins can render toxic effects.[141] Furthermore, other problematic is the codon usage of the Scocas9 coding gene itself. In nthCRISPRa, the sequence that codes for the Cas9 protein is codon optimized for *S. coelicor*. Although being both Gram-positive bacteria, *S. coelicor* and *L. lactis* have very dissimilar GC contents of approximately 70% and 35%, respectively[104,142], and given this disparity, the preferred codons and regulatory elements used by one organism in one end of the GC content spectrum may not be recognized by organisms on the other end.[143] If the editing system fails to work with nthCRISPRe, the Scocas9 coding sequence may need to be reconstructed using *L. lactis* codon usage bias. The commonly used SpyCas9 could also be used, as the GC content of *Streptococcus pyogenes* is of approximately 36%[144], resembling more the one observed in *L. lactis*.

The present expression and editing system also contains an oriT region. As previously discussed, transformation by electroporation in lactic acid bacteria is a laborious and non-replicative between different strains; therefore, using different DNA introduction methods may be advantageous. The presence of an oriT region in nthCRISPRa/nthCRISPRe conduces to the consideration that this plasmid(s) may be transformed in *L. lactis* LMG 19460 by conjugation, in the future, if transformation fails to work.

5 Overview, Final Remarks and Future Perspectives

To establish an efficient Lactic acid bacteria platform to produce high levels of high quality plasmid DNA, hurdles concerning products that withdraw this process must be exceeded. Furthermore, although being more suitable for gene therapy purposes, DNA production by LAB should be comparable to the one observed in *E. coli*, the most used host for DNA production worldwide. To reach that thesis and achieve high plasmid yields, the knockout of the endonuclease III coding gene, *nth*, was attempted using the CRISPR/Cas9 editing system, designed by Huang *et al.* 2015[97]. Previous to the present work, the system was redesigned for *nth* deletion in *L. lactis* LMG 19460. To achieve the main goal of this project, other knockouts should be tested in the future, including *recA*, *pgi* and *pyk*, knockouts which proved to be advantageous in *E. coli* in the past[121]. Other interesting deletion candidates were also predicted using the OptFlux Software and can be tested in a proximate future.

Transformation of lactic acid bacteria with the nthCRISPRa plasmid was attempted by electroporation using different pulses and voltages. However, transforming the nthCRISPRa plasmid did not prove to be an easy task and transformants were not obtained. The desired edition and the nthCRISPRa plasmid were not detected in the supposedly transformed cells that manifested growth, using previously defined minimal inhibitory concentrations and higher concentrations of apramycin. This led to the conclusion that apramycin is not the indicated antibiotic to be used in lactic acid bacteria selection, as medium containing this compound is either selecting non-sensible cells or spontaneous resistant organisms.

In what concerns PCR protocols for assessment of edition and plasmid presence after transformation, Nth_conf primers, that only generate amplicons in the presence of genomic DNA, have generated a fragment using purified plasmid DNA as template, which suggests contamination of pDNA with gDNA. One can conclude that using purified pDNA as template for the generation of amplicons that are present in genomic and plasmid DNA is not a good approach, at least when using the same protocol for pDNA extraction from gram-positive bacteria, used in the present work. New primers were designed as to generate an amplicon that is only present in the nthCRISPRa plasmid. Using the newly designed primers Cas9_conf and the primers Nth_conf, a fast assay to test the presence of edited cells and the presence of the plasmid inside *L. lactis* LMG 19460 was established, and can be used in further studies. The PCR procedure using 250,000 cells is faster than the previously used detection method, which required extraction of gDNA and pDNA, and independent PCR programmes, and revealed to generate reliable products. Nonetheless, it would be advantageous to use an alternative confirmation method or counterselection procedure, to increase the probability of testing an edition positive colony. Otherwise, much more candidates should be tested.

As we were convinced that the apramycin selection marker was not the most suitable for *L. lactis* LMG 19460 transformants selection, and sensibility tests had not been done in the present model organism

before this project implementation, it was decided to change the selection marker to a well-established one. The erythromycin selection marker must be introduced in the apramycin resistance region by Gibson assembly, as no restriction sites are available to cleanly remove the apramycin resistance coding sequence, without compromising the other elements present in the nthCRISPRa plasmid. PCR amplifications for Gibson assembly proven that PCR amplifications may be laborious when they are not optimized and that they depend on a variety of factors, such as template DNA quality and concentration, DNA polymerase, annealing temperature, co-factors presence, etc. Although using similar conditions, generation of amplicons were not always totally reproducible, indicating that each PCR reaction preparation may harbour small differences in sample preparations that cause changes in amplification profiles, when reactions are not totally optimized. Gibson assembly is yet to be achieved, and we may conclude that the ligation has not taken place properly as Scocas9 or erythromycin resistance internal fragments are detectable by colony PCR, but cells do not contain plasmid DNA after the growth protocol and plasmid DNA purification. The successful transformation of the linear fragments inside cells would explain this fact and the cell growth in erythromycin containing media. To ascertain fragments ligation before growing candidates in liquid medium, primers that prime forwardly one of the fragments and reversely other fragment can be designed.

Once the *nth* knockout of *L. lactis* LMG 19460 and the plasmid curing is accomplished, a quantitative Real-Time PCR (qRT-PCR) relative method should be performed, to quantify plasmid DNA inside cells. To analyse plasmid copy number, pTRKH3 could be used. A comparison between *L. lactis* LMG 10460 wild type cells and *L. lactis* LMG 10460 *Δnth* would be made to infer whether the knockout had improved pTRKH3 plasmid copy number. This qRT-PCR method would be made based on the ratio between the *erm* gene, present in pTRKH3, and the amplification of the single-copy genome reference gene, *feoA*. Furthermore, pDNA supercoiled fractions would be analysed by image analysis of agarose gel electrophoresis.

