



**UNIVERSIDADE DE LISBOA  
INSTITUTO SUPERIOR TÉCNICO**

**Lactic acid bacteria as cell factories:  
A synthetic biology approach for plasmid DNA and recombinant protein production**

**Sofia de Oliveira Dias Duarte**

**Supervisor: Doctor Gabriel António Amaro Monteiro**

**Co-Supervisor: Doctor Lígia Raquel Marona Rodrigues**

**Thesis approved in public session to obtain the PhD Degree in Biotechnology and Biosciences**

**Jury final classification: Pass with Distinction**



**UNIVERSIDADE DE LISBOA  
INSTITUTO SUPERIOR TÉCNICO**

**Lactic acid bacteria as cell factories:  
A synthetic biology approach for plasmid DNA and recombinant protein production**

**Sofia de Oliveira Dias Duarte**

**Supervisor: Doctor Gabriel António Amaro Monteiro  
Co-Supervisor: Doctor Lúcia Raquel Marona Rodrigues**

**Thesis approved in public session to obtain the PhD Degree in Biotechnology and Biosciences  
Jury final classification: Pass with Distinction**

**Jury**

**Chairperson: Doctor Duarte Miguel de França Teixeira dos Prazeres, Instituto Superior Técnico da Universidade de Lisboa**

**Members of the Committee:**

**Doctor Gabriel António Amaro Monteiro, Instituto Superior Técnico da Universidade de Lisboa**

**Doctor Lúcia Maria de Oliveira Martins, Instituto de Tecnologia Química e Biológica António Xavier da Universidade Nova de Lisboa**

**Doctor Leonilde de Fátima Morais Moreira, Instituto Superior Técnico da Universidade de Lisboa**

**Doctor Joaquina Teresa Gaudêncio Dias, Escola Superior Agrária de Bragança do Instituto Politécnico de Bragança**

**Funding Institution**

**Fundação para a Ciência e Tecnologia**

**2018**

**Título:** Bactérias ácido lácticas como fábricas celulares: Uma abordagem de biologia sintética para produção de DNA plasmídico e proteínas recombinantes

**Nome:** Sofia de Oliveira Dias Duarte

**Doutoramento em:** Biotecnologia e Biociências

**Orientador:** Gabriel António Amaro Monteiro

**Co-orientador:** Lígia Raquel Marona Rodrigues

## **Resumo**

O objetivo da presente tese é desenvolver uma plataforma de Bactérias Ácido Lácticas (LAB)/plasmídeos artificiais altamente eficiente para a produção de DNA plasmídico (pDNA) de grau farmacêutico e proteínas recombinantes para aplicações terapêuticas. Vacinas de DNA (ou plasmídeos) são ferramentas promissoras para a prevenção de doenças infecciosas, assim como para o tratamento de doenças adquiridas, tal como o cancro. As atuais plataformas *Escherichia coli*/plasmídeo levantam preocupações relativamente à sua segurança, devido à possível co-purificação de lipopolissacarídeos, o que aumentaria os custos de purificação, assim como o risco de disseminação de genes de resistência. Várias espécies de LAB têm um estatuto de grau alimentar e são geralmente reconhecidas como seguras (GRAS), tornando-as a escolha ideal quando o objetivo é produzir vacinas de DNA e proteínas recombinantes que poderão ser utilizadas como moléculas terapêuticas para vacinação mucosal usando LAB geneticamente modificadas.

A estirpe hospedeira de LAB escolhida para esta tese foi a *Lactococcus lactis* subsp. *lactis* LMG19460, pois os *L. lactis* têm uma elevada relevância industrial e são a estirpe modelo de LAB para uso laboratorial, estando disponíveis várias ferramentas de Biologia Molecular. Além disso, ao contrário doutras bactérias Gram-positivas, apenas secreta para o meio extracelular uma proteína principal (Usp45), simplificando os processos de purificação *downstream*. Esta espécie é também capaz de sobreviver pelo menos 2-3 dias no trato gastrointestinal humano, sem evocar respostas imunitárias fortes no hospedeiro, tornando-a ideal para ser usada como vetor vivo para vacinação mucosal.

Esta estirpe específica é de extrema relevância quando o objetivo é a clonagem de genes ou a transferência de DNA, devido ao facto de ser livre de plasmídeo, não existindo plasmídeos endógenos e/ou contaminantes patogénicos a serem copurificados. O facto de ser livre de plasmídeo também contribui para reduzir a carga metabólica associada à replicação do plasmídeo, permitindo que a célula redirecione os seus esforços para o plasmídeo de interesse, aumentando o rendimento de produção de plasmídeo e proteínas recombinantes associadas. Também é importante referir que o genoma desta estirpe foi sequenciado e publicado no contexto da presente tese.

O plasmídeo principal usado nesta tese foi o pTRKH3, um plasmídeo *shuttle* capaz de replicar em hospedeiros Gram-negativos e Gram-positivos, devido à presença das origens de replicação p15A e pAM $\beta$ 1, respetivamente, juntamente com os genes de resistência à tetraciclina e eritromicina.

Os objetivos principais da presente tese consistem em responder a três grandes desafios: 1) o baixo número de cópias dos plasmídeos disponíveis para LAB; 2) a presença dum elevado nível de atividade de endonucleases na maior parte

das estirpes; e 3) a implementação duma plataforma altamente eficiente de LAB/plasmídeo de elevado número de cópias.

O primeiro desafio tem dois objetivos específicos, que corresponde ao nível de engenharia do plasmídeo: 1.1) aumentar o número de cópias de plasmídeo (PCN) fazendo mutagénesis dirigida à sequência Shine-Dalgarno da origem de replicação; e 1.2) usar *Gibson Assembly* para construir um plasmídeo mínimo com partes melhoradas, para aumentar o PCN. Um plasmídeo com um elevado número de cópias é um dos principais requisitos para se alcançar uma plataforma industrialmente rentável para a produção de pDNA para vacinas de DNA ou proteínas heterólogas para serem utilizadas como antígenos em vacinação mucosal. Até agora, o maior PCN reportado em bactérias Gram-positivas para a replicação pAM $\beta$ 1 foi de cerca de 100 cópias por célula de *Bacillus subtilis*. No **capítulo 2**, o local de ligação ao ribossoma (RBS) do operão *repDE* do plasmídeo pTRKH3 foi alvo de mutagénesis dirigida, dando origem ao mutante de elevado PCN pTRKH3-b, que era capaz de produzir 215 cópias de plasmídeo por cromossoma após 10.5 horas de crescimento, correspondendo a um aumento de 3.5 vezes relativamente ao pTRKH3 não modificado (62 cópias por cromossoma). No **capítulo 3**, foi utilizada uma abordagem inovadora e versátil para construir um plasmídeo mínimo usando a tecnologia de *BioBricks*, com o objetivo final de obter um replicão seguro e de elevado número de cópias e, conseqüentemente, um elevado rendimento de proteínas recombinantes. Cada *BioBrick* foi introduzido junto a outro ou trocado usando a técnica de *Gibson Assembly*. Para além duma cassete de replicação em *E. coli*, as principais partes a introduzir são uma cassete de replicação em LAB que permita um elevado PCN e uma cassete de expressão de proteínas em eucariotas, para a produção de vacinas de DNA. Além disso, está prevista também a introdução duma cassete de expressão de proteínas recombinantes em LAB para vacinação mucosal. A cassete de expressão de proteínas em LAB será testada com uma via sintética de produção de curcumina, atualmente em estudo na presente tese.

O segundo desafio corresponde ao nível da engenharia de estirpe e também tem dois objetivos específicos: 2.1) proceder a uma análise genómica usando o software *Optiflux* para prever os melhores genes a deletar (*araT*, *nucA*, *yjH*, *gltB*, *nrdF*, *adhE*, *argH*, *argg* and *purH*), para que haja um aumento da biomassa e da produção de pDNA; e 2.2) minimizar a digestão não-específica e aumentar o rendimento e a qualidade do pDNA, usando um sistema de CRISPR/Cas9 para deletar o gene da endonuclease (*nth*). No **capítulo 4** optou-se por uma estratégia *one-step* de CRISPR/Cas9 para realizar a deleção, sendo que é a primeira tentativa alguma vez descrita em LAB. Até agora, já fomos capazes de desenhar e construir o plasmídeo baseado no pKCCas9dO, um sistema *one-step* originalmente desenhado para *Streptomyces*, para a deleção do gene da endonuclease de *L. lactis* LMG19460. De forma a aumentar a eficiência da plataforma, o gene de resistência à apramicina foi substituído pelo gene de resistência à eritromicina, e testou-se a indução do promotor *tipA* com o antibiótico thiostrepton. Até ao momento ainda não foi possível alcançar a prova-de-conceito de que é possível realizar uma deleção em *L. lactis* LMG19460 usando o método do CRISPR/Cas9.

O terceiro desafio insere-se no nível das fábricas celulares e tem como objetivo principal implementar uma plataforma altamente eficiente de LAB/plasmídeo de elevado número de cópias, usando uma via sintética de produção de curcumina. No **capítulo 5** é efetuada uma revisão da literatura sobre a importância da curcumina como um agente anti-cancerígeno seguro. Como a curcumina tem um *uptake* celular lento, uma baixa biodisponibilidade sistémica e doses orais elevadas da molécula são necessárias para alcançar uma concentração terapêutica no interior das células, a

possibilidade das LAB sintetizarem a curcumina *in situ*, no trato gastrointestinal humano, representa uma solução poderosa e inovadora para a terapia dos cancros gastrointestinais. Os passos de clonagem molecular para construir os clones necessários com os 3 genes da via da curcumina em operação (*4CL*, *DCS* e *CURSI*) já foram completados com sucesso. A estirpe *L. lactis* LMG19460 já foi transformada com todos os plasmídeos necessários, mas os clones não foram ainda capazes de produzir curcumina.

**Palavras-chave:** Bactérias ácido lácticas, DNA plasmídico, Proteínas recombinantes, Biologia sintética, Engenharia de estirpes

**Title:** Lactic acid bacteria as cell factories: A synthetic biology approach for plasmid DNA and recombinant protein production

## **Abstract**

This thesis aims to develop a highly efficient Lactic Acid Bacteria (LAB)/artificial plasmid platform for the production of pharmaceutical-grade plasmid DNA and recombinant proteins with therapeutic applications. DNA (or plasmid) vaccines are promising tools for the prevention of infectious diseases as well as treatment of acquired diseases such as cancers. The current *Escherichia coli*/plasmid-based platforms raise safety concerns such as co-purification of lipopolysaccharides, which increases purification costs, and the risk of spreading resistance genes. Several LAB species have a food-grade and Generally Recognized As Safe (GRAS) status, which make them the ideal choice when the goal is to produce DNA vaccines and recombinant proteins that can be used as therapeutic molecules in mucosal vaccination with engineered LAB.

The chosen GRAS LAB host strain used in the present thesis was *Lactococcus lactis* subsp. *lactis* LMG19460, since *L. lactis* has a high industrial relevance and is the model LAB laboratory strain, with several molecular tools already available. Also, unlike other Gram-positive bacteria, it only secretes one major protein (Usp45) into the extracellular medium, simplifying the downstream purification processes. They are also able to survive through the gastrointestinal tract of humans for at least 2-3 days without evoking strong host immune responses, making them ideal to be used as live vectors for mucosal vaccination.

This specific strain is of utmost relevance for gene cloning and DNA transfer purposes, since it is a plasmid-free strain, having no copurifying endogenous plasmids and/or pathogenic contaminants. The plasmid-free status also contributes to decrease the metabolic burden associated with plasmid replication, allowing the cell to redirect its efforts towards the plasmid of interest, increasing the plasmid and also the associated recombinant proteins yields. It is also important to refer that the genome of this strain was sequenced and published in the context of the present thesis.

The backbone plasmid used in this thesis was pTRKH3, a shuttle vector able to replicate both in Gram-negative and Gram-positive hosts, due to the presence of a p15A and pAM $\beta$ 1 origin of replication, respectively, together with tetracycline and erythromycin resistance genes.

The major goals of the present PhD thesis are to answer to three major challenges: 1) the low copy number of the available LAB plasmids; 2) the presence of a high level of endonuclease activity in the majority of the strains; and 3) the implementation of a highly efficient LAB/high copy number plasmid platform.

The first challenge has two main specific goals, which correspond to the plasmid engineering level: 1.1) increase the plasmid copy number (PCN) by site-directed mutagenesis of the Shine-Dalgarno sequence of the origin of replication; and 1.2) use Gibson Assembly to engineer a minimal plasmid with improved parts, to increase the PCN. A plasmid with a high PCN is one of the major requirements to achieve an industrially profitable platform for the production of pDNA for DNA vaccines or heterologous proteins to be used as antigens in mucosal vaccination. Until now, the highest

number of copies already reported in gram-positive bacteria for the highly stable pAM $\beta$ 1 replicon was around 100 copies per *Bacillus subtilis* cell. On **chapter 2**, the *repDE* Ribosome Binding Site (RBS) of the pTRKH3 plasmid was engineered by site-directed mutagenesis, generating the high PCN pTRKH3-b mutant, which was able to achieve 215 copies of plasmid per chromosome after 10.5 hours of growth, which corresponds to a 3.5 fold increase when compared to the non-modified pTRKH3 (62 copies per chromosome). On **chapter 3**, it was used an innovative and versatile approach to assemble a minimal plasmid, using BioBrick technology, with the final purpose of obtaining a safe and high copy replicon and, consequently, high yield of recombinant protein. Each BioBrick can be easily assembled together or exchanged using the Gibson Assembly technique, in order to test different alternatives. Besides an *E. coli* replication cassette, the main parts to assemble are a LAB replicative cassette that allow a high PCN and also an eukaryotic protein expression cassette, for DNA vaccines production. In addition, a LAB recombinant protein expression cassette for mucosal vaccination is encompassed. The LAB protein expression cassette will be tested with a curcumin synthetic pathway currently under study in the present thesis.

The second challenge corresponds to the strain engineering level and also has two main specific goals: 2.1) perform genomic analysis using Optflux software to predict the best candidate genes to knockout (*araT*, *nucA*, *yjhF*, *gltB*, *nrdF*, *adhE*, *argh*, *argg* and *purH*), with the goal to increase biomass and pDNA production; and 2.2) minimize non-specific digestion and improve yield and quality of pDNA by using a CRISPR/Cas9 system to knockout the endonuclease gene (*nth*). On **chapter 4** it was described the first attempt of performing a gene knockout using a one-step CRISPR/Cas9 approach in LAB. Until now we were able to design and construct the appropriate plasmid based on pKCcas9dO, a one-step system originally designed for *Streptomyces*, for the endonuclease gene knockout in *L. lactis* LMG19460. In order to increase the platform efficiency, the apramycin resistance gene was replaced for the erythromycin one, and *tipA* promoter induction with thioestrepton was tested. Until now the proof-of-concept that it is possible to perform a knockout in *L. lactis* LMG19460 using this method is yet to be achieved.

The third challenge is in the cell factory level and has as main goal to implement a highly efficient LAB/high copy number plasmid platform using a curcumin synthetic pathway. On **chapter 5**, it was reviewed the importance of curcumin as a safe anti-carcinogenesis agent. Since curcumin has a slow cellular uptake, low systemic bioavailability and repetitive oral doses of the molecule are required to achieve an effective therapeutic concentration inside the cells, the possibility of synthesizing it *in situ*, i.e. in the human gastrointestinal tract, by LAB represents a powerful and innovative solution for gastrointestinal cancer therapy. The molecular cloning steps to construct the necessary clones with the 3 genes of the curcumin pathway (*4CL*, *DCS* and *CURS1*) in operon were successfully achieved. The *L. lactis* LMG19460 strain was also successfully transformed with all the necessary plasmids, however the *L. lactis* clones failed to produce curcumin.

**Key-words:** Lactic Acid Bacteria, Plasmid DNA, Recombinant proteins, Synthetic biology, Strain engineering

**Título:** Bactérias ácido lácticas como fábricas celulares: Uma abordagem de biologia sintética para produção de DNA plasmídico e proteínas recombinantes

### **Resumo alargado em Português**

O objetivo da presente tese é desenvolver uma plataforma de Bactérias Ácido Lácticas (LAB)/plasmídeos artificiais altamente eficiente para a produção de DNA plasmídico (pDNA) de grau farmacêutico e proteínas recombinantes para aplicações terapêuticas. Vacinas de DNA (ou plasmídeos) são ferramentas promissoras para a prevenção de doenças infecciosas, assim como para o tratamento de doenças adquiridas, tal como o cancro. As atuais plataformas *Escherichia coli*/plasmídeo levantam preocupações relativamente à sua segurança, devido à possível co-purificação de lipopolissacarídeos, o que aumentaria os custos de purificação, assim como o risco de disseminação de genes de resistência. Várias espécies de LAB têm um estatuto de grau alimentar e são geralmente reconhecidas como seguras (GRAS), tornando-as a escolha ideal quando o objetivo é produzir vacinas de DNA e proteínas recombinantes que poderão ser utilizadas como moléculas terapêuticas para vacinação mucosal usando LAB geneticamente modificadas.

A estirpe hospedeira de LAB escolhida para esta tese foi a *Lactococcus lactis* subsp. *lactis* LMG19460, pois os *L. lactis* têm uma elevada relevância industrial e são a estirpe modelo de LAB para uso laboratorial, estando disponíveis várias ferramentas de Biologia Molecular. Além disso, ao contrário doutras bactérias Gram-positivas, apenas secreta para o meio extracelular uma proteína principal (Usp45), simplificando os processos de purificação *downstream*. Esta espécie é também capaz de sobreviver pelo menos 2-3 dias no trato gastrointestinal humano, sem evocar respostas imunitárias fortes no hospedeiro, tornando-a ideal para ser usada como vetor vivo para vacinação mucosal. No **capítulo 1** é apresentada uma extensa revisão de literatura acerca dos *L. lactis* serem utilizados como vetores de distribuição de antígenos, onde diferentes replicões foram usados com sucesso em vacinação mucosal.

Esta estirpe específica é de extrema relevância quando o objetivo é a clonagem de genes ou a transferência de DNA, devido ao facto de ser livre de plasmídeo, não existindo plasmídeos endógenos e/ou contaminantes patogénicos a serem copurificados. O facto de ser livre de plasmídeo também contribui para reduzir a carga metabólica associada à replicação do plasmídeo, permitindo que a célula redirecione os seus esforços para o plasmídeo de interesse, aumentando o rendimento de produção de plasmídeo e proteínas recombinantes associadas. É também uma estirpe interessante para produzir pDNA de grau alimentar e farmacêutico para utilizar em vacinas de DNA, visto que não possui os sistemas de restrição-modificação presentes originalmente nos plasmídeos endógenos, permitindo a produção de pDNA em maior quantidade e com melhor qualidade. Também é importante referir que o genoma desta estirpe foi sequenciado e publicado no contexto da presente tese.

O plasmídeo principal usado nesta tese foi o pTRKH3, um plasmídeo *shuttle* capaz de replicar em hospedeiros Gram-negativos e Gram-positivos, devido à presença das origens de replicação p15A e pAM $\beta$ 1, respetivamente, juntamente com os genes de resistência à tetraciclina e eritromicina. O replicão pAM $\beta$ 1 é constituído pelo operão *repDE*, cuja expressão está sob o controlo do promotor P<sub>DE</sub>, sendo que o gene *repE* codifica para a proteína limitante para o início

da replicação, essencial para o mecanismo de replicação unidirecional do tipo *theta*, enquanto a proteína RepD tem uma função desconhecida, não sendo essencial para a replicação do plasmídeo. A sua transcrição é estreitamente regulada por dois sistemas diferentes que em última instância controlam o número de cópias do plasmídeo: o repressor da transcrição CopF e um sistema de atenuação da transcrição mediado por RNA antisense (CT-RNA).

Os objetivos principais da presente tese consistem em responder a três grandes desafios: 1) o baixo número de cópias dos plasmídeos disponíveis para LAB; 2) a presença dum elevado nível de atividade de endonucleases na maior parte das estirpes; e 3) a implementação duma plataforma altamente eficiente de LAB/plasmídeo de elevado número de cópias.

O primeiro desafio tem dois objetivos específicos, que corresponde ao nível de engenharia do plasmídeo: 1.1) aumentar o número de cópias de plasmídeo (PCN) fazendo mutagénesis dirigida à sequência Shine-Dalgarno da origem de replicação; e 1.2) usar *Gibson Assembly* para construir um plasmídeo mínimo com partes melhoradas, para aumentar o PCN. Um plasmídeo com um elevado número de cópias é um dos principais requisitos para se alcançar uma plataforma industrialmente rentável para a produção de pDNA para vacinas de DNA ou proteínas heterólogas para serem utilizadas como antígenos em vacinação mucosal. Até agora, o maior PCN reportado em bactérias Gram-positivas para o replicão pAM $\beta$ 1 foi de cerca de 100 cópias por célula de *Bacillus subtilis*. No **capítulo 2**, o local de ligação ao ribossoma (RBS) do operão *repDE* do plasmídeo pTRKH3 foi alvo de mutagénesis dirigida, dando origem ao mutante de elevado PCN pTRKH3-b, que era capaz de produzir 215 cópias de plasmídeo por cromossoma após 10.5 horas de crescimento, correspondendo a um aumento de 3.5 vezes relativamente ao pTRKH3 não modificado (62 cópias por cromossoma). Uma análise de componentes principais mostrou que este mutante tem uma quantidade intermédia dos repressores da transcrição (CopF e CT-RNA), que juntamente com uma forte sequência do RBS que se encontra duplicado na sequência deste mutante e uma estrutura secundária do mRNA capaz de promover a ligação do ribossoma, contribui para um PCN tão elevado. No **capítulo 3**, foi utilizada uma abordagem inovadora e versátil para construir um plasmídeo mínimo usando a tecnologia de *BioBricks*, com o objetivo final de obter um replicão seguro e de elevado número de cópias e, consequentemente, um elevado rendimento de proteínas recombinantes. Cada *BioBrick* foi introduzido junto a outro ou trocado usando a técnica de *Gibson Assembly*. Para além duma cassette de replicação em *E. coli*, as principais partes a introduzir são uma cassette de replicação em LAB que permita um elevado PCN e uma cassette de expressão de proteínas em eucariotas, para a produção de vacinas de DNA. Além disso, está prevista também a introdução duma cassette de expressão de proteínas recombinantes em LAB para vacinação mucosal. A cassette de expressão de proteínas em LAB será testada com uma via sintética de produção de curcumina, atualmente em estudo na presente tese.

O segundo desafio corresponde ao nível da engenharia de estirpe e também tem dois objetivos específicos: 2.1) proceder a uma análise genómica usando o software *Optflux* para prever os melhores genes a deletar (*araT*, *nucA*, *yjH*, *gltB*, *nrdF*, *adhE*, *argh*, *argg* and *purH*), para que haja um aumento da biomassa e da produção de pDNA; e 2.2) minimizar a digestão não-específica e aumentar o rendimento e a qualidade do pDNA, usando um sistema de CRISPR/Cas9 para deletar o gene da endonuclease (*nth*). No **capítulo 4**, após várias tentativas preliminares de usar a mutagénesis aleatória e uma estratégia de *recombineering*, sem sucesso, optou-se por uma estratégia *one-step* de

CRISPR/Cas9 para realizar a deleção, sendo que é a primeira tentativa alguma vez descrita em LAB. Até agora, já fomos capazes de desenhar e construir o plasmídeo baseado no pKCcas9dO, um sistema *one-step* originalmente desenhado para *Streptomyces*, para a deleção do gene da endonuclease de *L. lactis* LMG19460. De forma a aumentar a eficiência da plataforma, o gene de resistência à apramicina foi substituído pelo gene de resistência à eritromicina, e testou-se a indução do promotor *tipA* usando o antibiótico thiostrepton. Até ao momento ainda não foi possível alcançar a prova-de-conceito de que é possível realizar uma deleção em *L. lactis* LMG19460 usando o método do CRISPR/Cas9.

O terceiro desafio insere-se no nível das fábricas celulares e tem como objetivo principal implementar uma plataforma altamente eficiente de LAB/plasmídeo de elevado número de cópias, usando uma via sintética de produção de curcumina. No **capítulo 5** é efetuada uma revisão da literatura sobre a importância da curcumina como um agente anticancerígeno seguro. Vários ensaios clínicos já terminados sugerem que a curcumina é útil na prevenção do cancro do cólon em humanos, sendo segura para as células saudáveis e originando efeitos secundários mínimos, sendo uma boa alternativa aos agentes quimioterapêuticos atuais. Como a curcumina tem um *uptake* celular lento, uma baixa biodisponibilidade sistémica e doses orais elevadas da molécula são necessárias para alcançar uma concentração terapêutica no interior das células, a possibilidade das LAB sintetizarem a curcumina *in situ*, no trato gastrointestinal humano, representa uma solução poderosa e inovadora para a terapia dos cancros gastrointestinais. Os passos de clonagem molecular para construir os clones necessários com os 3 genes da via da curcumina em operação (*4CL*, *DCS* e *CURSI*) já foram completados com sucesso. A estirpe *L. lactis* LMG19460 já foi transformada com todos os plasmídeos necessários, mas os clones não foram ainda capazes de produzir curcumina. Como sugestão de trabalho futuro, assim que os *L. lactis* LMG19460 forem capazes de produzir curcumina, o objectivo é obter a prova-de-conceito de que as LAB com a via sintética da curcumina são citotóxicas para as células cancerígenas.

**Palavras-chave:** Bactérias ácido lácticas, DNA plasmídico, Proteínas recombinantes, Biologia sintética, Engenharia de estirpes

## ACKNOWLEDGMENTS

First I would like to thank my supervisor Professor Gabriel Monteiro for all the motivation since the day I said I want to apply for a PhD scholarship and until the very final of the four years. Always had good advices and out-of-the-box ideas, but the most important always believed in me, even when I didn't, motivating me for always making even better, work even harder and pushing even more for myself.

To my co-supervisor Professor Lúcia Rodrigues, a special acknowledgment for all the support and constant interest in my work, and for always having kind words of motivation and good advices.

I also would like to thank Professor Miguel Prazeres that although not being my supervisor, always gave important advices that contributed immensely to my work.

The master students which work I help to supervise deserve also an acknowledgment, because every single one of them were not only my students, but turned out amazing friends. Sílvia Andrade, Maria Martins, Carolina Gonçalves and Inês Gonçalves, all of them contributed to the success of my work, not only with its own results, but with their constant friendship and support. I couldn't ask for a better "Lactic Acid Bacteria team"!

Other not less important acknowledgment goes to my "Lunch team", which every day contributed to good moments of relaxation and fun, but also important scientific brainstorming, helping to turn the hard work days more bearable. Marisa Santos, Teresa Esteves, Margarida Silva, Nuno Faria and Diana Cipriano, your friendship is one of the best results of my PhD thesis.

Cláudia Alves deserves a particular acknowledgment, since she was my "right arm" in the laboratory, my soft place to fall, my loyal friend and my biggest inspiration, since is one of the hardest working persons I've ever met.

To all the remaining current MoBiol team (Ana Rita Santos, Ana Rosa, João Filipe, Diogo Faria, Rui Silva, Marisa Faustino and Ana Rita Lourenço) I would like to acknowledge all the good vibes and happy work environment they provided me, which was an unevaluable help for accomplishing my thesis.

I also want to highlight some former MoBiol members that had a big positive impact in my PhD thesis, such as Jorge Paulo, Elisabete Freitas and Geisa Gonçalves, but especially Salomé Magalhães who was a big inspiration, a great friend and always believed in me.

I would also like to thank all the iBB members, with a special focus on Flávio Ferreira for its help with the HPLC.

Ricardo Pereira need to have a special place in this acknowledgment section, since he always helped me every time I needed and for being such a good friend. Dona Rosa was of inestimable help every single day and our conversations are truly unforgettable.

The last, but not the least, to my family (parents and grandparents) goes a big acknowledgment, for putting up with me and my PhD stress, and for giving me the unconditional love that I needed to pursue my dreams. Rui Cunha, my life partner and my love, was probably the person that needs the biggest thank for bearing with me every single day. Even

in the days where everything went wrong in the laboratory, I know that at home I had my soft place to fall, the person that gave me strength to accomplish my goals and believe in myself.

And I let to the end the most important, my son Tiago, my reason to live, my universe, my everything. I carried him for 9 months while working in the laboratory, turning my PhD thesis even more special. One single smile from him gave me strengths that I don't know I had. Without even knowing, he contributed immensely for the success of my thesis and that's why I want to dedicate it to him, my baby boy.

# TABLE OF CONTENTS

<b>CHAPTER 1 - General introduction and literature review</b> .....	1
1.1. Lactic Acid Bacteria .....	2
1.2. <i>L. lactis</i> as cell factories for plasmid DNA and recombinant protein production .....	4
1.2.1. Gene expression systems for expression of heterologous proteins.....	5
1.2.2. DNA vaccines .....	6
1.2.3. Live mucosal vaccination using <i>L. lactis</i> - Mucosal delivery of pDNA, antigens and cytokines 10	
1.2.3.1.Plasmids with a pWV01-based origin of replication .....	12
1.2.3.2.Plasmids with a pSH71-based origin of replication .....	14
1.2.3.3.Plasmids with a pAM $\beta$ 1-based origin of replication .....	17
1.3. Research aims .....	19
1.4. Thesis outline .....	19
<b>CHAPTER 2 - Increasing plasmid copy number of pTRKH3 in <i>Lactococcus lactis</i> by the modification of the <i>repDE</i> ribosome binding site</b> .....	21
2.1. Abstract .....	22
2.2. Introduction.....	23
2.3. Material and methods .....	25
2.3.1. Bacterial strains and plasmids.....	25
2.3.2. Media and growth conditions .....	26
2.3.3. pTRKH3 modification by site-directed mutagenesis .....	26
2.3.4. LAB electrotransformation with modified plasmids.....	28
2.3.5. Plasmid copy number determination by qPCR .....	28
2.3.5.1. Primer design.....	28
2.3.5.2. Samples and standards preparation.....	29
2.3.5.3. Real-time qPCR reaction.....	29
2.3.5.4. Plasmid copy number (PCN) quantification.....	30
2.3.5.5. qPCR validation .....	30
2.3.5.6. Statistical analysis .....	31
2.3.6. Relative translation initiation rate prediction.....	31
2.3.7. <i>feoA</i> , <i>repD</i> , <i>copF</i> and <i>CT-RNA</i> mRNA two-step quantification by qPCR.....	31
2.3.7.1. Primer design.....	31

2.3.7.2. RNA isolation and mRNA quantification by real-time qPCR.....	32
2.4. Results and discussion .....	32
2.4.1. Growth profile of <i>L. lactis</i> LMG19460 harboring wild-type and modified plasmids.....	32
2.4.2. Plasmid copy number (PCN) determination by real-time qPCR .....	34
2.4.3. <i>In silico</i> mRNA analysis .....	38
2.4.4. Considerations about the pTRKH3-b mutant.....	40
2.4.5. mRNA analysis.....	44
2.5. Conclusions.....	49
<b>CHAPTER 3 - Artificial plasmids for replication in LAB .....</b>	<b>51</b>
3.1. Abstract.....	52
3.2. Introduction.....	53
3.3. Materials and methods .....	61
3.3.1. Bacterial strains and plasmids.....	61
3.3.2. Media and growth conditions .....	62
3.3.3. Competent cells and transformation .....	63
3.3.3.1. <i>E. coli</i> DH5 $\alpha$ , GM2163 and SCS110.....	63
3.3.3.2. <i>L. lactis</i> LMG19460 .....	63
3.3.4. Biobricks <i>in silico</i> design.....	64
3.3.5. Biobricks amplification and assembly .....	66
3.3.6. Plasmid copy number determination by qPCR .....	68
3.3.6.1. <i>L. lactis</i> LMG19460 growth conditions .....	68
3.3.6.2. Primer design.....	69
3.3.6.3. Samples and standards preparation.....	69
3.3.6.4. Real-time qPCR reaction.....	69
3.3.6.5. Plasmid copy number (PCN) quantification.....	70
3.3.6.6. qPCR validation .....	70
3.3.6.7. Statistical analysis .....	71
3.4. Results and discussion .....	71
3.4.1. <i>In silico</i> design and plasmid assembly.....	71
3.4.1.1. <i>E. coli</i> and Gram-positive cassettes assembly .....	73
3.4.1.2. Assembly of a prokaryotic protein expression cassette .....	76
3.5. Conclusions.....	80

<b>CHAPTER 4 - <i>L. lactis</i> LMG19460 strain engineering</b> .....	83
4.1. Abstract.....	84
4.2. Introduction.....	85
4.2.1. ssDNA and dsDNA recombineering in LAB.....	85
4.2.2. CRISPR/Cas9 system.....	86
4.2.3. CRISPR/Cas9 applications in LAB.....	89
4.2.4. CRISPR/Cas9 system used in the present work.....	92
4.3. Materials and methods.....	94
4.3.1. Bacterial strains and plasmids.....	94
4.3.2. Media and growth conditions.....	95
4.3.3. Competent cells and transformation.....	95
4.3.3.1. <i>E. coli</i> DH5 $\alpha$ , GM2163 and SCS110.....	95
4.3.3.2. <i>L. lactis</i> LMG19460.....	96
4.3.4. <i>L. lactis</i> LMG19460 genome characterization.....	96
4.3.5. <i>In silico</i> genomic analysis using Optflux software.....	97
4.3.6. Gene knockout using random mutagenesis.....	97
4.3.7. Gene knockout using Datsenko and Wanner <sup>121</sup> strategy.....	98
4.3.8. Cloning strategy to construct the gene-specific CRISPR/Cas9 plasmid (pKCcas9dO_nth_apra)100	
4.3.8.1. <i>In silico</i> design.....	100
4.3.8.2. Fragments amplification, SOEing PCRs and molecular cloning of the gene-specific sgRNA and homologous arms in pKCcas9dO plasmid.....	102
4.3.9. Gene knockout using CRISPR/Cas9 strategy.....	104
4.3.10. Induction of the ScoCas9 expression using thiostrepton.....	105
4.3.11. Replacement of the apramycin resistance gene for the erythromycin resistance gene.....	106
4.3.12. Insertion of the <i>ScoCas9</i> gene, sgRNA and homologous arms into the pTRKH3 vector.....	109
4.4. Results and discussion.....	111
4.4.1. <i>L. lactis</i> LMG19460 genome characterization.....	111
4.4.2. Improvement of plasmid DNA and biomass production by <i>in silico</i> metabolic engineering....	115
4.4.3. Gene knockout using random mutagenesis.....	119
4.4.4. Gene knockout using Datsenko and Wanner <sup>121</sup> strategy.....	120
4.4.5. Gene knockout using CRISPR/Cas9 strategy.....	124
4.4.5.1. Induction of the ScoCas9 expression using thiostrepton.....	129
4.4.5.2. Replacement of the apramycin resistance gene for the erythromycin resistance gene.....	132
4.4.5.3. Insertion of the <i>ScoCas9</i> gene, sgRNA and homologous arms into the pTRKH3 vector.....	134
4.5. Conclusions.....	136

<b>CHAPTER 5 - Heterologous production of curcumin by probiotic lactic acid bacteria towards gastrointestinal cancer therapy</b> .....	139
5.1. Abstract .....	140
5.2. Introduction .....	141
5.2.1. Curcumin – a putative solution .....	141
5.2.1.1. Biological activities .....	141
5.2.1.2. Anti-tumor properties – mechanisms of action .....	142
5.2.1.3. Chemical properties of curcumin .....	145
5.2.1.4. Chemical stability and degradation .....	145
5.2.1.5. Uptake and metabolism .....	146
5.2.1.6. The bioavailability problem and possible solutions .....	148
5.2.2. Biosynthetic pathways to produce curcumin .....	149
5.2.2.1. Natural pathway .....	149
5.2.2.2. Artificial pathways .....	150
5.2.2.3. The artificial pathway proposed in the present PhD .....	152
5.2.3. LAB as cell factories to produce heterologous proteins .....	153
5.2.3.1. Metabolite secretion .....	153
5.3. Material and methods .....	154
5.3.1. Bacterial strains and plasmids .....	154
5.3.2. Media and growth conditions .....	155
5.3.3. Competent cells and transformation .....	156
5.3.3.1. <i>E. coli</i> DH5 $\alpha$ .....	156
5.3.3.2. <i>L. lactis</i> LMG19460 .....	156
5.3.4. Cloning strategy of the curcumin pathway genes .....	157
5.3.4.1. Curcumin pathway <i>in silico</i> design and codon optimization .....	157
5.3.4.2. Molecular cloning of the curcumin pathway genes in pTRKH3 under the control of a constitutive promoter .....	158
5.3.5. Curcumin production .....	159
5.3.6. Curcumin extraction .....	160
5.3.7. HPLC analysis .....	160
5.3.7.1. Standards preparation .....	161
5.3.7.2. Optimization of the analysis parameters .....	161
5.3.8. <i>CURS1</i> mRNA quantification by qPCR .....	161
5.4. Results and discussion .....	162

5.4.1.	Cloning of the curcumin pathway genes .....	162
5.4.1.1	<i>In silico</i> design and codon optimization .....	162
5.4.1.2.	Molecular cloning of the curcumin pathway .....	168
5.4.2.	LAB transformation .....	180
5.4.3.	Curcumin production and quantification.....	183
5.4.3.1.	Curcumin and ferulic acid calibration curves .....	183
5.4.3.2.	Optimization of the growth conditions .....	186
5.4.3.3.	Induction of curcumin production using different ferulic acid concentrations .....	188
5.4.3.4.	Induction of curcumin production using the optimized growth conditions .....	192
5.4.	Conclusions and future work .....	202
<b>CHAPTER 6 - Concluding remarks and perspectives for further research .....</b>		<b>205</b>
6.1.	General conclusions and future work.....	206
<b>CHAPTER 7 - References .....</b>		<b>2099</b>

# LIST OF TABLES

<b>Table 1.1</b> Plasmids used as antigen-expressing vectors in <i>L. lactis</i> host, for mucosal vaccination purposes against several diseases. ....	11
<b>Table 2.1.</b> Sequence, length and melting temperature of the primers used in site-directed mutagenesis procedures. F and R indicate forward and reverse primers, respectively. ....	27
<b>Table 2.2.</b> Sequence of the primers used for qPCR amplification of the target ( <i>erm</i> , <i>repDE</i> , <i>copF</i> and <i>CT-RNA</i> ) and reference ( <i>feoA</i> ) genes, as well as the accession number of the complete sequence and expected amplicon size. F and R indicate forward and reverse primers, respectively. ....	29
<b>Table 2.3.</b> Translation initiation rate (TIR) and $\Delta G_{\text{total}}$ values of <i>repD</i> and <i>repE</i> genes for pTRKH3 and the six modified plasmids. ....	39
<b>Table 2.4.</b> Start codon sites and translation initiation rate (TIR) of <i>repD</i> gene for pTRKH3-b mutant with and without the detected insertion. The start codons in-frame with the <i>repD</i> start codon are highlighted in green. ....	41
<b>Table 2.5.</b> Average values of PCN and PCN from <i>repDE</i> , <i>copF</i> and <i>CT-RNA</i> mRNAs for pTRKH3 and the six modified plasmids. The PCN results were obtained from real-time qPCR analysis of seven independent growths, while the remaining categories were the result of two independent experiments. The standard error of the mean (SEM) is represented for each point. ....	46
<b>Table 3.1.</b> Main characteristics of the strains and plasmids used in this chapter. ....	62
<b>Table 3.2.</b> Sequence, expected amplicon size and melting temperature of the primers used for <i>nth</i> CRISPR/Cas9 gene knockout strategy. It is also indicated where the primer overlaps (lowercase sequence), as well as the place where it anneals in the plasmid (uppercase sequence). Underlined letters corresponds to introduced restriction enzyme sites. ....	65
<b>Table 3.3.</b> Final sequences of the p15A origin, <i>ery</i> (erythromycin resistance gene), <i>erm</i> (ery promoter) and pAM $\beta$ 1 origin of replication. ....	65
<b>Table 3.4.</b> Mass and molar amount of each DNA fragment necessary for the several Gibson Assembly reactions <sup>111</sup> . ....	68
<b>Table 3.5.</b> Sequence of the primers used for qPCR amplification of the target ( <i>erm</i> ) and reference ( <i>feoA</i> ) genes, as well as the accession number of the complete sequence and expected amplicon size. F and R indicate forward and reverse primers, respectively. ....	69
<b>Table 3.6.</b> Mass and molar amount of each DNA fragment, as well as the vector to insert molar ratio used in the several Gibson Assembly reactions. ....	79
<b>Table 4.1.</b> Genome editing tools currently available for <i>Lactococcus</i> and <i>Lactobacillus</i> species <sup>123</sup> . ....	85
<b>Table 4.2.</b> Main characteristics of the strains and plasmids used in this chapter. ....	94
<b>Table 4.3.</b> Sequence of the <i>nth</i> gene from <i>L. lactis</i> LMG19460. ....	99
<b>Table 4.4.</b> Sequence, expected amplicon size and melting temperature of the primers used for <i>nth</i> gene knockout strategy. ....	100
<b>Table 4.5.</b> Sequences of the <i>nth</i> specific sgRNA and the two homologous arms. ....	102
<b>Table 4.6.</b> Sequence, expected amplicon size and melting temperature of the primers used for <i>nth</i> CRISPR/Cas9 gene knockout strategy. ....	102
<b>Table 4.7.</b> Sequence, expected amplicon size and melting temperature of the primers used for the replacement of the apramycin resistance gene for the erythromycin resistance gene in pKCCas9dO_ <i>nth</i> _ <i>apra</i> plasmid, resulting in the pKCCas9dO_ <i>nth</i> _ <i>ery</i> plasmid. The lowercase nucleotides correspond to the overlapping sequences with the adjacent fragment, necessary for the Gibson Assembly protocol. ....	107

<b>Table 4.8.</b> Sequence, expected amplicon size and melting temperature of the primers used for the cloning of the necessary CRISPR/Cas9 features into the pTRKH3 template. The lowercase nucleotides correspond to the overlapping sequences with the adjacent fragment, necessary for the Gibson Assembly protocol.....	110
<b>Table 4.9.</b> Characterization of the prophages in the <i>L. lactis</i> LMG19460 genome, detected by PHAST. ....	115
<b>Table 4.10.</b> Number of <i>L. lactis</i> LMG19460 counted after UV light exposure, at the 5 <sup>th</sup> dilution, which theoretically should have 1,000 total cells. ....	119
<b>Table 4.11.</b> OD <sub>600nm</sub> , pDNA concentration and pDNA specific yield from the several colonies exposed to different conditions of UV light and then grown for different time periods, starting with OD=0.1.....	119
<b>Table 4.12.</b> Optical density measured at 600 nm of <i>L. lactis</i> LMG19460 strain after 4 days of growth in M17 medium supplemented with different thioestrepton concentrations.....	129
<b>Table 5.1.</b> Main characteristics of the strains and plasmids used in this chapter.....	155
<b>Table 5.2.</b> Sequence, expected amplicon size and melting temperature of the primers used for promoters and genes amplification, as well as for confirmation sequencing procedures. ....	159
<b>Table 5.3.</b> Sequence, expected amplicon size and melting temperature of the primers used for <i>CURS1</i> mRNA quantification by real-time quantitative PCR.....	162
<b>Table 5.4.</b> <i>4CL</i> , <i>DCS</i> and <i>CURS</i> genes sequence synthesized by Nzytech, after codon optimization.....	164
<b>Table 5.5.</b> ODs achieved by <i>L. lactis</i> LMG19460 with pTRKH3 after overnight growth in M17 or modified M9 minimal salt medium (without thiamine, trace elements or CaCO <sub>3</sub> ), at 25°C or 30°C, with different ethanol concentrations. ....	188
<b>Table 5.6.</b> pDNA concentrations and specific yields after purification of 3 mL samples of <i>L. lactis</i> LMG19460 pTRKH3 and <i>L. lactis</i> LMG19460 pTRKH3_SLP_4CL_DCS_CURS1, grown in the presence of 0, 2 or 4 mM ferulic acid. The samples were taken after 24h, 3 days or 6 days of growth. ....	190
<b>Table 5.7.</b> <i>CURS1</i> mRNA quantification by real-time quantitative PCR. The Crossing Point (CP) values, as well as the melting temperature (T <sub>m</sub> ) values are presented).....	192

# LIST OF FIGURES

- Figure 1.1.** Gram-positive bacteria phylogenetic tree. The bar represents 10% sequence divergence. Figure from Wood and Holzapfel (1995)<sup>1</sup>. ..... 2
- Figure 1.2.** Phylogenetic tree of the low GC content subdivision of Gram-positive bacteria, as determined by similarity between ribosomal RNAs. Figure from Wood and Holzapfel (1995)<sup>1</sup>. ..... 3
- Figure 1.3.** Phylogenetic tree representing the relationships between the different species of the genus *Lactococcus*. The bar represents 10% sequence divergence and the tree was built based on 16S rRNA sequence analysis. Figure from Wood and Holzapfel (1995)<sup>1</sup>. ..... 3
- Figure 1.4.** Examples of inducible promoters in *L. lactis*. A) promoter induced by phage infection; B) NICE system, induced by nisin; C) promoter induced by temperature; D) promoter induced by pH shift<sup>5</sup>; E) promoter induced by zinc concentration<sup>7</sup>. Figures from D'Souza et al. (2012)<sup>5</sup> and Morello et al. (2007)<sup>7</sup>. ..... 5
- Figure 1.5.** Overview of the process of plasmid DNA vaccination, focusing on the direct and indirect routes of antigen presentation. Figure from Kemp (2017)<sup>19</sup>. ..... 8
- Figure 1.6.** Essential elements of a plasmid DNA vector. The propagation unit should contain the antibiotic resistance gene and the bacterial origin of replication, for growth and selection in bacteria. The expression unit should include the gene encoding the protein of interest, under the control of a viral promoter, and with transcription termination sequences (Poly A). ..... 9
- Figure 1.7.** Overview of the live mucosal vaccination process. (I-IV) Cloning of the gene coding for the antigen of interest and bacterial transformation. (V) Antigen expression in one of three different presentation forms: cytoplasmic, surface display or secreted to the extracellular medium. (VI) Humoral and cell-mediated immune response. (VII) Protection against the target pathogen. Figure from Lin et al. (2015)<sup>23</sup>. ..... 10
- Figure 1.8.** Main features of the plasmid pWV01. Coding regions are indicated by arrows and the inverted repeats (IR) are indicated by boxes. Relevant restriction sites are indicated. Figure from Seegers et al. (1995)<sup>49</sup>. ..... 12
- Figure 1.9.** Structure of A) pValac<sup>55</sup> plasmid and B) pCYT plasmid<sup>36</sup>. Figures from Guimarães et al. (2009)<sup>55</sup> and Beck et al. (2017)<sup>30</sup>. ..... 14
- Figure 1.10.** Map of A) pNZ8148, B) pNZ8149, C) pNZ8150<sup>58</sup> and D) pNZ8048 plasmids<sup>59</sup>. Figures from MoBiTec (2017)<sup>58</sup> and Frelet-Barrand et al. (2010)<sup>59</sup>. ..... 16
- Figure 1.11.** Map of A) pIL253, B) pOri253, C) pOri23<sup>66</sup>, D) pTRKH1, E) pTRKH2 and F) pTRKH3<sup>65</sup> plasmids. Figures from Que et al. (2000)<sup>66</sup> and O' Sullivan and Klaenhammer (1993)<sup>65</sup>. ..... 18
- Figure 2.1.** pAM $\beta$ 1 replication region of the pTRKH3 plasmid. P<sub>F</sub> – copF promoter; T<sub>F</sub> – copF terminator; P<sub>DE</sub> – repDE promoter; T<sub>DE</sub> – repDE terminator; P<sub>CT</sub> – countertranscript RNA promoter; T<sub>CT</sub> – countertranscript RNA terminator. .... 24
- Figure 2.2.** Sequence of the RBS region of the repDE promoter in the pAM $\beta$ 1 origin of replication, in the wild-type pTRKH3 and in the six modified plasmids. We considered the RBS sequences either starting 6 nucleotides (hypothesis 1) or 3 nucleotides (hypothesis 2) upstream the ATG codon. The nucleotides altered by site-directed mutagenesis are highlighted in grey. All the distances to the repD start codon are also represented. The predicted RBS sequence from pTRKH3-e is shifted upstream 3 nucleotides (6 nucleotides from the start codon). The number of equal nucleotides (with a maximum of 8) between the RBS sequence of each plasmid and the consensus RBS from *L. lactis* is indicated for both consensus RBS hypothesis. .... 27
- Figure 2.3.** Growth profiles of *L. lactis* LMG19460 cells harboring the non-modified pTRKH3 (filled circle with solid line) and the six plasmids with the designed mutations in the RBS sequence of the repDE promoter (filled circle for pTRKH3-a, filled square for pTRKH3-b, filled triangle for pTRKH3-c, empty circle for pTRKH3-d, empty square for pTRKH3-e and empty triangle for pTRKH3-f), during 10.5h of growth in 100 mL shake flasks with 75 mL of M-17 supplemented with 20 g/L glucose, 30°C, 100 rpm (initial OD<sub>600nm</sub>=0.1, starting from overnight inoculum). The profile was obtained from seven independent growths and the standard error of the mean (SEM) is represented for each point.

The specific growth rates ( $\mu$ ) were 0.40 ( $\pm 0.01$ ), 0.40 ( $\pm 0.01$ ), 0.44 ( $\pm 0.02$ ), 0.41 ( $\pm 0.01$ ), 0.41 ( $\pm 0.04$ ), 0.43 ( $\pm 0.01$ ) and 0.37 ( $\pm 0.01$ )  $\text{h}^{-1}$  for the wild-type pTRKH3 and the six mutants (pTRKH3-a to pTRKH3-f, respectively)..... 33

**Figure 2.4.** pH profiles of *L. lactis* LMG19460 cells harboring the non-modified pTRKH3 (filled circle with solid line) and the six plasmids with the designed mutations in the RBS sequence of the repDE promoter (filled circle for pTRKH3-a, filled square for pTRKH3-b, filled triangle for pTRKH3-c, empty circle for pTRKH3-d, empty square for pTRKH3-e and empty triangle for pTRKH3-f), during 10.5h of growth in 100 mL shake flasks with 75 mL of M-17 supplemented with 20 g/L glucose, 30°C, 100 rpm (initial  $\text{OD}_{600\text{nm}}=0.1$ , starting from overnight inoculum). The profile was obtained from seven independent growths and the standard error of the mean (SEM) is represented for each point. .... 34

**Figure 2.5.** pDNA real-time qPCR standard curve performed in quadruplicate, using four log serial dilutions of pTRKH3 pDNA (10, 100, 1,000 and 10,000 pg) spiked with *L. lactis* LMG19460 cells. The linear regression trendline is represented, along with the corresponding equation, coefficient of determination value ( $R^2$ ) and efficiency values. .... 35

**Figure 2.6.** gDNA real-time qPCR standard curve performed in triplicate using five log serial dilutions of *L. lactis* LMG19460 gDNA (1, 10, 100, 1,000 and 10,000 pg). The linear regression trendline is represented, along with the corresponding equation, coefficient of determination value ( $R^2$ ) and efficiency values. .... 35

**Figure 2.7.** Average PCN values for *L. lactis* LMG19460 cells harboring the wild-type and the six modified plasmids after 6h (light grey) and 10.5h of growth (dark grey). The results were obtained from real-time qPCR analysis of seven independent growths in 100 mL shake flasks with 75 mL of M-17 supplemented with 20 g/L glucose, 30°C, 100 rpm (initial  $\text{OD}_{600\text{nm}}=0.1$ , starting from an overnight inoculum). The standard error of the mean (SEM) is represented for each point, as well as the statistically significant differences at  $p < 0.05$  (\*) or  $p < 0.01$  (\*\*) between the non-modified pTRKH3 and the plasmids with the designed mutations in the RBS sequence of the repDE promoter. .... 36

**Figure 2.8.** Average PCN values for *L. lactis* LMG19460 cells harboring the wild-type pTRKH3 (light grey) and the highest plasmid producer pTRKH3-b (dark grey) at different growth times (3, 6, 8 and 10.5h). The PCN results were obtained from real-time qPCR analysis of seven independent growths and the standard error of the mean (SEM) is represented for each point. .... 37

**Figure 2.9.** Prediction of the repDE mRNA secondary structures for the wild-type pTRKH3 and the six modified plasmids. The start codon is highlighted with a box and the first nucleotide (position 19) of the hypothesis 1 consensus RBS is indicated with an arrow. The hypothesis 2 consensus RBS starts three nucleotides downstream the arrow. .. 38

**Figure 2.10.** pTRKH3-b mutant sequence showing the insertion of a duplicated sequence consisting in the end portion of the repDE terminator, the complete RBS and the beginning of the repD gene..... 40

**Figure 2.11.** Prediction of the repDE mRNA secondary structures for the pTRKH3-b with and without the detected insertion. The first start codon in frame with the repD start codon is highlighted with a solid box and the first nucleotide of the hypothesis 1 consensus RBS is indicated with an arrow. The hypothesis 2 consensus RBS starts three nucleotides downstream the arrow. .... 42

**Figure 2.12.** Growth profiles of *L. lactis* LMG19460 cells harboring the non-modified pTRKH3 (filled circle with solid line), the pTRKH3-b mutant without the insertion (filled triangle) and the pTRKH3-b mutant with the insertion (filled square), during 10.5h of growth in 100 mL shake flasks with 75 mL of M-17 supplemented with 20 g/L glucose, 30°C, 100 rpm (initial  $\text{OD}_{600\text{nm}}=0.1$ , starting from overnight inoculum). The profile was obtained from seven independent growths for the wild-type pTRKH3 and pTRKH3-b with the insertion (except for the 24h point that was the result of two independent experiments), and from one experiment for the pTRKH3-b without the insertion. The standard error of the mean (SEM) is represented for each point. The specific growth rates ( $\mu$ ) were 0.40 ( $\pm 0.01$ ) for pTRKH3, 0.44 ( $\pm 0.02$ ) for the pTRKH3-b with the insertion and 0.51 for the pTRKH3-b without the insertion..... 43

**Figure 2.13.** Average PCN values for *L. lactis* LMG19460 cells harboring the wild-type pTRKH3 (light grey), the pTRKH3-b without the insertion (white) and the highest plasmid producer pTRKH3-b with the insertion (dark grey) at different growth times (6, 8, 10.5 and 24h). The PCN results were obtained from real-time qPCR analysis of seven independent growths for the wild-type pTRKH3 and pTRKH3-b with the insertion (except for the 24h point that was the result of two independent experiments), and from one experiment for the pTRKH3-b without the insertion. The standard error of the mean (SEM) is represented for each point. .... 44

**Figure 2.14.** Real-time qPCR standard curves done with five log serial dilutions (1, 10, 100, 1,000 and 10,000 pg) of a) *L. lactis* LMG19460 gDNA using primers specific for *feoA* gene; and pTRKH3 pDNA using primers specific for b) *repDE*, c) *copF* and d) *CT-RNA* genes. The experiments were performed in duplicate, except the one with *repDE* primers that was performed in triplicate. The linear regression trendlines are represented, along with the corresponding equation, coefficient of determination value ( $R^2$ ) and efficiency values..... 45

**Figure 2.15.** Map of *copF* gene in the pAM $\beta$ 1 origin of replication of pTRKH3 plasmid. The promoter predicted with Softberry BPROM program (<http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb>) (Solovyev and Salamov 2011) is indicated in light blue, while the two putative start codons are identified in yellow. The translation of the original *copF* gene results in a protein with 83 amino acids, while the translation of the *copF* present in the pTRKH3 plasmid could result in two putative truncate versions of CopF. The longer version has 77 amino acids, 72 of which are in common with the original CopF (the beginning of the common sequence is highlighted in green, while the different amino acids are in red). The shorter truncated version only have 68 amino acids, all in common with the original CopF, starting in the sequence identified in orange. .... 47

**Figure 2.16.** Average PCN values and average ratio values between PCN and number of *repDE* (dark grey columns), *copF* (light grey columns) and *CT-RNA* (medium grey columns) mRNA molecules for *L. lactis* LMG19460 cells harboring the wild-type and the six modified plasmids. The results were obtained from two-step mRNA real-time qPCR quantification of two independent growths in 100 mL shake flasks with 75 mL of M-17 supplemented with 20 g/L glucose, 30°C, 100 rpm (initial OD<sub>600nm</sub>=0.1, starting from an overnight inoculum). The PCN data (black line) is the average of seven independent experiments. The standard error of the mean (SEM) is represented for each point. .... 48

**Figure 2.17.** Principal component analysis of the mRNA quantification results and the associated eigenvalues. .... 49

**Figure 3.1.** Overview of the disciplines that contribute for the Synthetic Biology field and its several applications<sup>101</sup>. Figure from Freemont and Kitney (2017)<sup>101</sup>. .... 53

**Figure 3.2.** pSB1C3 plasmid backbone used to harbor the parts available at the Registry of Standard Biological Parts<sup>102</sup>. Figure from Registry of Standard Biological Parts<sup>102</sup>. .... 54

**Figure 3.3.** General structure of a plasmid backbone with a BioBrick part compatible with BioBrick RFC[10] Standard<sup>102</sup>. Figure from Registry of Standard Biological Parts<sup>102</sup>. .... 55

**Figure 3.4.** Assembly method used in the BioBrick RFC[10] Standard<sup>102</sup>. Figure from Registry of Standard Biological Parts<sup>102</sup>. .... 55

**Figure 3.5.** Assembly method used in the Berkeley RFC[21] Standard<sup>104</sup>. .... 56

**Figure 3.6.** Assembly method used in the Freiburg RFC[25] Standard<sup>104</sup>. .... 57

**Figure 3.7.** Overview of the 3 Antibiotic (3A) Assembly method<sup>102,106</sup>. Figure from Registry of Standard Biological Parts<sup>102</sup>. .... 58

**Figure 3.8.** Gibson Assembly method overview<sup>104,108</sup>. Figure from NEB (2017)<sup>108</sup>. .... 60

**Figure 3.9.** Overview of the strategy applied to assemble a minimal genome (whole-genome synthesis), using the Gibson Assembly technique. Figure from Hutchison et al. (2016)<sup>110</sup>. .... 61

**Figure 3.10.** General design for the assembly of a synthetic minimal plasmid. .... 73

**Figure 3.11.** In silico design of the plasmids with the *E. coli* and Gram-positive cassettes, and representation of the primers used for the fragments amplification..... 74

**Figure 3.12.** Average PCN values for *L. lactis* LMG19460 cells harboring the non-modified pTRKH3, the pTRKH3-b, the pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_1, the pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_1, pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_2 and pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_2 after 6h (light green), 8h (blue), 10.5h (yellow) and 24h of growth (dark green). The results were obtained from real-time qPCR analysis of three independent growths in 100 mL shake flasks with 75 mL of M-17 supplemented with 20 g/L glucose, 30°C, 100 rpm (initial OD<sub>600nm</sub>=0.1, starting from an overnight inoculum). The standard error of the mean (SEM) is represented for each point, except for the 8h and 24h points, which were quantified only once<sup>111</sup>. .... 76

<b>Figure 3.13.</b> In silico design of the pBio_WT_CUR and pBio_mutb_CUR plasmids. The primers used for the fragments amplification are also represented. ....	77
<b>Figure 3.14.</b> Quality assessment of the template plasmids for the BioBricks amplification. Lanes: <b>1-</b> Nzyladder III; <b>3-</b> pBio_erm/ery_p15A_pAMβ1_WT_2; <b>4-</b> pBio_erm/ery_p15A_pAMβ1_mutb_2; <b>5-</b> pTRKH3_SLP_4CL_DCS_CURS1.....	78
<b>Figure 3.15.</b> Amplification of the vectors and insert fragments for the pBio_WT_CUR and pBio_mutb_CUR construction. Lanes: <b>1-</b> Nzyladder III; <b>2-</b> vector amplified from pBio_erm/ery_p15A_pAMβ1_WT_2; <b>3-</b> vector amplified from pBio_erm/ery_p15A_pAMβ1_mutb_2; <b>4-</b> insert amplified from the pTRKH3_SLP_4CL_DCS_CURS1 plasmid. ....	78
<b>Figure 3.16.</b> Colony PCR to check for the assembly of the vector (with the p15A origin of replication, erythromycin resistance gene and WT vs mutant b pAMβ1 origin of replication) and insert (3 genes of the curcumin pathway under the control of the SLP promoter) BioBricks. Lanes: <b>1-</b> Nzyladder III; <b>2-</b> colony 1; <b>3-</b> colony 2; <b>4-</b> colony 3; <b>5-</b> colony 4; <b>6-</b> colony 5; <b>7-</b> colony 6; <b>8-</b> colony 7; <b>9-</b> colony 8; <b>10-</b> colony 9; <b>11-</b> colony 10. ....	79
<b>Figure 4.1.</b> Two classes and six types of CRISPR/Cas systems. The classes differ on the nature of the effector proteins, while the types depends on the type of cleavage they perform on the DNA. Figure from Pijkeren and Barrangou (2017) <sup>131</sup> . ....	87
<b>Figure 4.2.</b> Overview of the <i>S. pyogenes</i> type II CRISPR/Cas9 system. A) The CRISPR locus in <i>S. pyogenes</i> genome, with the cas genes operon, tracrRNA and CRISPR repeat-spacer array; B) After locus expression, Cas9 associates with the tracrRNA:crRNA duplex (processed by ribonuclease III); C) The RNA-guided target DNA is cleaved by Cas9. Figure from Doudna and Charpentier (2014) <sup>132</sup> . ....	87
<b>Figure 4.3.</b> Distribution of the different types of CRISPR/Cas systems in bacterial and archaea sequenced genomes. Figure from Makarova et al. (2015) <sup>133</sup> . ....	88
<b>Figure 4.4.</b> Differences between A) the naturally occurring CRISPR/Cas9 systems and B) the engineered ones, where the tracrRNA:crRNA duplex is replaced by a sgRNA. Figure from Sander and Joung (2014) <sup>134</sup> . ....	89
<b>Figure 4.5.</b> CRISPR/Cas9 system used to eliminate a cryptic plasmid from <i>Leuconostoc citreum</i> CB2567 <sup>130</sup> . The authors engineered a synthetic vector with the cas9 gene from <i>Streptococcus pyogenes</i> and a chimeric gRNA with a 20 nucleotide guiding sequence that targeted the pCB42 cryptic plasmid. After strain transformation with this synthetic vector, Cas9 endonuclease cleaved the cryptic plasmid, guided by the gRNA. Figure from Jang et al. (2017) <sup>130</sup> . ....	90
<b>Figure 4.6.</b> Overview of the CRISPR/Cas9 <sup>D10A</sup> system used in <i>Lactobacillus casei</i> <sup>123</sup> . The plasmid pLCNICK containing the cas9 <sup>D10A</sup> gene, as well as the sgRNA and the homologous arms is used to transform <i>L. casei</i> by electroporation. The nickase Cas9 <sup>D10A</sup> is guided by the sgRNA to the target gene in the bacterial genome and generates a nick in the chromosome, triggering the cell homologous recombination-dependent repair process, where the homologous arms work as the repair donors. In the end, the pLCNICK plasmid is cured and the strain is ready for a new round of genome editing. Figure from Song et al. (2017) <sup>123</sup> . ....	91
<b>Figure 4.7.</b> Comparison between Cas9 nuclease and Cas9 <sup>D10A</sup> nickase. Figure from Doudna and Charpentier (2014) <sup>132</sup> . ....	91
<b>Figure 4.8.</b> Design and function of the pKCcas9dO plasmid. Scocas9: <i>S. coelicolor</i> codon-optimized cas9; Acc(3)IV: apramycin resistance gene; oriT: origin of transference for conjugation; Repts: thermo-sensitive origin of replication; j23119: promoter controlling the expression of sgRNA; sgRNA: single guide RNA constituted by a 20 bp guide sequence, a 42 bp Cas9 handle and a 41 bp terminator; homology arms: sequences homologous to the upstream and downstream regions of the target gene for knockout. Figure from Huang et al. (2015) <sup>135</sup> . ....	92
<b>Figure 4.9.</b> Overview of the knockout process in <i>S. coelicolor</i> using the pKCcas9dO plasmid. The plasmid enters the bacterial cells by conjugation, using <i>E. coli</i> . The Cas9 endonuclease associates with the sgRNA, recognizes the chromosomal site complementary to the guide sequence and performs a double strand break in the <i>S. coelicolor</i> genome, activating the cell homology-directed recombination repair system. In the end, the target gene has been removed by two rounds of crossovers and the plasmid is cured by increasing the growth temperature to 37°C. Figure from Huang et al. (2015) <sup>135</sup> . ....	93
<b>Figure 4.10.</b> Main steps of the recombineering strategy for gene knockout developed by Datsenko and Wanner (2000). Figure from Datsenko and Wanner (2000) <sup>121</sup> . ....	100

<b>Figure 4.11.</b> In silico design of the amplification, SOEing PCR and cloning strategies to construct the nth gene specific CRISPR/Cas9 plasmid (pKCcas9dO_nth_apra). .....	101
<b>Figure 4.12.</b> Schematic representation of the A) pKCcas9dO_nth_apra and B) pTRKH3 plasmids with the necessary primers for the fragments amplification. ....	108
<b>Figure 4.13.</b> Percentage of proteins of <i>L. lactis</i> LMG19460 belonging to each COG functional category, calculated by BASys. C- energy production and conversion, D- cell cycle control and mitosis, E- amino acid metabolism and transport, F- nucleotide metabolism and transport, G- carbohydrate metabolism and transport, H- coenzyme metabolism, I- lipid metabolism, J- translation, K- transcription, L- DNA replication, recombination and repair, M- cell wall/membrane/envelope biogenesis, N- Cell motility, O- post-translational modification, protein turnover, chaperone functions, P- Inorganic ion transport and metabolism, Q- secondary metabolites biosynthesis, transport and catabolism, T- signal transduction, R- general functional prediction only, S- function unknown. ....	112
<b>Figure 4.14.</b> Percentage of proteins of <i>L. lactis</i> LMG19460 with 100 amino acids or less, 101-500 amino acids, 501-1000 amino acids, 1001-1500 amino acids or more than 1500 amino acids, as predicted by BASys. ....	113
<b>Figure 4.15.</b> Percentage of proteins of <i>L. lactis</i> LMG19460 with each amino acid residue, calculated by BASys. A- alanine; C- cysteine; D- aspartic acid; E- glutamic acid; F- phenylalanine; G- glycine; H- histidine; I- isoleucine; K- lysine; L- leucine; M- methionine; N- asparagine; P- proline; Q- glutamine; R- arginine; S- serine; T- threonine; V- valine; W- tryptophan; Y- tyrosine. ....	114
<b>Figure 4.16.</b> Position of the yjhF and adhE genes (highlighted in red) from <i>L. lactis</i> IL1403 identified by the Optflux software <sup>140</sup> as possible knockout targets in the glycolysis/gluconeogenesis pathway. ....	116
<b>Figure 4.17.</b> Position of the purH and nrdF genes (highlighted in red) from <i>L. lactis</i> IL1403 identified by the Optflux software <sup>140</sup> as possible knockout targets in the purine metabolism pathway. ....	117
<b>Figure 4.18.</b> Position of the pgi gene (highlighted in red) from <i>L. lactis</i> IL1403 identified by the Optflux software <sup>140</sup> as a possible knockout target in the glycolysis/gluconeogenesis pathway. ....	118
<b>Figure 4.19.</b> 1 µg of each pDNA purified from the several colonies exposed to different UV light conditions. Lanes: 1- Nzyladder III; 2- colony 1 (6 cm, 24h of growth); 3- colony 2 (6 cm, 24h of growth); 4- colony 3 (6 cm, 24h of growth); 5- colony 4 (6 cm, 24h of growth); 6- colony 5 (6 cm, 16h of growth); 7- colony 6 (0 cm, 24h of growth); 8- colony 7 (0 cm, 24h of growth); 9- colony 8 (0 cm, 24h of growth); 10- colony 9 (0 cm, 24h of growth); 11- colony 10 (0 cm, 16h of growth); 12- colony 11 (no exposure, 24h of growth). ....	120
<b>Figure 4.20.</b> pKD13, pKD46 and pCP20 purified plasmids. Lanes: 1- Nzyladder III; 2- pKD13 plasmid (3434 bp); 3- pKD46 plasmid (6329 bp); 4- pCP20 (9400 bp). ....	121
<b>Figure 4.21.</b> Lanes: 1- Nzyladder III; 2- Positive colony PCR of <i>L. lactis</i> LMG19460 transformed with pKD46 (614 bp). ....	122
<b>Figure 4.22.</b> In silico representation of the linear recombination cassette amplified from pKD13, using the KO_nth_F and KO_nth_R primers. ....	122
<b>Figure 4.23.</b> Lanes: 1- Nzyladder III; 2- Linear recombination cassette amplified from pKD13 plasmid (1414 bp). ....	123
<b>Figure 4.24.</b> PCR result after transformation of <i>L. lactis</i> LMG19460 harboring the pKD46 plasmid with the linear recombination cassette. Lanes: 1- Nzyladder III; 2- Cassette incubated with 3 µM TTAB, colony 1; 3- Cassette incubated with 3 µM TTAB, colony 2; 4- Cassette incubated with 30 µM TTAB, colony 1; 5- Cassette incubated with 30 µM TTAB, colony 2; 6- Cassette incubated with 30 µM TTAB, colony 3. ....	123
<b>Figure 4.25.</b> Amplification of nth gene specific A) sgRNA and B) the two homologous arms. The correct band corresponding to the sgRNA should have 136 bp, while the bands corresponding to the two homologous arms should have 1,028 and 955 bp, respectively. The DNA ladder used in both agarose gels was Nzyladder III. ....	124
<b>Figure 4.26.</b> Agarose gels with the result of the A) first SOEing PCR to join both nth homologous arms (1983 bp) and the B) second SOEing PCR to join the previous fragment to the nth sgRNA (2119 bp). The DNA ladder used in both agarose gels was Nzyladder III. ....	125
<b>Figure 4.27.</b> Lanes: 1- Nzyladder III; 2- PCR product from amplification of the fragment containing the nth specific sgRNA and the two homologous arms, with the primers sgRNA_F and HA2_R_Apa (2131 bp). ....	125

<b>Figure 4.28.</b> A) Vector pKCcas9dO (12,041 + 1,341 bp) and B) fragment with nth specific sgRNA and two homologous arms (2,117 + 7 + 7 bp), digested with SpeI and ApaI. The DNA ladder used in both agarose gels was NzyLadder III. ....	126
<b>Figure 4.29.</b> Confirmation of the molecular cloning of pKCcas9dO (12,041 bp) with nth specific fragment containing sgRNA and two homologous arms (2,117 bp). Lanes: <b>1-</b> NzyLadder III; <b>2-</b> colony 1; <b>3-</b> colony 2. ....	126
<b>Figure 4.30.</b> Lanes: <b>1-</b> NzyLadder III; <b>2-</b> Non-digested final plasmid (pKCcas9dO cloned with the fragment containing nth specific sgRNA and two homologous arms); <b>3-</b> Final plasmid (pKCcas9dO cloned with the fragment containing nth specific sgRNA and two homologous arms) digested with EcoRI (8785 + 4372 + 989 bp).....	127
<b>Figure 4.31.</b> In silico representation of the pKCcas9dO_nth_apra plasmid engineered in this study, comparing with the parental pKCcas9dO plasmid. ....	127
<b>Figure 4.32.</b> PCR amplification of the gDNA purified from several colonies of <i>L. lactis</i> LMG19460 transformed with pKCcas9dO_nth_apra plasmid. Lanes: <b>1-</b> NzyLadder III; <b>2-</b> colony 1; <b>3-</b> colony 2; <b>4-</b> colony 3.....	128
<b>Figure 4.33.</b> PCR amplification of the linear fragment containing the two nth homologous arms. The molecular marker used was the NzyLadder III.....	130
<b>Figure 4.34.</b> Colony PCR results from the second thiostrepton induction experiment with nth_conf_F/R primers. Lanes: <b>1-</b> NzyLadder III; <b>2-</b> strain transformed with only the pKCcas9dO_nth_apra plasmid and induced with 0.01 µg/mL thiostrepton, colony 1; <b>3-</b> colony 2; <b>4-</b> colony 3; <b>5-</b> strain transformed with the the pKCcas9dO_nth_apra plasmid and the linear fragment and induced with 0.3 µg/mL thiostrepton, colony 1; <b>6-</b> NzyLadder III; <b>7-</b> strain transformed with only the pKCcas9dO_nth_apra plasmid and induced with 0.01 µg/mL thiostrepton, colony 4; <b>8-</b> colony 5; <b>9-</b> colony 6. ....	131
<b>Figure 4.35.</b> Colony PCR results from the second thiostrepton induction experiment with the Cas9_conf_F/R. Lanes: <b>1-</b> NzyLadder III; <b>2-</b> strain transformed with only the pKCcas9dO_nth_apra plasmid and induced with 0.01 µg/mL thiostrepton, colony 1; <b>3-</b> colony 2; <b>4-</b> colony 3; <b>5-</b> colony 4; <b>6-</b> positive control (pKCcas9dO_nth_apra purified from <i>E. coli</i> DH5α). ....	132
<b>Figure 4.36.</b> <b>1-</b> NzyLadder III; <b>2-</b> PCR amplification of the erythromycin resistance gene from pTRKH3 with Ery_F/R primers. ....	133
<b>Figure 4.37.</b> <b>1-</b> NzyLadder III; <b>2-</b> PCR amplification of the Apa_Ery fragment from pKCcas9dO_nth_apra plasmid with Apa_Ery_F/R primers. ....	133
<b>Figure 4.38.</b> PCR amplification of the shorter pTRKH3 plasmid that will work as the vector for assembling the necessary CRISPR/Cas9 components, with pT vector_F/R primers and using different annealing temperatures. <b>1-</b> NzyLadder III; <b>2-</b> 60°C; <b>3-</b> 72°C. ....	134
<b>Figure 4.39.</b> PCR amplification of the insert with the necessary CRISPR/Cas9 components, using different annealing temperatures. <b>1-</b> NzyLadder III; <b>2-</b> 56°C; <b>3-</b> 60°C; <b>4-</b> 68°C; <b>5-</b> 72°C. ....	135
<b>Figure 5.1.</b> Stages in tumor progression inhibited by curcumin. Figure from Hatcher et al. (2008) <sup>181</sup> . ....	142
<b>Figure 5.2.</b> Cyclin-dependent pathway by which curcumin can alter the deregulated cell cycle of cancer cells. Figure from Sa and Das (2008) <sup>180</sup> . ....	143
<b>Figure 5.3.</b> p53-dependent pathway by which curcumin can alter the deregulated cell cycle of cancer cells. Figure from Sa and Das (2008) <sup>180</sup> . ....	144
<b>Figure 5.4.</b> p53-independent pathway by which curcumin can alter the deregulated cell cycle of cancer cells. Figure from Sa and Das (2008) <sup>180</sup> . ....	144
<b>Figure 5.5</b> Chemical structures of the three major natural curcuminoids. Figure from Metzler et al. (2012) <sup>175</sup> . ....	145
<b>Figure 5.6.</b> Keto- and enol-tautomeric forms of curcumin. Figure from Basnet and Skalko-Basnet (2011) <sup>177</sup> . ....	145
<b>Figure 5.7.</b> Degradation products of curcumin through alkaline hydrolysis, autoxidation and irradiation. Figure from Metzler et al. (2012) <sup>175</sup> . ....	146
<b>Figure 5.8.</b> Main products of curcumin metabolism. Figure from Metzler et al. (2012) <sup>175</sup> . ....	147
<b>Figure 5.9.</b> Curcuminoid biosynthesis in <i>C. longa</i> . Figure from Katsuyama et al. (2009) <sup>120</sup> . ....	150

<b>Figure 5.10.</b> Artificial curcuminoid biosynthetic pathway. Figure from Katsuyama et al. (2008) <sup>119</sup> .....	151
<b>Figure 5.11.</b> Second artificial curcuminoid biosynthetic pathway and representation of the hydroxylation and methylation steps of carboxylic acids. Figure from Katsuyama et al. (2008) <sup>119</sup> .....	152
<b>Figure 5.12.</b> Artificial pathway proposed.....	152
<b>Figure 5.13.</b> 4CL sequence codon usage (red line) in <i>E. coli</i> K12 a) before codon optimization and b) after codon optimization using JCat online software <sup>200</sup> . The blue line represents the mean codon usage in <i>E. coli</i> K12, considering all of the organism's genes, and the grey lines depicts the corresponding standard deviation intervals. ....	163
<b>Figure 5.14.</b> DCS sequence codon usage (red line) in <i>E. coli</i> K12 a) before codon optimization and b) after codon optimization using JCat online software <sup>200</sup> . The blue line represents the mean codon usage in <i>E. coli</i> K12, considering all of the organism's genes, and the grey lines depicts the corresponding standard deviation intervals. ....	163
<b>Figure 5.15.</b> CURS1 sequence codon usage (red line) in <i>E. coli</i> K12 a) before codon optimization and b) after codon optimization using JCat online software <sup>200</sup> . The blue line represents the mean codon usage in <i>E. coli</i> K12, considering all of the organism's genes, and the grey lines depicts the corresponding standard deviation intervals. ....	164
<b>Figure 5.16.</b> Final design of the 4CL, DCS and CURS1 genes, with several flanking unique restriction enzyme sites, that was sent for synthesis by Nzytech.....	166
<b>Figure 5.17.</b> In silico design of the 4CL, DCS and CURS1 genes amplification step, with restriction enzyme sites compatible with cloning in the pTRKH3 plasmid.....	166
<b>Figure 5.18.</b> Cloning steps, starting with the backbone plasmid pTRKH3, followed by promoters and individual genes cloning, and finishing with the 3 genes cloned as operons under the control of each promoter. ....	167
<b>Figure 5.19.</b> Quality characterization and confirmation of the pTRKH3 plasmid purified from <i>E. coli</i> DH5 $\alpha$ . Lanes: 1- Nzyladder III; 2- Non-digested pTRKH3 sample A; 3- pTRKH3 sample A digested with HindIII and EcoRI (5265 + 1523 + 978 bp); 4- Non-digested pTRKH3 sample B; 5- pTRKH3 sample B digested with HindIII and EcoRI (5265 + 1523 + 978 bp).....	168
<b>Figure 5.20.</b> Two samples of genomic DNA from <i>L. acidophilus</i> ATCC4356. Lanes: 1- Nzyladder III; 2- gDNA sample A (1766 ng/ $\mu$ L); 3- gDNA sample B (1900 ng/ $\mu$ L).....	168
<b>Figure 5.21.</b> PCR products from the amplification of the A) SLP (305 bp) and B) LDH (296 bp) promoters from the gDNA of the <i>L. acidophilus</i> ATCC4356. Lanes: 1- Nzyladder III; 2- SLP promoter fragment; 3- Nzyladder III; 4- LDH promoter fragment. ....	169
<b>Figure 5.22.</b> Double digestion with EcoRV and BamHI of A) pTRKH3 and B) SLP and LDH promoters. Lanes: 1- Nzyladder III; 2- Non-digested pTRKH3; 3- pTRKH3 digested with EcoRV and BamHI; 4- Nzyladder III; 5- SLP promoter digested with EcoRV and BamHI; 6- LDH promoter digested with EcoRV and BamHI. ....	169
<b>Figure 5.23.</b> Colony PCR amplifications. Lanes: 1- Nzyladder III; 2- non-digested pTRKH3; 3- pTRKH3 ligated with the SLP promoter (3h, RT, colony 1); 4- pTRKH3 ligated with the SLP promoter (3h, RT, colony 2); 5- pTRKH3 ligated with the SLP promoter (3h, RT, colony 3); 6- pTRKH3 ligated with the SLP promoter (OV, 4°C, colony 1); 7- pTRKH3 ligated with the SLP promoter (OV, 4°C, colony 2); 8- pTRKH3 ligated with the LDH promoter (3h, RT, colony 1); 9- pTRKH3 ligated with the LDH promoter (3h, RT, colony 2); 10- pTRKH3 ligated with the LDH promoter (3h, RT, colony 3); 11- pTRKH3 ligated with the LDH promoter (OV, 4°C, colony 1); 12- pTRKH3 ligated with the LDH promoter (OV, 4°C, colony 2). RT- room temperature; OV- overnight.....	170
<b>Figure 5.24.</b> Colony PCR amplifications. Lanes: 1- Nzyladder III; 2- digested and recircularized pTRKH3; 3- pTRKH3 ligated with the LDH promoter (3h, RT, colony 1); 4- pTRKH3 ligated with the LDH promoter (3h, RT, colony 2); 5- pTRKH3 ligated with the LDH promoter (3h, RT, colony 3); 6- pTRKH3 ligated with the LDH promoter (O/N, 4°C, colony 1); 7- pTRKH3 ligated with the LDH promoter (O/N, 4°C, colony 2); 8- pTRKH3 ligated with the LDH promoter (O/N, 4°C, colony 3). RT- room temperature; O/N- overnight. ....	170
<b>Figure 5.25.</b> Alignment of the expected sequence of the pTRKH3 cloned with the SLP or LDH promoters against the real sequence. The gaps or mismatches are represented as white lines/gaps in the red arrow. ....	171
<b>Figure 5.26.</b> Characterization and confirmation of pUC57-4CL, DCS and CURS1 plasmids. Lanes: 1- Nzyladder III; 2- Non-digested pUC57-4CL; 3- pUC57-4CL digested with NdeI and BamHI (2458 + 1949 + 251 + 11 bp); 4- Non-digested pUC57-DCS; 5- pUC57-DCS digested with NheI and BamHI (2709 + 1208 + 11 bp); 6- pUC57-DCS digested	

with XmnI (2790 + 1138 bp); **7-** Non-digested pUC57-CURS1; **8-** pUC57-CURS1 digested with BamHI and AgeI (2709 + 1205 + 11 bp); **9-** pUC57-CURS1 digested with XmnI (3925 bp)..... 172

**Figure 5.27.** Digestion of the amplified and purified genes with BamHI and Sall. Lanes: **1-** Nzyladder III; **2-** 4CL gene fragment; **3-** DCS gene fragment; **4-** CURS1 gene fragment. .... 173

**Figure 5.28.** pTRKH3\_SLP and pTRKH3\_LDH A) non-digested and B) digested with BamHI and Sall. Lanes: **1-** Nzyladder III; **2-** Non digested pTRKH3\_SLP; **3-** Non-digested pTRKH3\_LDH; **4-** Nzyladder III; **5-** pTRKH3\_SLP digested with BamHI and Sall; **6-** pTRKH3\_LDH digested with BamHI and Sall. .... 173

**Figure 5.29.** Confirmation of the clones resulting from the ligation of the vectors pTRKH3\_SLP and pTRKH3\_LDH with the inserts 4CL, DCS and CURS1. The confirmation was performed by digestion of all the clones with BamHI and Sall. Lanes: **A) 1-** Nzyladder III; **2-** pTRKH3\_SLP\_4CL, colony 1; **3-** pTRKH3\_SLP\_DCS; **4-** pTRKH3\_SLP\_CURS1; **5-** pTRKH3\_LDH\_4CL; **6-** pTRKH3\_LDH\_CURS, colony 1; **B) 7-** Nzyladder III; **8-** pTRKH3\_LDH\_4CL, colony 1; **9-** pTRKH3\_LDH\_4CL, colony 2; **10-** pTRKH3\_LDH\_4CL, colony 3; **11-** pTRKH3\_LDH\_DCS, colony 1; **12-** pTRKH3\_LDH\_DCS, colony 2; **13-** pTRKH3\_LDH\_DCS, colony 3; **C) 14-** Nzyladder III; **15-** pTRKH3\_SLP\_4CL, colony 2; **16-** pTRKH3\_SLP\_4CL, colony 3; **17-** pTRKH3\_SLP\_4CL, colony 4; **18-** pTRKH3\_LDH\_CURS, colony 2; **19-** pTRKH3\_LDH\_CURS, colony 3; **20-** pTRKH3\_LDH\_CURS, colony 4; **21-** pTRKH3\_LDH\_CURS, colony 5. .... 174

**Figure 5.30.** Alignment performed with the SnapGene software between the expected pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1 sequences and the real ones obtained by sequencing. .... 175

**Figure 5.31.** Alignment performed with the SnapGene software between the expected pTRKH3\_LDH\_4CL, pTRKH3\_LDH\_DCS and pTRKH3\_LDH\_CURS1 sequences and the real ones obtained by sequencing. .... 176

**Figure 5.32.** Digestion with AvrII and Sall of the A) vectors pTRKH3\_SLP\_4CL and pTRKH3\_LDH\_4CL, B) and for the DCS gene insert. Lanes: **1-** Nzyladder III; **2-** pTRKH3\_LDH\_4CL digested with AvrII and Sall; **3-** pTRKH3\_SLP\_4CL digested with AvrII and Sall; **4-** Nzyladder III; **5-** DCS gene fragment digested with AvrII and Sall. .... 177

**Figure 5.33.** Colony PCR of colonies resulting from the ligation of pTRKH3\_SLP\_4CL and pTRKH3\_LDH\_4CL with the insert DCS. Lanes: **1-** Nzyladder III; **2-** pTRKH3\_LDH\_4CL\_DCS, colony 1; **3-** pTRKH3\_LDH\_4CL\_DCS, colony 2; **4-** pTRKH3\_SLP\_4CL\_DCS, colony 1; **5-** pTRKH3\_SLP\_4CL\_DCS, colony 2..... 177

**Figure 5.34.** Confirmation of the pTRKH3\_SLP\_4CL\_DCS clones. Lanes: **1-** Nzyladder III; **2-** Non-digested pTRKH3\_SLP\_4CL\_DCS, colony 1; **3-** Non-digested pTRKH3\_SLP\_4CL\_DCS, colony 2; **4-** pTRKH3\_SLP\_4CL\_DCS, colony 1, digested with AvrII and Sall; **5-** pTRKH3\_SLP\_4CL\_DCS, colony 2, digested with AvrII and Sall. .... 178

**Figure 5.35.** Alignment performed with the SnapGene software between the expected pTRKH3\_SLP\_4CL\_DCS sequence and the real one obtained by sequencing. .... 178

**Figure 5.36.** Confirmation of the pTRKH3\_LDH\_4CL\_DCS and pTRKH3\_SLP\_4CL\_DCS\_CURS1 clones. Lanes: **1-** Nzyladder III; **2-** pTRKH3\_SLP\_4CL\_DCS\_CURS1, colony 1; **3-** pTRKH3\_SLP\_4CL\_DCS\_CURS1, colony 2; **4-** pTRKH3\_SLP\_4CL\_DCS\_CURS1, colony 3; **5-** pTRKH3\_SLP\_4CL\_DCS\_CURS1, colony 4; **6-** pTRKH3\_LDH\_4CL\_DCS, colony 1; **7-** pTRKH3\_LDH\_4CL\_DCS, colony 2; **8-** pTRKH3\_LDH\_4CL\_DCS, colony 3; **9-** pTRKH3\_LDH\_4CL\_DCS, colony 4..... 179

**Figure 5.37.** Alignment performed with the SnapGene software between the expected pTRKH3\_LDH\_4CL\_DCS and pTRKH3\_SLP\_4CL\_DCS\_CURS1 sequences and the real ones obtained by sequencing. .... 179

**Figure 5.38.** Confirmation of the pTRKH3\_LDH\_4CL\_DCS\_CURS1 clone. Lanes: **1-** Nzyladder III; **2 to 4-** results from other experiment; **5-** Non-digested pTRKH3\_LDH\_4CL\_DCS\_CURS1, colony 1; **6-** Non-digested pTRKH3\_LDH\_4CL\_DCS\_CURS1, colony 2; **7-** Non-digested pTRKH3\_LDH\_4CL\_DCS\_CURS1, colony 3; **8 and 9-** results from other experiment; **10-** pTRKH3\_LDH\_4CL\_DCS\_CURS1 digested with AscI and Sall, colony 1; **11-** pTRKH3\_LDH\_4CL\_DCS\_CURS1 digested with AscI and Sall, colony 3..... 180

**Figure 5.39.** Alignment performed with the SnapGene software between the expected pTRKH3\_LDH\_4CL\_DCS\_CURS1 sequence and the real one obtained by sequencing. .... 180

**Figure 5.40.** A) C) Non-digested and B) BamHI + Sall digested plasmids after purification from *L. lactis* LMG19460. Lanes: **1-** Nzyladder III; **2-** non-digested pTRKH3\_SLP\_4CL; **3-** non-digested pTRKH3\_SLP\_DCS; **4-** non-digested

pTRKH3\_SLP\_CURS1; **5-** non-digested pTRKH3\_SLP\_4CL\_DCS\_CURS1; **6-** non-digested pTRKH3\_LDH\_4CL; **7-** non-digested pTRKH3\_LDH\_DCS; **8-** non-digested pTRKH3\_LDH\_4CL\_DCS; **9-** Nzyladder III; **10-** pTRKH3\_SLP\_4CL digested with BamHI and Sall (7587 + 1929 bp); **11-** pTRKH3\_SLP\_DCS digested with BamHI and Sall (7587 + 1196 bp); **12-** pTRKH3\_SLP\_CURS1 digested with BamHI and Sall (7587 + 1188 bp); **13-** pTRKH3\_SLP\_4CL\_DCS\_CURS1 digested with BamHI and Sall (7587 + 4281 bp); **14-** pTRKH3\_LDH\_4CL digested with BamHI and Sall (7578 + 1929 bp); **15-** pTRKH3\_LDH\_DCS digested with BamHI and Sall (7578 + 1196 bp); **16-** pTRKH3\_LDH\_4CL\_DCS digested with BamHI and Sall (7587 + 3110 bp); **17-** Nzyladder III; **18-** non-digested pTRKH3\_LDH\_4CL\_DCS\_CURS1; **19-** non-digested pTRKH3\_LDH\_CURS1, colony 1; **20-** non-digested pTRKH3\_LDH\_CURS1, colony 2. .... 181

**Figure 5.41.** *L. lactis* LMG19460 transformed with pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1. Lanes: **1-** Nzyladder III; **2-** colony 1, primer for pTRKH3\_SLP\_4CL; **3-** colony 1, primer for pTRKH3\_SLP\_DCS; **4-** colony 1, primer for pTRKH3\_SLP\_CURS1; **5-** colony 2, primer for pTRKH3\_SLP\_4CL; **6-** colony 2, primer for pTRKH3\_SLP\_DCS; **7-** colony 2, primer for pTRKH3\_SLP\_CURS1; **8-** colony 3, primer for pTRKH3\_SLP\_4CL; **9-** colony 3, primer for pTRKH3\_SLP\_DCS; **10-** colony 3, primer for pTRKH3\_SLP\_CURS1; **11-** colony 4, primer for pTRKH3\_SLP\_4CL; **12-** colony 4, primer for pTRKH3\_SLP\_DCS; **13-** colony 4, primer for pTRKH3\_SLP\_CURS1. .... 182

**Figure 5.42.** pDNA purified from *L. lactis* LMG19460 transformed with pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1. Lanes: **1-** Nzyladder III; **2-** colony 1; **3-** colony 2; **4-** colony 3; **5-** colony 4. .... 182

**Figure 5.43.** Lanes: **1-** Nzyladder III; **2-** pDNA purified from *E. coli* MG1655 harboring the plasmids pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1. .... 183

**Figure 5.44.** Curcumin standard curve and respective equation. .... 184

**Figure 5.45.** Ferulic acid standard curve and respective equation. .... 184

**Figure 5.46.** Chromatograms obtained for the highest concentrations of A) curcumin (400 µg/mL) and B) ferulic acid (200 µg/mL, corresponding to 1.03 mM). .... 185

**Figure 5.47.** Growth curves from the first curcumin production experiment, with *L. lactis* LMG19460 harboring pTRKH3\_SLP\_4CL\_DCS\_CURS1 (closed squares for initial OD of 0.1 and open squares for initial OD of 0.01), or the negative control pTRKH3 (closed circles for initial OD of 0.1 and open circles for initial OD of 0.01). The *E. coli* DH5α growth curves are represented with triangles, being closed when harbors pTRKH3 and open when harbors pTRKH3\_SLP\_4CL\_DCS\_CURS1. .... 186

**Figure 5.48.** Modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>) autoclaved with 20 mM (3,884 µg/mL) and 2 mM of ferulic acid (388.4 µg/mL). .... 187

**Figure 5.49.** Growth curves from the curcumin production experiment at 25°C in M9 minimal salt medium, with *L. lactis* LMG19460 harboring pTRKH3\_SLP\_4CL\_DCS\_CURS1 (open symbols) or the negative control pTRKH3 (closed symbols). Both were grown in the presence of three different ferulic acid concentrations: 0 mM (circles), 2mM (squares) or 4 mM (triangles). .... 188

**Figure 5.50.** Chromatograms obtained for ferulic acid detection after 3 days of growth, in *L. lactis* LMG19460 harboring A) pTRKH3\_SLP\_4CL\_DCS\_CURS1 or B) pTRKH3. .... 189

**Figure 5.51.** Agarose gel with 10 µL pDNA samples purified from *L. lactis* LMG19460 pTRKH3 and *L. lactis* LMG19460 pTRKH3\_SLP\_4CL\_DCS\_CURS1, grown in the presence of 0, 2 or 4 mM ferulic acid. The samples were taken after 24h, 3 days or 6 days of growth. .... 191

**Figure 5.52.** Growth curves from the growth experiment in optimized growth conditions (25°C, 2 mM ferulic acid) during 7 days of growth in A) M17, B) modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>) and C) complete modified M9 minimal salt medium (with thiamine and trace elements, and CaCO<sub>3</sub> in *E. coli* condition). The clones tested were *L. lactis* LMG19460 clones harboring pTRKH3 (closed circles), pTRKH3\_SLP\_4CL\_DCS\_CURS1 (closed squares), pTRKH3\_LDH\_4CL\_DCS\_CURS1 (closed triangles) and the 3 single gene clones (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1) (open circles), as well as *E. coli* K12 MG1655 (DE3) harboring the pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1 plasmids (not tested in M17 medium) (open squares). .... 193

**Figure 5.53.** Shake flasks from the growth experiment in optimized growth conditions (25°C, 2 mM ferulic acid) after 7 days of growth in A) M17, B) modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>) and C) complete modified M9 minimal salt medium (with thiamine and trace elements, and CaCO<sub>3</sub> in *E. coli* condition). The clones tested were photographed in the following order: *E. coli* K12 MG1655 (DE3) harboring the pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1 plasmids (not tested in M17 medium), as well as *L. lactis* LMG19460 clones harboring pTRKH3, pTRKH3\_SLP\_4CL\_DCS\_CURS1, pTRKH3\_LDH\_4CL\_DCS\_CURS1 and the 3 single gene clones (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1). ..... 194

**Figure 5.54.** Ferulic acid quantification by HPLC of the samples from the third growth experiment, after A) 24h of growth, B) 72h of growth and C) 7 days of growth. The data for the 3 media used (M9, complete M9 and M17) are presented. The strains tested were *L. lactis* LMG19460 harboring the pTRKH3 plasmid (dark blue columns), *L. lactis* LMG19460 harboring the pTRKH3\_SLP\_4CL\_DCS\_CURS1 (red columns), *L. lactis* LMG19460 harboring the pTRKH3\_LDH\_4CL\_DCS\_CURS1 (green columns), *L. lactis* LMG19460 with the 3 single-gene plasmids (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1) (purple columns) and *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1) (light blue columns). All the samples were the columns reached the top of the graph means that it exceeded the equipment detection limit. The yy axis was deliberately truncated by the 120 µg/mL level to make it easier to visualize the smaller concentrations. .... 195

**Figure 5.55.** Areas (a.u.) of the extra peak that appeared in the ferulic acid chromatogram from HPLC analysis of the samples from the third growth experiment, after A) 24h of growth, B) 72h of growth and C) 7 days of growth. The data for the 3 media used (M9, complete M9 and M17) are presented. The strains tested were *L. lactis* LMG19460 harboring the pTRKH3 plasmid (dark blue columns), *L. lactis* LMG19460 harboring the pTRKH3\_SLP\_4CL\_DCS\_CURS1 (red columns), *L. lactis* LMG19460 harboring the pTRKH3\_LDH\_4CL\_DCS\_CURS1 (green columns), *L. lactis* LMG19460 with the 3 single-gene plasmids (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1) (purple columns) and *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1) (light blue columns). .... 196

**Figure 5.56.** Chromatograms for ferulic acid obtained after 3 days of growth in modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>). The strains tested were A) *L. lactis* LMG19460 harboring the pTRKH3 plasmid, B) *L. lactis* LMG19460 harboring the pTRKH3\_SLP\_4CL\_DCS\_CURS1, C) *L. lactis* LMG19460 harboring the pTRKH3\_LDH\_4CL\_DCS\_CURS1, D) *L. lactis* LMG19460 with the 3 single-gene plasmids (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1) and E) *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1). ..... 197

**Figure 5.57.** Curcumin chromatogram for *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1), obtained after 3 days of growth in modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>). ..... 198

**Figure 5.58.** Areas (a.u.) of the curcumin peaks that appeared in the HPLC chromatogram of the *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1), after 0h, 24h, 72h and 7 days of growth, with A) M9 medium and B) complete M9 medium. The area values for the 2 different peaks (first peak in blue columns and second peak in red columns) that appeared in the chromatogram are presented. .... 198

**Figure 5.59.** Curcumin A) concentration and B) specific yield after analysis of the second peak of the curcumin HPLC chromatogram of the *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1) from the third growth experiment, after 0h, 24h, 72h and 7 days of growth, with M9 (blue columns) and complete M9 medium (red columns). ..... 199

**Figure 5.60.** Agarose gel with 1 µg pDNA samples purified after 24h, 3 days and 7 days of growth. Lanes: **M-** Nzyladder III; **1-** *L. lactis* LMG19460 harboring the pTRKH3 plasmid, grown in modified M9 medium or **2-** M17 medium; **3-** *L. lactis* LMG19460 harboring the pTRKH3\_SLP\_4CL\_DCS\_CURS1, grown in modified M9 medium or **4-** M17 medium; **5-** *L. lactis* LMG19460 harboring the pTRKH3\_LDH\_4CL\_DCS\_CURS1, grown in modified M9 medium or **6-** M17 medium; **7-** *L. lactis* LMG19460 with the 3 single-gene plasmids (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1), grown in modified M9 medium or **8-** M17 medium; **9-** *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1), grown in modified M9 medium or **10-** complete M9 medium. .... 200

**Figure 5.61.** CURS1 mRNA standard curve and respective equation. .... 201

**Figure 5.62.** Number of CURS1 mRNA molecules per cell, quantified by real-time quantitative PCR, after 3 days of growth. .... 202

## LIST OF ABBREVIATIONS

4CL	4-coumarate:CoA ligase
4-VG	4-vinylguaiacol
ACC	Acetyl-CoA carboxylase
ADP	Adenosine diphosphate
Amp <sup>R</sup>	Ampicillin resistance gene
ATP	Adenosine triphosphate
BCG	Bacille Calmette-Guerin
BER	Base excision repair
BGH	Bovine growth hormone
C3H	Cinnamate-3-hydroxylase
C4H	Cinnamate-4-hydroxylase
Cas9	CRISPR associated protein 9
cDNA	Complementary DNA
Cm <sup>R</sup>	Chloramphenicol resistance gene
COX	Cyclooxygenases
CP	Crossing point
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR-RNA
CTP	Cytidine triphosphate
CUR	Curcumin
CURS	Curcumin synthase
CUS	Curcuminoid synthase
DCS	Diketide-CoA synthase
DCW	Dry cell weight
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
dsDNA	Double-strand DNA
DSO	Double-strand origin
DTT	Dithiothreitol

Ery <sup>R</sup>	Erythromycin resistance gene
FLP	Flippase
FnBPA	Fibronectin binding protein A
FRT	FLP recognition target
gDNA	Genomic DNA
GDP	Guanosine 5'-diphosphate
GI	Gastrointestinal tract
GRAS	Generally recognized as safe
gRNA	Guide RNA
GSH	Glutathione
GST	Glutathione S-transferase
HA	Homologous arm
HCT	Hydroxycinnamoyl transferase
HDR	Homology-directed repair
HFA	Hydroferulic acid
HPLC	High performance liquid chromatography
IBD	Inflammatory bowel disease
IFN	Interferon
IL	Interleukin
IMP	Inosine monophosphate
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
IR	Inverted repeat
Kan <sup>R</sup>	Kanamycin resistance gene
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	Knockout
LAB	Lactic Acid Bacteria
LB broth	Luria-Bertani broth
LDH	Lactate dehydrogenase
LL-PAP	Human pancreatitis-associated protein I
LOX	Lipoxygenases
LPS	Lipopolysaccharides
MAM	Microbial anti-inflammatory molecule

MCS	Multiple cloning site
mRNA	Messenger RNA
MRS medium	De Man, Rogosa and Sharpe medium
NHEJ	Non-homologous end joining
NICE	Nisin Control Expression System
OD	Optical density
OmpA	Outer membrane protein A
OMT	O-methyltransferase
ORF	Open Reading frame
O/N	Overnigth
PAL	Phenylalanine ammonia-lyase
PAM	Protospacer adjacent motif
pCMV	Cytomegalovirus promoter
PCN	Plasmid copy number
PCR	Polymerase chain reaction
pDNA	Plasmid DNA
PEG	Polyethylene glycol
PKS	Polyketide synthases
polyA	Polyadenylation sequence from BGH
qPCR	Quantitative PCR
RC	Rolling-circle
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RT	Room temperature
sgRNA	Single guide RNA
SLP	Surface (S)-layer protein
SOEing PCR	Splicing by overlap extension PCR
ssDNA	Single-strand DNA
SSDR	ssDNA recombineering
SSO	Single-strand origin
Strep <sup>R</sup>	Streptomycin resistance gene
Tet <sup>R</sup>	Tetracycline resistance gene

TLR	Toll-like receptor
T <sub>m</sub>	Melting temperature
TNF	Tumor necrosis factor
tracrRNA	<i>Trans</i> -activating crRNA
TTAB	Tetradecyltrimethylammonium bromide
UV	Ultraviolet



# CHAPTER 1

**General introduction and literature review**

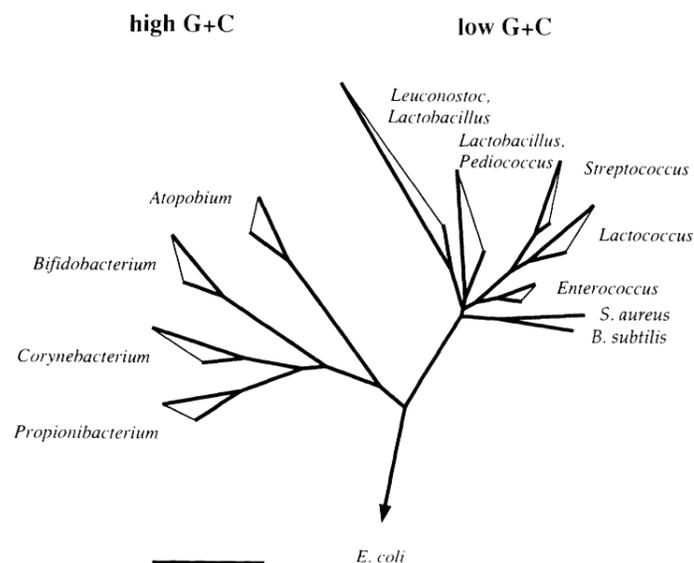
*Somewhere, something incredible is waiting to be known.* – **Carl Sagan**

The main goal of the present thesis is to use Synthetic Biology to develop a highly efficient Lactic Acid Bacteria (LAB)/high copy number plasmid platform for the production of pharmaceutical-grade plasmid DNA and recombinant proteins to be used in Gene Therapy.