The main concern of the present work is the system feasibility. As it was designed to suit *Streptomyces* genome edition, replicative, transcriptional, expression and regulatory elements of *L. lactis* LMG 19460 may not recognize some plasmid elements in the nthCRISPRa/nthCRISPRe, which is corroborated with the fact that *Streptomyces* and *Lactococcus* have very dissimilar GC contents. If the pSG5 replicative origin fails to work in *L. lactis*, the system highly decreases its efficiency as we would be using a non-replicative plasmid to express enough Scocas9 endonuclease enzyme to induce DSB and much less DNA template repair would be available for recombination events to occur. Furthermore, if *L. lactis* LMG does weakly recognise the expression system of the nthCRISPRa/nthCRISPRe plasmid, Scocas9 may be poorly or not expressed, which would impair the recombination event. Also, the uncertainty of the tipA promoter function without the induction of the thiostrepton antibiotic is still clearly known, and its diminished activity also impairs the DSB and, consequentially, the recombination event. Previously, the susceptibility of *L. lactis* to this antibiotic would have to be tested.

Alternatively, the system could be redesigned to incorporate native or previously tested regulatory and expression elements from *L. lactis* or from more closely related organisms. Other editing systems such as

CRISPR/Cas9 selection combined with ssDNA recombineering can also be used. The latter has indeed been successfully used for mutagenesis and deletions in *Lactobacillus reuteri*, a lactic acid bacterium.

CRISPR-based genome editing techniques were quickly adopted for eukaryotic applications. However, their use in prokaryotic applications have lagged behind, mainly due to technical shortcomings, as transformation efficiencies and the still lacking characterization of many prokaryotic regulatory mechanisms and optimal growth conditions. Additionally, there is a great need of characterization and development of expression and regulatory systems present in plasmids, as well as the elaboration of common vector systems for different organisms. Basic microbiology and molecular biology are very important to develop new genome editing techniques, and will allow the extended application of CRISPR/Cas applications. There is no definitive answer for which is the best genome editing technique, but CRISPR/Cas9 seems to be of great promising applications as soon as the system is established. One great possibility of this tool is the multiplexing, which will allow for the edition of different target sites in a single editing event.

6 References

1. Zhu, Y., Zhang, Y. & Li, Y. Understanding the industrial application potential of lactic acid bacteria through genomics. *Appl. Microbiol. Biotechnol.* **83,** 597–610 (2009).
2. Makarova, K. *et al.* Comparative genomics of the lactic acid bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 15611–15616 (2006).
3. *Biology of Microorganisms on Grapes, in Must and in Wine*. (Springer-Verlag Berlin Heidelberg, 2009).
4. Pontes, D. S. *et al. Lactococcus lactis* as a live vector: Heterologous protein production and DNA delivery systems. *Protein Expr. Purif.* **79,** 165–175 (2011).
5. Carr, F. J., Chill, D. & Maida, N. The lactic acid bacteria: a literature survey. *Crit. Rev. Microbiol.* **28,** 281–370 (2002).
6. Hayek, S. A. & Ibrahim, S. A. Current limitations and challenges with lactic acid bacteria: a review. *Food Nutr. Sci.* **4,** 73 (2013).
7. Cano-Garrido, O., Seras-Franzoso, J. & Garcia-Fruitós, E. Lactic acid bacteria : reviewing the potential of a promising delivery live vector for biomedical purposes. *Microb. Cell Fact.* **14,** 137 (2015).
8. Steidler, L. & Rottiers, P. Therapeutic drug delivery by genetically modified *Lactococcus lactis*. *Ann. N. Y. Acad. Sci.* **1072,** 176–186 (2006).
9. Panagiota, F.-P., Christaki, E. & Bonos, E. Lactic acid bacteria as source of functional ingredients. in *Lactic Acid Bacteria-R & D for Food, Health and Livestock Purposes* 489–614 (2013).
10. Pessione, E. Lactic acid bacteria contribution to gut microbiota complexity : lights and shadows. *Front. Cell. Infect. Microbiol.* **2,** (2012).
11. Wells, J. M. & Mercenier, A. Mucosal delivery of therapeutic and prophylactic molecules using lactic acid bacteria. *Nat. Rev. Microbiol.* **6,** 349–362 (2008).
12. Bermúdez-Humarán, L. G. *et al.* Engineering lactococci and lactobacilli for human health. *Curr. Opin. Microbiol.* **16,** 278–283 (2013).
13. Vesković Moračanin, S. M., Dukić, D. A. & Memiši, N. R. Bacteriocins produced by lactic acid bacteria a review article. *APCBEE Procedia* **2,** 50–56 (2012).
14. García-fruitós, E. Lactic acid bacteria: a promising alternative for recombinant protein production. *Microb. Cell Fact.* **11,** 157 (2012).
15. Ferrer-Miralles, N. & Villaverde, A. Bacterial cell factories for recombinant protein production; expanding the catalogue. 113 (2013).
16. Vos, W. M. De. Systems solutions by lactic acid bacteria: from paradigms to practice. *Microb. Cell Fact.* **10,** Suppl 1:S2 (2011).
17. Morello, E. *et al. Lactococcus lactis*, an efficient cell factory for recombinant protein production and secretion. *J. Mol. Microbiol. Biotechnol.* **14,** 48–58 (2008).
18. Daniel, C., Roussel, Y., Kleerebezem, M. & Pot, B. Recombinant lactic acid bacteria as mucosal biotherapeutic agents. *Trends Biotechnol.* **29,** 499–508 (2011).
19. Bermúdez-humarán, L. G., Kharrat, P., Chatel, J.-M. & Langella, P. Lactococci and lactobacilli as mucosal delivery vectors for therapeutic proteins and DNA vaccines. *Microb. Cell Fact.* **10,** S4 (2011).
20. Glenting, J. *et al.* Production of recombinant peanut allergen Ara h 2 using *Lactococcus lactis*. *Microb. Cell Fact.* **6,** 28 (2007).
21. Bermudez-Humaran, L., Cortes-Perez, N., L'Haridon, R. & Langella, P. Production of biological

active murine IFN-γ by recombinant *Lactococcus lactis*. *FEMS Microbiol. Lett.* **280,** 144–149 (2008).

22. Wells, J. M. Immunomodulatory mechanisms of lactobacilli. *Microb. Cell Fact.* **10,** S17 (2011).
23. LeBlanc, J. G. *et al.* Mucosal targeting of therapeutic molecules using genetically modified lactic acid bacteria: an update. *FEMS Microbiol. Lett.* **344,** 1–9 (2013).
24. Faurez, F., Dory, D., Moigne, V. Le, Gravier, R. & Jestin, A. Biosafety of DNA vaccines: New generation of DNA vectors and current knowledge on the fate of plasmids after injection. *Vaccine* **28,** 3888–3895 (2010).
25. Glenting, J. & Wessels, S. Ensuring safety of DNA vaccines. *Microb. Cell Fact.* **4,** 26 (2005).
26. Bahey-El-Din, M., Gahan, C. G. & Griffin, B. T. *Lactococcus lactis* as a cell factory for delivery of therapeutic proteins. *Curr. Gene Ther.* **10,** 34–45 (2010).
27. Mancha-Agresti, P. *et al.* A New Broad Range Plasmid for DNA Delivery in Eukaryotic Cells Using Lactic Acid Bacteria: In Vitro and In Vivo Assays. *Mol. Ther. Methods Clin. Dev.* **4,** 83–91 (2017).
28. Kay, M. A. State-of-the-art gene-based therapies: the road ahead. *Nat. Rev. Genet.* **12,** 316–328 (2011).
29. Gardlík, R. *et al.* Vectors and delivery systems in gene therapy. *Med. Sci. Monit.* **11,** RA110-RA121 (2005).
30. Gene Therapy Clinical Trials Worldwide. Available at: http://www.abedia.com/wiley/. (Accessed: 20th October 2017)
31. Kutzler, M. A. & Weiner, D. B. DNA vaccines : ready for prime time ? *DNA vaccines ready prime time?* **9,** 776–788 (2008).
32. Edelstein, M. L., Abedi, M. R., Wixon, J. & Edelstein, R. M. Gene therapy clinical trials worldwide 1989–2004—an overview. *J. Gene Med.* **6,** 597–602 (2004).
33. Anson, D. S. The use of retroviral vectors for gene therapy-what are the risks? A review of retroviral pathogenesis and its relevance to retroviral vector-mediated gene delivery. *Genet. Vaccines Ther.* **2,** 9 (2004).
34. Gardlík, R. & Pálffy, R. Vectors and delivery systems in gene therapy. (2014).
35. Niidome, T. & Huang, L. Gene therapy progress and prospects: nonviral vectors. *Gene Ther.* **9,** 1647 (2002).
36. Yin, H. *et al.* Non-viral vectors for gene-based therapy. *Nat. Rev. Genet.* **15,** 541–555 (2014).
37. Mairhofer, J. & Grabherr, Æ. R. Rational vector design for efficient non-viral gene delivery: challenges facing the use of plasmid DNA. *Mol. Biotechnol.* **39,** 97–104 (2008).
38. Walker, W. E., Porteous, D. J. & Boyd, A. C. The effects of plasmid copy number and sequence context upon transfection efficiency. *J. Control. release* **94,** 245–252 (2004).
39. Yin, W., Xiang, P. & Li, Q. Investigations of the effect of DNA size in transient transfection assay using dual luciferase system. *Anal. Biochem.* **346,** 289–294 (2005).
40. Remaut, K., Sanders, N. N., Fayazpour, F., Demeester, J. & Smedt, S. C. De. Influence of plasmid DNA topology on the transfection properties of DOTAP/DOPE lipoplexes. *J. Control. release* **115,** 335–343 (2006).
41. FDA. Guidance for Industry: Considerations for Plasmid DNA Vaccines for Infectious Disease Indications. Available at: https://www.fda.gov/. (Accessed: 28th October 2017)
42. Wells, J. Mucosal vaccination and therapy with genetically modified lactic acid bacteria. *Annu. Rev. Food Sci. Technol.* **2,** 423–445 (2011).
43. Chatel, J. M. *et al.* In vivo transfer of plasmid from food-grade transiting lactococci to murine epithelial cells. *Gene Ther.* **15,** 1184–1190 (2008).

44. Guimarães, V. D. *et al.* Internalin-expressing *Lactococcus lactis* is able to invade small intestine of guinea pigs and deliver DNA into mammalian epithelial cells. *Microbes Infect.* **7,** 836–844 (2005).

45. Gram, G. J., Fomsgaard, A., Thorn, M., Madsen, S. M. & Glenting, J. Immunological analysis of a *Lactococcus lactis*-based DNA vaccine expressing HIV gp120. *Genet. Vaccines Ther.* **5,** 3 (2007).

46. Erridge, C., Bennett-Guerrero, E. & Poxton, I. R. Structure and function of lipopolysaccharides. *Microbes Infect.* **4,** 837–851 (2002).

47. Son, Y. J., Ryu, A. J., Li, L., Han, N. S. & Jeong, K. J. Development of a high-copy plasmid for enhanced production of recombinant proteins in *Leuconostoc citreum*. *Microb. Cell Fact.* **15,** 12 (2016).

48. Gonçalves, G. A. L., Bower, D. M., Prazeres, D. M. F., Monteiro, G. A. & Prather, K. L. J. Rational engineering of Escherichia coli strains for plasmid biopharmaceutical manufacturing. 251–261 (2012). doi:10.1002/biot.201100062

49. Gonçalves, G. A. L., Prather, K. L. J., Monteiro, G. A., Carnes, A. E. & Prazeres, D. M. F. Plasmid DNA production with *Escherichia coli* GALG20, a pgi-gene knockout strain: Fermentation strategies and impact on downstream processing. *J. Biotechnol.* **186,** 119–127 (2014).

50. Nemudryi, A. A., Valetdinova, K. R., Medvedev, S. P. & Zakian, S. M. TALEN and CRISPR/Cas genome editing systems: tools of discovery. *Acta Naturae (англоязычная версия)* **6,** (2014).

51. Takata, M. *et al.* Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.* **17,** 5497–5508 (1998).

52. Szostak, J. W., Orr-weaver, T. L., Rothstein, R. J. & Stahl, F. W. The double-strand-break repair model for recombination. *Cell* **33,** 25–35 (1983).