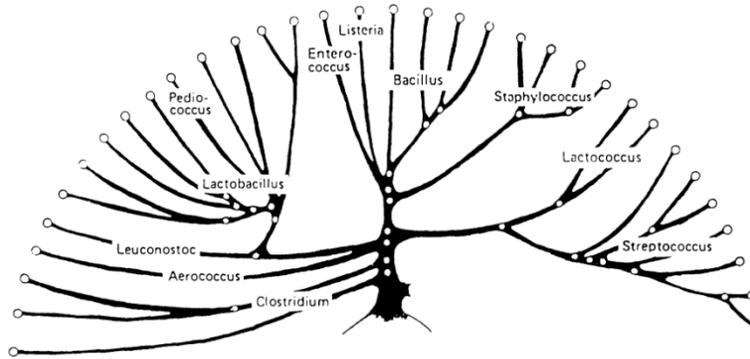
## 1.1. Lactic Acid Bacteria

LAB constitutes a heterogeneous group of Gram-positive bacteria that have in common the fact that they produce lactic acid as an important product from the fermentation of sugars. They are non-spore forming bacteria, chemo-organotrophic and can occur as cocci or rods. LAB can be found in fermented foods and beverages, plants, sewage and also in the respiratory, genital and intestinal tract of humans and animals<sup>1</sup>.

Phylogenetically, based on 16S and 23S sequence data, the Gram-positive bacteria can be divided in two main branches: one with less than 50% of GC content, also called the *Clostridium* branch, and the other with more than 50% GC content, also known as the actinomycetes branch. The typical LAB belong to the low GC content branch and includes *Carnobacterium*, *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Pediococcus* and *Streptococcus* genera (Figure 1.1). The Figure 1.2 shows the low GC content branch in more detail<sup>1</sup>.

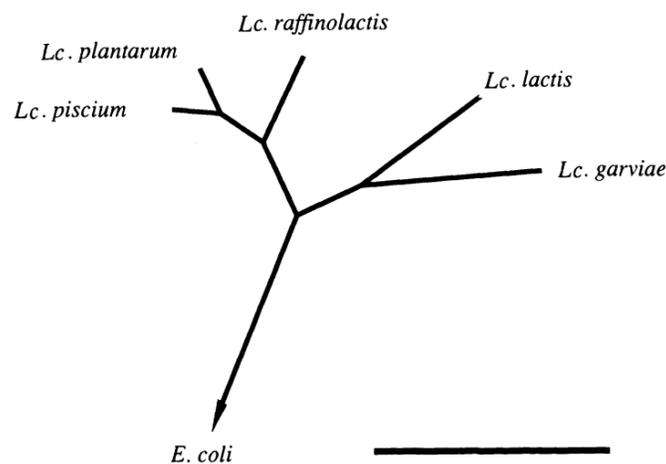


**Figure 1.1.** Gram-positive bacteria phylogenetic tree. The bar represents 10% sequence divergence. Figure from Wood and Holzapel (1995)<sup>1</sup>.



**Figure 1.2.** Phylogenetic tree of the low GC content subdivision of Gram-positive bacteria, as determined by similarity between ribosomal RNAs. Figure from Wood and Holzapfel (1995)<sup>1</sup>.

The genus *Lactococcus* includes the non-motile and mesophilic streptococci that carry a group N antigen, and was first described by Schleifer *et al.* (1985)<sup>2</sup>. The species included in this genus are *L. lactis*, *L. garviae*, *L. raffinolactis*, *L. plantarum* and *L. piscium* (Figure 1.3)<sup>1</sup>. More specifically, *L. lactis* is non-pathogenic and is the best characterized LAB due to its wide applicability in industry, mainly as starters for cheese fermentation. They are also found in plant's surfaces and animal's gastrointestinal tract, and used as laboratory strains as models for research in molecular biology, genetics, physiology and metabolism of LAB<sup>3</sup>. They are also recently used as probiotics, mucosal delivery vehicles of therapeutical molecules and recombinant protein producers, due to its food-grade and Generally Recognized As Safe (GRAS) status, which contributed to raise the industrial relevance of this species<sup>4</sup>. *L. lactis* has two subspecies: *L. lactis* spp. *lactis*, used for making soft cheeses and *L. lactis* spp. *cremoris*, preferred for making hard cheeses<sup>3</sup>.



**Figure 1.3.** Phylogenetic tree representing the relationships between the different species of the genus *Lactococcus*. The bar represents 10% sequence divergence and the tree was built based on 16S rRNA sequence analysis. Figure from Wood and Holzapfel (1995)<sup>1</sup>.

The first LAB genome being sequenced was the one from *L. lactis* spp. *lactis* IL1403, a commonly used laboratory strain. The circular chromosome had 2,365,589 bp, with 2310 open reading frames (ORFs)<sup>3</sup>. Other important sequencing was performed by our laboratory in the scope of the present thesis. The sequenced strain was the plasmid-free *L. lactis* spp. *lactis* LMG19460 strain. The parental strain was isolated from starter cultures of German cheese factories between 1971 and 1979, and then plasmid-cured by exposing it to high sublethal temperatures. These plasmids were important for the ability of the strain to work as a starter culture in the dairy industry, since they coded for carbohydrate and citrate metabolic functions. Although the strain had loosed this competence, it has several other advantages, such as being amenable for gene cloning and DNA transfer, since there are no conflicting plasmids from the same compatibility group. It is also an interesting strain to produce food- and pharmaceutical-grade pDNA to use in DNA vaccines, since it lacks the restriction-modification systems present originally in the endogenous plasmids, allowing for pDNA production in higher quantity and with better quality. This strain also allows a more cost-effective purification process, since there are no copurifying endogenous plasmids. The plasmid-free status also contributes to decrease the metabolic burden associated with plasmid replication, allowing the cell to redirect its efforts towards the plasmid of interest, increasing the plasmid and also the associated recombinant proteins yields. The genome from the *L. lactis* spp. *lactis* LMG19460 strain was quite similar to the one from *L. lactis* spp. *lactis* IL1403, having a slightly smaller genome with 2,260,841 bp and 2171 protein-coding sequences<sup>4</sup>.

## 1.2. *L. lactis* as cell factories for plasmid DNA and recombinant protein production

The GRAS status of *L. lactis*, together with its non-pathogenicity, non-invasiveness and safety, is crucial when the goal is to use them for production of therapeutic recombinant proteins or DNA vaccines<sup>5,6</sup>. Other important aspect is the fact they are able to survive through the gastrointestinal (GI) tract of humans for at least 2-3 days without evoking strong host immune responses<sup>5</sup>, making them ideal to be used as live vectors for mucosal vaccination.

*Escherichia coli* has been the major prokaryotic system used as cell factory, since it leads to the highest protein levels. However, LAB have a major advantage associated with its efficient protein secretion system, unlike *E. coli* for which the most commonly used production strategies are intracellular (periplasm or cytoplasm), thus leading to expensive downstream purification processes. Additionally, unlike *E. coli*, LAB do not produce the highly immunogenic lipopolysaccharides (LPS) that may be co-purified with the proteins of interest and should be removed before administration to humans, making the process more profitable and safe<sup>7</sup>.

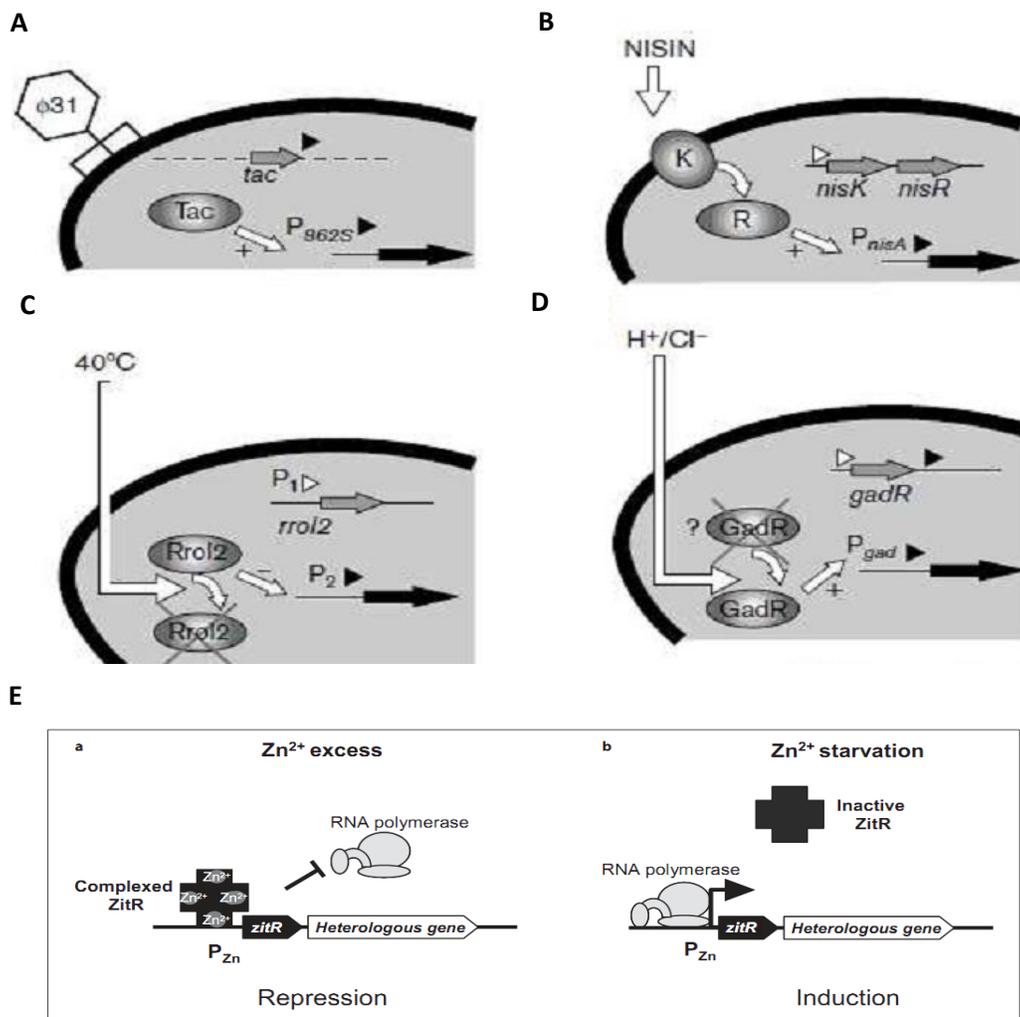
Concerning protein production, *L. lactis* is an attractive alternative to the model Gram-positive bacterium *Bacillus subtilis*, which main problem is that many heterologous proteins secreted by these bacteria are degraded by its complex extracellular proteolytic system<sup>7</sup>. *L. lactis* have only one major protein, Usp45, which is secreted into the medium, simplifying the downstream purification processes and the laboratory strains only have one exported housekeeping protease, HtrA, and a protease-free mutant is already available<sup>7</sup>.

Several heterologous proteins have already been successfully produced in *L. lactis*, such as reporter molecules, bacterial, eukaryotic and viral antigens, interleukins, allergens, virulence factors, bacteriocins and enzymes<sup>8</sup>.

### 1.2.1. Gene expression systems for expression of heterologous proteins

Several gene expression systems based on *L. lactis* had been developed, since it is an organism very easy to handle, with a known metabolism. The production of heterologous proteins has been obtained using lactococcal constitutive or inducible promoters, depending on the goal of the experiment. The constitutive promoters are the ones that express the proteins independently of environmental factors and P21, P23, P32, P44 and P59 were reported as being the most efficient promoters in *L. lactis*. P21, P23 and P59 are considered strong promoters, while P32 and P44 are weak promoters<sup>5</sup>.

But usually, inducible promoters (Figure 1.4) are more useful for the industrial applications, because they only express the proteins when there is some type of stimulation/stress from the environment<sup>5,7</sup>. This type of promoters are particularly important in the case of toxic genes, that are repressed until there is enough biomass and the induction occurs at a defined culture phase/time to produce the proteins of interest<sup>7</sup>.



**Figure 1.4.** Examples of inducible promoters in *L. lactis*. A) promoter induced by phage infection; B) NICE system, induced by nisin; C) promoter induced by temperature; D) promoter induced by pH shift<sup>5</sup>; E) promoter induced by zinc concentration<sup>7</sup>. Figures from D'Souza *et al.* (2012)<sup>5</sup> and Morello *et al.* (2007)<sup>7</sup>.

For example, there are promoters inducible by sugars, such as the high efficiency lactococcal promoter from the *L. lactis* sugar transport and catabolism operon, which is controlled by the autoregulated LacR repressor at the level of transcription initiation. Lactose works as inducer, through the intermediate tagatose-6-phosphate that inactivates the LacR repressor<sup>5</sup>. Since phage infection can induce some promoters (Figure 1.4A), there is an expression system based on this, where the expression of the gene of interest (under the control of the middle promoter of f31, P<sub>8625</sub>) is triggered by induction by bacteriophage f infection. What happens is that the infection results in the synthesis of Tac transcriptional activator that will in turn induce expression from P<sub>8625</sub>, resulting in protein over-production and lysis of the expression host<sup>5</sup>.

There are also promoters induced by temperature (Figure 1.4C), such as P<sub>1</sub> and P<sub>2</sub>, where the thermolabile RroI2 repressor (regulated by P<sub>1</sub>) of P<sub>2</sub> is inactive at 40°C, resulting in induction of gene expression controlled by P<sub>2</sub><sup>5</sup>. During growth, LAB acidifies the medium and there are several promoters that are expressed in response to the pH shift (Figure 1.4D). For example, the induction of the P<sub>gad</sub> by chloride (Cl<sup>-</sup>) or acid (H<sup>+</sup>) is mediated by GadR by an unknown mechanism. Other expression system of the same type is the P<sub>170</sub> system that has the advantage of self inducibility via lactic acid accumulation in the medium during growth<sup>5,7</sup>.

A very well-known inducible expression system is the NICE system (Figure 1.4B). Here, the sensor kinase NisK and the response regulator NisR respond to extracellular nisin, inducing gene expression from P<sub>nisA</sub>. Nisin is a post-translationally modified antimicrobial peptide, produced by several strains of *L. lactis*. This system has been extensively used in the industrial production due to its ease of use, high protein yield, tight control and large scale production. But on the other side, adding nisin increases the cost of the process and further protein purification is needed when the protein has a pharmaceutical purpose<sup>5</sup>.

Finally, a last example is the P<sub>(Zn)</sub>ZitR expression system (Figure 1.4E), which is putatively related with Zn<sup>2+</sup> uptake by the ABC transporter ZitSQP, and it is controlled by ZitR, a MarR family transcriptional regulator. The expression from the promoter P<sub>Zn</sub> is tightly regulated in response to zinc concentration in the medium<sup>5</sup>. Under extracellular zinc excess conditions, ZitR repressor binds to P<sub>Zn</sub> promoter, repressing expression by competing with RNA polymerase binding. In zinc starvation conditions (EDTA addition or growth in a zinc-poor medium), ZitR becomes inactive and allows RNA polymerase binding followed by transcription initiation. This production process is cheap and compatible with large-scale production<sup>7</sup>.

### 1.2.2. DNA vaccines

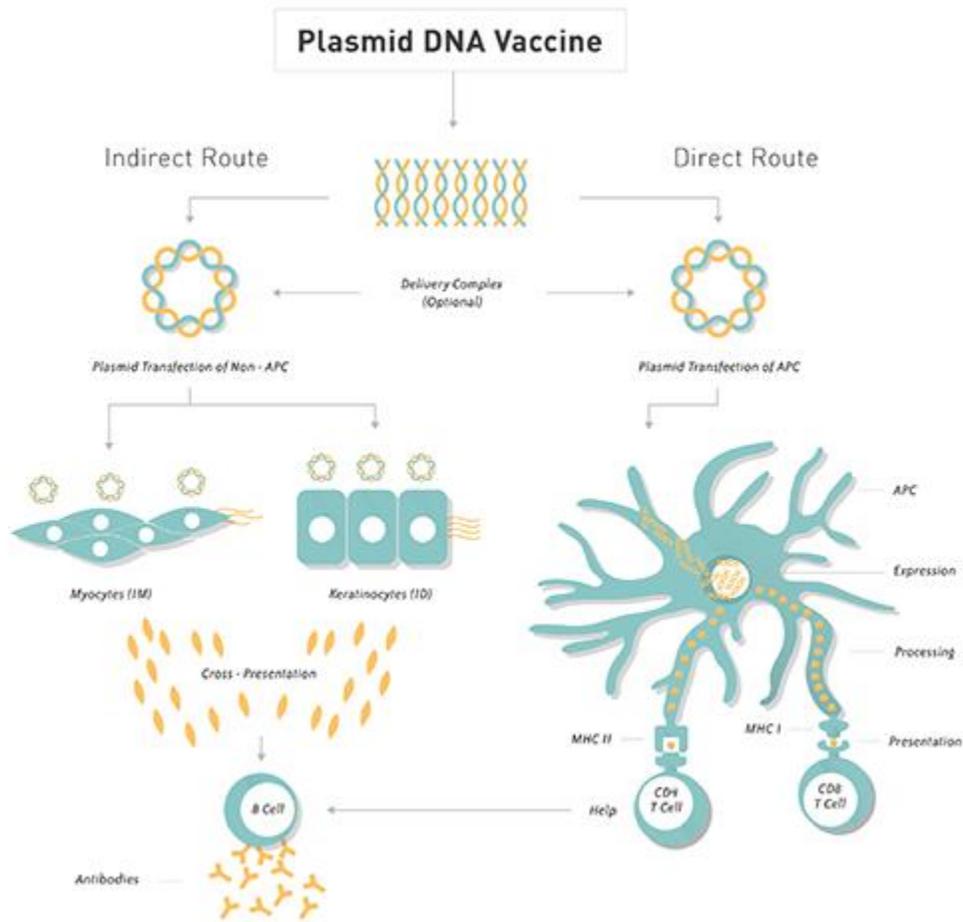
The development of vaccines against several pathogenic agents was one of the most important advances in medicine<sup>9</sup>. A vaccine is any preparation with the goal of inducing immunity against a disease by stimulating antibodies production<sup>10</sup>. There are many types of vaccines, including live weakened vaccines, inactivated, subunit, conjugated and recombinant vaccines<sup>11</sup>. Until now, only the live weakened type is efficient at inducing cellular immunity<sup>12</sup>. But this type of vaccine has some safety issues: it is not safe for people with compromised immune systems and could induce some adverse effects, such as the ones associated with the reversion of the microorganisms to virulence<sup>11</sup>.

The recent improvements in the molecular biology field triggered the development of DNA or gene-based vaccines. This type of vaccines is able to trigger both cytotoxic cellular and antibody (humoral) immune responses against pathogenic microorganisms, without the safety issues of the live attenuated type<sup>13</sup>. DNA vaccines cause prolonged expression of the antigen, generating significant immunological memory, and this antigen is expressed in the host on its natural form (exactly as it is expressed by the pathogen), which improves the immune response<sup>13</sup>. Their main disadvantage is the low immunogenicity registered, and until now this type of vaccines is only being used in the veterinary field<sup>13</sup>. But, several recent studies indicate their promising future potential application in treating or preventing inherited or acquired diseases, such as cancer and certain viral infections<sup>13</sup>.

DNA vaccines are composed of a vector that enables gene delivery into the nucleus of target cells, enabling the production of the desired protein (the antigen) *in situ*. This have an enormous advantage when compared with the recombinant protein therapy, where it is needed to deliver proteins, which are lost in huge amounts until they reach the final destiny. Several characteristics must be accomplished in order to DNA vectors being considered appropriate for clinical application: gene transcription in the cell must be ensured, gene degradation has to be avoided and safety must be guaranteed. From a commercial point of view, to be broadly applied, they should be inexpensive and easy to produce and purify, in high concentrations and quantity<sup>14</sup>.

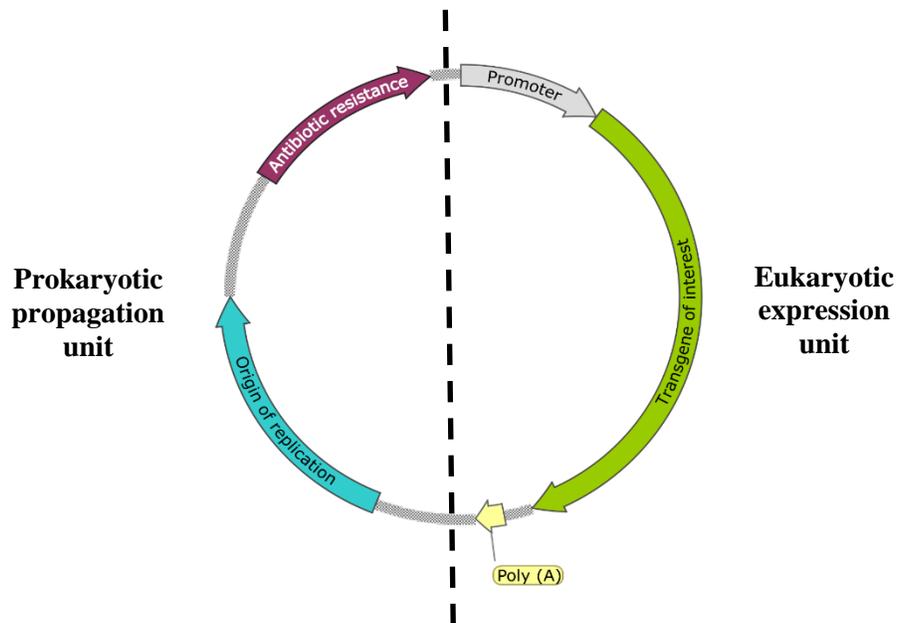
Until now, viral (retroviral, adenoviral and adeno-associated) and non-viral (plasmids) are the two main types of vectors that have been used as vehicles to gene transfer<sup>13</sup>. The viral systems are currently the most effective, in terms of DNA delivery to the nucleus and expression. But, this type raises several safety concerns: immunogenicity, toxicity, reversion to virulence, random integration into the genome, and activation of oncogenes or deactivation of tumour-suppressing genes<sup>13,15</sup>. As a safer alternative, the non-viral systems have been increasing its relevance, although they are still less efficient when compared with the viral systems. The non-viral systems can be non-toxic, non-pathogenic, and with little immune response, allowing the possibility of large dosage and/or repeated administration. Also, it is much easier to modify a non-viral vector to target a specific cell type, than a viral vector<sup>13,14,16</sup>.

Since naked plasmid DNA (pDNA) can transfect cells *in vivo*, this was a viable alternative to the viral vectors<sup>13,17</sup>. Plasmid DNA has several advantages as a gene delivery system, as a vaccine or as therapy, such as its easy manipulation, and can be produced and purified in high quantity. The ideal vaccine should incorporate features such as safety (concerns about integration, tolerance and auto-immunity), appropriated humoral and cellular elicitation (long-lived protection) (Figure 1.5), easy to administer, inexpensive to manufacture and stable in storing procedures (the supercoiled pDNA isoform used in vaccines is temperature stable, eliminating the need for expensive cold chains)<sup>13,18</sup>.



**Figure 1.5.** Overview of the process of plasmid DNA vaccination, focusing on the direct and indirect routes of antigen presentation. Figure from Kemp (2017)<sup>19</sup>.

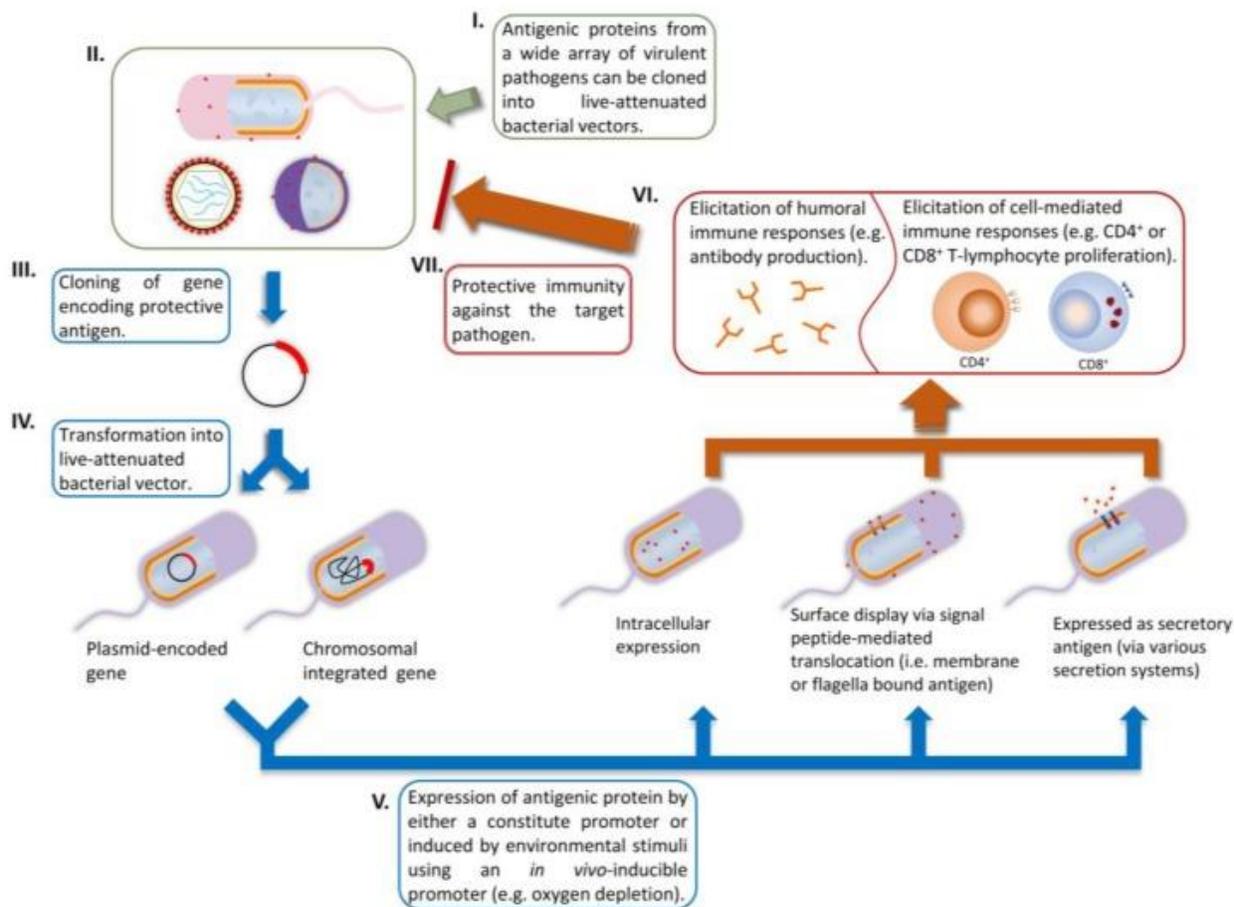
These plasmids are circular double-stranded DNA that must have elements for propagation in a bacterial host (bacterial origin of replication and antibiotic resistance gene) and for transgene expression in the final host (promoter, gene encoding the desired protein and transcription termination sequences)<sup>20</sup> (Figure 1.6).



**Figure 1.6.** Essential elements of a plasmid DNA vector. The propagation unit should contain the antibiotic resistance gene and the bacterial origin of replication, for growth and selection in bacteria. The expression unit should include the gene encoding the protein of interest, under the control of a viral promoter, and with transcription termination sequences (Poly A).

There are several studies using engineered *L. lactis* as a cell factory to produce pDNA to be used as DNA vaccines. The best example goes back to 2007, to a study where the authors constructed a plasmid expressing the HIV-1 gp 120 protein to be used as DNA vaccine. *L. lactis* was used as the cell factory host to produce this DNA vaccine and compared with pDNA produced by *E. coli*. The intramuscular immunization of mice with pDNA produced by *L. lactis* developed gp120 antibody titers comparable to the ones developed in mice after immunization with *E. coli* produced pDNA, showing that *L. lactis* are a suitable host for DNA vaccines production, with the advantage of increased safety, when compared with LPS-producing *E. coli*<sup>21</sup>.

But the majority of the studies with vaccination, instead of naked plasmid DNA, use *L. lactis* as an antigen delivery vector in live mucosal vaccination. The *L. lactis* strain is transformed with a plasmid encoding the antigen of interest and delivered to a mucosa (e.g. gastrointestinal tract). The antigen can be presented in one of three different ways: 1) cytoplasmic, which requires bacterial lysis for pDNA or antigen release and delivery to the target cells, but has the advantage of protecting the pDNA or the antigen from digestion in the gastrointestinal tract; 2) secreted to the gastrointestinal tract, where the antigen contacts directly with the mucosal epithelium and consequently the cells of interest; and 3) cell surface expression, in which the antigen is anchored at the cell membrane, protecting it from proteolytic degradation<sup>22</sup> (Figure 1.7).



**Figure 1.7.** Overview of the live mucosal vaccination process. (I-IV) Cloning of the gene coding for the antigen of interest and bacterial transformation. (V) Antigen expression in one of three different presentation forms: cytoplasmic, surface display or secreted to the extracellular medium. (VI) Humoral and cell-mediated immune response. (VII) Protection against the target pathogen. Figure from Lin *et al.* (2015)<sup>23</sup>.

### 1.2.3. Live mucosal vaccination using *L. lactis* - Mucosal delivery of pDNA, antigens and cytokines

The mucosal vaccination is a strategy where bacterial strains carrying antigen-coding pDNA are administered through host mucosal routes. In the past, the majority of the bacterial strains studied were attenuated pathogenic such as *Listeria monocytogenes*, *Salmonella typhi* and *Shigella flexneri*, due to its ability to naturally infect the mucosal surfaces, but they had a high risk associated with possible reversion to virulence and therefore they are considered not entirely safe for use in humans, especially in children and immunosuppressed patients. A food-grade alternative was imperative and the research turned to LAB as safe mucosal DNA delivery vehicles. More specifically, *Lactococcus lactis* had been extensively studied, due to its GRAS and non-pathogenic status, for antigen and cytokines production and delivery, as well as a vehicle for oral delivery of DNA vaccines<sup>24</sup>.

The most recent studies using *L. lactis* for mucosal vaccination are summarized in Table 1.1. The plasmids used to express the intended antigen in this species are also presented and can be grouped in three main replicons that will be

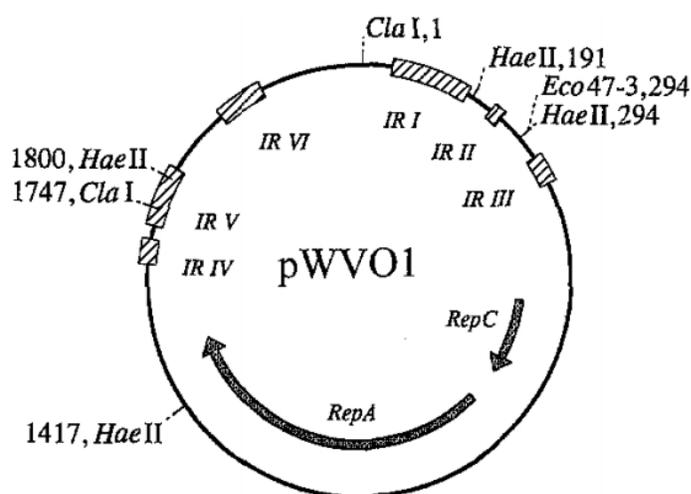
described in further detail: the pWV01- and pSH71-based replicons with a rolling-circle replication mechanism and the pAM $\beta$ 1-based replicon, which has a theta-type mode of replication.

**Table 1.1** Plasmids used as antigen-expressing vectors in *L. lactis* host, for mucosal vaccination purposes against several diseases.

Disease	Antigen/Cytokine	Plasmid	Replicon	Reference
Tuberculosis	<i>Mycobacterium tuberculosis</i> ESAT-6	pValac	pWV01	24,25
Inflammatory bowel disease	IL-4	pValac	pWV01	26
Intestinal mucositis	Human pancreatitis-associated protein I	pSEC	pWV01	27
Inflammatory bowel disease	<i>Mycobacterium</i> Hsp65	pSEC	pWV01	28
<i>Shigella dysenteriae</i> diarrhea	Outer membrane protein A of <i>S. dysenteriae</i>	pSEC	pWV01	29
Fish edwardsiellosis	OmpA and flagellar hook protein D antigens from <i>Edwardsiella tarda</i>	pCYT	pWV01	30
<i>Streptococcus iniae</i> infection	<i>S. iniae</i> M-like protein antigen	pCYT	pWV01	31
<i>Helicobacter pylori</i> vaccine adjuvant	Heat-labile enterotoxin B subunit	pNZ8149	pSH71	32
Bird flu H5N1	Hemagglutinin from an Avian H5N1 isolate	pNZ8150	pSH71	33
Avian coccidiosis ( <i>Eimeria tenella</i> infection)	<i>E. tenella</i> 3-1E protein	pTX8048 (pNZ8048 derived)	pSH71	34,35
Diabetes	Single-chain insulin (SCI-59) analog	pMRF5018 (pNZ8149 derived) and pMRF5019 (pNZ8048 derived)	pSH71	36
<i>Yersinia pseudotuberculosis</i> infection	LcrV from <i>Y. pseudotuberculosis</i>	pMEC237 (pNZYR derived)	pSH71	37
Thrombosis	Subtilisin QK-2	pRF01 (pNZ8149 derived ) and pRF03 (pNZ8048 derived)	pSH71	38
Shiga Toxin 1 B Subunit, from shiga toxin-producing bacteria, like enterohemorrhagic <i>E. coli</i> and <i>Shigella dysenteriae</i>	Albumin-binding domain variants	pNZ8148	pSH71	39
Inflammatory diseases, autoimmune diseases and cancer	Single-chain variable fragment antibody against mouse interleukin 6	pNZ8148	pSH71	40
DSS-induced colitis	Porcine insulin-like growth factor I	pNZ8148	pSH71	41
House dust mite allergy	House dust mite allergen	pNZ8148	pSH71	42
Human colon carcinoma	Bioactive kisspeptin (KiSS1)	pNZ401	pSH71	43
Colitis	cDNA of Microbial Anti-inflammatory Molecule (MAM) from <i>Faecalibacterium prausnitzii</i>	pILMAM (pIL253 derived)	pAM $\beta$ 1	44
Multiple sclerosis	Myelin peptides	pILMAM (pIL253 derived)	pAM $\beta$ 1	45
<i>Campylobacter jejuni</i> infection	<i>C. jejuni</i> antigens	pIL253	pAM $\beta$ 1	46
HPV-16-induced tumors	HPV-16-E7 antigen	pOri23	pAM $\beta$ 1	47
Parasitic worms	<i>Bacillus thuringiensis</i> crystal protein Cry5B	pTRK593 (pTRKH2 derived)	pAM $\beta$ 1	48

### 1.2.3.1. Plasmids with a pWV01-based origin of replication

The pWV01-based origin of replication originally comes from the 2.3 kb pWV01 cryptic plasmid isolated from *L. lactis* subsp. *cremoris* Wg2. It has a rolling-circle (RC) mode of replication and the replication initiator protein is the 27 kDa RepA. The replication of the RC plasmids proceeds via single-stranded intermediates and its conversion to the double-stranded form initiates at the single-strand origin (SSO), by a RNA polymerase-independent process. There are for groups of RC plasmids, distinguished based on similarities in the leading strand replication region. The pWV01 plasmid belongs to the pE194/pLS1 group, which have a double-strand origin (DSO) in the inverted repeat (IR) III, a SSO in the IR I and IR II, a *repA* gene that codes for the replication protein that introduces a single-strand nick in the plus origin, a *repC* gene coding for a 6 kDa protein involved in negative copy number control and a *repC/repA* terminator in the IR IV (Figure 1.8)<sup>49</sup>. There is other ORF, called ORFD, putatively involved in copy number regulation<sup>50</sup>. Due to the presence of several copy number control mechanisms, usually the pWV01-based vectors have a low plasmid copy number with around 5-10 copies of plasmid per cell<sup>51</sup>.



**Figure 1.8.** Main features of the plasmid pWV01. Coding regions are indicated by arrows and the inverted repeats (IR) are indicated by boxes. Relevant restriction sites are indicated. Figure from Seegers *et al.* (1995)<sup>49</sup>.

The pWV01 origin of replication has been used to create more than 20 different cloning vectors. pGK12 is the most widely used pWV01-based plasmid, replicating in a wide range of hosts, including *B. subtilis*, *E. coli*, *Lactococcus lactis* and several *Lactobacillus* species, but it has the big disadvantage of being unstable and having a low copy number in these species. A high copy number derivative (pBAV1K-T5) was engineered by deleting repeats IV to VI and the ORFD, leading to 357 and 251 copies per cell in different *E. coli* strains, compared to the pGK12 that had only 60 copies per *E. coli* cell<sup>50</sup>.

In brief, pWV01 is considered a broad-host-range plasmid, since it is able to replicate both in Gram-positive and Gram-negative bacteria. It is widely used as a cloning vector<sup>52,53</sup> and there are several recent examples of pWV01-based plasmids being used for DNA and mucosal vaccination. A first example was the work of Pereira *et al.* (2017)<sup>25</sup> with *Lactococcus lactis* carrying a DNA vaccine for tuberculosis, working as a live mucosal DNA vaccine. The plasmid

pValac coded for the ESAT-6 antigen, which is the 6 kDa early-secreted antigenic target, a protein present in pathogenic *Mycobacterium tuberculosis*, but absent in the Bacille Calmette-Guerin (BCG) vaccine. The *L. lactis* strain used expressed the *Staphylococcus aureus* fibronectin-binding protein A (FnBPA+), turning it into an invasive strain with improved pDNA transfer abilities to eukaryotic cells, place where the antigen will be expressed. This strain was shown to increase IL-17, IL-6, IFN- $\gamma$  and TNF- $\alpha$  cytokine secretion by the host and consequently boost the BCG vaccine immune response in mice, after oral immunization<sup>25</sup>.

*L. lactis* carrying the pValac expression vector coding for IL-4 was shown to reduce chemically-induced intestinal inflammation in inflammatory bowel diseases (IBD), such as Crohn's disease. This effect was due to an increase in the levels of IL-10-producing cells. Besides IL-4, the plasmid also contained a DNA nuclear targeting sequence. An ELISA assay showed that 48 h post-transfection the IL-4 was only detected in the supernatant of pValac::dts::IL-4 transfected cells ( $5.25 \pm 0.1$  ng/mL), when compared with non-transfected cells or transfected with the empty vector<sup>26</sup>.

Still in 2017, Carvalho *et al.*<sup>27</sup> studied *L. lactis* NZ9000 able to secrete the human pancreatitis-associated protein I (LL-PAP) as an efficient tool for the prevention of intestinal mucositis, a gastrointestinal inflammatory condition resulting from the use of chemotherapeutic drugs. The work has shown that was the joint action of the strain and of the expressed LL-PAP that led to the protective and anti-inflammatory effect against intestinal mucositis in mice. The plasmid with the ORF coding for the LL-PAP harbored a Nisin Control Expression System (NICE) and was a pSEC plasmid, meaning that its replicon was a pWV01<sup>27</sup>.

The IBD in mice was shown to be prevented by *Mycobacterium* Hsp65-producing *L. lactis*, through IL-10 and TLR2-dependent pathways<sup>28</sup>. Hsp65 is a heat shock protein produced in high quantities in inflammation sites, by intestinal immune cells and commensal microbiota. The Hsp65 coding plasmid was again a pSEC (pWV01-based replicon)<sup>28</sup>.

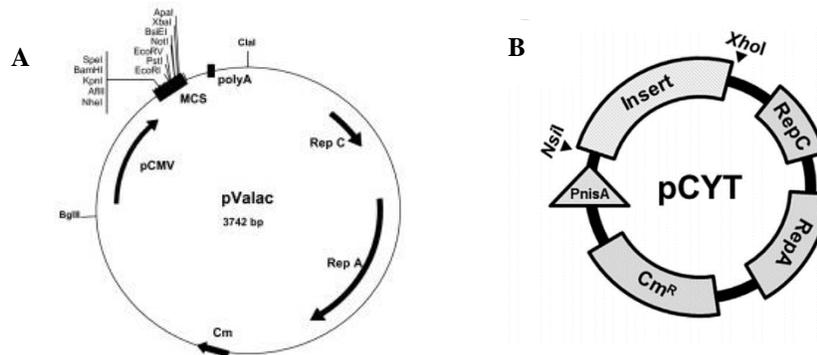
Both oral and nasal administration in mice of *L. lactis* expressing outer membrane protein A (OmpA) of *Shigella dysenteriae* type I were evaluated concerning its immunogenic potential. The oral route elicited higher systemic and mucosal immune response. No vaccine against this bacterial diarrhea and dysentery is currently available and many *Shigella* strains are resistant to multiple antibiotics, therefore the development of an effective and safe vaccine is of extreme priority. The *ompA* gene was cloned in a pSEC plasmid with a pWV01-based origin of replication<sup>29</sup>.

A DNA vaccine plasmid reporter, called pPERDBY and also derived from pSEC, with a pWV01 origin of replication, was developed specifically for DNA vaccination using *L. lactis*<sup>54</sup>.

The OmpA and flagellar hook protein D antigens from *Edwardsiella tarda* expressed by a recombinant *L. lactis* BFE920 contributed to an effective protection of the fish *Paralichthys olivaceus* from edwardsiellosis, a major fish disease. The recombinant *L. lactis* strain were administered to the fish as a feed vaccine and 82.5% of the immunized fish survived after being challenged with the pathogen, against 10% in the control group. The backbone plasmid was pCYT, a pWV01-based replicon with RepA and RepC genes. The plasmid also contained a nisin induced promoter and a chloramphenicol resistance gene<sup>30</sup>.

Other *P. olivaceus* fish feed vaccine was developed against *Streptococcus iniae* infection. The vaccine was constituted by *L. lactis* BFE920 expressing *S. iniae* M-like protein antigen, which contributed to an 82% relative percent survival rate of the vaccinated fish. The plasmid construction was identical to the one described for the previous work<sup>31</sup>.

In brief, the main pWV01-based replicons used in the most recent works for DNA and mucosal vaccination were pValac (pDNA is delivered to the eukaryotic host cells after bacteria internalization and the coded antigen is expressed by the host cells), pSEC (able to secrete the coding antigen proteins) and pCYT (express antigen proteins that remain at the cytoplasm). The pValac plasmid generally contains the replication origin genes (*repC* and *repA*), chloramphenicol resistance gene ( $Cm^R$ ), the cytomegalovirus promoter (pCMV) for expression in eukaryotic hosts, a multiple cloning site (MCS) and a BGH polyadenylation region (polyA) (Figure 1.9A)<sup>55</sup>. The pCYT plasmid is an integrative plasmid for cytosol expression, while pSEC is an integrative plasmid for secretion expression. Both have a *PnisA* that is an inducible promoter, followed by the insert of interest. The origin of replication and antibiotic resistance gene are the same as in pValac (Figure 1.9B)<sup>30</sup>. The additional feature present in pSEC plasmid is a signal peptide of *usp45* (*SPusp45*) from *L. lactis* located immediately downstream the *PnisA* promoter.



**Figure 1.9.** Structure of A) pValac<sup>55</sup> plasmid and B) pCYT plasmid<sup>36</sup>. Figures from Guimarães *et al.* (2009)<sup>55</sup> and Beck *et al.* (2017)<sup>30</sup>.

### 1.2.3.2. Plasmids with a pSH71-based origin of replication

The pSH71 replicon from the 2.1 kb *L. lactis* subsp. *lactis* 712 plasmid is closely related to the pWV01 replicon, also having the *repA* and *repC* genes and replicating by a rolling-circle mode. It is also able to replicate both in Gram-positive and Gram-negative bacteria. Several cloning plasmids were engineered from the pSH71 replicon, such as the high copy number pCK1, the high copy number pNZ12 and the low copy number pNZ121<sup>56</sup>. The plasmid copy number for the parental pSH71 was estimated as being 200 copies per cell<sup>57</sup>.

The pWV01 is very similar to pSH71, differing only in a few nucleotides and in the absence of a direct repeat. Like pWV01, the plasmid pSH71 has a high segregational stability in *L. lactis* and other Gram-positive hosts, with a plasmid loss being less than  $10^{-5}$  per generation, but a low stability in *E. coli* (loss of more than  $10^{-2}$  plasmids per generation)<sup>51</sup>.

One example of an application of a pSH71-derived replicon in *L. lactis* mucosal vaccination was the use of *L. lactis* engineered for producing of heat-labile enterotoxin B (LtB) subunit as adjuvant in oral vaccines formulation. When co-administered to mice with a *H. pylori* vaccine candidate expressing Lpp20 antigen, it significantly enhanced the

mucosal antibody responses against *H. pylori*. The backbone plasmid was a pNZ8149 with a *nisA* promoter, a signal sequence of *usp45* gene, the *ltb* gene and the replication origin (coding for the RepA and RepC proteins)<sup>32</sup>.

A recombinant *L. lactis* strain was engineered to express Hemagglutinin from an Avian H5N1 isolate, in order to be used for antigen production and delivery to chicken and mice mucosal surfaces, as a vaccine against bird flu. The *H5* gene was cloned under the control of the nisin-controlled gene expression system in the pNZ8150 plasmid, a pSH71-based replicon. The preliminary results from the animal trials showed that the oral delivery of live *L. lactis* cells producing H5 protein was able to elicit an immune response. The protein was engineered in order to be produced intracellularly, and not secreted, to protect the antigen from the passage through the stomach<sup>33</sup>.

*L. lactis* expressing surface anchored *Eimeria tenella* 3-1E protein induced protective immunity against infection by *E. tenella* (avian coccidiosis), after being administered orally to chicken. The original plasmid used in this study was the pTX8048, which derives from pNZ8048, a pSH71-based replicon<sup>34,35</sup>.

A single-chain insulin (SCI-59) analog was expressed in *L. lactis* for diabetes therapy, to overcome the problem of the low bioavailability of orally administered insulin. Both secretory and surface displayed protein in *L. lactis* were tested, being the later considered more stable. Since LAB are able to survive the passage through the stomach, it makes them the ideal vehicle for mucosal delivery of SCI-59. The plasmids used in this study were pMRF5018 and pMRF5019, both having the *SCI-59* gene and the *Usp45* signal peptide to allow its secretion, and the later having the gene that coded for three LysM repeats, allowing the SCI-59 surface display. All the construction were under the control of the NICE system. The pMRF5018 was a plasmid derived from pNZ8149, while the pMRF5019 were derived from pNZ8048, being both pSH71-based replicons, with *repA* and *repC* genes<sup>36</sup>.

Mice vaccinated with an LcrV-secreting strain of *L. lactis* evidenced a protective immune response against *Yersinia pseudotuberculosis* infection. LcrV is a virulence determinant of pathogenic *Yersinia* and is an essential component of a type III secretion system. When used in the vaccine, it stimulates both humoral and cellular immune responses, providing protection against *Yersinia* infections. The plasmid containing the PCR generated *lcrV* gene from *Y. pseudotuberculosis* fused to the secretion signal of the lactococcal *Usp45* protein was the pMEC237, a pNZYR-derived plasmid that itself derives from the pSH71 plasmid<sup>37</sup>.

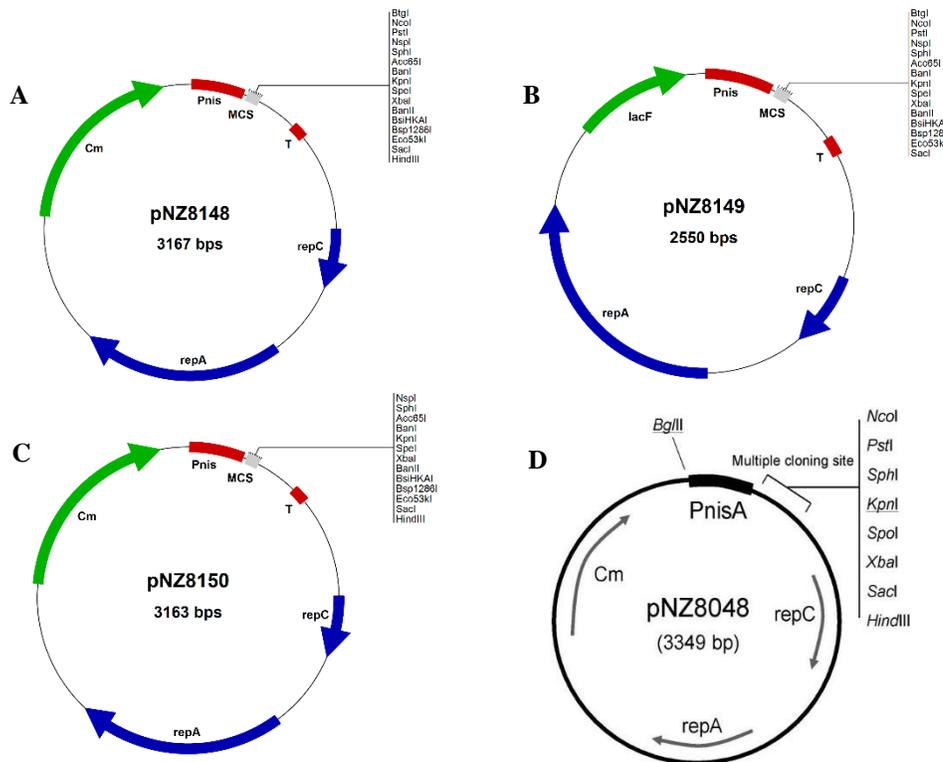
Subtilisin QK-2 is a type of effective thrombolytic reagent that was expressed in *L. lactis* as a secretory protein or as surface-display protein. This *L. lactis* produced surface-displayed antigen was shown to be more efficient in preventing and controlling thrombosis diseases. The plasmids pRF01 and pRF03 were derived from pNZ8149 and pNZ8048, respectively. Both are pSH71-derived plasmids<sup>38</sup>.

A recombinant *L. lactis* was developed to display albumin-binding domain variants in its surface, as a mucosal vaccine against Shiga Toxin 1 B Subunit, from shiga toxin-producing bacteria, like enterohemorrhagic *E. coli* and *Shigella dysenteriae*. The backbone plasmid was the pSH71-derived pNZ8148, containing the chloramphenicol resistance gene and the *nisA* promoter for a nisin-controlled expression<sup>39</sup>. Other three recent studies of *L. lactis* mucosal vaccines also used the pNZ8148 as a backbone plasmid. One of the studies used *L. lactis* engineered to secrete a single-chain variable fragment antibody against mouse interleukin 6 (IL-6), which can be applied to treat inflammatory diseases,

autoimmune diseases and cancer<sup>40</sup>. The same plasmid was used to express porcine insulin-like growth factor I in *L. lactis*, which was shown to attenuate the symptoms and progression of DSS-induced colitis in mice<sup>41</sup>. A third study was performed in *L. lactis* expressing house dust mite allergen. This strain was used as an oral vaccine for prevention and treatment of house dust mite allergy and the plasmid backbone was also the pNZ8148<sup>42</sup>.

Other study used *L. lactis* able to secrete a bioactive kisspeptin (KiSS1) that was shown to inhibit the proliferation and migration of human colon carcinoma HT-29 cells. The maximum detected KiSS1 levels were 19.7 and 27.9 µg/ml for  $5 \times 10^8$  and  $5 \times 10^9$  CFU/ml cell densities of *L. lactis* NZ9000 harboring the plasmid pNZ401-kiss1, respectively<sup>43</sup>.

In brief, the main pSH71-derived backbone plasmids used in mucosal vaccination studies with *L. lactis* were pNZ8148, pNZ8149, pNZ8150 and pNZ8048 (Figure 1.10). The pNZ8148 plasmid contains the nisin A promoter (PnisA) followed by a MCS that includes the NcoI restriction site for translational fusions at the ATG. After the MCS it contains a terminator. This broad-host-range vector has resistance to chloramphenicol and a replication origin with the replication genes A and C (*repA* and *repC*). The only difference in pNZ8149 is that it has the *L. lactis lacF* gene as food-grade selection marker, instead of the antibiotic resistance gene. For transformants selection, this vector needs a host strain with the lactose operon without the *lacF* gene. The pNZ8150 is very similar to pNZ8148, with the difference that the translation fusions are performed at the ScaI site, precisely at the ATG<sup>58</sup>. pNZ8048 is identical to pNZ8148, except in the existence of additional 60 bp that corresponds to residual *B. subtilis* DNA and were deleted during pNZ8148 construction<sup>59</sup>.



**Figure 1.10.** Map of A) pNZ8148, B) pNZ8149, C) pNZ8150<sup>58</sup> and D) pNZ8048 plasmids<sup>59</sup>. Figures from MoBiTec (2017)<sup>58</sup> and Frelet-Barrand *et al.* (2010)<sup>59</sup>.

### 1.2.3.3. Plasmids with a pAM $\beta$ 1-based origin of replication

The pAM $\beta$ 1 plasmid was isolated from *Enterococcus faecalis* and replicates by a unidirectional theta mechanism. This replicon has a high structural stability that allow cloning large fragments and has a wide host range within Gram-positive bacteria<sup>60</sup>. The highest number of copies reported for the pAM $\beta$ 1 replicon was around 100 copies per *B. subtilis* cell, after inactivation of the transcriptional repressor *copF* that represses the plasmid-encoded replication initiation protein RepE<sup>60</sup>. Upstream the *repE* gene, the replicon has an ORF coding for the RepD protein, which have an unknown function, not essential for the plasmid replication<sup>60</sup>. Both *repD* and *repE* genes are under the control of the P<sub>DE</sub> promoter and they have their transcription, and ultimately the plasmid copy number, tightly regulated by two different systems: the transcriptional repressor protein CopF and an antisense RNA-mediated transcription attenuation system<sup>61</sup>.

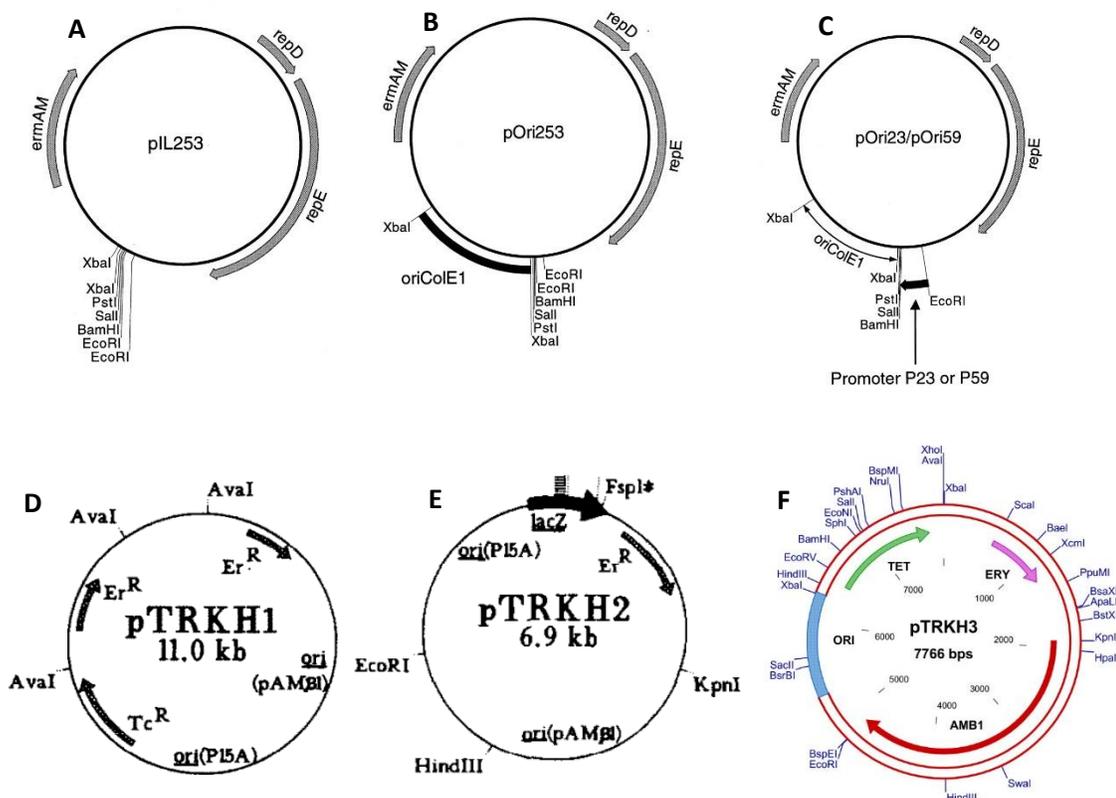
Several *L. lactis* mucosal vaccination studies were performed used pAM $\beta$ 1-derived backbone plasmids. *L. lactis* carrying a plasmid with the cDNA of Microbial Anti-inflammatory Molecule (MAM) from *Faecalibacterium prausnitzii* had shown anti-inflammatory properties in mice colitis models. The plasmid used was pILMAM, which derives from pIL253, a pAM $\beta$ 1-based plasmid<sup>44</sup>. Other study that used the same plasmid was performed with *L. lactis* expressing myelin peptides and showed that its oral administration could induce oral tolerance for the treatment of multiple sclerosis<sup>45</sup>. A third study using the pIL253 as the backbone plasmid involved *L. lactis* expressing *Campylobacter jejuni* antigens anchored to its cell wall<sup>46</sup>.

Since pAM $\beta$ 1-derived plasmids are unable to replicate in *E. coli*, the research has turned to the creation of pAM $\beta$ 1-based shuttle vectors, by cloning other origins of replication in the same plasmid. One example is the pOri253 plasmid, derived from pIL253, a 5.2 kb LAB cloning vector, by addition of the Gram-negative origin of replication *ColEI*. The pOri253 plasmid has the erythromycin resistance gene that confers resistance both in Gram-positive and Gram-negative bacteria<sup>62</sup>. A recent study using *L. lactis* for DNA delivery to eukaryotic cells used this plasmid as backbone and the results shown that CHO cells were able to express the GFP reporter protein as early as 12h after bacterial inoculation<sup>63</sup>. A different study, with recombinant *L. lactis* expressing a cell-surface anchored FnBPA from *Staphylococcus aureus* (LL-FnBPA) and at the same time the HPV-16-E7 antigen (LL-E7), showed improved protection of the vaccinated mice when challenged with HPV-16-induced tumors. The LL-E7 antigen was cloned in a pGK plasmid (pWV01-based plasmid), while LL-FnBPA was cloned in pOri23 plasmid, derived from the pAM $\beta$ 1 replicon<sup>47</sup>.

The pTRK plasmid family is other collection of plasmids engineered from the pAM $\beta$ 1 plasmid, by addition of a second origin of replication to pIL253, which in this case was p15A (low copy number) from *E. coli*<sup>64,65</sup>. The presence of both origins of replication turned this plasmid into a shuttle vector, allowing its replication both in Gram-positive and Gram-negative hosts. An example of an application of this vector family was in a study where the *Bacillus thuringiensis* crystal protein Cry5B was expressed intra and extracellularly in *L. lactis*, in order to be used as an anthelmintic. The gene coding for the Cry5B protein was cloned into pTRK593 plasmid, a pTRKH2-based vector<sup>48</sup>.

In brief, the main pAM $\beta$ 1-based plasmids used in mucosal vaccination using *L. lactis* as a host were derived from pIL253 (Figure 1.11A). This vector has an erythromycin resistance gene and a pAM $\beta$ 1 origin of replication, with *repD*

and *repE* genes that replicates in Gram-positive hosts. It is able to replicate at high copy number (45-85 copies per cell) and is able to stably maintain large DNA inserts<sup>66</sup>. From the pIL253 backbone, three main shuttle vectors, able to replicate in Gram-positive and Gram-negative hosts, were engineered. The addition of the Gram-negative replicon is of utmost importance, since it allows the molecular cloning techniques to be performed in a simpler and faster way using the well-studied model organism *E. coli*. The Gram-positive bacteria can then be transformed with the final plasmid, avoiding more time- and resource-consuming techniques that are necessary for applying molecular biology techniques with this Gram-positive hosts. The pOri253 plasmid (Figure 1.11B) was constructed by insertion of the *ColE1* origin of replication from Gram-negative bacteria in the pIL253 MCS. After the insertion of the *P23* lactococcal promoter, the new expression vector was called pOri23 (Figure 1.11C)<sup>66</sup>. The pTRK family of shuttle vectors also derived from pIL253 after insertion of a Gram-negative p15A origin of replication, resulting in the pTRKH1 vector (11 kb) (Figure 1.11D), which also harbored the tetracycline resistance gene for selection in Gram-positive hosts (erythromycin works in both Gram-negative and Gram-positive hosts). The pTRKH1 plasmid copy number in Gram-positive bacteria was similar to the parental plasmid and in *E. coli* it had a medium copy number (30-40 plasmids per cell). pTRKH2 plasmid (6.9 kb) resulted from the improvement of the pTRKH1 vector, after incorporation of a *lacZ* cassette and removal of non-essential Gram-negatives sequences (Figure 1.11E). Also from pTRKH1, the pTRKH3 (7.8 kb) (Figure 1.11F) (used in the present thesis) was constructed after removal of non-essential sequences, remaining with erythromycin and tetracycline resistance genes and both origins of replication (pAM $\beta$ 1 for Gram-positive and p15A for Gram-negative hosts)<sup>65</sup>.



**Figure 1.11.** Map of A) pIL253, B) pOri253, C) pOri23<sup>66</sup>, D) pTRKH1, E) pTRKH2 and F) pTRKH3<sup>65</sup> plasmids. Figures from Que *et al.* (2000)<sup>66</sup> and O' Sullivan and Klaenhammer (1993)<sup>65</sup>.

### 1.3. Research aims

The implementation of a highly efficient LAB/high copy number plasmid platform faces two major issues that the present PhD thesis aims to solve: 1) the low copy number of the available LAB plasmids; 2) the presence of a high level of endonuclease activity in the majority of the strains.

The first challenge has two main specific goals to be settled at plasmid engineering level: 1.1) increase the plasmid copy number by site-directed mutagenesis of the Shine-Dalgarno sequence of a protein involved in replication; and 1.2) use Gibson Assembly to engineer a minimal plasmid with improved parts, to increase the plasmid copy number.

The second challenge corresponds to the strain engineering level and also has two main specific goals: 2.1) perform genomic analysis using Optflux software to predict the best genes to knockout; and 2.2) minimize non-specific digestion and improve yield and quality of pDNA by using a CRISPR/Cas9 system to knockout the endonuclease gene.

A final integrative cell factory challenge is to implement a highly efficient LAB/high copy number plasmid platform using a curcumin synthetic pathway.

### 1.4. Thesis outline

The thesis was structured in seven chapters that cover the research aims described above:

- In **Chapter 1** is presented a general introduction to the thesis work and a literature review of the most relevant topics in the context of the present thesis. This chapter starts with an overview about LAB, followed by a more specific analysis of *L. lactis* as cell factories for plasmid DNA and recombinant protein production, with a focus in its use for DNA vaccines and mucosal vaccination. The plasmids used for antigens and cytokines production in mucosal vaccination are discussed in more detail, as well as the gene expression systems for production of heterologous proteins. The research aims and the structure of the thesis are also outlined in the present chapter.

The different sections of **Experimental Results** are presented from **Chapter 2** to **Chapter 5**. In each chapter a brief Introduction, Materials and methods, Results and discussion, and Conclusions are given.

- The increase of the plasmid copy number of the plasmid pTRKH3 in *L. lactis* by the modification of the *repDE* ribosome binding site is addressed in the **Chapter 2**. The site-directed mutagenesis strategy is explained and the plasmid copy number determination by real-time quantitative PCR results are presented. In order to further understand the PCN results, the relative translation initiation rates of each plasmid were predicted and the mRNA from different transcriptional repressors were quantified.
- In **Chapter 3** the goal is to use Gibson Assembly to engineer a minimal plasmid with improved parts, to increase the plasmid copy number. A brief overview about Synthetic Biology is given in this chapter, followed by the plasmid engineering and assembly steps. The success of the results are measured in terms of PCN, determined by quantitative real-time PCR.

- **Chapter 4** concerns the strain engineering level, being addressed not only the Optflux software simulations of the best candidate genes to knockout in order to increase both the biomass and the pDNA production, but also the CRISPR/Cas9 system employed to knockout the endonuclease (*nth*) gene in *L. lactis* LMG 19460. It is also presented an *in silico* analysis of the *L. lactis* LMG19460 genome, recently sequenced by our laboratory in the context of the present thesis.
- The last experimental results chapter is the **Chapter 5** that describes the construction of a highly efficient LAB/high copy number plasmid platform used for curcumin production. A brief review about the importance of curcumin is presented, followed by the construction of the synthetic pathway proposed in this work. The curcumin production, extraction and quantification by HPLC is also reported.
- **Chapter 6** contains the concluding remarks and perspectives for further research in this topic.
- Finally, the **Chapter 7** compiles the key references used in this work.

## **CHAPTER 2**

**Increasing plasmid copy number of pTRKH3 in *Lactococcus lactis* by the modification of the *repDE* ribosome binding site**

*All our dreams can come true if we have the courage to pursue them.* – **Walt Disney**

## 2.1. Abstract

The increasing demanding for safer alternatives in the DNA vaccination field was the motivation for using the Generally Recognized As Safe *Lactococcus lactis* as plasmid producer for therapeutic purposes, instead of the conventional but lipopolysaccharide-producing *Escherichia coli*. In order to establish a plasmid purification procedure to be industrially profitable, one of the major requirements is to have a high-copy number plasmid, but the highest number of copies already reported in gram-positive bacteria for the pAM $\beta$ 1 replicon was around 100 copies per *Bacillus subtilis* cell. Although the rolling circle-based plasmids are able to achieve higher plasmid copy numbers, the pAM $\beta$ 1 theta-based replicon has the advantage of being highly structurally stable for cloning purposes and having a wide host range, turning it into a more appealing template for the present work. The purpose of this work was to engineer by site-directed mutagenesis the *repDE* Ribosome Binding Site (RBS) of the pTRKH3 plasmid, whose replication depends on RepE protein, to increase the plasmid copy number (PCN) in *L. lactis* LMG19460 cells.

Several candidates were constructed and the pTRKH3-b mutant was the most promising candidate, achieving 215 copies of plasmid per chromosome after reaching the late exponential/early stationary phase of growth, which corresponds to a 3.5 fold increase when compared to the non-modified pTRKH3 (62 copies per chromosome). We suggest that the highest PCN obtained with pTRKH3-b is probably due to a combination of a stronger RBS sequence, an mRNA secondary structure that promotes the ribosome binding and an ideal intermediate amount of transcriptional repressors (CopF and CT-RNA) present in *L. lactis* cells. Also the presence of a duplicated region that added an additional RBS sequence and seven new alternative start codons should be contributing for the high PCN value of the pTRKH3-b mutant.

## 2.2. Introduction

A key point for the development of DNA vaccines is the production of pharmaceutical-grade plasmid DNA in high quantities and, at the same time, in a cost-effective manner. The safe status along with the ability of elicit both humoral and cellular immune responses, gives DNA vaccines several at least theoretical advantages over other vaccine types, and make them efficient tools to prevent infectious diseases and as therapeutic tools against autoimmune diseases and cancers<sup>67</sup>.

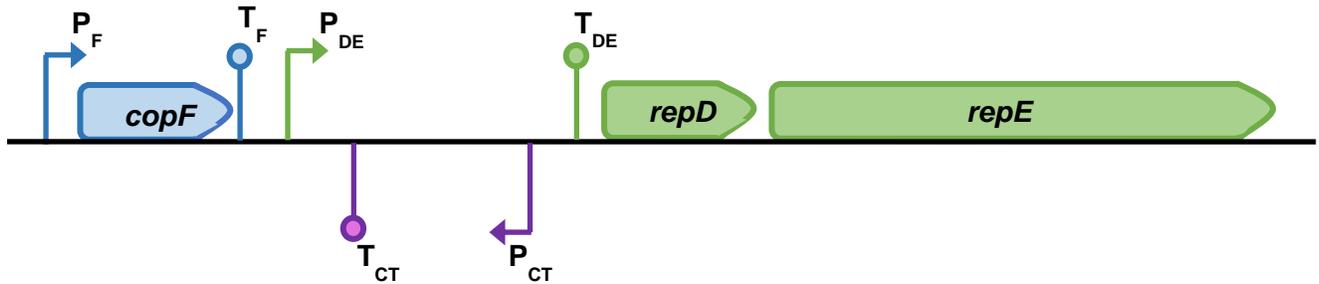
The traditional host to produce pDNA is *Escherichia coli*, being a model organism for which the upstream and downstream procedures for pDNA manufacture are well established. The majority of plasmids were engineered in order to replicate and express recombinant proteins in this Gram-negative host, having been attained high supercoiled pDNA yields in a robust manner<sup>68</sup>. However, the copurification of lipopolysaccharides (LPS) present in *E. coli* outer membrane, raises several safety concerns owing to its extremely immunogenic properties that could lead to unwanted reactions in humans and potentially death<sup>69</sup>. In order to assure that pDNA is endotoxin-free, several LPS-removing purification steps must be attained, which increases substantially the manufacture cost<sup>68</sup>.

Gram-positive bacteria are safer alternative hosts for pDNA production and purification, since they are naturally LPS-free<sup>21</sup>, allowing simpler and more profitable downstream procedures. Lactic Acid Bacteria (LAB) are a particular group of Gram-positive bacteria with even more interesting features to attain our goal, due to its GRAS (Generally Recognized As Safe) and food-grade status, as well as to its probiotic abilities<sup>21</sup>. On account of being broadly used in the dairy industry, *Lactococcus lactis* is the most well characterized LAB so far, since several strains already have its genome sequenced and some genetic tools have been specifically developed for this species<sup>70</sup>. But besides safety and in order to achieve a more profitable process for pDNA vaccines production, another issue that must be addressed is the plasmid copy number in LAB. When the goal is to obtain pDNA as the final product or a high amount of recombinant protein, the ideal is to have a plasmid with an origin of replication that allows to generate a high number of copies<sup>71</sup>.

Shuttle vectors, able to replicate both in Gram-negative and Gram-positive hosts, are generally required when working with LAB, since genetic manipulation of the vector is much easier using *E. coli*. Whereas in *E. coli* high plasmid copy numbers can be achieved (500-700 copies per cell for pUC vectors)<sup>72</sup>, the origins of replication for LAB usually produce a much lower number of plasmid copies per cell. Several strategies have already been pursued to increase plasmid copy number in LAB by changing the type of origin of replication (theta or rolling circle)<sup>73</sup> or by modification of specific features of the replicon, such as deletion of auxiliary factors<sup>74</sup>, without much success. The theta-type pAMβ1 replicon is promising due to its high structural stability that allow cloning large fragments, wide host range within Gram-positive bacteria and well-studied replication mechanism<sup>60</sup> and copy number control elements<sup>61</sup>. The highest number of copies reported for this replicon was around 100 copies per *Bacillus subtilis* cell, after inactivation of *copF*, a transcriptional repressor of the replication initiation protein RepE<sup>60</sup>.

The pAMβ1 origin of replication (Figure 2.1), which is present in pTRKH3 plasmid, is composed of different sequences that are responsible for the control of plasmid copy number (PCN) value. The *repE* gene codes for the rate-limiting replication initiation protein essential for the unidirectional theta-type replication mechanism, while the

upstream encoded RepD protein have an unknown function, not essential for the plasmid replication<sup>60</sup>. Both genes are under the control of the P<sub>DE</sub> promoter and their transcription is tightly regulated by two different systems that ultimately control the number of copies of the plasmid: the transcriptional repressor protein CopF and an antisense RNA-mediated transcription attenuation system.



**Figure 2.1.** pAM $\beta$ 1 replication region of the pTRKH3 plasmid. P<sub>F</sub> – copF promoter; T<sub>F</sub> – copF terminator; P<sub>DE</sub> – repDE promoter; T<sub>DE</sub> – repDE terminator; P<sub>CT</sub> – countertranscript RNA promoter; T<sub>CT</sub> – countertranscript RNA terminator.

CopF negatively regulates the PCN by binding to the 31-nucleotide’s operator located upstream of the -35 box of the P<sub>DE</sub> promoter, being responsible for a 10-fold repression of the transcription from this promoter, by directly interfering with the RNA polymerase binding<sup>61</sup>.

Between the *copF* and *repD* genes there is a 350 bp intergenic region that besides P<sub>DE</sub> promoter also harbors the P<sub>CT</sub> promoter in the complementary DNA strand (and its rho-independent transcription terminator T<sub>CT</sub>), which is responsible for the constitutive transcription of a small countertranscript RNA (CT-RNA) that controls PCN by a transcriptional attenuation system<sup>61,75</sup>. The CT-RNA is complementary to the leader region of the *repDE* mRNA causing its premature transcription termination (attenuation) at the rho-independent T<sub>DE</sub> terminator located downstream of the P<sub>DE</sub> promoter, but upstream the *repDE* genes<sup>75</sup>. There are no specific information about the pAM $\beta$ 1 transcription attenuation mechanism, but since pAM $\beta$ 1 and pIP501 plasmids belong to the same *inc18* family is assumed that both transcription attenuation mechanisms are quite similar and we could use the pIP501 origin of replication as model<sup>76</sup>. During transcription, the leader region (5'-UTR) of the *repDE* mRNA could adopt two structures that are mutually exclusive, depending on the presence of the CT-RNA. CT-RNA molecules can bind to a complementary sequence, located upstream of the ribosome binding site (RBS), of the nascent *repDE* mRNA forming the T<sub>DE</sub> attenuator that is responsible for the premature transcription termination of the *repDE* mRNA. Only a shorter mRNA is synthesized that does not contain the information for the Rep proteins<sup>76</sup>.

Besides repressing the transcription from the P<sub>DE</sub> promoter, CopF indirectly increases transcription initiation from the antisense P<sub>CT</sub> promoter<sup>76</sup>. More specifically, and considering the high homology between CopF and CopR (from pIP501 plasmid) proteins and analogous mechanisms between *inc18* family, CopF prevent the convergent transcription from the P<sub>DE</sub> and the P<sub>CT</sub> promoters (Figure 2.1), only possible due to its supercoiling-sensitive characteristics<sup>76</sup>. When CopF is not bind to the operator, a higher PCN is expected due to increased transcription from P<sub>DE</sub> promoter through

the antisense supercoiling-sensitive P<sub>CT</sub> promoter, resulting in a decreased transcription from the antisense promoter and accordingly a lower attenuation rate (lower CT-RNA/PCN ratio)<sup>76</sup>. The increased quantity of *repDE* mRNA also titrates the remaining CT-RNA, reducing even more its attenuation effect. In the situation when CopF is attached to the operator, the opposite happens, with RNA polymerase does not being able to transcribe through P<sub>DE</sub> promoter and indirectly allowing a higher transcription initiation rate from the P<sub>CT</sub> promoter, which increases the attenuation effect (higher CT-RNA/PCN ratio)<sup>76</sup>.

All of these mechanisms control the copy number fluctuations, in order to prevent loss of the plasmid or, on the other side, to avoid wasting energy producing an excessive quantity of plasmids. When PCN is too high, CT-RNA quantity also increases because it is constitutively synthesized, increasing *repDE* transcription attenuation and consequently decreasing replication. In the scenario when PCN is turning too low, CopF concentration is also lower, allowing transcription from the P<sub>DE</sub> promoter and at the same time avoiding the convergent transcription from the P<sub>CT</sub> promoter. In that way, it will reduce the attenuation mechanism and the PCN will increase<sup>76</sup>. The two copy number control systems seems to have additive properties, because it results in a 100-fold repression of *repE* transcription, but when both are inactivated it was never achieved a PCN higher than 100 copies<sup>61</sup>.

In the present work we propose a different approach for modifying the plasmid copy number of the pAMβ1 replicon, by altering the ribosome binding site (RBS) or Shine-Dalgarno sequence of the *repDE* operon that accordingly will influence the strength of ribosome binding to the mRNA and ultimately the production of the RepE protein. Besides the strong influence of the RBS on the translation initiation rate, which are known to affect the amount of protein produced<sup>77,78</sup>, there are other features known to affect the expression level, such as the stability and secondary structures of the transcripts, the sequences upstream and downstream of the ORF, the strength of the start codon<sup>77</sup> and the existence of regulatory proteins and codon usage<sup>78</sup>. Thereat, besides analyzing the plasmid copy number of the several RBS mutants, the mRNA secondary structures and corresponding predicted translation initiation rates will be calculated, and the *repD* and regulators (*copF* and *CT-RNA*) mRNA quantities will be assessed.

## 2.3. Material and methods

### 2.3.1. Bacterial strains and plasmids

*E. coli* DH5α (Invitrogen) was used for the site-directed mutagenesis steps being the resulting plasmids then transferred by electroporation to *L. lactis* LMG19460 (BCCM Culture Collection, Belgium).

The plasmid used as backbone for the subsequent modifications by site-directed mutagenesis was the pTRKH3 (BCCM/LMBP Plasmid Collection, Belgium), a shuttle cloning vector able to replicate both in *E. coli* and several Gram-positive genera (*Lactococcus*, *Enterococcus*, *Streptococcus* and *Lactobacillus*). To make that possible, pTRKH3 harbors two distinct replication origins, p15A with a medium copy number (30-40 copies per cell) in *E. coli* and pAMβ1 with a high copy number (45-85 copies per cell) in Gram-positive hosts, as well as two different selection markers, namely a tetracycline resistance gene for selection in *E. coli* and a erythromycin resistance gene that allows selection in both taxonomic groups<sup>65</sup>.

### 2.3.2. Media and growth conditions

*E. coli* DH5 $\alpha$  transformed with the pTRKH3 was grown at 37°C, 250 rpm in 20 g/L of LB medium (Nzytech), supplemented with 500  $\mu$ g/mL erythromycin (Sigma). *L. lactis* LMG19460 was microaerophilically grown at 30°C, 100 rpm in 54.3 g/L MRS (pH 6.2, containing 20 g/L of glucose) (LiofilChem), which was supplemented with 5  $\mu$ g/mL erythromycin (Sigma) when the strain was transformed with the plasmids. The erythromycin concentration used for both strains were determined experimentally for the strains present in our laboratory.

*L. lactis* strains harboring the wild-type or the plasmids modified by site-directed mutagenesis were grown in a systematic way, starting with an overnight inoculum grown at 30°C, 100 rpm in 5 mL MRS<sup>79</sup> (15 mL Falcon tube). Each inoculum was used to start a culture with an initial optical density of 0.1, using the previously determined correlation of 1 OD unit at 600 nm being equivalent to  $7 \times 10^8$  cells/ml<sup>80</sup>, which was confirmed for our strain. The cultures were performed in 100 mL shake flasks with 75 mL of M-17 medium<sup>81</sup> (pH 7.0, containing 5 g/L of lactose) (Fluka) supplemented with 20 g/l of glucose and 5  $\mu$ g/ml of erythromycin, at 30°C, 100 rpm. At several time points (3h, 6h, 8h, and 10.5h) samples were collected for OD and pH measurements, for real-time qPCR analysis, for pDNA purification (OD=20 equivalent to  $1.4 \times 10^{10}$  cells) to assess plasmid quantity and quality, and for total RNA extraction ( $1.0 \times 10^9$  cells) for further analyses of *repDE*, *feoA*, *copF* and *CT-RNA* mRNAs. Since a considerable amount of sampling volume was needed to those analyses, four parallel shake flasks (one per time point) were performed in each one of the independent cultures, to maintain the culture volume and guarantee the same growth conditions to all cells.