53. Rouet, P., Smih, F. & Jasin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl. Acad. Sci.* **91,** 6064–6068 (1994).

54. Maeder, M. L. & Gersbach, C. A. Genome-editing technologies for gene and cell therapy. *Mol. Ther.* **24,** 430–446 (2016).

55. Lieber, M. R., Ma, Y., Pannicke, U. & Schwarz, K. Mechanism and regulation of human non-homologous DNA end-joining. *Nat. Rev. Mol. cell Biol.* **4,** (2003).

56. Chen, F. *et al.* High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat. Methods* **8,** 753–755 (2011).

57. Bedell, V. M. *et al.* In vivo genome editing using a high-efficiency TALEN system. *Nature* **491,** 114–118 (2012).

58. Xiao, A. *et al.* Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Res.* **41,** e141–e141 (2013).

59. Şöllü, C. *et al.* Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Res.* **38,** 8269–8276 (2010).

60. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169,** 5429–5433 (1987).

61. Haft, D. H., Selengut, J., Mongodin, E. F. & Nelson, K. E. A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput. Biol.* **1,** e60 (2005).

62. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151,** 2551–2561 (2005).

63. Mojica, F. J. M., Dıez-Villasenor, C., García_Martínez, J. & Soria, E. Intervening sequences of

regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **60,** 174–182 (2005).

64. Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* **8,** 172 (2007).

65. Makarova, K. S. *et al.* Evolution and classification of the CRISPR–Cas systems. *Nat. Rev. Microbiol.* **9,** 467–477 (2011).

66. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science (80-.).* **346,** 1258096 (2014).

67. Barrangou, R. & Pijkeren, J.-P. van. Exploiting CRISPR–Cas immune systems for genome editing in bacteria. *Curr. Opin. Biotechnol.* **37,** 61–68 (2016).

68. Mojica, F. J. M., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155,** 733–740 (2009).

69. Swarts, D. C., Mosterd, C., Passel, M. W. J. Van & Brouns, S. J. J. CRISPR interference directs strand specific spacer acquisition. *PLoS One* **7,** e35888 (2012).

70. Bondy-Denomy, J., Pawluk, A., Maxwell, K. l. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493,** 429–432 (2013).

71. Makarova, K. S., Aravind, L., Wolf, Y. I. & Koonin, E. V. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct* **6,** 38 (2011).

72. Stoddard, B. L. & Fox, K. Editorial : CRISPR in Nucleic Acids Research. *Nucleic Acids Res.* **44,** 4989–4990 (2016).

73. Jinek, M. *et al.* A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science (80-.).* **337,** 816–821 (2012).

74. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science (80-.).* **339,** 823–826 (2013).

75. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science (80-.).* **339,** 819–823 (2013).

76. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31,** 230–232 (2013).

77. Boettcher, M. & McManus, M. T. Choosing the right tool for the job: RNAi, TALEN, or CRISPR. *Mol. Cell* **58,** 575–585 (2015).

78. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157,** 1262–1278 (2014).

79. O'Geen, H., Henry, I. M., Bhakta, M. S., Meckler, J. F. & Segal, D. J. A genome-wide analysis of Cas9 binding specificity using ChIP-seq and targeted sequence capture. *Nucleic Acids Res.* **43,** 3389–3404 (2015).

80. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32,** 279–284 (2014).

81. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci.* **109,** E2579–E2586 (2012).

82. Sander, J. D. & Joung, J. K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* **32,** 347–355 (2014).

83. Gilbert, L. A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159,** 647–661 (2014).

84. Mougiakos, I., Bosma, E. F., de Vos, W. M., van Kranenburg, R. & van der Oost, J. Next generation prokaryotic engineering: the CRISPR-Cas toolkit. *Trends Biotechnol.* **34,** 575–587 (2016).

85. Tian, P. *et al.* Fundamental CRISPR-Cas9 tools and current applications in microbial systems. *Synth. Syst. Biotechnol.* **2,** 219–225 (2017).

86. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31,** 233–239 (2013).

87. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Coupling the CRISPR/Cas9 system with lambda red recombineering enables simplified chromosomal gene replacement in *Escherichia coli*. *Appl. Environ. Microbiol.* **81,** 5103–5114 (2015).

88. Jakočiūnas, T., Jensen, M. K. & Keasling, J. D. CRISPR/Cas9 advances engineering of microbial cell factories. *Metab. Eng.* **34,** 44–59 (2016).

89. Reisch, C. R. & Prather, K. L. J. The no-SCAR (Scarless Cas9 Assisted Recombineering) system for genome editing in *Escherichia coli*. *Sci. Rep.* **5,** 15096 (2015).

90. Oh, J.-H. & van Pijkeren, J.-P. CRISPR–Cas9-assisted recombineering in *Lactobacillus reuteri*. *Nucleic Acids Res.* **42,** e131–e131 (2014).

91. van Pijkeren, J.-P. & Britton, R. A. Precision genome engineering in lactic acid bacteria. *Microb. Cell Fact.* **13,** S10 (2014).

92. Jiang, Y. *et al.* Multigene editing in the Escherichia coli genome via the CRISPR-Cas9 system. *Appl. Environ. Microbiol.* **81,** 2506–2514 (2015).

93. Standage-Beier, K., Zhang, Q. & Wang, X. Targeted large-scale deletion of bacterial genomes using CRISPR-nickases. *ACS Synth. Biol.* **4,** 1217–1225 (2015).

94. Wang, Y. *et al.* Markerless chromosomal gene deletion in *Clostridium beijerinckii* using CRISPR/Cas9 system. *J. Biotechnol.* **200,** 1–5 (2015).

95. Xu, T. *et al.* Efficient genome editing in *Clostridium cellulolyticum* via CRISPR-Cas9 nickase. *Appl. Environ. Microbiol.* **81,** 4423–4431 (2015).

96. Ravachol, J. *et al.* Combining free and aggregated cellulolytic systems in the cellulosome-producing bacterium *Ruminiclostridium cellulolyticum*. *Biotechnol. Biofuels* **8,** 114 (2015).