### 2.3.3. pTRKH3 modification by site-directed mutagenesis

In order to achieve different plasmid copy numbers in *L. lactis* cells, the RBS sequence of the *repDE* promoter of the pAM $\beta$ 1 replication origin of the plasmid pTRKH3 was modified by site-directed mutagenesis.

There are two likely overlapped RBS sequences (Figure 2.2) at the 5' untranslated region (5'-UTR) of the *repDE* mRNA. Taking that in account we designed new RBS sequences based in the two theoretically strongest sequences of the prokaryotic RBS Anderson library (Registry of Standard Biological Parts)<sup>82</sup> (pTRKH3-a and pTRKH3-c) and also a sequence with intermediate strength (pTRKH3-b). In addition, we designed two full complementarity consensus RBS sequences<sup>83,84</sup> (pTRKH3-d and pTRKH3-e), but while pTRKH3-d keeps the same distance to the start codon as the previous mutants, the increased distance in pTRKH3-e could theoretically improve the translation initiation rate<sup>57</sup>; and finally a negative control (pTRKH3-f) was designed with the purpose of being less complementary to the 16S rRNA.

consensus RBS (hypothesis 1)	A A G G A G G T	Number of equal nucleotides (max. 8)	
consensus RBS (hypothesis 2)	A A G G A G G T	Hyp 1	Hyp 2
pTRKH3	A A G G T G G A G T T T G T A T G	6	5
pTRKH3-a	A A G G <b>G</b> G G A <b>C A A</b> T G T A T G	6	3
pTRKH3-b	A A G <b>C T</b> G G A G <b>C C</b> T G T A T G	5	4
pTRKH3-c	A A G <b>C A</b> G G A <b>C C C</b> T G T A T G	6	4
pTRKH3-d	A A G <b>A A</b> G G A G <b>G</b> T T G T A T G	6	8
pTRKH3-e	A A G G <b>A</b> G G <b>T</b> G T T T G T A T G	8	5
pTRKH3-f	A A G <b>C T C</b> G A <b>C</b> T T T G T A T G	4	3

**Figure 2.2.** Sequence of the RBS region of the *repDE* promoter in the pAM $\beta$ 1 origin of replication, in the wild-type pTRKH3 and in the six modified plasmids. We considered the RBS sequences either starting 6 nucleotides (hypothesis 1) or 3 nucleotides (hypothesis 2) upstream the ATG codon. The nucleotides altered by site-directed mutagenesis are highlighted in grey. All the distances to the *repD* start codon are also represented. The predicted RBS sequence from pTRKH3-e is shifted upstream 3 nucleotides (6 nucleotides from the start codon). The number of equal nucleotides (with a maximum of 8) between the RBS sequence of each plasmid and the consensus RBS from *L. lactis* is indicated for both consensus RBS hypothesis.

The features of each new RBS sequence are represented in Figure 2.2. The primers used to site-directed mutagenesis procedure were designed using the QuickChange Primer Design program<sup>85</sup> and their sequences are shown in Table 2.1.

**Table 2.1.** Sequence, length and melting temperature of the primers used in site-directed mutagenesis procedures. F and R indicate forward and reverse primers, respectively.

Primers	Sequence	Length (nt)	T <sub>m</sub> (°C)
pTRKH3-a F	CGTTTGATTGCTTTTTTTGTATTCATTTATAGAAGGGGGACAA TGTATGAATCATGATGAATGTAAAACCTTATATAAAA	79	77
pTRKH3-a R	TTTTATATAAGTTTACATTCATCATGATTCATACATTGTCCCC CTTCTATAAATGAATACAAAAAAGCAATCAAACG	79	77
pTRKH3-b F	TTTATATAAGTTTACATTCATCATGATTCATACAGGCTCCAG CTTCTATAAATGAATACAAAAAAGCAATCAAACG	78	78
pTRKH3-b R	CGTTTGATTGCTTTTTTTGTATTCATTTATAGAAGCTGGAGCC TGTATGAATCATGATGAATGTAAAACCTTATATAAA	78	78
pTRKH3-c F	TTTACATTCATCATGATTCATACAGGGTCTGCTTCTATAAAT GAATACAAAAAAG	57	73
pTRKH3-c R	CTTTTTTTGTATTCATTTATAGAAGCAGGCCCTGTATGAATC AT GATGAATGTA	57	73
pTRKH3-d F	TATATAAGTTTACATTCATCATGATTCATACAAACCTCCTTCTT CTATAAATGAATACAAAAAAGCAATCAAAC	75	75
pTRKH3-d R	GTTTGATTGCTTTTTTTGTATTCATTTATAGAAGGAGGTT GTATGAATCATGATGAATGTAAAACCTTATATA	75	75
pTRKH3-e F	GTTTTACATTCATCATGATTCATACAAACACCTCCTTCTATAA ATGAATACAAAAAAGCAA	62	74
pTRKH3-e R	TTGCTTTTTTTGTATTCATTTATAGAAGGAGGTGTTTGTATGA ATCATGATGAATGTAAAAC	62	74
pTRKH3-f F	TAAGTTTTACATTCATCATGATTCATACAAAGTCGAGCTTCTA TAAATGAATACAAAAAAGCAATCA	68	75
pTRKH3-f R	TGATTGCTTTTTTTGTATTCATTTATAGAAGCTCGACTTTGTA TGAATCATGATGAATGTAAAACCTTA	68	75

The mutagenesis reaction was prepared using the KOD Hot Start DNA Polymerase kit (Novagen), by mixing 50 ng of template pDNA, 0.02U/ $\mu$ L of KOD DNA polymerase, 2.3 mM of  $MgSO_4$ , 1x buffer, 0.2 mM of dNTPs, 0.5  $\mu$ M of each primer and completed with PCR-grade water to a final volume of 25  $\mu$ L. The cycling conditions were: initial denaturation for 3 min at 95°C, followed by 30 cycles of 1 min at 95°C, 1 min at 65°C and 8 min at 70°C. After amplification, the PCR products were incubated with 20 U of DpnI (Promega) for 2h at 37°C, to digest the template pDNA. The modified plasmids were used to transform chemically competent DH5 $\alpha$  cells by heat shock. The colonies grown in selective LB agar supplemented with 5  $\mu$ g/mL erythromycin were analyzed for the presence of the desired mutations in the pTRKH3, by sequencing of the pDNA (Stabvida), previously purified using the High Pure Plasmid Isolation Kit (Roche). The quality of each modified plasmid was evaluated by measurement in a Nanodrop Spectrophotometer (Nanovue Plus, GE), restriction analysis and agarose gel electrophoresis.

#### **2.3.4. LAB electrotransformation with modified plasmids**

Electrocompetent *L. lactis*, prepared by a protocol in which the cells were grown in a high NaCl concentration<sup>86</sup>, were transformed with the wild-type and the modified pTRKH3 plasmids. Aliquots of 100  $\mu$ L with  $2 \times 10^9$  cells were mixed with 500 ng of pDNA and, after 30 min on ice, electroporated with 3-5 pulses of 18 kV/cm. Cells were allowed to recover for 3 hours with 900  $\mu$ L MRS at 30°C, 100 rpm and then were plated on MRS agar supplemented with 5  $\mu$ g/mL erythromycin.

The presence of the desired plasmids in *L. lactis* was validated by sequencing of the pDNA purified with Nucleospin Plasmid Kit (Macherey-Nagel), from the selected colonies grown in liquid MRS medium.

#### **2.3.5. Plasmid copy number determination by qPCR**

##### **2.3.5.1. Primer design**

Two sets of primers were designed, one specific for the erythromycin resistance gene (*erm*) present in the pTRKH3 plasmid and the other specific for the ferrous iron transport protein A (*feoA*) chromosomal single copy gene of *L. lactis* (Table 2.2).

The genome of the *L. lactis* LMG19460 strain used in the present study was not yet sequenced at the time the experiments were performed. Therefore, it was assumed that the sequence of the *feoA* gene was similar to *L. lactis* ssp. *lactis* IL1403. This assumption was verified by analysis of the product length after amplification using *L. lactis* LMG19460 genomic DNA as template. Primers were synthesized by Stabvida.

**Table 2.2.** Sequence of the primers used for qPCR amplification of the target (*erm*, *repDE*, *copF* and *CT-RNA*) and reference (*feoA*) genes, as well as the accession number of the complete sequence and expected amplicon size. F and R indicate forward and reverse primers, respectively.

	Gene	Primer sequence	Accession number	Product size (bp)
Plasmid	<i>erm</i>	F: CCATGGGTCTGACATCTATCT R: CTGTGGTATGGCGGTAAGT	LMBP 4462	190
Chromosome	<i>feoA</i>	F: TCAGACGCCGCTTGATGGAC R: AGTTCAAGAGGGTCCGCAAGTG	AE005176	89
Plasmid	<i>repDE</i>	F: TCAGATAAGTATTTTTCTTCGGAGG R: TTAATTAAGTTGGTTTCTTTTCA	LMBP 4462	171
Plasmid	<i>copF</i>	F: CTAAGTCAACGCTAGTAGTGG R: CTTGTTCTGATTCTGTTTCTGG	LMBP 4462	75
Plasmid	<i>CT-RNA</i>	F: TGCATAGCCCATAAGATTG R: CGAGGGAAGAGTTCATCTGA	LMBP 4462	63

### 2.3.5.2. Samples and standards preparation

Genomic DNA from *L. lactis* LMG19460 was purified using Wizard Genomic DNA Purification Kit (Promega), quantified by Nanodrop measurement and its quality assessed by agarose gel electrophoresis. The pTRKH3 was purified from *L. lactis* LMG19460 using Nucleospin Plasmid Kit (Macherey-Nagel). Then, genomic and plasmid DNAs were serially diluted in PCR-grade water to be used as real-time qPCR standards (1, 10, 100, 1,000 and 10,000 pg per reaction).

The samples of *L. lactis* LMG19460 harboring the wild-type and the modified plasmids were collected as described above and then diluted in PCR-grade water in order to add 1,000 cells to each PCR capillary. To ensure that the samples and the pDNA standards have the same putatively inhibitors present, the capillaries for the pDNA standard curve were spiked with the same number of *L. lactis* cells (1,000) without plasmid. The same was not applied to the gDNA standards, because the *L. lactis* genome of the cells used to spike the capillaries could mislead the gDNA quantification.

### 2.3.5.3. Real-time qPCR reaction

The real-time qPCR reactions were performed in the LightCycler (Roche) detection system using the FastStart DNA Master SYBR Green I kit (Roche). Each 20 µL reaction mixture had 2 µl of 10x SYBR Green Mix, 1 µl of each primer (final concentration 0.5 µM), 1.6 µl of MgCl<sub>2</sub> (final concentration 3 mM), 10.4 µl of PCR-grade water and the remaining volume was completed with the samples (4 µL with 1,000 cells containing the different plasmids) or with pDNA (2 µL of pDNA plus 2 µL with 1,000 spiking cells) or gDNA (4 µL) standards. Negative controls without cells and without pDNA or gDNA were performed in all cycling reactions, as well as at least one point of each standard curve for experimental validation.

The cycling conditions were as follows: initial denaturation for 10 min at 95°C, followed by 40 cycles of 10s at 95°C, 5s at 57°C and 14s at 72°C. The fluorescence signal was detected and quantified automatically by the instrument at the end of each extension step. To confirm that only the desired specific targets were amplified, melting curve analyses

were performed by making a temperature gradient of 0.05° C/s from 70 to 95°C. Lastly, the samples were cooled down to 40°C for 30s.

#### 2.3.5.4. Plasmid copy number (PCN) quantification

A relative quantification method developed by Skulj et al. (2008)<sup>87</sup> was used for PCN quantification using the real-time qPCR data. The crossing point (CP) values were determined automatically by the LightCycler software, using the Second Derivative Maximum method.

First, the standard curves for pDNA and gDNA were determined by plotting the logarithm of each concentration against the CP values. After estimation of each linear regression equation, the slope of each standard curve was used to determine the amplification efficiency (E) according to Equation 1:

$$E = 10^{\left(-\frac{1}{\text{slope}}\right)} \quad (1)$$

Knowing the amplification efficiencies of both the pDNA (E<sub>p</sub>) and gDNA (E<sub>g</sub>) standards, as well as the CP values for each sample amplified independently with the two primer sets (CP<sub>p</sub> with plasmid primer set and CP<sub>g</sub> with gDNA primer set), the PCN of each sample can be calculated using the Equation 2<sup>87</sup>:

$$\text{PCN} = \frac{E_g^{\text{CP}_g}}{E_p^{\text{CP}_p}} \quad (2)$$

The pDNA and gDNA standard curves were performed in triplicate. Concerning the *L. lactis* samples harboring the wild-type and modified pTRKH3 plasmids, they were collected from 4 to 7 independent growths. Several growth samples were analyzed in two different cycling reactions, allowing the assessment of inter- and intra-assay variability.

#### 2.3.5.5. qPCR validation

To verify the specificity of the real-time qPCR amplification, the melting temperature (T<sub>m</sub>) of each amplicon was checked. Also, the content of each capillary was run in a 1% agarose gel, being visible the amplicons with the correct sizes for each set of primers, confirming the specificity of all the performed reactions.

Negative controls without cells and without pDNA or gDNA were performed in all cycling reactions and resulted in CPs much higher (>than 35 CPs) than all the samples, therefore background amplification could be considered negligible.

The validity of the qPCR results was also evaluated by the coefficient of determination (R<sup>2</sup>) and by the percentage of amplification efficiency of each standard curve, which was calculated using Equation 3<sup>87</sup>:

$$E(\%) = \left(10^{\left(-\frac{1}{\text{slope}}\right)} - 1\right) \times 100 \quad (3)$$

The quality and integrity of the pDNA in each analyzed sample was determined in purified plasmid samples with 1.4 x10<sup>10</sup> cells (OD=20) using the High Pure Plasmid Isolation Kit (Roche) and subsequent quantification in a Nanodrop

Spectrophotometer (Nanovue Plus, GE) and isoform analysis by agarose gel electrophoresis. The agarose gel electrophoresis performed shown that the wild-type as well as the modified derivatives of pTRKH3 did not undergo any significant changes in its quality from 3h to 10.5h of growth, preserving a majority of supercoiled isoform, followed by open circular, and only a vestigial amount of linear isoform. Samples were also taken at 24h of growth, to ascertain if the PCN value increased when compared with the 10.5h samples, but the pDNA was degraded when visualized on a 1% agarose gel, making it unfeasible to take any conclusions from these later points of stationary phase of growth.

#### 2.3.5.6. Statistical analysis

For each independent growth, average PCN from the different cycling reactions was determined and then used to calculate the average and inter-assay variability by determination of the standard error of the mean (SEM). When necessary, a t-student statistical test was performed and differences were considered statistically significant when  $p < 0.05$ .

#### 2.3.6. Relative translation initiation rate prediction

The translation initiation rates (TIR) and consequently the predicted protein expression levels of each designed RBS sequences were evaluated using the RBS Calculator (<https://salislab.net/software/>, Salis Lab, Penn State University)<sup>88,89</sup>. This online tool allows a “reverse engineering” TIR calculation, having in consideration the strength of several molecular interactions in the hybridization of the 16S rRNA to the RBS sequence and of the tRNA<sup>fMet</sup> to the start codon in each mRNA transcript (given as the Gibbs free energy change  $\Delta G_{\text{total}}$ ), as well as the presence of secondary structures in the mRNA that could affect the 16S rRNA binding site or standby site, and the distance between the actual 16S rRNA binding site and the start codon.

For the thermodynamic calculations of the *repD* transcripts, 74 nucleotides’ long mRNA sequences (32 nucleotides of the 5’-UTR region and 42 nucleotides from the coding region) were used as input for the original pTRKH3 and the six modified plasmids, since longer inputs did not alter the prediction results. For the *repE* transcripts, longer inputs of 400 nucleotides (consisting of 32 nucleotides of the 5’-UTR region followed by 368 nucleotides of the coding regions) were necessary in order to analyze the TIR and  $\Delta G_{\text{total}}$  predictions for *repD* and *repE* as an operon.

#### 2.3.7. *feoA*, *repD*, *copF* and *CT-RNA* mRNA two-step quantification by qPCR

##### 2.3.7.1. Primer design

Four sets of primers were designed to be used both for the first strand cDNA synthesis and for the qPCR real-time quantification. One set of primers is the same that had been used in the PCN quantification, to detect the ferrous iron transport protein A (*feoA*) chromosomal single-copy gene of *L. lactis* (Table 2.2). The other sets of primers were specific for three different pTRKH3 plasmid genes: the first ORF of the replication initiation operon (*repD*), the

transcriptional repressor of the RepE protein (*copF*) and the countertranscript RNA responsible for the RepE transcription attenuation system (*CT-RNA*) (Table 2.2). Primers were synthesized by Stabvida.

### 2.3.7.2. RNA isolation and mRNA quantification by real-time qPCR

Total RNA from *L. lactis* LMG19460 harboring the pTRKH3 and the six modified plasmids was extracted using the High Pure RNA Isolation Kit (Roche) according to the manufacturer's instructions. The extracted RNA was quantified using a Nanodrop Spectrophotometer (Nanovue Plus, GE) and 700 ng of total RNA was used for reverse transcription. This step was performed with the First Strand cDNA Synthesis Kit for RT-PCR (AMV) (Roche), using the primers *feoA\_R*, *repD\_R*, *copF\_R* and *CT-RNA\_F* (Table 2.2). The real-time qPCR reactions were performed in the LightCycler (Roche) detection system using the FastStart DNA Master SYBR Green I kit (Roche). Each 20  $\mu$ L reaction mixture were performed as described above, but instead of cells, a fixed volume (2  $\mu$ L) of cDNA first strand was used per capillary for quantification by qPCR using the complete sets of primers (Table 2.2), with *feoA* being used again as the reference gene.

The four cDNA qPCR standard curves were obtained using genomic and plasmid DNAs purified as described above and serially diluted in PCR-grade water (1, 10, 100, 1,000 and 10,000 pg).

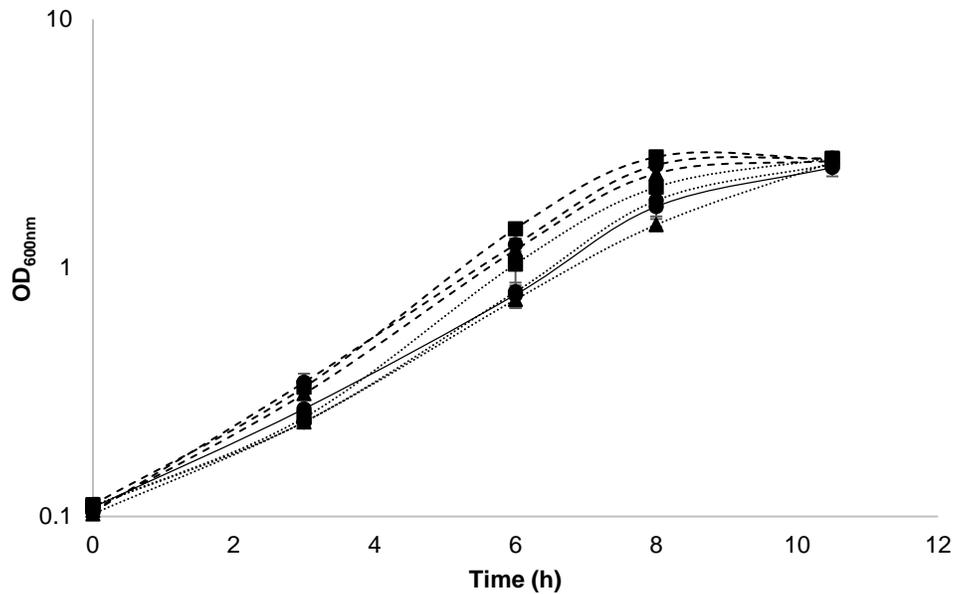
The cycling conditions were as follows: initial denaturation for 10 min at 95°C, followed by 40 cycles of 10s at 95°C, 5s at 55°C and 14s at 72°C. To confirm that only the desired specific targets were amplified, melting curve analyses were performed by making a temperature gradient of 0.05° C/s from 70 to 95°C.

The number of *repD*, *feoA*, *copF* and *CT-RNA* mRNA molecules were calculated as described above, using the relative quantification method developed by Skulj et al. (2008)<sup>87</sup>.

## 2.4. Results and discussion

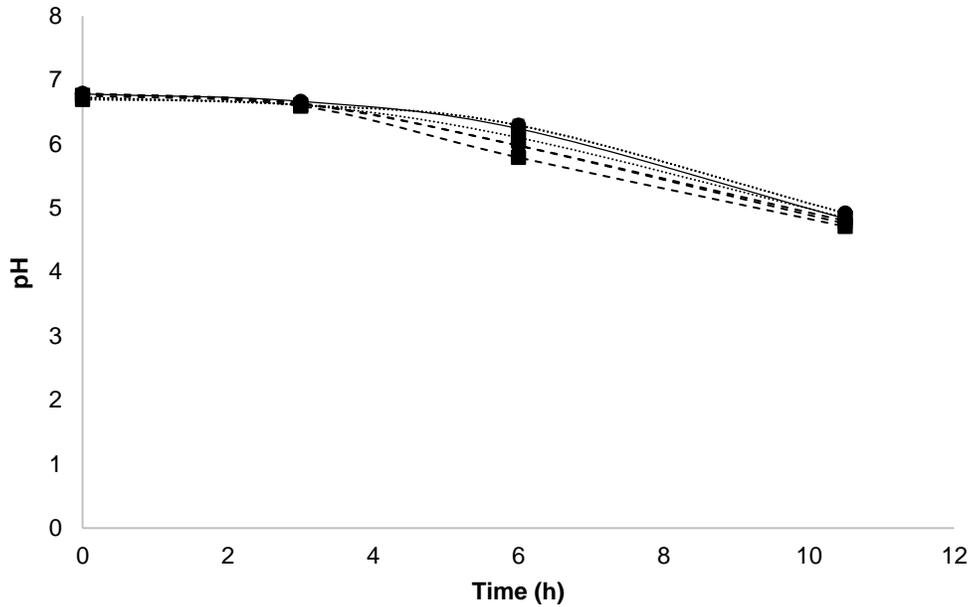
### 2.4.1. Growth profile of *L. lactis* LMG19460 harboring wild-type and modified plasmids

*L. lactis* LMG19460 cells harboring the non-modified pTRKH3 and the plasmids with the designed mutations were grown under previously optimized conditions (75 mL in 100 mL shake flasks of M-17 with 20 g/L glucose at 30°C, 100 rpm). The specific growth rates of cells containing the modified plasmids were not significantly affected (Figure 2.3) ( $p > 0.05$ ).



**Figure 2.3.** Growth profiles of *L. lactis* LMG19460 cells harboring the non-modified pTRKH3 (filled circle with solid line) and the six plasmids with the designed mutations in the RBS sequence of the *repDE* promoter (filled circle for pTRKH3-a, filled square for pTRKH3-b, filled triangle for pTRKH3-c, empty circle for pTRKH3-d, empty square for pTRKH3-e and empty triangle for pTRKH3-f), during 10.5h of growth in 100 mL shake flasks with 75 mL of M-17 supplemented with 20 g/L glucose, 30°C, 100 rpm (initial  $OD_{600nm}=0.1$ , starting from overnight inoculum). The profile was obtained from seven independent growths and the standard error of the mean (SEM) is represented for each point. The specific growth rates ( $\mu$ ) were  $0.40 (\pm 0.01)$ ,  $0.40 (\pm 0.01)$ ,  $0.44 (\pm 0.02)$ ,  $0.41 (\pm 0.01)$ ,  $0.41 (\pm 0.04)$ ,  $0.43 (\pm 0.01)$  and  $0.37 (\pm 0.01) h^{-1}$  for the wild-type pTRKH3 and the six mutants (pTRKH3-a to pTRKH3-f, respectively).

A drop in the pH values over time was observed as expected, derived from the lactic acid produced by *L. lactis* during growth (Figure 2.4). The pH decrease pattern was not statistically different between the cells harboring the diverse plasmids ( $p > 0.05$ ).

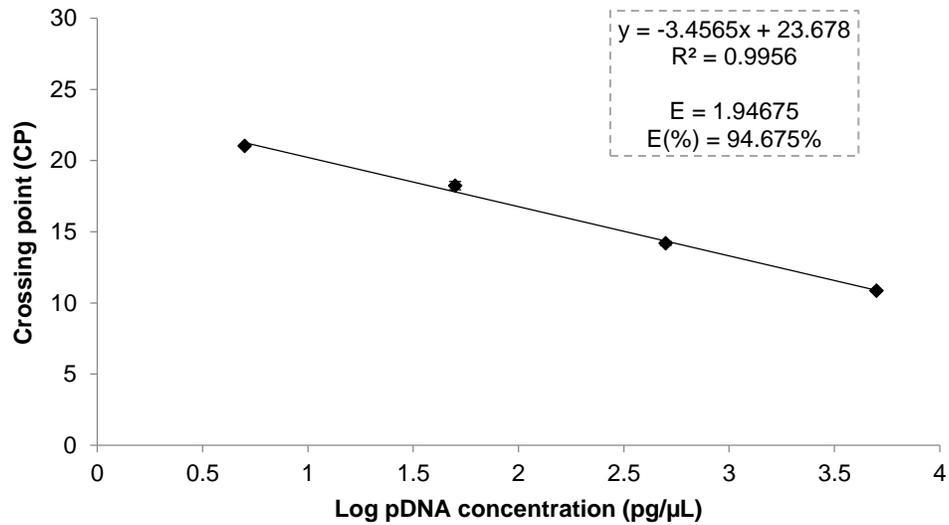


**Figure 2.4.** pH profiles of *L. lactis* LMG19460 cells harboring the non-modified pTRKH3 (filled circle with solid line) and the six plasmids with the designed mutations in the RBS sequence of the *repDE* promoter (filled circle for pTRKH3-a, filled square for pTRKH3-b, filled triangle for pTRKH3-c, empty circle for pTRKH3-d, empty square for pTRKH3-e and empty triangle for pTRKH3-f), during 10.5h of growth in 100 mL shake flasks with 75 mL of M-17 supplemented with 20 g/L glucose, 30°C, 100 rpm (initial  $OD_{600nm}=0.1$ , starting from overnight inoculum). The profile was obtained from seven independent growths and the standard error of the mean (SEM) is represented for each point.

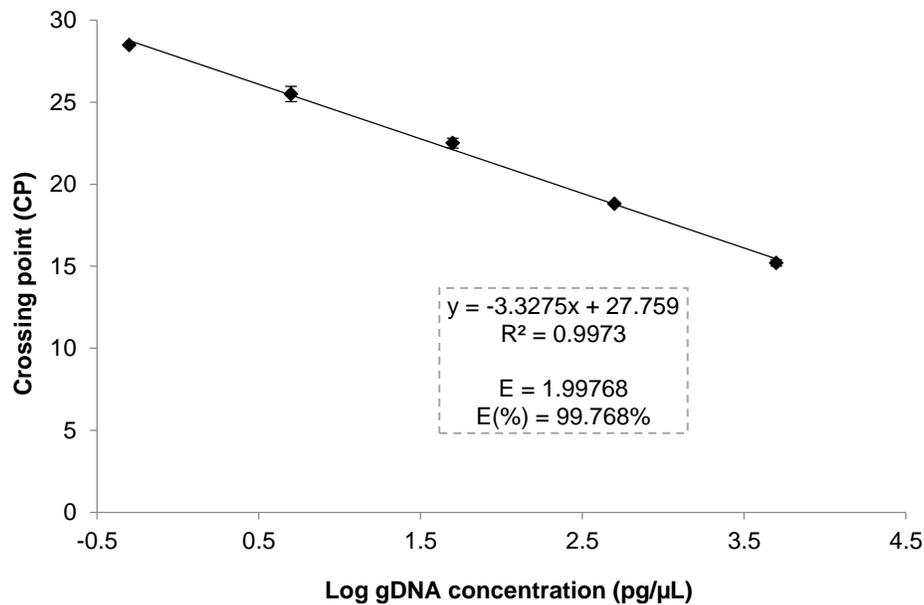
These results suggest that the modifications in the *repDE* RBS sequence of pTRKH3 do not appear to influence negatively the cell overall metabolism and the next step was to evaluate the impact of those mutations in the plasmid copy number.

#### 2.4.2. Plasmid copy number (PCN) determination by real-time qPCR

The efficiencies for pDNA and gDNA amplification were calculated using Equation 1 as 94.7% and 99.8%, respectively (pDNA and gDNA standards curves in Figure 2.5 and 2.6, respectively), which is well above the recommended 90%<sup>90</sup>, meaning that both DNA samples are highly pure, not containing impurities and contaminants that could inhibit the qPCR reaction<sup>90</sup>. The coefficients of determination ( $R^2$ ) were above 0.99 in both cases, which also indicate a high precision in the PCR performance<sup>91</sup> (Figure 2.5 and 2.6).



**Figure 2.5.** pDNA real-time qPCR standard curve performed in quadruplicate, using four log serial dilutions of pTRKH3 pDNA (10, 100, 1,000 and 10,000 pg) spiked with *L. lactis* LMG19460 cells. The linear regression trendline is represented, along with the corresponding equation, coefficient of determination value ( $R^2$ ) and efficiency values.

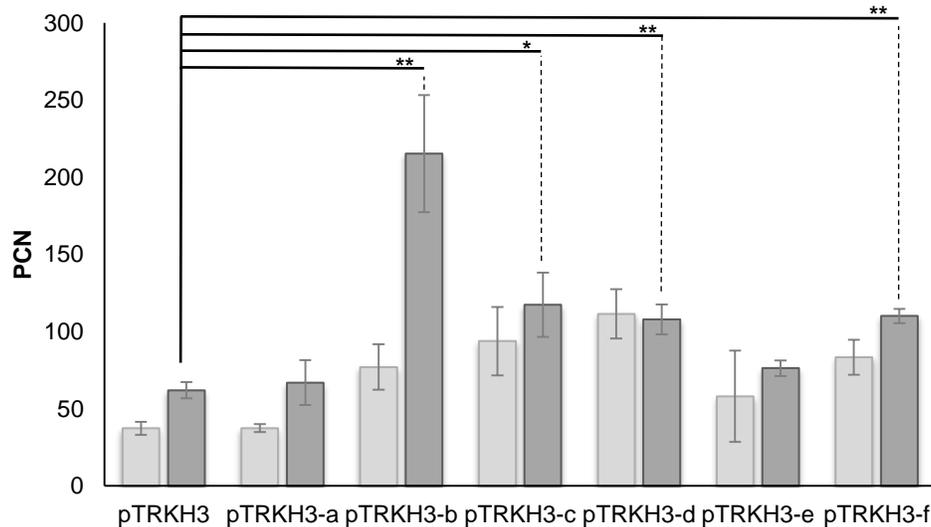


**Figure 2.6.** gDNA real-time qPCR standard curve performed in triplicate using five log serial dilutions of *L. lactis* LMG19460 gDNA (1, 10, 100, 1,000 and 10,000 pg). The linear regression trendline is represented, along with the corresponding equation, coefficient of determination value ( $R^2$ ) and efficiency values.

The PCN values were calculated using Equation 2 that takes into account not only with the CP values of each sample, but also with the pDNA and gDNA standard curves average efficiencies. This is of paramount importance because when the target gene (*erm* in plasmid) and the reference gene (*feoA* in genome) are not amplified with the same efficiency, the PCN estimates will not be accurate<sup>87</sup>. In fact, when the standard curve traditional method was applied

to these results an overestimation of almost 10-fold of PCN values in all samples was obtained, as already reported by other authors<sup>92</sup>. Only the improved method that considers the discrepancy between pDNA and gDNA amplification allowed a precise and reproducible determination of the plasmid copy number per chromosome.

The average PCN values for wild-type and modified plasmids at different growth phases (6 and 10.5h) are represented in Figure 2.7. At 10.5h of growth (Figure 2.7), that corresponds to the late exponential/early stationary phase, the average PCN of all the modified plasmids, except pTRKH3-a and pTRKH3-e were statistically higher ( $p < 0.05$  for pTRKH3-c and  $p < 0.01$  for pTRKH3-b, pTRKH3-d and pTRKH3-f) than the PCN obtained for the wild-type pTRKH3, and from all pTRKH3-b showed the highest average PCN value. The plasmid pTRKH3-b also achieved a statistically higher PCN than pTRKH3-a and pTRKH3-e ( $p < 0.05$ ), but the same does not happened for the other plasmids. The differences at the middle exponential phase (6h of growth) are less noticeable, but again all the modified plasmids, except pTRKH3-a and pTRKH3-e had a statistically higher ( $p < 0.05$ ) average PCN than the wild-type pTRKH3. Also, at middle exponential phase of growth it was pTRKH3-d and not pTRKH3-b that showed the highest PCN (statistically higher than pTRKH3, pTRKH3-a and pTRKH3-b,  $p < 0.05$ ).

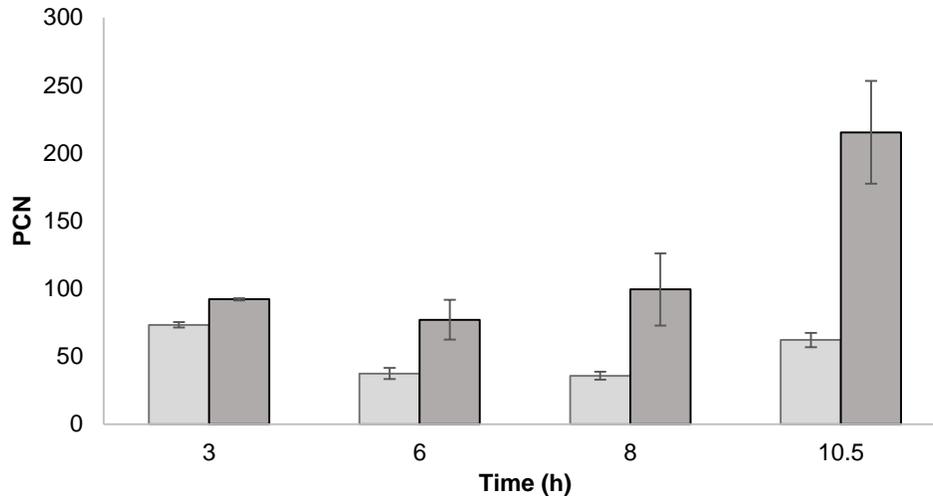


**Figure 2.7.** Average PCN values for *L. lactis* LMG19460 cells harboring the wild-type and the six modified plasmids after 6h (light grey) and 10.5h of growth (dark grey). The results were obtained from real-time qPCR analysis of seven independent growths in 100 mL shake flasks with 75 mL of M-17 supplemented with 20 g/L glucose, 30°C, 100 rpm (initial  $OD_{600nm}=0.1$ , starting from an overnight inoculum). The standard error of the mean (SEM) is represented for each point, as well as the statistically significant differences at  $p < 0.05$  (\*) or  $p < 0.01$  (\*\*) between the non-modified pTRKH3 and the plasmids with the designed mutations in the RBS sequence of the *repDE* promoter.

Samples at intermediate growth phases (3 and 8 hours) were analyzed for the original pTRKH3 and the highest plasmid producer pTRKH3-b (Figure 2.8), to evaluate if different growth phases alter plasmid production.

The increase in the PCN value of pTRKH3-b towards the late exponential/early stationary phase (Figure 2.7 and 2.8) is consistent with the fact that during exponential phase the cells redirect the metabolism mainly for biomass rather than for plasmid production. As the stationary phase approaches, the resources are available to shift metabolism

towards an increased investment in plasmid production, which is noticeable also for the majority of the remaining plasmids (except pTRKH3-d), although in a smaller scale.



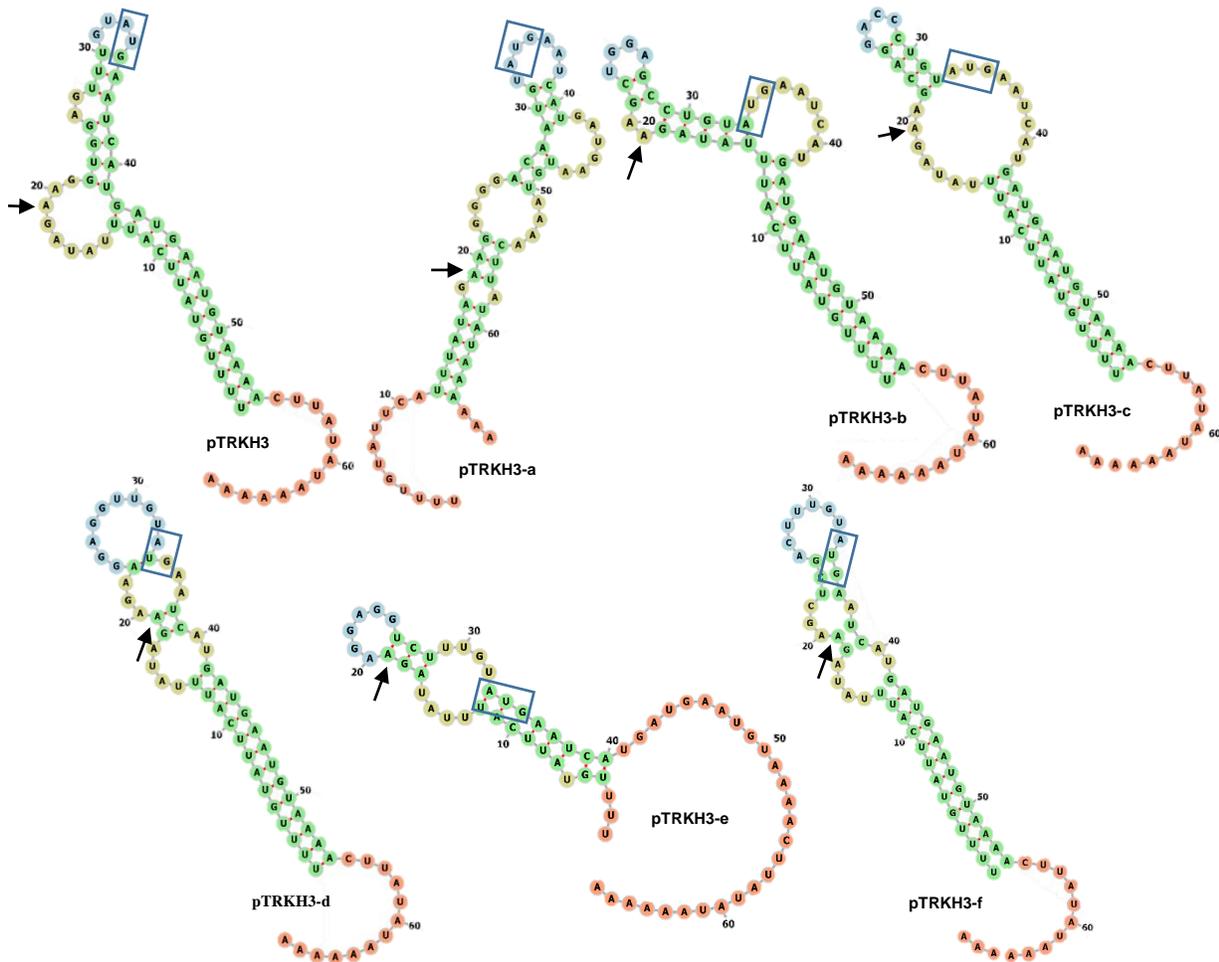
**Figure 2.8.** Average PCN values for *L. lactis* LMG19460 cells harboring the wild-type pTRKH3 (light grey) and the highest plasmid producer pTRKH3-b (dark grey) at different growth times (3, 6, 8 and 10.5h). The PCN results were obtained from real-time qPCR analysis of seven independent growths and the standard error of the mean (SEM) is represented for each point.

The non-modified pTRKH3 is described in literature as having 45-85 copies in streptococcal and lactococcal hosts<sup>65</sup>, which is consistent with the average PCN value of 62 ( $\pm 5$ ) copies per genome (10.5h of growth) obtained in this study, strengthening the validity of the method and of the remaining results. The pTRKH3-a, pTRKH3-c, pTRKH3-d, pTRKH3-e and pTRKH3-f modifications increased the average PCN values to 67 ( $\pm 14$ ), 117 ( $\pm 21$ ), 108 ( $\pm 10$ ), 76 ( $\pm 5$ ) and 110 ( $\pm 5$ ) copies per genome, respectively, while pTRKH3-b achieved 215 ( $\pm 38$ ) copies per genome, representing a 3.5 fold increase over the wild-type plasmid. This value is also much higher than the maximum of 100 copies obtained for the pAM $\beta$ 1 replicon in previous studies<sup>60</sup>.

pTRKH3-b plasmid neither has the most similar RBS to the 16S rRNA sequence nor the strongest RBS predicted by the Anderson library (*in silico* generated RBS sequences organized by its predicted expression strength). Apparently, the RBS present in pTRKH3-b allow an increase in RepD and RepE proteins needed for the replication of the plasmid or there are other mechanisms that contribute to the plasmid copy number.

### 2.4.3. *In silico* mRNA analysis

The unexpected highest PCN of the pTRKH3-b when compared with the other modified plasmids (pTRKH3-a, pTRKH3-c, pTRKH3-d and pTRKH3-e) that theoretically should had a better performance led us to search for explanations by analyzing putative differences between their predicted *repDE* mRNA likely secondary structures (Figure 2.9) and relative translation initiation rates (TIR) (Table 2.3). The predicted structures may help to understand the effect of the different sequences on the 16S rRNA binding site at the 5'-UTR, which influence the TIR value and ultimately the RepE protein expression levels that control the PCN values.



**Figure 2.9.** Prediction of the *repDE* mRNA secondary structures for the wild-type pTRKH3 and the six modified plasmids. The start codon is highlighted with a box and the first nucleotide (position 19) of the hypothesis 1 consensus RBS is indicated with an arrow. The hypothesis 2 consensus RBS starts three nucleotides downstream the arrow.

**Table 2.3.** Translation initiation rate (TIR) and  $\Delta G_{\text{total}}$  values of *repD* and *repE* genes for pTRKH3 and the six modified plasmids.

Plasmid	<i>repD</i> TIR (A.U.)	<i>repD</i> $\Delta G_{\text{total}}$ (kcal/mol)	<i>repE</i> TIR (A.U.)	<i>repE</i> $\Delta G_{\text{total}}$ (kcal/mol)
pTRKH3	13949	-5.39	4248	-2.75
pTRKH3-a	27630	-6.91	4607	-2.93
pTRKH3-b	1431	-0.33	2555	-1.62
pTRKH3-c	1222	0.02	1232	0
pTRKH3-d	126571	-10.29	5155	-3.18
pTRKH3-e	72218	-9.04	2134	-1.22
pTRKH3-f	897	0.71	4248	-2.75

Long stem-loops are thought to be responsible for lower protein expression levels by reducing the translation initiation rate, while unpaired nucleotides near the start codon lead to higher expression levels<sup>93,94</sup>. When the 16S rRNA is unable to bind the RBS it could reduce not only the protein expression levels, but also the own mRNA levels, due to its increased exposure to nuclease activity<sup>93</sup>. Solely considering the number of predicted unpaired nucleotides (versus long hairpins) in the RBS (Figure 2.9) it is clear that this attribute *per se* is not enough to explain the PCN value variation between the six mutant plasmids. For example, pTRKH3-b transcript is very similar to the pTRKH3-c but the PCN obtained with each are very different.

The TIR and  $\Delta G_{\text{total}}$  values (Table 2.3) are also important factors to take in consideration. Translation initiation is the rate-limiting step in bacterial translation processes and TIR, as already explained in the Material and Methods section, takes in consideration several factors, including the Gibbs free energy change  $\Delta G_{\text{total}}$  (sum of all binding free energies between mRNA and ribosome). When  $\Delta G_{\text{total}}$  is negative means that exists attractive interactions on the mRNA sequence surrounding the start codon and is more likely that the ribosome binds to mRNA and initiates translation<sup>88,89</sup>.

Starting with the *repD* mRNA, the predicted TIR and  $\Delta G_{\text{total}}$  values are in accordance with the theoretically expected when the mutations in the RBS of the *repDE* promoter were designed, except for pTRKH3-c, but again, not with the PCN. The plasmids pTRKH3-d and pTRKH3-e have full complementary sequences to the 16S rRNA sequence of the consensus Shine-Dalgarno region in *L. lactis* (considering the RBS consensus hypothesis 2 and 1, respectively) and were the ones with the highest TIR values for *repD*. The distance of the RBS to the start codon (hypothesis 1) in pTRKH3-e could theoretically improve the translation initiation rate<sup>57</sup>, but apparently, the shorter distance of pTRKH3-d (hypothesis 2) is favorable. The analysis for the *repE* mRNA is not so straightforward, since pTRKH3-d have the highest TIR, which makes sense with the stronger designed RBS (hypothesis 2), but pTRKH3-e comes in fifth place, having a TIR even lower than the negative control pTRKH3-f (designed to have the minimal complementarity to the RBS consensus hypothesis 2).

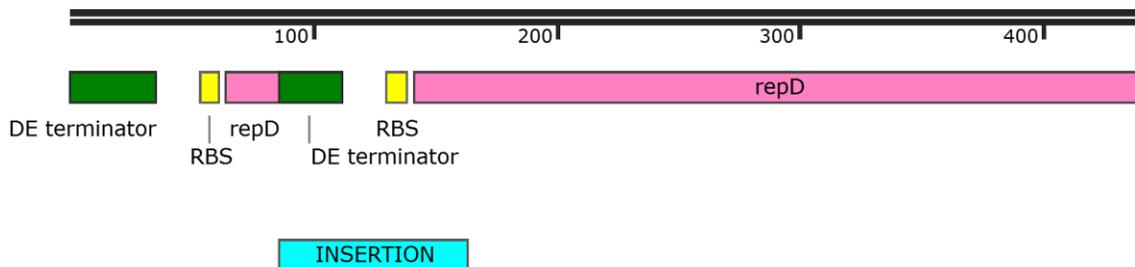
The lack of correlation in the TIR values between the *repD* and *repE* mRNA is explained by the fact that predicting the expression levels of genes organized as an operon is an extremely challenging task, since many factors affect the correlation between the expression level of the first gene and the second gene in the operon. Most of the times the translation efficiency of genes within an operon is interdependent, a phenomenon called translational coupling<sup>77</sup>. The gene expression level of the downstream gene is always enhanced by the expression level of the first gene, but varies from almost no coupling up to ten-fold, depending on the distance between the stop and start codons of the neighboring

genes, the sequence of the upstream gene, the formation of inhibitory secondary structures in the downstream gene and the sequence of both translation initiation regions<sup>77</sup>.

The most unexpected result is when the TIR values are analyzed side by side with the PCN values, because the highest producer pTRKH3-b was expected to have one of the highest TIR values, what would correspond to high RepD and RepE protein expression levels and consequently high PCN values, but instead showed a 10 and 2 fold lower TIR than the wild type for *repD* and *repE*, respectively. This phenomenon is probably due to the fact that the correlation between the mRNA and the respective protein levels depends on several factors at the translation level, such as the similarity of the Shine-Dalgarno region to the 16S rRNA sequence, the TIR value, the presence of secondary structures in the mRNA, the stability of the transcript, codon bias and the presence of regulatory proteins and sRNAs that could regulate translation<sup>77,78,94</sup>. In addition, although the mRNA structure and TIR prediction software tools may be useful, a critical analysis should be performed on their predictive outputs.

#### 2.4.4. Considerations about the pTRKH3-b mutant

Since the increased plasmid copy number of the pTRKH3-b mutant could not be explained by the RBS similarity to the 16S rRNA sequence nor by the Anderson library predicted strength neither by the mRNA secondary structures, other explanations were sought. The analysis of the sequencing results of all the six modified plasmids was already performed right after the plasmids construction, but a more in depth posterior analysis showed that all the plasmids had the correct sequence with the expected altered RBS sequences, except pTRKH3-b that presented an unexpected sequence, namely an insertion with 77 bp. The highest plasmid producer pTRKH3-b probably suffered this insertion during the site-directed mutagenesis PCR amplification. More specifically, a duplicated region consisting in the end portion of the *repDE* premature terminator, the complete RBS and the beginning of the *repD* gene was inserted 22 bp after the first nucleotide of the *repD* gene (Figure 2.10).



**Figure 2.10.** pTRKH3-b mutant sequence showing the insertion of a duplicated sequence consisting in the end portion of the *repDE* terminator, the complete RBS and the beginning of the *repD* gene.

This insertion lead to the existence of two alternative ribosome binding sites and more possible start codons. The already explained phenomenon called translational coupling, which happens when the gene expression level of the downstream gene is enhanced by the expression level of the first gene<sup>77</sup> could be the reason behind the increased plasmid copy number of the pTRKH3-b mutant. Although the protein coded by the *repD* gene has an unknown function, apparently not related with the plasmid replication or copy number control, its translation rate and start/stop site will affect the translation rate of the downstream *repE* gene<sup>95</sup>.

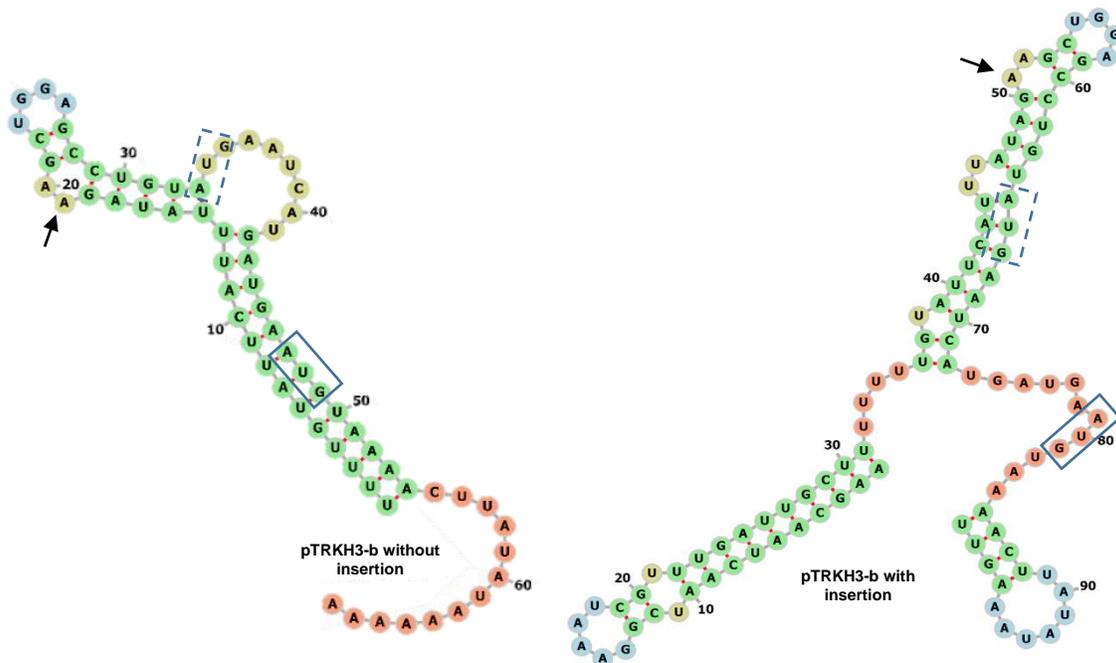
The analysis of the relative translation initiation rates helped to shed some light over these unexpected results (Table 2.4). Comparing the pTRKH3-b with and without the inserted region, the first evidence that draws our attention is the fact that the insertion created seven new alternative start codons. Although only two of the total of possible start codons (start sites 78 and 141) are *in-frame* with the *repD*, the predicted translation from the additional start codon (position 141) is two times higher than from the pTRKH3-b without insertion (start at 78). The presence of a second start site at the position 141 could lead to an increase in the *repE* expression by the translational coupling phenomenon and, consequently, to an increase in the plasmid copy number.

It is also interesting to analyze the remaining possible start codon sites, even they are not *in-frame*, because it is not essential to have a functional *repD* protein. Alternative start codons could lead to stop codons further downstream, decreasing the intergenic region distance between the stop codon of the first gene and the start codon of the second gene, increasing the probability and efficiency of the ribosome binding<sup>95</sup>. For example, the first start codon downstream the first RBS is in the position 64, having a high TIR value, but leading to a truncated version of the *repD* protein, not decreasing the distance to the *repE* gene.

**Table 2.4.** Start codon sites and translation initiation rate (TIR) of *repD* gene for pTRKH3-b mutant with and without the detected insertion. The start codons *in-frame* with the *repD* start codon are highlighted in green.

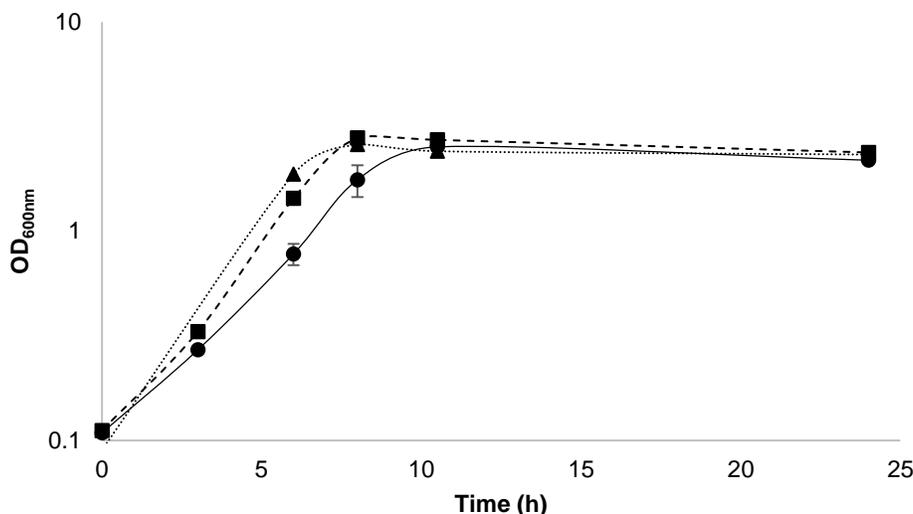
pTRKH3-b without insertion		pTRKH3-b with insertion	
Start site	<i>repD</i> TIR (A.U.)	Start site	<i>repD</i> TIR (A.U.)
21	0.4	21	0.4
25	0.8	25	0.8
34	0.7	34	0.7
64	1,472	64	2,008
71	1,642	71	1,642
74	826	74	826
78	290	78	290
		98	16
		102	17
		111	5
		141	567
		148	1,100
		151	553
		155	302

Looking directly to the mRNA secondary structures, further conclusions can be drawn from the first start codon after the RBS that is *in-frame* with the *repD* gene start codon (box with solid line, Figure 2.11). Considering that unpaired nucleotides near the start codon lead to higher expression levels<sup>93,94</sup>, the pTRKH3-b with the insertion meets this criteria, as opposed to the same mutant without the insertion. Analyzing the remaining mutants, only the pTRKH3-e has the same pattern around this specific start codon as the pTRKH3-b without the insertion. This similarity of the mRNA secondary structures between the pTRKH3-e and the pTRKH3-b without the insertion will probably be responsible for similarly low PCN values.



**Figure 2.11.** Prediction of the *repDE* mRNA secondary structures for the pTRKH3-b with and without the detected insertion. The first start codon *in frame* with the *repD* start codon is highlighted with a solid box and the first nucleotide of the hypothesis 1 consensus RBS is indicated with an arrow. The hypothesis 2 consensus RBS starts three nucleotides downstream the arrow.

The next logic step was to engineer the pTRKH3-b without the insertion, by repeating the site-directed mutagenesis procedure and confirming the absence of the insertion by sequencing. The pTRKH3-b without insertion (with only 1 RBS sequence) was then grown in parallel with the wild-type pTRKH3 and the pTRKH3-b with the insertion (with 2 RBS sequences), using the standard conditions described in the Materials and Methods section. Figure 2.12 shows the growth curves, being the specific growth rates of the pTRKH3-b with and without insertion similar, meaning that the insertion has no effect in the metabolic burden of the cells.

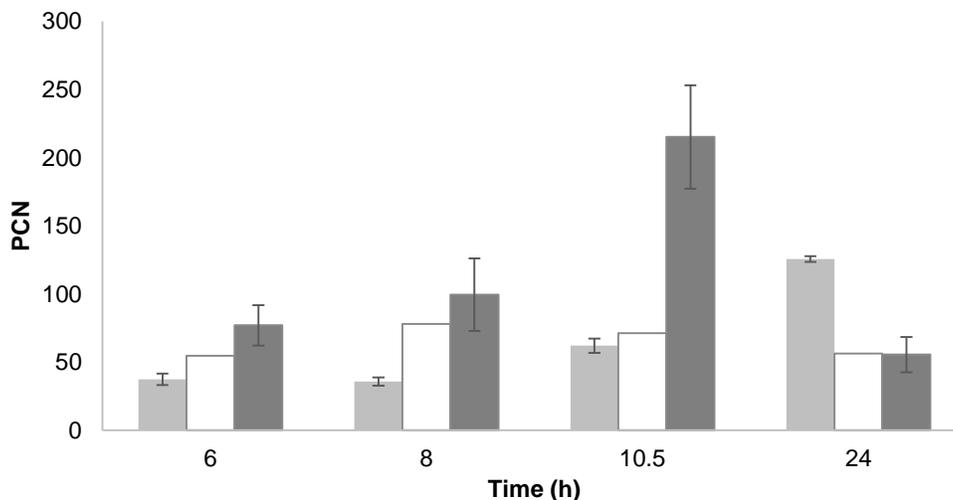


**Figure 2.12.** Growth profiles of *L. lactis* LMG19460 cells harboring the non-modified pTRKH3 (filled circle with solid line), the pTRKH3-b mutant without the insertion (filled triangle) and the pTRKH3-b mutant with the insertion (filled square), during 10.5h of growth in 100 mL shake flasks with 75 mL of M-17 supplemented with 20 g/L glucose, 30°C, 100 rpm (initial  $OD_{600nm}=0.1$ , starting from overnight inoculum). The profile was obtained from seven independent growths for the wild-type pTRKH3 and pTRKH3-b with the insertion (except for the 24h point that was the result of two independent experiments), and from one experiment for the pTRKH3-b without the insertion. The standard error of the mean (SEM) is represented for each point. The specific growth rates ( $\mu$ ) were  $0.40 (\pm 0.01)$  for pTRKH3,  $0.44 (\pm 0.02)$  for the pTRKH3-b with the insertion and  $0.51$  for the pTRKH3-b without the insertion.

The PCN analysis showed that the pTRKH3-b without the insertion did not have the same high copy number property after 10.5 hours of growth as its counterpart with the detected insertion (Figure 2.13). This means that it is the insertion that is responsible for the increase in the plasmid copy number and not solely the sequence of this specific RBS by itself. We hypothesize that it could be the joint effect of this RBS sequence with the increase in the number of RBS sites and start codons with higher TIRs that is contributing to the increase in the plasmid copy number of the pTRKH3-b mutant with the insertion.

An interesting and unexpected result from this experiment concerns the wild-type pTRKH3 that achieved a maximum value of PCN after 24 hours of growth, as opposed to pTRKH3-b with the insertion, which achieved its maximum PCN at the late exponential/early stationary phase (10.5h of growth) (Figure 2.13). But, even at the 24h point, the non-modified pTRKH3 only achieved 126 copies per chromosome, which is much lower when compared with the maximum value of 215 copies per chromosome for pTRKH3-b with the insertion after 10.5h of growth. The pTRKH3-b without the insertion achieved its peak PCN value after 8 hours of growth with 78 copies of plasmid per chromosome. When looking at the pDNA quality in an agarose gel, it is clearly visible a decrease in plasmid quality and quantity from 10.5h to 24h of growth.

In the remaining chapter, where the pTRKH3-b is designated it corresponds to the pTRKH3-b mutant with the insertion, since it is the highest plasmid producer.

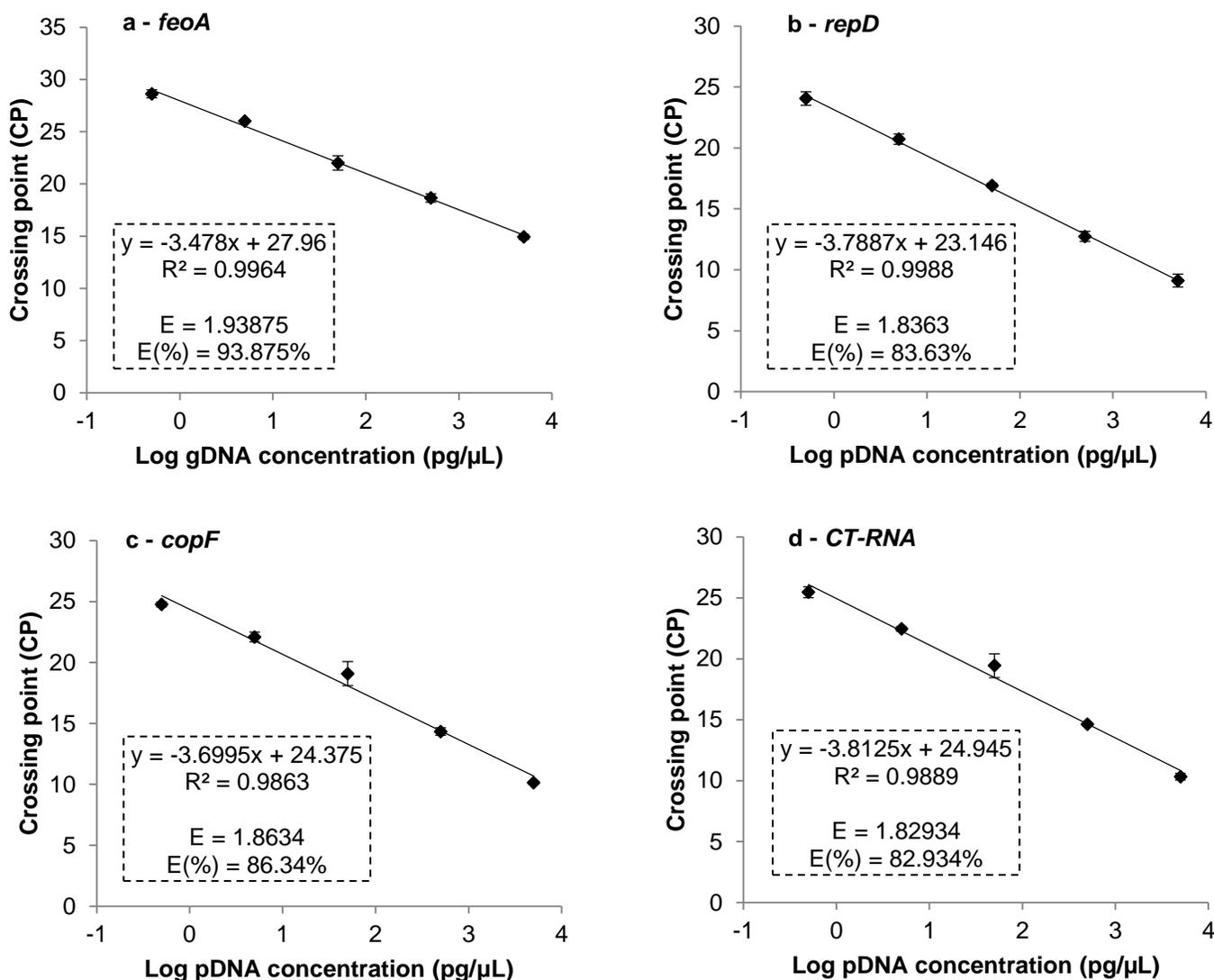


**Figure 2.13.** Average PCN values for *L. lactis* LMG19460 cells harboring the wild-type pTRKH3 (light grey), the pTRKH3-b without the insertion (white) and the highest plasmid producer pTRKH3-b with the insertion (dark grey) at different growth times (6, 8, 10.5 and 24h). The PCN results were obtained from real-time qPCR analysis of seven independent growths for the wild-type pTRKH3 and pTRKH3-b with the insertion (except for the 24h point that was the result of two independent experiments), and from one experiment for the pTRKH3-b without the insertion. The standard error of the mean (SEM) is represented for each point.

#### 2.4.5. mRNA analysis

In order to further explore the gap between the PCN values and the mRNA TIR predictions, the *repD* mRNA was quantified as well as the two transcripts involved in the control of plasmid copy number, the transcriptional repressor *copF* and the CT-RNA responsible for antisense RNA-mediated transcription attenuation, using a two-step quantification by real-time qPCR. The *feoA*, *repD*, *copF* and *CT-RNA* standard curves necessary for the upcoming mRNA quantification assays are shown in Figure 2.14. The coefficient of determination ( $R^2$ ) was at least 0.99 in all cases, indicating a high precision in the PCR performance<sup>91</sup>. The efficiencies for each gene amplification were calculated from the slope of each linear regression equation, using equation 1. The real-time qPCR amplified the *feoA*, *repD*, *copF* and *CT-RNA* standards with an efficiency of 93.9%, 83.6%, 86.4% and 82.9%, respectively.

*repD* was chosen instead of *repE*, notwithstanding the fact that RepE is the functional replication protein, because *repD* is at the 5'-end of the mRNA molecule, preventing underestimation errors that could arise from mRNA degradation. Since *repD* and *repE* are organized as an operon, a single polycistronic mRNA is transcribed, so there is no disadvantage in quantifying the *repD* instead of the *repE* gene. The correlation between *repD* and *repE* mRNA quantification was validated previously obtaining similar results for both (not shown) and from now on the mRNA will be called *repDE*.



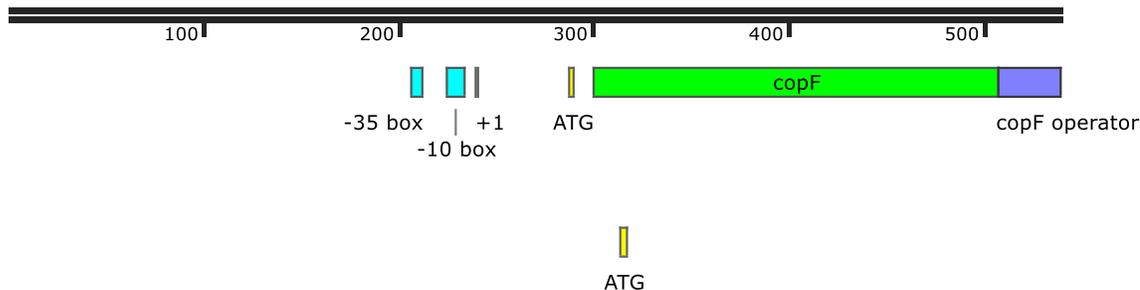
**Figure 2.14.** Real-time qPCR standard curves done with five log serial dilutions (1, 10, 100, 1,000 and 10,000 pg) of a) *L. lactis* LMG19460 gDNA using primers specific for *feoA* gene; and pTRKH3 pDNA using primers specific for b) *repDE*, c) *copF* and d) *CT-RNA* genes. The experiments were performed in duplicate, except the one with *repDE* primers that was performed in triplicate. The linear regression trendlines are represented, along with the corresponding equation, coefficient of determination value ( $R^2$ ) and efficiency values.

The results were analyzed in terms of the PCN/*repDE*, PCN/*copF* and PCN/*CT-RNA* ratios that will give information about how many *repDE*, *copF* and *CT-RNA* mRNA molecules are in the cells harboring the wild-type and the six modified plasmids when compared with the number of copies of plasmid. The absolute values are shown in Table 2.5.

**Table 2.5.** Average values of PCN and PCN from *repDE*, *copF* and *CT-RNA* mRNAs for pTRKH3 and the six modified plasmids. The PCN results were obtained from real-time qPCR analysis of seven independent growths, while the remaining categories were the result of two independent experiments. The standard error of the mean (SEM) is represented for each point.

Plasmid	PCN		PCN <i>repDE</i> mRNA		PCN <i>copF</i> mRNA		PCN <i>CT-RNA</i> mRNA	
	Average	SEM	Average	SEM	Average	SEM	Average	SEM
pTRKH3	62.0	5.24	103.8	47.11	19.6	1.86	9.8	0.31
pTRKH3-a	67	14.48	10.2	3.11	6.2	1.01	6.7	0.35
pTRKH3-b	215.3	37.93	35.0	7.47	14.5	0.55	13.4	5.38
pTRKH3-c	117.4	20.78	45.2	11.74	19.9	1.53	15.8	7.41
pTRKH3-d	107.9	9.67	103.0	2.53	22.6	4.62	18.6	10.29
pTRKH3-e	76.3	5.04	4.2	0.53	1.9	0.19	2.0	0.04
pTRKH3-f	110.1	4.66	23.8	2.59	8.2	0.63	7.3	0.98

In pTRKH3, the *copF* gene is incomplete, missing its promoter and part of the ORF, due to a transposon insertion<sup>96</sup> occurred in the parental plasmid (pMT9). Nevertheless, the *copF* mRNA was analyzed, since the sequence has a predicted promoter<sup>97</sup> (Softberry BROM program (<http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb>)<sup>98</sup> region in the beginning that could allow the expression of a partial protein that could still act as a repressor. Two putative start codons could allow the translation of a longer truncated CopF version with 77 amino acids (72 of which in common with the 83 amino acids of the original CopF protein) or a shorter truncated CopF version with 68 amino acids (all in common with the original CopF protein) (Figure 2.15). It is important to consider that this pAM $\beta$ 1 derivative with a transposon insertion was shown to have a PCN ~20-fold higher when compared with the parental plasmid with the complete *copF* gene<sup>96</sup>, meaning that if a truncated version exists, it is probably produced in lower quantity and/or have lower activity.



### Original CopF (83 aa)

MRGTKSKEKFSQELEMSRSNYSRIESGKSDPTIKTLEQIAKLTNSTLVVDLIPNEPTEPEPETESEQVTLDLEM  
EEEKSNDFV\*

### Longer truncated CopF (77 aa – 72 aa in common with original CopF)

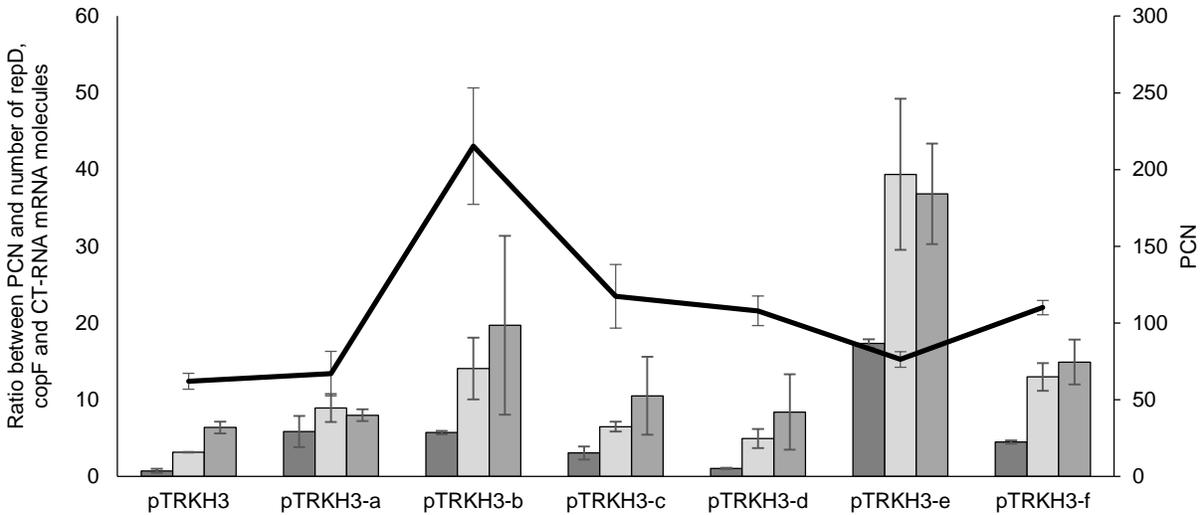
MLAVPQELMSRSNYSRIESGKSDPTIKTLEQIAKLTNSTLVVDLIPNEPTEPEPETESEQVTLDLEMEEEEKS  
NDFV\*

### Shorter truncated CopF (68 aa)

SRSNYSRIESGKSDPTIKTLEQIAKLTNSTLVVDLIPNEPTEPEPETESEQVTLDLEMEEEEKSNDFV\*

**Figure 2.15.** Map of *copF* gene in the pAMβ1 origin of replication of pTRKH3 plasmid. The promoter predicted with Softberry BPROM program (<http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb>) (Solovyev and Salamov 2011) is indicated in light blue, while the two putative start codons are identified in yellow. The translation of the original *copF* gene results in a protein with 83 amino acids, while the translation of the *copF* present in the pTRKH3 plasmid could result in two putative truncate versions of CopF. The longer version has 77 amino acids, 72 of which are in common with the original CopF (the beginning of the common sequence is highlighted in green, while the different amino acids are in red). The shorter truncated version only have 68 amino acids, all in common with the original CopF, starting in the sequence identified in orange.

The highest PCN/*repDE* value was obtained for the pTRKH3-e plasmid, having 17-fold more copies of plasmid than copies of *repDE* mRNA (Figure 2.16), which means that the transcription efficiency for *repDE* was very low. Ultimately, this could be responsible for a decreased *repE* expression and consequently for the low number of plasmid copies obtained for this modified plasmid. Analyzing the PCN/*copF* and PCN/*CT-RNA* values, they were again the highest ratio values obtained when compared with the other plasmids. pTRKH3-e had 39-fold and 37-fold more copies of plasmid than copies of *copF* and *CT-RNA*, respectively (Figure 2.16). Probably it was happening the copy number control mechanism described above that when the PCN is too low, CopF concentration is accordingly lower and the convergent transcription from the P<sub>CT</sub> promoter is avoided. If more time were given, it could had happen an increase in the PCN that was not detected at 10.5h of growth.

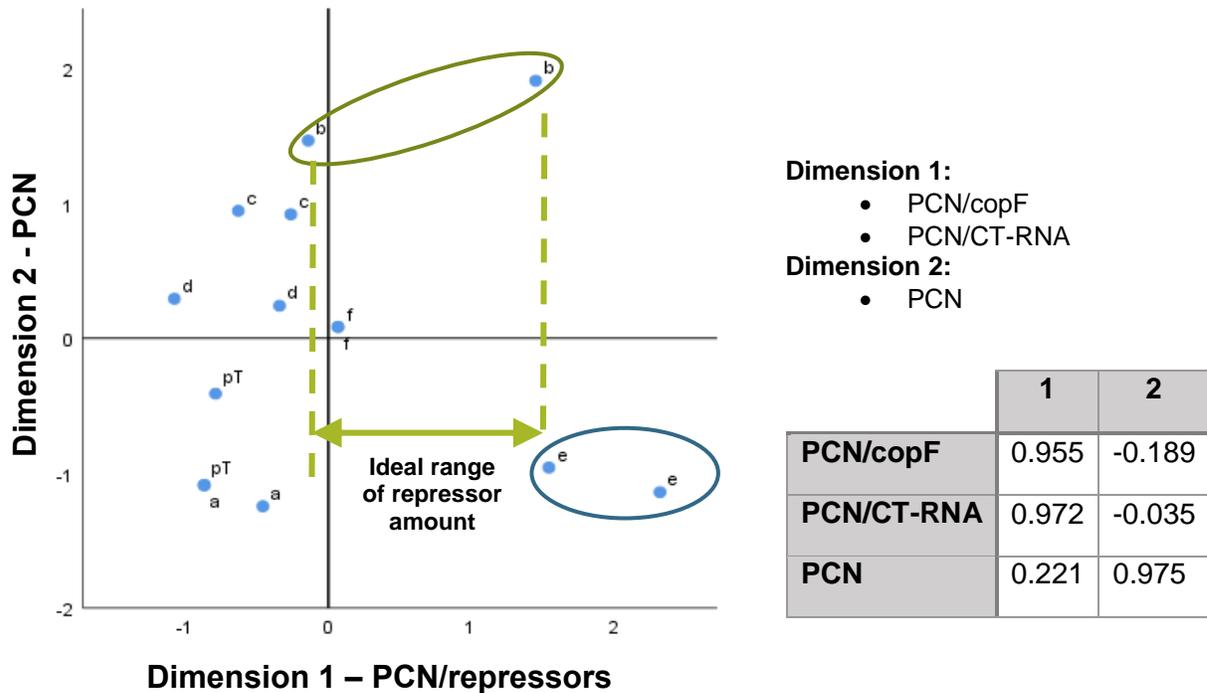


**Figure 2.16.** Average PCN values and average ratio values between PCN and number of *repDE* (dark grey columns), *copF* (light grey columns) and *CT-RNA* (medium grey columns) mRNA molecules for *L. lactis* LMG19460 cells harboring the wild-type and the six modified plasmids. The results were obtained from two-step mRNA real-time qPCR quantification of two independent growths in 100 mL shake flasks with 75 mL of M-17 supplemented with 20 g/L glucose, 30°C, 100 rpm (initial OD<sub>600nm</sub>=0.1, starting from an overnight inoculum). The PCN data (black line) is the average of seven independent experiments. The standard error of the mean (SEM) is represented for each point.

The plasmid pTRKH3-b was the one with the highest PCN value and shows high values of each ratio, when compared with the wild-type and the other modified plasmids. It had 6-fold (third highest), 14-fold (second highest) and 20-fold (second highest) more copies of plasmid than *repDE*, *copF* and *CT-RNA* mRNA (Figure 2.16). The transcription efficiency of the three sequences seems much higher in this plasmid, when compared with pTRKH3-e, which contributed to its high number of plasmid copies. The moderately high quantities of *copF* and *CT-RNA* are probably explained by the same mechanism than pTRKH3-e, with the only difference that we were able to detect the expected increase in the PCN value.

Considering the wild-type pTRKH3, it had the lowest values for all the ratios, having 3-fold and 6-fold more copies of plasmid than *copF* and *CT-RNA* mRNA (Figure 2.16), implying that both regulators were present in high quantities inside the cells. This is contradictory with the very low value of the PCN/*repDE* ratio (0.72), which means that there were more copies of *repDE* mRNA than copies of plasmid that is in accordance with the absolute PCN values (Table 2.5). Therefore, the *repDE* transcription did not appear to be repressed, because the mRNA was being produced, but indeed this was not reflected in the PCN value. Although the intermediate concentrations of *copF* and *CT-RNA* mRNA inside the cells must have allowed an increased quantity of *repDE* mRNA (Table 2.5), something could be happening at the translation level, related or not with the RBS strength, that prevents the increase in the number of plasmid copies.

A principal component analysis was performed on these results in order to make it easier to visualize and interpret them (Figure 2.17). The first principal component, which accounts for the largest possible variance, is explained mainly by the variables PCN/copF and PCN/CT-RNA that had the highest eigenvalues. This first dimension will be identified as PCN/repressors. The second dimension is explained by the PCN alone.



**Figure 2.17.** Principal component analysis of the mRNA quantification results and the associated *eigenvalues*.

Figure 2.17 shows clearly that exists an ideal range of repressor amount, that leads to the highest PCN of the pTRKH3-b mutant. Since both repressors are constitutively expressed, if its amount is too high, as happens with pTRKH3-d, the PCN will immediately decrease because both CopF and CT-RNA are repressing the transcription from the  $P_{DE}$  promoter<sup>76</sup>.

When the regulators amount is too low that can only means that the PCN is correspondingly low, expressing only small amounts of each repressor, as happens with pTRKH3-e.

The intermediate amount of CopF and CT-RNA repressors present on the pTRKH3-b seems to allow the perfect balance between transcription initiation and its repression to achieve the highest PCN, together with the presence of two alternative RBS sequences and several alternative start codons.

## 2.5. Conclusions

Although a low-copy plasmid could be indispensable when the goal is to express a protein that in high amount is toxic for the cells or repress an important pathway, a high-copy plasmid is also required when high amounts of a certain

heterologous protein coded in the plasmid sequence or even the plasmid itself is needed. If the final purpose is to produce DNA vaccines using food-grade LAB strains, a high copy-number shuttle plasmid is economically appealing for industry, since the pDNA purification yield increases, adding to the fact that the LPS-free status of LAB itself contributes for reducing the purification costs.

With that goal in mind, several site-directed mutagenesis were performed in the RBS controlling the expression of the *repDE* genes of the pAM $\beta$ 1 replication origin of the shuttle plasmid pTRKH3. The copy number analysis of each modified plasmid was performed by real-time qPCR and compared with the wild-type pTRKH3. The pTRKH3-b exhibited a 3.5 fold increase in the number of copies in *L. lactis* LMG19460 strain, having an average of 215 copies per chromosome, comparatively to the non-modified plasmid that has an average of only 62 copies per chromosome.

The analysis of the optical density and pH profiles during 10.5h of growth showed that these modifications in the RBS sequences apparently did not interfere with the overall cell metabolism.

Considering the mRNA analysis, the high copy number mutant pTRKH3-b had 6-fold more copies of plasmid than *repDE* mRNA, but had the second highest values of the PCN/*copF* (14) and PCN/*CT-RNA* (20) ratios, meaning that there were much more copies of plasmid than copies of *copF* and *CT-RNA* mRNA. What probably happened is that the pAM $\beta$ 1 copy number control systems ensure that when PCN is turning too low, CopF concentration is lower, allowing transcription from the P<sub>DE</sub> promoter, but avoiding the convergent transcription from the P<sub>CT</sub> promoter, reducing the CT-RNA concentration and increasing PCN<sup>76</sup>. The principal component analysis showed that the ideal is for the cells to have an intermediate amount of repressors present, in order to have higher PCN, such as happens with pTRKH3-b.

The sequencing results helped to clarify further why pTRKH3-b had such a high number of copies of plasmid per chromosome. These results showed an insertion in the pTRKH3-b sequence 22 bp downstream the first nucleotide of the *repD* gene. This insertion was indeed a duplicated region consisting in the end portion of the *repDE* terminator, the complete RBS and the beginning of the *repD* gene, meaning that the pTRKH3-b had an additional RBS sequence and seven new start codons, when compared with the same mutant without the insertion. The only additional *in-frame* start codon had a higher TIR and a more favorable mRNA secondary structure than the original *in-frame* start codon already present in the pTRKH3-b without the insertion. When looking at the PCN quantification results, the pTRKH3-b without the insertion did not showed the same increase in the PCN after 10.5h of growth as its counterpart with the insertion, meaning that this duplicated region must be contributing to increase the PCN.

Therefore, pTRKH3-b is a very promising high copy number shuttle plasmid that will contribute to bring LAB to the same level as *E. coli* as DNA vaccines producers, with the additional advantage of being endotoxin free.

# CHAPTER 3

## Artificial plasmids for replication in LAB

*What I cannot create, I do not understand.* – **Richard Feynman**

### 3.1. Abstract

The hierarchical characteristic of biology is what makes it possible to begin with small parts or BioBricks (sequences of DNA coding for a definable function) that could be assembled in complex new systems. The majority of the preexistent BioBricks and vector backbones were designed to work in *Escherichia coli*, yet some research has been performed in Gram-positive bacteria. More specifically in LAB, those studies focused only on the nisin biosynthesis pathway and were never applied to such a complex goal as the one proposed in this chapter. Here, several essential and optimized parts were synthesized and assembled in an artificial high-copy number plasmid, which would allow an efficient LAB protein expression, namely antigens for DNA vaccines or enzymes for the production of novel therapeutic metabolites.

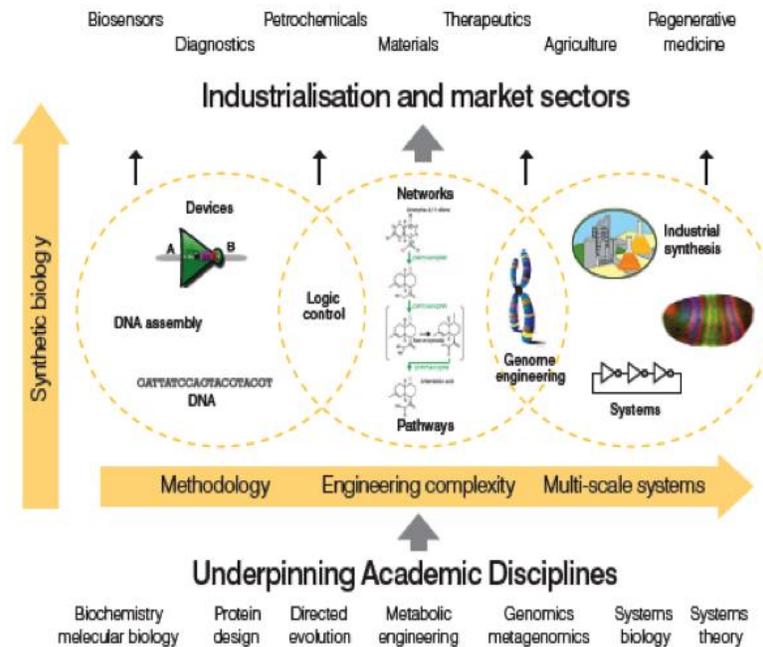
The plasmids assembled by Gibson Assembly technique were based on the already available *E. coli*/Gram-positive shuttle plasmid, pTRKH3, and should contain the following BioBrick compatible parts: an *E. coli* cassette (useful for the initial steps of molecular cloning), a Gram-positive cassette and two protein expression cassettes, one for prokaryotic applications such as using LAB for producing proteins or metabolites, and the other when the goal is the expression in eukaryotic cells as it happens in DNA vaccination. The *E. coli* and the Gram-positive cassettes should include appropriate antibiotic resistance genes and replication origins to allow plasmid maintenance in each type of bacteria.

Until now, the *E. coli* and the Gram-positive cassettes (with both wild-type pAM $\beta$ 1 and the high PCN pAM $\beta$ 1-b origins of replication) were successfully assembled. The plasmid copy number (PCN) quantification by real-time quantitative PCR showed a decrease in the PCN values when compared with the original pTRKH3 and pTRKH3-b plasmids, probably due to recombination events detected during assembly confirmation. The assembly of a LAB protein expression cassette with the 3 genes of the curcumin pathway under the control of the SLP promoter were attempted, without success. Further efforts must be made to accomplish this assembly, in order to increase plasmid size, hoping to avoid the recombination events.

Several alternatives of future work are discussed in the end of the present chapter.

### 3.2. Introduction

Synthetic biology could be defined as “a) the design and construction of new biological parts, devices and systems and b) the re-design of existing natural biological systems for useful purposes”<sup>99</sup>. It has several possible applications in different areas, including therapeutics: disease modeling, drug discovery and production, vaccine development, treatment of infectious diseases and cancer<sup>100</sup> (Figure 3.1).



**Figure 3.1.** Overview of the disciplines that contribute for the Synthetic Biology field and its several applications<sup>101</sup>. Figure from Freemont and Kitney (2017)<sup>101</sup>.

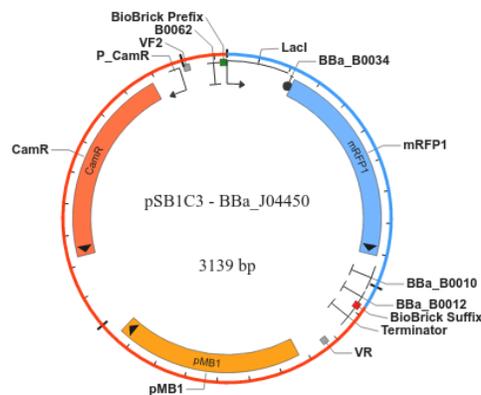
Synthetic biology is based in standard parts, which are standardized sequences of DNA that code for specific biological functions and cannot be divided further into smaller parts. These parts are functional units that could exist in nature or been *de novo* synthesized, and posteriorly assembled in biological devices or systems<sup>102</sup>. Examples of parts are promoters, ribosome binding sites (RBS), terminators, protein coding regions, etc., and these parts could be assembled together in new, more complex and longer parts by using Assembly Standards. The Registry of Standard Biological Parts has an online library with thousands of available ready-to-use standardized parts, which had been tested and characterized.

There are several different Assembly Standards, but all meet three criteria: ensure that the parts are compatible, define how they should be assembled and guarantee that newly composed parts are still compatible with future assemblies. The plasmid backbones used on Assembly Standards must have a prefix and a suffix, upstream and downstream each part, respectively. The prefix and suffix contain restriction enzyme sites and when the goal is to assemble a new part, is just necessary to digest with the appropriate enzymes and connect with a DNA ligase. The new longer part will still maintain the prefix and suffix, allowing for more new assemblies. A big disadvantage of these Assembly Standards is

that for the part to be compatible with them, it cannot have in its sequence the restriction enzyme sites from the prefix and suffix, which limits the number of parts that can be used with each Standard. These methods also leave a scar in the place where DNA ligase acts, but there are also scarless methods, which enables assembly without necessity for prefix and suffix restriction sites<sup>102</sup>.

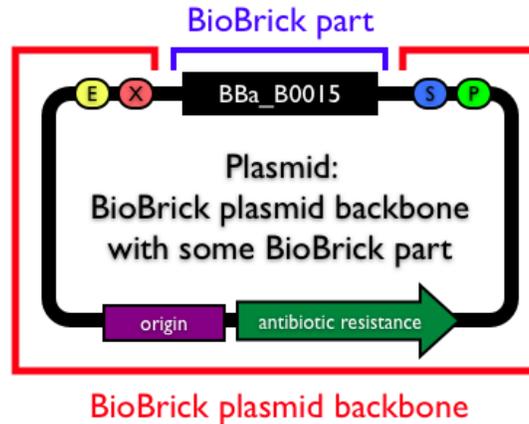
The parts available on the Registry of Standard Biological Parts are compatible with at least one of the five Registry supported Assembly Standards, which are RFC[10] (the BioBrick Standard), RFC[12] (BioBrick BB-2 Standard), RFC[21] (Berkeley Standard), RFC[23] (Silver Standard) and RFC[25] (Freiburg Standard)<sup>102</sup>. All the available parts are in the pSB1C3 plasmid backbone (Figure 3.2), which itself is compatible only with the RFC[10] Assembly Standard that is the Registry's current *de facto* standard. If compatibility with a different standard is necessary, the part can be moved to other plasmid backbone. pSB1C3 is a high copy number plasmid, carrying the pUC19-derived pMB1 origin of replication (100-300 copies of plasmid per cell), and the chloramphenicol resistance gene<sup>102</sup>.

The BioBrick RFC[10] Standard was developed by Tom Knight in 2007 and is the most commonly used assembly standard, simple the majority of the parts and plasmid backbones available are compatible with this assembly standard. For protein coding regions, this standard uses an alternative shorter prefix to allow for optimum spacing between the Ribosome Binding Site (RBS) and Coding DNA Sequence (CDS), leaving a 6 bp scar instead of 8 bp. To assemble protein fusions, this is not a useful standard, because the traditional 8 bp scar creates a frame-shift, while the 6 bp creates a stop codon, being preferable a scarless assembly method or DNA synthesis<sup>102</sup>.

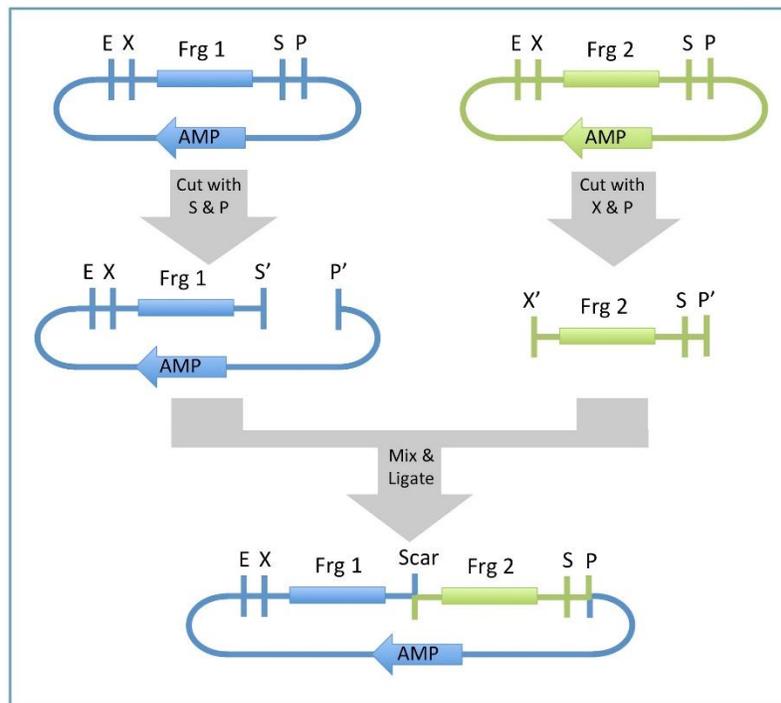


**Figure 3.2.** pSB1C3 plasmid backbone used to harbor the parts available at the Registry of Standard Biological Parts<sup>102</sup>. Figure from Registry of Standard Biological Parts<sup>102</sup>.

The prefix in BioBrick RFC[10] Standard is constituted by EcoRI, NotI and XbaI, while the suffix is constituted by SpeI, NotI and PstI. The general structure of a plasmid backbone with a BioBrick part compatible with this standard is represented in Figure 3.3. To join two parts together, the first part must be digested with SpeI and PstI, while the second part is digested with XbaI and PstI (Figure 3.4). After ligation, the assembly of two different parts leaves the following 8 bp scar: 5' [partA] TACTAGAG [part B] 3'. When using the shorter prefix for the assembly between a RBS and a CDS, the resulting 6 bp scar is the following: 5' [part A] TACTAG [part B] 3'<sup>102,103</sup>.



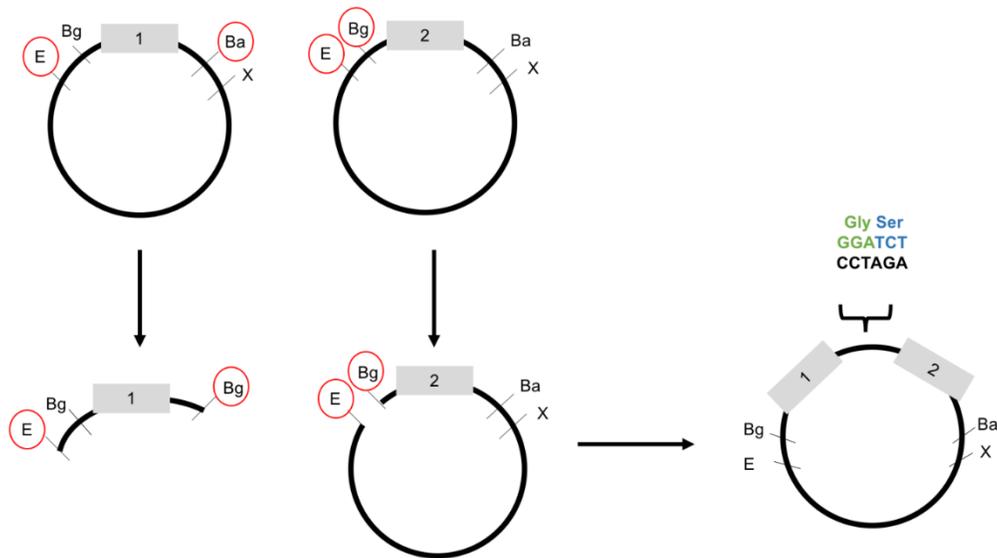
**Figure 3.3.** General structure of a plasmid backbone with a BioBrick part compatible with BioBrick RFC[10] Standard<sup>102</sup>. Figure from Registry of Standard Biological Parts<sup>102</sup>.



**Figure 3.4.** Assembly method used in the BioBrick RFC[10] Standard<sup>102</sup>. Figure from Registry of Standard Biological Parts<sup>102</sup>.

The BioBrick BB-2 RFC[12] Standard also created by Tom Knight in 2008 has an improvement when compared with the RFC[10] Standard, since it allows for *in-frame* protein assembly. The prefix is composed by EcoRI, NotI and SpeI, while the suffix has NheI, NotI and PstI. In order to join two different parts, the first one must be digested with EcoRI and NheI, and the second with EcoRI and SpeI. After the ligation of the two parts, the 6 bp scar (5' [part A] GCTAGT [part B] 3') codes for two amino acids (alanine and serine), instead of a stop codon, as happened in the RFC[10] Standard<sup>102</sup>.

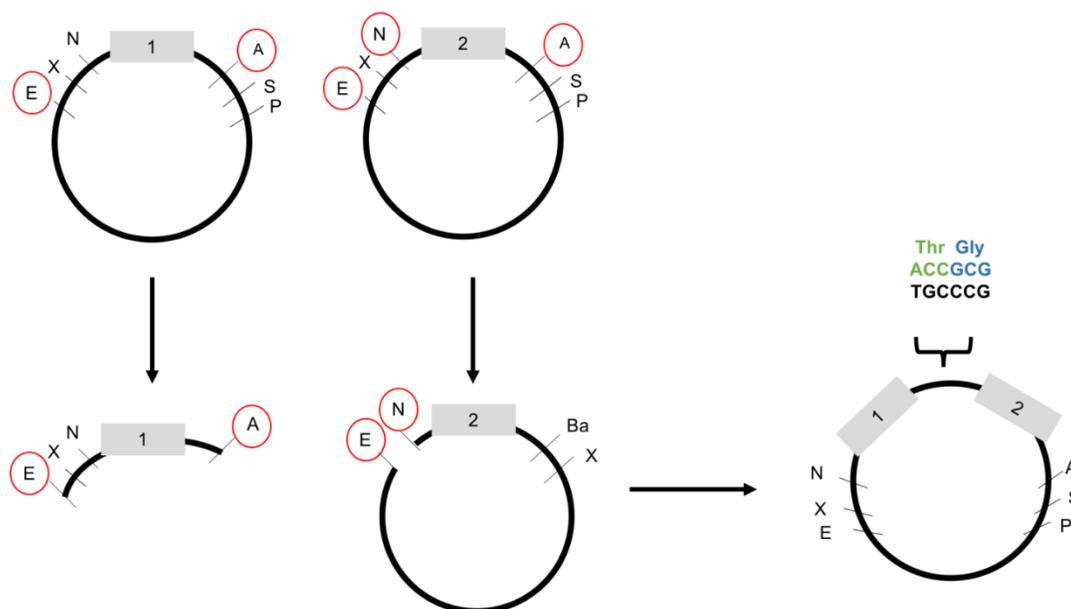
The Berkeley RFC[21] Standard is quite similar to the RFC[12] Standard, also allowing for *in-frame* protein parts assembly, with the only difference being in the enzymes used in the suffix and prefix. The enzymes in the prefix are EcoRI and BglII and in the suffix are BamHI and XhoI. To join two parts, the first one should be digested with EcoRI and BamHI (or BglII and XhoI), while the second part must be digested with EcoRI and BglII (or BamHI and XhoI), resulting in the following scar: 5' [part A] GGATCT [part B] 3' (Figure 3.5). This scar codes for a glycine and a serine. It was created in 2009 by Christopher Anderson *et al.* and is also known as BglBricks Assembly Standard. Its main disadvantage is that it uses the BglII enzyme that is not heat-inactivated, preventing the use of the 3A Assembly method<sup>102,104,105</sup>.



**Figure 3.5.** Assembly method used in the Berkeley RFC[21] Standard<sup>104</sup>.

The Silver RFC[23] Standard also allows *in-frame* assembly of proteins parts, but leaves a 6 bp scar, including a codon for arginine, which is known as being destabilizing and contribute to a faster protein degradation. It was created by Ira Philips and Pamela Silver in 2006, as a modification of the RFC[10] Standard. The prefix includes the enzymes EcoRI, NotI and XbaI, while the suffix includes SpeI, NotI and PstI. The assembly method is the same as the one for RFC[10] Standard. When compared with RFC[10], both prefix and suffix have 1 less bp each, allowing for the 6 bp scar (coding for a threonine and an arginine). These missing spacer nucleotides prevented Dam DNA methylation in RFC[10], so in RFC[23] methylation could be a problem<sup>102</sup>.

Finally, the Freiburg RFC[25] Standard was proposed by the Freiburg team in 2007 as an alternative to RFC[23] and is also compatible with RFC[10] Standard. Besides the enzymes already present in RFC[10], the prefix has a start codon and a NgoMIV restriction site, while the suffix starts with an AgeI restriction site followed by a stop codon. To join two parts together, the same method of RFC[10] Standard can be used here, but results in a longer scar sequence. In alternative, to assemble fusion proteins, the first part can be digested with EcoRI and AgeI, while the second part with EcoRI and NgoMIV, forming a 6 bp scar coding for a threonine and a glycine (5' [part A] ACCGGC [part B] 3') (Figure 3.6)<sup>102</sup>.

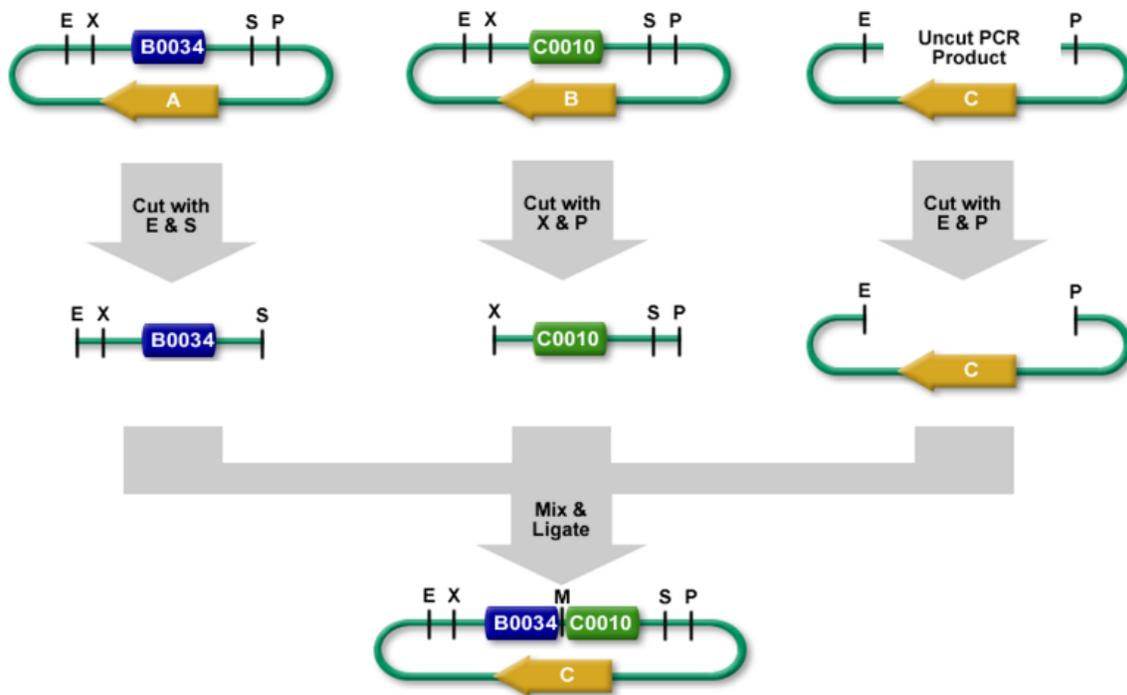


**Figure 3.6.** Assembly method used in the Freiburg RFC[25] Standard<sup>104</sup>.

Within each Assembly Standard, the parts could be assembled together using different assembly methods/technologies. The most widely used is called Standard Assembly and is the one that had been described for each Assembly Standard, using normal cloning techniques and making the appropriate digestions in the prefix and suffix. The resulting insert and vector must be purified from an agarose gel and ligated, regenerating again the complete prefix and suffix, which allows for successive assemblies. Since gel purification many times has a low yield, leading to ligation low success rate, the Registry of Standard Biological Parts no longer recommends the use of this Assembly method<sup>102,103</sup>.

The 3 Antibiotic (3A) Assembly method was developed as an alternative to the Standard Assembly method and also uses the restriction sites on prefix and suffix for assembling different parts, and the new composite part will still have the same prefix and suffix, as well as an additional scar. But it has a difference in the selection for the correct assemblies, which is performed using antibiotic selection, instead of gel purification of the digested parts and colony PCR of the colonies resulting from the ligation step. The antibiotic selection removes the background colonies and around 97% should have the desired assembly. This assembly method is compatible with RFC[10], RFC[12], RFC[23] and RFC[25] Standards. Explaining in a brief manner, the protocol involves digesting the first part with EcoRI and

SpeI, the second part with XbaI and PstI and a PCR amplified linear backbone plasmid with EcoRI and PstI. Each plasmid has a different antibiotic resistance. In the end, all restriction reactions are heated to inactivate the restriction enzymes. A ligation reaction is performed with equimolar quantities of the 3 digestion products. In this ligation reaction, the first part SpeI's overhang should ligate with the second part Xba's overhang, resulting in a scar not digested with any of the intervening enzymes. This new composite part should ligate to the EcoRI and PstI sites of the linear plasmid backbone (Figure 3.7). The ligation is transformed into cells and grown in the presence of the antibiotic C, which guarantees that only the cells with the correct assembly survive<sup>102,106</sup>.



**Figure 3.7.** Overview of the 3 Antibiotic (3A) Assembly method<sup>102,106</sup>. Figure from Registry of Standard Biological Parts<sup>102</sup>.

Like the previous method, the type IIS Assembly methods (e.g. MoClo, Golden Gate Assembly) are also based in restriction enzymes. But it has a big difference from the 3A Assembly method that concerns the type of restriction enzymes used. Type IIS restriction enzymes like BsaI cut the DNA outside of its recognition site, allowing for the creation of custom overhangs and consequently allowing for the assembly of multiple parts at the same time<sup>102</sup>.

Besides methods that rely in restriction enzymes, there are assembly methods that does not use them and are considered scarless methods, which are useful for assembling proteins. The big advantage of these methods is that parts coming from backbones compatibles with different standards can be assembled together. One of the most used scarless methods is DNA synthesis, which can be used not only to create new parts, but also to assemble them.

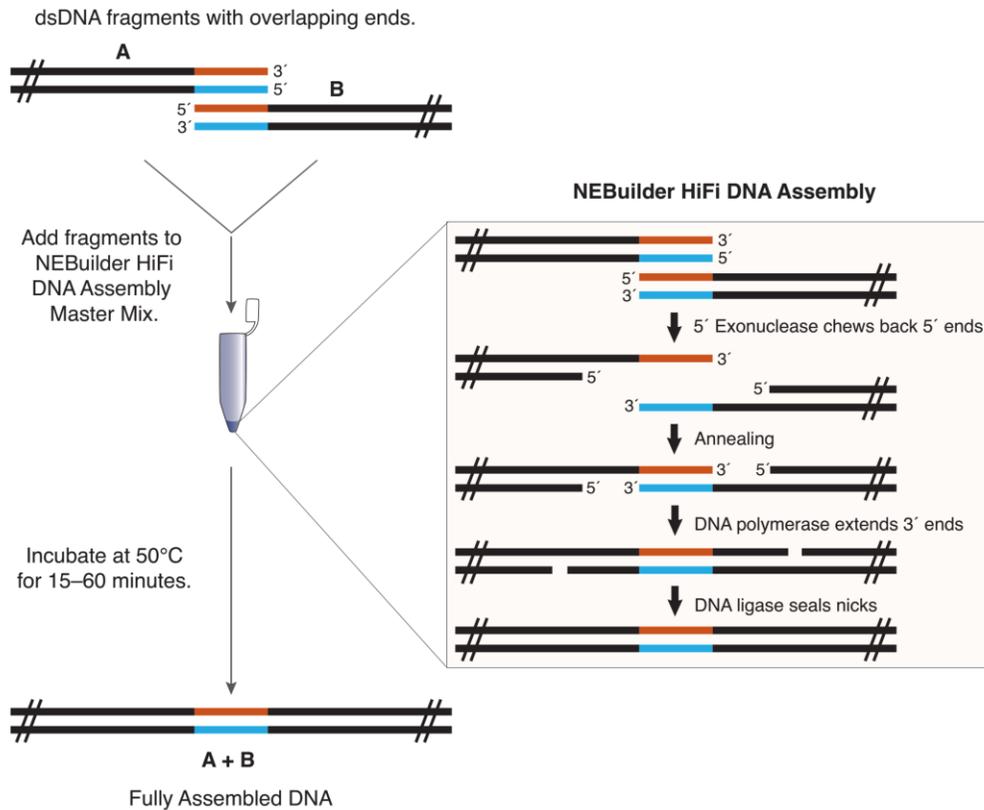
An alternative method is the Gibson Assembly that was the one chosen to be used in this chapter. Gibson Assembly is currently being evaluated by the Registry of Standard Biological Parts in order to have resources for this method available soon. One of the main advantages of this method is the fact that it does not rely in restriction endonucleases

or T4 ligase reactions to join the DNA molecules (which can range from 100 bp to 100 kb), avoiding the need for compatible restriction enzyme recognition sites<sup>107</sup>. Just by using a single isothermal step performed in a standard thermocycler, this *in vitro* recombination system is able to assemble and repair overlapping DNA molecules, starting with only nanograms amounts of input DNA. All the necessary enzymes and reagents are commercially available, turning this into a very easy and accessible method<sup>107</sup>.

The DNA sequences to be assembled must overlap with each other (20-150 bp). To create these overlapping regions between BioBrick parts, custom PCR primers are used that usually overlaps 20 bp of each neighboring BioBrick part. The fragments must be amplified using a high-fidelity DNA polymerase. The Gibson Assembly enzyme mix contains a T5 exonuclease, a DNA polymerase with proofreading abilities and a DNA ligase, all simultaneously active in the same isothermal reaction<sup>107</sup>. The T5 exonuclease removes the nucleotides from the 5' ends of linear double-stranded fragments, creating single-stranded regions that allow the two parts to anneal each other, with the advantage that this enzyme does not compete with the DNA polymerase activity. The DNA polymerase extends the 3' ends, closing the gaps between the two parts, and the DNA ligase seals the nicks<sup>107,108</sup> (Figure 3.8). The T5 exonuclease is heat-labile, being inactivated during the 50°C reaction, avoiding the digestion of all the DNA molecules<sup>109</sup>.

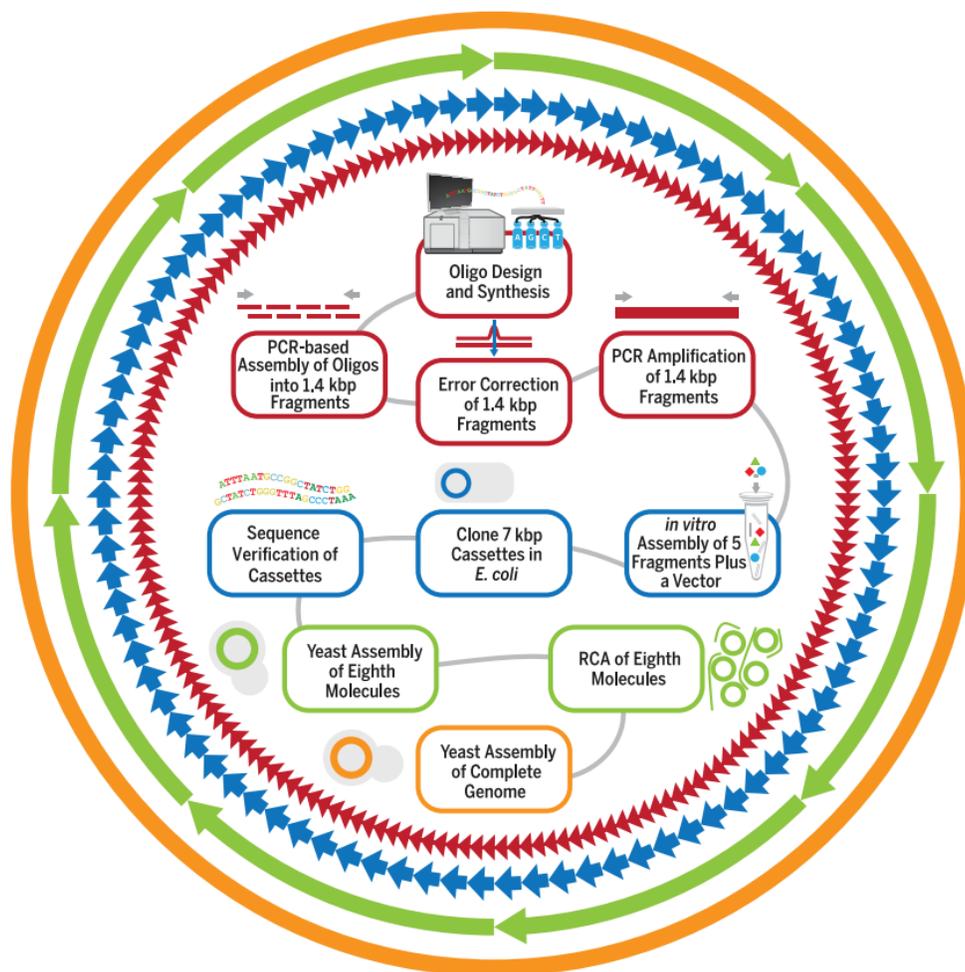
No restriction site scar will remain in the end and more than 10 different DNA fragments can be assembled at the same time in one single reaction, turning this methods cheaper, faster, and more efficient and accurate than conventional cloning and restriction enzymes-based methods. The vector could be linearized by restriction digestion or PCR amplified with a high-fidelity DNA polymerase. The DNA fragments and the linearized vector are mixed with the Gibson Assembly Master Mix containing the three necessary enzymes and incubated at 50°C for 15 min to 1 hour, depending on the number of fragments. The resulting assembly product could be used to transform directly bacterial cells<sup>107,108</sup>.

One alternative to Gibson Assembly is the SOEing PCR, a method where overlapping DNA fragments can be assembled during a PCR reaction, just by using primers complementary to the ends of the assembly region. Nevertheless, this method does not allow the assembly of large fragments, as opposed to Gibson Assembly that can join fragments up to 100 kb<sup>109</sup>.



**Figure 3.8.** Gibson Assembly method overview<sup>104,108</sup>. Figure from NEB (2017)<sup>108</sup>.

Several important accomplishments had been achieved using the Gibson Assembly technique. As the best example, a minimal 531 kb genome with 473 genes, based on the genome from *Mycoplasma mycoides*, was designed and the oligonucleotides were chemically synthesized and assembled using Gibson Assembly<sup>110</sup>. The overlapping oligonucleotides were first synthesized and assembled into 1.4 kb fragments. Each five of these fragments were then PCR amplified and assembled into 7 kb cassettes, which in turn was assembled into eight molecules. These eight molecules were then assembled in yeast in order to generate the complete genome (Figure 3.9). After transplantation of this assembled genome to the cytoplasm, it was able to produce a viable cell<sup>110</sup>. This whole-genome assembling clearly exemplifies the important applicability and the extreme efficiency of the Gibson Assembly method in cutting-edge Synthetic Biology applications.



**Figure 3.9.** Overview of the strategy applied to assemble a minimal genome (whole-genome synthesis), using the Gibson Assembly technique. Figure from Hutchison *et al.* (2016)<sup>110</sup>.

### 3.3. Materials and methods

#### 3.3.1. Bacterial strains and plasmids

*Escherichia coli* DH5 $\alpha$  (Invitrogen), *E. coli* GM2163 (New England Biolabs) or *E. coli* SCS110 (Stratagene) were used for all the molecular cloning steps performed in this chapter and the resulting plasmids were transferred for *L. lactis* LMG19460 by electroporation (BCCM Culture Collection, Belgium). Both *E. coli* GM2163 and SCS110 were used because of its ability to generate pDNA free from Dcm and Dam methylation, meaning that the DNA can be digested by methylation-sensitive restriction enzymes and LAB transformation efficiency could be improved.

The template plasmid for the several parts used in the Gibson Assembly was the pTRKH3 (Figure 1.11F) (BCCM/LMBP Plasmid Collection, Belgium), a shuttle cloning vector able to replicate both in *E. coli* and several Gram-positive genera (*Lactococcus*, *Enterococcus*, *Streptococcus* and *Lactobacillus*). To make that possible, pTRKH3

harbors two distinct replication origins, p15A with a medium copy number (30-40 copies per cell) in *E. coli* and pAM $\beta$ 1 with a high copy number (45-85 copies per cell) in Gram-positive hosts, as well as two different selection markers, namely a tetracycline resistance gene for selection in *E. coli* and an erythromycin resistance gene that allows selection in both taxonomic groups<sup>65</sup>.

The characteristics of all the strains and plasmids used are summarized in table 3.1.

**Table 3.1.** Main characteristics of the strains and plasmids used in this chapter.

Strains	Relevant genotype	Source
<i>E. coli</i> DH5 $\alpha$	F- $\Phi$ 80 <i>lacZ</i> $\Delta$ M15 $\Delta$ ( <i>lacZYA-argF</i> ) U169 <i>recA1 endA1 hsdR17</i> (rK-, mK+) <i>phoA supE44 <math>\lambda</math>- thi-1 gyrA96 relA1</i>	Invitrogen
<i>E. coli</i> GM2163	<i>ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2</i> <i>galT22 mcrA dcm-6 hisG4 rfbD1</i> <i>R(zgb210::Tn10)TetS endA1 rpsL136</i> <i>dam13::Tn9 xylA-5 mtl-1 thi-1 mcrB1 hsdR2</i>	New England Biolabs (ER2925)
<i>E. coli</i> SCS110	<i>rpsL (Strr) thr leu endA thi-1 lacY galK galT</i> <i>ara tonA tsx dam dcm supE44 <math>\Delta</math>(lac-proAB) [F'</i> <i>traD36 proAB lacIq Z<math>\Delta</math>M15]</i>	Stratagene (200247)
<i>L. lactis</i> LMG19460	Plasmid-free strain; isolated from starter cultures of German cheese factories	LMG/BCCM
Plasmids	Construct	Source
pTRKH3	P15A <i>ori</i> , pAM $\beta$ 1 <i>ori</i> , Tet <sup>R</sup> , Ery <sup>R</sup>	LMBP/BCCM (LMBP 4462)
pTRKH3-b	P15A <i>ori</i> , pAM $\beta$ 1 <i>ori</i> (mutant b), Tet <sup>R</sup> , Ery <sup>R</sup>	Chapter 2
pBio_ery_p15A pBio_erm/ery_p15A	P15A <i>ori</i> , Ery <sup>R</sup>	<sup>111</sup>
pBio_ery_p15A_pAM $\beta$ 1_WT pBio_erm/ery_p15A_pAM $\beta$ 1_WT_1 pBio_erm/ery_p15A_pAM $\beta$ 1_WT_2	P15A <i>ori</i> , pAM $\beta$ 1 <i>ori</i> , Ery <sup>R</sup>	<sup>111</sup>
pBio_ery_p15A_pAM $\beta$ 1_mutb pBio_erm/ery_p15A_pAM $\beta$ 1_mutb_1 pBio_erm/ery_p15A_pAM $\beta$ 1_mutb_2	P15A <i>ori</i> , pAM $\beta$ 1 <i>ori</i> (mutant b), Ery <sup>R</sup>	<sup>111</sup>
pBio_WT_CUR pBio_mutb_CUR	P15A <i>ori</i> , pAM $\beta$ 1 <i>ori</i> (wild-type vs mutant b), Ery <sup>R</sup> , 3 genes of the curcumin synthetic pathway ( <i>4CL</i> , <i>DCS</i> and <i>CURS1</i> ) under the control of the SLP promoter	This study

### 3.3.2. Media and growth conditions

20 g/L of Luria-Bertani (LB) broth (Nzytech) supplemented with 500  $\mu$ g/mL erythromycin (Sigma) was used for growing *E. coli* DH5 $\alpha$ , GM2163 or SCS110 harboring the different plasmids, at 37°C, 250 rpm.

*L. lactis* LMG19460 were grown at 30°C, 100 rpm, using M-17 (pH 7.0, Fluka) supplemented with 20 g/L glucose (Fisher Scientific), and 5  $\mu$ g/mL erythromycin (Sigma) when transformed with one of the several plasmids.

When necessary, *E. coli* cells were grown in LB agar (2%) plates with the appropriate antibiotic, at 37°C, while *L. lactis* cells were grown in solid regeneration medium<sup>112</sup> containing (per liter) 10 g of tryptone (BD Biosciences), 5 g of yeast extract (Liofilchem), 200 g of sucrose (Fisher Scientific), 10 g of glucose (Fisher Scientific), 25 g of gelatin (Merck), 15 g of agar (JMV Pereira), 2.5 mM of MgCl<sub>2</sub>·6H<sub>2</sub>O (Fagron), 2.5 mM of CaCl<sub>2</sub> (V. Reis) and 5 mg erythromycin, at 30°C.

Master and working cell banks of all the strains were made with a final glycerol concentration of 20%, and stored at -80°C.

### 3.3.3. Competent cells and transformation

#### 3.3.3.1. *E. coli* DH5 $\alpha$ , GM2163 and SCS110

Chemically competent *E. coli* DH5 $\alpha$ , GM2163 or SCS110 cells were prepared using cells grown until reach an OD<sub>600nm</sub> of 1, in a 100 mL shake flask with 20 mL LB, at 37°C and 250 rpm (initial OD<sub>600nm</sub>=0.1, from an overnight inoculum). Freshly grown cells were centrifuged (1,000 g, 10 min, 4°C) and resuspended in 0.1 volumes (2 mL) of chilled TSS medium (20 g/L LB, 5% DMSO, 50 mM MgCl<sub>2</sub>, 10% PEG 8000 (w/v), pH=6.5). After 10 min on ice, 100  $\mu$ L aliquots were made and stored at -80°C.

In order to transform DH5 $\alpha$ , GM2163 or SCS110 cells with pDNA by heat shock, for every 100  $\mu$ L of chemically competent cells, the appropriate amount of pDNA was added, accordingly with each experiment. The mixture was incubated for 30 min on ice, and then submitted to 42°C for 60 s. The cells were immediately incubated for 2 min on ice, and then resuspended with 900  $\mu$ L of LB medium. After 1 hour of recuperation at 37°C, the cells were plated on LB agar with 500  $\mu$ g/mL erythromycin (Sigma), and incubated overnight at 37°C.

#### 3.3.3.2. *L. lactis* LMG19460

The procedure for making electrocompetent *L. lactis* LMG19460 were adapted and optimized from Holo & Nes (1989) protocol<sup>112</sup>. An overnight growth in 5 mL of M-17 supplemented with 0.5% glucose (GM-17), 30°C, 100 rpm, is used to inoculate a 100 mL shake flask with 75 mL of GM-17 with a starting OD of 0.1, using the previously determined correlation of 1 OD unit at 600 nm being equivalent to 7x10<sup>8</sup> cells/ml<sup>80</sup>. After growing at 30°C, 100 rpm until reaching an OD<sub>600nm</sub>=0.5-0.8, the whole broth was diluted 100-fold to fresh GM-17 medium (750  $\mu$ L to 75 mL) supplemented with 0.5M sucrose (SGM-17) and 1-2% glycine (Sigma). The cells were grown at 30°C, 100 rpm until reach an OD<sub>600nm</sub>=2-2.5 and harvested by centrifugation at 6,000g, 3 min, 4°C. Next, the cells were washed twice in an ice cold solution containing 0.5 M sucrose and 10% glycerol (VWR Chemicals) (washing solution). Finally, they were resuspended in 1/100 of the initial culture volume of the washing solution and stored at -80°C.

For transformation of the electrocompetent *L. lactis* LMG19460 cells, 40  $\mu$ L aliquots containing approximately 1x10<sup>9</sup> cells were mixed with the appropriate volume of pDNA and transferred to a 2 mm gap width electroporation cuvette. After 30 min on ice, 2-3 pulses of 1,000 V were given using an electroporator (BTX ECM399), resulting in

time constants of 4 to 5 ms. To recover from the electroporation, 960  $\mu\text{L}$  of recovery medium (ice-cold SGM-17 supplemented with 20 mM  $\text{MgCl}_2$  and 2 mM  $\text{CaCl}_2$ ) was immediately added. After resting 5 min on ice, the mixture was incubated 3h at 30°C without agitation. Then, the cells were collected by centrifugation and used to inoculate 5 mL of M-17 supplemented with 20 g/L glucose and a sub-lethal erythromycin concentration<sup>113</sup> (2.5  $\mu\text{g}/\text{mL}$ ), at 30°C without agitation. After an overnight growth, the cells were again collected by centrifugation and plated in regeneration medium<sup>112</sup> containing (per liter) 10 g of tryptone, 5 g of yeast extract, 200 g of sucrose, 10 g of glucose, 25 g of gelatin, 15 g of agar, 2.5 mM of  $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$ , 2.5 mM of  $\text{CaCl}_2$  and 5 mg erythromycin, at 30°C.

### 3.3.4. Biobricks *in silico* design

Several different plasmids were designed *in silico* in order to be assembled by a Gibson Assembly strategy, using Biobrick parts. All the Biobrick parts were designed using the pTRKH3 plasmid as template, with the goal of engineering a minimal working plasmid, with improved features. The idea is that if the plasmid becomes smaller by discarding non-essential parts, the plasmid copy number (PCN) could increase, since the metabolic burden should decrease. Also, the use of Gibson Assembly allows to test different alternative parts, such as the mutant pAM $\beta$ 1 origin of replication, from pTRKH3-b, which was developed to have an increased PCN.

The *in silico* design and assembly was performed using the APE<sup>114</sup> and SnapGene<sup>115</sup> softwares, while the primers necessary for the Gibson Assembly were designed using the online software NEBuilder Assembly Tool<sup>116</sup>.

The first Biobrick assembled was pBio\_ery\_p15A, consisting of a p15A Gram-negative origin of replication and the erythromycin resistance gene, both amplified from the backbone pTRKH3 using primers with overlapping sequences (p15A\_F/R and eryR\_F/R) (Table 3.2). The final p15A origin of replication and erythromycin resistance gene sequences are represented in Table 3.3. The primers were designed in order to introduce a XbaI restriction enzyme site between the end of the erythromycin resistance gene and the beginning of the p15A origin, and a SalI restriction enzyme site between the end of the p15A origin and the beginning of the erythromycin resistance gene, in order to have alternative ways to assemble the final plasmid, if necessary<sup>111</sup>.

The pBio\_ery\_p15A\_pAM $\beta$ 1\_WT and pBio\_ery\_p15A\_pAM $\beta$ 1\_mutb plasmids were assembled with the addition of the wild-type or mutant b Gram-positive pAM $\beta$ 1 origin of replication, respectively, to the previous plasmid, using the primers indicated in Table 3.2. The final pAM $\beta$ 1 sequence is represented in Table 3.3. A new restriction enzyme site (PmeI) was introduced between the pAM $\beta$ 1 and the p15A origins of replication<sup>111</sup>.

Two different approaches were pursued in order to assemble the erythromycin resistance gene promoter (erm): 1) introducing directly in the pBio\_ery\_p15A\_pAM $\beta$ 1\_WT and pBio\_ery\_p15A\_pAM $\beta$ 1\_mutb plasmids, generating the pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_1 and pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_1; or 2) introducing first in the pBio\_ery\_p15A, generating the pBio\_erm/ery\_p15A, followed by the assembly of the wild-type and mutant pAM $\beta$ 1 origins of replication (pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_2 and pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_2)<sup>111</sup>. The final design is quite similar, differing only in some nucleotides, due to the necessity of using different primers in each situation. The primers used in each approach are described in Table 3.2 and the erm sequence in Table 3.3.

The pBio\_WT\_CUR and pBio\_mutb\_CUR plasmids were the result of the assembly of the 3 genes (*4CL*, *DCS* and *CURSI*) of the curcumin synthetic pathway under the control of the SLP promoter, into the vectors pBio\_erm/ery\_p15A\_pAMβ1\_WT\_2 and pBio\_erm/ery\_p15A\_pAMβ1\_mutb\_2, using the primers indicated in the Table 3.2. The sequences of each gene are described in Table 5.2 (Chapter 5).

**Table 3.2.** Sequence, expected amplicon size and melting temperature of the primers used for nth CRISPR/Cas9 gene knockout strategy. It is also indicated where the primer overlaps (lowercase sequence), as well as the place where it anneals in the plasmid (uppercase sequence). Underlined letters corresponds to introduced restriction enzyme sites.

Primer name	Overlaps	Sequence (5' → 3') <sup>1</sup>	Anneals	Product size (bp)	T <sub>m</sub> (°C)
<b>pBio_ery_p15A<sup>III</sup></b>					
<b>p15A_F</b>	F eryR	gaggaaataatctagaTAGCGGAGTGATACTGG	p15A	909	58.6
<b>P15A_R</b>	R eryR	tatTTTTgttcatgtcgcacACAACCTTATATCGTATGGGG	p15A	60.8	
<b>eryR_F</b>	F p15A	tataagttgtcgcacATGAACAAAAATATAAAATATTCTCA AAAC	eryR	738	58.5
<b>eryR_R</b>	R p15A	cactccgctatctagaTTATTTCCTCCCGTTAAATAATAG	eryR	59.9	
<b>pBio_erm/ery_p15A<sup>III</sup></b>					
<b>pBio_erm/ery_p15A_pAMβ1_WT_1<sup>III</sup></b>					
<b>pBio_erm/ery_p15A_pAMβ1_mutb_1<sup>III</sup></b>					
<b>bioery_F</b>	F -	ATGAACAAAAATATAAAATATTCTCAAAACTTTTT AAC	Plasmids	Variable	53.1
<b>bioery_R</b>	R -	GTCGACACAACCTTATATCGTATGGG	Plasmids	55.1	
<b>erm_F</b>	F Plasmids	acgatataagttgtcgcacAGTCTAGAATCGATACGATTTTG	Erm	541	61.9
<b>erm_R</b>	R Plasmids	aatattttatattttgttcatGTAATCACTCCTTCTTAATTACAAA TTTTTAG	Erm	58.6	
<b>pBio_ery_p15A_pAMβ1_WT<sup>III</sup></b>					
<b>pBio_erm/ery_p15A_pAMβ1_WT_2<sup>III</sup></b>					
<b>pBio_ery_p15A_pAMβ1_mutb<sup>III</sup></b>					
<b>pBio_erm/ery_p15A_pAMβ1_mutb_2<sup>III</sup></b>					
<b>p15A_ery_F</b>	F pAMβ1	aatagaattcgtttaaacTAGCGGAGTGATACTGG	Plasmids	Variable	59.1
<b>P15A_ery_R</b>	R pAMβ1	agagcgcctagTCTAGATTATTTCTCCCG	Plasmids	60.5	
<b>pAMβ1_F</b>	F Plasmids	ataatctagaCTAGCGCTCTTATCATGG	pAMβ1	3,478	55.6
<b>pAMβ1_R</b>	R Plasmids	cactccgctagtttaaacGAATTCTATTTAATCACTTTG ACTAG	pAMβ1	60.6	
<b>pBio_WT_CUR</b>					
<b>pBio_mutb_CUR</b>					
<b>Biovector_F</b>	F 3 genes	gctgtactagcactgtTAGCGGAGTGATACTGG	Vectors	5,706	64.4
<b>Biovector_R</b>	R 3 genes	tacttaccacactagtGAATTCTATTTAATCACTTTGACTAG	Vectors	58.8	
<b>CUR_3genes_F</b>	F Vectors	aatagaattcactagtGTGGTAAGTAATAGGACGTG	3 genes	4,563	58.7
<b>CUR_3genes_R</b>	R Vectors	cactccgctacactgtCTAGTACAGCGGCATAGAAC	3 genes	66.8	

<sup>1</sup> F and R indicate forward and reverse primers, respectively.

**Table 3.3.** Final sequences of the p15A origin, *ery* (erythromycin resistance gene), *erm* (ery promoter) and pAMβ1 origin of replication.

	Sequence (5' → 3')
<b>p15A origin</b>	tagcggagtgtatactgcttactatgttggcactgatgaggggtcagtgaaagtcttcatgtgcccagagaaaaagcctgaccggtgctgctcagcagaat atgtgatacaggatataatccgcttctcctcactgactcgetacgctcgttcgactcgcggcagcggaaatgcttacgaacggggcggagatttctt gaaagatgccaggaagataacttaacagggaagtgaagggcccgccgaaagccgttttccataggtccgcccctgacaagcatacgaatctgac gctcaaatcagtggtggcgaacccgacaggactataaagatacaggcgtttccccctggcggctccctcgtgcgctcctgttctgcttctggttaccg gtgtcattccgctgttatggccgctttgtctcattccacgctgacactcagttccggtaggagctcctccaagctggactgtatgcacgaacccccgttc agtcgaccgctgcttccgcttatccgtaatactatctgtcttgccaacccgaaagacatgcaaaagcaccactggcagcagccactgtaattgattagagg agttagtctgaaatcagcgcgggtaaggctaaactgaaaggacaagtttggtagctgctcctccaagccagttacctcgggtcaaaaggtgtagctc agagaacctcgaaaaaccgcccctgcaaggcgttttcttccagagcaagagattacgcgagacaaaacgatctcaagaagatcatcttattaatcag ataaaatattctagattcagtgcaatttatcttcaaatgtagcactgaaagtcagccccatagataaagttgt
<b><i>ery</i> (erythromycin resistance gene)</b>	atgaacaaaaataaaatattctcaaaactttttaaagagtgaaaaagtactcaacaaaataaaaaaattgaatttaaaagaaaccgatacgttttacgaaat tggaacaggtaaaggcatttaacgacgaaactggctaaataagtaaacaggtacgtctattgaaatgacagctcatcttcaactatcgtcagaaaaatta

	<pre> aaactgaatactgtgtcacttfaatcaccagatattctacagttcaattccctaacaacagaggtataaaattgtgggagttaccctaccatftaagcacac aaatattaaaaagtggttttgaagccatgcgctgacatctatctgattgftgaagaaggattcacaagcgtaccttggatattaccgaaactaggggtgc tcttgacactcaagctcgttcgcaatcttaagctgccagcggatgctttcatcctaaacaaaagtaaacagtgcttataaaacttaccggcaatacc acagatgttccagataaaatattggaagctatatactgactttgttcaaaatgggtcaatcagagaatcgtcaactgtttactaaaaactgattcaagcaatga aacacgccaaagtaacaattaaagctaccgttactatgagcaaggtattgctattttaaatagttatctattttaacgggaggaataa </pre>
<b>erm (ery promoter)</b>	<pre> agtctagaatcgatactgttgaagtggaacagataaaaaaaagcagttttaaattgttctgaacttttaaaacaaagcaatacaatcattgctgcacagat agcgacagagaagggcaaaacattccctggctgatcattcataaagcaaatgctctttctaaagataaacgtataaaagactatggatcaatggttagaaaa gatgctgacccgtagcgggtttcaaaattgcaaccaggaatgaattactccctttatcaagaagcgaacaaagaaaaacgaaatgataccaactcagtgcaa aaaaagataaatgggagataagcggctgctgctgactgacccatataaaaaatcgaacagcaagaatggcggaaactgaaaaagagttat ggaaataagacttagaagcaaacftaaagtggtgtagatgcaagatctttaaattgtataataggaattgaagftaaattagatgctaaaaattgttaataag aaggagtgattac </pre>
<b>pAMβ1 origin</b>	<pre> ctagcgtcttatcatggggaagctggatcatatgcaagacaaaaataaacctgcaacagcacttggagaatgggacgaatcagaaaaaccttttacct ggattacatatataaaagccgtaaggagaggggttcaaaaagggtttaaataaggagagaacaaatcaatcattagctagaactatatttttggacaactggg agaatttagaagcgtgctctcaagaccgttacaagagctagctactaaacataatttaaacgctataagtggtggaacactgtatataggaaaaagc cgtagaagaataaaagcaagaggagaatttagagaagatttaagccatgctggtggcgttaggagggaaacatatacaatttcttgagaatacaaaattga aggattacatgacactggcgaatgaattacgtctttacgtataaaagagccgttttattctaataaacggctcttttatagaaaaactcctagctggtgttttt tcgaaatgctgctggtaccctaaagaaatagaatgagtagatcaaatattacgaatagaatcagaaaaatcagaaactgaaacaaactagaaacaa ttgcaaaagtaactaactcaacgctagtagtgatgatttaacccaaatgagccaacagaaccagagccagaacagaatcagaacaagtaaacattggatttagaa atggaagaaagaaaaagcaatgactctgtgtaataatgcacgaaatcgttctatttttttaaaagcgttatactagataaacgaaacaaacgaactgaaata gaaacgaaaaagagccatgacacattataaaatgtttgacgacatttataaaatgcatagcccagataagttgcaaaacacgcttactagttagtcagatga acttcccctgtaagaagttatttaactttgttgaagcggatataaccgtactactattatagggaaatcagagagtttcaagttactaagctactgaa tttaagaattgtaagcaatcaatcggaaatcgttattgctttttgtattcattatagaaggtggagttgtatgaatcatgatgaatgtaaaactatataaaaaa tagttattggagataagaaaattagcaaatctatacactagaacgtttaaagaaagagttagaagaaagaaatattctactagaacaaaaactagataagatt ttctcggagggggagattataataagtttaagaaataacaaaaatatttactgattagtgaaaaaaattgactataaaggaaaaaaactcttttcaaa acatgcaatattgaaacagttgaaatgaaaaagcaaaccaagtttaataacaacactattttataggattataggaagcgaacagctgaaatgaatccccctttg ttgagaactgtgcttcatgacggctgttaagtagcaaaatttaaaaatagtaaaatcgcctcaatcactaccaagccaggtaaaaagcaaaagggcctattttgcg tatcgtcaaaatcaagcatgattggcggctgctggtgttctgactccgaggaaagcattcaagaaatcaagatacatttacacattggacaccaacgcttt atcgttatggaacgtatgcagacgaaacccgttatacagaaagacattcgaacaaatitaaagcaaaatcaatcctcttattgattttgatattcacacgg caaaagaaactattcagcaagcgtattttaaaccgctattgattaggtttatgctactatgattcaaatctgataaagggtatcaacatgacatattttgtttag aaacgccagtctgactcaaaatcagaatttaaaactgtcaaaagcagccaaataatttcgcaaaatctccgagaatctttgaaagctcttggcagttgac taacgtgtaactattgtgattgctgcatacgaacggcaaatgtaaaattttgatccttaattaccgttatttcaaaagaatggcaagattggtcttcaaa caaacagataataaggccttactgttcaagctaaacgggtttaaagcgtgacagaagcgaacaaacaaagtagatgaacctggttattctctattgacagaaa cgaattttcagagaaaagggttataagggcgtataaacgctatgtttaccctcttttagcctacttttagctaggtcagctatcaatcgaacgtgcaataataat gtttgagtttaataatcgattagatcaacccttagaagaaaaagaaatcaaaatgtagaagtgctattcagaaaaactcaagggcctaatagggaatc attaccattcttgcagaactgggtatcaagtttaaccagtaagattttgtccgtcaagggtgtttaaactcaagaaaaaaagaaagcgaacgtcaacg tgtcattgtcagaatggaaagatttaagcttataatgcaaaaaagcgtatataaacgcttatttagtgcagcaaaaaagagattagagaagtg ctaggcattcctgaacggacattagataaattgctgaagtgactgaagcgaatcagaaatctttaaagtaaacaggaagaaatggtggcattcaact ctagtgtttaaactattgtgctatcgtatcattaaagtaaaaaagaaagaaagctatataaagcgtgacaaaacttttgacttagcagacatacattcatt caagagactttaaacaagctagcagaacgcctaaaacggacacacaactcgaattgttagctatgatacaggctgaaaaataaaaccgactatgccattac atttatctatgatacgtgtttttcttctgttttagcgaatgattagcaaaaatatacagagtaagattttaaataattattagggggagaagggagagatgag cccgaaaacttttagtggcttggactgaacgaagtgaggaaagcactaaaacgtcgaagggcagtgagagcgaagcgaacactgatttttaattttcta cttttatagtcattagatatactattgtcctataaaactatttagcagcataatagatttattgaatagctttaaagttgagcatattagaggaagaaacttg gagaaatattgaaagaccgattacatggttggattgctgtgttacgtgttttaactaaaagtagtgaatttttgaatttttggctgtgctgtctgtgttagt attgctagctcaaaagtgattaaatagaattc </pre>

### 3.3.5. Biobricks amplification and assembly

PCR amplification of the p15A origin and erythromycin resistance genes Biobrick parts, to assemble the pBio\_ery\_p15A plasmid, was performed using the KOD Hot Start DNA Polymerase kit (Novagen), by mixing 0.5 ng of purified pTRKH3, 0.02U/μL of KOD DNA Polymerase, 2.3 mM of MgSO<sub>4</sub>, 1x buffer, 0.2 mM of dNTPs, 0.3 μM of each primer (p15A\_F/R or eryR\_F/R) and completed with PCR-grade water to a final volume of 25 μL. The cycling conditions for the 2 reactions were: initial denaturation for 2 min at 95°C, followed by 40 cycles of 1 min at 95°C, 1 min at 59°C and 2 min at 70°C<sup>11</sup>.

For the assembly of the pBio\_ery\_p15A\_pAMβ1\_WT and pBio\_ery\_p15A\_pAMβ1\_mutb plasmids, it was necessary to amplify the wild-type and mutant b pAMβ1 origin of replication, starting with 0.5 ng of the template plasmids (pTRKH3 and pTRKH3-b, respectively), using the pAMβ1\_F/R primers. The cycling conditions for the 2 reactions were: initial denaturation for 2 min at 95°C, followed by 40 cycles of 1 min at 95°C, 1 min at 60°C and 3.5 min at 70°C. Next, the pBio\_ery\_p15A plasmid was amplified using the p15A\_ery\_F/R primers, starting with 0.5 ng of the template

plasmid. The plasmid was amplified with the following conditions: 2 min at 95°C, followed by 40 cycles of 1 min at 95°C, 1 min at 60°C and 2 min at 70°C<sup>111</sup>.

To generate the pBio\_erm/ery\_p15A, pBio\_erm/ery\_p15A\_pAMβ1\_WT\_1 and pBio\_erm/ery\_p15A\_pAMβ1\_mutb\_1, first the erm Biobrick was amplified from the pTRKH3 plasmid, starting with 0.5 ng of the template and using the erm\_F/R primers. The cycling conditions were: initial denaturation for 2 min at 95°C, followed by 40 cycles of 1 min at 95°C, 1 min at 58°C and 1 min at 70°C. The plasmids pBio\_ery\_p15A, pBio\_ery\_p15A\_pAMβ1\_WT and pBio\_ery\_p15A\_pAMβ1\_mutb plasmids were amplified with the bioery\_F/R primers, from 5 ng of template previously linearized with Sall restriction enzyme, using the following cycling conditions: 2 min at 95°C, followed by 40 cycles of: 1 min at 95°C, 1 min at 55.3°C and 1 min 42 s (for the first template) or 5 min 12 s (for the remaining templates), at 70°C<sup>111</sup>.

The second approach to generate similar final plasmids, started with the assembly of the pBio\_erm/ery\_p15A, followed by the assembly of the wild-type and mutant pAMβ1 origins of replication (pBio\_erm/ery\_p15A\_pAMβ1\_WT\_2 and pBio\_erm/ery\_p15A\_pAMβ1\_mutb\_2). The pTRKH3 and pTRKH3-b pAMβ1 replicons were amplified as already described above, while the pBio\_erm/ery\_p15A was amplified starting with 0.5 ng of pDNA and the p15A\_ery\_F/R primers, with the following cycling conditions: 2 min at 95°C, followed by 40 cycles of 1 min at 95°C, 1 min at 60°C and 2 min at 70°C<sup>111</sup>.

In order to generate the pBio\_WT\_CUR and pBio\_mutb\_CUR plasmids, the vector and insert fragments were amplified starting with 0.5-10 ng of pDNA, using the KOD Hot Start DNA Polymerase kit (Novagen), with the following cycling conditions: 2 min at 95°C, followed by 40 cycles of 1 min at 95°C, 1 min at 60°C and 6 min at 70°C.

After amplification, the PCR products were run in a 1% agarose gel in order to retrieve the specific fragment without UV or ethidium bromide exposure, to keep the DNA quality and integrity. The agarose gel fragment was purified using NZYGelPure (Nzytech) and its concentration quantified using a Nanodrop Spectrophotometer (Nanovue Plus, GE). Finally, the purified product was digested with 10U DpnI (Promega) for 2 hours at 37°C, to remove the parental plasmids<sup>111</sup>.

To perform the Assembly of the different fragments it was applied the Gibson Assembly technique, using for that purpose the NEBuilder HiFi DNA Assembly Cloning Kit. The manufacturer instructions were followed and in some steps optimized to the specificities and requirements of the present work. Each 20 μL Gibson Assembly reaction consisted of 10 μL of NEBuilder HiFi DNA Assembly Master Mix (constituted by the three necessary enzymes and a reaction buffer with PEG-8000, Tris-HCl pH 7.5, MgCl<sub>2</sub>, DTT, dNTPs and NAD<sup>109</sup>) and the necessary volume of each fragment to obtain a maximum of 0.3 pmol of DNA in each assembly reaction. The remaining volume was completed with PCR grade water. The DNA mass ratio of vector:insert tested were between 1:1 and 1:17. Each reaction was incubated at 50°C for different incubation times (15 min and 1 hour), after which chemically competent *E. coli* DH5α were transformed with 2 μL of the reaction mixture. The exact amounts of DNA fragments used in each assembly reaction are displayed in Table 3.4<sup>111</sup>.

Several colonies were picked from the LB plates supplemented with 500 µg/ml erythromycin and grown in liquid medium at 37°C, in order to purify pDNA using High Pure Plasmid Isolation Kit (Roche) and test if the assembly was successful, using the appropriate restriction enzymes<sup>111</sup>.

The result from each successful assembly will be first sent to sequencing and then used to transform *L. lactis* LMG19460, in the case when they had the pAMβ1 origin of replication. The resulting colonies must be tested for the plasmid presence, by purifying pDNA using the NucleoSpin Plasmid kit (Macherey-Nagel) and visualizing it in a 1% agarose gel, after digesting it with the appropriate enzymes<sup>111</sup>.

**Table 3.4.** Mass and molar amount of each DNA fragment necessary for the several Gibson Assembly reactions<sup>111</sup>.

Plasmid	DNA fragments	DNA mass (ng)	DNA molar amount (pmol)
pBio_ery_p15A	p15A	92.0	0.150
	eryR	75.1	0.150
pBio_erm/ery_p15A	pBio_ery_p15A	50.0	0.046
	erm	100.0	0.260
pBio_ery_p15A_pAMβ1_WT	pBio_ery_p15A	164.5	0.150
pBio_ery_p15A_pAMβ1_mutb	pAMβ1/ pAMβ1 mutb	341.3	0.150
pBio_erm/ery_p15A_pAMβ1_WT_1	erm	100.0	0.260
pBio_erm/ery_p15A_pAMβ1_mutb_1	pBio_ery_p15A_pAMβ1_WT/mutb	50.0	0.015
pBio_erm/ery_p15A_pAMβ1_WT_2	pBio_erm/ery_p15A	60.0	0.050
pBio_erm/ery_p15A_pAMβ1_mutb_2	pAMβ1/ pAMβ1 mutb	95.0	0.050

### 3.3.6. Plasmid copy number determination by qPCR

#### 3.3.6.1. *L. lactis* LMG19460 growth conditions

*L. lactis* strains harboring the different assembled plasmids were grown in a systematic way, starting with an overnight inoculum grown at 30°C, 100 rpm in 5 mL MRS<sup>79</sup> (15 mL Falcon tube). Each inoculum was used to start a culture with an initial optical density of 0.1, using the previously determined correlation of 1 OD unit at 600 nm being equivalent to  $7 \times 10^8$  cells/ml<sup>80</sup>, which was confirmed for our strain. The cultures were performed in 100 mL shake flasks with 75 mL of M-17 medium<sup>81</sup> (pH 7.0, containing 5 g/L of lactose) (Fluka) supplemented with 20 g/l of glucose and 5 µg/ml of erythromycin, at 30°C, 100 rpm. At several time points (6h, 8h, 10.5h and 24h) samples were collected for OD measurements, for real-time qPCR analysis, for pDNA purification (OD=20 equivalent to  $1.4 \times 10^{10}$  cells) to assess plasmid quantity and quality, and for total RNA extraction ( $1.0 \times 10^9$  cells) for further mRNA analyses. Since a considerable amount of sampling volume was needed to those analyses, four parallel shake flasks (one per time point) were performed in each one of the independent cultures, to maintain the culture volume and guarantee the same growth conditions to all cells.

### 3.3.6.2. Primer design

Two sets of primers were designed, one specific for the erythromycin resistance gene (*erm*) present in the pTRKH3 plasmid and the other specific for the ferrous iron transport protein A (*feoA*) chromosomal single copy gene of *L. lactis* (Table 3.5).

**Table 3.5.** Sequence of the primers used for qPCR amplification of the target (*erm*) and reference (*feoA*) genes, as well as the accession number of the complete sequence and expected amplicon size. F and R indicate forward and reverse primers, respectively.

	Gene	Primer sequence	Accession number	Product size (bp)
Plasmid	<i>erm</i>	F: CTTGTTATGATTTTACA R: CAATATCAACAATTCCAT	LMBP 4462	190
Chromosome	<i>feoA</i>	F: TCAGACGCCGCTTGATGGAC R: AGTTCAAGAGGGTCGCCAAGTG	AE005176	89

The genome of the *L. lactis* LMG19460 strain used in the present study was not yet sequenced at the time the experiments were performed. Therefore, it was assumed that the sequence of the *feoA* gene was similar to *L. lactis* ssp. *lactis* IL1403. This assumption was verified by analysis of the product length after amplification using *L. lactis* LMG19460 genomic DNA as template and latter corroborated by *L. lactis* LMG19460 sequence analysis. Primers were synthesized by Stabvida.

### 3.3.6.3. Samples and standards preparation

Genomic DNA from *L. lactis* LMG19460 was purified using Wizard Genomic DNA Purification Kit (Promega), quantified by Nanodrop measurement and its quality assessed by agarose gel electrophoresis. The pTRKH3 was purified from *L. lactis* LMG19460 using Nucleospin Plasmid Kit (Macherey-Nagel). Then, genomic and plasmid DNAs were serial diluted in PCR-grade water to be used as real-time qPCR standards (1, 10, 100, 1,000 and 10,000 pg per reaction).

The samples of *L. lactis* LMG19460 harboring the different assembled plasmids were collected as described above and then diluted in PCR-grade water in order to add 1,000 cells to each PCR capillary. To ensure that the samples and the pDNA standards have the same putatively inhibitors present, the capillaries for the pDNA standard curve were spiked with the same number of *L. lactis* cells (1,000) without plasmid. The same was not applied to the gDNA standards, because the *L. lactis* genome of the cells used to spike the capillaries could mislead the gDNA quantification.

### 3.3.6.4. Real-time qPCR reaction

The real-time qPCR reactions were performed in the LightCycler (Roche) detection system using the FastStart DNA Master SYBR Green I kit (Roche). Each 20  $\mu$ L reaction mixture had 2  $\mu$ l of 10x SYBR Green Mix, 1  $\mu$ l of each primer (final concentration 0.5  $\mu$ M), 1.6  $\mu$ l of MgCl<sub>2</sub> (final concentration 3 mM), 10.4  $\mu$ l of PCR-grade water and the remaining volume was completed with the samples (4  $\mu$ L with 1,000 cells containing the different plasmids) or with

pDNA (2 µL of pDNA plus 2 µL with 1,000 spiking cells) or gDNA (4 µL) standards. Negative controls without cells and without pDNA or gDNA were performed in all cycling reactions, as well as at least one point of each standard curve for experimental validation.

The cycling conditions were as follows: initial denaturation for 10 min at 95°C, followed by 40 cycles of 10s at 95°C, 5s at 57°C and 14s at 72°C. The fluorescence signal was detected and quantified automatically by the instrument at the end of each extension step. To confirm that only the desired specific targets were amplified, melting curve analyses were performed by making a temperature gradient of 0.05° C/s from 70 to 95°C. Lastly, the samples were cooled down to 40°C for 30s.

### 3.3.6.5. Plasmid copy number (PCN) quantification

A relative quantification method developed by Skulj *et al.* (2008)<sup>87</sup> was used for PCN quantification using the real-time qPCR data. The crossing point (CP) values were determined automatically by the LightCycler software, using the Second Derivative Maximum method.

First, the standard curves for pDNA and gDNA were determined by plotting the logarithm of each concentration against the CP values. After estimation of each linear regression equation, the slope of each standard curve was used to determine the amplification efficiency (E) according to Equation 1:

$$E = 10^{\left(-\frac{1}{\text{slope}}\right)} \quad (1)$$

Knowing the amplification efficiencies of both the pDNA (E<sub>p</sub>) and gDNA (E<sub>g</sub>) standards, as well as the CP values for each sample amplified independently with the two primer sets (CP<sub>p</sub> with plasmid primer set and CP<sub>g</sub> with gDNA primer set), the PCN of each sample can be calculated using the Equation 2 (Skulj *et al.* 2008)<sup>87</sup>:

$$\text{PCN} = \frac{E_g^{\text{CP}_g}}{E_p^{\text{CP}_p}} \quad (2)$$

### 3.3.6.6. qPCR validation

To verify the specificity of the real-time qPCR amplification, the melting temperature (T<sub>m</sub>) of each amplicon was checked. Also, the content of each capillary was run in a 1% agarose gel, being visible the amplicons with the correct sizes for each set of primers, confirming the specificity of all the performed reactions.

Negative controls without cells and without pDNA or gDNA were performed in all cycling reactions and resulted in CPs much higher (>than 35 CPs) than all the samples, therefore background amplification could be considered negligible.

The validity of the qPCR results was also evaluated by the coefficient of determination (R<sup>2</sup>) and by the percentage of amplification efficiency of each standard curve, which was calculated using Equation 3<sup>87</sup>:

$$E(\%) = \left(10^{\left(-\frac{1}{\text{slope}}\right)} - 1\right) \times 100 \quad (3)$$

The quality and integrity of the pDNA in each analyzed sample was determined in purified plasmid samples with  $1.4 \times 10^{10}$  cells (OD=20) using the High Pure Plasmid Isolation Kit (Roche) and subsequent quantification in a Nanodrop Spectrophotometer (Nanovue Plus, GE) and isoform analysis by agarose gel electrophoresis.

### 3.3.6.7. Statistical analysis

For each independent growth, average PCN from the different cycling reactions was determined and then used to calculate the average and inter-assay variability by determination of the standard error of the mean (SEM). When necessary, a t-student statistical test was performed and differences were considered statistically significant when  $p < 0.05$ .

## 3.4. Results and discussion

### 3.4.1. *In silico* design and plasmid assembly

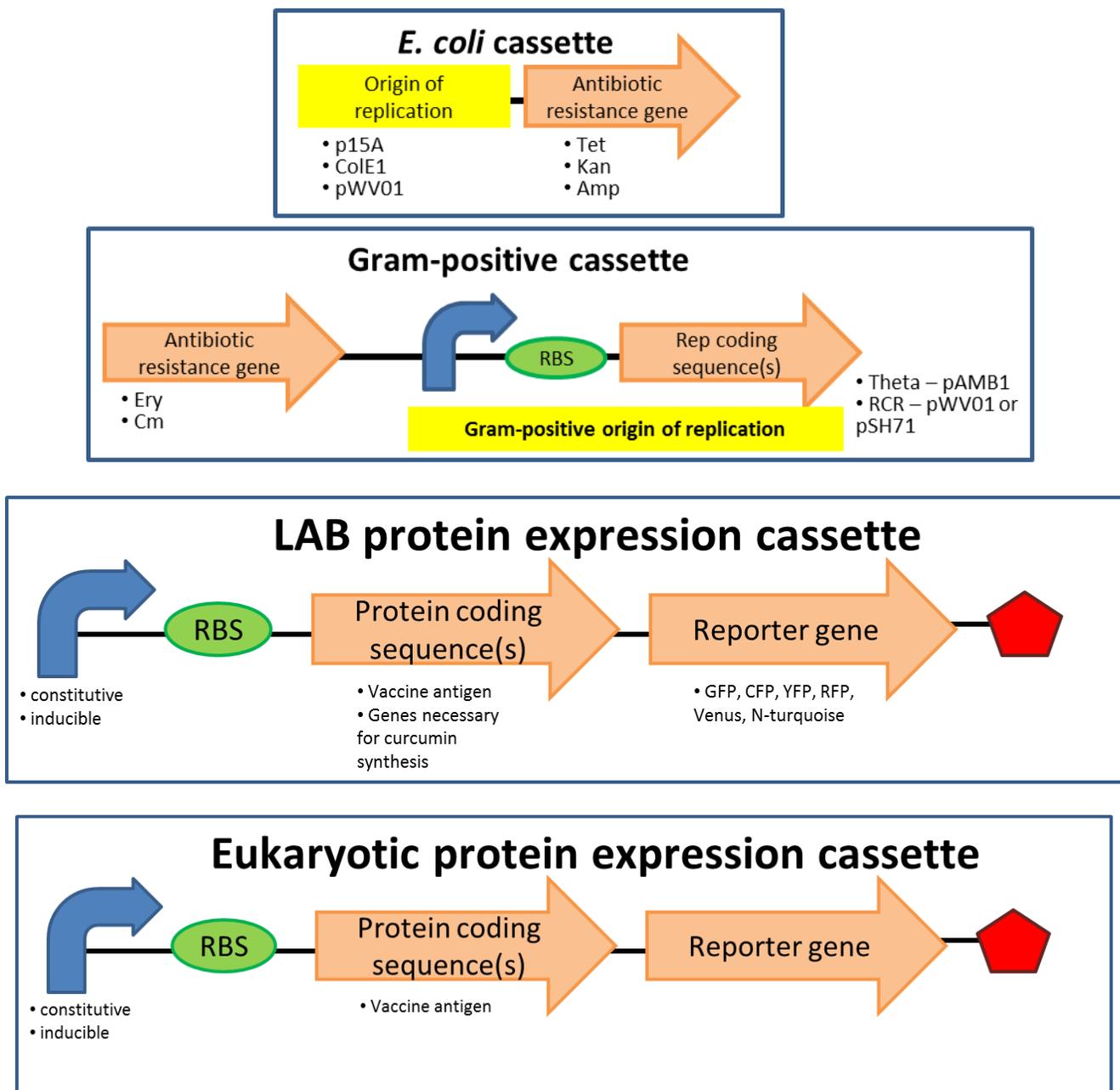
The general idea for this chapter was to design a plasmid with different cassettes, each one constituted by several different BioBrick parts. A synthetic biology approach based on a Gibson assembly strategy will allow the construction of these new artificial expression minimal plasmids. This cutting-edge strategy is easier and more flexible than the traditional synthetic biology approach using isocaudomers restriction.

The assembled plasmids will be based on the already available *E. coli*/Gram-positive shuttle plasmid pTRKH3<sup>65</sup> and should contain the following BioBrick compatible parts (Figure 3.10): an *E. coli* cassette (useful for the initial steps of molecular cloning), a Gram-positive cassette and two protein expression cassettes, one for prokaryotic applications such as using LAB for producing proteins or metabolites, and the other when the goal is the expression in eukaryotic cells as it happens in DNA vaccination. The *E. coli* and the Gram-positive cassettes should include appropriate antibiotic resistance genes and replication origins to allow plasmid maintenance in each type of bacteria. High and low plasmid copy number (PCN) replication origins for LAB should be considered, depending if the goal is to achieve a tight control of the expression or a high plasmid yield, respectively. The RCR replication type plasmids<sup>64</sup> are generally high-copy-number, but do not allow cloning of large inserts<sup>117</sup>. Theta replicating plasmids are segregationally more stable<sup>73</sup> and can hold larger DNA inserts, but are low-copy-number<sup>118</sup>. Besides changing the entire origin of replication, we propose to alter the Ribosome Binding Site (RBS) of the existing promoter, which controls the *repDE* genes and specifying PCN. On Chapter 2, it was already showed that some modifications in the RBS strongly influence the PCN of pTRKH3. The goal is to use this optimized RBS as a BioBrick part in the new plasmid.

The protein expression cassettes could include inducible or constitutive promoters, according to the expression tightness that is necessary for each cloned gene. In the prokaryotic protein expression cassette, the promoter should be one that works in LAB (such as the constitutive P21, P23, P32, P44 and P59 promoters, or the nisin-inducible NICE

promoter), while in the eukaryotic protein expression cassette should be a eukaryotic one (such as the constitutive CMV, SV40 and EF1a promoters or the inducible TRE and Ac5 promoters). A reporter gene, such as GFP, CFP, YFP, RFP, Venus or N- turquoise, could be added to the cassettes, to facilitate clone screening.

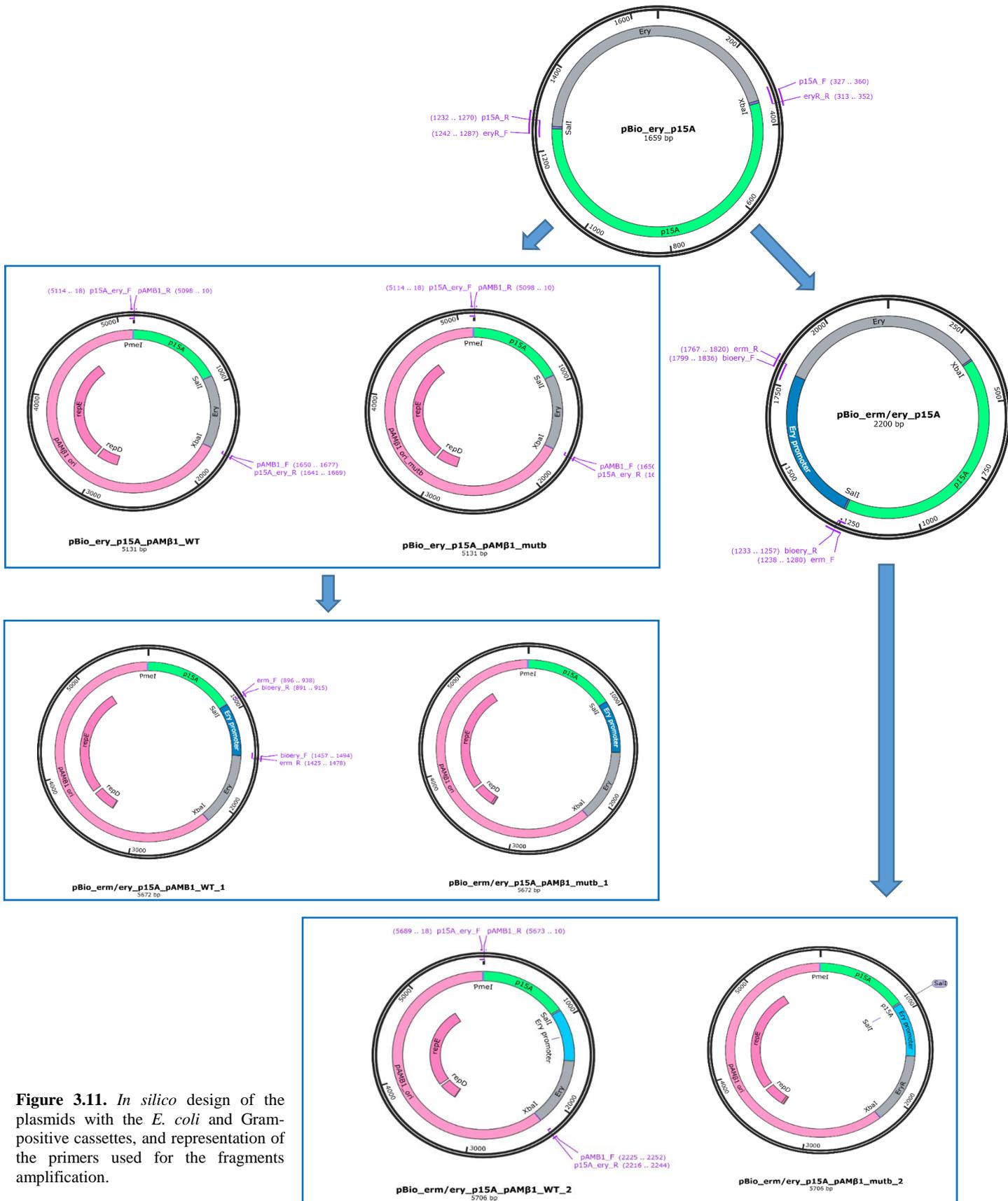
Although I performed all the *in silico* design of the minimal plasmid and the design of the necessary primers, the actual plasmid assembly experiments of the *E. coli* and Gram-positive cassettes were performed by a master student under my co-supervision<sup>111</sup> and her main results will be summarized in the following section, followed by my own attempts to assemble a prokaryotic protein expression cassette.



**Figure 3.10.** General design for the assembly of a synthetic minimal plasmid.

### 3.4.1.1. *E. coli* and Gram-positive cassettes assembly

The *in silico* design of all the plasmids, after fragments amplification and assembly are represented in Figure 3.11, together with the primers used for the fragments amplification. The primer design contemplated the introduction of restriction enzyme sites between two Biobrick parts, in order to enable plasmid confirmation by digestion and as an alternative to linearize the plasmid, if PCR amplification was not suitable.



**Figure 3.11.** *In silico* design of the plasmids with the *E. coli* and Gram-positive cassettes, and representation of the primers used for the fragments amplification.

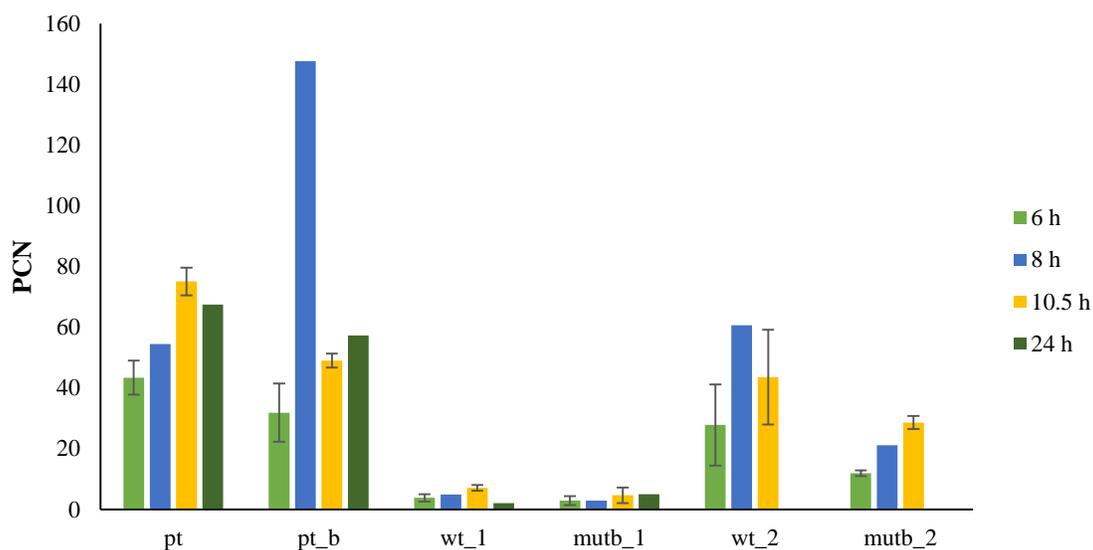
The first assembled plasmid contained two different Biobrick parts: the erythromycin resistance gene and the p15A Gram-negative origin of replication, being able to replicate only in *E. coli* strains. The next assembly step consisted in the addition of the pAM $\beta$ 1 (wild-type or the high PCN mutant b) Gram-positive origin of replication to the pBio\_ery\_p15A plasmid, generating the pBio\_ery\_p15A\_pAM $\beta$ 1\_WT and pBio\_ery\_p15A\_pAM $\beta$ 1\_mutb plasmids<sup>111</sup>. Since both p15A and pAM $\beta$ 1 origins of replication are present in these plasmids, they should replicate both in Gram-negative (*E. coli*) and Gram-positive (*L. lactis*) hosts, being possible to use the erythromycin resistance gene as selection marker for both.

It was noticed that the annotation of the erythromycin resistance gene had a problem, lacking its promoter (erm) in the newly assembled plasmids, but not in the parental pTRKH3 plasmid<sup>111</sup>. In order to overcome the absence of the erm promoter, two different approaches were pursued in order to introduce it in the assembled plasmids: 1) introducing directly in the pBio\_ery\_p15A\_pAM $\beta$ 1\_WT and pBio\_ery\_p15A\_pAM $\beta$ 1\_mutb plasmids, generating the pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_1 and pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_1; or 2) introducing first in the pBio\_ery\_p15A, generating the pBio\_erm/ery\_p15A, followed by the assembly of the wild-type and mutant pAM $\beta$ 1 origins of replication (generating the pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_2 and pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_2 plasmids)<sup>111</sup>.

It was expected that such small plasmids had a high transformation efficiency, even better than the parental pTRKH3, but this assumption was not verified, since the transformation efficiency was very low for almost all plasmids<sup>111</sup>. Other problem after *L. lactis* LMG19460 transformation was related with the appearance of recombined forms of all the plasmids<sup>111</sup>. Actually, this strain is a wild-type strain, being RecA<sup>+</sup>, which means it has an active recombinase able to recombine exogenous DNA such as our plasmid. This undesired effect of the RecA protein is the reason why in common laboratory strains for cloning like *E. coli* DH5 $\alpha$  the *recA* gene is deleted. The mechanisms by which these plasmids are recombined in *L. lactis* were not disclosed in this work. Since the recombination only occurs in the newly assembled plasmids, but not in the parental pTRKH3<sup>111</sup>, one can think that the reason is the smaller size of the assembled plasmids. Actually, the pIL253 plasmid, which gave origin to the pTRKH3 plasmid, tends to recombine in dimers and trimers when used to transform *L. lactis* IL1403, a RecA<sup>+</sup> strain like the *L. lactis* LMG19460 used in this study. The pIL253 is smaller than the pTRKH3 plasmid, with only 4.8 Kb, instead of 7.7 Kb. The smaller size could exacerbate the effect of the presence of direct or inverted repeats in the sequences. It could be anticipated that the assembly of the protein expression cassette, with the associated increase in the plasmid size, could solve this problem without further modifications<sup>111</sup>.

The pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_1 and pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_1 plasmids had additional problems, since the sequencing results revealed mutations that could decrease or even inhibit the *repE* gene expression and/or alter the RepE proteins' functionality, lowering the PCN. However, the pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_2 and pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_2 plasmids were purified with acceptable concentrations, do not have the problematic mutations in the pAM $\beta$ 1 origin of replication, and still have the correct non-recombined plasmid present together with the recombined forms<sup>111</sup>.

Concerning the PCN quantification by real-time quantitative PCR, the average results from three growth experiments can be visualized in Figure 3.12<sup>111</sup>. The newly assembled plasmids had always lower PCNs when compared with the parental plasmids pTRKH3 and pTRKH3-b, which could be probably due to the presence of the recombined forms that are inhibiting at some extent the plasmid replication. This low PCN is even more evident in the pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_1 and pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_1, which makes sense if one considers the several problematic mutations identified in both plasmids. As expected from the results from Chapter 2, the parental pTRKH3-b achieved the highest PCN (147.6 copies per chromosome), but after 8 h of growth<sup>111</sup>, instead of the 10.5 hours reported earlier. Some condition(s) in the growth experiments must have been uncontrollably different this time, to affect this parameter.



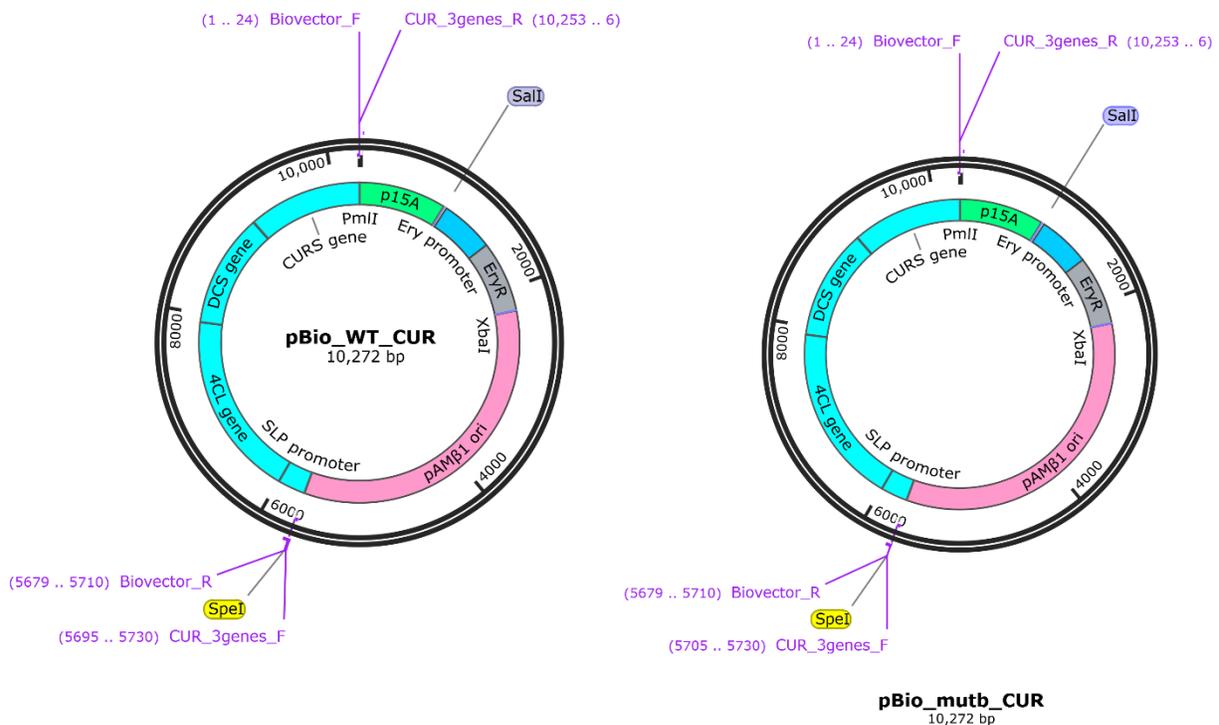
**Figure 3.12.** Average PCN values for *L. lactis* LMG19460 cells harboring the non-modified pTRKH3, the pTRKH3-b, the pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_1, the pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_1, pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_2 and pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_2 after 6h (light green), 8h (blue), 10.5h (yellow) and 24h of growth (dark green). The results were obtained from real-time qPCR analysis of three independent growths in 100 mL shake flasks with 75 mL of M-17 supplemented with 20 g/L glucose, 30°C, 100 rpm (initial OD<sub>600nm</sub>=0.1, starting from an overnight inoculum). The standard error of the mean (SEM) is represented for each point, except for the 8h and 24h points, which were quantified only once<sup>111</sup>.

### 3.4.1.2. Assembly of a prokaryotic protein expression cassette

After the master student under my co-supervision had assembled the previously described plasmids<sup>111</sup>, I attempted to assemble a prokaryotic protein expression cassette to the pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_2 and pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_2 plasmids, to construct the pBio\_WT\_CUR and the pBio\_mutb\_CUR. The unavoidable increase in the plasmid size after assembly of the new cassette could eventually solve the recombination problem verified after *L. lactis* LMG19460 transformation with the previously assembled plasmids. The bigger size could decrease the effect of the presence of direct or inverted repeats in the sequences that otherwise on smaller plasmids could improve the probability of occurring recombination events.

The chosen prokaryotic protein expression cassette was the one described on Chapter 5, used to produce curcumin, an anti-carcinogenic metabolite, in LAB. The minimum curcumin pathway comprises 3 genes: *4CL* (4-coumarate: CoA ligase)<sup>119</sup>, *DCS* (diketide-CoA synthase) and *CURS1* (curcumin synthase)<sup>120</sup>. The sequence of the first gene was obtained from *Lithospermum erythrorhizon* (Accession number D49366.1), while the other two genes were from *Curcuma longa* (Accession number AB495006.1 for *DCS* and AB495007.1 for *CURS1*). Ferulic acid is the substrate that has to be supplied to the growth media, in order to the enzyme expressed by the *4CL* gene to produce feruloyl-CoA. This resulting product will further act as substrate for the *DCS* enzyme, generating feruloyl-diketide-CoA, which is used as substrate by *CURS1* enzyme to produce curcumin. The 3 genes pathway is under the control of the surface (S)-layer protein (SLP) promoter, completing the prokaryotic protein expression cassette.

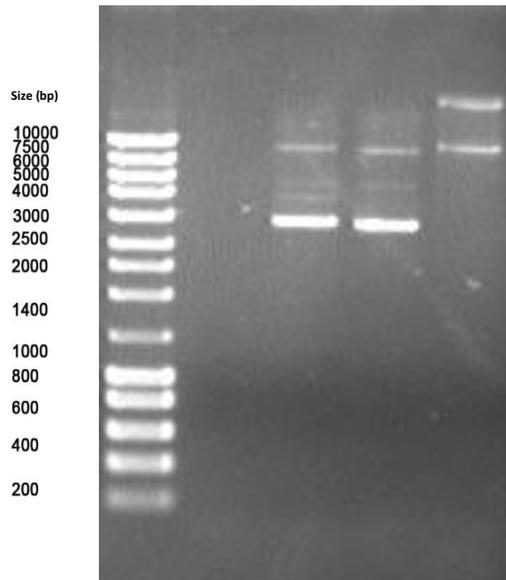
The *in silico* design of the pBio\_WT\_CUR and the pBio\_mutb\_CUR plasmids and the primers used to amplify each BioBrick part are represented in Figure 3.13.



**Figure 3.13.** *In silico* design of the pBio\_WT\_CUR and pBio\_mutb\_CUR plasmids. The primers used for the fragments amplification are also represented.

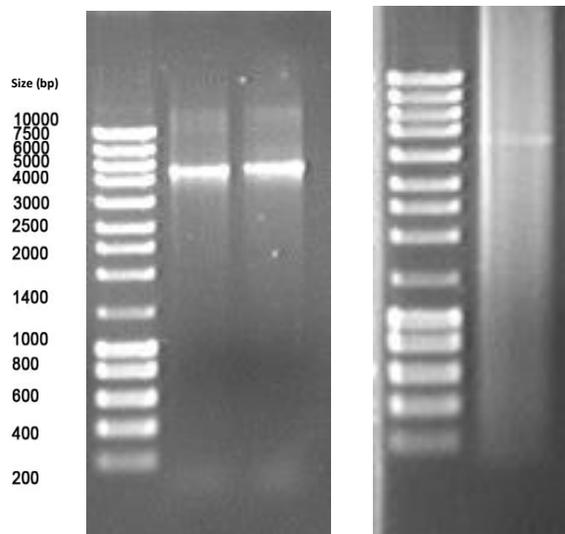
The first step was to check the integrity and quality of the template plasmids for the BioBricks amplification (Figure 3.14). The pBio\_erm/ery\_p15A\_pAMβ1\_WT\_2 and pBio\_erm/ery\_p15A\_pAMβ1\_mutb\_2 plasmids were the templates for the amplification of the vector constituted by the p15A Gram-negative origin of replication, the erythromycin resistance gene and the wild-type versus mutant b pAMβ1 origin of replication. The three genes cassette under the control of the SLP promoter was amplified from the pTRKH3\_SLP\_4CL\_DCS\_CURS1 plasmid. As can be visualized in the agarose gel, the template plasmids for the WT and mutb vectors had a good quality, with a majority

of the supercoiled isoform, while the template plasmid for the insert had less supercoiled isoform, but at least any of the plasmid was in the linear conformation.



**Figure 3.14.** Quality assessment of the template plasmids for the BioBricks amplification. Lanes: **1-** Nzyladder III; **3-** pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_2; **4-** pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_2; **5-** pTRKH3\_SLP\_4CL\_DCS\_CURS1.

The vector fragments (p15A Gram-negative origin of replication, the erythromycin resistance gene and the wild-type versus mutant b pAM $\beta$ 1 origin of replication), as well as the insert fragment with the three genes of the curcumin pathway under the control of the SLP promoter had the correct size (5,706 bp and 4,563 bp, respectively), as can be visualized in Figure 3.15.



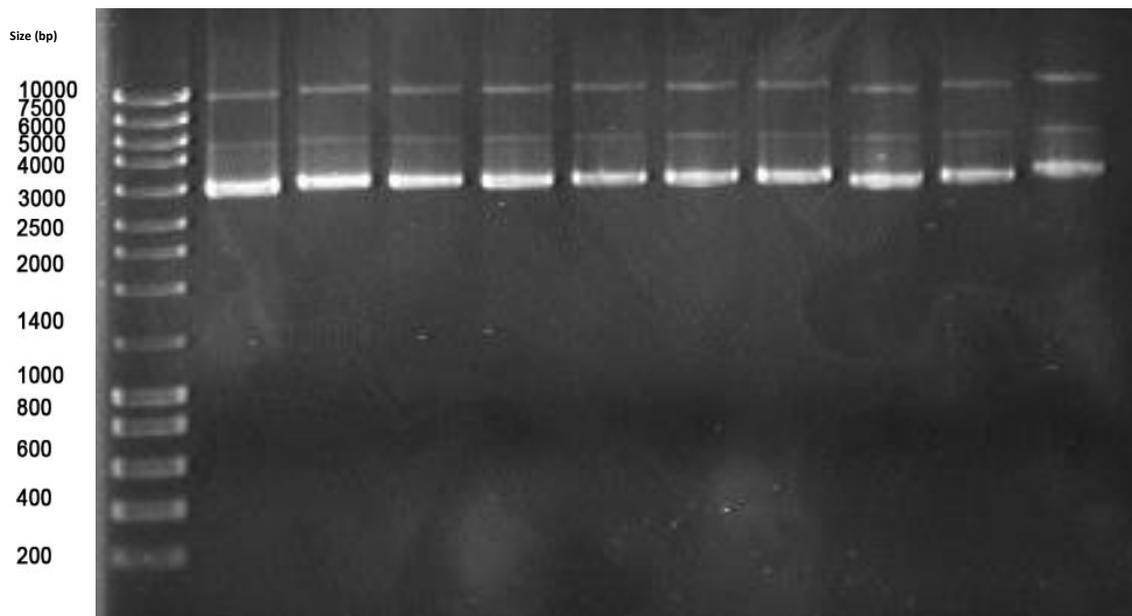
**Figure 3.15.** Amplification of the vectors and insert fragments for the pBio\_WT\_CUR and pBio\_mutb\_CUR construction. Lanes: **1-** Nzyladder III; **2-** vector amplified from pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_WT\_2; **3-** vector amplified from pBio\_erm/ery\_p15A\_pAM $\beta$ 1\_mutb\_2; **4-** insert amplified from the pTRKH3\_SLP\_4CL\_DCS\_CURS1 plasmid.

The vectors and insert fragments were excised and purified from the agarose gel, and then used to perform separate Gibson Assembly reactions. Several conditions and parameters of the Gibson Assembly reaction were optimized, such as reaction time (15 min, 1 hour or 4 hours), vector and insert DNA mass and molarity (Table 3.6).

**Table 3.6.** Mass and molar amount of each DNA fragment, as well as the vector to insert molar ratio used in the several Gibson Assembly reactions.