97. Huang, H., Zheng, G., Jiang, W., Hu, H. & Lu, Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in Streptomyces. *Acta Biochim. Biophys. Sin. (Shanghai).* **47,** 231–243 (2015).

98. Cobb, R. E., Wang, Y. & Zhao, H. High-efficiency multiplex genome editing of Streptomyces species using an engineered CRISPR/Cas system. *ACS Synth. Biol.* **4,** 723–728 (2014).

99. Zeng, H. *et al.* Highly efficient editing of the actinorhodin polyketide chain length factor gene in *Streptomyces coelicolor* M145 using CRISPR/Cas9-CodA (sm) combined system. *Appl. Microbiol. Biotechnol.* **99,** 10575–10585 (2015).

100. Tong, Y., Charusanti, P., Zhang, L., Weber, T. & Lee, S. Y. CRISPR-Cas9 based engineering of actinomycetal genomes. *ACS Synth. Biol.* **4,** 1020–1029 (2015).

101. Francke, C., Siezen, R. J. & Teusink, B. Reconstructing the metabolic network of a bacterium from its genome. *Trends Microbiol.* **13,** 550–558 (2005).

102. Becker, S. A. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat. Protoc.* **2,** 727–738 (2007).

103. Rocha, I. *et al.* OptFlux: an open-source software platform for *in silico* metabolic engineering. *BMC Syst. Biol.* **4,** 45 (2010).

104. Silva, I., Duarte, S., Moreira, L. M. & Monteiro, G. A. Draft Genome Sequence of the Plasmid-Free *Lactococcus lactis* subsp. lactis Strain LMG 19460. *Genome Announc.* **5,** e00210-17 (2017).

105. Bolotin, A. *et al.* The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp.

lactis IL1403. *Genome Res.* **11,** 731–753 (2001).

106. Jones, S. E. & Versalovic, J. Probiotic Lactobacillus reuteri biofilms produce antimicrobial and anti-inflammatory factors. *BMC Microbiol.* **9,** 35 (2009).

107. Martins, M. S. C. Redesign of Lactic Acid Bacteria for plasmid DNA production A strategy for increase the plasmid copy number Examination Committee. (Instituto Superior Técnico, Universidade de Lisboa, 2015).

108. Holmes, D. S. & Quigley, M. A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* **114,** 193–197 (1981).

109. NEBuilder Assembly Tool. Available at: http://nebuilder.neb.com/. (Accessed: 3rd April 2017)

110. Oliveira, A. P., Nielsen, J. & Förster, J. Modeling Lactococcus lactis using a genome-scale flux model. **15,** 1–15 (2005).

111. National Center for Biotechnological Information. Available at: https://www.ncbi.nlm.nih.gov/. (Accessed: 20th January 2017)

112. Guillot, A., Gitton, C., Anglade, P. & Mistou, M.-Y. Proteomic analysis of *Lactococcus lactis*, a lactic acid bacterium. *Proteomics* **3,** 337–354 (2003).

113. Rijnen, L., Bonneau, S. & Yvon, M. Genetic characterization of the major lactococcal aromatic aminotransferase and its involvement in conversion of amino acids to aroma compounds. *Appl. Environ. Microbiol.* **65,** 4873–4880 (1999).

114. KEGG: Kyoto Encyclopedia of Genes and Genomes. Available at: http://www.kegg.jp/kegg/pathway.html#metabolism. (Accessed: 18th January 2017)

115. Borja, G. M. *et al.* Engineering *Escherichia coli* to increase plasmid DNA production in high cell-density cultivations in batch mode. *Microb. Cell Fact.* **11,** 132 (2012).

116. Schleef, M. Plasmid-Based Medicinal Products – Focus on pFAR: A Miniplasmid Free of Antibiotic Resistance Markers. in *Minicircle and miniplasmid dna vectors: the future of non-viral and viral gene transfer* 37–58 (John Wiley & Sons, 2013).

117. Subramanian, G. endA and recA. in *Biopharmaceutical Production Technology, 2 Volume Set* 12 (John Wiley & Sons, 2012).

118. Miller, R. V & Kokjohn, T. A. General microbiology of recA: environmental and evolutionary significance. *Annu. Rev. Microbiol.* **44,** 365–394 (1990).

119. Duwat, P., Ehrlich, S. D. & Gruss, A. The recA gene of Lactococcus lactis: characterization and involvement in oxidative and thermal stress. *Mol. Microbiol.* **17,** 1121–1131 (1995).

120. Hua, Q., Yang, C., Baba, T., Mori, H. & Shimizu, K. Responses of the central metabolism in Escherichia coli to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts. *J. Bacteriol.* **185,** 7053–7067 (2003).

121. Gonçalves, G. A. L., Prazeres, D. M. F., Monteiro, G. A. & Prather, K. L. J. De novo creation of MG1655-derived *E. coli* strains specifically designed for plasmid DNA production. *Appl. Microbiol. Biotechnol.* **97,** 611–620 (2013).

122. Cunningham, D. S. *et al.* Pyruvate kinase-deficient Escherichia coli exhibits increased plasmid copy number and cyclic AMP levels. *J. Bacteriol.* **191,** 3041–3049 (2009).

123. New England BioLabs. Cleavage Close to the End of DNA Fragments. Available at: https://www.neb.com/. (Accessed: 20th July 2017)

124. Spectinomycine. Available at: http://www.sigmaaldrich.com/catalog/product/sigma/s0692?lang=pt&region=PT. (Accessed: 2nd February 2017)

125. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular cloning: a laboratory manual. Molecular cloning: a laboratory manual.* (Cold Spring Harbor Laboratory Press, 1989).

126. Palomino, M. M., Allievi, M. C., Prado-Acosta, M., Sanchez-Rivas, C. & Ruzal, S. M. New method for electroporation of Lactobacillus species grown in high salt. *J. Microbiol. Methods* **83,** 164–167 (2010).

127. Adamovic, N. Optimization of electrotransformation of industrial lactic acid bacteria strains. (Universitat Wien, 2011).

128. Karapetkov, N., Georgieva, R., Rumyan, R. & Karaivanova, E. Antibiotic susceptibility of different lactic acid bacteria strains. *Benef. Microbes* **2,** 335–339 (2011).

129. Roux, K. H. Optimization and troubleshooting in PCR. *Cold Spring Harb. Protoc.* **4,** pdb-ip66 (2009).

130. Lorenz, T. C. Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *J. Vis. Exp. JoVE* (2012).

131. Owczarzy, R., Moreira, B. G., You, Y., Behlke, M. A. & Wälder, J. A. Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochemistry* **47,** 5336–5353 (2008).

132. Muth, G., Nußbaumer, B., Wohlleben, W. & Pühler, A. A vector system with temperature-sensitive replication for gene disruption and mutational cloning in streptomycetes. *Mol. Gen. Genet. MGG* **219,** 341–348 (1989).

133. Foley, S., Bron, S., Venema, G., Daly, C. & Fitzgerald, G. F. Molecular Analysis of the Replication Origin of the *Lactococcus lactis* Plasmid pCI305. *Plasmid* **36,** 125–141 (1996).

134. Leenhouts, K. J. *et al.* Nucleotide sequence and characterization of the broad-host-range lactococcal plasmid pWVO1. *Plasmid* **26,** 55–66 (1991).

135. Maguin, E., Duwat, P., Hege, T., Ehrlich, D. & Gruss, A. New thermosensitive plasmid for gram-positive bacteria. *J. Bacteriol.* **174,** 5633–5638 (1992).

136. Holmes, D. J., Caso, J. L. & Thompson, C. J. Autogenous transcriptional activation of a thiostrepton-induced gene in *Streptomyces lividans*. *EMBO J.* **12,** 3183 (1993).

137. Murakami, T., Holt, T. G. & Thompson, C. J. Thiostrepton-induced gene expression in Streptomyces lividans. *J. Bacteriol.* **171,** 1459–1466 (1989).

138. Yun, B.-S., Hidaka, T., Furihata, K. & Seto, H. Promothiocins A and B, novel thiopeptides with a tip A promoter inducing activity produced by Streptomyces sp. SF2741. *J. Antibiot. (Tokyo).* **47,** 510–514 (1994).

139. Schmitt-John, T. & Engels, J. W. Promoter constructions for efficient secretion expression in *Streptomyces lividans*. *Appl. Microbiol. Biotechnol.* **36,** 493–498 (1992).

140. Zhu, D. *et al.* Isolation of strong constitutive promoters from *Lactococcus lactis* subsp. lactis N8. *FEMS Microbiol. Lett.* **362,** (2015).

141. Jiang, W., Brueggeman, A. J., Horken, K. M., Plucinak, T. M. & Weeks, D. P. Successful transient expression of Cas9 and single guide RNA genes in Chlamydomonas reinhardtii. *Eukaryot. Cell* **13,** 1465–1469 (2014).

142. Wright, F. & Bibb, M. J. Codon usage in the G+ C-rich Streptomyces genome. *Gene* **113,** 55–65 (1992).

143. Primrose, S. B. & Twyman, R. Cloning in Gram-positive bacteria. in *Principles of Gene Manipulation and Genomics* 189–190 (John Wiley & Sons, 2013).

144. Maruyama, F., Watanabe, T. & Nakagawa, I. Streptococcus pyogenes Genomics. in *Streptococcus pyogenes: Basic Biology to Clinical Manifestations* 1–80 (2016).

Supplemental Data

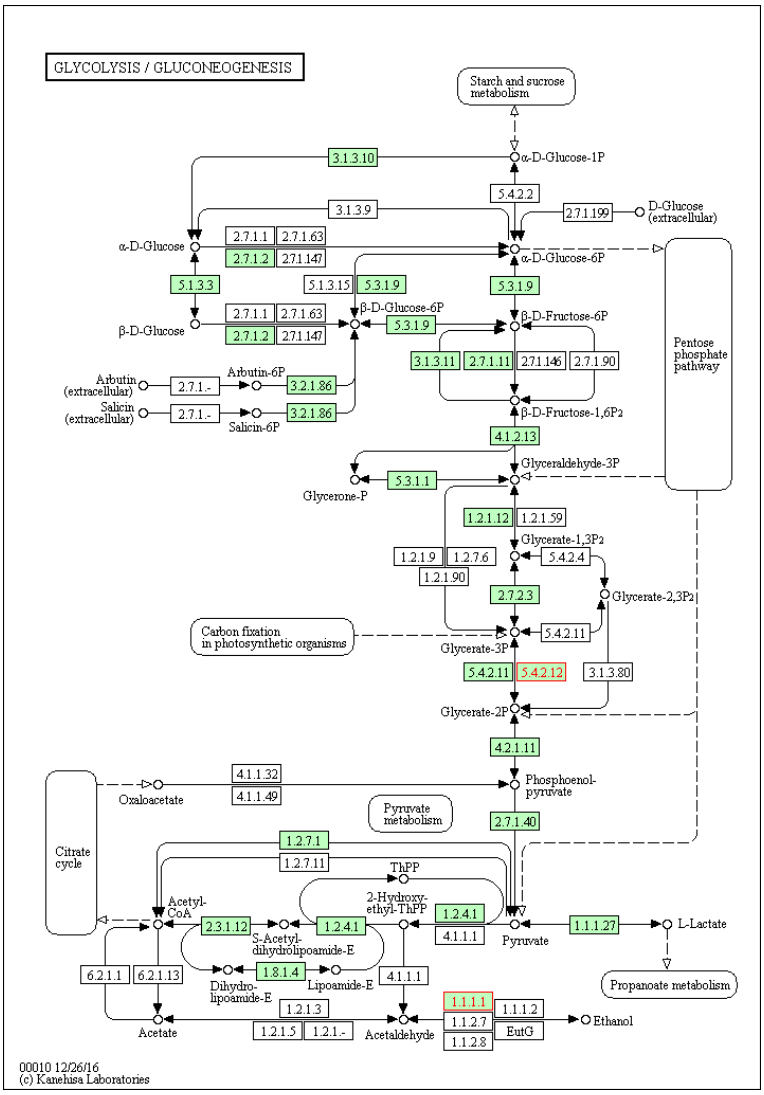

Figure S 1 Position of the yjhF (5.4.2.12) and adhE (1.1.1.1) genes from L. lactis IL1403 identified by the Optflux software as possible knockout targets in the glycolysis/gluconeogenesis pathway.[114]