Experiment #	DNA fragments	DNA mass (ng)	DNA molar amount (pmol)	Vector:insert ratio
1	Vector WT	100.8	0.027	1:1
	Vector mutb	100.1	0.027	
	Insert	80.1	0.027	
2	Vector WT	239.2	0.065	1:1
	Vector mutb	239.2	0.065	
	Insert	190.9	0.065	
3	Vector WT	575	0.155	1:1
	Vector mutb	575	0.155	
	Insert	460	0.155	

After *E. coli* DH5 $\alpha$  transformation and colony screening by colony PCR, any of the conditions tested resulted in a successful assembly, as can be visualized in Figure 3.16, where the correctly assembled plasmids should have 10,272 bp, but clearly had the vector only size (5,706 bp). The vector fragment was enough to confer the resistance to erythromycin, being difficult to select for the colonies with the correctly assembled plasmids.



**Figure 3.16.** Colony PCR to check for the assembly of the vector (with the p15A origin of replication, erythromycin resistance gene and WT vs mutant b pAM $\beta$ 1 origin of replication) and insert (3 genes of the curcumin pathway under the control of the SLP promoter) BioBricks. Lanes: **1-** Nzy ladder III; **2-** colony 1; **3-** colony 2; **4-** colony 3; **5-** colony 4; **6-** colony 5; **7-** colony 6; **8-** colony 7; **9-** colony 8; **10-** colony 9; **11-** colony 10.

Further work is necessary on this section, in order to successfully obtain the pBio\_WT\_CUR and pBio\_mutb\_CUR plasmids. Different vector to insert molar ratios could be tested, as well as increased molarity values. Also, the solution could be to design new primers for the BioBricks amplification reactions, with longer overlapping regions.

After achieving the final plasmids, they should be used to transform *L. lactis* LMG19460. The new LAB strains should be grown in a systematic way, as the one described in the Materials and Methods section 3.3.6.1. The samples collected from this growth should be analyzed for its PCN, to assess if the increased size of the plasmids had solved the recombination issue verified in the LAB hosts and to confirm if the mutant b at least maintain its high PCN. Since this synthetically assembled plasmid is smaller than the original pTRKH3-b engineered on Chapter 2, lacking non-essential sequences from the wild-type pTRKH3, it is expected to have an even higher PCN. The plasmid size is not the only factor able to influence the PCN of this plasmid and other factors such as an alteration in the mRNA secondary structures could increase or decrease this value. This means that the future assembled plasmids should be rigorously analyzed at the gene, transcription and translation levels, besides the PCN quantification.

If the recombination problem does not subside only by the increase in plasmid size, the work proposed in Chapter 4, for the knockout of the recombinase genes should be first accomplished. It is expected that a Rec<sup>-</sup> strain do not recombine the exogenous plasmid DNA, solving the recombination issue. If the pDNA does not suffer recombination, its PCN will also probably increase to the expected values determined on Chapter 3.

If the final minimal plasmid with this prokaryotic curcumin expression cassette have indeed an increased PCN, it is expected to overcome some of the problems found on Chapter 5, namely the absence of curcumin production.

### 3.5. Conclusions

The main goal of this chapter was to engineer a minimal plasmid using Biobrick parts joined together by Gibson Assembly. The idea is to construct a high copy number minimal plasmid, suitable to be used with therapeutic purposes, such as for DNA vaccination or mucosal vaccination, using LAB as live delivery vectors. Gibson Assembly is a powerful technique to achieve this goal, since it enables the assembly of several Biobrick parts at the same time, with a high degree of efficiency and precision, and do not rely in restriction enzymes.

The *E. coli* and LAB replication cassettes (both wild-type pAM $\beta$ 1 and high PCN pAM $\beta$ 1-b were tested) were successfully assembled<sup>111</sup>. After transformation of *L. lactis* LMG19460 with the assembled plasmids, it was noticed a problem with plasmid recombination<sup>111</sup>, probably due to the fact that the strain is RecA<sup>+</sup>. This recombination is probably potentiated by the presence of inverted or direct repeats, which becomes more problematic when the plasmids have smaller sizes. The newly assembled plasmids presented lower PCN than the parental plasmids, probably due to the same recombination problems, and also mutations in some of the plasmids<sup>111</sup>, not being a favorable alternative to the pTRKH3-b already achieved in Chapter 2.

The next step was to assemble the LAB protein expression cassette with the 3 genes of the curcumin pathway under the control of the SLP promoter, to increase the plasmid size in the hope of overcome the recombination problem.

After several Gibson Assembly attempts, the correctly assembled plasmid is yet to be obtained and further work is necessary. If the recombination issue is not sorted out in this way, further analysis of the recombination mechanisms should be performed. In the future, when the LAB protein expression cassette is correctly assembled to the remaining minimal plasmid, its PCN should be evaluated and, if the recombination problem is solved solely by the plasmid size increase, it is expected higher PCN values. This could solve the problem of absence of curcumin production found on Chapter 5. If the recombination problem still persists, one possible solution is to first engineer a RecA<sup>-</sup> *L. lactis* LMG19460 strain, as purposed in Chapter 4.

Other proposal for future work relates with the testing of alternative parts in order to try to increase the PCN even further. The modulation of PCN is crucial not only for plasmid production but also for recombinant protein production. A low PCN is desirable when a tight control of the protein expression is needed, but a high PCN is preferred in view of higher plasmid yields. Our strategies to vary the copy number will rely on i) the change of the entire origin of replication (e.g. pWV01, pSH71), and ii) change specific parts of the pAMβ1 origin of replication, such as eliminate the truncated *copF* gene or the *CT-RNA* sequence.

Finally, an eukaryotic protein expression cassette could be assembled to the pBio\_erm/ery\_p15A\_pAMβ1\_WT\_2 and pBio\_erm/ery\_p15A\_pAMβ1\_mutb\_2 plasmids, for expression of vaccine antigens for DNA vaccination. Currently, LAB are very desirable hosts for DNA vaccines production due to its increased safety when compared with the traditional *E. coli* hosts, and to its usefulness as live delivery vectors.



# CHAPTER 4

*L. lactis* LMG19460 strain engineering

*Imagination will often carry us to worlds that never were. But without it we go nowhere. – Carl Sagan*

## 4.1. Abstract

The final goal of this chapter is to engineer the *Lactococcus lactis* LMG19460 strain by performing knockout or overexpression of the genes suggested by the *in silico* analysis, using the one-step CRISPR/Cas9 system. In the end we expect to have strains able to produce pDNA and recombinant proteins with high yields and quality.

To bias metabolism of wild-type Lactic Acid Bacteria (LAB) strains to overproduce pDNA and recombinant proteins, key genes like *endA* or *recA* genes that degrade or recombine exogenous DNA, or *htrA* that is responsible for the degradation of heterologous proteins, should be altered. Overexpression and/or knockout of other genomic LAB genes could also be beneficial as largely referred for *Escherichia coli*. The candidate genes were chosen using *in silico* analysis performed with the Optflux software, with the purpose of being modified using the CRISPR/Cas9 system.

After several preliminary attempts to use random mutagenesis and the Datsenko and Wanner recombineering strategy<sup>121</sup>, without success, the CRISPR/Cas9 system was chosen as the most suitable to perform the knockouts in this strain. Until now we were able to design and construct the appropriate plasmid based on pKCcas9dO, a one-step system originally designed for *Streptomyces*, for the endonuclease gene (*nth*) knockout in *L. lactis* LMG19460. The first constructed plasmid was the pKCcas9dO\_nth\_apra containing the apramycin resistance gene, which was not considerate appropriate for *L. lactis* LMG19460 selection, due to its high natural resistance. A second plasmid was designed, pKCcas9dO\_nth\_ery, with the erythromycin resistance gene instead the apramycin one. The successful construction of this plasmid is yet to be confirmed, as well as its ability to transform *L. lactis* LMG19460. The induction of the *tipA* promoter of the pKCcas9dO\_nth\_apra plasmid with the antibiotic thiostrepton was also attempted without success. Several possible problems and suggestions of overcoming them are discussed in the end of the chapter.

To the best of our knowledge this is the first attempt of performing a gene knockout using a one-step CRISPR/Cas9 approach in LAB.

## 4.2. Introduction

Although lactic acid bacteria are widely used as starters for food and beverages fermentations, the strains used are usually wild-type, without genetic modifications. Its wild-type status could difficult some molecular biology processes and desired modifications. For example, if the strain has active genes for endonucleases, its transformation with a synthetic plasmid will be very difficult or probably impossible. If the goal is for the strain to produce a heterologous protein, the production of proteases from the wild-type strain is not desirable. There are several synthetic biology tools that turn it possible to re-design and optimize the bacteria, by knocking-out non-essential genes, or overexpressing others, in order to redirect the Lactic Acid Bacteria (LAB) metabolism for high quality plasmid or protein production.

For *Lactococcus* and *Lactobacillus* genome editing there are currently several different tools available (Table 4.1). These systems can be used to perform genetic modifications in probiotic strains in order to improve its probiotic features, as well as to turn them able to deliver vaccines or modulate the host immune response, increasing even more its therapeutic potential<sup>122</sup>.

**Table 4.1.** Genome editing tools currently available for *Lactococcus* and *Lactobacillus* species<sup>123</sup>.

Tools	Species	Characteristics	References
Plasmid-based allelic exchange	- <i>Lactococcus lactis</i> - <i>Lactobacillus casei</i>	- Efficient transformation - Marker-free - Time consuming	124,125
Recombinase-mediated dsDNA recombineering	- <i>Lactobacillus plantarum</i>	- High efficiency for deletion and insertion - Marker-dependent	126
ssDNA recombineering	- <i>Lactococcus lactis</i> - <i>Lactobacillus reuteri</i> - <i>Lactobacillus plantarum</i> - <i>Lactobacillus gasseri</i>	- Genomic mutagenesis - Marker-free	127,128
CRISPR/Cas-assisted recombineering	- <i>Lactobacillus reuteri</i>	- High efficiency for small deletions (<1 Kb) - Marker-free	129
CRISPR/Cas9	- <i>Lactobacillus casei</i>	- Gene deletion and insertion - Simple editing procedure - Marker-free	121

### 4.2.1. ssDNA and dsDNA recombineering in LAB

Recombineering stands for recombination-mediated genetic engineering and could use single-strand DNA (ssDNA) or double-strand DNA (dsDNA). In ssDNA recombineering (SSDR) several proteins are necessary. The recombinases Beta from phage-derived lambda *red* locus and RecT from the Rac prophage are ssDNA binding proteins that promote complementary DNA annealing. The exonucleases Exo (from phage-derived lambda *red* locus) and RecE (from the Rac prophage) transform dsDNA molecules in ssDNA to allow Beta and RacT binding<sup>128</sup>. The expression of the recombinase (Beta or RecT) should be induced in the host cells and an oligonucleotide containing the desired mutations should be used to transform the same cells. In this way, SSDR could be used to introduce genomic point mutations without need for antibiotic selection, using for example PCR amplification instead<sup>128</sup>.

This type of recombineering has been widely used in *E. coli*<sup>121</sup>, but there are at least two reports of being used in LAB, namely in *Lactobacillus reuteri* and *Lactococcus lactis*, where small mutations were introduced in the cell genome with efficiencies ranging from 0.3% to 20%. Several RecT recombinases are available in LAB to perform SSDR and since its expression must be induced (high level expression reduced cell viability), a system like NICE (nisin controlled gene expression) should be used<sup>128</sup>.

dsDNA recombineering allowed DNA with lengths up to 2.5 kb to be inserted, using small homologous arms<sup>128</sup>. This system was successfully implemented in *Lactobacillus plantarum* to perform marker-free genetic manipulations by using homologous recombination between a dsDNA substrate and the cell genomic DNA. This system requires 3 proteins: the exonuclease Lp\_0642, the host nuclease inhibitor Lp\_0640 and the single strand annealing protein Lp\_0641, which have homology with Exo, Gam and RecT<sup>126</sup>.

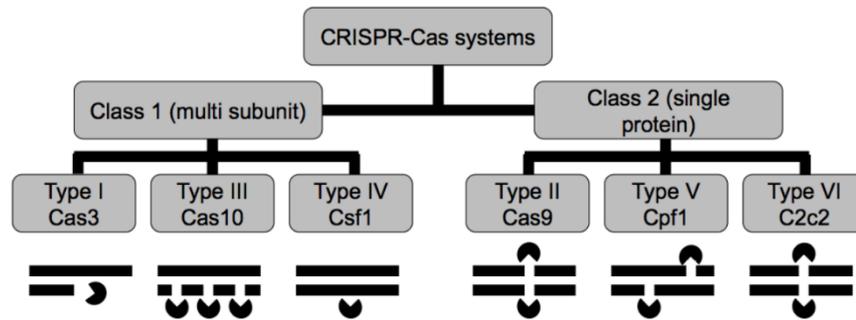
#### 4.2.2. CRISPR/Cas9 system

One of the available molecular tools is the CRISPR/Cas9, a cutting edge and precise technology for genome engineering and gene knockout<sup>128</sup> and specifically in LAB and even Gram-positive bacteria there were only few studies<sup>129,130</sup>.

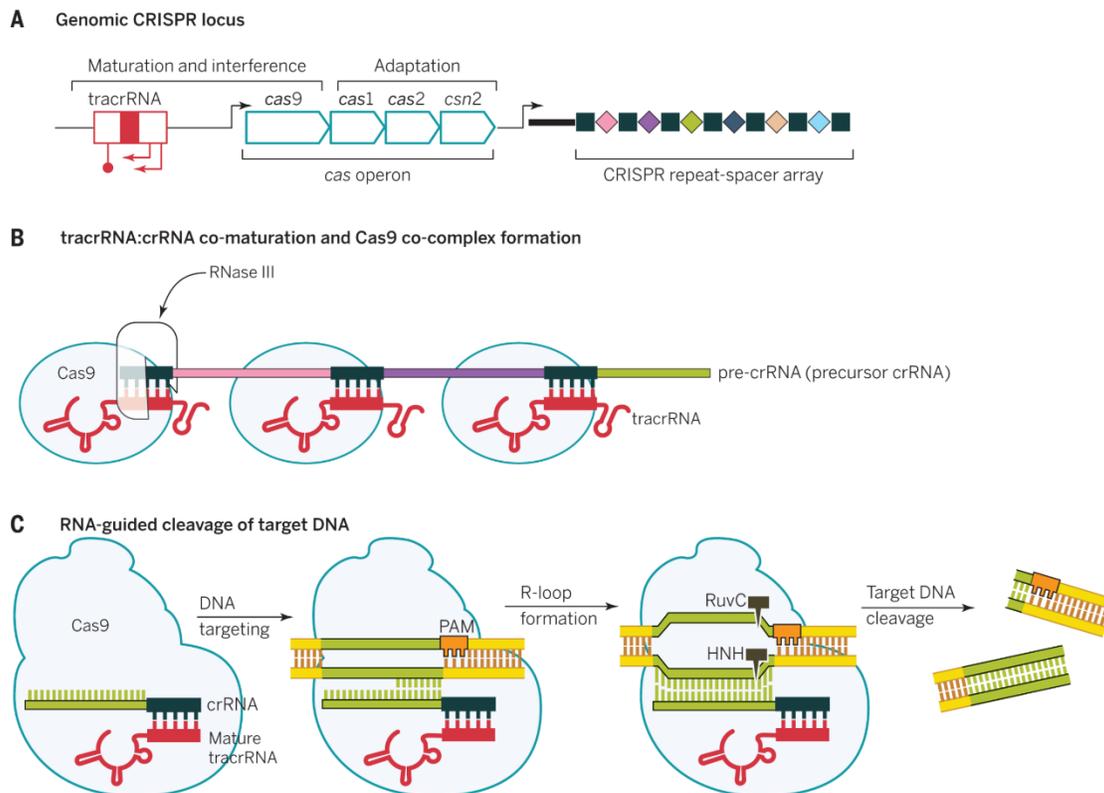
CRISPR accounts for clustered interspaced short palindromic repeats, while Cas is the CRISPR-associated enzyme<sup>131</sup>.

In its essence, the CRISPR/Cas system is a type of adaptive immune system naturally present in bacteria and archaea, which cleaves foreign DNA (bacteriophages or plasmids) when it enters the cells<sup>130,131</sup>. When exogenous DNA enters the bacterial cells, what happens is the incorporation of small DNA fragments between the CRISPR repeats. When occurs the transcription of the CRISPR locus, a small CRISPR-RNA (crRNA) is expressed, leading to the formation of foreign DNA-crRNA duplexes in a sequence specific manner. The crRNA will guide the Cas nuclease towards cleavage of the DNA complementary target sequence, based on the Watson-Crick base pairing between the guide sequence of crRNA and protospacer of target DNA, featured with an adjacent protospacer adjacent motif (PAM). In the end, it will avoid the foreign DNA replication inside the host cell<sup>123,131</sup>.

There are two classes and six different types of CRISPR/Cas systems, being the type II the most widely used in the genome engineering field (Figure 4.1), due to its high programmability, precision and efficiency<sup>122,130</sup>. The two classes differ in the nature of its effector proteins responsible for the targeting: if it is a multisubunit complex, it belongs to the class 1; if it is a single protein, belongs to the class 2 system. In the most studied type II system, the Cas9 endonuclease complexes with a dual-RNA, constituted by a crRNA that directs the Cas nuclease to the target site and a *trans*-activating crRNA (tracrRNA), required for Cas9 nuclease activity. This dual-RNA is necessary for the Cas9 to recognize and generate a double strand blunt break in a specific target sequence (Figure 4.2)<sup>130,132</sup>. On type VI the cleavage of the DNA is also blunt, while the type V performs a sticky-end dual nicking. Considering the class 1 effector proteins, the type I generates single-strand exonucleolytic cleavages, while type III completely shreds the DNA and type IV acts in an unknown way<sup>131</sup>.



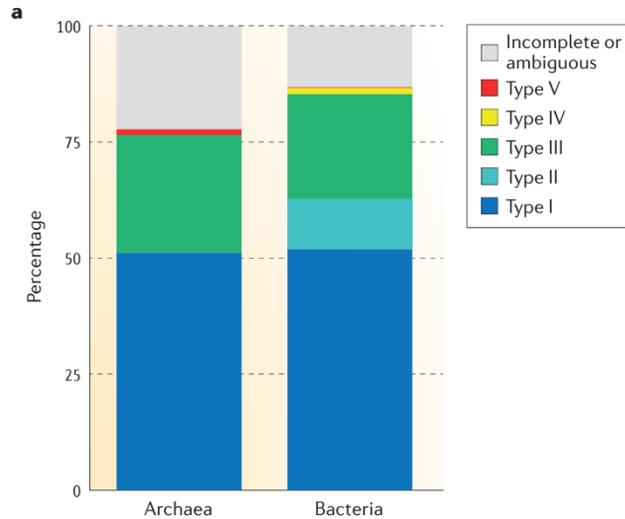
**Figure 4.1.** Two classes and six types of CRISPR/Cas systems. The classes differ on the nature of the effector proteins, while the types depends on the type of cleavage they perform on the DNA. Figure from Pijkeren and Barrangou (2017)<sup>131</sup>.



**Figure 4.2.** Overview of the *S. pyogenes* type II CRISPR/Cas9 system. A) The CRISPR locus in *S. pyogenes* genome, with the *cas* genes operon, tracrRNA and CRISPR repeat-spacer array; B) After locus expression, Cas9 associates with the tracrRNA:crRNA duplex (processed by ribonuclease III); C) The RNA-guided target DNA is cleaved by Cas9. Figure from Doudna and Charpentier (2014)<sup>132</sup>.

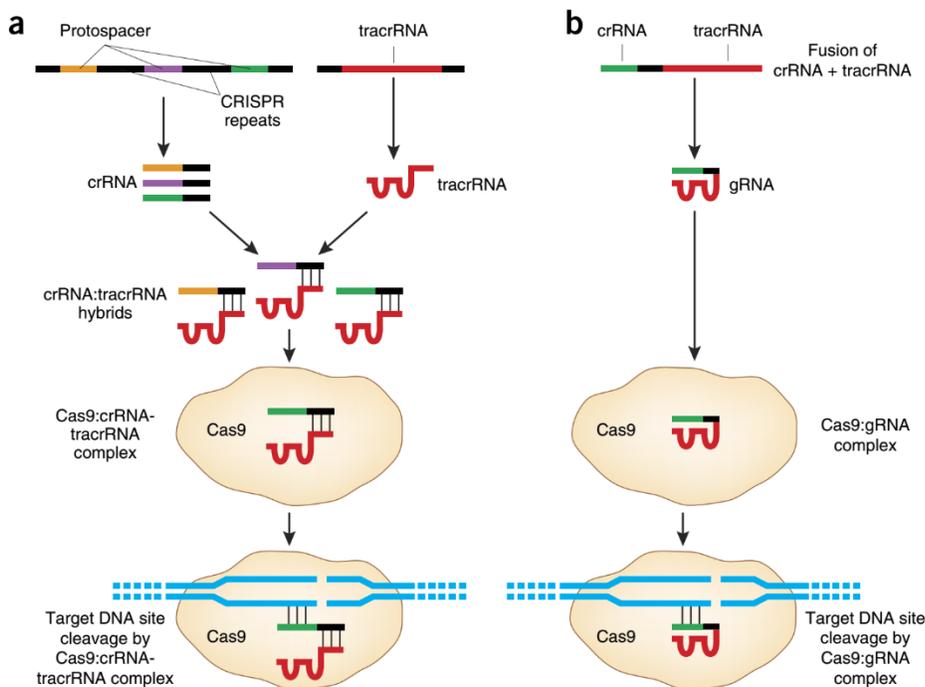
The double strand blunt break performed by the type II Cas9 nuclease is then repaired by the cell DNA repair machinery, namely non-homologous end joining (NHEJ) or homology-directed repair (HDR)<sup>132</sup>.

The different types of CRISPR/Cas systems is not distributed in even amount throughout the different bacterial strains, existing a predominance of type I and III systems, followed in third place by the most studied type II (Figure 4.3)<sup>133</sup>. 50% of all the bacterial genomes already sequenced contained CRISPR/Cas systems, against 87% of the archaeal genomes<sup>133</sup>.



**Figure 4.3.** Distribution of the different types of CRISPR/Cas systems in bacterial and archaea sequenced genomes. Figure from Makarova *et al.* (2015)<sup>133</sup>.

In the genetic engineering field, this system is used for genome editing, by using the cell repair machinery for replacement of the wild-type DNA target sequence by an edited one<sup>131</sup>. More specifically, the protospacer sequence is replaced by the sequence of interest, which could be for example an antigenic construct that will alter the bacterial cell surface composition, to be used as probiotic oral vaccination<sup>122,123</sup>. Usually, the crRNA:tracrRNA duplex is replaced by an artificially designed chimeric single guide RNA (sgRNA)<sup>123,134</sup> and could be designed in order to direct the Cas9 nuclease activity for a sequence of interest (Figure 4.4).



**Figure 4.4.** Differences between A) the naturally occurring CRISPR/Cas9 systems and B) the engineered ones, where the tracrRNA:crRNA duplex is replaced by a sgRNA. Figure from Sander and Joung (2014)<sup>134</sup>.

It could be also repurposed and used as a strain-specific antimicrobial, since many bacteria lack efficient non-homologous end joining systems, being unable to repair the damage inflicted by the Cas nuclease if the target sequence is in its genome<sup>131</sup>.

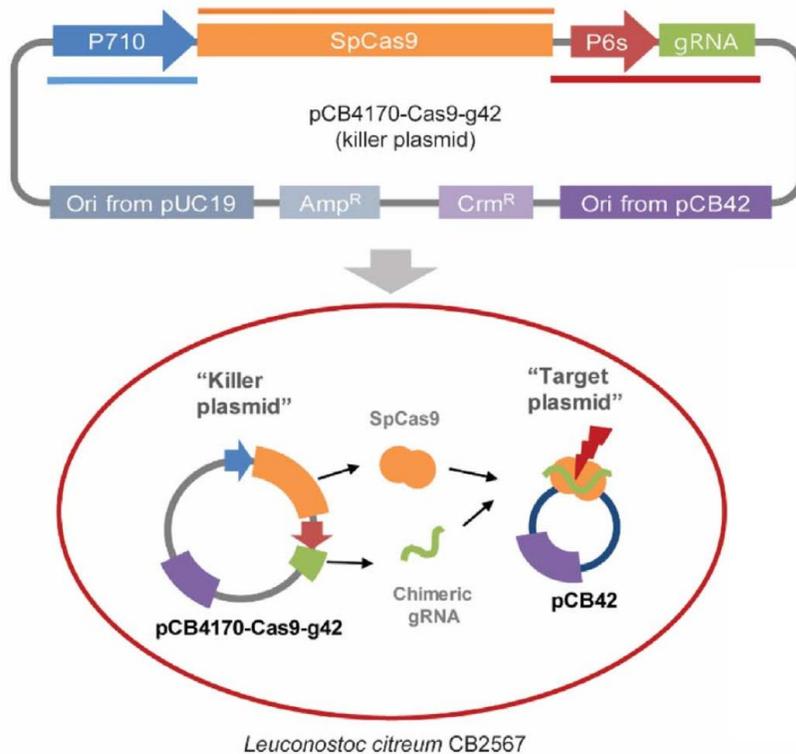
While CRISPR/Cas9 are already widely studied in eukaryotic systems, with the goal of curing human, animal and plant diseases, studies in bacterial genome editing are still quite scarce.

#### 4.2.3. CRISPR/Cas9 applications in LAB

The CRISPR/Cas systems are widely distributed in LAB, since they are frequently object of horizontal gene transfer and phage attacks in dairy environments, gastrointestinal tract and fermented food. Around 63% of all *Lactobacillus* sequenced genomes encodes CRISPR/Cas systems, against 46% of the total bacterial genomes<sup>131</sup>. This means that *Lactobacillus* strains are a good source of native Cas enzymes and also good targets for genome editing. But on the other side, there are limited genetic tools available to use in LAB genome editing, which had impaired the investigation and biotechnological application of many LAB strains<sup>123</sup>. Curiously, these CRISPR/Cas systems are rarely present in *L. lactis* strains<sup>122</sup>.

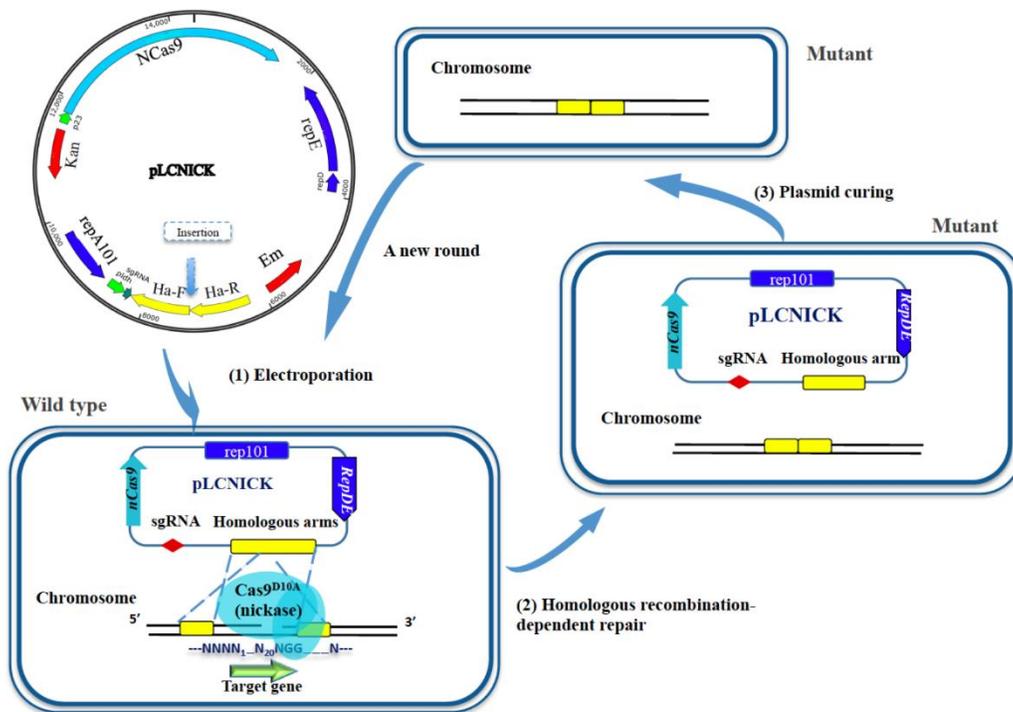
One of the most recent uses of the CRISPR/Cas9 system in LAB were in *Leuconostoc citreum* CB2567, for eliminating the non-curable cryptic plasmid pCB42, which presence could impair the transformation with other plasmids of

interest<sup>130</sup>. First, the strain was transformed with a killer plasmid, containing the *Streptococcus pyogenes* system, with the *cas9* endonuclease gene and a guide RNA (gRNA) that targets for a specific sequence of the cryptic plasmid (Figure 4.5). Once the cryptic plasmid was successfully eliminated, the killer plasmid was cured by serial subculturing without antibiotic. If an editing DNA template and a homologous recombining system could be added to the plasmid harboring the *cas9* and the gRNA, the double strand DNA break in the genome caused by Cas9 could be repaired and new DNA introduced, being a powerful genome engineering tool<sup>130</sup>.

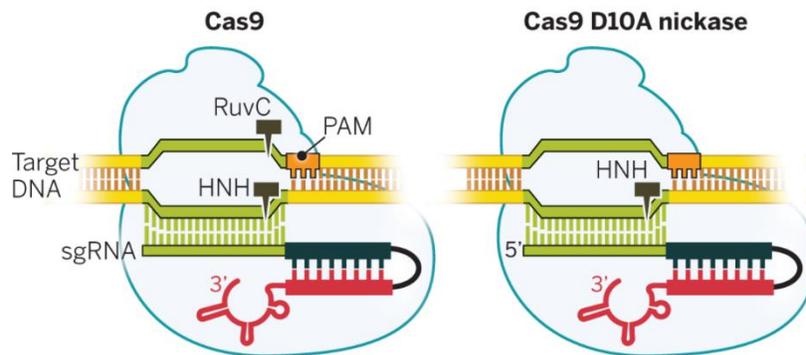


**Figure 4.5.** CRISPR/Cas9 system used to eliminate a cryptic plasmid from *Leuconostoc citreum* CB2567<sup>130</sup>. The authors engineered a synthetic vector with the *cas9* gene from *Streptococcus pyogenes* and a chimeric gRNA with a 20 nucleotide guiding sequence that targeted the pCB42 cryptic plasmid. After strain transformation with this synthetic vector, Cas9 endonuclease cleaved the cryptic plasmid, guided by the gRNA. Figure from Jang *et al.* (2017)<sup>130</sup>.

Other recent application of a CRISPR/Cas9 system to a LAB strain was for genome editing in *Lactobacillus casei*<sup>123</sup>. The authors engineered the genome of this strain by using the plasmid pLCNICK that harbored the *cas9*<sup>D10A</sup> gene, the sgRNA expression cassette and homologous arms of the target gene as homologous recombination repair templates (Figure 4.6). This Cas9<sup>D10A</sup> performs single-strand breaks (nicks) in the target DNA (Figure 4.7) and the cell repair them by the high-fidelity base excision repair pathway (BER) or homology-directed repair (HDR), being less toxic and lethal for the cell when compared with NHEJ and HDR of chromosomal double strand breaks. This plasmid proved to be a more precise, effective and rapid tool for genome editing, than the classic homologous recombination-dependent double crossover strategy<sup>123</sup>. The disadvantages of this plasmids relies in its inability to perform deletions of sequences bigger than 5 Kb.



**Figure 4.6.** Overview of the CRISPR/Cas9<sup>D10A</sup> system used in *Lactobacillus casei*<sup>123</sup>. The plasmid pLCNICK containing the *cas9<sup>D10A</sup>* gene, as well as the sgRNA and the homologous arms is used to transform *L. casei* by electroporation. The nickase Cas9<sup>D10A</sup> is guided by the sgRNA to the target gene in the bacterial genome and generates a nick in the chromosome, triggering the cell homologous recombination-dependent repair process, where the homologous arms work as the repair donors. In the end, the pLCNICK plasmid is cured and the strain is ready for a new round of genome editing. Figure from Song *et al.* (2017)<sup>123</sup>.



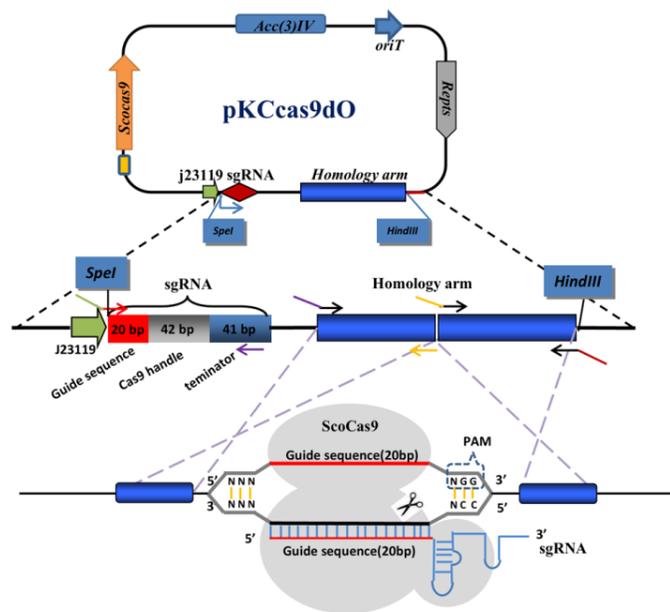
**Figure 4.7.** Comparison between Cas9 nuclease and Cas9<sup>D10A</sup> nickase. Figure from Doudna and Charpentier (2014)<sup>132</sup>.

A less recent but also relevant example of CRISPR/Cas9 use in LAB is the employment of a CRISPR/Cas9 system combined with the ssDNA recombineering in *Lactobacillus reuteri*<sup>129</sup> to perform codon saturation mutagenesis in *L. reuteri* chromosome. The authors used CRISPR/Cas9 combined with ssDNA recombineering to select and remove the unedited cells. What happens is that the CRISPR/Cas9 system kills the wild-type cells and the population becomes enriched in the low efficiency mutants<sup>129</sup>.

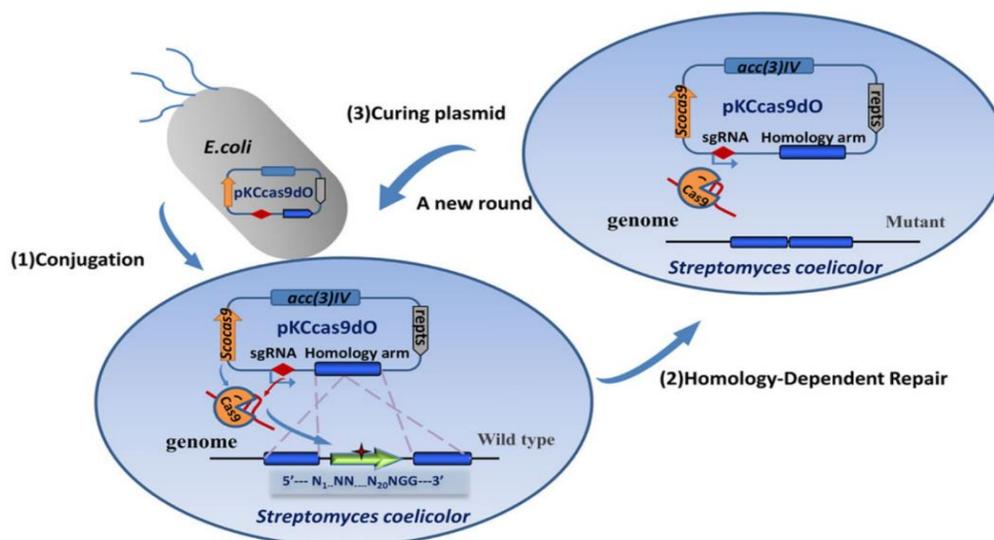
#### 4.2.4. CRISPR/Cas9 system used in the present work

The CRISPR/Cas9 system used in the present work is based on the system engineered and developed by Huang *et al.* (2015)<sup>135</sup>. Although *Streptomyces coelicolor* do not belong to the LAB group, they are all Gram-positive bacteria, and this system had the advantages of being commercially available and being a highly efficient and fast one-step system. The authors achieved high efficiencies (60-100%) for single gene deletion, large-size gene cluster deletion, multiple gene deletions and introduction of nucleotide point mutations. It is also a fast system, since only one step of transformation is required, since all the necessary elements are in a single plasmid, and in the end of each genome editing round, the plasmid can be cured by increasing growth temperature to 37°C<sup>135</sup>.

The plasmid pKCcas9dO (Addgene no. 62552) includes a *S. coelicolor* codon-optimized *cas9* gene (*Scocas9*), a target specific gRNA and two homology-directed repair templates (Homology arms 1 and 2) (Figure 4.8). The plasmid also has a thermo-sensitive replication origin, to allow an easy curing, an origin of transference for conjugation and an apramycin resistance gene<sup>135</sup>. Gene-specific sgRNA and homology arms could be cloned between *SpeI* and *HindIII* restriction enzyme sites. The sgRNA is under the control of the J23119 promoter and is constituted by a gene-specific 20 bp guide sequence, a 42 bp Cas9 handle and a 41 bp terminator. The *ScoCas9* recognizes the chromosomal sequence complementary to the 20 bp guide sequence, which harbors upstream the PAM sequence in *S. coelicolor* genome, and performs a double strand cut, which is then repaired by the cell homology-dependent repair system. In the end of each round, the plasmid is cured by growing at 37°C (Figure 4.9). Both *Scocas9* and sgRNA are constitutively expressed in the cells<sup>135</sup>.



**Figure 4.8.** Design and function of the pKCcas9dO plasmid. *Scocas9*: *S. coelicolor* codon-optimized *cas9*; *Acc(3)IV*: apramycin resistance gene; *oriT*: origin of transference for conjugation; *Repts*: thermo-sensitive origin of replication; *j23119*: promoter controlling the expression of sgRNA; sgRNA: single guide RNA constituted by a 20 bp guide sequence, a 42 bp Cas9 handle and a 41 bp terminator; homology arms: sequences homologous to the upstream and downstream regions of the target gene for knockout. Figure from Huang *et al.* (2015)<sup>135</sup>.



**Figure 4.9.** Overview of the knockout process in *S. coelicolor* using the pKCcas9dO plasmid. The plasmid enters the bacterial cells by conjugation, using *E. coli*. The Cas9 endonuclease associates with the sgRNA, recognizes the chromosomal site complementary to the guide sequence and performs a double strand break in the *S. coelicolor* genome, activating the cell homology-directed recombination repair system. In the end, the target gene has been removed by two rounds of crossovers and the plasmid is cured by increasing the growth temperature to 37°C. Figure from Huang *et al.* (2015)<sup>135</sup>.

#### 4.2.5. Strain improvement methods without the use of recombinant DNA technology

Strain improvement (removal of undesirable traits or acquisition of desirable characteristics) could be performed without resort to recombinant DNA technology. One of the most used methods is the random mutagenesis, by using specific chemicals (e.g. ethidium bromide) or UV light exposure. On the down side, it is a method where one cannot control the mutations that will appear and a large number of mutants must be screened in order to find one with the desired mutation<sup>136</sup>.

Other technique is directed evolution that consists in slowly adapt a strain to specific growth conditions and, after several generations, the population becomes enriched in strains with the desired property. But unintended mutations can also accumulate, being the major disadvantage<sup>136</sup>.

Dominant selection is another technique, where selection of strains with a specific desired trait or mutation is promoted, but a profound knowledge in bacterial physiology is necessary<sup>136</sup>.

These techniques had been widely used, mainly in the food industry, with different goals, such as increasing bacteriophage resistance, remove antibiotic resistance, elimination of unwanted metabolic pathways, improve growth yields during fermentation, increase ethanol or bile tolerance, reduce product acidification to increase its stability, improve strains for enhancement of texture properties in yogurts and improve stress tolerance<sup>136</sup>.

## 4.3. Materials and methods

### 4.3.1. Bacterial strains and plasmids

*Escherichia coli* DH5 $\alpha$  (Invitrogen), *E. coli* GM2163 or *E. coli* SCS110 were used for all the molecular cloning steps performed in this chapter and the resulting plasmids were transferred for *L. lactis* LMG19460 by electroporation (BCCM Culture Collection, Belgium). Both *E. coli* GM2163 and SCS110 were used because of its ability to generate pDNA free from Dcm and Dam methylation, meaning that the DNA can be digested by methylation-sensitive restriction enzymes and LAB transformation efficiency could be improved.

The plasmids used in the recombineering strategy to generate knockouts in *L. lactis* LMG19460 were pKD13, pKD46 and pCP20. The pKD13 plasmid contains the Kan-cassette flanked by FRT sites, while pKD46 contains the  $\lambda$ -Red recombination system under the control of an arabinose inducible promoter and pCP20 has the yeast *flp* recombinase gene.

The plasmid employed as the main backbone for the CRISPR/Cas9 system was pKCcas9dO (Addgene no. 62552) (Figure 4.8)<sup>135</sup>. This plasmid has a temperature-sensitive pSG5 origin of replication that allows for an easy curing procedure, by simply growing the host strain at 37°C. pKCcas9dO also harbors an origin of transfer (*oriT*), necessary for transferring DNA from a bacterial host to a recipient by a process called conjugation. The authors that developed this plasmid used conjugation to transfer the engineered plasmid from *E. coli* to the final host *S. coelicolor*, but in the present work this strategy will not be used, although the *oriT* will be maintained in the plasmid. It also have an apramycin resistance gene, which confers resistance to the antibiotic apramycin, belonging to the class of the aminoglycosides that are inhibitors of protein synthesis initiation.

The same plasmid have several components essential to the CRISPR/Cas9 system, namely the *cas9* gene that was codon-optimized for *S. coelicolor* (*Scocas9*) and which expression is under the control of the *tipA* promoter. This *tipA* promoter is inducible by thiostrepton, an antibiotic naturally produced by some strains of *Streptomyces*. The gene-specific sgRNA is under the control of the *j23119* constitutive promoter. The sgRNA is constituted by three different parts: a gene-specific 20 bp guide sequence, a 42 bp Cas9 handle and a 41 bp terminator. Two homology arms were cloned downstream the sgRNA sequence, having homology to the flanking sequences of the gene of interest.

The characteristics of all the strains and plasmids used are summarized in table 4.2.

**Table 4.2.** Main characteristics of the strains and plasmids used in this chapter.

Strains	Relevant genotype	Source
<i>E. coli</i> DH5 $\alpha$	F- $\Phi$ 80 <i>lacZ</i> $\Delta$ M15 $\Delta$ ( <i>lacZYA-argF</i> ) U169 <i>recA1 endA1 hsdR17</i> (rK-, mK+) <i>phoA supE44</i> $\lambda$ - <i>thi-1 gyrA96 relA1</i>	Invitrogen
<i>E. coli</i> GM2163	<i>ara-14 leuB6 fhuA31 lacY1 tsx78 glnV44 galK2</i> <i>galT22 mcrA dcm-6 hisG4 rfbD1</i> <i>R(zgb210::Tn10)TetS endA1 rpsL136 dam13::Tn9</i> <i>xylA-5 mtl-1 thi-1 mcrB1 hsdR2</i>	New England Biolabs (ER2925)

<i>E. coli</i> SCS110	<i>rpsL (Strr) thr leu endA thi-1 lacY galK galT ara tonA tsx dam dcm supE44 Δ(lac-proAB) [F' traD36 proAB lacIq ZΔM15]</i>	Stratagene (200247)
<i>L. lactis</i> LMG19460	Plasmid-free strain; isolated from starter cultures of German cheese factories	LMG/BCCM
<b>Plasmids</b>	<b>Construct</b>	<b>Source</b>
pKD13	oriR6Kgamma, Kan <sup>R</sup> flanked by FRT sites, Amp <sup>R</sup>	121
pKD46	oriR101 w/repA101ts, Amp <sup>R</sup> , Gam-beta-exo proteins under the control of arabinose inducible promoter	121
pCP20	FLP <sup>+</sup> , λ cI857 <sup>+</sup> , λ p <sub>R</sub> Rep <sup>ts</sup> , Cm <sup>R</sup> , Amp <sup>R</sup>	137
pKCcas9dO	pSG5 ori, Apm <sup>R</sup>	Addgene (62552)
pTRKH3	P15A ori, pAMβ1 ori, Tet <sup>R</sup> , Ery <sup>R</sup>	LMBP/BCCM (LMBP 4462)
pKCcas9dO_nth_apra	pKCcas9dO derivative with specific sequences for <i>nth</i> gene knockout; Apm <sup>R</sup>	This study
pKCcas9dO_nth_ery	pKCcas9dO derivative with specific sequences for <i>nth</i> gene knockout; Ery <sup>R</sup>	This study

#### 4.3.2. Media and growth conditions

20 g/L of Luria-Bertani (LB) broth (Nzytech) supplemented with 25 µg/mL apramycin (Sigma) was used for growing *E. coli* DH5α, GM2163 or SCS110 harboring the pKCcas9dO or the pKCcas9dO\_nth\_apra plasmids, at 30°C, 100 rpm. The growth of *E. coli* DH5α with the pTRKH3 or pKCcas9dO\_nth\_ery plasmids was performed with the same media, but supplemented with 500 µg/mL erythromycin (Sigma) and at 37°C, 250 rpm.

*L. lactis* LMG19460 were grown at 30°C, 100 rpm, using M-17 (pH 7.0, Fluka) supplemented with 20 g/L glucose (Fisher Scientific), and 500 µg/mL apramycin (Sigma) when transformed with the pKCcas9dO or the pKCcas9dO\_nth\_apra plasmids.

When necessary, *E. coli* cells were grown in LB agar (2%) plates with the appropriate antibiotic, at 30°C, while *L. lactis* cells were grown in solid regeneration medium<sup>112</sup> containing (per liter) 10 g of tryptone (BD Biosciences), 5 g of yeast extract (Liofilchem), 200 g of sucrose (Fisher Scientific), 10 g of glucose, 25 g of gelatin (Merck), 15 g of agar (JMV Pereira), 2.5 mM of MgCl<sub>2</sub>·6H<sub>2</sub>O (Fagron), 2.5 mM of CaCl<sub>2</sub> (V. Reis) and 500 mg apramycin, at 30°C.

Master and working cell banks of all the strains were made with a final glycerol concentration of 20%, and stored at -80°C.

#### 4.3.3. Competent cells and transformation

##### 4.3.3.1. *E. coli* DH5α, GM2163 and SCS110

Chemically competent *E. coli* DH5α, GM2163 or SCS110 cells were prepared using cells grown until reach an OD<sub>600nm</sub> of 1, in a 100 mL shake flask with 20 mL LB, at 37°C and 250 rpm (initial OD<sub>600nm</sub> =0.1, from an overnight pre-inoculum). Freshly grown cells were centrifuged (1,000 g, 10 min, 4°C) and resuspended in 0.1 volumes (2 mL) of

chilled TSS medium (20 g/L LB, 5% DMSO, 50 mM MgCl<sub>2</sub>, 10% PEG 8000 (w/v), pH=6.5). After 10 min on ice, 100 µL aliquots were made and stored at -80°C.

In order to transform DH5α, GM2163 or SCS110 cells with pDNA by heat shock, for every 100 µL of chemically competent cells, the appropriate amount of pDNA was added, accordingly with each experiment. The mixture was incubated for 30 min on ice, and then submitted to 42°C for 60 s. The cells were immediately incubated for 2 min on ice, and then resuspended with 900 µL of LB medium. After 1 hour of recuperation at the correct temperature, the cells were plated on LB agar with the appropriate antibiotic, accordingly with the experiment, and incubated overnight at the correct temperature.

#### **4.3.3.2. *L. lactis* LMG19460**

The procedure for making electrocompetent *L. lactis* LMG19460 were adapted and optimized from Holo & Nes (1989) protocol<sup>112</sup>. An overnight growth in 5 mL of M-17 supplemented with 0.5% glucose (GM-17), 30°C, 100 rpm, is used to inoculate a 100 mL shake flask with 75 mL of GM-17 with a starting OD of 0.1, using the previously determined correlation of 1 OD unit at 600 nm being equivalent to 7x10<sup>8</sup> cells/ml<sup>80</sup>. After growing at 30°C, 100 rpm until reaching an OD<sub>600nm</sub>=0.5-0.8, the whole broth is diluted 100-fold to fresh GM-17 medium (750 µL to 75 mL) supplemented with 0.5 M sucrose (SGM-17) and 1-2% glycine (Sigma). The cells were grown at 30°C, 100 rpm until reach an OD<sub>600nm</sub>=2-2.5 and harvested by centrifugation at 6,000g, 3 min, 4°C. Next, the cells were washed 2 times in an ice cold solution containing 0.5 M sucrose and 10% glycerol (VWR Chemicals) (washing solution). Finally, they were resuspended in 1/100 of the initial culture volume of the washing solution and stored at -80°C.

For transformation of the electrocompetent *L. lactis* LMG19460 cells, 40 µL aliquots containing approximately 1x10<sup>9</sup> cells were mixed with the appropriate volume of pDNA and transferred to a 2 mm electroporation cuvette. After 30 min on ice, 2-3 pulses of 1,000 V were given using an electroporator (BTX ECM399), resulting in time constants of 4 to 5 ms. To recover from the electroporation, 960 µL of recovery medium (ice-cold SGM-17 supplemented with 20 mM MgCl<sub>2</sub> and 2 mM CaCl<sub>2</sub>) was immediately added. After resting 5 min on ice, the mixture was incubated 3h at 30°C without agitation. Then, the cells were collected by centrifugation and used to inoculate 5 mL of M-17 supplemented with 20 g/L glucose and a sub-lethal apramycin concentration<sup>113</sup> (250 µg/mL), at 30°C without agitation. After an overnight growth, the cells were again collected by centrifugation and plated in regeneration medium<sup>112</sup> containing (per liter) 10 g of tryptone, 5 g of yeast extract, 200 g of sucrose, 10 g of glucose, 25 g of gelatin, 15 g of agar, 2.5 mM of MgCl<sub>2</sub>.6H<sub>2</sub>O, 2.5 mM of CaCl<sub>2</sub> and 500 mg apramycin, at 30°C.

#### **4.3.4. *L. lactis* LMG19460 genome characterization**

The *L. lactis* subsp. *lactis* LMG19460 genome was sequenced in the context of the present thesis, in cooperation with the Instituto Gulbenkian de Ciência (Portugal) (accession number MUBH01000000)<sup>4</sup>. The 41 generated contigs were further analyzed with BASys<sup>138</sup>, an online platform that allows the annotation of bacterial genomes, determining

several features such as COG function, protein length and amino acid residues distribution. The presence of prophages was determined using PHAST (PHAge Search Tool)<sup>139</sup>.

#### 4.3.5. *In silico* genomic analysis using Optflux software

In order to find genomic candidate genes from *L. lactis* LMG19460 for knockout or overexpression, relevant for plasmid production or protein expression, simulations were performed using Optflux 3 software (SilicoLife, Universidade do Minho, Portugal)<sup>140</sup>, using the iAP358 metabolic model from *Lactococcus lactis* subsp. *lactis* IL1403. This software allows *in silico* metabolic engineering and the relevant module for the present task was the strain optimization module, which allowed to simulate which gene or reaction knockout leads to the highest increase in biomass and pDNA production. The input for the substrate was glucose (R\_Ex\_glc\_LPAREN) and the software was asked to maximize the objective function Biomass-Product Coupled Yield, where was chosen to maximize both biomass and cytidine triphosphate (CTP – R\_CTPS1\_1 and R\_CTPS2) or NrdD\_1. CTP is a pyrimidine nucleoside triphosphate, while NrdD\_1 codes for an anaerobic ribonucleoside-triphosphate reductase that is involved in the pyrimidine and purine synthesis pathways, meaning that production of both products is positively correlated with pDNA synthesis. When desired, the software automatically ignored knockouts of genes or reactions essential for cell viability (critical genes/reactions). Also drain and transport reactions were excluded from the simulation. A maximum of 6 simultaneous gene knockouts and 5000 solutions evaluations were imposed to the software calculations.

The genes obtained after the simulation were then analyzed using Kyoto Encyclopedia of Genes and Genomes (KEGG), more specifically the KEGG pathway module<sup>141,142,143</sup>, in order to identify the pathways were they intervene.

#### 4.3.6. Gene knockout using random mutagenesis

*L. lactis* LMG19460 harboring the pTRKH3 plasmid was grown overnight in 5 mL of De Man, Rogosa and Sharpe (MRS) medium supplemented with 5 µg/mL erythromycin, at 30°C, 100 rpm. Several serial dilutions were performed in order to plate 10,000, 1,000, 100 and 10 cells in MRS agar supplemented with 5 µg/mL erythromycin. The plates were exposed to 254 nm UV light during 60s, at 6 cm away from the UV source. Immediately afterwards the plates were kept at dark, at 30°C. From the 1,000 cells condition, two additional control plates were made, one negative control that were not exposed to the UV light and other plate that was exposed in the same conditions as described above, except for the distance that was 0 cm. This random mutagenesis protocol was adapted from Kadam *et al.* (2006)<sup>144</sup> and Patel *et al.* (2010)<sup>145</sup>.

After two days growing at 30°C, the colonies in the plates were counted and grown in 5 mL of MRS medium supplemented with 5 µg/mL erythromycin, at 30°C, 100 rpm. After achieving the growth stationary phase, the pDNA was purified using the NucleoSpin Plasmid kit (Macherey-Nagel). Its concentration was quantified using a Nanodrop Spectrophotometer (Nanovue Plus, GE) and the pDNA quality was visualized by placing 1 µg of each sample in an 1% agarose gel.

#### 4.3.7. Gene knockout using Datsenko and Wanner<sup>121</sup> strategy

The first non-random method chosen to create the *nth* (endonuclease gene) knockout is named recombineering (homologous recombination-mediated genetic engineering). This technique provides an efficient way to generate knockout mutations directly on the bacterial chromosome and the constructs are designed to the base pair and not dependent on suitable restriction sites<sup>146</sup>.

The first step of the procedure is to design the appropriate PCR primers to make the linear insertion cassette. The knockout's (KO) primers (forward and reverse) were designed in order to incorporate stretches of homology ( $\approx 50$  nts) to the target gene and to part of an appropriate drug resistance cassette ( $\approx 20$  nts). DNA sequences of the genes from *L. lactis* LMG19460 were retrieved from the sequence of the reference genome of *L. lactis* IL1403 (NC\_002662.1 from NCBI). Afterwards, the strain LMG19460 was sequenced by our laboratory<sup>4</sup> and the sequence of the *nth* gene was confirmed as being exactly homologous to the *L. lactis* IL1403 one. The complete *nth* gene sequence is in Table 4.3. The drug resistance cassette chosen was the kanamycin gene from plasmid pKD13<sup>121</sup>. The designed primers contained, at the 5' end, the homology to the gene and at the 3' end the bases that will prime synthesis of the drug cassette. Primers used are listed in Table 4.4.

In the next step of the procedure, a PCR with the chimeric primers and the plasmid pKD13 was performed in order to make the linear recombination cassette, using Platinum PCR Supermix High Fidelity (Invitrogen), by mixing the following components: 10 ng of template pDNA, 0.2  $\mu$ M of each primer (KO\_*nth*\_F and KO\_*nth*\_R, Table 3.3), 1x Platinum<sup>TM</sup> PCR SuperMix High Fidelity and completed with PCR-grade water to a final volume of 50  $\mu$ L. The cycling conditions were: initial denaturation for 2 min at 94°C, followed by 30 cycles of 25 s at 94°C, 25 s at 55°C and 1.5 min at 68°C. Then, the PCR product was run in a 1% agarose gel with molecular weight markers to confirm its size. If the size was in agreement with the expected, the PCR product was extracted from the gel (without exposure to UV or ethidium bromide) with the QIAquick Gel Extraction Kit (Qiagen), and the DNA concentration of the fragment was determined.

In parallel, *L. lactis* LMG19460 was turned electrocompetent by growing in the presence of high NaCl concentration, by adaptation of the method described in Palomino *et al.* (2010)<sup>25</sup>. More specifically, *L. lactis* LMG19460 was grown overnight in 5 mL MRS, at 30°C, 100 rpm. This growth was used to inoculate a shake flask containing 100 mL MRS supplemented with 0.7M NaCl, starting with an initial OD of 0.1. After an overnight growth, the cells were harvested by centrifugation at 6,000g, 3 min, 4°C and washed 3 times in ice cold water. Finally, they were resuspended in a volume of 20% glycerol that allowed  $2 \times 10^{10}$  cells per 100  $\mu$ L aliquot and stored at -80°C. For transformation of the electrocompetent *L. lactis* LMG19460 cells, the 100  $\mu$ L aliquots were mixed with the appropriate volume to had 1  $\mu$ g of pKD46 plasmid and transferred to a 2 mm electroporation cuvette. After 30 min on ice, 2 pulses of 1250 V were given using an electroporator (BTX ECM399), resulting in time constants of 4 to 5 ms. To recover from the electroporation, 900  $\mu$ L of MRS was immediately added and the mixture was incubated 3h at 30°C without agitation. Then, the cells were plated in MRS agar supplemented with 150  $\mu$ g/mL ampicillin and incubated at 30°C.

The transformation of the bacteria with pKD46 was confirmed by colony-PCR, using the NovaTaq Hot Start Master Mix Kit (Novagen). Each colony PCR reaction contained 10  $\mu$ L of the supernatant of thermally lysed cells, together

with 1x NovaTaq Hot Start Master Mix, 0.1  $\mu\text{M}$  of each primer (Ampl\_pKD46\_F and Ampl\_pKD46\_R) completed with PCR-grade water to a final volume of 50  $\mu\text{L}$ . The cycling conditions were: initial denaturation for 9 min at 95°C, followed by 30 cycles of 30 s at 94°C, 30 s at 52°C and 1 min at 72°C, and a final extension step of 10 min at 72°C. The result from the colony PCR was visualized in a 1% agarose gel.

After confirmation if *L. lactis* LMG19460 was transformed with the pKD46 plasmid, the new strain was turned electrocompetent using the high salt protocol described above, with the only difference that during growth in the presence of 0.7 M of NaCl was also necessary to supplement with 0.2% L-arabinose to induce the expression of the  $\lambda$ -Red recombinase genes and with 150  $\mu\text{g}/\text{mL}$  ampicillin. The electrocompetent cells were transformed with 500 ng of the linear recombination cassette, previously incubated with 30  $\mu\text{M}$  of tetradecyltrimethylammonium bromide (TTAB), for 30 min at room temperature, to help in the DNA compactation and consequently in the transformation procedure. The electroporation was performed with 3 pulses of 1800V in 1 mm cuvettes and the bacteria then recovered with 900  $\mu\text{L}$  of MRS, during 3h, at 30°C, 100 rpm. Finally, the cells were plated in MRS agar supplemented with 1000  $\mu\text{g}/\text{mL}$  neomycin and incubated at 30°C.

To check if the  $\lambda$ -Red recombinase enabled the integration of the cassette in the genome and deleted the wild-type sequence between the regions of homology, a colony PCR was performed using the Conf\_nth\_F and Conf\_nth\_R primers. The only difference from the colony PCR described above was that it was necessary to use as template 200 ng of purified gDNA from the colonies grown overnight in liquid medium.

Ultimately, positive clones were transformed with pCP20 (replicon is resistant to 75  $\mu\text{g}/\text{mL}$  chloramphenicol and temperature sensitive) and grown at 43°C to induce flippase (FLP) recombinase that will remove the KanR cassette and cure all helper plasmids used. After that step, the selected colony was restreaked on plates with MRS supplemented with 150  $\mu\text{g}/\text{mL}$  ampicillin, 1000  $\mu\text{g}/\text{mL}$  neomycin and 75  $\mu\text{g}/\text{mL}$  chloramphenicol and grown at 30°C overnight to confirm that all helper plasmids and KanR cassette were fully cured. In the end, other colony PCR primed to the region surrounding the modification was made to ensure the entire KanR cassette removal.

The general steps of this recombineering strategy are described in Figure 4.10.

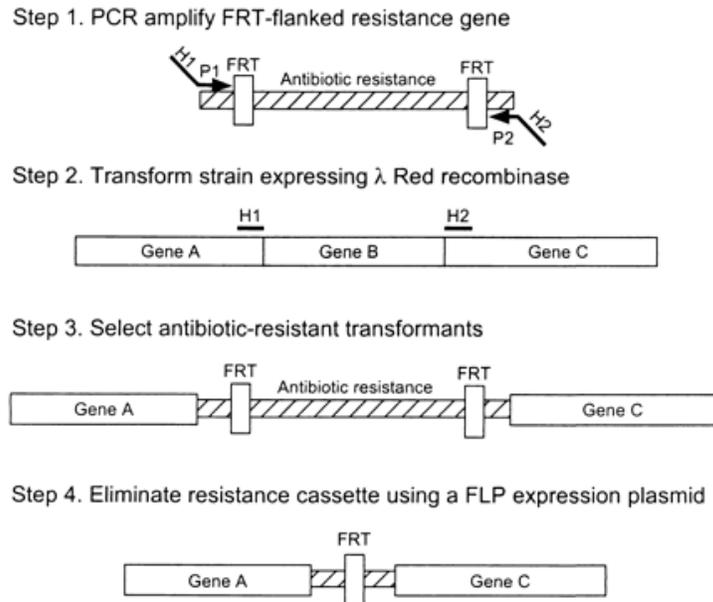
**Table 4.3.** Sequence of the *nth* gene from *L. lactis* LMG19460.

Sequence (5' → 3')	
<b><i>nth</i> gene</b>	atgtaagtaaaaaagatatcttgaagccttagaattatgaagacatgtcccgcaagctcacggagaattagatgggaaactcctttca actattgattgcgacaattttatctgcccaagcaactgacaaaggagtaataaggcgacgccggctcttttgccactttccagatgcacaaa caatgtcacaggctaaagtgggaagaaattgaaaagctgattcggacaattggctctttacaagactaaagctaaaaatatttacggacatcac agatgttagtgaccgactttggggacctttgctgattgccaaggataaaaagggtttacaactttaccgggggtgggtcgaaaaacag caaatgtgttttagcagaagcctacggaattccggggattcggttgatacacatgttgaacgatttctaacggttgatattcctcccaa aaagcaacagtttagaagttgaagaaaagttgatgaagctgattccgcaagaaaaatgggtacaagcacatcaccatctcatttcttgggc ggtatcattgtacagcaaaaaagccgaaatgtgcagactgtcctgttttagattattgtaaatggaaagaaatattaaaaaatgaatga

**Table 4.4.** Sequence, expected amplicon size and melting temperature of the primers used for *nth* gene knockout strategy.

Primer name		Sequence (5' → 3') <sup>1</sup>	Product size (bp)	T <sub>m</sub> (°C)
KO_ <i>nth</i> _F	F	AGAGAAAGAAACCACAAGAAGATTTTTATATTCCTTTGGAT GGACCATGGAATAGTTAATgtgtaggctggagctgcttc	1,414	82
KO_ <i>nth</i> _R	R	GGTCTGAGCCAATATCAGCAAGTCTTGCTCCATTATCAACAT AATTAGCTACTGCTTTCAAtcgtcgacctgcagtt		85
Conf_ <i>nth</i> _F	F	GTCCTCAATCGTAAGGTATC	1,007 (wild-type <i>nth</i> gene)	43
Conf_ <i>nth</i> _R	R	CTTTAACCACTTCTCCCGCTACC	1,604 (cassette insertion)	55
Ampl_pKD46_F	F	GCGATCTGTCTATTTTCGTTC	614	47
Ampl_pKD46_R	R	GTCTGTATGTGGCGCGGT		57

<sup>1</sup> F and R indicate forward and reverse primers, respectively.



**Figure 4.10.** Main steps of the recombineering strategy for gene knockout developed by Datsenko and Wanner (2000). Figure from Datsenko and Wanner (2000)<sup>121</sup>.

#### 4.3.8. Cloning strategy to construct the gene-specific CRISPR/Cas9 plasmid (pKCcas9dO\_*nth*\_apra)

##### 4.3.8.1. *In silico* design

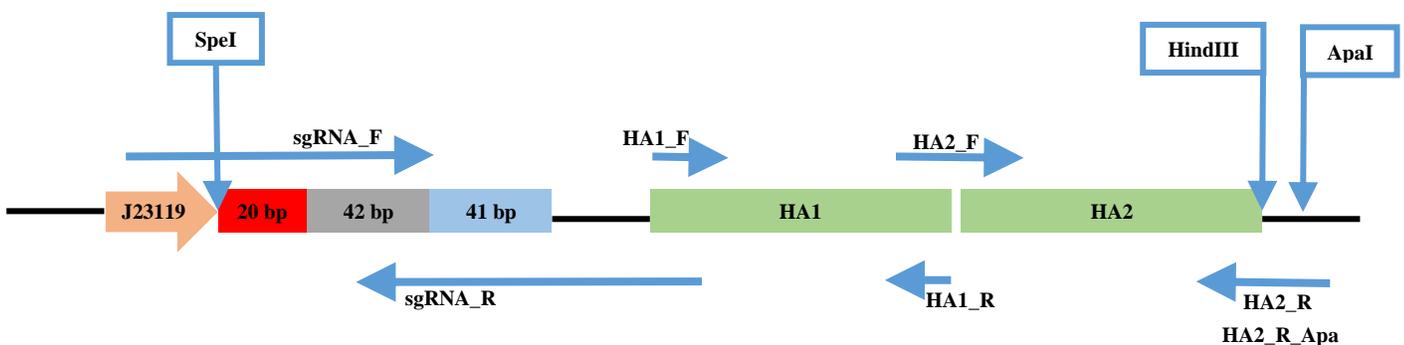
The original pKCcas9dO plasmid already had *actII-orf4* gene specific guide sequence and homologous arms, cloned between SpeI and HindIII restriction enzyme sites. For *nth* gene knockout, *nth* gene specific parts must be amplified and cloned together.

The sgRNA must contain a gene specific guide sequence with 20 bp, followed by a conserved 42 bp Cas9 handle and a 41 bp terminator. Immediately upstream the sgRNA sequence is the SpeI restriction site, meaning that the necessary J23119 promoter that will control the sgRNA expression will already be present in the vector. Downstream the terminator, must be cloned two homologous arms, corresponding to the genome flanking regions of the *nth* gene, with a HindIII restriction site in the end, to allow cloning in the remaining pKCcas9dO vector (Figure 4.8).

The guide sequence was chosen considering the mandatory rule that should be a 20 bp sequence immediately upstream a PAM sequence in the genome. The PAM sequence is 5'-NGG-3' and there are softwares available online<sup>147</sup>, that automatically finds possible guide sequences and give them a score based on its %GC and melting temperature of the sgRNA:DNA duplex. The guide sequence with the highest score found by the software<sup>147</sup> was chosen.

The *nth* gene specific sgRNA was designed to be amplified with two semi-overlapping primers, containing together the whole necessary sequence. The forward primer also contained a region homologous to a segment of the J23119 promoter, while the reverse primer had a region of homology with the beginning of the first homologous arm. Each homologous arm were designed to be amplified by a set of primers that should had overlapping regions with the upstream and downstream segments. The primers to amplify the first homologous arms does not needed to have this overlapping regions, because the sgRNA reverse primer and the second homologous arm forward primer already harbored the necessary homologous segments. The forward primer to amplify the second homologous arm should had an homologous region to the end of the first arm, while the reverse primer should had an homologous region with the beginning of the pKCcas9dO vector.

The complete *in silico* design is represented in Figure 4.11, while the sequences of each region are in Table 4.5. The three resulting fragments (sgRNA and the two homologous arms) from the PCR amplifications with the primers represented in Table 4.6 will be joined together by SOEing (Splicing by Overlap Extension) PCR. Due to the fact that the sequence of one of the homologous arms had a HindIII restriction site, a final amplification of the complete insert was performed with sgRNA\_F and HA2\_R\_Apa primers to insert an ApaI restriction site. In the end, the insert will be cloned between the SpeI and ApaI restriction sites of the pKCcas9dO vector.



**Figure 4.11.** *In silico* design of the amplification, SOEing PCR and cloning strategies to construct the *nth* gene specific CRISPR/Cas9 plasmid (pKCcas9dO\_nth\_apra).

**Table 4.5.** Sequences of the *nth* specific sgRNA and the two homologous arms.

Sequence (5' → 3')	
<b><i>nth</i> sgRNA</b>	gcagaagcctacggaattccggttttagagctagaataagcaagttaaaataaggctagccgttatcaactgaaaaagtgaccaggctgctgttttt
<b><i>nth</i> homologous arm 1</b>	gaaccgttagagtagtaagtcattaaagtgaagtgataatttctaataatctttgattgatttttctactttttgagtaatttttctgtttgctttcattgtttatctt ttcaaatgtccctttgtttataagtagatataatcaaggttgcctgctaatttctacaaaaatgtaattgaagaaaaaatgctttaggaaataaggtctgaaatta aatagtataaaaataaattgattaaaaatttctaaagttaactgatttaagtagtttatgtgaagttcattgaaccataatgacgcgatagctataatgctctatgaat tattacgatgaatatagcaaggcctctgtcttaccaaatgtgattitagcacattttagtaagcttttccctctgctttgatttttgattgggttatactttttgaaa atcaagatattgcacctagtgaattgctgaaaaaacaggtaaaaaattatcggaggtcaatctggccattgataatctgcaaaattgggcaaatgaaggtga ctttaattgaaattgatggagacatggaacttctttgatattgcccggcctttaagcacttagatgaaatttgggtgctgttggcagttcaactccagaga atgaaaaagttcaggaagggcagcttaagaalttagtagcatgtttgaagcagaatggcctgatttctctgtacaaatggaagaattgagagcttgg ctgtttgaagatggtatgacttttctgttattaaacaagcttgcgtgaagcagtcctcaatcgaaggtatctttaaataacataaaagctattttacgaaattggaa aaatgatggggtgaattcttcaagcattgagagtcgacaatggaacggaagaaagtaagaagaagaaccacaagaagattttatattcttggatg gacatggaatgtaataagataatggacgtgcacg
<b><i>nth</i> homologous arm 2</b>	aattaaacttcaaacgtttgaaagcagtagctaattatggtgataatggagcaagactgctgatattggctcagaccatgcttattaccaacttattgatgcaa aaattgggtgattgatttgcgtagcgggagaagtggtttaaaggcctttgaaattgctaaaaatcatgatacagaagctgattaaagtatcggattgaggttcg attggccaatggcttaggtctattgaaatactgacaagattgatacaactgctgattgctggtatgggaggaatattaattccgaaattttagaggcaggtaaa gaaaattgtagccatgtaaacgtttgattttgcagcctaataatcatgaggaaagcctcctcagtggttagtgaatcacaattgctcattaaaaaagagat ttattggaggtgctgtaagttttacgagattattgtccgaaccactctcaaaatgacagaaaaagttgctcagtaaatgatttaacattggccctttttactaa agaaaaatcaactgtttttcaacaaaaatgcaaaaaggaattaaactttgaaataaattattgctcgtttgcctgaagaacaagccgaaaaagcgaagaggtt ttaaatacaattgcgagaattgaggaaagtgctcgtgataaaattctgattttatgcttgaatacgaataaaatttgccecaagaattagctgttgaaggtgatcctgt aggtttacagttgggaaccaaacgatgaactaacaagaattttgtaaccttggatattcgtgagcaaacagtgccagaagcaaaaagcgtgggtgtaat ctgattattgctaaacatccttattttccgccccttctgcattgacttcaatgaatgatcaggaataatgattttgacttggcgcgagctg

**Table 4.6.** Sequence, expected amplicon size and melting temperature of the primers used for *nth* CRISPR/Cas9 gene knockout strategy.

Primer name		Sequence (5' → 3') <sup>1</sup>	Product size (bp)	T <sub>m</sub> (°C)
<b>sgRNA_F</b>	F	atatatactagtGCAGAAGCCTACGGAATTCGGTTTTAGAGCTAG AAATAGCAAGTTAAAATAAGGCTAGTCCG	136	78
<b>sgRNA_R</b>	R	gacttactctacaacgggtcAAAAAAGCACCAGCTCGGTGCCACTTTT TCAAGTTGATAACGGACTAGCCTTATTTTAAC		83
<b>HA1_F</b>	F	gaaccgttagagtagtaagtc	1,028	40
<b>HA1_R</b>	R	ctgtcacgtccattatctct		50
<b>HA2_F</b>	F	agagataatggacgtgcacgaattaacttcaaaaagctt	979	70
<b>HA2_R</b>	R	atatataagcttcagctcgcgccaagtcctcaaa		67
<b>HA2_R_Apa</b>	R	atatatggcccatatataagcttcagctcgcgccaagtcctcaaa	991	67
<b><i>nth</i>_conf_F</b>	F	caaggcgaagtcataataac	2,790 (wild- type <i>nth</i> gene)	49
<b><i>nth</i>_conf_R</b>	R	gtatggctggtatagacagca	2133 (knockout)	54
<b>Cas9_conf_F</b>	F	cccaggtcaacatcgtcaag	241	57
<b>Cas9_conf_R</b>	R	tccatgatggtgatgccag		58

<sup>1</sup> F and R indicate forward and reverse primers, respectively.

#### 4.3.8.2. Fragments amplification, SOEing PCRs and molecular cloning of the gene-specific sgRNA and homologous arms in pKCas9dO plasmid

PCR amplification of both homologous arms was performed using the KOD Hot Start DNA Polymerase kit (Novagen), by mixing 200 ng of template gDNA from *L. lactis* LMG19460, 0.02U/μL of KOD DNA Polymerase, 1 mM of MgSO<sub>4</sub>, 1x buffer, 0.2 mM of dNTPs, 0.3 μM of each primer and completed with PCR-grade water to a final volume of 25 μL. The reaction was the same for the sgRNA amplification, but without template DNA, only for primer annealing. The cycling conditions for the 3 reactions were: initial denaturation for 2 min at 95°C, followed by 35 cycles of 1 min at 95°C, 1 min at 65°C and 1 min at 70°C. After amplification, the PCR products were run in a 1% agarose gel in order to retrieve the specific band without UV or ethidium bromide exposure, to keep the DNA quality and integrity. The

agarose gel fragment was purified using NZYGelPure (Nzytech) and its concentration quantified using a Nanodrop Spectrophotometer (Nanovue Plus, GE).

The first SOEing PCR was performed between the two homologous arms by mixing 100 ng of each purified fragment, 0.02U/ $\mu$ L of KOD DNA Polymerase, 1 mM of  $MgSO_4$ , 1x buffer, 0.2 mM of dNTPs, 0.3  $\mu$ M of each primer (HA1\_F and HA2\_R) and completed with PCR-grade water to a final volume of 25  $\mu$ L. The cycling conditions were: initial denaturation for 2 min at 95°C, followed by 35 cycles of 1 min at 95°C, 1 min at 68°C and 2 min at 70°C. The PCR product was again run in a 1% agarose gel, and the specific band was excised and the DNA purified and quantified.

Using 100 ng of the DNA purified from the first SOEing PCR as one template and 100 ng of the amplified sgRNA as the second template, a second SOEing PCR was performed using the usual conditions for the KOD polymerase, with the primers sgRNA\_F and HA2\_R. The cycling conditions were: initial denaturation for 2 min at 95°C, followed by 35 cycles of 1 min at 95°C, 1 min at 62°C and 2.5 min at 70°C. The PCR product was run again in a 1% agarose gel, and the DNA purified from the gel band.

A final PCR was performed using the result of the second SOEing PCR as template and sgRNA\_F and HA2\_R\_Apa as primers, in order to introduce the ApaI restriction site. The PCR conditions mix was the same as the previous one, while the cycling conditions were: initial denaturation for 2 min at 95°C, followed by 35 cycles of 1 min at 95°C, 1 min at 68°C and 2.5 min at 70°C. The following procedure was the usual, with the PCR product being visualized in a 1% agarose gel and the correct band being excised and the DNA purified.

The next step was to proceed to the molecular cloning strategy, starting by digesting both vector (pKCcas9dO) and insert (fragment with sgRNA and both homologous arms) with SpeI and ApaI (Promega), for 3h at 37°C. The correct bands were cut from a 1% agarose gel, the DNA purified using NZYGelPure (Nzytech) and its concentration quantified using a Nanodrop Spectrophotometer (Nanovue Plus, GE). After the gel purification step, the samples were concentrated using a DNA SpeedVac Concentrator. The ligation between the vector and insert was prepared with 2  $\mu$ L of T4 DNA ligase 3 u/ $\mu$ L (Promega) and T4 ligase buffer (300 mM Tris-HCl (pH 7.8), 100 mM  $MgCl_2$ , 100 mM DTT, 10 mM ATP) diluted 10 times in a final volume of 20  $\mu$ L. Therefore, for the reaction mixture, a certain quantity of vector (100 ng) and insert were mixed maintaining a specific mass of vector, considering a 1:1 or 3:1 insert/vector molar ratio, according to the equation:

$$ng\ of\ insert = \frac{ng\ of\ vector \times kb\ size\ of\ insert}{kb\ size\ of\ vector} \times molar\ ratio\ of\ \frac{insert}{vector}$$

The ligation mixtures were incubated for 3 hours at room temperature and overnight at 4°C. After the first 3 hours of incubation, chemically competent *E.coli* DH5 $\alpha$  cells were transformed with 10  $\mu$ L of the total mix volume, and the remaining 10  $\mu$ L of ligation mixture were incubated at 4°C overnight, and used to transform another aliquot of *E.coli* DH5 $\alpha$  cells by heat shock.

Several colonies were picked from the LB plates supplemented with 25 µg/ml apramycin and grown in liquid medium at 30°C, in order to purify pDNA using High Pure Plasmid Isolation Kit (Roche) and test if the cloning was successful, using the appropriate restriction enzymes.

The final pKCcas9dO\_nth\_apra plasmid with the *nth* gene specific sgRNA and homologous arms was also used to transform two other *E. coli* strains: GM2163 or SCS110. These strains have the advantage of being both Dam<sup>-</sup> and Dcm<sup>-</sup>, meaning that the pDNA will not be methylated, which could be advantageous in LAB transformation.

#### 4.3.9. Gene knockout using CRISPR/Cas9 strategy

*L. lactis* LMG19460 was transformed with the final pKCcas9dO\_nth\_apra plasmid with the *nth* gene specific sgRNA and homologous arms, purified from the 3 different *E. coli* strains, using the strategy described above. The resulting colonies were grown in liquid medium (M17 supplemented with 20 g/L glucose and 500 µg/ml apramycin), at 30°C, 100 rpm, and its pDNA purified using NucleoSpin Plasmid (Macherey-Nagel) and the gDNA purified using the Wizard Genomic DNA Purification Kit (Promega).

The pDNA was visualized in a 1% agarose gel to confirm the strain transformation with the correct plasmid, and also used to perform a PCR to detect its presence inside *L. lactis* LMG19460 cells. The PCR was performed using the KOD Hot Start DNA Polymerase kit (Novagen) and mixing 100 ng of purified pDNA or 250,000 cells, 0.02U/µL of KOD DNA Polymerase, 1 mM of MgSO<sub>4</sub>, 1x buffer, 0.2 mM of dNTPs, 0.3 µM of each primer (HA1\_F/R or Cas9\_conf\_F/R complementary to the *ScoCas9* sequence) and completed with PCR-grade water to a final volume of 25 µL. The cycling conditions were: initial denaturation for 2 min at 95°C, followed by 35 cycles of 30 s at 95°C, 30 s at 65°C (HA1\_F/R) or 56°C (Cas9\_conf\_F/R) and 3 min at 70°C. If the strain was successfully transformed, the expected band should have 1028 bp.

The gDNA was used to perform a PCR using the KOD Hot Start DNA Polymerase kit (Novagen), by mixing 200 ng of template gDNA or 250,000 cells from *L. lactis* LMG19460 transformed with the CRISPR/Cas9 *nth* specific plasmid, 0.02U/µL of KOD DNA Polymerase, 1 mM of MgSO<sub>4</sub>, 1x buffer, 0.2 mM of dNTPs, 0.3 µM of each primer (nth\_conf\_F and nth\_conf\_R, complementary to regions in the *L. lactis* LMG19460 genome immediately upstream and downstream the homologous arms) and completed with PCR-grade water to a final volume of 25 µL. The cycling conditions were: initial denaturation for 2 min at 95°C, followed by 35 cycles of 30 s at 95°C, 30 s at 53°C and 3 min at 70°C. If the genome had the wild-type *nth* gene, the band visualized in a 1% agarose gel should had 2790 bp, while the band corresponding to the *nth* gene knockout should had 2133 bp.

After knockout confirmation, the correct colony should be plated in regeneration medium<sup>112</sup> without antibiotic and grown at 37°C, to cure the plasmid.

#### 4.3.10. Induction of the ScoCas9 expression using thiostrepton

Since it was not possible to obtain the correct knockout by simply transforming the *L. lactis* LMG19460 strain with the pKCCas9dO\_nth\_apra, other alternative approaches were pursued. A possible problem could be the promoter tipA that controls the expression of ScoCas9 and is inducible by thiostrepton and thiostrepton-like antibiotics. In *Streptomyces lividans* the expression of the genes cloned under the control of this type of promoter is induced at least 200-fold after the addition of thiostrepton to the growth media<sup>148</sup>. There is always a basal expression from this promoter in the absence of the antibiotic, but it is highly inducible in its presence<sup>148,149</sup>. In the hope that this basal expression was enough to express the ScoCas9 in our *L. lactis* LMG19460 platform, we did not use thiostrepton in the growth media, but it could be a reason why the system is not working properly. In this section the thiostrepton was added to the growth media as an attempt to induce the maximum *ScoCas9* expression.

First, since it is not possible to foresee if, on one side, the *L. lactis* LMG19460 strain is naturally resistant to this antibiotic or if, on the other side, the presence of thiostrepton will impair the strain correct growth, different thiostrepton concentrations were used to grow the strain at 30°C, 100 rpm: 0, 0.001, 0.01, 0.08, 0.15, 0.3, 0.625, 1.25, 2.5, 5, 10, 25 and 50 µg/mL.

The thiostrepton was reconstituted in two different solvents to test which was less aggressive for the cell growth: DMSO and glacial acetic acid.

To increase the recombination probability, the two homologous arms sequence were amplified as a linear fragment to be used to transform the *L. lactis* LMG19460 strain together with the pKCCas9dO\_nth\_apra plasmid. PCR amplification of both homologous arms was performed using the the KOD Hot Start DNA Polymerase kit (Novagen), by mixing 10 ng of pKCCas9dO\_nth\_apra plasmid as template, 0.02U/µL of KOD DNA Polymerase, 1 mM of MgSO<sub>4</sub>, 1x buffer, 0.2 mM of dNTPs, 0.3 µM of each primer (HA1\_F and HA2\_R) and completed with PCR-grade water to a final volume of 25 µL. The cycling conditions were: initial denaturation for 2 min at 95°C, followed by 35 cycles of 1 min at 95°C, 1 min at 65°C and 1 min at 70°C. After amplification, the PCR product was digested with 10U DpnI (Promega) for 2 hours at 37°C and run in a 1% agarose gel in order to retrieve the specific band without UV or ethidium bromide exposure, to keep the DNA quality and integrity. The agarose gel fragment was purified using NZYGelPure (Nzytech) and its concentration quantified using a Nanodrop Spectrophotometer (Nanovue Plus, GE).

The induction protocol consisted in transforming the *L. lactis* LMG19460 strain with a) 1 µg of the pKCCas9dO\_nth\_apra plasmid or b) 1 µg of the pKCCas9dO\_nth\_apra plasmid together with 100 ng of the HA1/HA2 linear fragment, using the strategy described above. After the 3 hours recuperation at 30°C with the recovery medium, the cells were used to inoculate 5 mL of M-17 supplemented with 20 g/L glucose, a sub-lethal apramycin concentration<sup>113</sup> (250 µg/mL) and different thiostrepton concentrations, which were solubilized with DMSO or acetic acid, at 30°C without agitation. Two negative controls were performed, without thiostrepton and without thiostrepton and apramycin.

After an overnight growth, the cells were again collected by centrifugation and plated in regeneration medium<sup>116</sup> supplemented with 500 mg apramycin, and grown at 30°C. Several resulting colonies were picked from the plates in

order to perform a colony PCR to confirm if the *L. lactis* LMG19460 strain was successfully transformed with the plasmid pKCCas9dO\_nth\_apra, using Cas9\_conf\_F/R primers that amplified a portion of the *ScoCas9* gene. To confirm if the strain had the correct knockout, the same strategy was used, but with the nth\_conf\_F/R primers, which are complementary to regions in the *L. lactis* LMG19460 genome immediately upstream and downstream the homologous arms. The colony PCR was performed using the NovaTaq Hot Start Master Mix Kit (Novagen) and each reaction contained 10 µL of cells, together with 1x NovaTaq Hot Start Master Mix, 0.1 µM of each primer (Cas9\_conf\_F/R) completed with PCR-grade water to a final volume of 25 µL. The cycling conditions were: initial denaturation for 2 min at 95°C, followed by 40 cycles of 30 s at 95°C, 30 s at 56°C and 3 min at 70°C. The result from the colony PCR was visualized in a 1% agarose gel.

If the genome had the wild-type *nth* gene, the band visualized in a 1% agarose gel should had 2790 bp, while the band corresponding to the *nth* gene knockout should had 2133 bp. Whenever the pKCCas9dO\_nth\_apra plasmid was present inside the *L. lactis* LMG19460 cells, a band with 241 bp appeared in the agarose gel.

The positive colonies were grown in liquid medium (M17 supplemented with 20 g/L glucose and 500 µg/ml apramycin), at 30°C, 100 rpm, and its pDNA purified using NucleoSpin Plasmid (Macherey-Nagel). After confirmation of the presence of the plasmid by agarose gel electrophoresis, the correct clone should be plated in regeneration medium<sup>112</sup> without antibiotic and grown at 37°C, to cure the plasmid.

#### **4.3.11. Replacement of the apramycin resistance gene for the erythromycin resistance gene**

Since the pKCCas9dO\_nth\_apra revealed some issues in the steps of *L. lactis* LMG19460 transformation and selection of the correct clones, the next approach consisted in replacing the apramycin resistance gene by the erythromycin resistance gene. The procedure was performed using PCR amplification of each desired fragment, followed by Gibson Assembly to join them as plasmid pKCCas9dO\_nth\_ery. Two separate fragments were amplified from pKCCas9dO\_nth\_apra using Apa\_Ery\_F/R and Ery\_Apa\_F/R primers, instead of a single fragment, in order to reduce the size of the plasmid by discarding the final sequence of the parental homologous arms that remained in the pKCCas9dO\_nth\_apra and at the same time to allow the replacement of the antibiotic resistance gene from apramycin to erythromycin. The primers Apa\_Ery\_F/R amplified the region from the ApaI restriction site, in the end of the homologous arm 2, until the beginning of the antibiotic resistance gene, while the primers Ery\_Apa\_F/R amplified starting on the end of the apramycin resistance gene, until the end of the truncated parental homologous arm (Figure 4.12A). The erythromycin resistance gene, including its promoter region, was amplified from the pTRKH3 plasmid using the Ery\_F/R primers (Figure 4.12B).

The sets of primers used for the PCR amplifications are described in Table 4.7 and were designed following the Gibson Assembly requirements, namely the need for overlapping sequences between adjacent fragments. The online software used to design the necessary primers was the NEB Builder Assembly tool v1.12.16<sup>116</sup>.

The fragments corresponding to the erythromycin resistance gene, from the pTRKH3 template, and the one starting in the ApaI restriction site and finishing upstream the antibiotic resistance gene, from the pKCCas9dO\_nth\_apra template,

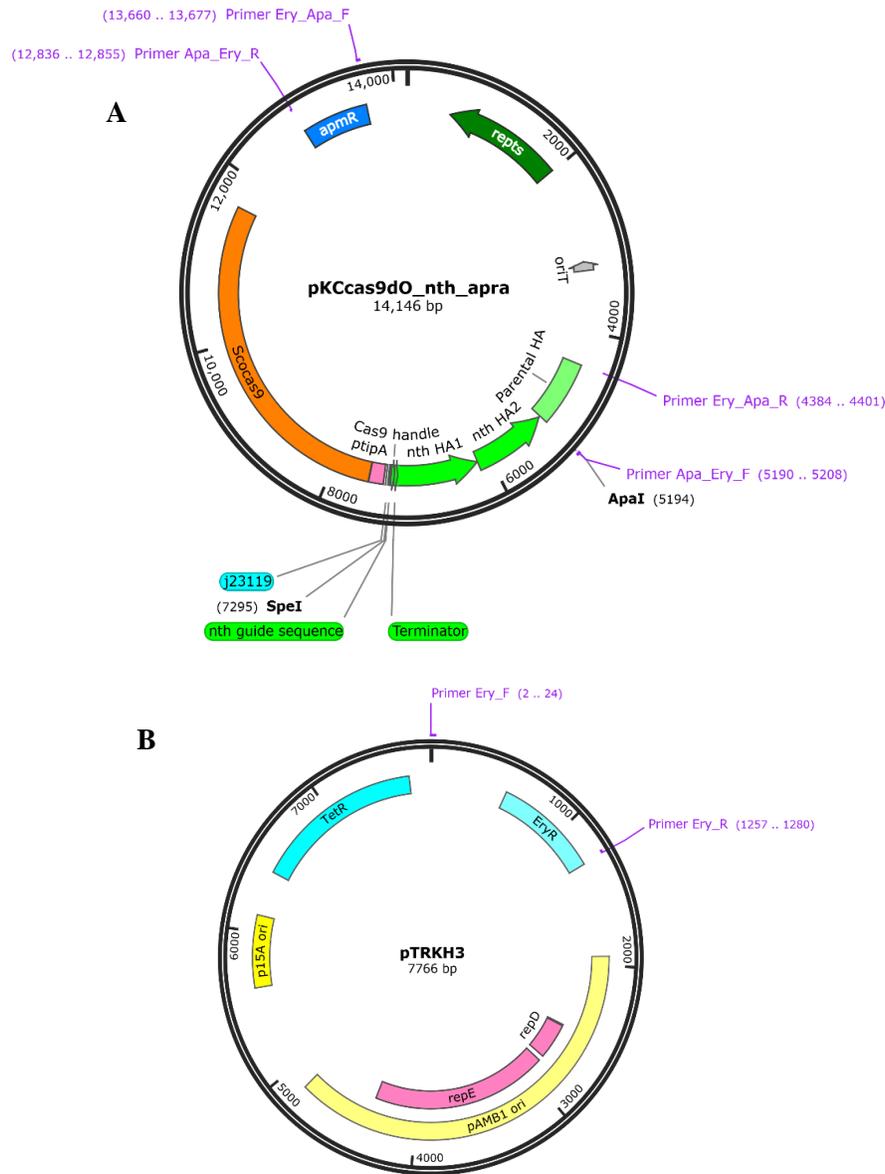
were both amplified using the KOD Hot Start DNA Polymerase kit (Novagen), by mixing 0.5 ng of template pDNA, 0.02U/ $\mu$ L of KOD DNA Polymerase, 2.3 mM of MgSO<sub>4</sub>, 1x buffer, 0.2 mM of dNTPs, 0.3  $\mu$ M of each primer (Ery\_F/R or Apa\_Ery\_F/R) and completed with PCR-grade water to a final volume of 25  $\mu$ L. The cycling conditions were: initial denaturation for 2 min at 95°C, followed by 40 cycles of 1 min at 95°C, 1 min at 59°C for Ery\_F/R or 62°C for Apa\_Ery\_F/R and 2 or 8 min at 70°C, for Ery\_F/R or Apa\_Ery\_F/R, respectively. The PCR products were visualized in a 1% agarose gel (1299 bp band for the erythromycin resistance gene fragment and 7686 bp band for the Apa\_Ery fragment).

The fragment starting downstream the antibiotic resistance gene and ending downstream the truncated parental homologous arm was amplified from pKCcas9dO\_nth\_apra, using KOD Hot Start DNA Polymerase kit (Novagen). The reaction contained 0.5-10 ng of pDNA, together with 0.02U/ $\mu$ L of KOD DNA Polymerase, 1 mM of MgSO<sub>4</sub>, 1x buffer, 0.2 mM of dNTPs, 0.3  $\mu$ M of each primer (Ery\_Apa\_F/R) and completed with PCR-grade water to a final volume of 25  $\mu$ L. The cycling conditions were: initial denaturation for 2 min at 95°C, followed by 40 cycles of 1 min at 95°C, 1 min at 54°C and 5 min at 70°C. The 4,908 bp band was visualized in a 1% agarose gel.

**Table 4.7.** Sequence, expected amplicon size and melting temperature of the primers used for the replacement of the apramycin resistance gene for the erythromycin resistance gene in pKCcas9dO\_nth\_apra plasmid, resulting in the pKCcas9dO\_nth\_ery plasmid. The lowercase nucleotides correspond to the overlapping sequences with the adjacent fragment, necessary for the Gibson Assembly protocol.

Primer name		Sequence (5' $\rightarrow$ 3') <sup>1</sup>	Product size (bp)	T <sub>m</sub> (°C)
Ery_F	F	tgatcgactgAGTCTAGAATCGATACGATTTTG	1,299	57.9
Ery_R	R	gctcatgagcTTATTTCTCCCGITAAATAATAG		57.9
Apa_Ery_F	F	tgccaagcttGGGCCCATATATAAGCTTC	7,686	60.3
Apa_Ery_R	R	attctagactCAGTCGATCATAGCACGATC		60.3
Ery_Apa_F	F	gaggaaataaGCTCATGAGCGGAGAACGAGATGACGTT	4,908	65.1
Ery_Apa_R	R	atatggcccAAGCTTGGCACTGGCCGTCGTTTTACAA		69.6
Ery_conf_F	F	CCATGCGTCTGACATCTATCT	190	55.2
Ery_conf_R	R	CTGTGGTATGGCGGTAAGT		55.2

<sup>1</sup> F and R indicate forward and reverse primers, respectively.



**Figure 4.12.** Schematic representation of the A) pKCcas9dO\_nth\_apra and B) pTRKH3 plasmids with the necessary primers for the fragments amplification.

All the PCR products were digested with 10U DpnI (Promega) for 2 hours at 37°C, to remove the parental plasmids and isolate the fragments of interest. DpnI inactivation was accomplished by submitting the samples to 80°C for 20 minutes. The PCR products were purified from the gel using the NZYGelPure Kit (Nzytech) and its concentration quantified using a Nanodrop Spectrophotometer (Nanovue Plus, GE).

If the three fragments were correctly obtained (Ery, Apa\_Ery and Ery\_Apa) the goal was to assemble them together using the Gibson Assembly technique, employing for that purpose the NEBuilder HiFi DNA Assembly Cloning Kit. The colonies resulting from the Gibson Assembly should be picked from the LB plates supplemented with 500 µg/ml

erythromycin in order to perform a colony PCR to confirm if the *E. coli* strain was successfully transformed with the new plasmid pKCcas9dO\_nth\_ery, using Ery\_conf\_F/R primers that amplified a portion of the erythromycin resistance gene. To further confirm the presence of the plasmid, amplifications with Cas9\_conf\_F/R primers could also be performed, using the conditions already described above.

The putative result from a successful assembly will be used to transform *L. lactis* LMG19460, which resulting colonies must be tested for the knockout presence using nth\_conf\_F/R primers.

#### **4.3.12. Insertion of the *ScoCas9* gene, sgRNA and homologous arms into the pTRKH3 vector**

Since the Ery\_Apa fragment was never obtained even after several optimization attempts of the PCR protocol, a new alternative solution was pursued. Also, other possible reason for the recombination not being working after *L. lactis* LMG19460 transformation with the pKCcas9dO\_nth\_apra plasmid could be related with the pSG5 origin of replication present in the backbone pKCcas9dO plasmid that, to the best of our knowledge, was never tested in LAB. If the strain is transformed with the intended plasmid, but does not allow its replication, the amounts of *ScoCas9* and sgRNA being produced would decrease dramatically, and consequently decreasing the probability of a knockout event to occur.

The present alternative encompass cloning the necessary components for the CRISPR/Cas9 knockout to occur (the *ScoCas9* gene under the control of the *tipA* promoter, the sgRNA under the control of the J23119 promoter and the two homologous arm sequences) in the pTRKH3 template, which harbours the well studied pAM $\beta$ 1 origin of replication able to replicate in LAB. The primers used to amplify the pTRKH3 vector were designed in order to exclude the tetracycline resistance gene sequence, which is not essential for the present purpose and allowed a plasmid with a smaller size.

The primers used to amplify both the pTRKH3 vector and the insert with all the necessary CRISPR/Cas9 components are described in the Table 4.8. The primers were designed in order to have overlapping sequences to allow perform Gibson Assembly between both fragments.

**Table 4.8.** Sequence, expected amplicon size and melting temperature of the primers used for the cloning of the necessary CRISPR/Cas9 features into the pTRKH3 template. The lowercase nucleotides correspond to the overlapping sequences with the adjacent fragment, necessary for the Gibson Assembly protocol.

Primer name		Sequence (5' → 3') <sup>1</sup>	Product size (bp)	T <sub>m</sub> (°C)
pT vector_F	F	ATGGAAGCCGGCGGCACC	6,575	72
pT vector_R	R	ACACGGTGCCTGACTGCG		72
CRISPR insert_F	F	aacgcagtcaggcaccgtgtCAGCTCGCGCCAAGTCCAAAAC	6,480	72
CRISPR insert_R	R	gaggtgccgccggcttccatTCAGTCGCCGCCAGCTG		72

<sup>1</sup> F and R indicate forward and reverse primers, respectively.

The pTRKH3 vector fragment was amplified from the original pTRKH3 purified from *E. coli* DH5 $\alpha$ , using KOD Hot Start DNA Polymerase kit (Novagen). The reaction contained 10 ng of pDNA, together with 0.02U/ $\mu$ L of KOD DNA Polymerase, 2.3 mM of MgSO<sub>4</sub>, 1x buffer, 0.2 mM of dNTPs, 0.3  $\mu$ M of each primer (pT vector\_F/R) and completed with PCR-grade water to a final volume of 25  $\mu$ L. The cycling conditions were: initial denaturation for 2 min at 95°C, followed by 40 cycles of 30 s at 95°C, 30 s at 60°C and 7 min at 70°C.

The insert containing the *ScoCas9* gene under the control of the *tipA* promoter, the sgRNA under the control of the J23119 promoter and the two homologous arm sequences was amplified from the original pKCcas9dO\_nth\_apra plasmid purified from *E. coli* DH5 $\alpha$ , using KOD Hot Start DNA Polymerase kit (Novagen). The reaction contained 100 ng of pDNA, together with 0.02U/ $\mu$ L of KOD DNA Polymerase, 2.3 mM of MgSO<sub>4</sub>, 1x buffer, 0.2 mM of dNTPs, 0.3  $\mu$ M of each primer (CRISPR insert\_F/R) and completed with PCR-grade water to a final volume of 25  $\mu$ L. The cycling conditions were: initial denaturation for 3 min at 95°C, followed by 40 cycles of 1 min at 95°C, 1 min at 72°C and 7 min at 70°C.

The 6,575 bp and 6,480 bp bands from the vector and the insert, respectively, were visualized in an 1% agarose gel and the PCR products were purified from the gel using the NZYGelPure Kit (Nzytech) and its concentration quantified using a Nanodrop Spectrophotometer (Nanovue Plus, GE).

After the two fragments being correctly obtained the goal was to assemble them together using the Gibson Assembly technique, employing for that purpose the NEBuilder HiFi DNA Assembly Cloning Kit. The resulting colonies in *E. coli* should be tested for the presence of the correctly assembled plasmid by colony PCR. The putative result from a successful assembly will be used to transform *L. lactis* LMG19460, which resulting colonies must be tested for the knockout presence using nth\_conf\_F/R primers.

## 4.4. Results and discussion

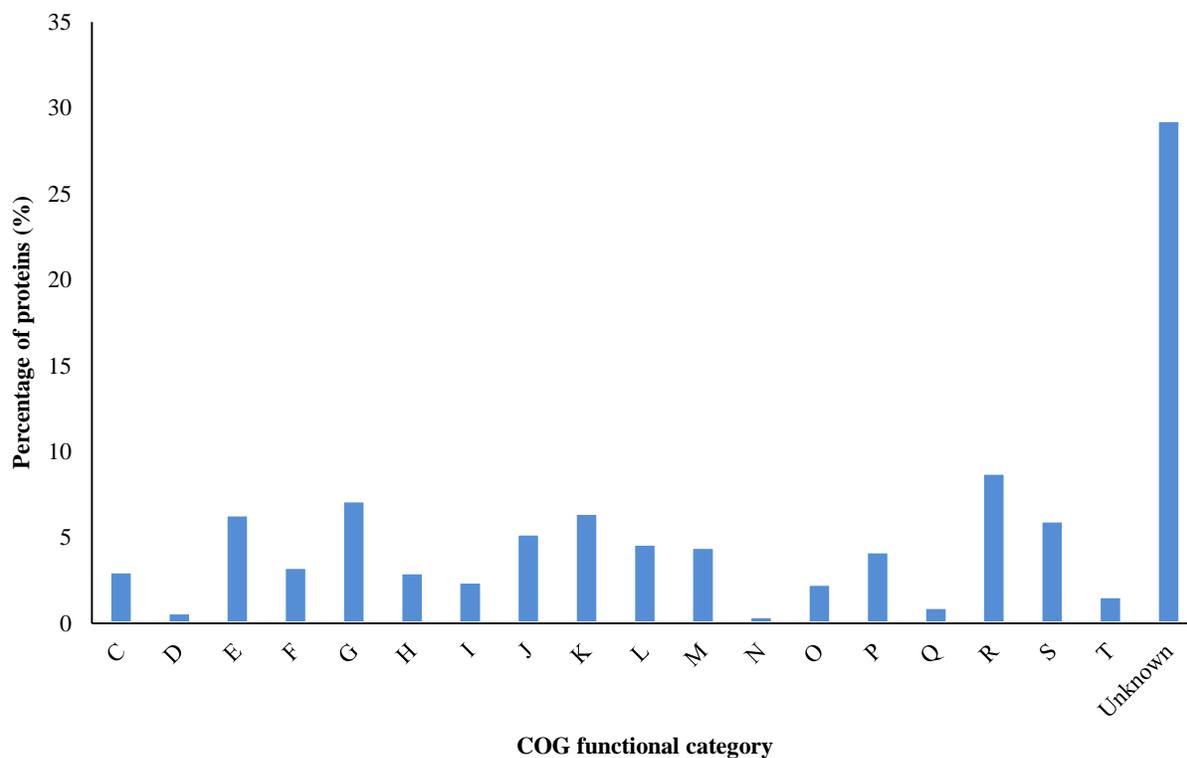
### 4.4.1. *L. lactis* LMG19460 genome characterization

The *L. lactis* subsp. *lactis* LMG 19460 genome was sequenced by our laboratory and Instituto Gulbenkian de Ciência (Portugal) using the Illumina MiSeq platform<sup>4</sup>. This genome have 2,260,841 bp and an estimated G+C content of 35.1%. The NCBI Prokaryotic Genome Annotation Pipeline predicted a total of 2,310 genes, from which 2,171 are protein-coding sequences, 78 pseudogenes, six rRNAs, and 51 tRNAs<sup>4,150</sup>. This strain has a slightly smaller genome when compared with the average value of 18 genomes of *L. lactis* subsp. *lactis* strains (2,451,000 bp)<sup>151</sup>. Comparing the protein-coding sequences, the present strain also had less than the *L. lactis* subsp. *lactis* core genome (2,364 genes). The eighteen *L. lactis* subsp. *lactis* genomes had more genes coding for tRNAs and rRNAs, but less pseudogenes, than the *L. lactis* LMG19460 strain, with an average of 60 tRNAs, 18 rRNAs and 31 pseudogenes. The G+C content of the present strain was similar to the one from the core genome. The eighteen genomes of *L. lactis* subsp. *lactis* had an average of 2.3 cryptic plasmids, while the *L. lactis* LMG19460 is a plasmid-free strain<sup>151</sup>.

A common system to organize gene information from newly sequenced prokaryotic genomes is to categorize the genes into Clusters of Orthologous Groups (COGs) of proteins, which confer a functional annotation platform. In the Evolutionary Biology field, the genes could have two different types of homologous relationships: paralogy and orthology, and the latter is the one that matters in this context. When the homologous genes derive from a single ancestral gene in the last common ancestor are considered orthologous, and typically occupy the same functional niche in different species. On the opposite, paralogs are homologous genes which have evolved by duplication of an ancestral gene and usually evolve new functions<sup>152</sup>.

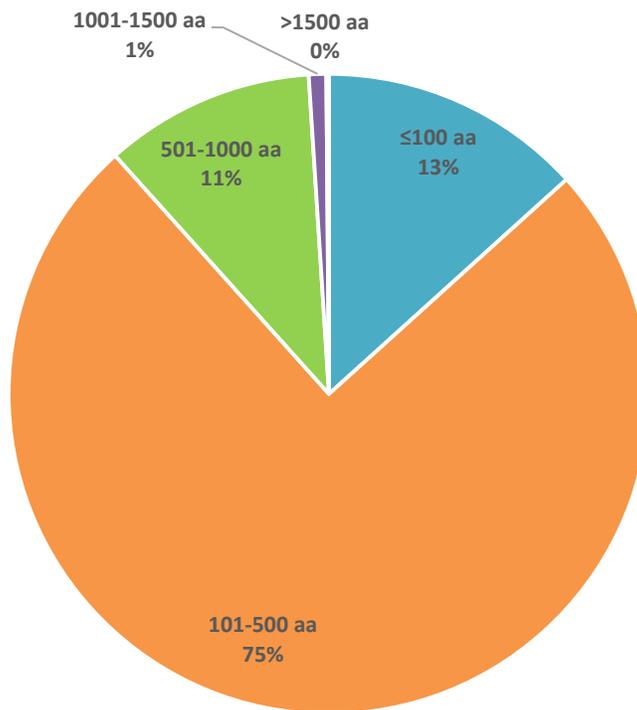
Looking at the distribution of the *L. lactis* LMG19460 proteins for the several COG functional categories (Figure 4.13), it is clear that the majority of the proteins fall into unknown function categories (S and unknown), followed by the general biochemical activity (R) category. Excluding these ones, the category with the highest percentage of proteins concerns the carbohydrate metabolism and transport, with 7.2% of the total proteins falling in this COG. The transcription (6.4%) followed closely by amino acid metabolism and transport (6.3%) COGs are the next categories with the highest number of proteins. When comparing with the *L. lactis* subsp. *lactis* core genome, the highest number of proteins fall in the same three categories (again excluding the unknown and general biochemical functions, which account for the majority of the proteins), but in a different order, with amino acid metabolism and transport having the highest fraction of proteins, followed by carbohydrate metabolism and transport, and finally transcription<sup>151</sup>.

An unexpected result concerns the translation category, which in the *L. lactis* LMG19460 is the COG with the fourth highest percentage of proteins (5.2%), while in the *L. lactis* subsp. *lactis* core genome is one of the three categories with the lowest number of proteins<sup>151</sup>.



**Figure 4.13.** Percentage of proteins of *L. lactis* LMG19460 belonging to each COG functional category, calculated by BASys. C- energy production and conversion, D- cell cycle control and mitosis, E- amino acid metabolism and transport, F- nucleotide metabolism and transport, G- carbohydrate metabolism and transport, H- coenzyme metabolism, I- lipid metabolism, J- translation, K- transcription, L- DNA replication, recombination and repair, M- cell wall/membrane/envelope biogenesis, N- Cell motility, O- post-translational modification, protein turnover, chaperone functions, P- Inorganic ion transport and metabolism, Q- secondary metabolites biosynthesis, transport and catabolism, T- signal transduction, R- general functional prediction only, S- function unknown.

The BASys software also give information about the length of the proteins of the *L. lactis* LMG19460 strain (Figure 4.14). The majority (75%) of the proteins had between 101 and 500 amino acids, which is consistent with the data from 226 unique sequences of *E. coli* where the reported median value for protein length was 277 amino acids<sup>153</sup>. While prokaryotes usually have proteins with lengths closer to 300 amino acids, the ones from eukaryotes are bigger, with around 400 amino acids<sup>153</sup>.

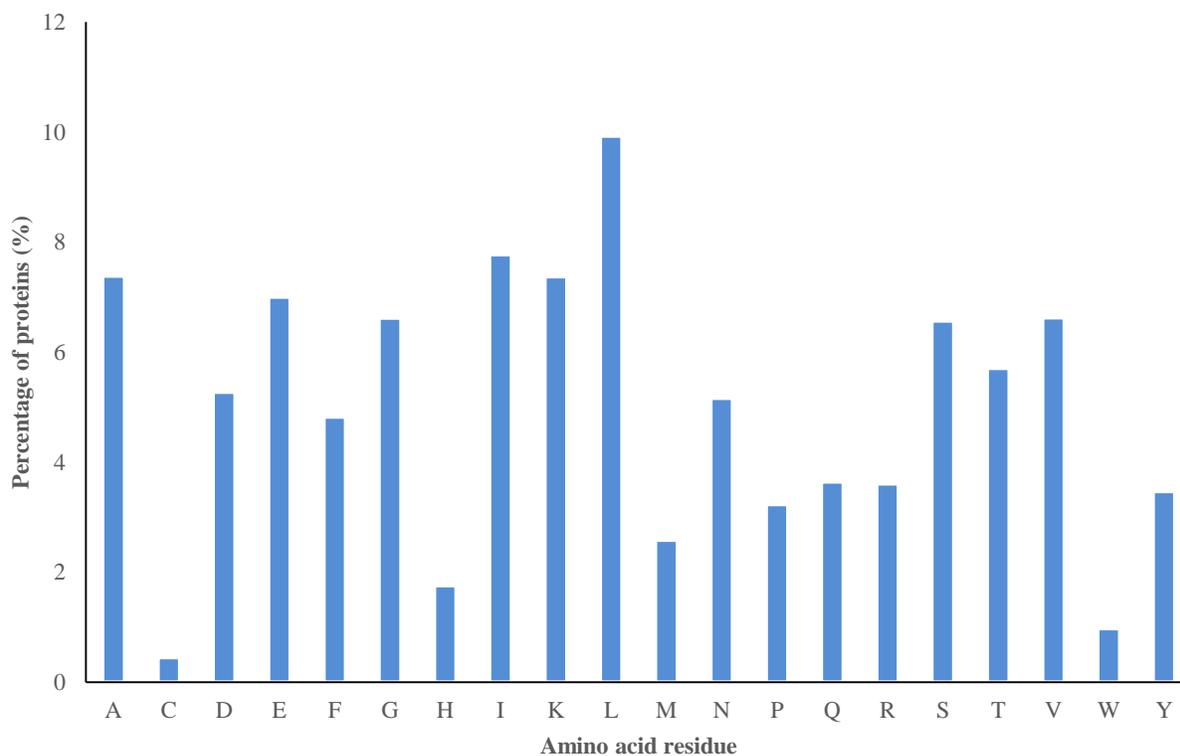


**Figure 4.14.** Percentage of proteins of *L. lactis* LMG19460 with 100 amino acids or less, 101-500 amino acids, 501-1000 amino acids, 1001-1500 amino acids or more than 1500 amino acids, as predicted by BASys.

Concerning the percentage of proteins of *L. lactis* LMG19460 with each amino acid residue, its distribution is very similar to the one described for the eubacteria group<sup>154</sup>, with a predominance of leucine, isoleucine and alanine residues, and a shortage of cysteine and tryptophan residues (Figure 4.15).

It is also of utmost importance to analyze the presence of prophages, bacteriophage genomes integrated into the *L. lactis* LMG19460 genome, since prophages are often a major source of new genes and/or new functions, related for example with virulence, toxin biosynthesis and secretion, fitness cost, genomic variations, and evolution<sup>155</sup>. These new functions could increase the adaptativeness of bacteria to new environments, which has an increased relevance in the dairy or pharmaceutical industry fields, where LAB strains are subjected to several different stress sources.

Using PHAST software, 4 complete, 5 incomplete and 1 questionable prophages were found in *L. lactis* LMG19460 genome, totalizing 187,600 bp (Table 4.9). The eighteen genomes of *L. lactis* subsp. *lactis* have an average of 3 complete intact prophages and 4 partial/remnant prophages<sup>151</sup>, meaning that the present strain under study is enriched in prophages when compared with its core genome.



**Figure 4.15.** Percentage of proteins of *L. lactis* LMG19460 with each amino acid residue, calculated by BASys. A- alanine; C- cysteine; D- aspartic acid; E- glutamic acid; F- phenylalanine; G- glycine; H- histidine; I- isoleucine; K- lysine; L- leucine; M- methionine; N- asparagine; P- proline; Q- glutamine; R- arginine; S- serine; T- threonine; V- valine; W- tryptophan; Y- tyrosine.

Analyzing the possible phages in more detail (Table 4.9), the bIL285, bIL310, bIL311 and bIL312 are in common with the *L. lactis* IL1403 strain, which is the best well characterized LAB laboratory strain and the first *L. lactis* subsp. *lactis* being completely sequenced and annotated<sup>3,156</sup>. As previously described, bIL285 is considered as having a large genome, while bIL310, bIL311 and bIL312 have smaller genomes, probably lacking the genes required for phage morphogenesis and host cell lysis, relying on the bigger phage for multiplication purposes<sup>156</sup>. The phage 98201 was also already sequenced from *L. lactis* dairy strains<sup>157</sup>.

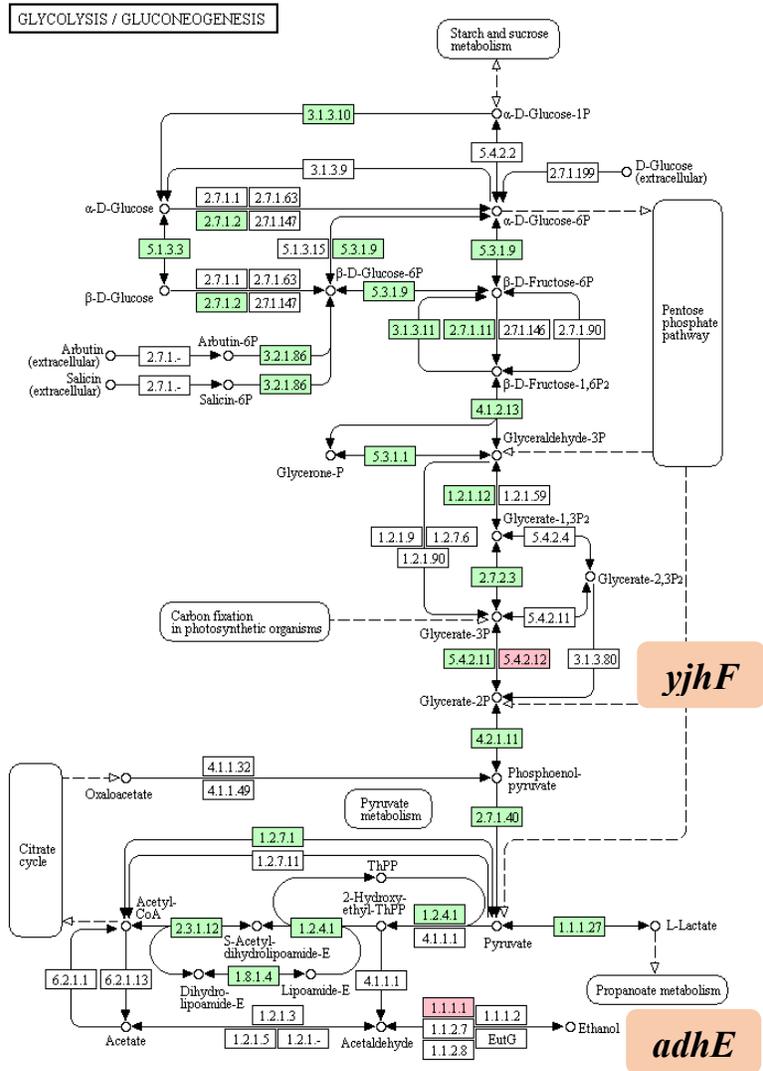
**Table 4.9.** Characterization of the prophages in the *L. lactis* LMG19460 genome, detected by PHAST.

Contig	Number of prophages	Prophage status	Prophage size (kb)	Possible phage
1	1	Complete	43.7	bIL285
2	1	Incomplete	11.4	P_SSM2
4	1	Complete	18.6	bIL312
5	1	Incomplete	11.7	phiARI0923
6	1	Incomplete	6.4	Flavob_2A
8	1	Complete	40.7	98201
10	1	Questionable	18	bIL311
13	1	Incomplete	8.7	Bacill_G
16	1	Complete	18.1	bIL310
23	1	Incomplete	10.3	P_SSM2

#### 4.4.2. Improvement of plasmid DNA and biomass production by *in silico* metabolic engineering

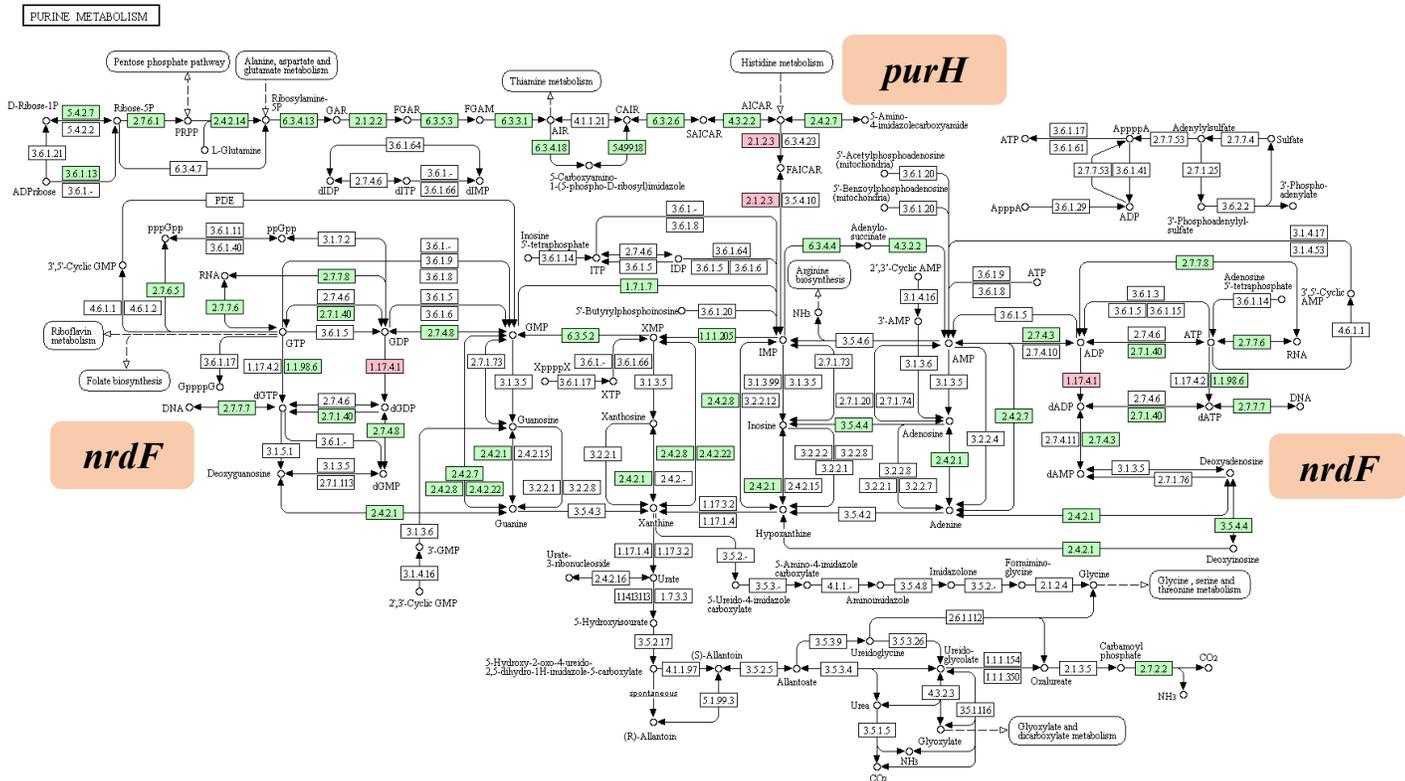
The *in silico* simulations using Optflux<sup>140</sup> software had predicted that the knockout of several genes could improve significantly the biomass and pDNA production. The software did not have the *L. lactis* LMG19460 strain, so all the simulations were performed using the most similar genome, from *L. lactis* ssp. *lactis* IL1403 strain. The genes were then analyzed using KEGG pathway module<sup>141,142,143</sup> and two main pathways appeared as more relevant: glycolysis/gluconeogenesis and purine metabolism.

Starting with the glycolysis/gluconeogenesis pathway, the Optflux software simulated that the knockout of the *yjhF* and *adhE* genes should increase biomass and pDNA production (Figure 4.16). The *yjhF* gene codes for a phosphoglycerate mutase responsible for the conversion of the glycerate-3P into glycerate-2P, while the *adhE* gene codes for an acetaldehyde-CoA/alcohol dehydrogenase, which is responsible for the conversion of acetaldehyde in ethanol and for the reverse reaction. In both cases, the gene inactivation could redirect the carbon flux to the pentose phosphate pathway, leading to an increase in nucleotide synthesis and pDNA production.



**Figure 4.16.** Position of the *yjhF* and *adhE* genes (highlighted in red) from *L. lactis* IL1403 identified by the Optflux software<sup>140</sup> as possible knockout targets in the glycolysis/gluconeogenesis pathway.

The second relevant pathway was the one responsible for the purine metabolism (Figure 4.17). The knockout of the *purH* and *nrdF* genes were predicted as leading to an increase in the biomass and pDNA production. The gene *purH* codes for a phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase, which knockout will reduce the flow from the histidine metabolism, increasing it towards nucleotide production. A similar mechanism happens with the gene *nrdF* coding for a ribonucleotide-diphosphate reductase, which knockout should increase the ADP pool towards RNA production and the GDP pool towards DNA production.



**Figure 4.17.** Position of the *purH* and *nrdF* genes (highlighted in red) from *L. lactis* IL1403 identified by the Optflux software<sup>140</sup> as possible knockout targets in the purine metabolism pathway.

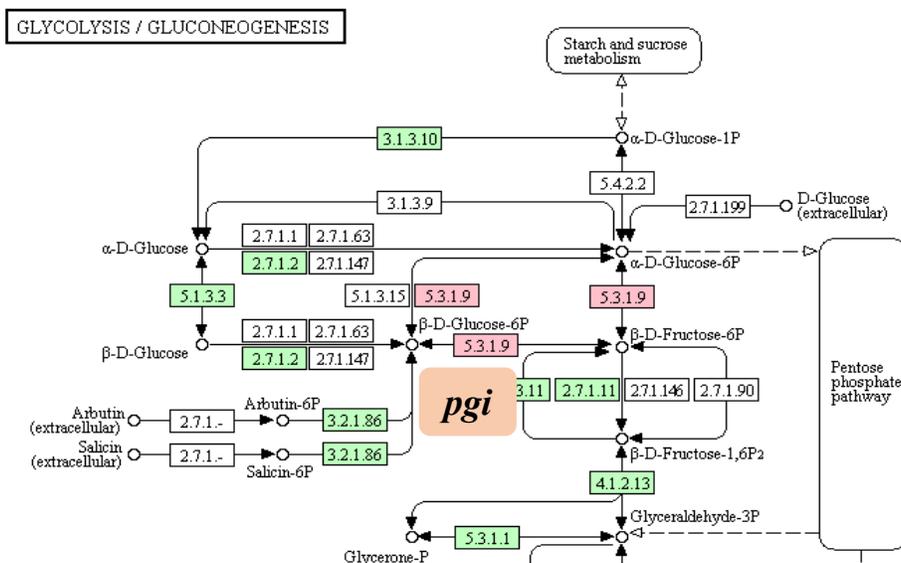
Other genes identified by the Optflux software were *araT*, *nucA*, *gltB*, *argG* and *argH*. The gene *araT* codes for an aromatic amino acid aminotransferase and intervenes in the lysine biosynthesis, while the gene *gltB* codes for the large subunit of a glutamate synthase, and the genes *argG* (argininosuccinate synthase) and *argH* (argininosuccinate lyase) enter the arginine biosynthesis pathway. The exact consequences of these knockouts that should lead to the increase in the pDNA production are not clear.

The nucleotidase coded for the *nucA* gene is responsible for the degradation of foreign DNA, meaning that its knockout is of extreme relevance in order to avoid degradation of the pDNA transformed into the cells.

Other relevant genes to perform knockout were chosen solely based on literature and previous studies, such as *nth*, *recA*, *pgi*, *htrA* and *pyk*. The *nth* gene codes for an endonuclease III responsible for the degradation of exogenous DNA by non-specific digestion<sup>3</sup>, meaning that its knockout will allow *L. lactis* LMG19460 transformation with pDNA, resulting in higher plasmid yields and better plasmid quality. The RecA protein (expressed from the *recA* gene) is a recombinase A that is able to recombine exogenous DNA<sup>158</sup>. Its knockout should allow obtaining pDNA with higher quality, without undesired recombinations.

The *pgi* gene codes for a glucose-6-phosphate isomerase that intervenes in several steps of the glycolysis/gluconeogenesis pathway<sup>159</sup> (Figure 4.18). The *pgi* gene knockout will redirect the carbon flow to the pentose phosphate pathway, increase nucleotide production and consequently pDNA production.

The triple knockout  $\Delta endA\Delta recA\Delta pgi$  was already described for *E. coli* (*E. coli* GALG20) by Gonçalves *et al.* (2013)<sup>160</sup>, which showed that it was able to produce 25-fold more plasmid DNA (19.1 mg/g dry cell weight, DCW) than its parental strain (MG1655 $\Delta endA\Delta recA$ ) (0.8 mg/g DCW), when glucose was the primary carbon source. The same effect of increasing the pDNA production is expected when the knockout be performed in *L. lactis* LMG19460.



**Figure 4.18.** Position of the *pgi* gene (highlighted in red) from *L. lactis* IL1403 identified by the Optflux software<sup>140</sup> as a possible knockout target in the glycolysis/gluconeogenesis pathway.

HtrA is an exported serine protease responsible for the degradation of secreted hybrid proteins<sup>161</sup>, so its gene knockout is of utmost importance if the goal is to have recombinant proteins being produced and secreted in high quantity and with good quality. Finally, Pyk is a pyruvate kinase that catalyzes the formation of phosphoenolpyruvate from pyruvate in the glycolysis/gluconeogenesis pathway<sup>162</sup>. Its gene knockout should act similar to the *pgi* knockout, with the carbon flow being redirected to the pentose phosphate pathway.

Since the main problem with *L. lactis* LMG19460 has been the quality and quantity of the pDNA after transformation, is absolutely essential to start with the knockout of the *nth* gene, responsible for the expression of an endonuclease III.

#### 4.4.3. Gene knockout using random mutagenesis

Random mutagenesis is a simple and quite efficient method to perform modification in bacterial genomes, but with the big disadvantage of not being possible to control which mutations will appear, having a low precision, when compared with the recombinering DNA methods. Before experiment these more time-consuming techniques, the random mutagenesis using UV light was applied to *L. lactis* LMG19460 containing the pTRKH3 plasmid. It is important that the strain has the plasmid inside when exposed to the mutagenic UV light, in order to be possible to screen for the colonies that lead to higher pDNA quantities and with better quality.

The number of colonies obtained after UV light exposure are summarized in Table 4.10.

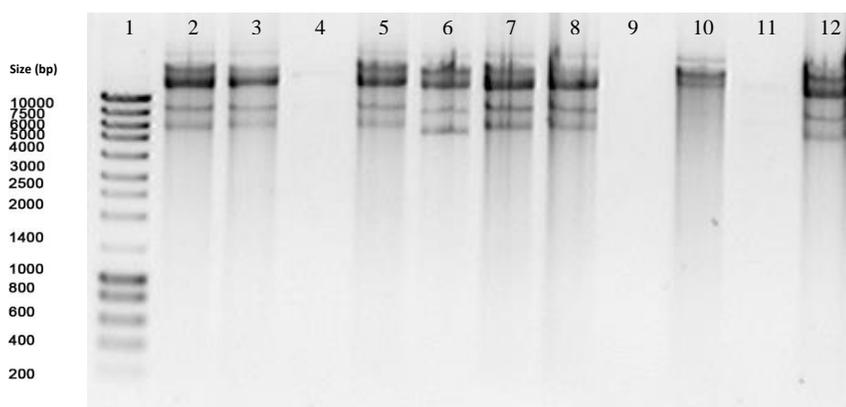
**Table 4.10.** Number of *L. lactis* LMG19460 counted after UV light exposure, at the 5<sup>th</sup> dilution, which theoretically should have 1,000 total cells.

Dilution number	UV exposure (6 cm distance from UV light)	UV exposure (0 cm distance from UV light)	No UV exposure
Real number of colonies	40	100	40

Several colonies were grown in liquid medium, more specifically 5 colonies from the plate with UV exposure at 6 cm of distance, 5 colonies from the plate with UV exposure at 0 cm of distance and 1 colony from the negative control without UV exposure. After reaching the growth stationary phase (16h vs 24h of growth), the pDNA was purified, its concentration measured (Table 4.11) and 1 µg was visualized in a 1% agarose gel (Figure 4.19) to assess its quality.

**Table 4.11.** OD<sub>600nm</sub>, pDNA concentration and pDNA specific yield from the several colonies exposed to different conditions of UV light and then grown for different time periods, starting with OD=0.1.

Colony number	Distance (cm) to UV light	Time of growth (h)	OD <sub>600nm</sub>	pDNA concentration (ng/µL)	pDNA specific yield (ng/µL/OD)
1	6	24	2.69	188	70
2	6	24	2.72	155	57
3	6	24	2.73	124	45
4	6	24	2.66	175	66
5	6	16	2.72	399	147
6	0	24	2.65	261	99
7	0	24	2.76	218	79
8	0	24	2.69	62.5	23
9	0	24	2.64	190	72
10	0	16	2.7	66	24
11	No exposure	24	2.67	278	104



**Figure 4.19.** 1 µg of each pDNA purified from the several colonies exposed to different UV light conditions. Lanes: **1-** Nzy ladder III; **2-** colony 1 (6 cm, 24h of growth); **3-** colony 2 (6 cm, 24h of growth); **4-** colony 3 (6 cm, 24h of growth); **5-** colony 4 (6 cm, 24h of growth); **6-** colony 5 (6 cm, 16h of growth); **7-** colony 6 (0 cm, 24h of growth); **8-** colony 7 (0 cm, 24h of growth); **9-** colony 8 (0 cm, 24h of growth); **10-** colony 9 (0 cm, 24h of growth); **11-** colony 10 (0 cm, 16h of growth); **12-** colony 11 (no exposure, 24h of growth).

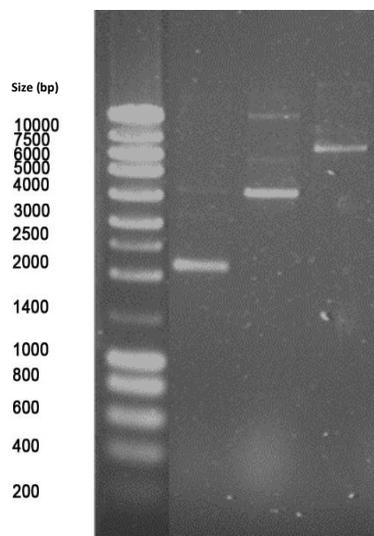
Looking at the pDNA specific yield results, only the colony 5 obtained a higher yield when compared to the negative control that was not exposed to the mutagenic UV light. This could be a good candidate for having an interesting mutation that could improve pDNA quantity and quality, but when analyzing the gel, the quality of the pDNA from colony 5 looks worse than other colonies and than the control.

Since no candidates appeared from this round of random mutagenesis, a more directed and precise approach was pursued.

#### 4.4.4. Gene knockout using Datsenko and Wanner<sup>121</sup> strategy

The first non-random knockout strategy pursued was one adapted from Datsenko and Wanner<sup>121</sup> protocol for gene knockout. Their strategy was designed for Gram-negative bacteria, more specifically *E. coli* and in this section the goal was to adapt it to work in Gram-positive bacteria.

First, all the intermediate and final knockout constructions were designed *in silico* in order to delineate the necessary primers for all the steps (Table 4.6). All the necessary plasmids (pKD13, pKD46 and pCP20) were purified and visualized in a 1% agarose gel (Figure 4.20). The three plasmids had the correct size (3,434 bp for pKD13, 6,329 bp for pKD46 and 9,400 bp for pCP20) and a majority in the supercoiled isoform, confirming its high quality.

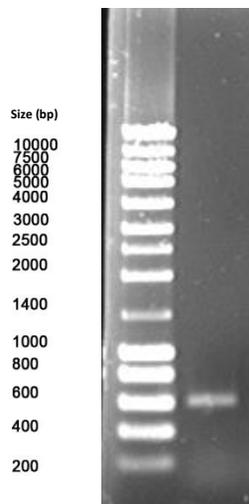


**Figure 4.20.** pKD13, pKD46 and pCP20 purified plasmids. Lanes: **1-** NzyLadder III; **2-** pKD13 plasmid (3434 bp); **3-** pKD46 plasmid (6329 bp); **4-** pCP20 (9400 bp).

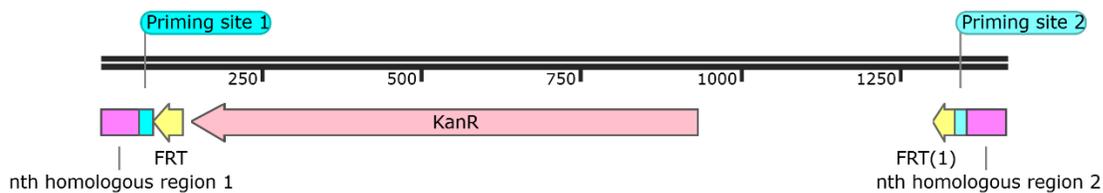
It was also necessary to optimize the antibiotic concentrations to use with *L. lactis* LMG19460 cells, since Gram-positive strains have some natural antibiotic resistances and higher antibiotic concentrations may be required, when compared with Gram-negative bacteria. One of the antibiotics was kanamycin, and it soon turned out that *L. lactis* LMG19460 were highly resistant to this antibiotic. Neomycin was the alternative found, since it is also an aminoglycoside antibiotic such as kanamycin, that binds to the 30S subunit of the ribosome, inhibiting the translation of mRNA into proteins. The kanamycin resistance gene is also able to confer resistance to neomycin, in prokaryotes. The concentration experimentally determined as optimal for the selection of correct *L. lactis* LMG19460 transformants was 1,000 µg/mL of neomycin. In case of ampicillin and chloramphenicol antibiotics, necessary for selection of pKD46 and pCP20 transformants, the determined optimal concentrations were 150 µg/mL and 75 µg/mL, respectively.

The pKD46 plasmid was then used to transform electrocompetent *L. lactis* LMG19460 cells and the resulting colonies were tested by colony-PCR. The Figure 4.21 shows an example of an agarose gel with a colony PCR positive result, harboring the 614 bp band corresponding to the amplification of a section of the ampicillin resistance gene from pKD46.

After the successful transformation of the bacterial strain with the pKD46 plasmid harboring the λ-Red recombinase system, the linear recombination cassette containing the kanamycin resistance gene was amplified from pKD13. The *in silico* representation of the amplified linear recombination cassette could be visualized in Figure 4.22. The primers KO\_nth\_F and KO\_nth\_R contained an upstream and downstream *nth* gene homologous regions and the priming sites sequences of the pKD13 plasmid. The *nth* homologous regions are necessary for the homologous recombination step performed by the λ-Red recombinase system coded by pKD46, where the linear recombination cassette will be inserted in the genome, removing the *nth* gene site. The amplified kanamycin resistance cassette was flanked by FRT (FLP recognition target) sites, where the FLP recombinase expressed by the pCP20 plasmid will act to excise the antibiotic resistance gene after the knockout had been performed.

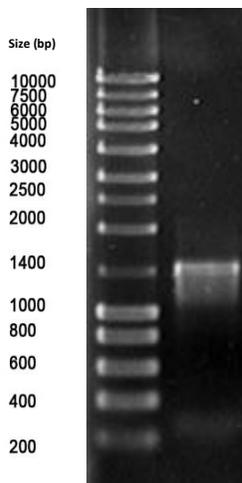


**Figure 4.21.** Lanes: 1- Nzy ladder III; 2- Positive colony PCR of *L. lactis* LMG19460 transformed with pKD46 (614 bp).

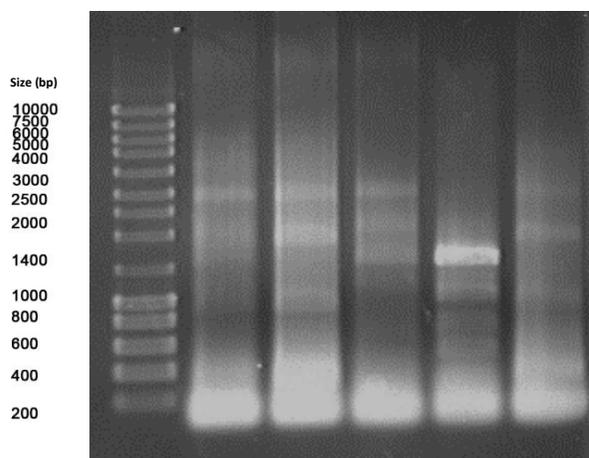


**Figure 4.22.** *In silico* representation of the linear recombination cassette amplified from pKD13, using the KO\_nth\_F and KO\_nth\_R primers.

The result of the PCR amplification of the linear recombination cassette from pKD13 plasmid was visualized in a 1% agarose gel (Figure 4.23). The perfectly visible 1,414 bp band was excised from the gel (without exposure to UV or ethidium bromide) and purified with the QIAquick Gel Extraction Kit (Qiagen). The purified fragment was used to transform the electrocompetent *L. lactis* LMG19460 harboring the pKD46 plasmid and previously induced with arabinose to express the  $\lambda$ -Red recombinase genes. The resulting colonies were grown and its gDNA purified using the Wizard Genomic DNA Purification Kit (Promega). The insertion of the linear recombination cassette in the *L. lactis* LMG19460 genome was confirmed by PCR of the colonies gDNA using Conf\_nth\_F and Conf\_nth\_R primers. If the cassette was successfully inserted, removing the wild-type *nth* gene, the agarose gel should had a band with 1,604 bp. But if the wild-type gene still remained in the genome, the band should had 1,007 bp. In Figure 4.24 is visible the agarose gel with the result of a PCR from the gDNA of several colonies resulting from the transformation of *L. lactis* LMG19460 with the linear recombination cassette. Only colonies from lanes 4 and 5 generated the band correspondent to the cassette insertion, but the colony from lane 5 should had been a mixed colony, since it also presented the band from the wild-type gene. For that reason, the colony from lane 4 was the one selected to proceed with the experiment.



**Figure 4.23.** Lanes: **1-** Nzyladder III; **2-** Linear recombination cassette amplified from pKD13 plasmid (1414 bp).



**Figure 4.24.** PCR result after transformation of *L. lactis* LMG19460 harboring the pKD46 plasmid with the linear recombination cassette. Lanes: **1-** Nzyladder III; **2-** Cassette incubated with 3 μM TTAB, colony 1; **3-** Cassette incubated with 3 μM TTAB, colony 2; **4-** Cassette incubated with 30 μM TTAB, colony 1; **5-** Cassette incubated with 30 μM TTAB, colony 2; **6-** Cassette incubated with 30 μM TTAB, colony 3.

The next step was to transform the cells from the chosen colony with the helper plasmid pCP20 and then plate it at 43°C, in order for FLP recombinase to excise the antibiotic resistance cassette between the FRT sites. If the excision successfully occurs, only a band with 430 bp should be visualized in a 1% agarose gel. But, after several attempts, no colonies had grown at 43°C. Other experiments were conducted at 37°C, but no scar was detected on the *L. lactis* LMG19460 genome.

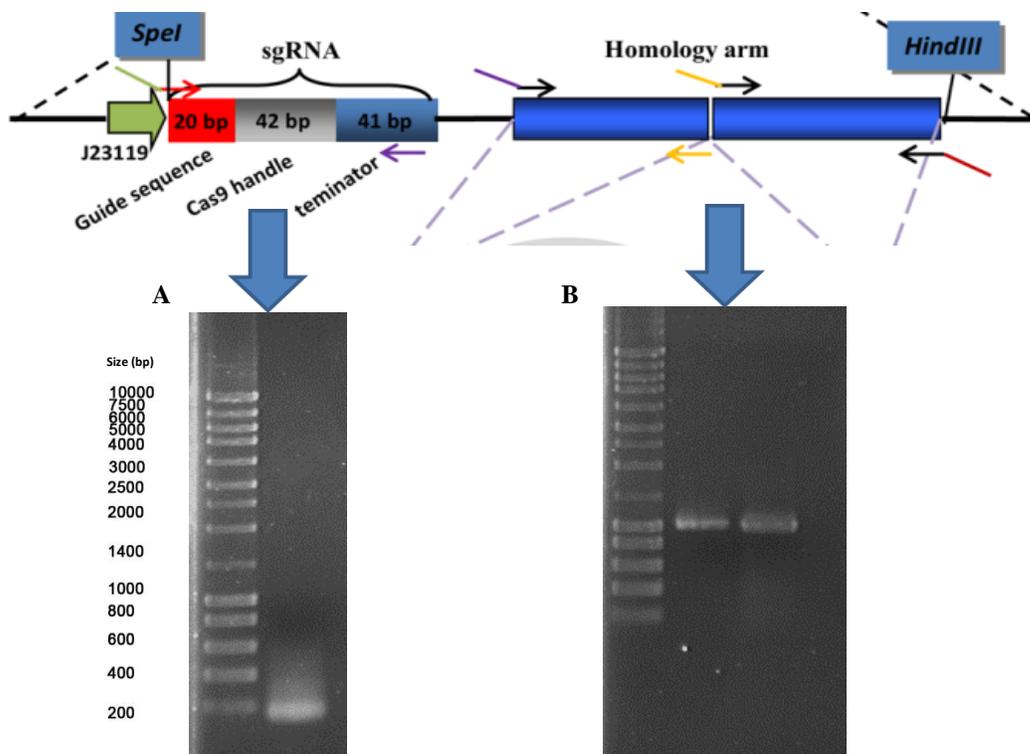
It seems that this strain is not able to grow at 43°C and the alternative at 37°C is not enough to induce the FLP recombinase activity, making this strategy unfeasible in this specific strain. Other problem could be in the promoter controlling the FLP recombinase expression, from which there is no information about its activity in Gram-positive bacteria.

This negative result was obtained after three independent experiments, creating the necessity for finding other non-random strategy to perform the *nth* knockout in *L. lactis* LMG19460 strain.

#### 4.4.5. Gene knockout using CRISPR/Cas9 strategy

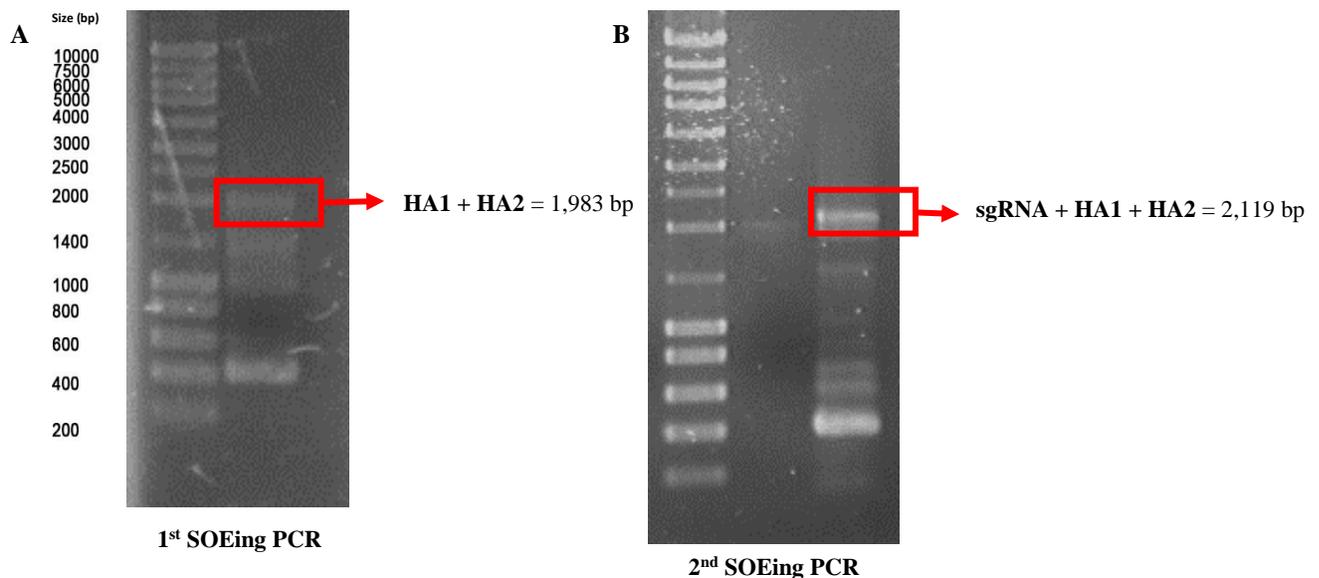
The alternative approach for *nth* gene knockout in *L. lactis* LMG19460 was to use the CRISPR/Cas9 system. The strategy was adapted for this strain from the one developed by Huang et al. (2015)<sup>135</sup> for *Streptomyces coelicolor*.

The first step was to convert the pKCCas9dO plasmid into a plasmid containing the specific *nth* sequences (guide sequence and the two homologous arms), since the original plasmid was constructed specifically for the knockout of the *actII-orf4* gene. The sgRNA fragment was amplified using the semi-overlapping primers sgRNA\_F and sgRNA\_R, which annealing resulted in the complete sgRNA sequence. The result of this amplification could be visualized in Figure 4.25A, where the expected fragment with 136 bp was clearly visible. Both homologous arms were amplified from the genome of *L. lactis* LMG19460, resulting in 1,028 and 955 bp fragments, respectively (Figure 4.25B).



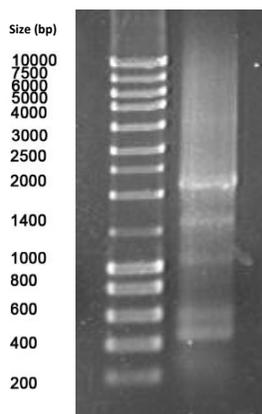
**Figure 4.25.** Amplification of *nth* gene specific A) sgRNA and B) the two homologous arms. The correct band corresponding to the sgRNA should have 136 bp, while the bands corresponding to the two homologous arms should have 1,028 and 955 bp, respectively. The DNA ladder used in both agarose gels was Nzyladder III.

After the three *nth* gene specific fragments had been amplified, it was necessary to join them by performing two successive PCR reactions. In the first SOEing PCR, the two homologous arms were joined together resulting in several unspecific fragments and the expected fragment with 1,983 bp (Figure 4.26A), which was excised and purified from the gel, to become the template for the second SOEing PCR. The second SOEing PCR, to join the fragment with the two homologous arms with the sgRNA fragment, also resulted in a high number of unspecific fragments, but it was possible to isolate and purify the DNA of the correct band with 2,119 bp (Figure 4.26B).



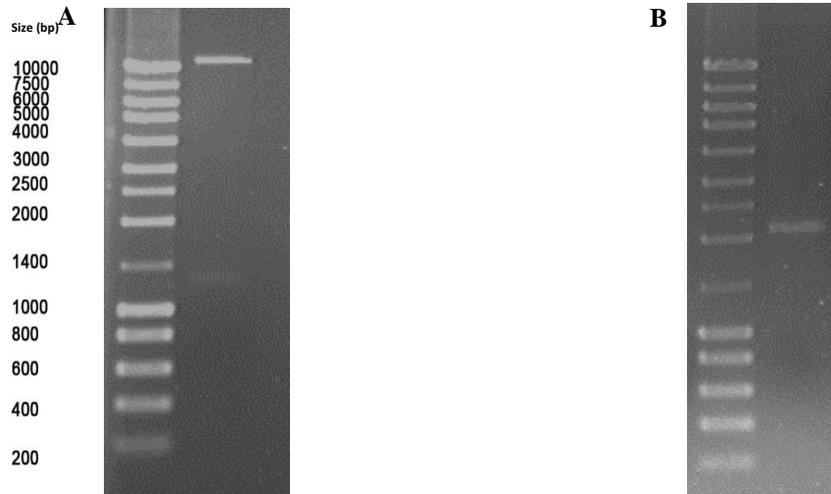
**Figure 4.26.** Agarose gels with the result of the A) first SOEing PCR to join both *nth* homologous arms (1983 bp) and the B) second SOEing PCR to join the previous fragment to the *nth* sgRNA (2119 bp). The DNA ladder used in both agarose gels was Nzyladder III.

Since the HindIII restriction enzyme is able to cut inside the sequence of the first homologous arm, to prevent its fragmentation in the restriction enzyme digestion step that anticipates the final molecular cloning procedure, it was necessary to introduce an ApaI restriction site downstream the HindIII restriction site. The result of the amplification of the complete fragment purified from the previous SOEing PCR with the primers sgRNA\_F and HA2\_R\_Apa is represented in Figure 4.27. From all the visible bands, the correct was the one with 2,131 bp, which was excised and purified from the gel, to be used in the molecular cloning steps.



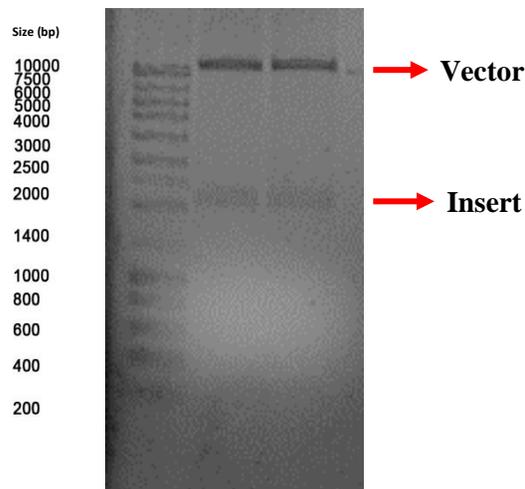
**Figure 4.27.** Lanes: 1- Nzyladder III; 2- PCR product from amplification of the fragment containing the *nth* specific sgRNA and the two homologous arms, with the primers sgRNA\_F and HA2\_R\_Apa (2131 bp).

The complete fragment, together with the vector pKCcas9dO, were digested with SpeI and ApaI and the expected bands (12,041 bp from the vector and 2,117 bp from the insert) (Figure 4.28) were purified from the gel and ligated using the T4 ligase protocol.



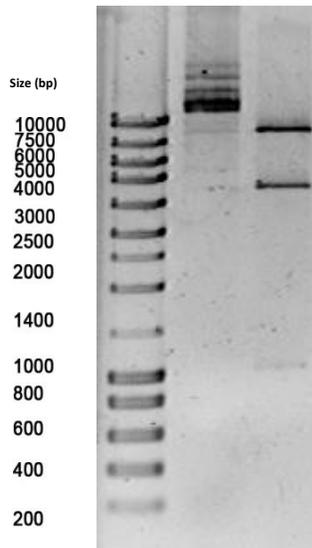
**Figure 4.28.** A) Vector pKCcas9dO (12,041 + 1,341 bp) and B) fragment with *nth* specific sgRNA and two homologous arms (2,117 + 7 + 7 bp), digested with SpeI and ApaI. The DNA ladder used in both agarose gels was Nzyladder III.

The resulting colonies were grown, its DNA purified and digested to check for the presence of both vector and insert with the correct sizes (12,041 and 2,117 bp, respectively) (Figure 4.29). The two colonies tested had the correct band pattern in the agarose gel.



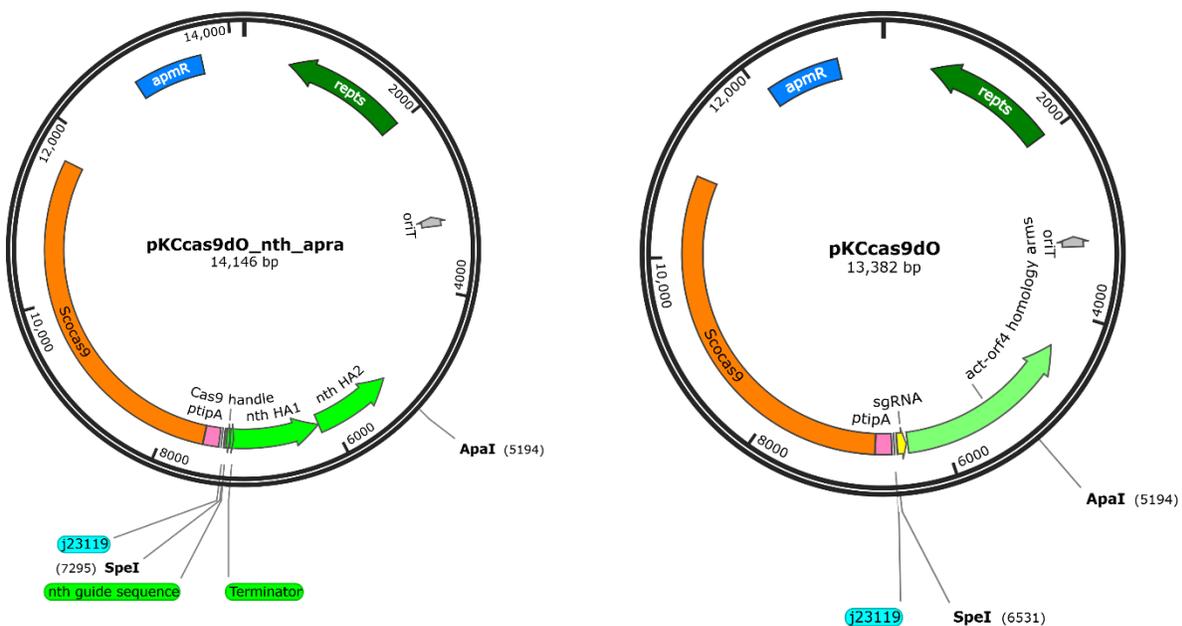
**Figure 4.29.** Confirmation of the molecular cloning of pKCcas9dO (12,041 bp) with *nth* specific fragment containing sgRNA and two homologous arms (2,117 bp). Lanes: 1- Nzyladder III; 2- colony 1; 3- colony 2.

An additional confirmation of the whole construction was performed by digesting the chosen clone with EcoRI, which cut generated the expected 8,785, 4,372 and 989 bp bands (Figure 4.30). One of the restriction sites was inside the sgRNA sequence, allowing the confirmation of its presence in the construction.



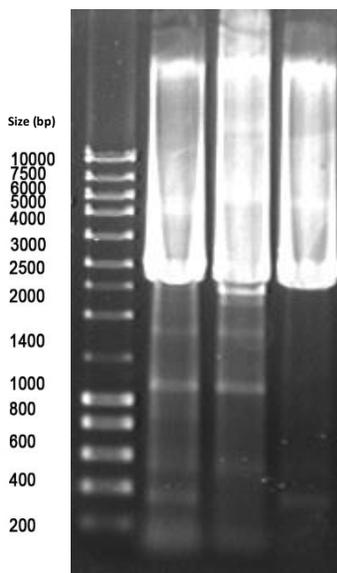
**Figure 4.30.** Lanes: 1- Nzy ladder III; 2- Non-digested final plasmid (pKCcas9dO cloned with the fragment containing *nht* specific sgRNA and two homologous arms); 3- Final plasmid (pKCcas9dO cloned with the fragment containing *nht* specific sgRNA and two homologous arms) digested with EcoRI (8785 + 4372 + 989 bp).

The final pKCcas9dO\_nth\_apra construction could be compared with the original pKCcas9dO in Figure 4.31. A small portion of the original homologous arms are still present downstream the *nht* HA2.



**Figure 4.31.** *In silico* representation of the pKCcas9dO\_nth\_apra plasmid engineered in this study, comparing with the parental pKCcas9dO plasmid.

Besides DH5 $\alpha$ , the final construction was then used to transform two additional *E. coli* strains, GM2163 and SCS110, in order to be able to purify non-methylated pDNA to use in *L. lactis* LMG19460 transformation. The colonies resulting from LAB transformation were grown in liquid medium and its pDNA and gDNA extracted, in order to confirm the presence of the CRISPR/Cas9 plasmid from the pDNA and to check for the presence of the *nth* knockout from gDNA amplification with *nth\_conf\_F* and *nth\_conf\_R*. From the pDNA alone, after 3 different attempts, the plasmid was never seen on a 1% agarose gel, but when the pDNA is amplified using primers that target the homologous arm 1, the correct band appears, proving that the strain is transformed with the necessary plasmid, but probably it is with a very low copy number. The gDNA amplification using the confirmation primers showed that all the colonies tested had the band corresponding to the wild-type *nth* gene, but the second colony also had the band with the size expected for the knockout (Figure 4.32), showing that colony 2 must be a mixed colony.



**Figure 4.32.** PCR amplification of the gDNA purified from several colonies of *L. lactis* LMG19460 transformed with pKCCas9dO\_nth\_apra plasmid. Lanes: **1-** Nzy ladder III; **2-** colony 1; **3-** colony 2; **4-** colony 3.

Several attempts to re-isolate this mixed colony were performed by restreaking in agar plates, but the band with the size expected for the knockout was never obtained again. Other relevant issue was connected with LAB transformation with the pKCCas9dO\_nth\_apra, since the purified pDNA was never visualized in an agarose gel, only the result of its amplification by PCR. Since the amplification was performed with the HA1\_F/R primers, and part of each primer has homology to the *L. lactis* LMG19460 genome, if the purified pDNA had even a trace contamination with gDNA, it could give a positive amplification. This could mislead to the conclusion that the LAB strain is transformed with the pKCCas9dO plasmid, when actually the PCR is only detecting the contaminant gDNA. Indeed, when the purified pDNA from transformed *L. lactis* LMG19460 was amplified with the *nth\_conf\_F/R* primers, it appeared a band with approximately 1,000 bp. Although the wild-type *nth* gene should amplify a 2,790 bp fragment, it is not possible to exclude the hypothesis that the pDNA could be contaminated with gDNA.

In order to confirm if the pKCCas9dO\_nth\_apra plasmid has entered the *L. lactis* LMG19460 cell, a different set of primers (Cas9\_conf\_F/R) were used that amplified a small portion of the *ScoCas9* gene, generating a 241 bp fragment

when the plasmid is present. The colony PCR to the colonies resulting from the transformation of *L. lactis* LMG19460 with pKCcas9dO did not render a positive result, showing that LAB transformation efficiency with this plasmid should be very low.

Other possible explanation for the difficulty in detecting the plasmid presence inside the *L. lactis* LMG19460 transformed strain could be the probable low copy number of the pKCcas9dO\_nth\_apra plasmid. If the plasmid has a low number of copies, the quantity of the ScoCas9 protein and of the sgRNA should be also very low, decreasing the knockout efficiency. The promoter tipA that controls the expression of ScoCas9 is inducible by thiostrepton and thiostrepton-like antibiotics, meaning that the expression of the *ScoCas9* gene could be further impaired due to the absence of thiostrepton in the growth media. In *Streptomyces lividans* the expression of the genes cloned under the control of thiostrepton-induced promoters is induced at least 200-fold after the addition of thiostrepton to the growth media<sup>148</sup>. In the absence of this antibiotic there was always a basal expression<sup>148,149</sup>, which could not exist in our specific strain, being responsible for the system not being working as expected.

The next logic step was then to add thiostrepton to the growth media as an attempt to induce the maximum *ScoCas9* expression.

#### 4.4.5.1. Induction of the ScoCas9 expression using thiostrepton

The natural resistance or susceptibility of the *L. lactis* LMG19460 strain to the thiostrepton antibiotic was completely unknown, not existing any reference in the literature concerning LAB. The wild-type strain was grown with different thiostrepton concentrations (0, 0.001, 0.01, 0.08, 0.15, 0.3, 0.625, 1.25, 2.5, 5, 10, 25 and 50 µg/mL) and the optical density was followed during 4 days in order to infer about the strain resistance to this antibiotic (Table 4.12).

**Table 4.12.** Optical density measured at 600 nm of *L. lactis* LMG19460 strain after 4 days of growth in M17 medium supplemented with different thiostrepton concentrations.

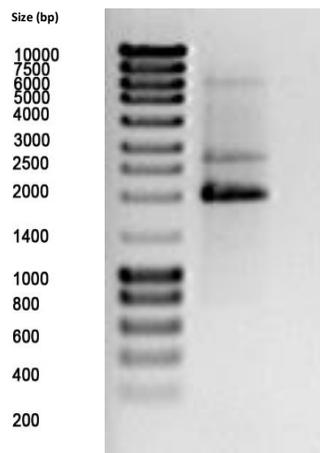
Thiostrepton concentration (µg/mL)	OD <sub>600nm</sub>
0	3.61
0.001	3.46
0.01	2.07
0.08	1.94
0.15	2.39
0.3	0.08
0.625	0.95
1.25	0.23
2.5	0
5	0
10	0
25	0
50	0

The *L. lactis* LMG19460 were only able to grow in the presence of thiostrepton concentrations under 1.25 µg/mL. The DMSO used to solubilize the thiostrepton and make the serial dilutions is in itself toxic for the cells. Even at low thiostrepton concentrations, sometimes the high DMSO volume needed of a certain antibiotic dilution was enough to inhibit the growth, explaining why the optical density did not follow a linear increase with the decrease of the antibiotic concentration.

Due to that fact, in the first thiostrepton induction experiment it was used thiostrepton both solubilized and diluted in DMSO and acetic acid, in the hope that the later had a less negative impact in the cell growth. The thiostrepton concentrations tested were 25, 2.5 and 0.3 µg/mL, since they were not yet tested with thiostrepton diluted in acetic acid, but after 7 days of growth the plates yielded no colonies.

One possible explanation for the absence of knockout could be the supercoiled state of the pKCcas9dO\_nth\_apra plasmid that could be turning the homologous arms sequences inaccessible to the cell machinery, and by transforming with the linear fragment containing these sequences, one could improve the probability of the recombination to occur. Therefore, in a second experiment, besides lowering the thiostrepton concentrations to values where the *L. lactis* LMG19460 is able to grow (0.3 and 0.01), a different approach was also attempted, which consisted in transform the strain both with the pKCcas9dO\_nth\_apra plasmid and with a linear fragment containing the two homologous arms, in the hope of increasing the probability of the recombination event.

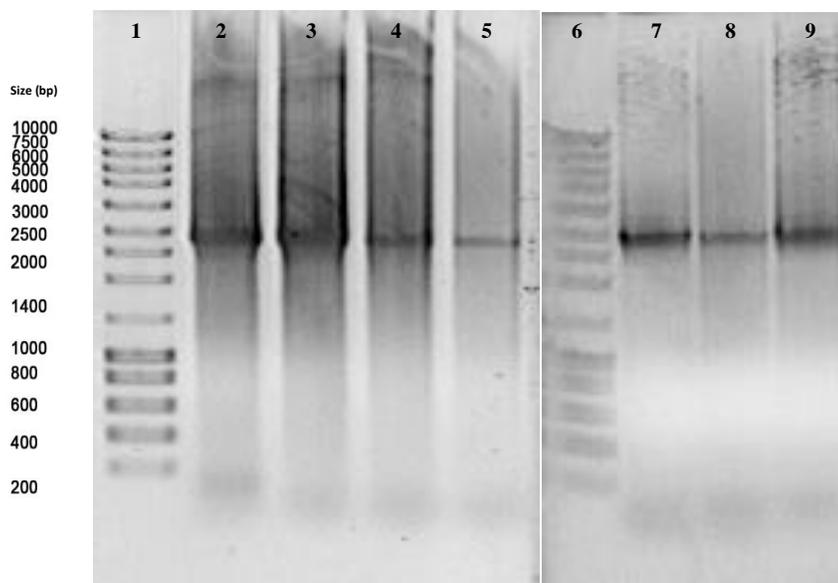
A linear fragment containing both homologous arms was amplified from the pKCcas9dO\_nth\_apra template and then the PCR mixture digested with DpnI to remove all the parental plasmid (Figure 4.33). The expected fragment with 2,007 bp was extracted and purified from the gel. After, 100 ng of the linear fragment together with 1 µg of the pKCcas9dO\_nth\_apra were used to transform *L. lactis* LMG19460 with 0, 0.01 or 0.3 µg/mL thiostrepton to induce the tipA promoter, since concentrations as low as 10<sup>-9</sup>M (0.0017 µg/mL) were shown to work in *S. lividans*<sup>163</sup>. Control conditions without apramycin and without the linear fragment in the transformation mixture were also performed.



**Figure 4.33.** PCR amplification of the linear fragment containing the two *nth* homologous arms. The molecular marker used was the Nzyladder III.

The conditions without thiostrepton and/or without apramycin rendered a confluent growth. The condition where the strain was transformed with only the pKCcas9dO\_nth\_apra plasmid and induced with 0.01 µg/mL thiostrepton resulted in a plate with around 200 isolated colonies, while the condition where the strain where transformed also with the linear fragment and induced with 0.3 µg/mL thiostrepton had one isolated colony in the final plate. No other experimental condition resulted in colonies. Several colony PCRs were performed to assess if any colony had the correct knockout. If the *nth* gene knockout occurred, the amplification with the nth\_conf\_F/R primers should result in a 2,133 bp band, while the wild-type strain without the knockout should result in a 2,790 bp band.

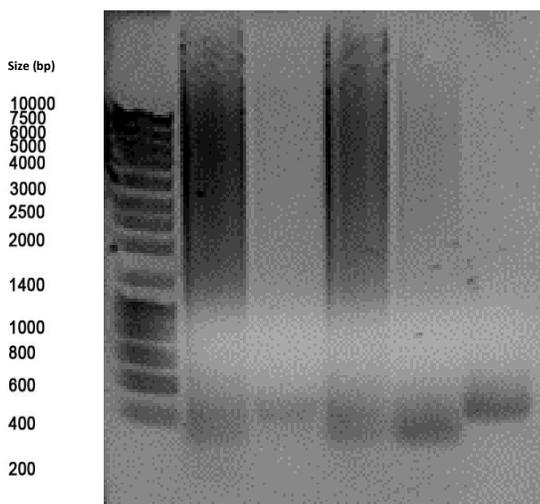
The agarose gels performed after two different colony PCRs are represented in Figure 4.34. It is clearly visible that any colony had the band corresponding to the *nth* knockout.



**Figure 4.34.** Colony PCR results from the second thiostrepton induction experiment with nth\_conf\_F/R primers. Lanes: **1-** Nzy ladder III; **2-** strain transformed with only the pKCcas9dO\_nth\_apra plasmid and induced with 0.01 µg/mL thiostrepton, colony 1; **3-** colony 2; **4-** colony 3; **5-** strain transformed with the the pKCcas9dO\_nth\_apra plasmid and the linear fragment and induced with 0.3 µg/mL thiostrepton, colony 1; **6-** Nzy ladder III; **7-** strain transformed with only the pKCcas9dO\_nth\_apra plasmid and induced with 0.01 µg/mL thiostrepton, colony 4; **8-** colony 5; **9-** colony 6.

To check if the problem could be related to the strain not being transformed with the pKCcas9dO\_nth\_apra plasmid, a colony PCR with the Cas9\_conf\_F/R primers was performed (Figure 4.35). Because the sequence of the linear fragment is included in the plasmid, there were no way to check by colony PCR if the linear fragment was inside the transformed cells. Only the colony 2 had been successfully transformed with the plasmid, having the expected fragment of 241 bp, such as the positive control lane (pKCcas9dO\_nth\_apra plasmid purified from *E. coli* DH5α). When the 4 colonies were picked to M17 liquid medium, again only the colony 2 were able to grow, confirming the colony PCR results. These results show that the reason behind the negative knockout results does not seem to be related with an inability to successfully transform the strain with the plasmid of interest. But, again, when the pDNA is purified from

the colony after growing in liquid medium, any band was visible in an agarose gel, pointing out that the pKCcas9dO\_nth\_apra plasmid must replicate in very low copy number inside the *L. lactis* LMG19460 cells.



**Figure 4.35.** Colony PCR results from the second thiostrepton induction experiment with the Cas9\_conf\_F/R. Lanes: **1-** Nzyladder III; **2-** strain transformed with only the pKCcas9dO\_nth\_apra plasmid and induced with 0.01  $\mu\text{g}/\text{mL}$  thiostrepton, colony 1; **3-** colony 2; **4-** colony 3; **5-** colony 4; **6-** positive control (pKCcas9dO\_nth\_apra purified from *E. coli* DH5 $\alpha$ ).

One possible explanation for the negative result could be that the thiostrepton concentration used was not enough to induce the *tipA* promoter, although concentrations as low as  $10^{-9}\text{M}$  (0.0017  $\mu\text{g}/\text{mL}$ ) were shown to work in *S. lividans*<sup>163</sup>, but it was verified that higher concentrations were toxic for cell growth. These reasons turns it inevitable, as a future work goal, to change this thiostrepton-induced promoter to a more user friendly promoter to express the *ScoCas9* gene, such as the native inducible nisin A promoter or constitutive promoters such as P2 or P3<sup>164</sup>.

Other problem encountered during these experiments was the low transformation efficiency, and one possible reason behind this could be the antibiotic resistance gene present in pKCcas9dO\_nth\_apra. Since LAB are highly resistant to apramycin, probably we were getting a majority of false positive colonies, which makes it unfeasible to use apramycin as the selection marker.

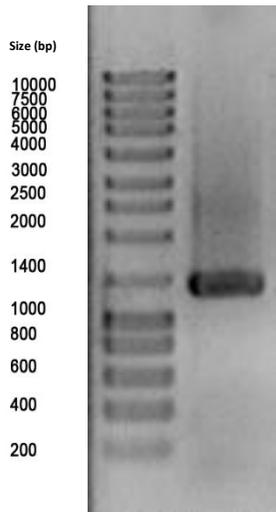
#### 4.4.5.2. Replacement of the apramycin resistance gene for the erythromycin resistance gene

An obvious choice of an alternative antibiotic resistance gene for LAB is erythromycin, since it is one of the most widely used antibiotics for LAB strains. It was not feasible to simply use molecular cloning to replace the antibiotic resistance gene, due to lack of available unique restriction enzyme sites, so a Gibson Assembly approach was pursued.

First, the three necessary fragments had to be amplified from the correct template pDNAs. The first fragment, corresponding to the complete erythromycin resistance gene and its promoter upstream, was amplified from the pTRKH3 plasmid, using the Ery\_F/R primers. The amplification yielded a high fragment concentration with 1,299 bp,

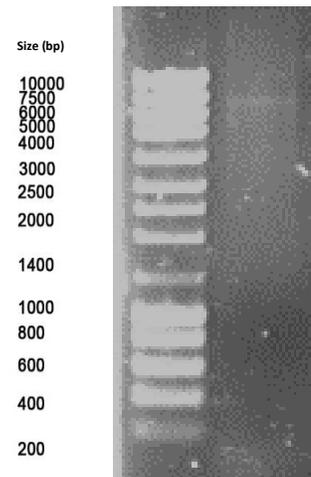
as could be visualized in Figure 4.36. After DpnI digestion, the PCR product was purified using NZYGelPure (Nzytech).

The two other fragments were amplified from pKCcas9dO\_nth\_apra, corresponding to the region between the end of the antibiotic resistance gene and the end of the truncated parental homologous arm (Ery\_Apa\_F/R primers), and to the region between the ApaI restriction enzyme site and the beginning of the antibiotic resistance gene (Apa\_Ery\_F/R primers) (Figure 4.12).



**Figure 4.36.** 1- Nzyladder III; 2- PCR amplification of the erythromycin resistance gene from pTRKH3 with Ery\_F/R primers.

The Apa\_Ery fragment was successfully amplified from pKCcas9dO\_nth\_apra plasmid generating the correct 7,686 bp band (Figure 4.37). After DpnI digestion, the PCR product was extracted from a 1% agarose gel to exclude the parental plasmid and purified using NZYGelPure (Nzytech).



**Figure 4.37.** 1- Nzyladder III; 2- PCR amplification of the Apa\_Ery fragment from pKCcas9dO\_nth\_apra plasmid with Apa\_Ery\_F/R primers.

Unfortunately, the Ery\_Apa fragment was never successfully amplified, even after optimization of the annealing temperatures (50°, 54°, 58°, 60°, 62°C), quantity of template pDNA (0.5-10 ng), origin of the template pDNA (DH5 $\alpha$  or GM2163 strains) and MgSO<sub>4</sub> concentration in the PCR mixture (1-2.3 M).

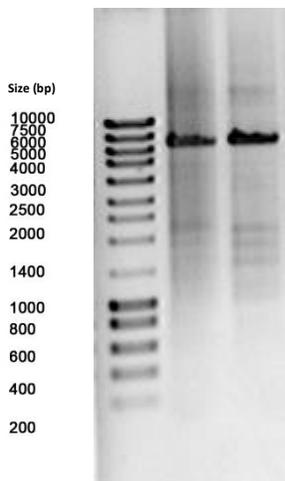
Since the main idea was to replace the apramycin resistance gene for the erythromycin one, and the Ery\_Apa fragment amplification turned out to be very challenging, a different approach was addressed.

#### 4.4.5.3. Insertion of the *ScoCas9* gene, sgRNA and homologous arms into the pTRKH3 vector

The new approach consisted in inserting the necessary CRISPR/Cas9 components for the *nth* knockout to occur (the *ScoCas9* gene under the control of the *tipA* promoter, the sgRNA under the control of the J23119 promoter and the two homologous arm sequences) into the pTRKH3 backbone plasmid. The most obvious advantage of this approach is that the pTRKH3 plasmid harbours the erythromycin gene, avoiding the more difficult selection using the apramycin antibiotic, to whom the LAB are naturally resistant. Other issue relates to the fact that the pSG5 origin of replication present in the pKCCas9dO plasmid, to the best of our knowledge, was never tested in LAB and could not work properly in our specific strain or even in the whole group of LAB. This could be impairing the pKCCas9dO replication inside the *L. lactis* LMG19460 cells, even if they are successfully transformed with the correct plasmid, and consequently decreasing dramatically the amounts of *ScoCas9* and sgRNA being produced. All these problems will undoubtedly decrease the probability of a knockout event to occur.

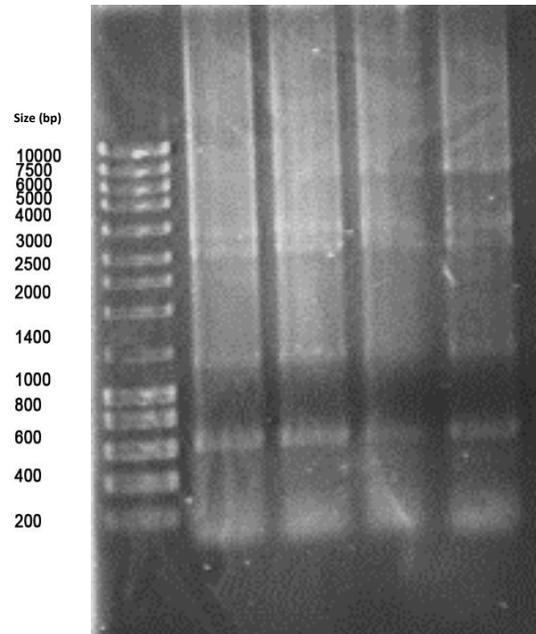
Transferring the necessary components from the pKCCas9dO plasmid to the pTRKH3 plasmid should avoid these problems, since the pTRKH3 has the pAM $\beta$ 1 origin of replication that is known to replicate with a medium plasmid copy number in LAB. While designing the necessary primers to amplify and then assemble the fragments, the tetracycline resistance gene sequence was excluded from the final plasmid, in order to decrease the plasmid size to increase transformation efficiency in LAB.

The smaller pTRKH3 vector was successfully amplified and further digested with DpnI, as visible in Figure 4.38. The correct 6,575 bp band was excised and purified from the gel without UV or ethidium bromide exposure, for further Gibson Assembly procedure.



**Figure 4.38.** PCR amplification of the shorter pTRKH3 plasmid that will work as the vector for assembling the necessary CRISPR/Cas9 components, with pT vector\_F/R primers and using different annealing temperatures. 1- Nzy ladder III; 2- 60°C; 3- 72°C.

The fragment containing the necessary CRISPR/Cas9 components (the *ScoCas9* gene under the control of the *tipA* promoter, the sgRNA under the control of the J23119 promoter and the two homologous arm sequences) was also amplified, generating the expected 6,480 bp fragment when higher annealing temperatures were applied (Figure 4.39). The band was also purified from the gel and its DNA concentration quantified.



**Figure 4.39.** PCR amplification of the insert with the necessary CRISPR/Cas9 components, using different annealing temperatures. 1- Nzyladder III; 2- 56°C; 3- 60°C; 4- 68°C; 5- 72°C.

As a future work, the next step should be to assemble the two fragments, using a Gibson Assembly approach and transform *E. coli* DH5 $\alpha$  with the assembled plasmids. The resulting colonies must be tested for the presence of the correctly assembled plasmids and this pDNA used to transform *L. lactis* LMG19460 with the goal of generating the CRISPR/cas9-based *nth* knockout.

Further future work could be performed if this approach does not work, such as using the pWV01-derivative pVE6002 thermo-sensitive replicon, which becomes inactive at temperatures above 35°C, being of utmost importance for the plasmid curing step<sup>57</sup>.

The fact that the *ScoCas9* gene was codon optimized for *S. coelicolor* could be other reason that is impairing the present system, since *L. lactis* LMG19460 may be having difficulties in translating the codons from this sequence. Although both species are Gram-positive, the preferred codons could be quite different, inasmuch its GC contents are very dissimilar (70% and 35% for *S. coelicolor* and *L. lactis*, respectively)<sup>4,165</sup>. A future work proposition will be to optimize this gene sequence for the codon usage bias of *L. lactis*.

Other possible approach is to generate and assemble a new *nth* guide sequence, in order to increase the probability of successful genome pairing and Cas9 recognition and cutting.

## 4.5. Conclusions

The final goal of this chapter was to turn the *L. lactis* LMG19460 into a strain able to produce pDNA and recombinant proteins with high yields and quality, by performing knockout of the genes suggested by the *in silico* analysis, using the one-step CRISPR/Cas9 system.

The production of pDNA and recombinant proteins in LAB is likely to be increased if the *nth* (endonuclease) and *recA* (recombinase) genes are deleted to minimize non-specific digestion of DNA and recombination, respectively. Additionally, in *L. lactis* the inactivation of the *htrA* gene encoding a protease abolished the degradation of heterologous proteins. *In silico* simulations using Optflux were used to find other candidate genomic genes, important for protein expression or plasmid production, for knockout or overexpression. A preliminary analysis showed that the removal of specific genes (*araT*, *nucA*, *yjhF*, *gltB*, *nrdF*, *adhE*, *argh*, *argg* and *purH*) should increase biomass and pDNA production, being excellent candidates for knockout.

The first gene chosen for knockout was the *nth* gene, due to the utmost importance of having a strain that produces pDNA in high quantity and with high quality for molecular biology purposes, and if the final goal is to produce DNA vaccines and recombinant proteins for live delivery therapeutic applications. The majority of the plasmids available for LAB have a low PCN, being essential to increase the plasmid quantities that the strains are able to produce, without impairing its growth.

Two preliminary methods for gene knockout were first performed, due to its simplicity and availability in our laboratory, namely the random mutagenesis and the Datsenko & Wanner recombineering strategy<sup>121</sup>. None of this approaches had the pretended result of generating the *nth* knockout.

The next approach was to create the LAB mutants using the CRISPR/Cas9 system, a recent technology for precision genome engineering that allows site-specific insertions, deletions or replacement of genetic material. The CRISPR/Cas9 plasmid available in our laboratory is the pKCCas9dO, originally designed for *Streptomyces coelicolor* genome editing. To the best of our knowledge this is the only one-step plasmid designed for Gram-positive bacteria, without necessity of helper plasmids. This was the first test of a CRISPR/Cas9 one-step system in *L. lactis*, highlighting the relevance of this chapter.

We were able to design and replace the necessary parts of the pKCCas9dO, namely the two homology-directed repair templates and the target-specific guide RNA (pKCCas9dO\_nth\_apra plasmid), with the goal of making the knockout of the *nth* gene in *L. lactis* LMG19460. After transformation of *L. lactis* LMG19460 with the pKCCas9dO\_nth\_apra plasmid, no recombinants were found. Besides the low copy number of the pKCCas9dO\_nth\_apra plasmid that does not allowed the confirmation that the strain was successfully transformed, a problem concerning the *tipA* promoter was addressed. This promoter controls the expression of the *ScoCas9* gene and is inducible by thiostrepton, although it was already described as could having a basal constitutive expression in *Streptomyces*.

In order to try to increase the probability of the *nth* knockout event to occur, after *L. lactis* LMG19460 transformation with the pKCCas9dO\_nth\_apra plasmid and/or a linear fragment containing both homologous arms, the *tipA* promoter was induced with different non-lethal concentrations of thiostrepton. There was not obtained any positive knockout

result, but it was proved that the strain was successfully transformed with the correct plasmid, which was an improvement when compared with the previous results. It was probably necessary a higher thiostrepton concentration to induce this promoter in a way to produce enough *ScoCas9* for the knockout to occur, but the problem was that this antibiotic was highly lethal for the *L. lactis* LMG19460 strain when used in higher concentrations than the ones we used in the present experiment. As a future work it is inevitable to change this thiostrepton-induced promoter to a more adequate promoter to express the *ScoCas9* gene, such as the native inducible nisin A promoter or constitutive promoters such as P2 or P3<sup>164</sup>.

The low transformation efficiency with the pKCcas9dO\_nth\_apra plasmid has been a problem encountered across all the different CRISPR/Cas9 experiments, and the hypothesis that was placed to explain this result was related with the high natural resistance of the strain to the antibiotic apramycin, which could impair the correct recombinant selection. Probably we were getting a majority of false positive colonies, which makes it unfeasible to use apramycin as the selection marker, and the solution goes through replacing this gene for the well-studied erythromycin resistance gene. A critical issue with one of the necessary fragments amplification was the reason why this approach was abandoned and an alternative approach was pursued.

Since the template pKCcas9dO plasmid was originally engineered to work in *Streptomyces*, and not in *L. lactis*, although both strains are Gram-positive bacteria, the origin of replication pSG5 could not work, turning the plasmid into a non-replicative plasmid. The next approach was to insert the necessary CRISPR/Cas9 components into the pTRKH3 plasmid, which have the well-studied pAM $\beta$ 1 origin of replication that has a medium plasmid copy number in LAB. Both fragments were successfully amplified, but further work will be necessary to assemble them by Gibson assembly. If in the future the assembly could be correctly achieved, *L. lactis* LMG19460 should be transformed with the plasmid. After selecting the correct transformant, the *nth* gene knockout should be confirmed by PCR amplification and the plasmid should be cured. It is important to transform the *Anth* strain with a known plasmid, such as pTRKH3, to ascertain if the new PCN increased, when compared with the PCN in the wild-type strain, which could be quantified by quantitative real-time PCR. The purified pTRKH3 from the wild-type and the *Anth* strains should be visualized and compared in an agarose gel, to check for the increase in the plasmid quality, namely the increase of the supercoiled fraction.