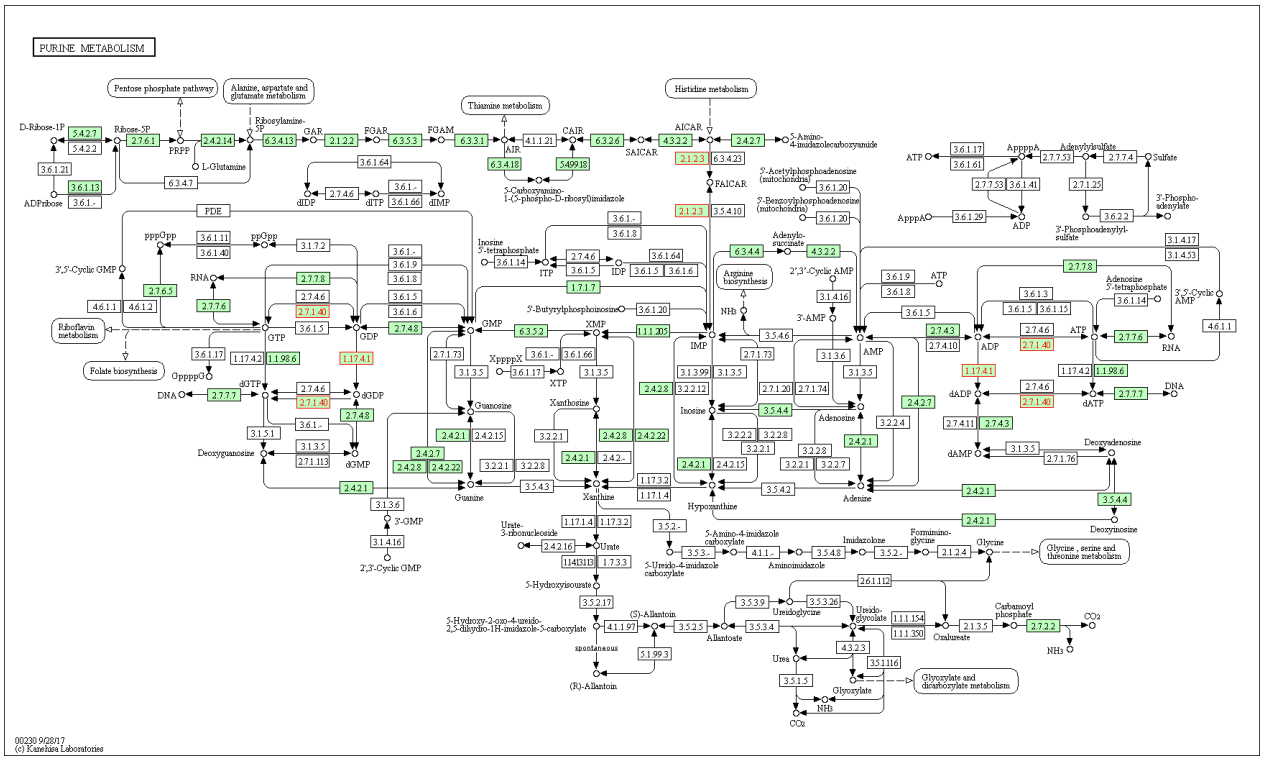

Figure S 2 Position of the purH (2.1.2.3) and nrdF (1.17.4.1) genes in the purine metabolism pathway from L. lactis IL1403 identified by the Optflux software as possible knockout targets and position of the pyk (2.7.1.40) gene in the purine metabolism pathway as a theoretically possible knockout target.[114]

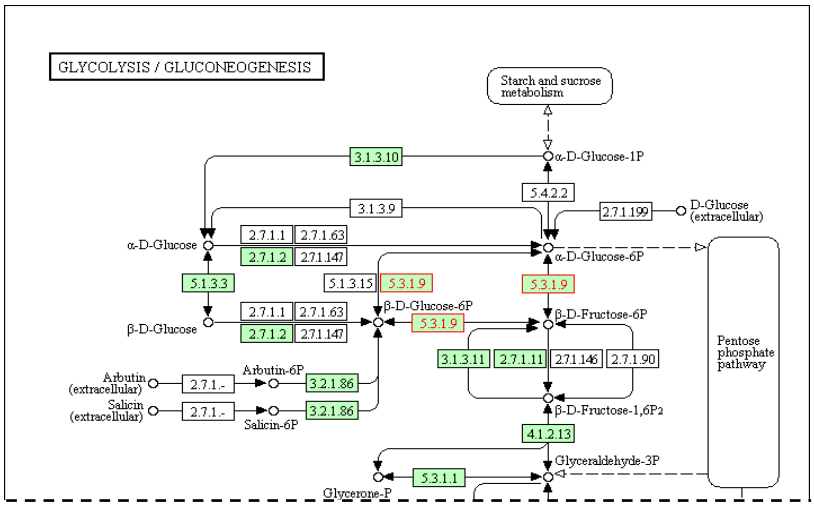

Figure S 3 Position of the pgi gene (5.3.1.9) in the glycolysis/gluconeogenesis pathway from L. lactis IL1403 theoretically proposed as a possible knockout target.[114]

Table S 1 Primer sequences for new constituent fragments amplification of the nthCRISPR plasmid for future CRISPR/Cas9-based gene knockouts. Highlighted nucleotides correspon to newly added restriction sites for insertion in the vector plasmid: light grey highlights correspond to restriction site sequences SpeI (A^CTAGT) and ApaI (GGGCC^C) and dark gre highlights correspond to added nucleotides for aided restriction enzyme docking and digestion efficiency. Lowercase nucleotide representative letters correspond to overlappin sequences for SOEing PCR.