Other relevant aspect is that the sequence of the *ScoCas9* itself was codon-optimized for *Streptomyces* and as a future work it should be optimized for the codon usage bias of *L. lactis* to try to increase its expression in this strain. Also the guide sequences could be replaced by others in order to try to increase the knockout event probability.



# CHAPTER 5

**Heterologous production of curcumin by probiotic lactic acid bacteria towards gastrointestinal cancer therapy**

*I seem to have been only like a boy playing on the seashore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me. - Isaac  
Newton*

## 5.1. Abstract

Cancer is the leading cause of death in developed countries and the cancers in the gastrointestinal tract appear in the top five of the most killing cancers. Several research efforts have been conducted towards the development of effective treatments, including the search for new drugs, since the current chemotherapeutic agents have several disadvantages.

Curcumin is the yellow pigment from *Curcuma longa* and pre-clinical studies have shown its ability to inhibit carcinogenesis in several cell lines. Completed clinical trials suggest that curcumin is useful for the chemoprevention of colon cancer in humans, being safe for normal cells and with minimal adverse effects, making it a good complement to the current chemotherapeutic agents.

The aim of the present chapter is to engineer an artificial curcumin pathway in Lactic Acid Bacteria (LAB), since they are efficient producers of heterologous proteins. Curcumin has a slow cellular uptake, leading to low systemic bioavailability. Therefore, repetitive oral doses of the molecule are required to achieve an effective therapeutic concentration inside the cells, and the possibility of synthesizing it *in situ*, i.e. in the human gastrointestinal tract, by LAB represents a powerful and innovative solution for gastrointestinal cancer therapy.

A minimal pathway with three genes for curcumin biosynthesis will be cloned in an *Escherichia coli*/LAB shuttle plasmid vector, under the control of a constitutive promoter. LAB are Generally Recognized As Safe (GRAS) and most of them act as probiotics, which make them ideal vectors to deliver specific agents. Therefore, the possibility of using LAB as cell factories to produce curcumin in the human gastrointestinal tract, constitutes a promising solution for gastrointestinal cancer therapy, which may allow overcoming the problem of the poor oral bioavailability of curcumin.

The proof-of-concept that *L. lactis* LMG19460 were able to produce curcumin was not yet achieved and several hypotheses were investigated.

## 5.2. Introduction

Cancer is the leading cause of death in developed countries, but even in developing countries is the second leading cause of death<sup>166</sup>. About 14.1 million cancer cases and 8.2 million cancer deaths are estimated to have occurred in 2012 worldwide<sup>167</sup>.

Although breast and lung cancer are the most frequently diagnosed and the leading cause of cancer death in females and males, respectively, the cancers in the GI tract (esophagus, stomach, colorectal) appear in the top five of the most killing cancers<sup>166</sup>.

The incidence rates of colorectal cancer are increasing in several areas worldwide, probably due to a combination of factors such as changes in dietary patterns, smoking and obesity. It is the third most commonly diagnosed cancer in males and second in females. Males have a higher probability of being diagnosed with the other types of GI tract cancer, with esophagus cancer being 3 to 4 times and stomach cancer 2 times more frequent among males than females<sup>166</sup>.

The standard treatment options includes surgery<sup>168,169</sup>, adjuvant chemotherapy<sup>169,170</sup> and adjuvant radiation therapy<sup>171</sup>. Several research efforts have been conducted towards the development of effective treatments, including the search for new drugs<sup>172,173</sup>.

Besides its high cost, the main problems of usual chemotherapeutic agents are its toxicity for tumor cells as well as for normal cells, which results in several major side effects, and the inability to prevent cancer<sup>174</sup>.

### 5.2.1. Curcumin – a putative solution

Curcumin is a polyphenol and the major yellow-orange pigment of turmeric, a spice derived from the rhizome of the Indian plant *Curcuma longa*<sup>175</sup>. The compound was isolated first in 1815 and its chemical structure was determined in 1973<sup>176</sup>.

#### 5.2.1.1. Biological activities

Curcumin has been used for centuries in Indian and Chinese traditional medicines and is considered safe by the Food and Drug Administration among other organizations, but has only entered extensive preclinical studies and scientific phase I and II/III clinical trial levels in the last 10–15 years<sup>177</sup>.

Curcuminoids have several biological activities and health benefits, such as anti-inflammatory and anti-cancer properties<sup>174</sup>. Pre-clinical studies have shown the ability of curcumin to inhibit carcinogenesis in several cell lines including breast, colon, gastric, hepatic, leukemia, oral epithelial, pancreatic, and prostate cancer<sup>178</sup>. Completed clinical trials suggest that curcumin is useful for the prevention of colon cancer in humans and has minimal adverse effects, which are two of the main advantages when compared with the usual chemotherapeutic agents. It is also safe for normal

cells, since curcumin induce apoptosis only in cancer cells<sup>178</sup>. All these properties of curcumin make it a promising alternative to the current chemotherapeutic drugs.

In the cancer treatment field, it was shown that curcumin downregulates several pro-inflammatory cytokine expressions such as tumor necrosis factor (TNF- $\alpha$ ), interleukins (IL-1, IL-2, IL-6, IL-8, IL-12) and chemokines. This molecule also has anti-inflammatory properties, since it decreases the inflammation associated with experimental colitis, including a substantial reduction of the rise in myeloperoxidase activity, an established marker for inflammatory cells and TNF- $\alpha$ <sup>177</sup>.

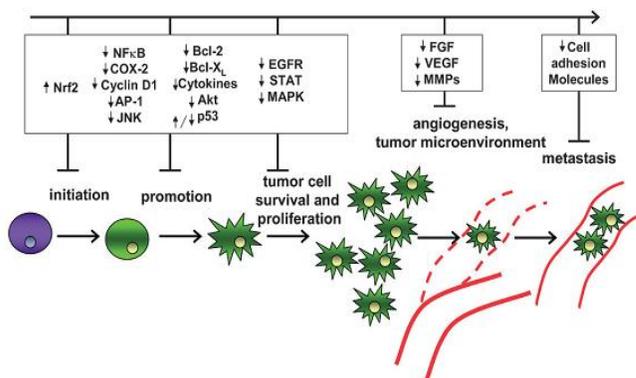
Curcumin is also a potent anti-oxidant due to its ability to scavenge directly the reactive oxygen species (such as O<sub>2</sub><sup>-</sup>, OH, NO and ONOO<sup>-</sup> radicals)<sup>177</sup>. It is also capable of modulating the immune system, by activating T-cells, B-cells, macrophages, neutrophils, natural killer cells and dendritic cells, and at low doses can also enhance antibody responses<sup>177</sup>.

It is the joint effect of anti-oxidant, anti-inflammatory and immune system modulator characteristics that contributes for curcumin anti-cancer properties<sup>177</sup>. But, although it has a very promising biological activity, its therapeutic effectiveness remains a challenge due to its low solubility and poor bioavailability, as will be discussed ahead.

### 5.2.1.2. Anti-tumor properties – mechanisms of action

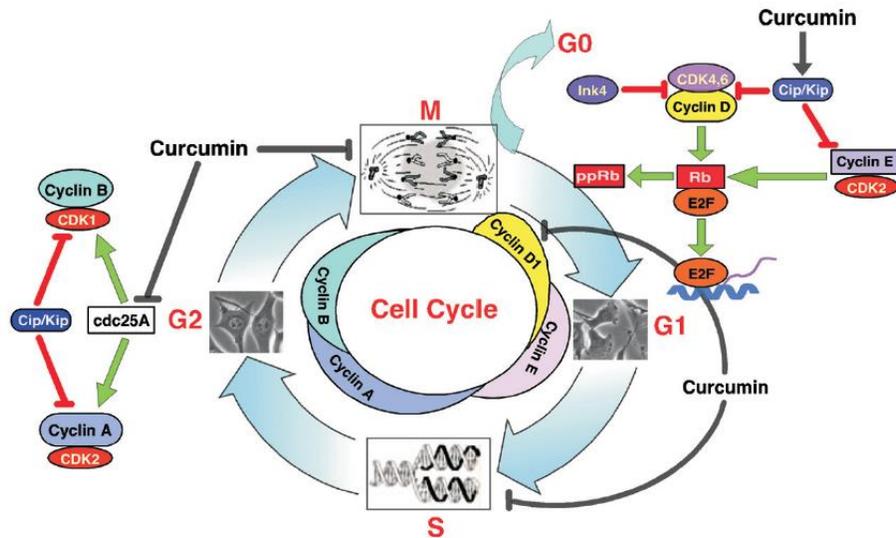
Cancer is a hyperproliferative disorder and those type of cells differ from normal cells in six main aspects: self-sufficiency in growth signals, insensitivity to growth suppressors, evasion of apoptosis, limitless replicative potential, sustained angiogenesis, and tissue invasion and metastasis<sup>174,179</sup>.

In tumor cells the balance between growth promoting and growth restraining signals is disrupted, meaning that there are continued cell proliferation and loss of differentiation. Also, the normal process of programmed cell death that exists in normal cells may no longer operate in the cancer cells. The anti-tumor effect of curcumin has been attributed to the suppression of cell proliferation, reduction of tumor load and induction of apoptosis (Figure 5.1). One of the main advantages of this compound is that preferably induces apoptosis in highly proliferating cells, meaning that death is much more pronounced in tumor cells than in normal cells, making it a relatively safe anti-cancer agent. In fact, curcumin arrests non-malignant cells in G<sub>0</sub> phase reversibly but do not induce apoptosis in these normal cells<sup>180</sup>.



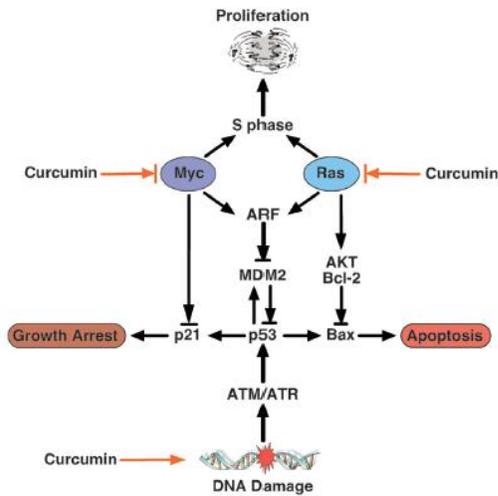
**Figure 5.1.** Stages in tumor progression inhibited by curcumin. Figure from Hatcher *et al.* (2008)<sup>181</sup>.

The mammalian cell cycle is divided into four distinct phases (G1, S, G2, and M) and when the integrity of one of the signaling pathways that regulates the cell cycle is lost, it can lead to a hyper-proliferative state of cells, manifested as cancer. Curcumin can alter the deregulated cell cycle in cancer cells via cyclin-dependent (Figure 5.2), p53-dependent (Figure 5.3) and p53-independent pathways (Figure 5.4)<sup>180</sup>. Curcumin modulates CKIs, CDK-cyclin and Rb-E2F complexes to render G1-arrest and alters CDK/cyclin B complex formation to block G2/M transition, and because of that is considered one of the major natural anticancer agents exerting anti-neoplastic activity in various types of cancer cells (Figure 5.2)<sup>180</sup>.



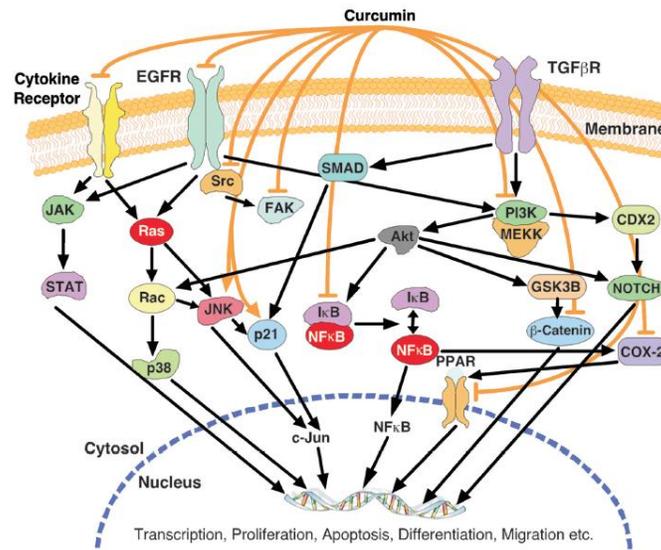
**Figure 5.2.** Cyclin-dependent pathway by which curcumin can alter the deregulated cell cycle of cancer cells. Figure from Sa and Das (2008)<sup>180</sup>.

Curcumin can also induce p53-dependent apoptosis in various cancer cell lines, as well as p53-dependent and independent G2/M phase arrest. p53 controls apoptotic pathways but is also a key cell cycle regulatory protein, since it can activate cell cycle inhibitors like Cip1 on the event of DNA damage during proliferation and when the damage is irreparable it induces apoptosis by inducing the expression of pro-apoptotic proteins like Bax (Figure 5.3). Curcumin induces p53 dramatically at G2 phase of cell cycle and enhances p53 DNA-binding activity resulting in apoptosis at this phase. But, in non-malignant cells curcumin reversibly up-regulates Cip1 expression and inactivates pRb, arresting them in G0 phase of cell cycle, without inducing apoptosis (Figure 5.2)<sup>180</sup>.



**Figure 5.3.** p53-dependent pathway by which curcumin can alter the deregulated cell cycle of cancer cells. Figure from Sa and Das (2008)<sup>180</sup>.

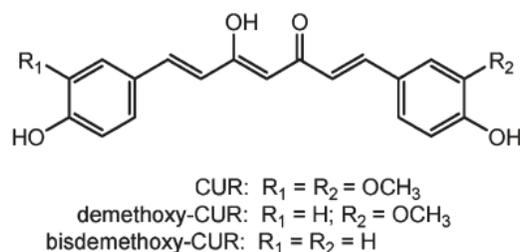
Curcumin is also able to block cell cycle progression or even induce apoptosis in a p53-independent manner (Figure 5.4), for example by activating caspase-8 and caspase-3 via Fas receptor aggregation, blocking NFκB cell survival pathway and suppressing the apoptotic inhibitor XIAP<sup>180</sup>.



**Figure 5.4.** p53-independent pathway by which curcumin can alter the deregulated cell cycle of cancer cells. Figure from Sa and Das (2008)<sup>180</sup>.

### 5.2.1.3. Chemical properties of curcumin

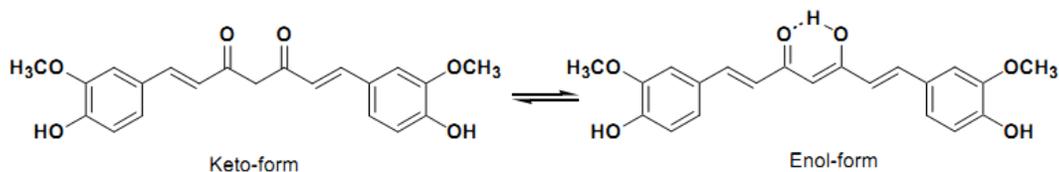
The alcoholic extract of turmeric and the commercially available curcumin contains a mixture of three curcuminoids, namely curcumin (CUR, 77%), and in smaller amounts, demethoxycurcumin (18%) and bisdemethoxycurcumin (5%)<sup>175,177</sup>. The chemical structures of these three major natural curcuminoids are represented in Figure 5.5. Studies have demonstrated that commercial grade curcumin has the same inhibitory effects in preclinical models of carcinogenesis, as pure curcumin<sup>182</sup>.



**Figure 5.5** Chemical structures of the three major natural curcuminoids. Figure from Metzler *et al.* (2012)<sup>175</sup>.

Curcumin, also known as (E,E)-1,7-bis(4 hydroxy-3-methoxyphenyl)-1,6-heptadiene-3,5 dione ( $C_{21}H_{20}O_6$ ) is a bis- $\alpha,\beta$ -unsaturated  $\beta$ -diketone, with a molecular weight of  $368.38 \text{ g}\cdot\text{mol}^{-1}$  and a melting point of  $179\text{--}183^\circ\text{C}$ <sup>177</sup>. The two aromatic rings containing phenolic groups are connected by two  $\alpha,\beta$ -unsaturated carbonyl groups<sup>176</sup>. Under physiological conditions, these carbonyl groups form a diketone which exists both in an enol and a bis-keto tautomeric form that coexist in equilibrium (Figure 5.6). The keto form predominates in acidic and neutral solutions as well as in the cell membrane, acting as a potent donor of H-atoms<sup>176,177</sup>. The enolic form predominates in alkaline conditions ( $\text{pH}>8$ ), and the phenolic part of the molecule plays the principal role as an electron donor, which is a very typical mechanism for the scavenging activity of phenolic antioxidants<sup>176,177,182</sup>.

Curcumin is hydrophobic and very soluble in dimethylsulfoxide (DMSO), ethanol or acetone, but its solubility in water and ether is very low<sup>177</sup>.



**Figure 5.6.** Keto- and enol-tautomeric forms of curcumin. Figure from Basnet and Skalko-Basnet (2011)<sup>177</sup>.

### 5.2.1.4. Chemical stability and degradation

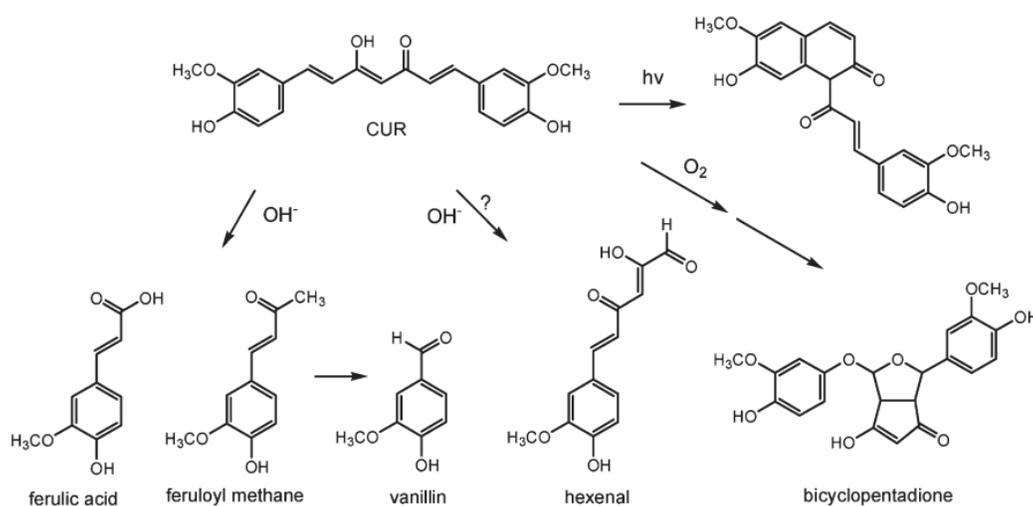
Curcumin is sensitive to alkaline pH, oxygen, and irradiation with ultraviolet and visible light, therefore it is not a chemically stable molecule. The decrease in the pH increases the half-life of curcumin, being reasonably stable at

pH < 6.5<sup>175</sup>. Addition of antioxidants such as ascorbic acid, N-acetylcysteine or glutathione to the culture media also inhibited its degradation, even at alkaline pH<sup>177</sup>. Demethoxy-CUR and bisdemethoxy-CUR are much more stable than curcumin in solution<sup>175</sup>.

When curcumin is degraded through alkaline hydrolysis, autoxidation or irradiation, several products can arise, that are represented in Figure 5.7. The alkaline degradation results in the formation of ferulic acid, feruloyl methane and vanillin. The *trans*-6-(4'-hydroxy-3'-methoxyphenyl)-2,4-dioxo-5-hexenal is a putative major degradation product, that still needs further confirmation<sup>175</sup>.

The main product of auto-oxidative transformation of curcumin is a bicyclopentadione derivative. The autoxidation phenomenon can account for the instability of curcumin dissolved in cell culture media or body fluids. Indeed, bicyclopentadiones are much more prevalent than ferulic acid and feruloylmethane as degradation products<sup>175</sup>.

Curcumin is also unstable when irradiated with light and one of the products identified after irradiation of curcumin solutions is represented. This should not be a problem for the stability of the molecule in biological systems, although it prevents its use in pharmaceutical coatings<sup>175</sup>.



**Figure 5.7.** Degradation products of curcumin through alkaline hydrolysis, autoxidation and irradiation. Figure from Metzler *et al.* (2012)<sup>175</sup>.

### 5.2.1.5. Uptake and metabolism

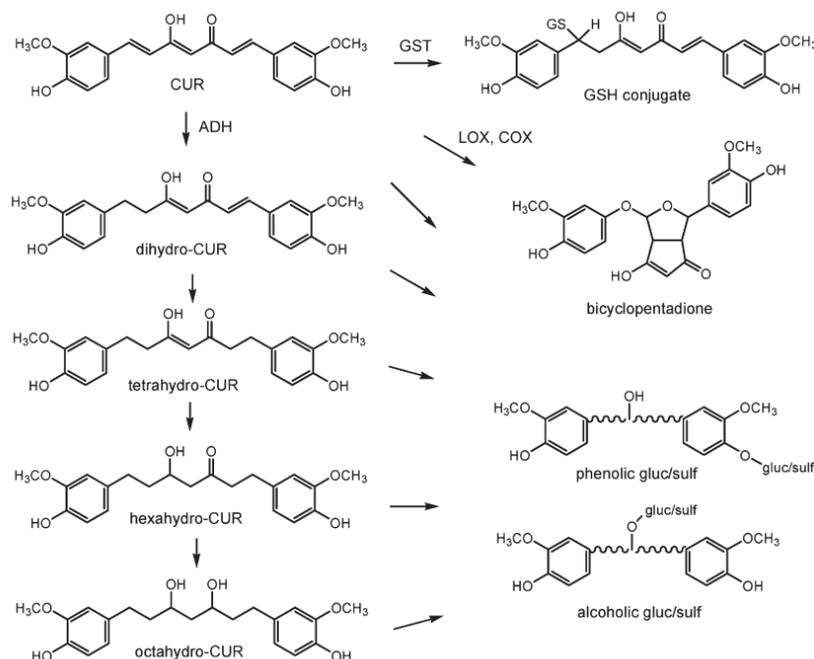
The information about the chemical structures of curcumin metabolites were first obtained from studies on the fate of tritium-labeled curcumin in the rat. After intravenous injection of <sup>3</sup>H-CUR, a mass spectrometry analysis identified tetrahydro-CUR and hexahydro-CUR as the major metabolites<sup>175</sup>. More recent information from *in vitro* studies<sup>183</sup> with hepatic and intestinal cells, and from clinical studies with cancer patients, identified dihydro-CUR and octahydro-CUR as minor metabolites. This allowed to unveil the phase I metabolism of curcumin (introduction of reactive and polar groups) (Figure 5.8), that comprises the successive reduction of the four double bonds of the heptadiene-3,5-dione system by enzymes (including alcohol dehydrogenases) in the cytosol of liver and intestinal cells<sup>175</sup>. Demethoxy-

CUR and bisdemethoxy-CUR undergo a very similar reductive metabolism, with the hexahydro product as being the major metabolite<sup>175</sup>.

Concerning the phase II metabolism (activated metabolites are conjugated with charged species) (Figure 5.8), curcumin and its reductive metabolites appear to be easily conjugated *in vivo*. There are reports of monoglucuronides, monosulfates, and mixed glucuronide/sulfates as conjugates. The predominating pathway of conjugation is glucuronidation and the major metabolite of curcumin found in body fluids, organs and cells is the glucuronide of hexahydro-CUR. In humans, the enzymes of the GI tract contribute substantially to the glucuronidation of curcuminoids. The formation of phenolic and alcoholic monoglucuronides is also possible, due to the fact that curcumin and its reduced metabolites they all have two aromatic (phenolic) and one aliphatic (alcoholic) hydroxyl groups. Until now, little is known about the exact chemical structures and the stability of sulfate conjugates of curcuminoids<sup>175</sup>.

Recently, a new type of conjugate was reported, with curcumin being conjugated with glutathione (GSH) by action of a glutathione S-transferase (GST) (Figure 5.8). This GSH conjugates are unstable and have a half-life of about 4 hours, being rapidly decomposed into CUR again and other unidentified degradation products<sup>184</sup>.

Bicyclopentadiones are not only degradation products, but can also be generated through enzymatic reaction by lipoxygenases (LOX) and cyclooxygenases (COX) (Figure 5.8), although little is known about its stability or metabolism<sup>175</sup>.



**Figure 5.8.** Main products of curcumin metabolism. Figure from Metzler *et al.* (2012)<sup>175</sup>.

The Joint FAO/WHO Expert Committee on Food Additives indicate as acceptable a daily intake level of 0.1–3 mg.kg<sup>-1</sup> body weight of curcumin. But studies with oral doses of curcumin up to 3,500 mg.kg<sup>-1</sup> body weight for up to 90 days in rats, dogs and monkeys, showed no adverse effects or toxicity. Even when administered 8,000 mg daily for three months in human cancer patients, no adverse effects were detected<sup>177</sup>.

A study in rats showed that curcumin is poorly absorbed from the GI tract after oral intake (the oral bioavailability of curcumin was estimated at about 1%) and mostly excreted unchanged with the feces. Even after high doses, the levels of curcumin in blood plasma, bile and urine are extremely low, probably due to its chemical and metabolic instability. Thus, any effects in peripheral tissues must be assumed to be mediated by degradation products and/or metabolites<sup>175</sup>. In short, curcumin has low oral bioavailability and may undergo rapid first-pass intestinal metabolism<sup>182</sup>.

Few pharmacokinetic data are available from human studies. For example, a study where up to 8,000 mg of curcumin was administered daily for three months to patients with high-risk premalignant conditions, showed that peak serum curcumin concentrations were around 1.75 μM achieved 1–2 hours after oral intake and gradually declined within 12 hours<sup>177</sup>. In a different study, when given 3,600 mg curcumin/day, urinary levels varied between 0.1 and 1.3 μM (curcumin), 19 and 45 nM (curcumin sulphate), and 210 and 510 nM (curcumin glucuronide)<sup>177</sup>, but the levels of drug and conjugates in plasma were near the limit of detection of the assays used<sup>182</sup>. The concentrations of curcumin in normal and malignant colorectal tissue were 12.7 and 7.7 nmol/g tissue, respectively<sup>177</sup>. In summary, it seems that curcumin achieves pharmacologically efficacious levels in colorectal tissue, but outside the gut its distribution is negligible<sup>182</sup>.

All these results showed that curcumin has low systemic bioavailability following oral dosing in humans and animals. This can be explained by the efficient intestinal metabolism of curcumin, particularly glucuronidation and sulphation<sup>182</sup>.

#### **5.2.1.6. The bioavailability problem and possible solutions**

It is well established that the main limitation of broader use of curcumin-based formulations is its poor solubility and fast metabolism<sup>177</sup>. After oral administration, the systemic bioavailability of curcumin is very low and several efforts have been made in order to increase its solubility, stability and pharmacological activities, such as improving formulations and delivery systems<sup>175,177</sup>.

One of the first approaches to enhance the bioavailability of curcumin was to increase its solubility, since water is the simplest and safest vehicle for delivery of drugs *in vivo*. For example, the use of heat can increase 12-fold the solubility of curcumin. The complexes of curcumin with metal ions (Zn<sup>2+</sup>, Cu<sup>2+</sup>, Mg<sup>2+</sup> and Se<sup>2+</sup>) were readily soluble in water-glycerol (1:1; w/w) and also quite stable to light and heat. The complexation with serum albumin also increased curcumin solubility<sup>177</sup>.

Concerning the delivery systems, curcumin can be formulated with adjuvants, nanoparticles, liposomes, and micelles<sup>175,177,185</sup>. In general, these delivery systems enhance stability, bioavailability and cellular uptake of curcumin, but none of them have yet reached clinical evaluation.

The co-administration of curcumin with other agents, such as piperine (an alkaloid found on black pepper) increased systemic bioavailability by as much as 154%, but the mechanism underlying this remains to be elucidated<sup>177,185</sup>.

Several attempts have also been made to prepare chemically modified curcumin in order to have curcumin derivatives or analogues with increased solubility and improved activity against cancer<sup>177,185</sup>.

In this work the aim is to engineer an artificial curcumin pathway<sup>186</sup> in LAB, belonging to the *L. lactis* LMG19460 strain, since gram-positive bacteria are efficient producers of heterologous proteins<sup>7</sup>.

LAB are able to survive through the GI tract of humans with a retention time of 2-3 days in gut<sup>5</sup>, which enables the oral administration of curcumin-producer LABs to GI cancer patients or for prevention purposes to the general population<sup>187</sup>.

Since curcumin has a slow cellular uptake and repetitive oral doses of the molecule are required to achieve an effective therapeutic concentration inside the cells<sup>185,188</sup>, the possibility of synthesizing it *in situ* by LAB represents a powerful and innovative solution for GI cancer therapy.

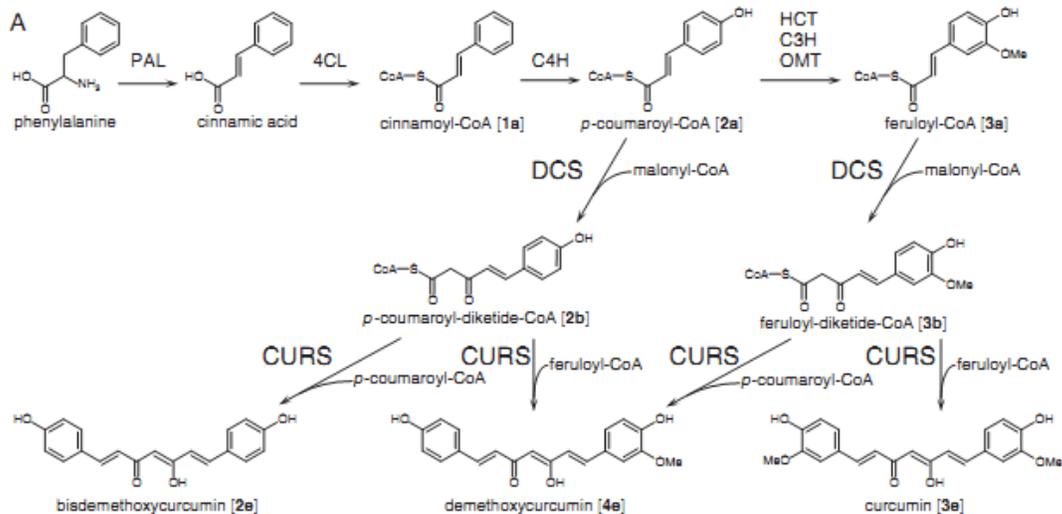
The biosynthetic pathways to produce curcumin will be discussed in the next section, starting with the natural pathway, followed by the artificial pathways already engineered in *E. coli* and the proposed artificial pathway.

## 5.2.2. Biosynthetic pathways to produce curcumin

### 5.2.2.1. Natural pathway

In *C. longa*, the biosynthesis of the three main curcuminoids is represented in Figure 5.9. The first step is catalyzed by phenylalanine ammonia-lyase (PAL) that synthesizes cinnamic acid from phenylalanine. The enzyme 4-coumarate:CoA ligase (4CL) is responsible for converting cinnamic acid in cinnamoyl-CoA. This compound can be further modified by cinnamate-4-hydroxylase (C4H) in p-coumaroyl-CoA, that in turn can originate feruloyl-CoA by the activity of hydroxycinnamoyl transferase (HCT), cinnamate-3-hydroxylase (C3H) and O-methyltransferase (OMT). Diketide-CoA synthase (DCS) catalyzes the formation of diketide-CoAs by condensing p-coumaroyl-CoA and feruloyl-CoA with malonyl-CoA. Curcumin synthase (CURS) catalyzes formation of curcuminoids (curcumin, demethoxycurcumin, bisdemethoxycurcumin) by condensing diketide-CoAs with starter substrates (p-coumaroyl-CoA and feruloyl-CoA)<sup>120</sup>.

DCS and CURS are type III polyketide synthases (PKS), which are compounds that usually play an important role in the biosynthesis of most plant polyketides. Both enzymes prefer feruloyl-CoA as a starter substrate, meaning that curcumin is probably synthesized via feruloyl-CoA<sup>120</sup>. When feruloyl-CoA and p-coumaroyl-CoA were present in the same reaction, demethoxycurcumin and a trace amount of bisdemethoxycurcumin were also produced, in addition to curcumin. But the concentrations of the starter substrates p-coumaroyl-CoA and feruloyl-CoA may not be equal *in vivo*, which can account for the difference of the product profile of the curcuminoids of the *in vitro* reaction from that of an ethyl acetate extract of the rhizome of turmeric<sup>120</sup>, with the rhizome of turmeric containing relatively larger amount of bisdemethoxycurcumin than the *in vitro* reaction<sup>189</sup>.



**Figure 5.9.** Curcuminoid biosynthesis in *C. longa*. Figure from Katsuyama *et al.* (2009)<sup>120</sup>.

CUS is also a type III PKS from rice (*Oriza sativa*)<sup>190</sup> that is able to catalyze both steps that are catalyzed by DCS and CURS. This means that the CUS system is simpler than the DCS/CURS system, making in theory CUS a better enzyme for the purpose of metabolic engineering of curcuminoids in microorganisms<sup>191</sup>. But CUS synthesizes curcuminoids accompanied by triketide pyrone formation, which is naturally avoided in turmeric by using DCS/CURS. DCS/CURS also has the advantage of an efficient regulation by allosteric regulation of DCS<sup>120</sup>. One dimer of DCS possesses two active sites. When the starter substrate binds to the active site in one subunit may affect the structure of the active site in the other subunit (homotropic activation) during the DCS reaction. Homotropic activation decreases activity of DCS as the substrate concentration becomes lower, leaving some feruloyl-CoA to be consumed by CURS<sup>120</sup>.

Two additional type III PKS, named CURS2 and CURS3, were already identified. *In vitro* analysis showed that CURS2 preferred feruloyl-CoA as a starter substrate, while CURS3 preferred both feruloyl-CoA and p-coumaroyl-CoA. This means that CURS2 can synthesize curcumin or demethoxycurcumin and CURS3 synthesizes the three curcuminoids. The first identified CURS is now called CURS1<sup>189</sup>.

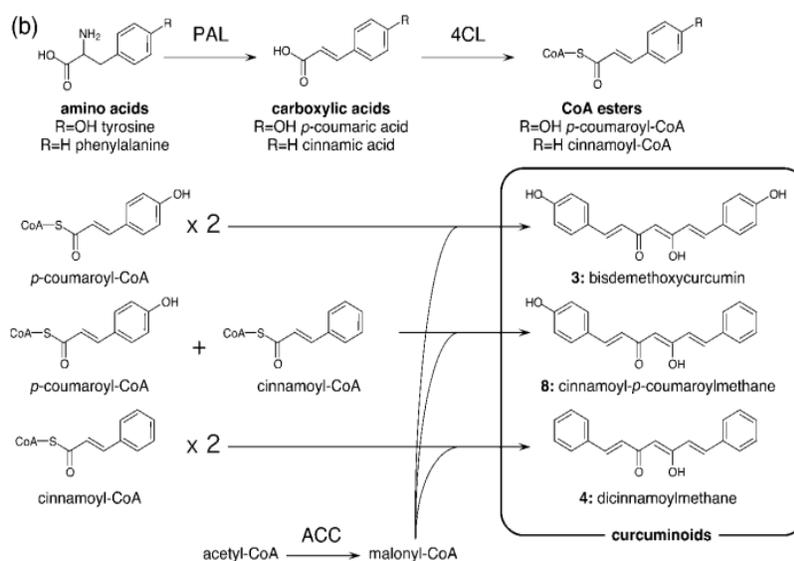
### 5.2.2.2. Artificial pathways

Recent advances in metabolic engineering, synthetic biology and systems biology have made it possible to produce plants specific polyketides by employing non-plant organisms as hosts<sup>119</sup>.

An artificial curcuminoid biosynthetic pathway was constructed in *E. coli* for the production of curcuminoids. This was the first study demonstrating the production of curcuminoids in an organism other than *Curcuma* species. *E. coli* harbored several plasmids with different genes important for the artificial pathway: genes *PAL* (phenylalanine ammonia-lyase from the yeast *Rhodotorula rubra*) and *LE4CL-1* (4-coumarate: CoA ligase from the plant

*Lithospermum erythrorhizon*) were on the plasmid pCDF-PAL/LE4CL-1, gene *CUS* (curcuminoid synthase from rice *O. sativa*) was on pET-CUS, and gene *ACC* (acetyl-CoA carboxylase from *Corynebacterium glutamicum*) on plasmid pRSF-ACC. All genes were under the control of an IPTG-inducible T7 promoter and a synthetic ribosome-binding sequence<sup>119</sup>.

First, by removing the amino group, PAL converts tyrosine and phenylalanine (amino acids) to the corresponding carboxylic acids, p-coumaric acid and cinnamic acid, respectively, which are then activated to CoA esters by 4CL. After that, it is necessary a polyketide synthesis step for condensation of the CoA esters with malonyl-CoA yielding curcuminoids (bisdemethoxycurcumin, dicinnamoylmethane and cinnamoyl-p-coumaroylmethane) performed by CUS. ACC catalyses conversion of acetyl-CoA into malonyl-CoA and is maintained in the cells to increase the intracellular pool of malonyl-CoA (Figure 5.10)<sup>119</sup>.

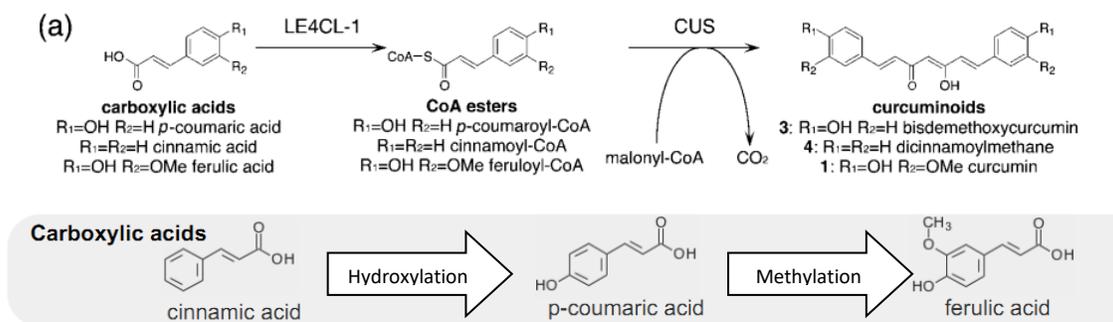


**Figure 5.10.** Artificial curcuminoid biosynthetic pathway. Figure from Katsuyama *et al.* (2008)<sup>119</sup>.

The results showed that the deamination catalyzed by PAL was the rate-limiting step for producing p-coumaroyl-CoA in high yield, and not the reaction catalyzed by LE4CL-1. Therefore, the addition of amino acids to the culture caused no acceleration of the reaction generating p-coumaroyl-CoA by PAL and 4CL<sup>119</sup>.

Another *E. coli* system carrying *4CL* and *CUS* genes was also used for high-yield production of curcuminoids (bisdemethoxycurcumin, dicinnamoylmethane and curcumin) from exogenously supplemented carboxylic acids: p-coumaric acid, cinnamic acid and ferulic acid. When carboxylic acids were directly supplied to this system carrying *4CL* and *CUS* genes (and also *ACC*), the yields of curcuminoids produced reached 90–110 mg.l<sup>-1</sup>, at a molar yield of approximately 60%. What happens was that the removal of the PAL step increased the p-coumaroyl-CoA concentration in *E. coli*<sup>119</sup>. But, the direct supply of each carboxylic acid only produced the respective curcuminoid, meaning that the hydroxylation and methylation steps are not catalyzed in *E. coli* (Figure 5.11)<sup>119</sup>.

In 2010 some of the authors of the previous paper studied the precursor-directed biosynthesis of curcumin analogs by using *E. coli* harboring *4CL*, *CUS* and *ACC*. By using this artificial pathway and feeding various analogs of p-coumaric acid to *E. coli* they succeeded in the production of 15 unnatural curcuminoids<sup>191</sup>.

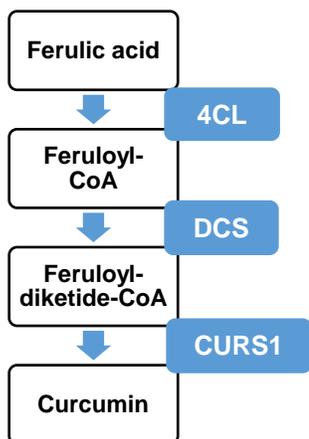


**Figure 5.11.** Second artificial curcuminoid biosynthetic pathway and representation of the hydroxylation and methylation steps of carboxylic acids. Figure from Katsuyama *et al.* (2008)<sup>119</sup>.

### 5.2.2.3. The artificial pathway proposed in the present PhD

Considering the previous results, the minimum synthetic pathway to produce curcumin that is proposed in the present PhD consists in *4CL*<sup>119</sup>, *DCS* and *CURS1* genes<sup>120</sup>. The sequence of the first gene can be obtained from *L. erythrorhizon*, while the other two can be obtained from *C. longa*. Since this is a minimal pathway, in order to obtain curcumin it is required that ferulic acid is supplied to the bacterial growth media. Ferulic acid is the substrate for the enzyme expressed by the *4CL* gene. The resulting product is feruloyl-CoA, which will further act as substrate for the *DCS* enzyme. This reaction generates feruloyl-diketide-CoA, which is used as substrate by *CURS1* enzyme to produce curcumin (Figure 5.12). Ferulic acid is a phytochemical and a component of several herbal health supplements, being known by its antioxidant and therapeutic effects against cancer, thus being a powerful ally to curcumin<sup>192</sup>.

A very similar pathway was already tested in *E. coli* K12 MG1655, with each gene cloned in different plasmids and using the *4CL* gene from *Arabidopsis thaliana*. They obtained 70 mg/L of curcumin after supplementation with ferulic acid<sup>193</sup>.



**Figure 5.12.** Artificial pathway proposed.

### 5.2.3. LAB as cell factories to produce heterologous proteins

LAB are gram positive bacteria that include *lactococci*, *streptococci*, and *lactobacilli*. All of these bacteria are widely used as starters for food and beverages fermentations. The GRAS status of *L. lactis* and the remaining LAB, together with the fact that they are non-pathogenic, non-invasive and safe, is important when the goal is to use them as cell factories for production of therapeutic recombinant proteins or DNA vaccines<sup>5,6</sup>.

They constitute an excellent alternative to attenuated pathogens, which until now have been the most popular live vectors being used due to its ability to infect naturally the mucosal surfaces. But these attenuated bacteria can recover their pathogenic potential and therefore they are considered not entirely safe for use in humans, especially in children and immunosuppressed patients<sup>6</sup>.

LAB is able to survive through the GI tract of humans, with a retention time of 2-3 days in gut, without invading or colonizing the mucous, and without evoking strong host immune responses<sup>5</sup>.

Among prokaryotic systems, using *E. coli* as the cell factory as resulted in the highest protein levels. However LAB can easily secrete proteins into the extracellular medium, unlike *E. coli* for which the most commonly used production strategies are intracellular (periplasm or cytoplasm), thus leading to expensive downstream purification processes. Additionally, it is well-known that *E. coli* contains highly immunogenic lipopolysaccharides (LPS) that may be co-purified with the proteins of interest and should be removed before being administered to humans. On the other hand, LAB do not produce LPS, thus making the process more profitable and safe<sup>7</sup>.

*Bacillus subtilis* is the model gram positive bacterium and have been used to produce several enzymes, being endotoxin free and considered a GRAS microorganism. Its main problem is that many heterologous proteins secreted by these bacteria are degraded by its complex extracellular proteolytic system<sup>7</sup>.

*L. lactis* is an attractive alternative to *B. subtilis*, being widely used in the dairy industry<sup>5,7</sup>. One advantage of *L. lactis* when compared with *B. subtilis* is the fact that only one major protein, Usp45, is secreted into the medium, which simplifies the downstream purification processes; the laboratory strains only have one exported housekeeping protease, HtrA, and a protease-free mutant is already available; and the growth of *L. lactis* under fermentation conditions allows an easy scale-up<sup>7</sup>.

Many heterologous proteins have already been produced in *L. lactis*, such as reporter molecules, bacterial, eukaryotic and viral antigens, interleukins, allergens, virulence factors, bacteriocins and enzymes<sup>6,194,195</sup>. LAB can also be used as DNA delivery vehicles for DNA vaccination and several studies in this field have already been performed<sup>6,21</sup>.

#### 5.2.3.1. Metabolite secretion

In short, for curcumin to be produced inside the LAB, the substrate (ferulic acid) has to be translocated from the environment to the cytoplasm. It is known that ferulic acid can be uptaken by bacteria, since *E. coli* incubated with this substrate in the growth media can produce curcuminoids<sup>177</sup>. Then, the substrate has to be converted into products

by the enzymes coded by the 3 heterologous genes, until the final metabolite, which is curcumin. In order to perform its biological activity, this metabolite has to be excreted to the exterior of the bacteria to reach the gastrointestinal cells.

There are several mechanisms by which bacteria can release metabolites to the environment. Attempts have been made to increase the rate of product secretion through manipulation of the physical properties of the cytoplasmic membrane, since it is the main barrier that metabolites have to cross<sup>196</sup>. There is almost no information about how curcumin can be secreted by bacteria, so the several possible systems known for other molecules will be considered.

The passive transport is the simplest mechanism, since solutes cross the membrane without the involvement of a specific transport protein and the driving force is given by the chemical gradient of the metabolite. It is known that hydrophobic molecules, like benzene are able to diffuse across the plasma membrane<sup>197</sup>. Since curcumin is hydrophobic, when the concentration inside the cell increases it can diffuse through the phospholipid bilayer, from the cytoplasm where it is in high concentration to the extracellular environment where the concentration of the molecule is lower.

But, since it is a relatively large molecule, other types of transport can be involved. Facilitated diffusion involves the movement of molecules in the direction determined by their relative concentrations inside and outside of the cell, without an external source of energy, but its passage is mediated by proteins (carrier proteins and channel proteins)<sup>197</sup>. In active transport, energy provided by another coupled reaction (such as the hydrolysis of ATP) is used to drive the transport of molecules in an energetically unfavorable direction. The most prominent examples of this type of transport are the ion pumps and the ABC transporters. But other molecules can be transported against their concentration gradients using energy derived from the coupled transport of a second molecule in the energetically favorable direction<sup>197</sup>. In principle, all these transport systems can function in both directions, playing a role in the secretion of metabolites<sup>196</sup>.

For example, many secondary metabolites (e.g. antibiotics and mycotoxins) are toxic to the microorganisms that produce them and can be exported by two different types of efflux systems: ATP-binding cassette transporters and major facilitator superfamily exporters<sup>198</sup>, and the same can happen in the case of curcumin.

## 5.3. Material and methods

### 5.3.1. Bacterial strains and plasmids

*E. coli* DH5 $\alpha$  (Invitrogen) was used for all the molecular cloning steps performed in this chapter and the resulting plasmids were transferred for *L. lactis* LMG19460 by electroporation (BCCM Culture Collection, Belgium). The plasmid employed as the main backbone was the previously characterized pTRKH3. *E. coli* K12 MG1655 (DE3) carrying pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1 was used as positive control of the curcumin production.

The characteristics of all the strains and plasmids used are summarized in table 5.1.

**Table 5.1.** Main characteristics of the strains and plasmids used in this chapter.

Strains	Relevant genotype	Source
<i>E. coli</i> DH5 $\alpha$	F <sup>-</sup> $\Phi$ 80 <i>lacZ</i> $\Delta$ M15 $\Delta$ ( <i>lacZYA-argF</i> ) U169 <i>recA1 endA1 hsdR17</i> (rK <sup>-</sup> , mK <sup>+</sup> ) <i>phoA supE44 <math>\lambda</math>-thi-1 gyrA96 relA1</i>	Invitrogen
<i>E. coli</i> K12 MG1655 (DE3)	F <sup>-</sup> $\lambda$ <sup>-</sup> <i>ilvG<sup>-</sup> rfb-50 rph-1</i> $\lambda$ (DE3)	<sup>199</sup>
<i>L. lactis</i> LMG19460	Plasmid-free strain; isolated from starter cultures of German cheese factories	LMG/BCCM
Plasmids	Construct	Source
pTRKH3	P15A <i>ori</i> , pAM $\beta$ 1 <i>ori</i> , Tet <sup>R</sup> , Ery <sup>R</sup>	LMBP/BCCM (LMBP 4462)
pAC-4CL1	P15A <i>ori</i> , <i>Plac</i> , Cm <sup>R</sup> , pACYC184-derived plasmid carrying 4CL1 from <i>Arabidopsis thaliana</i>	Addgene (35947)
pCDFDuet_DCS	CloDF13 <i>ori</i> , <i>lacI</i> , double T7 <i>lac</i> , Strep <sup>R</sup> ; pCDFDuet-1 carrying codon-optimized DCS from <i>Curcuma longa</i>	<sup>193</sup>
pRSFDuet_CURS1	RSF <i>ori</i> , <i>lacI</i> , double T7 <i>lac</i> , Kan <sup>R</sup> ; pRSFDuet-1 carrying codon-optimized CURS1 from <i>C. longa</i>	<sup>193</sup>
pUC57_4CL	pMB1 <i>ori</i> , Amp <sup>R</sup> ; pUC57 carrying the codon-optimized 4CL	Nzytech
pUC57_DCS	pMB1 <i>ori</i> , Amp <sup>R</sup> ; pUC57 carrying the codon-optimized DCS	Nzytech
pUC57_CURS1	pMB1 <i>ori</i> , Amp <sup>R</sup> ; pUC57 carrying the codon-optimized CURS	Nzytech
pTRKH3_SLP_4CL	pTRKH3 carrying the codon-optimized 4CL under the control of the SLP promoter	This study
pTRKH3_SLP_DCS	pTRKH3 carrying the codon-optimized DCS under the control of the SLP promoter	This study
pTRKH3_SLP_CURS1	pTRKH3 carrying the codon-optimized CURS1 under the control of the SLP promoter	This study
pTRKH3_SLP_4CL/DCS/CURS1	pTRKH3 carrying the codon-optimized 4CL, DCS and CURS1 under the control of the SLP promoter	This study
pTRKH3_LDH_4CL/DCS/CURS1	pTRKH3 carrying the codon-optimized 4CL, DCS and CURS1 under the control of the LDH promoter	This study

### 5.3.2. Media and growth conditions

The original JM110 strain harboring the pTRKH3 plasmid was grown accordingly with the instructions from the culture collection (Brain Heart Infusion medium supplemented with 10  $\mu$ g/mL tetracycline and 150  $\mu$ g/mL erythromycin, 37°C, 250 rpm).

20 g/L of Luria-Bertani (LB) broth (Nzytech) supplemented with 500  $\mu$ g/mL erythromycin (Sigma) was used for growing *E. coli* DH5 $\alpha$  harboring the different constructs, at 37°C, 250 rpm. The positive control *E. coli* K12 MG1655 (DE3) harboring the pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1 plasmids was grown in 20 g/L LB supplemented with 30  $\mu$ g/mL of chloramphenicol (Sigma-Aldrich), 100  $\mu$ g/mL of spectinomycin (Sigma-Aldrich) and 30  $\mu$ g/mL of kanamycin (Sigma-Aldrich), at 37°C, 250 rpm.

*L. lactis* LMG19460 were grown at 30°C, 100 rpm, using M-17 (pH 7.0, Fluka) supplemented with 20 g/L glucose, and 5 µg/mL erythromycin (Sigma) when transformed with one or several plasmids.

When necessary, *E. coli* cells were grown in LB agar (2%) plates with the appropriate antibiotic, at 37°C, while *L. lactis* cells were grown in solid regeneration medium<sup>112</sup> containing (per liter) 10 g of tryptone, 5 g of yeast extract, 200 g of sucrose, 10 g of glucose, 25 g of gelatin, 15 g of agar, 2.5 mM of MgCl<sub>2</sub>·6H<sub>2</sub>O, 2.5 mM of CaCl<sub>2</sub> and 5 mg erythromycin, at 30°C.

Master and working cell banks of all the strains were made with a final glycerol concentration of 20%, and stored at -80°C.

### 5.3.3. Competent cells and transformation

#### 5.3.3.1. *E. coli* DH5α

Chemically competent *E. coli* DH5α cells were prepared using cells grown until reach an OD<sub>600nm</sub> of 1, in a 100 mL shake flask with 20 mL LB, at 37°C and 250 rpm (initial OD<sub>600nm</sub> =0.1, from an overnight pre-inoculum). Freshly grown cells were centrifuged (1,000 g, 10 min, 4°C) and resuspended in 0.1 volumes (2 mL) of chilled TSS medium (20 g/L LB, 5% DMSO, 50 mM MgCl<sub>2</sub>, 10% PEG 8000 (w/v), pH=6.5). After 10 min on ice, 100 µL aliquots were made and stored at -80°C.

In order to transform DH5α cells with pDNA by heat shock, for every 100 µL of chemically competent cells, the appropriate amount of pDNA was added, accordingly with each experiment. The mixture was incubated for 30 min on ice, and then submitted to 42°C for 60 s. The cells were immediately incubated for 2 min on ice, and then resuspended with 900 µL of LB medium. After 1 hour of recuperation at the correct temperature, the cells were plated on LB agar with the appropriate antibiotic, accordingly with the experiment, and incubated overnight at the correct temperature.

When necessary, pDNA was purified from a fixed volume of cells using High Pure Plasmid Isolation Kit (Roche), its concentration measured using a Nanodrop Spectrophotometer (Nanovue Plus, GE) and its quality visualized in a 1% agarose gel.

#### 5.3.3.2. *L. lactis* LMG19460

The procedure for making electrocompetent *L. lactis* LMG19460 were adapted and optimized from Holo & Nes (1989) protocol<sup>112</sup>. An overnight growth in 5 mL of M-17 supplemented with 0.5% glucose (GM-17), 30°C, 100 rpm, is used to inoculate a 100 mL shake flask with 75 mL of GM-17 with a starting OD of 0.1. After growing at 30°C, 100 rpm until reaching an OD<sub>600nm</sub>=0.5-0.8, the whole broth is diluted 100-fold to fresh GM-17 medium (750 µL to 75 mL) supplemented with 0.5M sucrose (SGM-17) and 1-2% glycine. The cells were grown at 30°C, 100 rpm until reach an OD<sub>600nm</sub>=2-2.5 and harvested by centrifugation at 6,000g, 3 min, 4°C. Next, the cells were washed 2 times in an ice cold solution containing 0.5 M sucrose and 10% glycerol (washing solution). Finally, they were resuspended in 1/100 of the initial culture volume of the washing solution and stored at -80°C.

For transformation of the electrocompetent *L. lactis* LMG19460 cells, 40  $\mu$ L aliquots containing approximately  $1 \times 10^9$  cells were mixed with the appropriate volume of pDNA and transferred to a 2 mm electroporation cuvette. After 30 min on ice, 2-3 pulses of 1000 V were given using an electroporator (BTX ECM399), resulting in time constants of 4 to 5 ms. To recover from the electroporation, 960  $\mu$ L of recovery medium was immediately added. The recovery medium is composed by ice-cold SGM-17 supplemented with 20 mM  $MgCl_2$  and 2 mM  $CaCl_2$ . After resting 5 min on ice, the mixture was incubated 3h at 30°C without agitation. Then, the cells were collected by centrifugation and used to inoculate 5 mL of M-17 supplemented with 20 g/L glucose and a sub-lethal erythromycin concentration<sup>113</sup> (2.5  $\mu$ g/mL), at 30°C without agitation. After an overnight growth, the cells were again collected by centrifugation and plated in regeneration medium<sup>112</sup> containing (per liter) 10 g of tryptone, 5 g of yeast extract, 200 g of sucrose, 10 g of glucose, 25 g of gelatin, 15 g of agar, 2.5 mM of  $MgCl_2 \cdot 6H_2O$ , 2.5 mM of  $CaCl_2$  and 5 mg erythromycin, at 30°C.

When necessary, pDNA was purified from a fixed volume of cells using Nucleospin Plasmid Kit (Macherey-Nagel), its concentration measured using a Nanodrop Spectrophotometer (Nanovue Plus, GE) and its quality visualized in a 1% agarose gel.

### 5.3.4. Cloning strategy of the curcumin pathway genes

#### 5.3.4.1. Curcumin pathway *in silico* design and codon optimization

The chosen minimum curcumin pathway comprises 3 genes: *4CL* (4-coumarate: CoA ligase)<sup>119</sup>, *DCS* (diketide-CoA synthase) and *CURSI* (curcumin synthase)<sup>120</sup>. The sequence of the first gene was obtained from *Lithospermum erythrorhizon* (Accession number D49366.1), while the other two genes were from *Curcuma longa* (Accession number AB495006.1 for *DCS* and AB495007.1 for *CURSI*). When using these genes it's necessary to supply ferulic acid to the bacterial growth media, to work as the substrate for the enzyme expressed by the *4CL* gene. The product that results from this is feruloyl-CoA, which in turn acts as the substrate of the enzyme expressed by the *DCS* gene. This reaction generates feruloyl-diketide-CoA, where the enzyme expressed by the *CURSI* gene acts to produce curcumin.

Genes had been *in silico* designed with several unique restriction enzyme sites on both ends, in order to allow cloning each gene individually in different plasmids, or all the genes sequentially in the same plasmid. The designed sequences didn't have a stop codon, in order to allow the transcription/translation of a fusion protein if necessary. It was also taken into account that the genes should be cloned in the correct orientation, in frame with the start codon (ATG) that already existed and with each other when arranged as operon, and that the genes of the same operon shouldn't be separated by more than 20 bases. The plasmid design and *in silico* cloning was performed using APE<sup>114</sup> and SnapGene<sup>115</sup> softwares.

In the end, codon optimization for bacteria (*E. coli* K12) was performed in all genes, using Java Codon Optimization Tool<sup>200</sup>, in order to putatively increase the expression rate. After codon optimization, the genes were synthesized and cloned in the EcoRV site of different pUC57 plasmids, by Nzytech (Table 5.4).

#### 5.3.4.2. Molecular cloning of the curcumin pathway genes in pTRKH3 under the control of a constitutive promoter

The first step in the molecular cloning procedure was to insert a constitutive promoter in pTRKH3. The first alternative promoter is the surface (S)-layer protein (SLP) promoter, which was amplified from the genome of *Lactobacillus acidophilus* ATCC4356, using Ampl\_SLP\_F/R primers (Table 5.2). The second alternative constitutive promoter is the lactate dehydrogenase (LDH) promoter that was amplified also from the gDNA of *L. acidophilus* ATCC4356, using Ampl\_LDH\_F/R primers (Table 5.2). LDH was described as being a highly efficient promoter<sup>201</sup>, while SLP is considered as being responsible for a high level of transcription<sup>202</sup>.

The genomic DNA was purified after growing the cells overnight in 5 mL of MRS, at 37°C, without agitation, using the Wizard Genomic DNA Purification Kit (Promega, Madison, USA). The gDNA concentration was assessed using a Nanodrop Spectrophotometer (Nanovue Plus, GE) and its quality visualized in a 1% agarose gel. The PCR amplification reaction was prepared using the KOD Hot Start DNA Polymerase kit (Novagen), by mixing 200 ng of template gDNA, 0.02U/μL of KOD DNA Polymerase, 1 mM of MgSO<sub>4</sub>, 1x buffer, 0.2 mM of dNTPs, 0.3 μM of each primer and completed with PCR-grade water to a final volume of 25 μL. The cycling conditions were: initial denaturation for 2 min at 95°C, followed by 35 cycles of 1 min at 95°C, 1 min at 58°C (for SLP promoter) or 55°C (for LDH promoter) and 1 min at 68°C. After amplification, the PCR products were run in a 1% agarose gel in order to retrieve the specific band without UV or ethidium bromide exposure, to keep the DNA quality and integrity. The agarose gel fragment was purified using NZYGelPure (Nzytech) and its concentration quantified using a Nanodrop Spectrophotometer (Nanovue Plus, GE).

A total of 3 μg of both the amplified inserts as well as the cloning vector (pTRKH3) were digested using appropriate restriction enzymes (EcoRV and BamHI, from Promega, Madison, USA). After a second gel purification step, the samples were concentrated using a DNA SpeedVac Concentrator. The ligation between the vector and insert was prepared with 2 μL of T4 DNA ligase 3 u.μL<sup>-1</sup> (Promega) and T4 ligase buffer (300 mM Tris-HCl (pH 7.8), 100 mM MgCl<sub>2</sub>, 100 mM DTT, 10 mM ATP) diluted 10 times in a final volume of 20 μL. Therefore, for the reaction mixture, a certain quantity of vector and insert were mixed maintaining a specific mass of vector, considering a 1:1 or 3:1 insert/vector molar ratio, according to the equation:

$$ng\ of\ insert = \frac{ng\ of\ vector \times kb\ size\ of\ insert}{kb\ size\ of\ vector} \times molar\ ratio\ of\ \frac{insert}{vector}$$

The ligation mixtures were incubated for 3 hours at room temperature and overnight at 4°C. After the first 3 hours of incubation, chemically competent *E.coli* DH5α cells were transformed with 10 μL of the total mix volume, and the remaining 10 μL of ligation mixture were incubated at 4°C overnight, and used to transform another aliquot of *E.coli* DH5α cells by heat shock.

Several colonies were picked from the plates and grown in liquid medium, in order to purify pDNA using High Pure Plasmid Isolation Kit (Roche) and test if the cloning was successful, using the appropriate restriction enzymes. In order to further confirm the presence of the desired clones, each plasmid was sent to Stabvida to be sequenced using the Check\_SLP\_F and Check\_LDH\_F primers (Table 5.2).

After the promoters being cloned in pTRKH3, the next step was to clone the three genes from the curcumin pathway, individually in each plasmid, as well as sequentially as an operon. The genes *4CL*, *DCS* and *CURS1* were amplified from the pUC57 plasmids harboring each codon-optimized gene by mixing the following: 10 ng of template pDNA, 0.02U/ $\mu$ L of KOD DNA Polymerase, 2.3 mM of MgSO<sub>4</sub>, 1x buffer, 0.2 mM of dNTPs, 0.5  $\mu$ M of each primer (Ampl\_4CL\_F/R, Ampl\_DCS\_F/R or Ampl\_CURS\_F/R; Table 5.2) and completed with PCR-grade water to a final volume of 25  $\mu$ L. The cycling conditions were: initial denaturation for 3 min at 95°C, followed by 30 cycles of 1 min at 95°C, 1 min at 65°C and 3 min at 70°C. After the PCR reaction, the parental DNA was digested using 2  $\mu$ L of DpnI, for 2h at 37°C, followed by purification with NZYGelPure (Nzytech). The three amplified fragments, together with the vectors pTRKH3\_SLP and pTRKH3\_LDH, were digested (3  $\mu$ g) with the appropriate restriction enzymes (BamHI and Sall, Promega, Madison, USA) and the correct bands purified from the gel. The remaining concentration, ligation and clone testing procedures were performed as described above. Each construct was finally confirmed by sequencing (Stabvida), using the Check\_pTRKH3\_F, Check\_DCS\_F and Check\_CURS\_F primers (Table 5.2).

After confirmation by sequencing, all the constructs were used to transform *L. lactis* LMG19460 by electroporation, using the protocol described above.

**Table 5.2.** Sequence, expected amplicon size and melting temperature of the primers used for promoters and genes amplification, as well as for confirmation sequencing procedures.

Primer name		Sequence (5' → 3') <sup>1</sup>	Product size (bp)	T <sub>m</sub> (°C)
Ampl_SLP_F	F	gagagagatcGTGGTAAGTAATAGGACGTGC	305	
Ampl_SLP_R	R	agagaggatccGCTAACAGTAGATACAGC		
Ampl_LDH_F	F	gagagagatcTTTAGTCCAATGCCCTTC	296	
Ampl_LDH_R	R	agagaggatccAAGTCTCCTTTTATTAGTG		
Ampl_4CL_F	F	TAATAAAAAAGGAGACTTggatccATGGACACCCAGACAAA	1959	73
Ampl_4CL_R	R	atatatgctgacttcctagcctaGTAGTTGTGAACACCGTTAG		63.3
Ampl_DCS_F	F	atatatggatcccctaggATGGAAGCTAACGGTTACCG	1214	64.3
Ampl_DCS_R	R	atatatgctgactttggcgcgcctaGTAGTTCAGACGGCAAGAGT		68.9
Ampl_CURS_F	F	atatatggatccggcgcgccaATGGCTAACCTGCACGCTCT	1206	71.2
Ampl_CURS_R	R	atatatgctgacctaGTACAGCGCATAGAACGCA		63.6
Check_SLP_F	F	CTGAACCTATGGCCTATTAC		49.76
Check_LDH_F	F	CCAAACCTGCAATTATCTTC		49.29
Check_pTRKH3_F	F	AAATTGCTAACGCAGTCAGG		53.5
Check_pTRKH3_R	R	CATACCCACGCCGAAACAAGC		
Check_DCS_F	F	CGACGGTGACTACGGTGTG		58
Check_CURS_F	F	CCTGGACGAAATGCGTAACC		56.27

<sup>1</sup> F and R indicate forward and reverse primers, respectively.

### 5.3.5. Curcumin production

For curcumin production, an overnight pre-inoculum of *L. lactis* LMG19460 with the plasmid(s) of interest was grown with 5 mL M-17 supplemented with 20 g/L glucose and 5  $\mu$ g/mL erythromycin, at 30°C, 100 rpm. 750  $\mu$ L of the pre-inoculum was used to start a culture in 100 mL shake flasks containing 75 mL of M-17 supplemented with 20 g/l of

glucose and 5 µg/ml of erythromycin, at 25°C, 100 rpm. The inoculum was grown until reach an optical density of 0.1 (1 OD unit at 600 nm equivalent to  $7 \times 10^8$  cells/ml)<sup>80</sup> and the cell pellet transferred to new shake flasks containing 75 mL of modified M9 minimal salt medium<sup>193</sup> containing 40 g/L glucose, 6 g/L Na<sub>2</sub>HPO<sub>4</sub> (Panreac), 3 g/L KH<sub>2</sub>PO<sub>4</sub> (Panreac), 1 g/L NH<sub>4</sub>Cl (Chem-Lab), 0.5 g/L NaCl (Panreac), 17 mg/L CaCl<sub>2</sub>, 50 mg/L MgSO<sub>4</sub> (Labchem) and 340 mg/L thiamine (Sigma-Aldrich), and 5 g/L of CaCO<sub>3</sub> (Merck) where applied. The medium was also supplemented with 5 µg/ml of erythromycin, 2 mM of ferulic acid and 1 mL/L trace element solution, containing 27 g/L FeCl<sub>3</sub> (Sigma-Aldrich), 2 g/L ZnCl<sub>2</sub> (Sigma-Aldrich), 2 g/L CoCl<sub>3</sub>\*6H<sub>2</sub>O (Sigma-Aldrich), 2 g/L Na<sub>2</sub>MoO<sub>4</sub>\*2H<sub>2</sub>O (Sigma-Aldrich), 1 g/L CaCl<sub>2</sub>\*2H<sub>2</sub>O (Sigma-Aldrich), 1.3 g/L CuCl<sub>2</sub>\*2H<sub>2</sub>O (Sigma-Aldrich), 0.5 g/L N<sub>3</sub>BO<sub>3</sub> (Sigma-Aldrich) and 100 mL/L HCl 1.2 M. The growth conditions were the same as the inoculum (25°C, 100 rpm) and were maintained for 7 days. Samples for OD, HPLC analysis (1 mL of whole broth with cells) and pDNA purification (OD of 20) were collected at 0h, 6h and each 24h during the 7 days' growth.

The positive control *E. coli* K12 MG1655 (DE3) harboring the pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1 plasmids was grown overnight in 20 g/L LB supplemented with 30 µg/mL of chloramphenicol, 100 µg/mL of spectinomycin and 30 µg/mL of kanamycin, at 37°C, 250 rpm. 750 µL of the pre-inoculum was used to start a culture in 100 mL shake flasks containing 50 mL of LB supplemented with the appropriate antibiotics, at 25°C, 100 rpm. Again, the inoculum was grown until reach an optical density of 0.1 and the cell pellet transferred to new shake flasks containing 50 mL of modified M9 minimal salt medium. Compared with the M9 minimal salt medium described for *L. lactis* LMG19460, the only difference was in the antibiotics used and in the necessity of supplementing with 1 mM IPTG. The remaining steps were as reported for *L. lactis* LMG19460.

### 5.3.6. Curcumin extraction

The pH of the 1 mL samples of whole broth with cells collected at different time points was adjusted to 3.0, using HCl 5M. An equal volume (1 mL) of ethyl acetate was added to the whole broth for ferulic acid and curcumin extraction. After 1 min of homogenization using a vortex, the aqueous and organic phases were separated by centrifugation at 17,000g, 5 min. The upper organic phase was collected to a new 1.5 mL tube and the extracts were concentrated by solvent evaporation in a fume hood. The concentrated extracts were resuspended in 200 µL of acetonitrile (Fisher Scientific, Loughborough, UK) and centrifuged 6,000g for 3 min to separate possible debris. 150 µL of the extract was collected for the HPLC vials inside appropriate microinserts.

### 5.3.7. HPLC analysis

The quantification of the ferulic acid and curcumin present in the samples was performed using high-performance liquid chromatography (HPLC) analysis, using a ELITE La Chrom system (VWR Hitachi Autosampler L-2200, VWR Hitachi Diode Array Detector L-2456 and Merck Hitachi UV-Detector L-2400) and a Macherey-Nagel Nucleosil C18 column (5 µm, 250 mm x 4 mm, 120 Å).

### 5.3.7.1. Standards preparation

Ferulic acid was purchased from Acros Organics (Geel, Belgium) and curcumin from Fisher Scientific (Loughborough, UK).

Serial dilutions of the ferulic acid 10 mg/mL stock solution were prepared with acetonitrile (200, 100, 50, 25, 12.5, 6.25, 3.125, 1.5625, 0.78125, 0.390625 and 0.195313  $\mu\text{g/mL}$ ) and analyzed in the HPLC. After calculating the area (and the corresponding retention time), this parameter was plotted against the concentration to obtain the ferulic acid calibration curve and its equation.

The same was performed to obtain a curcumin calibration curve, using the following standards diluted with acetonitrile: 400, 200, 100, 50, 25, 12.5, 6.25, 3.125, 1.5625, 0.78125, 0.390625, 0.195313 and 0.097657  $\mu\text{g/mL}$ .

### 5.3.7.2. Optimization of the analysis parameters

The HPLC methods for analyzing ferulic acid and curcumin were first optimized in order to have a simpler and less reagent consuming method, using as starting point the method described in Rodrigues *et al.* (2015)<sup>193</sup>. To accomplish that, intermediate concentrations of ferulic acid (50  $\mu\text{g/mL}$ ) and curcumin (12.5  $\mu\text{g/mL}$ ) were analyzed and different H<sub>2</sub>O/acetonitrile ratios, use of gradient vs. isocratic method, flow rates and time of analysis were tested.

The optimized method for ferulic acid analysis was set as 10 min, at a flow rate of 1 mL/min, with 50% H<sub>2</sub>O (with 0.1% formic acid) and 50% acetonitrile (with 0.1% formic acid) as the mobile phases, being detected at 310 nm. For curcumin, the composition of the mobile phases and the flow rate were the same, but it was necessary a 15 min method and was detected at 425 nm. The retention times were 2.6 min and 9.3 min for ferulic acid and curcumin, respectively.

### 5.3.8. *CURS1* mRNA quantification by qPCR

Total RNA from *L. lactis* LMG19460 harboring the plasmid of interest was extracted using the High Pure RNA Isolation Kit (Roche) according to the manufacturer's instructions. The extracted RNA was quantified using a Nanodrop Spectrophotometer (Nanovue Plus, GE) and 200 ng of total RNA was used for reverse transcription. This step was performed with the First Strand cDNA Synthesis Kit for RT-PCR (AMV) (Roche), using the primer mRNA\_CURS\_R (Table 5.3). The real-time qPCR reactions were performed in the Roche LightCycler detection system using the NZYSpeedy qPCR Green (Nzytech). Each 20  $\mu\text{L}$  reaction mixture had 10  $\mu\text{l}$  of 2x NZYSpeedy qPCR Green Master Mix, 0.8  $\mu\text{l}$  of each primers (final concentration 0.4  $\mu\text{M}$ ) (mRNA\_CURS\_F/R, Table 5.3), 6.4  $\mu\text{l}$  of PCR-grade water and the remaining volume were completed with a fixed volume (2  $\mu\text{L}$ ) of cDNA first strand. Negative controls without cells and without cDNA or gDNA were performed for experimental validation.

The five cDNA qPCR points of the standard curve were obtained using plasmid DNA purified as described above and serially diluted in PCR-grade water (1, 10, 100, 1,000 and 10,000 pg).

The cycling conditions were as follows: initial denaturation for 2 min at 95°C, followed by 40 cycles of 10s at 95°C, 5s at 55°C and 14s at 72°C. To confirm that only the desired specific targets were amplified, melting curve analyses were performed by making a temperature gradient of 0.05° C/s from 70 to 95°C.

The number of *CURSI* molecules were calculated by the absolute quantification method using the standard curve.

**Table 5.3.** Sequence, expected amplicon size and melting temperature of the primers used for *CURSI* mRNA quantification by real-time quantitative PCR.

Primer name		Sequence (5' → 3') <sup>1</sup>	Product size (bp)	T <sub>m</sub> (°C)
mRNA_CURS_F	F	ACCATCATGGCTATCGGTAC	93	54.1
mRNA_CURS_R	R	GTCGTCAGAGTTGGTAACACG		55.5

<sup>1</sup> F and R indicate forward and reverse primers, respectively.

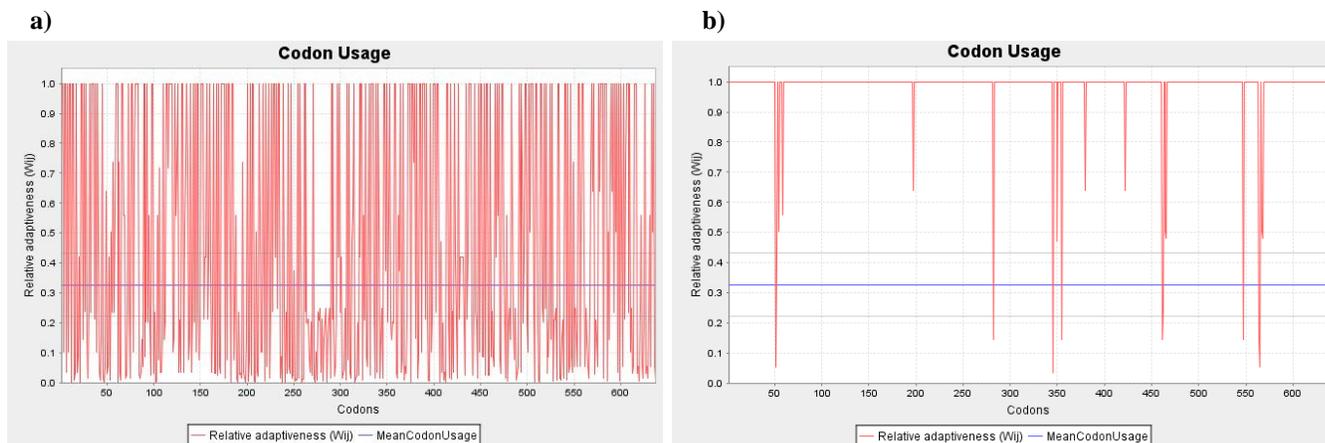
## 5.4. Results and discussion

### 5.4.1. Cloning of the curcumin pathway genes

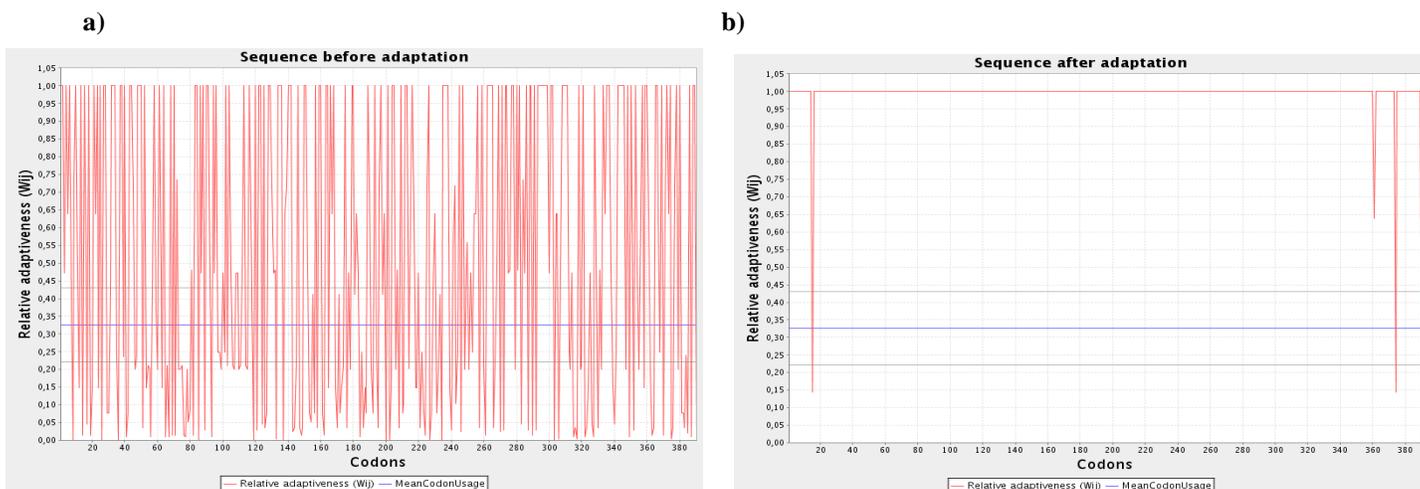
#### 5.4.1.1. *In silico* design and codon optimization

The codon optimization of the selected genes of the curcumin pathway is an important task, since they originally come from plants and the codons preferred by these eukaryotic organisms are very different from the ones used by bacteria. Before proceeding to the synthesis and cloning of the three genes of the curcumin pathway in the pTRKH3 plasmid, the sequences were designed and tested *in silico*, using the APE<sup>114</sup> and SnapGene<sup>115</sup> softwares.

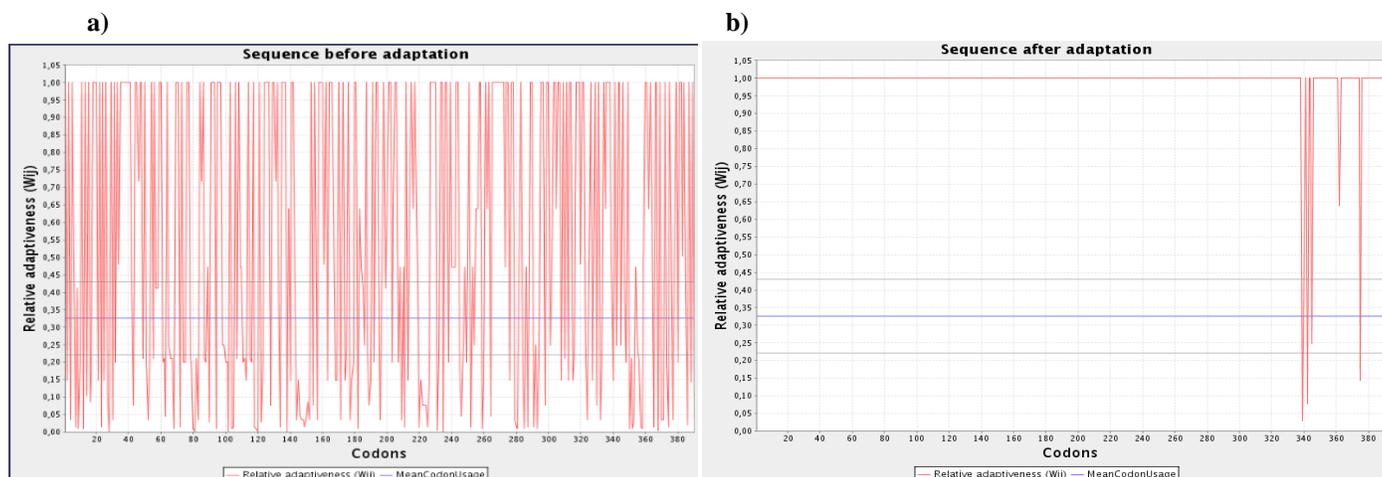
First, the sequence of the three genes was retrieved from Genbank (*4CL* with the accession number D49366.1, *DCS* with AB495006.1, and *CURSI* with AB495007.1). Only the sequence between the start and stop codons was chosen and codon optimized for *E. coli* K12 using Java Codon Optimization Tool<sup>200</sup> online software. For each gene sequence the codon usage before and after optimization with the software are represented in Figures 5.13, 5.14 and 5.15. In the end, the codon usage was not fully optimized, because the software was not allowed to remove several important restriction enzyme sites, as well as to generate stop codons and transcription terminators. The codon optimized sequences were depicted on Table 5.4.



**Figure 5.13.** 4CL sequence codon usage (red line) in *E. coli* K12 a) before codon optimization and b) after codon optimization using JCat online software<sup>200</sup>. The blue line represents the mean codon usage in *E. coli* K12, considering all of the organism's genes, and the grey lines depicts the corresponding standard deviation intervals.



**Figure 5.14.** DCS sequence codon usage (red line) in *E. coli* K12 a) before codon optimization and b) after codon optimization using JCat online software<sup>200</sup>. The blue line represents the mean codon usage in *E. coli* K12, considering all of the organism's genes, and the grey lines depicts the corresponding standard deviation intervals.



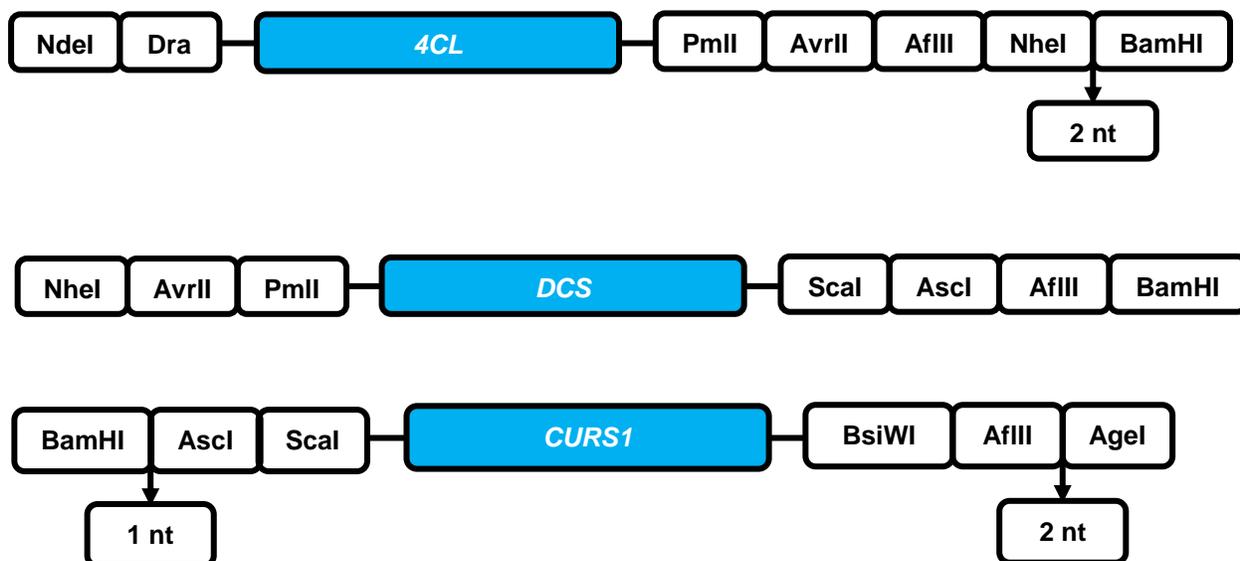
**Figure 5.15.** *CURS1* sequence codon usage (red line) in *E. coli* K12 a) before codon optimization and b) after codon optimization using JCat online software<sup>200</sup>. The blue line represents the mean codon usage in *E. coli* K12, considering all of the organism's genes, and the grey lines depicts the corresponding standard deviation intervals.

**Table 5.4.** *4CL*, *DCS* and *CURS* genes sequence synthesized by Nzytech, after codon optimization.

Gene name	Synthesized sequence
<i>4CL</i>	<p> <u>gggggcatatgtttaa</u>ATGGACACCCAGACCAAAACCCGACCAGAAAGACATCATCTTCCGTT  CTAAACTGCCGACATCTACATCCCGAAACACCTGCCGCTGCACTCTTACTGCGGTG  AAAACATCTCTCAGTTCTTCTCGTCCGTGCCTGATCAACGGTTCTAATGATCGTGT  ATACACCTACGCAGAAGTTGAAATCACCTCTCGTAAAGTTGCTGCTGGTCTGCACAA  ACACGGTATCAAACAGACCGAAACCATCATGCTGCTGCTGCCGAACTGCCCGGAATT  CGTTTTCGCTTTCCTGGGTGCTTCTTACATCGGTGCTGTTTCTACCACCGCTAACCCGT  TCTTCACCTCTTCTGAAATCATCAAACAGGCTAAAGCTTCTAAAACCAAATGATCAT  CACCGTTTCTACCACCGTTCCGAAACTGAAAGACTTCTCTCAGGAAAACCACGTTAA  AATCATGTGCATCGACGACAAAATCGACGGTTGCCTGCACTTCTTCTGACCTGGA  AAACTCTGACGAAACCACCTGCCGGACGTTGAAATCCGTCCGGACGACGTTGTTGC  TCTGCCGTA CTCTTCTGGTACCACCGGCTGCCGAAAGGTGTTATGCTGACCCACAA  AGGTCTGGTTACCTCTGTTGCTCAGCAGGTTGACGGTGACAACGCTAACCTGTACAT  GCACCACGAAGACGTTGTTATGTGCACCTGCCGCTGTTCCACATCTACTCTATGAAC  TCTATCCTGCTGTGCGGTCTGCGTGTGGTGCTGCTATCCTGCTGATGCACAAATTCG  AAATCGTTACCTTCTGGAAGTATCCAGCGTTACAAAGTTACCATCGGTCCATTCGT  TCCGCCGATCGTTCTGGCTATCGCTAAATCTAACGTTGTTGACCAGTACGACCTGTCT  ACCGTTCGTACCGTTATGTCTGGTGCTGCTCCGCTGGGTTCTGAACTGGAAGACGCTG  TTCGTGCTAAATTCGGAACGCTAAACTGGGTGACGGTTACGGTATGACCGAAGCTG  GTCCGTTTCTCGCTATGTGCCTGGCGTTCGCTAAAGAACCATTGAAATCAAATCTG  GTGCTTGCGGTACCGTTGTTGTA ACTCTGAAATGAAAATCATCGACACCGAAACCG  GCGTTTCTGCGCGTAACCAAGTCTGGTGAATCTGCATCCGTGGTGACAGATCA  TGAAAGGTTACCTGAACGACCCGGAAGCTACCGAACGTACCATCGACAAAGAAGGT  TGGTGCACACCGGCGACATCGGTTACATCGACGACGACGACGAACTGTTTCATCGTT  GACCGTCTGAAAGA ACTGATCAAATACAAAGTTTCCAGGTTGCTCCGCCGGA ACTG  GAAGCTCTGCTGGTTCCACATCCGAACGTAAGCGACGCTGCTGTTTCTATGAAA  GACGAAGGTGCTGGTGAAGTCCGGTTGCTTTCGTTGTTTCGTTCTAACGGTTCTACCA  CCACCGAAGACGAAATCAAACAGTTTCGTTTCTAAACAGGTTATCTTCTACAAACGTA  TCAACCGTGTTCGTTGTTGACTCTATCCCGAAATCTCCGCTGCTGGTAAAATCGTTTCG  TAAAGACCTGCGTGCTAAACTGGCTGCTCGTTTCTGAAACGGTCCAACCAACGTT  TGTTCCGAACGGTGGTAACGTTGCTAAAGACAACGTTCCAAATGGTGTAAGCAACGG </p>

	TGTTTCTAAAGCTAACGGTGGTGTGCTAAAGAAGGTGTTGCTAACGGTGTTCGGAC CGACGGTACTACGGTGTGCTACCAAAGGTGTTGCTAACGGTATCTCTAACGGTGT TTACAAACAGGTTTCTAACGGTGTGTTTCTAACGGTGTGCTAACGGTATCGTTTCT AACGGTATCGCTAACGGTGTTCACAACCTACcagctgctaggcttaaggctagcaaggatccggggg
<i>DCS</i>	<u>ggggggctagccctaggcacgtg</u> ATGGAAGCTAACGGTACCCTATCACCCACTCTGCTGACGG TCCAGCTACCATCCTGGCTATCGGTACCGCTAACCCGACCAACGTTGTTGACCAGAA CGTTACCCGGACTTCTACTTCCGTGTTACCAACTCTGAATACCTGCAGGAACCTGAAA GCTAAATTCCGTCGTATCTGCGAAAAAGCTGCTATCCGTA AACGTCACCTGTACCTG ACCGAAGAAATCCTGCGTGA AAAACCCGTCTCTGCTGGCTCCGATGGCTCCGTCTTTC GACGCTCGTCAGGCTATCGTTGTTGAAGCTGTTCCGAAACTGGCTAAAGAAGCTGCT GAAAAAGCTATCAAAGAATGGGGTCGTCCGAAATCTGACATCACCCACCTGGTTTTTC TGCTCTGCTTCTGGTATCGACATGCCGGGTTCTGACCTGCAGCTGCTGAAACTGCTGG GTCTGCCGCCGTCTGTTAACCGTGTATGCTGTACAACGTTGGTTGCCACGCTGGTGG TACCGCTCTGCGTGTGCTAAAGACCTGGCTGAAAACAACCGTGGTCTGCTGTTCT GGCTGTTTGTCTGTAAGTTACCGTCTGTCTTACCCTGGTCCGACCCGGCTCACATC GAATCTCTGTTGCTCAGGCTCTGTTCCGGTGACGGTGTGCTGCTGCTGGTTGTTGGTT CTGACCCGGTTGACGGTGTGAAACGTCCGATCTTCGAAATCGCTTCTGCTTCTCAGGT TATGCTGCCGGAATCTGCTGAAGCTGTTGGTGGTCACCTGCGTGA AATCGGTCTGAC CTTCCACCTGAAATCTCAGCTGCCGTCTATCATCGCTTCTAACATCGAACAGTCTCTG ACCACCGTTGCTCTCCGCTGGGTCTGTCTGACTGGAACCAGCTGTTCTGGGCTGTTC ACCCGGGTGGTCGTGCTATCCTGGACCAGGTTGAAGCTCGTCTGGGTCTGAAAAAAG ACCGTCTGGCTGCTACCCGTCACGTTCTGTCTGAATACGGTAACATGCAGTCTGCTAC CGTTCTGTTTCATCCTGGACGAAATGCGTAACCGTCTGCTGCTGAAGGTCACGCTACC ACCGGCGAAGGTCTGGACTGGGGTGTCTGCTGGGTTTCGGTCCAGGTCTGTCTATC GAAACCGTTGTTCTGCACTCTTGCCGTCTGAACTACagctactggcgcgcccttaaggatccggggg
<i>CURSI</i>	<u>gggggggatccggcgcgcccaagctact</u> ATGGCTAACCTGCACGCTCTGCGTCGTGAACAGCGTGCT CAGGGTCCGGTACCATCATGGCTATCGGTACCGCTACCCCGCCGAACCTGTACGAA CAGTCTACCTTCCCGACTTCTACTTCCGTGTTACCAACTCTGACGACAAACAGGAAC TGAAAAAATAATTCCGTCGTATGTGCGAAAAAACCATGGTTAAAAAACGTTACCTGC ACCTGACCGAAGAAATCCTGAAAGAACGTCCGAAACTGTGCTTTACAAAGAAGCTT CTTTCGACGACCGTCAGGACATCGTTGTTGAAGAAATCCCGCTCTGGCTAAAGAAG CTGCTGAAAAAGCTATCAAAGAATGGGGTCCGAAATCTGAAATCACCCACCTGG TTTTCTGCTCTATCTCTGGTATCGACATGCCGGGTGCTGACTACCGTCTGGCTACCCT GCTGGGTCTGCCGCTGACCGTTAACCGTCTGATGATCTACTCTCAGGCTTGCCACATG GGTGCTGCTATGCTGCGTATCGCTAAAGACCTGGCTGAAAACAACCGTGGTGTCTCGT GTTCTGGTTGTTGCTTGC GAAATCACCGTCTGTCTTTCCGTGGTCCGAACGAAGGTG ACTTCGAAGCTCTGGCTGGTCAGGCTGGTTTCGGTGACGGTGTGCTGGTGTGTTGTTGT TGGTGTGACCCGCTGGAAGGTATCGAAAAACCGATCTACGAAATCGTGTCTGCTAT GCAGGAAACCGTTGCTGAATCTCAGGGTGTGTTGGTGGTCACCTGCGTGTCTTTCGG TTGGACCTTCTACTTCTGAACCAGCTGCCGGCTATCATCGTGTGACAACCTGGGTCTG TCTCTGGAACGTGCTCTGGCTCCGCTGGGTGTTTCGTGAATGGAACGAGTTTTCTGGG TTGCTCACCCGGTA ACTGGGCTATCATCGACGCTATCGAAGCTAAACTGCAGCTGT CTCCGACAAACTGTCTACCGCTCGTCACGTTTTACCGAATACGGTAACATGCAGT CGGCGACCGTCTACTTCGTGATGGACGAACTGCGTAAACGTTCTGCTGTTGAAGGTC GTTCTACCACCGGCGACGGTCTGCAGTGGGGTGTCTGCTGGGTTTCGGTCCAGGTCT GTCTATCGAAACCGTTGTTCTGCGTTCTATGCCGCTGTACcgtactgcttaagaaacgggtggggg

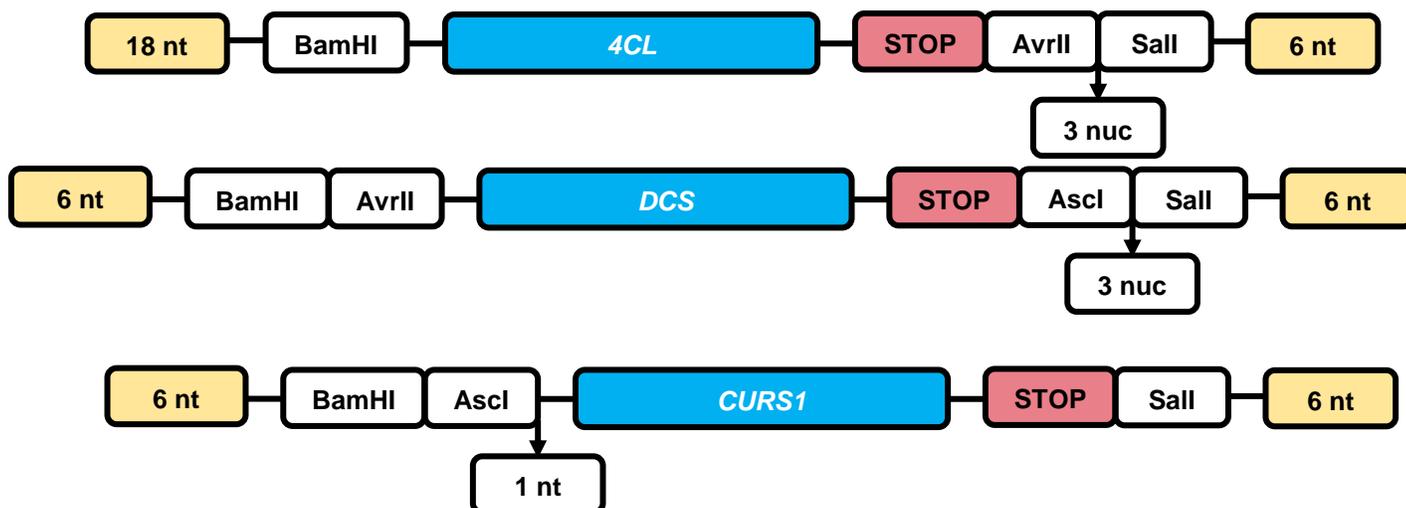
In addition, several unique restriction enzyme sites were added to each end of the 3 genes, following the strategy already described in the Materials and Methods section. The final design for the synthesis of each gene by Nzytech is represented in Figure 4.16. In order to keep the sequences *in frame* when the genes are cloned as an operon, several nucleotides had to be added between some restriction enzyme sites, as stated on the Figure 5.16.