Gene designation	Primers (5'-3')			Product length (bp)	
recA	sgRNA	Forward	CGACTAGTGGGTGCAGGTGGATATCCTAAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	132	
		Reverse	acttctggtaactctggcatAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAAC		
	HA1	Forward	ATGCCAGAGTTACCAGAAGTTG	972	
		Reverse	ATTTTATCTCCGTTATTCTT		
	HA2	Forward	aagaataacggagataaaatAAAAACAGCTGATTATTCAG	838	
		Reverse	CGGGCCCCACCAAAGGCTTCGAGTTCT		
	KO_conf	Forward	CCATCAGCTCGCGTCTTATC	w/o KO	3,545
		Reverse	CCAAGATCCCAATCAACATTCC	w/ KO	2,381
pgi	sgRNA	Forward	CGACTAGTGAATCCAATGGCAAATCGACGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	132	
		Reverse	atgatagcagagttcactcgAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAAC		
	HA1	Forward	CGAGTGAACTCTGCTATCATGG	965	
		Reverse	TCTTTGATTAAAAGTGTCAT		
	HA2	Forward	atgacacttttaatcaaagATTTAATTCCTTTCAATTTCT	965	
		Reverse	CGGGCCCCCTTGGGACATTCACTGACA		
	KO_conf	Forward	GGCTCCTCTAATCGTCTCATG	w/o KO	3,663
		Reverse	GTAGTCGCTGCTGATGACAA	w/ KO	2,316
pyk	sgRNA	Forward	CGACTAGTGGACCAGCGATTTCTTCAGCGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	132	
		Reverse	cgactgctaccttggtcaacAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAAC		
	HA1	Forward	GTTGACCAAGGTAGCAGTCG	940	
		Reverse	TTTGTCAATTAAGAATCTTA		
	HA2	Forward	taagattcttaattgacaaaTTTGTGTTTTTCTCCTATAA	824	
		Reverse	CGGGCCCGTGTCGTTGTAATCGGTGGTGAT		
	KO_conf	Forward	CCTTCAGCACCAACTACAGC	w/o KO	3,341
		Reverse	GGACAACTTGCTGGGATTGAG	w/ KO	1,832

Table S 2 Electroporation conditions of L. lactis LMG 19460, antibiotic concentration selection, number of transformation candidates obtained and analysed and KO and pDNA presence assessment methods.

DNA mass (ng)	Pulse conditions (# of pulses × kV/cm)	Apramicin ON recovery C (ug/mL)	Inoculum Volume	Apramicin selection C (ug/mL)	# of transformation candidates	# of transformation candidates analysed	*nth* KO		pDNA nthCRISPRa		# of attempts
							Assessment method	KO	Presence assessment method	Presence	
10	3 × 10	500	100 μL	500	>1000	2	PCR(Nth_conf)	No	PCR(HA1), V	Yes?, No	1
		250	T	1,000	0	-	-	-	-	-	1
100	1 × 10	300	T	500	0	-	-	-	-	-	1
	3 × 10	250	T	1,000	1	1	C_PCR(Nth_conf)	No	C_PCR(Cas9_conf)	No	1
		300	T	500	>1000	3	PCR(Nth_conf)	No	PCR(HA1), V	Yes?, No	1
			T	500	200	3	PCR(Nth_conf)	No	PCR(HA1), V	Yes?, No	1
			T	500	0	-	-	-	-	-	2
		500	T	500	0	-	-	-	-	-	1
			T	1,000	400	2	C_PCR(Nth_conf)	No	C_PCR(Cas9_conf)	No	1
			T	1,000	406	8	C_PCR(Nth_conf)	No	C_PCR(Cas9_conf)	No	1
		No ON	T	1,000	311	16	C_PCR(Nth_conf)	No	C_PCR(Cas9_conf)	No	1
	3 × 20	No ON	T	1,000	23	5	C_PCR(Nth_conf)	No	C_PCR(Cas9_conf)	No	1
		No ON	T	2,000	25	2	C_PCR(Nth_conf)	No	C_PCR(Cas9_conf)	No	1
250	3 × 10	500	100 μL	500	0	-	-	-	-	-	1
500	1 × 10	300	T	500	>1000	0	-	-	-	-	1
	3 × 10	300	T	500	0	-	-	-	-	-	1
		250	T	1000	7	1	C_PCR(Nth_conf)	No	C_PCR(Cas9_conf)	No	1

T. Total volume (Pelleted cells) | PCR. Polymerase Chain Reaction | C_PCR. Colony PCR | Nth_conf. Nth_conf primers | HA1. HA1 primers | Cas9_conf. Cas9_conf primers | V. pDNA visualization after gel electrophoresis.

Table S 3 Gibson Assembly reactions and transformation events conditions and results review.

Gibson Assembly #	Transformation by:		Number of Transformation events	Erythromycin selection (ug/mL)	# of total cfu obtained	# of total cfu analysed	Colony PCR		Liquid medium growth	pDNA presence
	Heat shock	Electroporation					Cas9_conf	Ery_conf		
1	✓		2	500	22	3	-	-	✓(3/3)	× (3/3)
		✓	1	500	3	3	× (3/3)	-	-	-
2		✓	1	500	3	2	× (2/2)	-	-	-
	✓		1	250	742	15	✓(11*/12)	-	✓(6/11*)(3**)	× (6/6)(3**)
	✓		1	350	42	1	-	-	✓	× (1/1)
3		✓	1	250	60	4	✓ (4/4)	-	× (3/3)	-
		✓	1	500	1	1	✓(1/1)	-	× (1/1)	-
	✓		1	250	0	-	-	-	-	-
	✓		1	500	1	1	✓(1/1)	-	× (1/1)	-
4		✓	1	250	253	9	-	✓(9/9)	✓(1*/9)	× (1/1*)
		✓	1	500	0	-	-	-	-	-
5	✓		1	250	18	0	-	-	-	-
	✓		1	500	7	1	-	-	✓(1*/7)	× (1/1*)
6	✓		1	250	3	0	-	-	-	-
	✓		1	500	16	0	-	-	× (16/16)	-