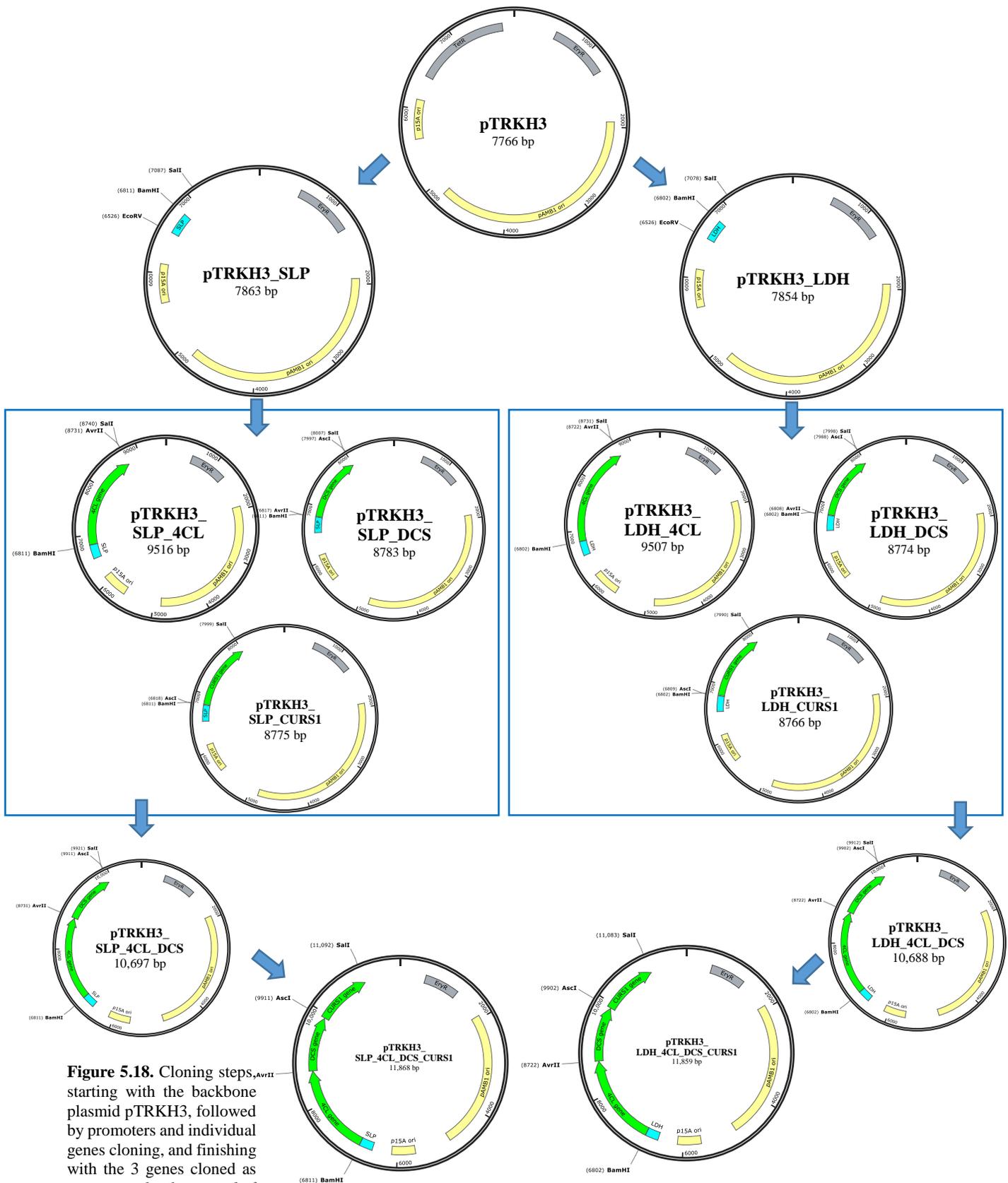
**Figure 5.16.** Final design of the *4CL*, *DCS* and *CURS1* genes, with several flanking unique restriction enzyme sites, that was sent for synthesis by Nzytech.

After being synthesized by Nzytech, the genes cloned in pUC57 plasmids had to be amplified to be subcloned in the pTRKH3 plasmid. The primers were designed in order to introduce compatible restriction enzyme sites, as well as stop codons at the end of each gene, to allow for a correct translation step, without fusion proteins. Several additional nucleotides were considered between some restriction sites to keep the design *in frame* when the 3 genes were cloned sequentially as an operon. Other nucleotides were added in each end of the 3 fragments to allow the correct performance of the restriction enzymes (Figure 5.17).

The several cloning steps were also simulated *in silico*, as represented in Figure 5.18. First, both SLP and LDH promoters were cloned in the pTRKH3 plasmid. Next, each gene was cloned individually downstream each promoter. Finally, the genes were cloned sequentially, as an operon. The relevant enzymes for each cloning step are also shown.



**Figure 5.17.** *In silico* design of the *4CL*, *DCS* and *CURS1* genes amplification step, with restriction enzyme sites compatible with cloning in the pTRKH3 plasmid.

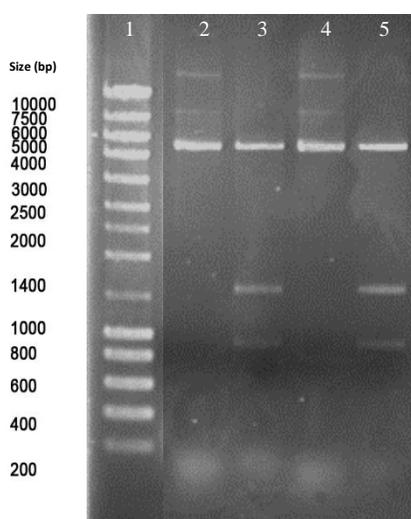


**Figure 5.18.** Cloning steps, starting with the backbone plasmid pTRKH3, followed by promoters and individual genes cloning, and finishing with the 3 genes cloned as operons under the control of each promoter.

#### 5.4.1.2. Molecular cloning of the curcumin pathway

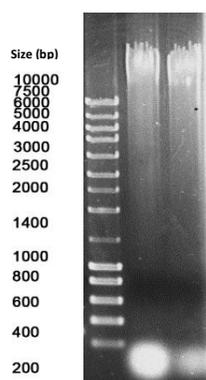
The backbone vector for all the cloning procedures was the pTRKH3 plasmid, whose main features are represented in Figure 4.18. This plasmid was ordered from the LMG/BCCM collection (Ghent, Belgium) and was originally inside *E. coli* JM110 strain. Since the amount of pDNA obtained from JM110 strain was low ( $\pm 15$  ng/ $\mu$ L/OD) the purified pTRKH3 was used to transform *E. coli* DH5 $\alpha$ , which allowed the purification of 10 fold more plasmid than the obtained in the JM110 strain.

In Figure 5.19 are represented two independent pTRKH3 plasmids purified from the *E. coli* DH5 $\alpha$  strain. The non-digested samples showed the three typical isoforms, with supercoiled being the predominant isoform, which confirms the high quality of the plasmid, and with the linear isoform having the expected length (7,766 bp). In the lanes where the pTRKH3 was digested with HindIII and EcoRI, appeared the three expected bands with 5,265, 1,523 and 978 bp.



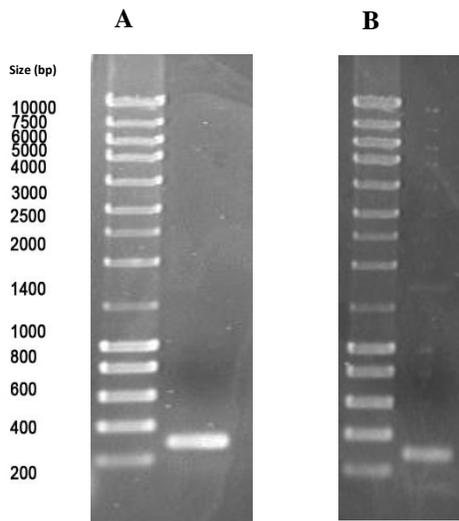
**Figure 5.19.** Quality characterization and confirmation of the pTRKH3 plasmid purified from *E. coli* DH5 $\alpha$ . Lanes: 1- Nzy ladder III; 2- Non-digested pTRKH3 sample A; 3- pTRKH3 sample A digested with HindIII and EcoRI (5265 + 1523 + 978 bp); 4- Non-digested pTRKH3 sample B; 5- pTRKH3 sample B digested with HindIII and EcoRI (5265 + 1523 + 978 bp).

The SLP and LDH constitutive promoters were first cloned in the backbone vector pTRKH3, after PCR amplification of purified gDNA of *Lactobacillus acidophilus* ATCC4356 (Figure 5.20), using the conditions described in the Materials and Methods section 5.3.4.2.



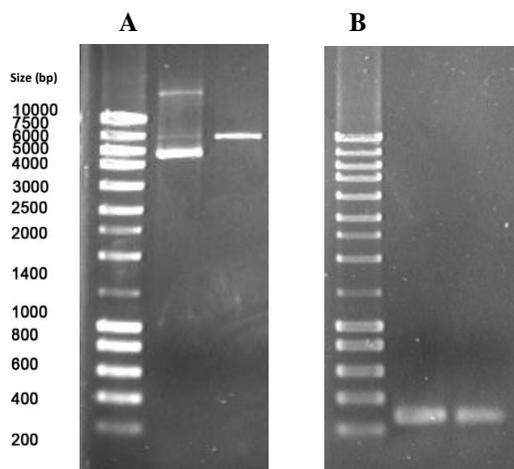
**Figure 5.20.** Two samples of genomic DNA from *L. acidophilus* ATCC4356. Lanes: 1- Nzy ladder III; 2- gDNA sample A (1766 ng/ $\mu$ L); 3- gDNA sample B (1900 ng/ $\mu$ L).

After amplification, the PCR products were visualized in a 1% agarose gel (Figure 5.21) and both had the expected lengths (305 bp for SLP and 296 bp for LDH promoter). Each fragment was recovered from the gel, without exposing it to ethidium bromide or UV light, and was purified and again visualized on a 1% agarose gel to confirm that the correct band was cut. Although after purification from the gel the concentrations were much lower, the bands had the expected size and the cloning could proceed.



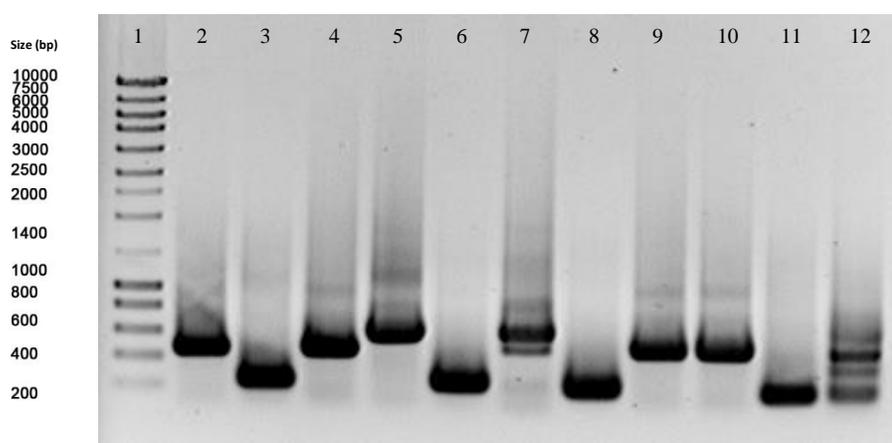
**Figure 5.21.** PCR products from the amplification of the A) SLP (305 bp) and B) LDH (296 bp) promoters from the gDNA of the *L. acidophilus* ATCC4356. Lanes: 1- Nzy ladder III; 2- SLP promoter fragment; 3- Nzy ladder III; 4- LDH promoter fragment.

A total of 3 µg of both the amplified inserts as well as the cloning vector (pTRKH3) were digested using appropriate restriction enzymes (EcoRV and BamHI). Figure 5.23A shows the non-digested pTRKH3 next to the digested one. After digestion with EcoRV and BamHI, pTRKH3 should have shown 2 bands (7,576 and 190 bp), but the smaller one is not visible, probably due to its low concentration. The Figure 5.22B holds the digestion of both the SLP and LDH promoters' fragments with the same enzymes.



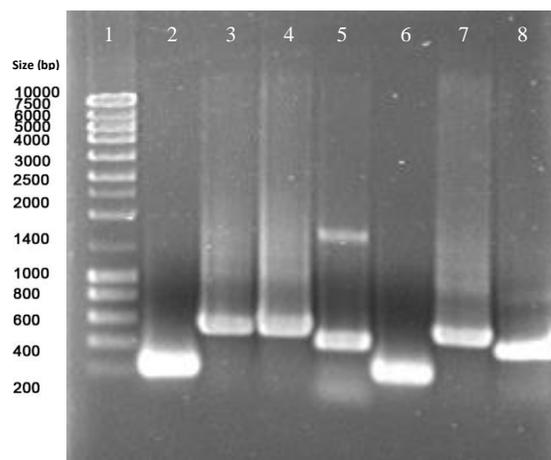
**Figure 5.22.** Double digestion with EcoRV and BamHI of A) pTRKH3 and B) SLP and LDH promoters. Lanes: 1- Nzy ladder III; 2- Non-digested pTRKH3; 3- pTRKH3 digested with EcoRV and BamHI; 4- Nzy ladder III; 5- SLP promoter digested with EcoRV and BamHI; 6- LDH promoter digested with EcoRV and BamHI.

After a second gel purification step, the samples were ligated using the protocol described in the Materials and Methods section 5.3.4.2. A colony PCR was performed over the colonies obtained after *E. coli* DH5a transformation, using the Check\_pTRKH3\_F/R primers (Table 5.3). The amplification of the non-digested pTRKH3 should generate a 460 bp band, the digested but recircularized pTRKH3 should amplify a 270 bp band, while the desired clones' pTRKH3 with the SLP or the LDH promoter should amplify a 557 or 548 bp band, respectively. The PCR products from a colony PCR to five clones with each promoter were run in a 1% agarose gel (Figure 5.23). The second lane corresponds to the amplification of the non-digested pTRKH3 as a control, which had the expected 460 bp band. The majority of the clones presented the band corresponding to the non-digested or to the digested but recircularized pTRKH3. Only the clone of the lane 5 had a single correct band of 557 bp that matches the one expected for the clone with the SLP promoter.



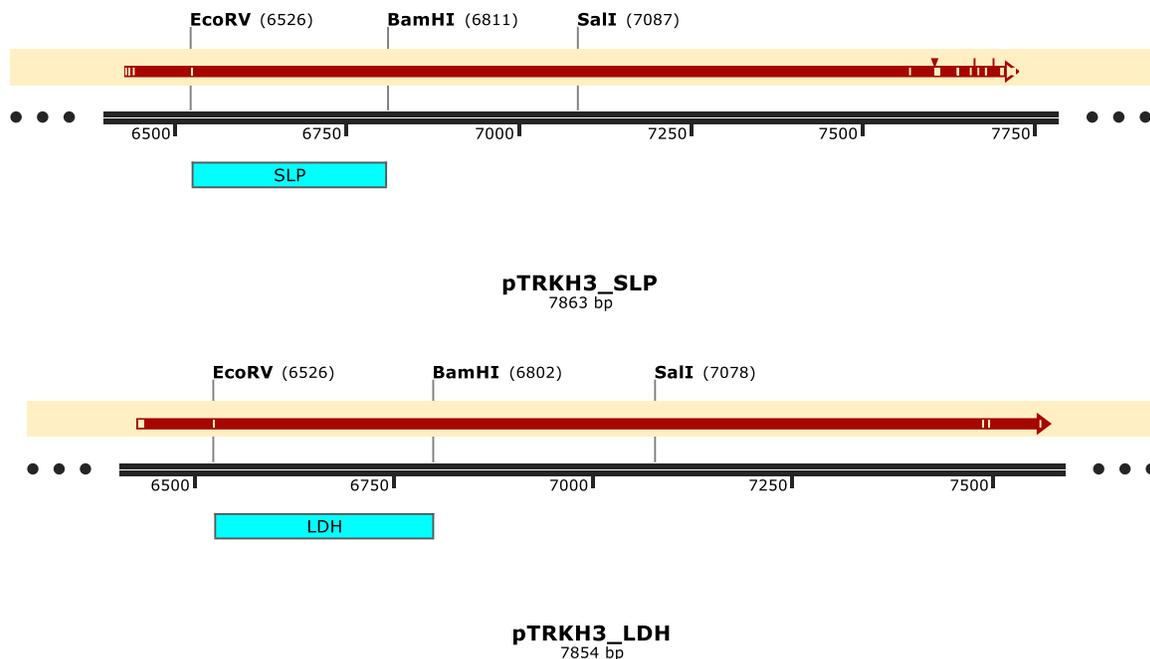
**Figure 5.23.** Colony PCR amplifications. Lanes: **1-** Nzyladder III; **2-** non-digested pTRKH3; **3-** pTRKH3 ligated with the SLP promoter (3h, RT, colony 1); **4-** pTRKH3 ligated with the SLP promoter (3h, RT, colony 2); **5-** pTRKH3 ligated with the SLP promoter (OV, 4°C, colony 1); **6-** pTRKH3 ligated with the SLP promoter (OV, 4°C, colony 2); **7-** pTRKH3 ligated with the LDH promoter (3h, RT, colony 1); **8-** pTRKH3 ligated with the LDH promoter (3h, RT, colony 2); **9-** pTRKH3 ligated with the LDH promoter (OV, 4°C, colony 1); **10-** pTRKH3 ligated with the LDH promoter (OV, 4°C, colony 2); **11-** pTRKH3 ligated with the LDH promoter (OV, 4°C, colony 1); **12-** pTRKH3 ligated with the LDH promoter (OV, 4°C, colony 2). RT- room temperature; OV- overnight.

A second colony PCR was performed to several other colonies resulting from the ligation with the LDH promoter, in order to find a correct clone. The Figure 5.24 shows the gel of the PCR products, being noticeable that only lane 3 and 4 had a band with the expected length of 548 bp.



**Figure 5.24.** Colony PCR amplifications. Lanes: **1-** Nzyladder III; **2-** digested and recircularized pTRKH3; **3-** pTRKH3 ligated with the LDH promoter (3h, RT, colony 1); **4-** pTRKH3 ligated with the LDH promoter (3h, RT, colony 2); **5-** pTRKH3 ligated with the LDH promoter (3h, RT, colony 3); **6-** pTRKH3 ligated with the LDH promoter (O/N, 4°C, colony 1); **7-** pTRKH3 ligated with the LDH promoter (O/N, 4°C, colony 2); **8-** pTRKH3 ligated with the LDH promoter (O/N, 4°C, colony 3). RT- room temperature; O/N- overnight.

The correct colonies from both colony PCRs (pTRKH3 ligated with the SLP promoter, 3h, RT, colony 3 and pTRKH3 ligated with the LDH promoter, 3h, RT, colony 2) were grown in liquid medium, in order to purify pDNA using High Pure Plasmid Isolation Kit (Roche) and sent to Stabvida to sequencing using the Check\_SLP\_F and Check\_LDH\_F primers (Table 5.3). Both pDNAs had the expected sequence (Figure 5.25), with the exception of a 2 nucleotides gap in the EcoRV site upstream the promoters. Since this gap do not affect the promoters sequence itself, the next cloning step was prepared over that clones.

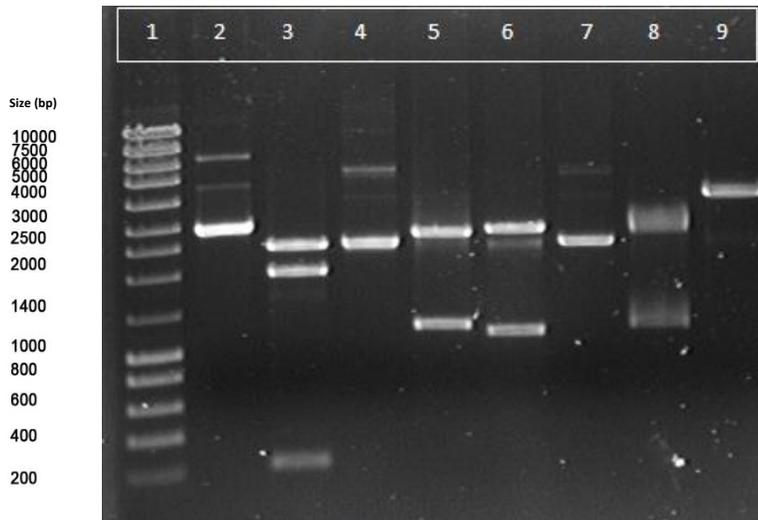


**Figure 5.25.** Alignment of the expected sequence of the pTRKH3 cloned with the SLP or LDH promoters against the real sequence. The gaps or mismatches are represented as white lines/gaps in the red arrow.

After having the SLP and the LDH promoters cloned individually in the pTRKH3, the next step was to clone the curcumin pathway genes. First, the genes *4CL*, *DCS* and *CURS1* were synthesized by Nzytech and sent after being cloned in the EcoRV site of different pUC57 plasmids. When these purified DNAs were received, different aliquots of electrocompetent DH5 $\alpha$  cells were transformed. The plasmid DNA purified from those cells was analyzed and characterized using restriction analysis and agarose gel electrophoresis (Figure 5.26). The pUC57 plasmid with the *4CL* gene (lane 2) presented the 3 isoforms, with the supercoiled being the most abundant. When subjected to a double digestion with NdeI and BamHI, appeared 3 bands in the gel (2,458, 1,949 and 251 bp) with the correct lengths (lane 3). The digestion should also form an 11 bp band, that wasn't visible in the gel. The 1,949 bp band corresponds to the *4CL* gene.

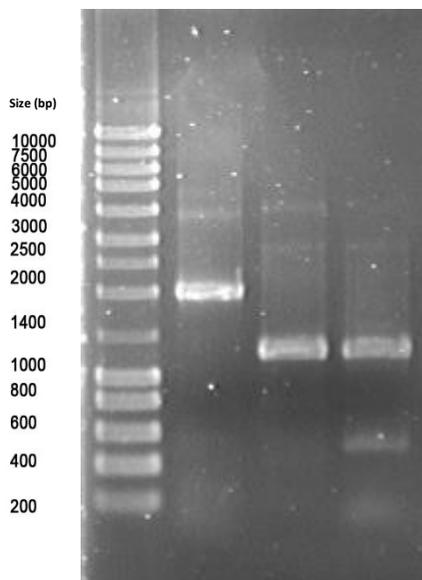
The puc57-*DCS* and pUC57-*CURS1* plasmids also had the 3 isoforms present, with supercoiled being the most relevant (lanes 4 and 7). The double digestion of pUC57-*DCS* with NheI and BamHI originated the 2,709 and 1,208 bp bands, and an expected 11 bp is again not visible in the gel (lane 5). When the pUC57-*CURS1* was digested with BamHI and

AgeI, the expected bands (lane 8) were equal to those from pUC57-DCS, except the band that corresponds to the gene that had 1208 bp for *DCS* and 1205 bp for *CURS1* gene. Since the restriction pattern were so identical for the last two plasmids, an additional digestion with XmnI was performed. In pUC57-DCS, it generated 2 bands, one with 2,790 and other with 1,138 bp (lane 6), while in pUC57-CURS1 generated only one band with 3,925 bp (lane 9), as expected.



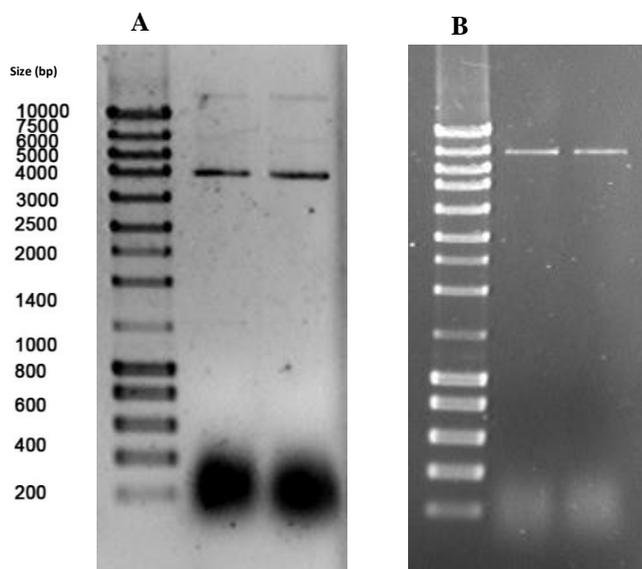
**Figure 5.26.** Characterization and confirmation of pUC57-4CL, DCS and CURS1 plasmids. Lanes: **1-** Nzy ladder III; **2-** Non-digested pUC57-4CL; **3-** pUC57-4CL digested with NdeI and BamHI (2458 + 1949 + 251 + 11 bp); **4-** Non-digested pUC57-DCS; **5-** pUC57-DCS digested with NheI and BamHI (2709 + 1208 + 11 bp); **6-** pUC57-DCS digested with XmnI (2790 + 1138 bp); **7-** Non-digested pUC57-CURS1; **8-** pUC57-CURS1 digested with BamHI and AgeI (2709 + 1205 + 11 bp); **9-** pUC57-CURS1 digested with XmnI (3925 bp).

The 3 genes were amplified using the Ampl\_4L\_F/R, Ampl\_DCS\_F/R and Ampl\_CURS1\_F/R primers (Table 5.3), as described in the appropriate Materials and Methods section. After amplification, the PCR products were digested with DpnI, followed by purification with NZYGelPure (Nzytech). The three purified fragments were digested with the appropriate restriction enzymes (BamHI and Sall) and the bands corresponding to each gene (1,929 bp for *4CL*, 1,196 bp for *DCS* and 1,188 bp for *CURS1*) were purified from the gel (Figure 5.27).



**Figure 5.27.** Digestion of the amplified and purified genes with BamHI and Sall. Lanes: **1-** Nzy ladder III; **2-** *4CL* gene fragment; **3-** *DCS* gene fragment; **4-** *CURSI* gene fragment.

The non-digested vectors pTRKH3\_SLP and pTRKH3\_LDH, as well as its digestion with BamHI and Sall were visualized in a 1% agarose gel (Figure 5.28). The majority of the supercoiled isoform in the non-digested lanes confirms the quality of the plasmids. The digested pTRKH3\_SLP and pTRKH3\_LDH presented the expected bands: 7,587 + 276 bp and 7,578 + 276 bp, respectively.

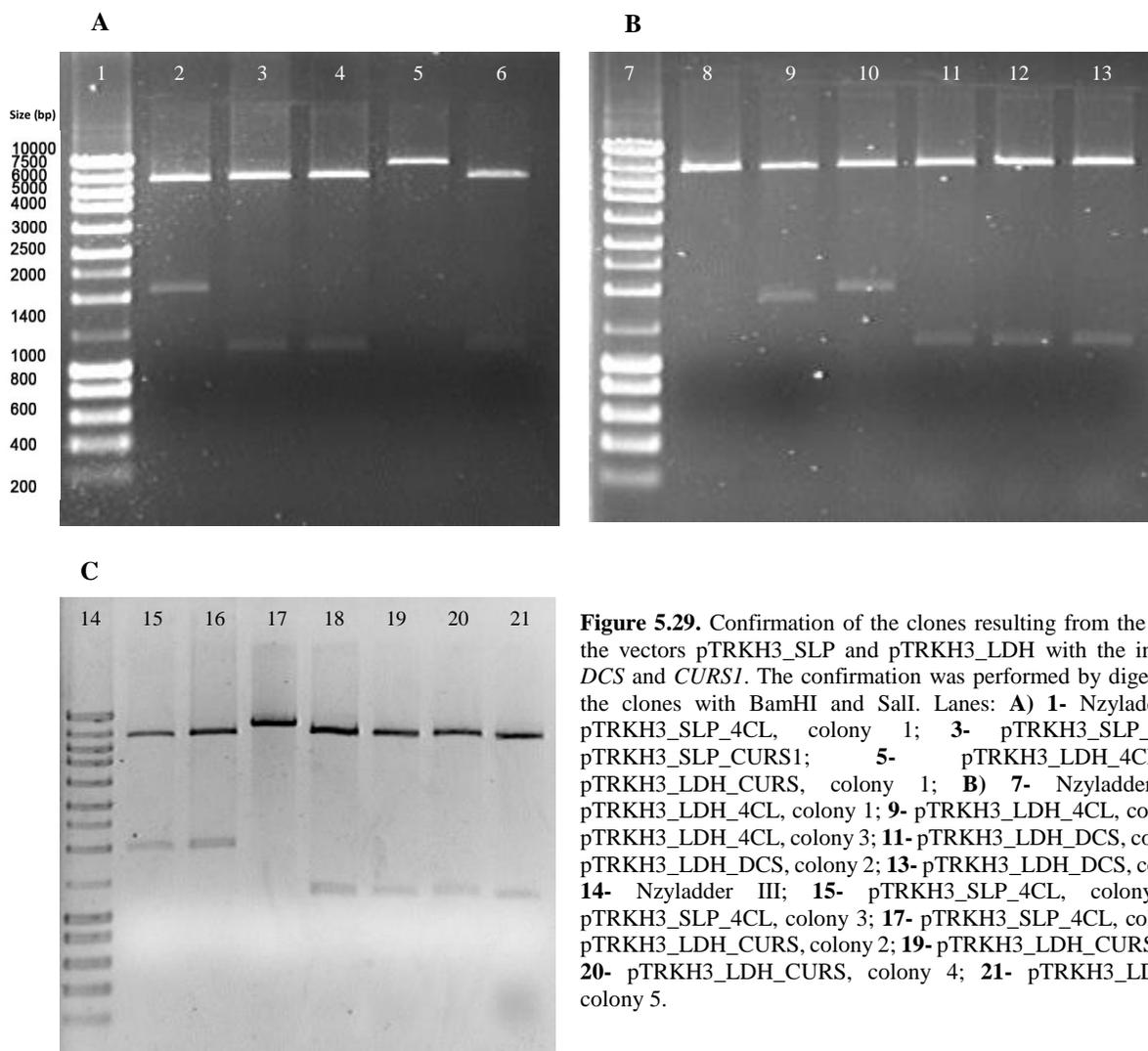


**Figure 5.28.** pTRKH3\_SLP and pTRKH3\_LDH A) non-digested and B) digested with BamHI and Sall. Lanes: **1-** Nzy ladder III; **2-** Non digested pTRKH3\_SLP; **3-** Non-digested pTRKH3\_LDH; **4-** Nzy ladder III; **5-** pTRKH3\_SLP digested with BamHI and Sall; **6-** pTRKH3\_LDH digested with BamHI and Sall.

For the subsequent cloning step, the bands from Figure 5.27 and 5.28, corresponding to the inserts (*4CL*, *DCS* and *CURSI* genes) and the vectors (7,587 and 7,578 bp bands), respectively, were excised and purified from the gel. The

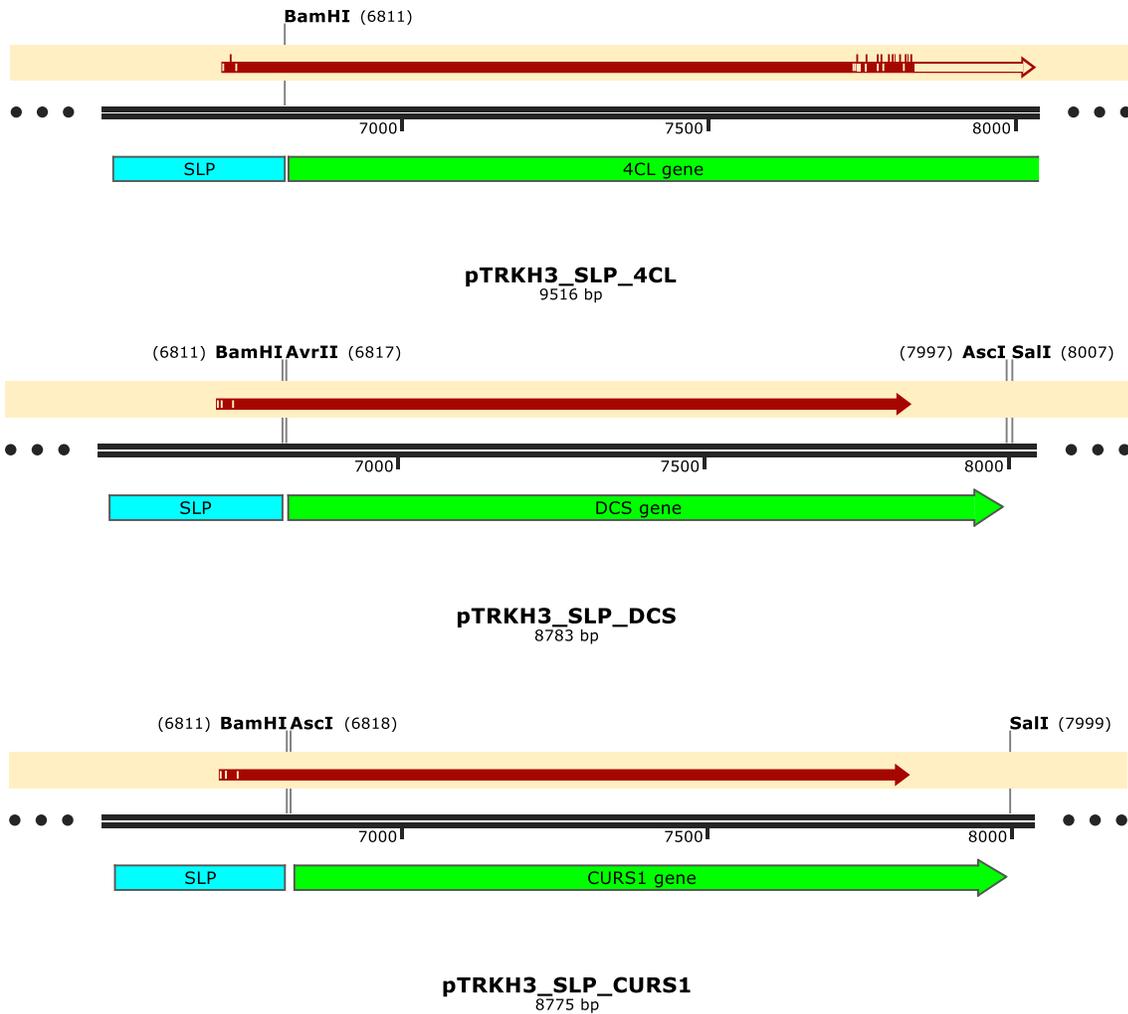
colonies resultant from the ligation and transformation procedures were grown in liquid medium and its pDNA purified. Each pDNA was digested with BamHI and SalI to confirm the presence of the correct clone (Figure 5.29).

The correct pTRKH3\_SLP\_4CL clone can be visualized on lane 16, with the 7,587 and 1,929 bp expected bands. The pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1 had the correct digestion pattern in the lanes 3 (7,587 + 1,196 bp) and 4 (7,587 + 1,188 bp), respectively. On lane 10 there was the correct pTRKH3\_LDH\_4CL with the two bands of 7,578 and 1,929 bp. In the case of pTRKH3\_LDH\_DCS and pTRKH3\_LDH\_CURS1 there were more than one lane with the correct patterns: lanes 11-13 for pTRKH3\_LDH\_DCS (7,578 + 1,196 bp) and lanes 18-21 for pTRKH3\_LDH\_CURS1 (7,578 + 1,188 bp).

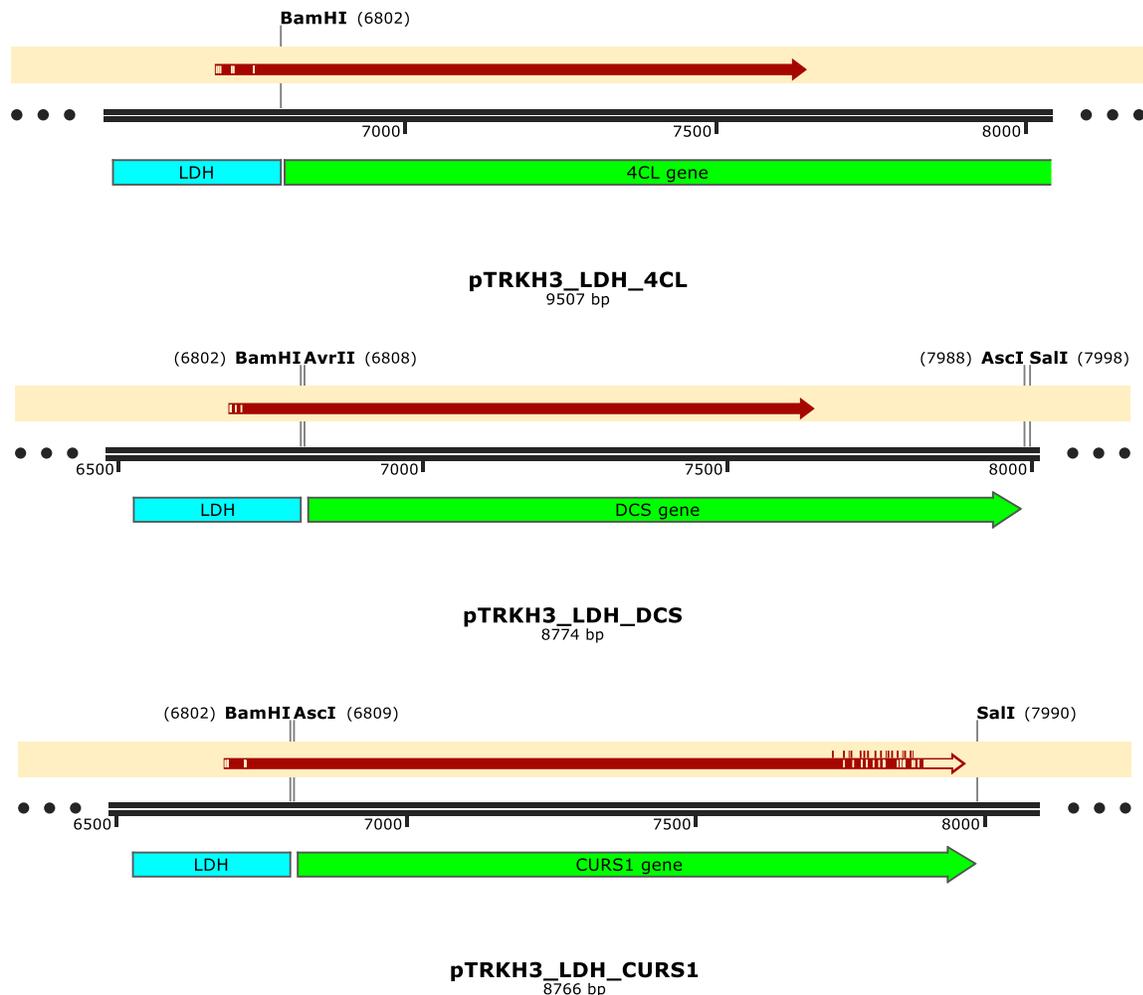


**Figure 5.29.** Confirmation of the clones resulting from the ligation of the vectors pTRKH3\_SLP and pTRKH3\_LDH with the inserts *4CL*, *DCS* and *CURS1*. The confirmation was performed by digestion of all the clones with BamHI and SalI. Lanes: **A)** 1- Nzyladder III; 2- pTRKH3\_SLP\_4CL, colony 1; 3- pTRKH3\_SLP\_DCS; 4- pTRKH3\_SLP\_CURS1; 5- pTRKH3\_LDH\_4CL; 6- pTRKH3\_LDH\_CURS, colony 1; **B)** 7- Nzyladder III; 8- pTRKH3\_LDH\_4CL, colony 1; 9- pTRKH3\_LDH\_4CL, colony 2; 10- pTRKH3\_LDH\_4CL, colony 3; 11- pTRKH3\_LDH\_DCS, colony 1; 12- pTRKH3\_LDH\_DCS, colony 2; 13- pTRKH3\_LDH\_DCS, colony 3; **C)** 14- Nzyladder III; 15- pTRKH3\_SLP\_4CL, colony 2; 16- pTRKH3\_SLP\_4CL, colony 3; 17- pTRKH3\_SLP\_4CL, colony 4; 18- pTRKH3\_LDH\_CURS, colony 2; 19- pTRKH3\_LDH\_CURS, colony 3; 20- pTRKH3\_LDH\_CURS, colony 4; 21- pTRKH3\_LDH\_CURS, colony 5.

The correct clones were sent for sequencing and all come out correct, as represented in Figure 5.30 and Figure 5.31.

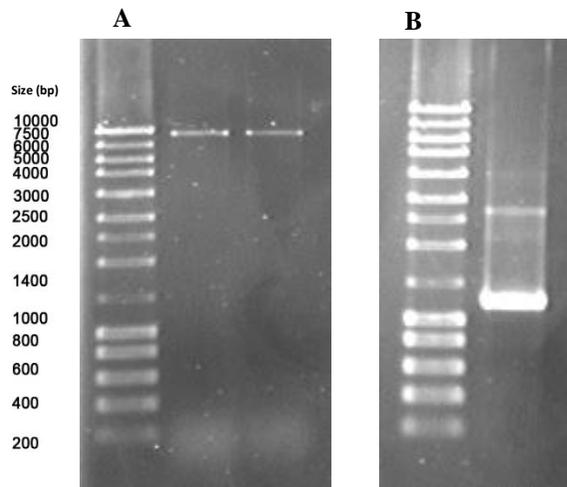


**Figure 5.30.** Alignment performed with the SnapGene software between the expected pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1 sequences and the real ones obtained by sequencing.



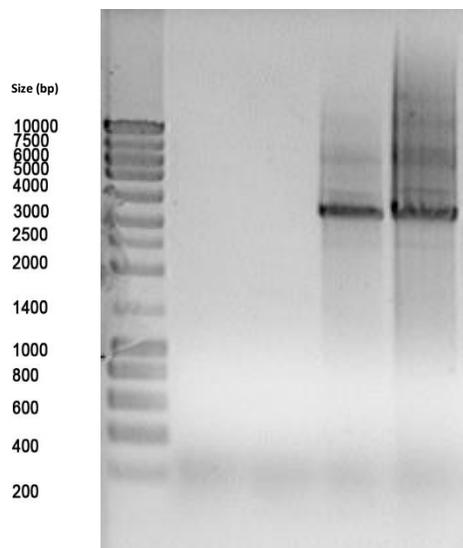
**Figure 5.31.** Alignment performed with the SnapGene software between the expected pTRKH3\_LDH\_4CL, pTRKH3\_LDH\_DCS and pTRKH3\_LDH\_CURS1 sequences and the real ones obtained by sequencing.

After the single gene clones were correctly constructed, the next step was to obtain the two- and three-fused genes clones. First, the *DCS* gene was added to the pTRKH3\_SLP\_4CL and pTRKH3\_LDH\_4CL clones. Both the vectors and the insert were digested with the appropriate enzymes (AvrII and SalI) and the resulting bands are shown in Figure 5.32. The correct bands corresponding to both vectors (9,516 bp for pTRKH3\_SLP\_4CL and 9,507 bp for pTRKH3\_LDH\_4CL) and to the *DCS* gene insert (1,190 bp) were excised and purified from the gel, followed by T4 ligase procedure, as described in the Materials and Methods section 5.3.4.2.

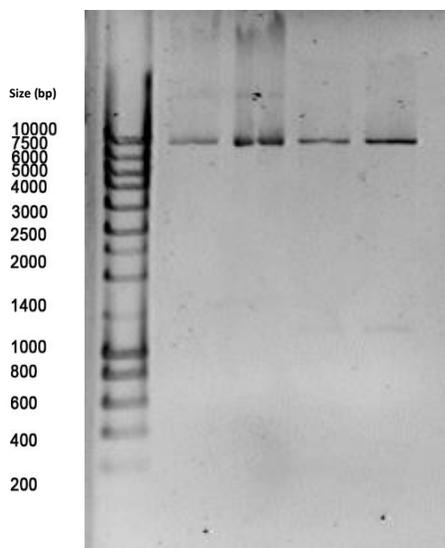


**Figure 5.32.** Digestion with AvrII and SalI of the A) vectors pTRKH3\_SLP\_4CL and pTRKH3\_LDH\_4CL, B) and for the DCS gene insert. Lanes: **1-** Nzy ladder III; **2-** pTRKH3\_LDH\_4CL digested with AvrII and SalI; **3-** pTRKH3\_SLP\_4CL digested with AvrII and SalI; **4-** Nzy ladder III; **5-** DCS gene fragment digested with AvrII and SalI.

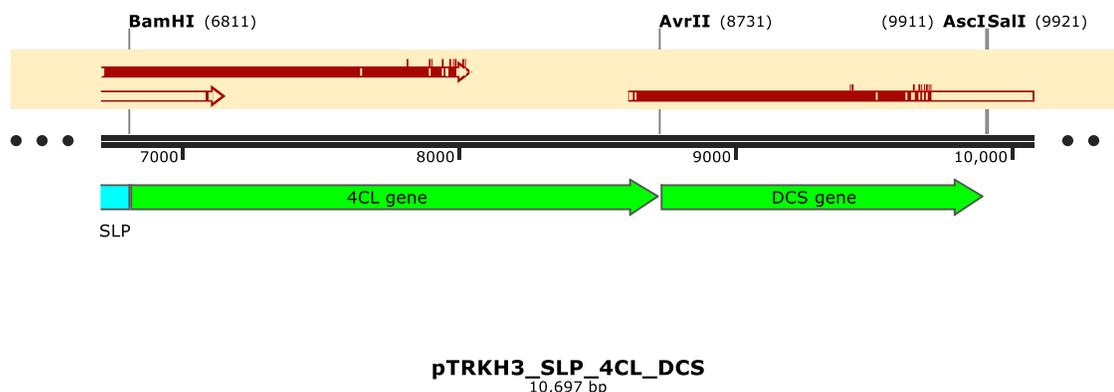
The resulting colonies were tested by colony PCR and the result is shown on Figure 5.33. After amplification with Check\_pTRKH3\_F and Ampl\_DCS\_R, the 3510 bp expected band for pTRKH\_SLP\_4CL\_DCS appeared on the colonies of the lanes 4 and 5. These two non-digested pDNAs, as well as its digestion with AvrII and SalI, were also visualized on a 1% agarose gel (Figure 5.34), where were visible the two correct bands with 9,507 and 1,190 bp. One of the clones was further confirmed by sequencing (Figure 5.35).



**Figure 5.33.** Colony PCR of colonies resulting from the ligation of pTRKH3\_SLP\_4CL and pTRKH3\_LDH\_4CL with the insert DCS. Lanes: **1-** Nzy ladder III; **2-** pTRKH3\_LDH\_4CL\_DCS, colony 1; **3-** pTRKH3\_LDH\_4CL\_DCS, colony 2; **4-** pTRKH3\_SLP\_4CL\_DCS, colony 1; **5-** pTRKH3\_SLP\_4CL\_DCS, colony 2.

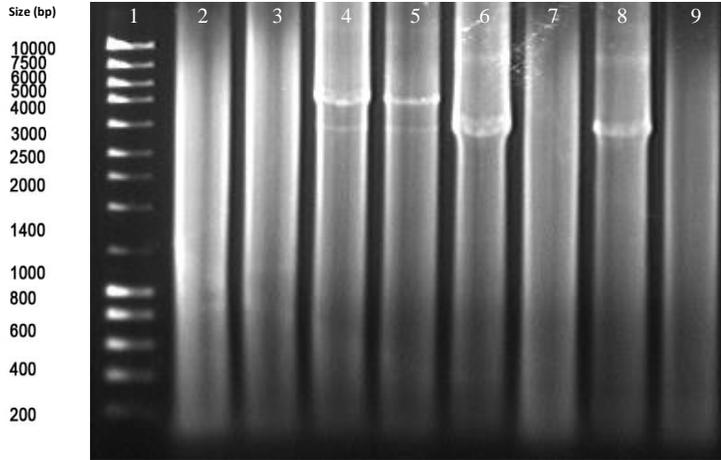


**Figure 5.34.** Confirmation of the pTRKH3\_SLP\_4CL\_DCS clones. Lanes: **1-** Nzy ladder III; **2-** Non-digested pTRKH3\_SLP\_4CL\_DCS, colony 1; **3-** Non-digested pTRKH3\_SLP\_4CL\_DCS, colony 2; **4-** pTRKH3\_SLP\_4CL\_DCS, colony 1, digested with AvrII and Sall; **5-** pTRKH3\_SLP\_4CL\_DCS, colony 2, digested with AvrII and Sall.

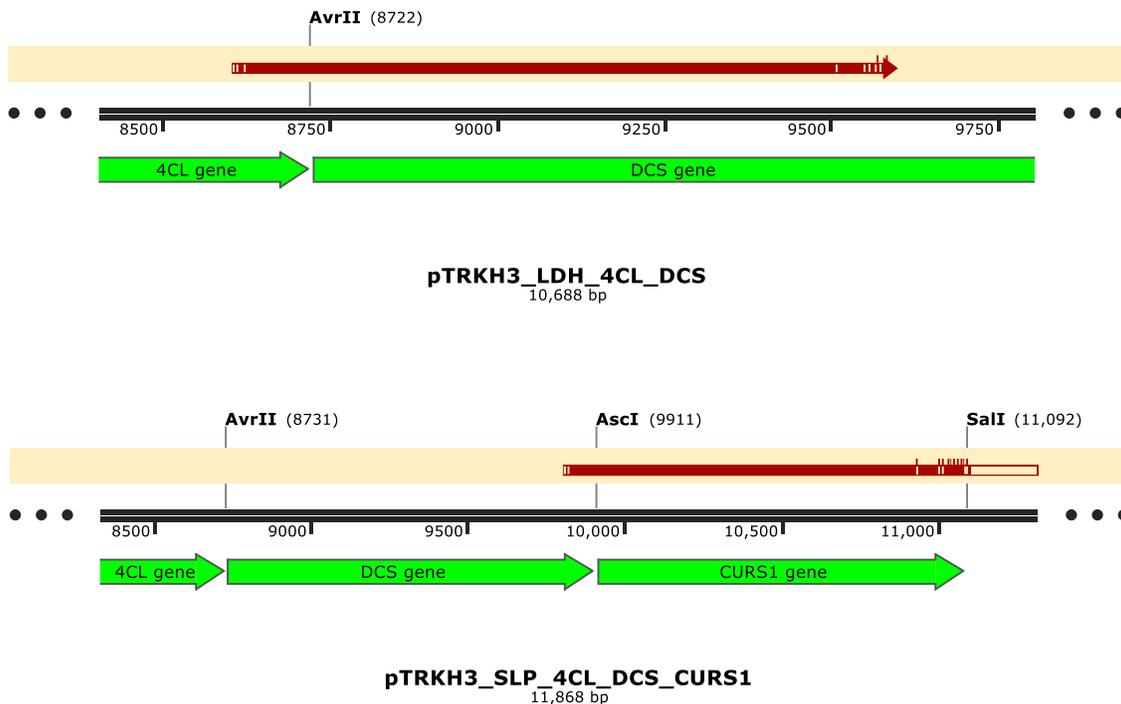


**Figure 5.35.** Alignment performed with the SnapGene software between the expected pTRKH3\_SLP\_4CL\_DCS sequence and the real one obtained by sequencing.

The pTRKH3\_LDH\_4CL\_DCS was harder to get, being necessary several ligation optimization procedures in order to obtain that double gene clone. Figure 5.36 have the gel with the colony PCR result, with the correct 3,510 bp band corresponding to the amplification of the pTRKH3\_LDH\_4CL\_DCS correct clones, in lanes 6 and 8. In the same Figure, on lanes 4 and 5, it's visible the 4,692 band corresponding to the 3 genes SLP clone (pTRKH3\_SLP\_4CL\_DCS\_CURS1), after amplification with Check\_pTRKH3\_F and Ampl\_CURS1\_R primers. The confirmation of both clones by sequencing is represented on Figure 5.37.

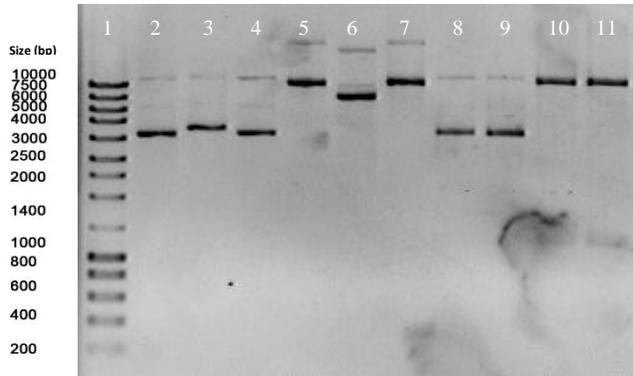


**Figure 5.36.** Confirmation of the pTRKH3\_LDH\_4CL\_DCS and pTRKH3\_SLP\_4CL\_DCS\_CURS1 clones. Lanes: **1-** Nzy ladder III; **2-** pTRKH3\_SLP\_4CL\_DCS\_CURS1, colony 1; **3-** pTRKH3\_SLP\_4CL\_DCS\_CURS1, colony 2; **4-** pTRKH3\_SLP\_4CL\_DCS\_CURS1, colony 3; **5-** pTRKH3\_SLP\_4CL\_DCS\_CURS1, colony 4; **6-** pTRKH3\_LDH\_4CL\_DCS, colony 1; **7-** pTRKH3\_LDH\_4CL\_DCS, colony 2; **8-** pTRKH3\_LDH\_4CL\_DCS, colony 3; **9-** pTRKH3\_LDH\_4CL\_DCS, colony 4

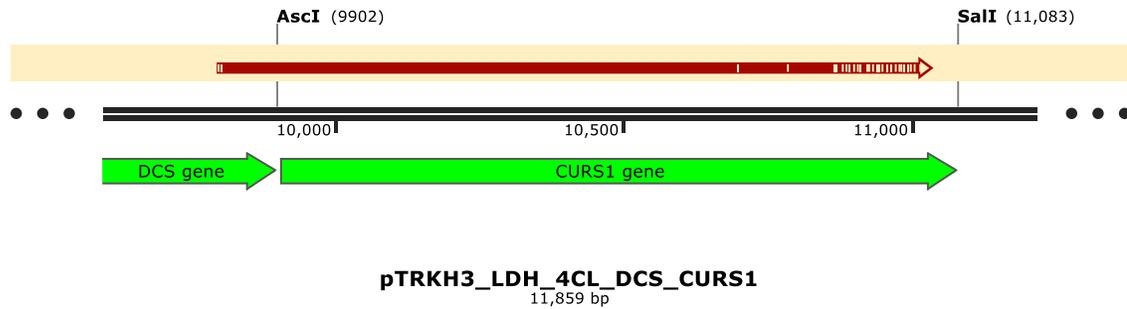


**Figure 5.37.** Alignment performed with the SnapGene software between the expected pTRKH3\_LDH\_4CL\_DCS and pTRKH3\_SLP\_4CL\_DCS\_CURS1 sequences and the real ones obtained by sequencing.

The last necessary clone, pTRKH3\_LDH\_4CL\_DCS\_CURS1, was obtained and confirmed by restriction digestion with AscI and Sall (10,686 + 1,182 bp bands on lane 11) (Figure 5.38) and sequencing (Figure 5.39).



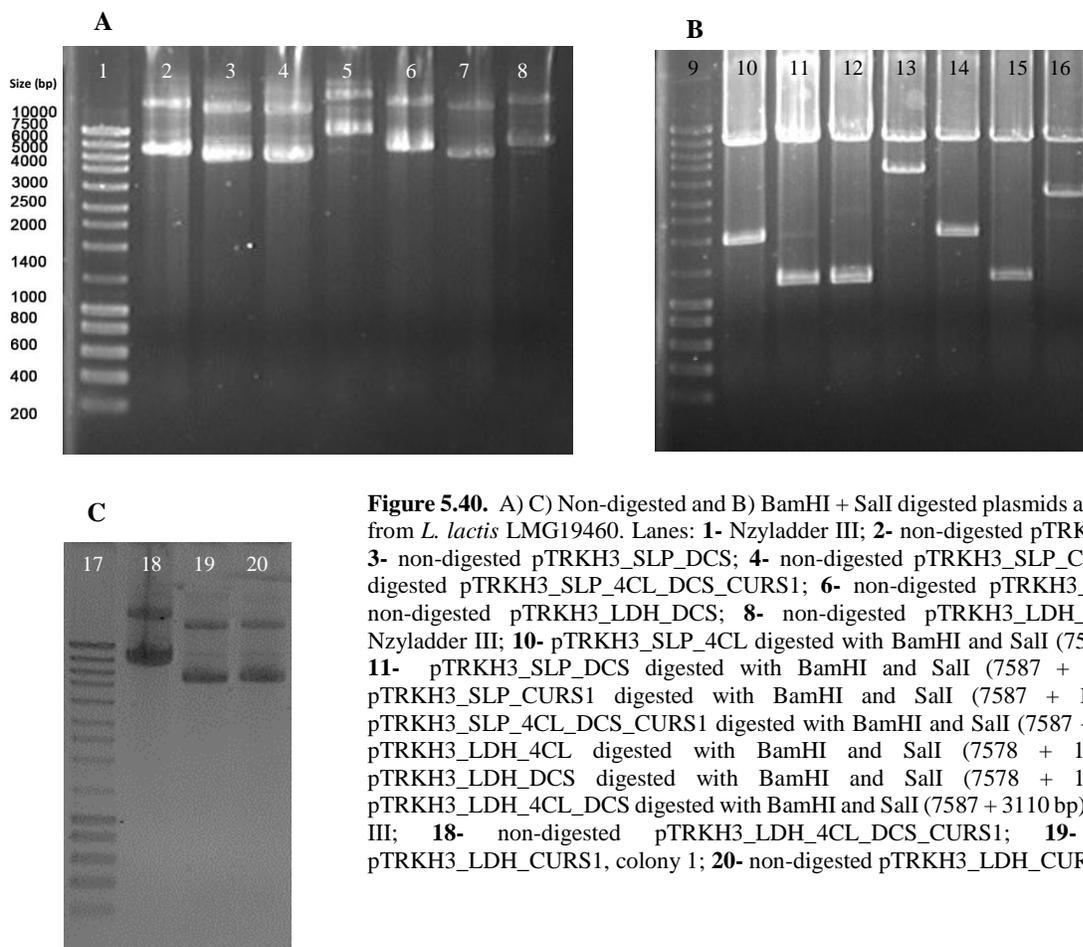
**Figure 5.38.** Confirmation of the pTRKH3\_LDH\_4CL\_DCS\_CURS1 clone. Lanes: **1-** Nzy ladder III; **2 to 4-** results from other experiment; **5-** Non-digested pTRKH3\_LDH\_4CL\_DCS\_CURS1, colony 1; **6-** Non-digested pTRKH3\_LDH\_4CL\_DCS\_CURS1, colony 2; **7-** Non-digested pTRKH3\_LDH\_4CL\_DCS\_CURS1, colony 3; **8 and 9-** results from other experiment; **10-** pTRKH3\_LDH\_4CL\_DCS\_CURS1 digested with AscI and Sall, colony 1; **11-** pTRKH3\_LDH\_4CL\_DCS\_CURS1 digested with AscI and Sall, colony 3.



**Figure 5.39.** Alignment performed with the SnapGene software between the expected pTRKH3\_LDH\_4CL\_DCS\_CURS1 sequence and the real one obtained by sequencing.

#### 5.4.2. LAB transformation

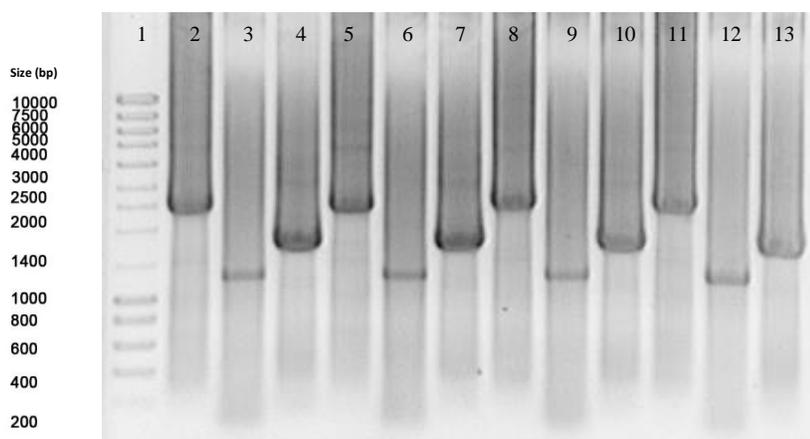
After obtaining all the necessary clones, with one, two or the three genes (*4CL*, *DCS* and *CURS1*) cloned under the control of one of the constitutive promoters (SLP or LDH), the next step was to transform *L. lactis* LMG19460 with each plasmid. Figure 5.40 shows the non-digested, as well as the BamHI + Sall digested pDNAs, which allowed the confirmation that LAB were transformed with the correct plasmids.



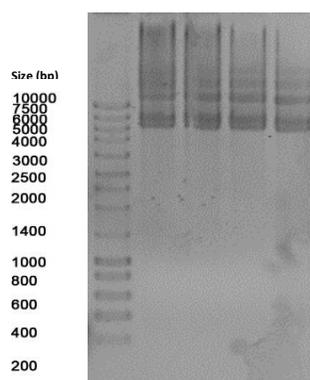
**Figure 5.40.** A) C) Non-digested and B) BamHI + SalI digested plasmids after purification from *L. lactis* LMG19460. Lanes: **1-** Nzy ladder III; **2-** non-digested pTRKH3\_SLP\_4CL; **3-** non-digested pTRKH3\_SLP\_DCS; **4-** non-digested pTRKH3\_SLP\_CURS1; **5-** non-digested pTRKH3\_SLP\_4CL\_DCS\_CURS1; **6-** non-digested pTRKH3\_LDH\_4CL; **7-** non-digested pTRKH3\_LDH\_DCS; **8-** non-digested pTRKH3\_LDH\_4CL\_DCS; **9-** Nzy ladder III; **10-** pTRKH3\_SLP\_4CL digested with BamHI and SalI (7587 + 1929 bp); **11-** pTRKH3\_SLP\_DCS digested with BamHI and SalI (7587 + 1196 bp); **12-** pTRKH3\_SLP\_CURS1 digested with BamHI and SalI (7587 + 1188 bp); **13-** pTRKH3\_SLP\_4CL\_DCS\_CURS1 digested with BamHI and SalI (7587 + 4281 bp); **14-** pTRKH3\_LDH\_4CL digested with BamHI and SalI (7578 + 1929 bp); **15-** pTRKH3\_LDH\_DCS digested with BamHI and SalI (7578 + 1196 bp); **16-** pTRKH3\_LDH\_4CL\_DCS digested with BamHI and SalI (7587 + 3110 bp); **17-** Nzy ladder III; **18-** non-digested pTRKH3\_LDH\_4CL\_DCS\_CURS1; **19-** non-digested pTRKH3\_LDH\_CURS1, colony 1; **20-** non-digested pTRKH3\_LDH\_CURS1, colony 2.

As an alternative strategy, the three single gene clones under the control of the SLP promoter were used to transform simultaneously *L. lactis* LMG19460 cells. The transformation procedure was similar to the one described in the appropriate Materials and Methods section, with the only difference that instead a single plasmid, it was used a mixture of 1,000 ng of each one of the three plasmids (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1) to transform the cells.

The resulting colonies were tested by colony PCR to check for the presence of each plasmid (Figure 5.41). In all the four colonies tested, the 3 plasmids were present, with the correct bands appearing on the gel: 2,331 bp for pTRKH3\_SLP\_4CL, 1,170 bp for pTRKH3\_SLP\_DCS and 1,599 bp for pTRKH3\_SLP\_CURS1. The pDNA purified from the same colonies was visualized on a 1% agarose gel (Figure 5.42), being visible several indistinguishable bands in each lane, corresponding to the different isoforms of each plasmid.

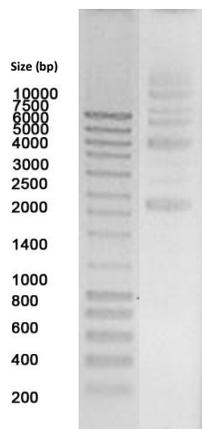


**Figure 5.41.** *L. lactis* LMG19460 transformed with pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1. Lanes: **1-** Nzy ladder III; **2-** colony 1, primer for pTRKH3\_SLP\_4CL; **3-** colony 1, primer for pTRKH3\_SLP\_DCS; **4-** colony 1, primer for pTRKH3\_SLP\_CURS1; **5-** colony 2, primer for pTRKH3\_SLP\_4CL; **6-** colony 2, primer for pTRKH3\_SLP\_DCS; **7-** colony 2, primer for pTRKH3\_SLP\_CURS1; **8-** colony 3, primer for pTRKH3\_SLP\_4CL; **9-** colony 3, primer for pTRKH3\_SLP\_DCS; **10-** colony 3, primer for pTRKH3\_SLP\_CURS1; **11-** colony 4, primer for pTRKH3\_SLP\_4CL; **12-** colony 4, primer for pTRKH3\_SLP\_DCS; **13-** colony 4, primer for pTRKH3\_SLP\_CURS1.



**Figure 5.42.** pDNA purified from *L. lactis* LMG19460 transformed with pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1. Lanes: **1-** Nzy ladder III; **2-** colony 1; **3-** colony 2; **4-** colony 3; **5-** colony 4.

As a positive control, *E. coli* MG1655 harboring the necessary plasmids for curcumin production was also used. This strain was kindly donated by Joana Rodrigues from Universidade do Minho<sup>193</sup>. The strain harbors the plasmids pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1, being necessary to supplement the medium with IPTG in order to induce the expression of each gene. The pDNA purified for this strain is visible on Figure 5.43, with several indistinguishable bands corresponding to the different isoforms of the 3 plasmids.



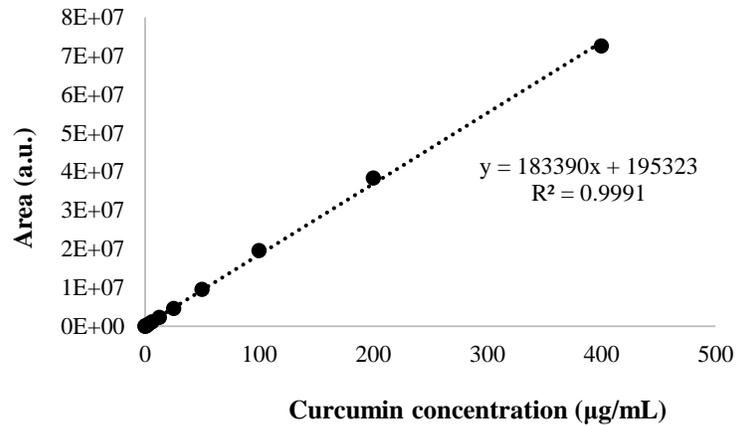
**Figure 5.43.** Lanes: 1- Nzyladder III; 2- pDNA purified from *E. coli* MG1655 harboring the plasmids pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1.

### 5.4.3. Curcumin production and quantification

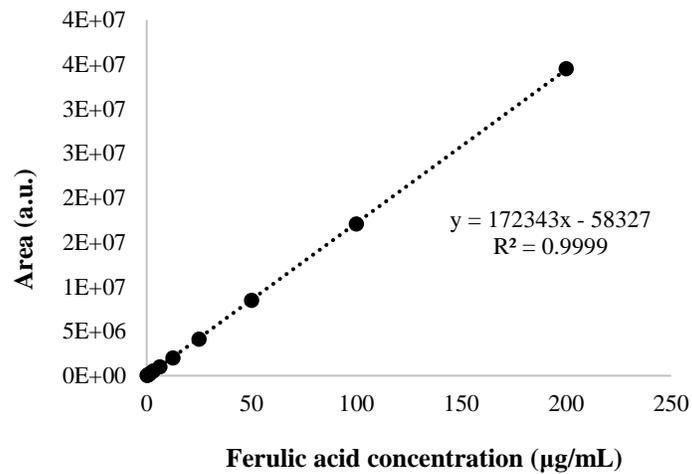
#### 5.4.3.1. Curcumin and ferulic acid calibration curves

After optimizing the HPLC running conditions, the curcumin and ferulic acid calibration curves were performed. To obtain the curcumin calibration curve, the following standards were diluted with acetonitrile and analyzed in the HPLC: 400, 200, 100, 50, 25, 12.5, 6.25, 3.125, 1.5625, 0.78125, 0.390625, 0.195313 and 0.097657  $\mu\text{g}/\text{mL}$ . After calculating the area (and the corresponding retention time), this parameter was plotted against the concentration to obtain the curcumin calibration curve and its equation (Figure 5.44). The average retention time determined for curcumin was 9.284 min.

The same was performed to ferulic acid, using the following serial dilutions of the 10 mg/mL stock solution prepared with acetonitrile: 200, 100, 50, 25, 12.5, 6.25, 3.125, 1.5625, 0.78125, 0.390625 and 0.195313  $\mu\text{g}/\text{mL}$  (Figure 5.45). The average retention time determined for ferulic acid was 2.598 min.

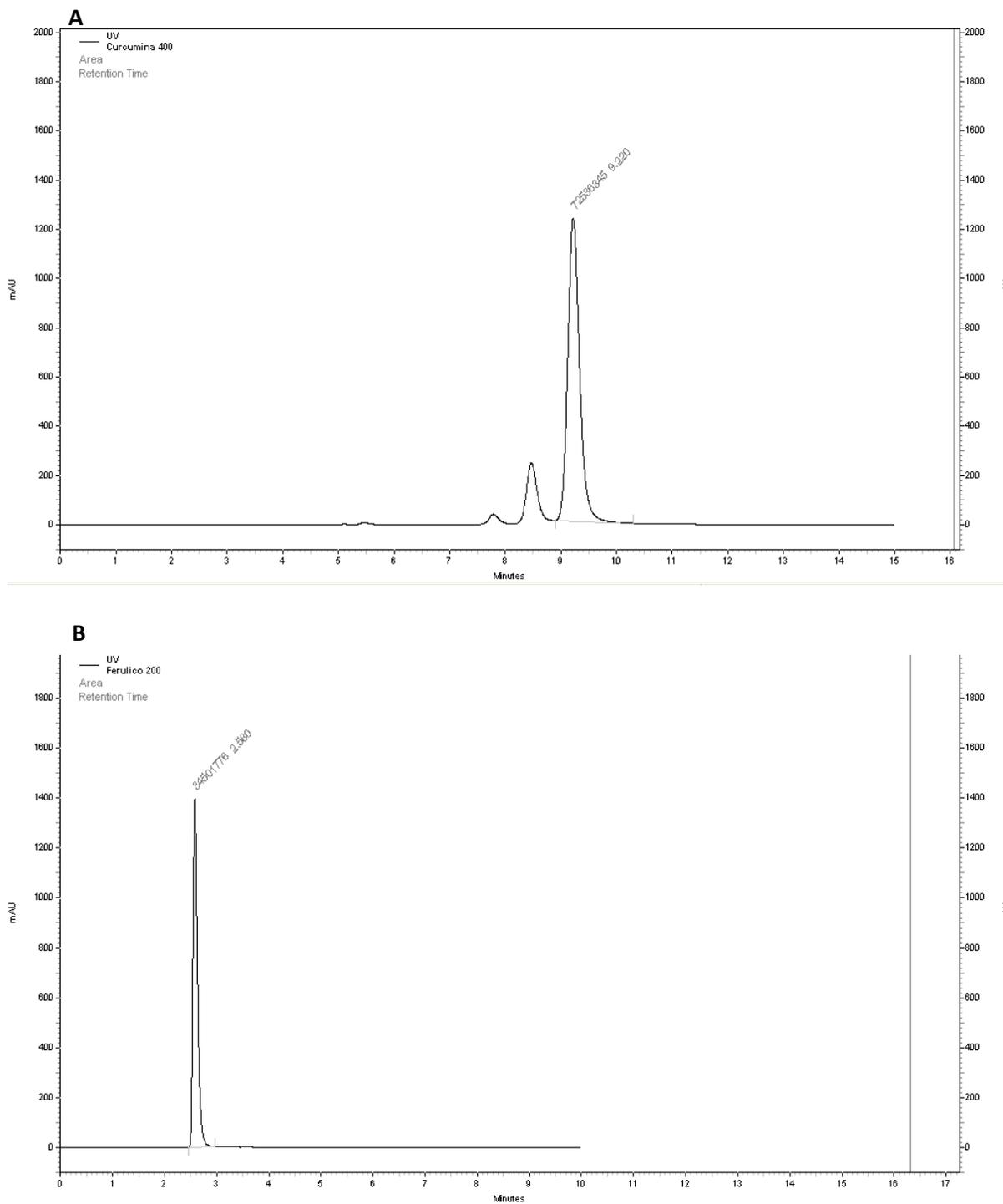


**Figure 5.44.** Curcumin standard curve and respective equation.



**Figure 5.45.** Ferulic acid standard curve and respective equation.

The chromatograms obtained for the highest concentrations of curcumin (400  $\mu\text{g/mL}$ ) and ferulic acid (200  $\mu\text{g/mL}$ , corresponding to 1.03 mM) are represented in Figure 5.46.



**Figure 5.46.** Chromatograms obtained for the highest concentrations of A) curcumin (400  $\mu\text{g/mL}$ ) and B) ferulic acid (200  $\mu\text{g/mL}$ , corresponding to 1.03 mM).

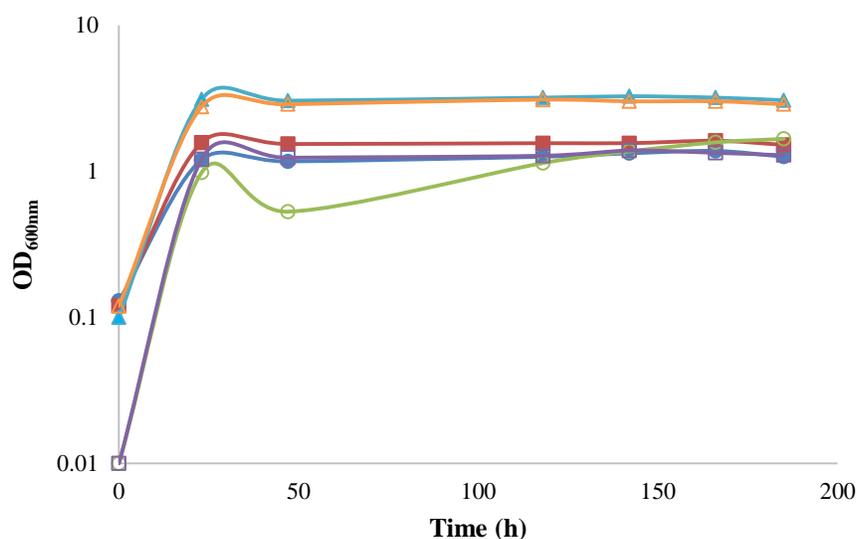
### 5.4.3.2. Optimization of the growth conditions

Several growth conditions were tested in order to obtain quantifiable curcumin production in *L. lactis* LMG19460, namely the starting OD, media used, lysozyme pre-treatment, composition of the ferulic acid solution and temperature.

*L. lactis* LMG19460 harboring the pTRKH3\_SLP\_4CL\_DCS\_CURS1 plasmid, which has the three necessary genes for curcumin production, was grown in parallel with the same strain containing the negative control plasmid pTRKH3 and the *E. coli* DH5 $\alpha$  strains containing the same plasmids. For the *L. lactis* strains, two different growth conditions were tested, starting with OD of 0.1 versus OD of 0.01<sup>203</sup> in the shake flasks with 75 mL of modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>). The growth was followed for 8 days, being taken samples for OD measurements and HPLC quantification.

The growth curves are represented in Figure 5.47, being noticeable that *E. coli* DH5 $\alpha$  reached approximately the double of the OD when compared with *L. lactis* LMG19460 strains. The conditions where the initial OD was 0.01 reached slightly lower OD values after 24h of growth, compared with the initial OD of 0.1 counterparts. Due to that fact, the subsequent experiments were performed using modified M9 minimal salt medium and started with an OD=0.1. *L. lactis* LMG19460 with the curcumin producing plasmid reached higher specific growth rates and higher maximum ODs, when compared with the negative control with pTRKH3 plasmid alone, showing that a bigger plasmid was not impairing cell growth by increasing the cells metabolic burden.

All the strains reached its maximum OD after around 24h, maintaining a stationary phase in the remaining collection points.



**Figure 5.47.** Growth curves from the first curcumin production experiment, with *L. lactis* LMG19460 harboring pTRKH3\_SLP\_4CL\_DCS\_CURS1 (closed squares for initial OD of 0.1 and open squares for initial OD of 0.01), or the negative control pTRKH3 (closed circles for initial OD of 0.1 and open circles for initial OD of 0.01). The *E. coli* DH5 $\alpha$  growth curves are represented with triangles, being closed when harbors pTRKH3 and open when harbors pTRKH3\_SLP\_4CL\_DCS\_CURS1.

The samples were analyzed by HPLC, following the protocol indicated in the Materials and Methods section 5.3.7, but no curcumin was detected in any sample. Considering the ferulic acid analysis, at time 0h the amount was above the detection limit (116.4  $\mu\text{g/mL}$ ), meaning that the analytic procedure was correctly detecting the 2 mM (388.4  $\mu\text{g/mL}$ ) ferulic acid that was added to each growth medium. After 5 days of growth, the same pattern appeared on DH5 $\alpha$  samples, but smaller peaks appeared on LMG19460 samples together with a second peak with a higher retention time, meaning that *L. lactis* LMG19460 is consuming ferulic acid and producing an unidentified metabolite. There are several reports from LAB being able to metabolize ferulic acid, producing other compounds. For example, *L. plantarum* is able to convert ferulic acid into 4-vinylguaiacol and hydroferulic acid<sup>204</sup>.

To discard the hypothesis that absence of curcumin was related to the extraction process, due to incomplete Gram-positive cell lysis, an upstream step of 30 min incubation with 10 mg/mL lysozyme was implemented before the ethyl acetate extraction process. The HPLC analysis did not detect any curcumin again, meaning that the problem is not in the extraction process, but probably in the production by the cells itself.

Further growth experiments were idealized using higher concentrations of ferulic acid with the goal of increasing the curcumin production, by supplementing an excess of substrate, but a first test with 20 mM (3,884  $\mu\text{g/mL}$ ) ferulic acid was performed without success. Since ferulic acid solutions should be prepared in ethanol, due to its low solubility in aqueous solutions, one of the main concerns were that the excess ethanol in the medium could impair the cell growth. In order to try to overcome that problem, the ferulic acid was incorporated as powder during media preparation, avoiding the presence of ethanol. The Figure 5.48 shows the appearance of the media with 20 mM (3,884  $\mu\text{g/mL}$ ) versus 2 mM (388.4  $\mu\text{g/mL}$ ) of ferulic acid. But after 5 days there were no growth in the 2 mM nor in the 20 mM condition, showing that just solubilizing the ferulic acid by heat (autoclave) is not a solution.



**Figure 5.48.** Modified M9 minimal salt medium (without thiamine, trace elements or  $\text{CaCO}_3$ ) autoclaved with 20 mM (3,884  $\mu\text{g/mL}$ ) and 2 mM of ferulic acid (388.4  $\mu\text{g/mL}$ ).

Since the ferulic acid should be solubilized in ethanol, the resistance of *L. lactis* LMG19460 with pTRKH3 to different ethanol concentrations (0%, 0.8% and 1.6%, which corresponds to the 0, 2 and 4 mM of ferulic acid solutions, respectively) was tested for cells growing in M17 and modified M9 minimal salt medium (without thiamine, trace elements or  $\text{CaCO}_3$ ), as well as at 30°C and also at 25°C, since lower temperatures are considered better when the goal is to produce proteins in high quantity and quality<sup>205</sup>. After overnight growth, starting with the same volume from a -80°C cell bank, the achieved ODs are represented in table 5.5.

**Table 5.5.** ODs achieved by *L. lactis* LMG19460 with pTRKH3 after overnight growth in M17 or modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>), at 25°C or 30°C, with different ethanol concentrations.

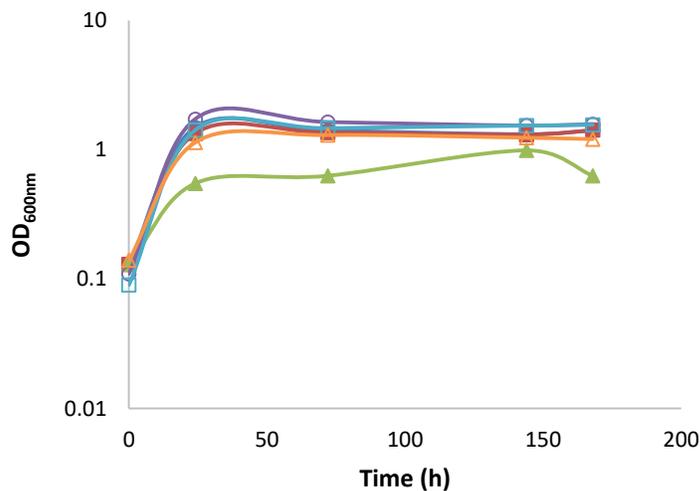
Medium	OD <sub>600nm</sub> at 25°C	OD <sub>600nm</sub> at 30°C
M17, 0% ethanol	1.69	1.55
M17, 0.8% ethanol	1.83	1.48
M17, 1.6% ethanol	1.73	1.57
M9, 0% ethanol	1.21	1.30
M9, 0.8% ethanol	1.13	1.15
M9, 1.6% ethanol	1.07	0.89

From the analysis of the Table 5.5, it is noticeable that *L. lactis* LMG19460 with pTRKH3 had grown better in M17 than in modified M9 minimal salt medium. In the range of the tested ethanol concentrations, the growth was not affected significantly, except when using 1.6% ethanol in modified M9 minimal salt medium, at 30°C.

#### 5.4.3.3. Induction of curcumin production using different ferulic acid concentrations

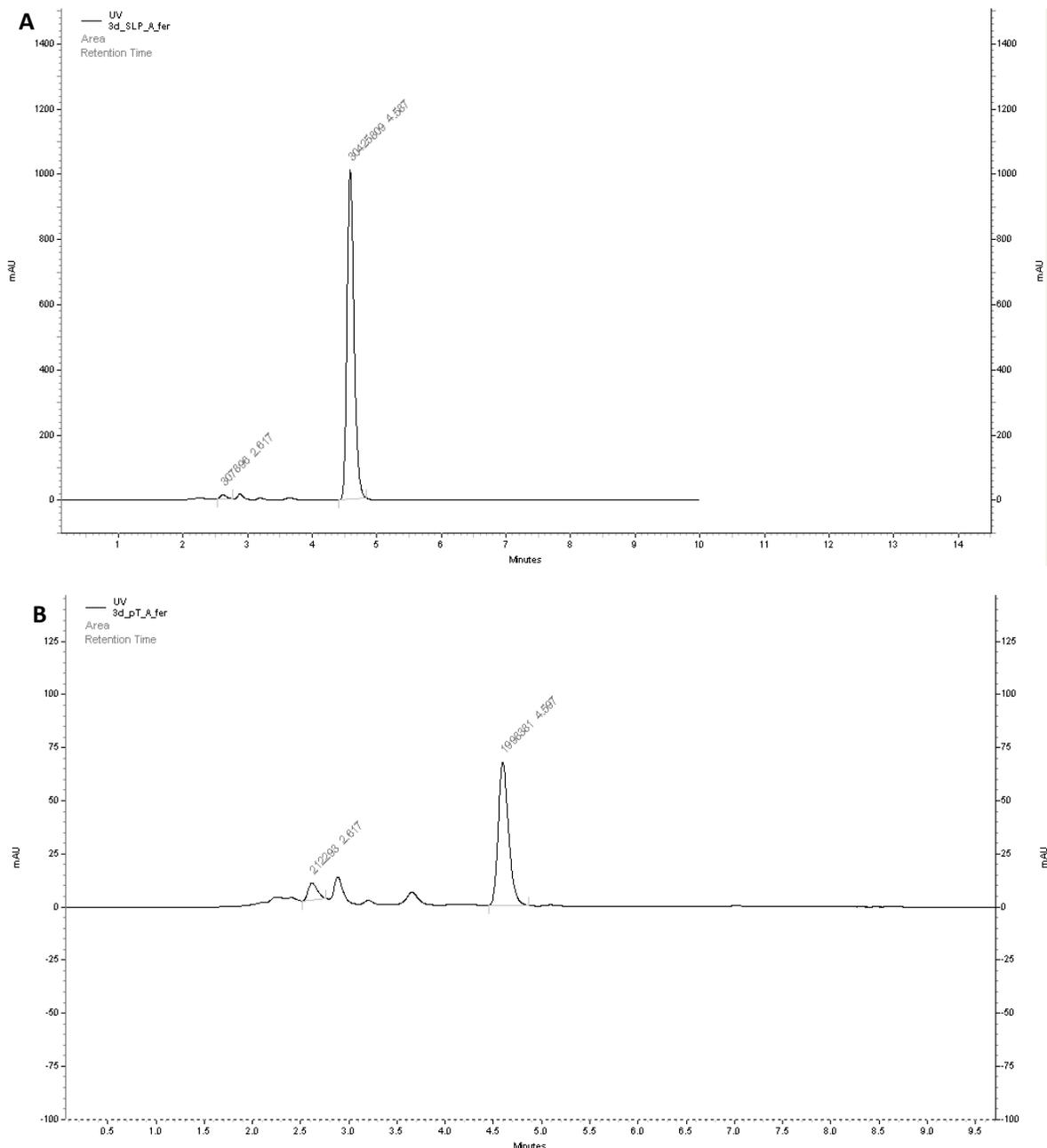
Based on the previous results, the next growth experiment to determine which ferulic acid concentration is best to induce the curcumin production was performed at 25°C because of its beneficial effect in the protein production and with the modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>), since this medium was already described as appropriate for protein production<sup>203</sup>, although did not reach the higher OD values.

*L. lactis* LMG19460 with pTRKH3 and pTRKH3\_SLP\_4CL\_DCS\_CURS1 were grown in the presence of 0, 2 or 4 mM of ferulic acid, corresponding to 0, 388.4 and 3,884 µg/mL, and samples were taken during a 7 days period. The measured growth curves are depicted in Figure 5.49.



**Figure 5.49.** Growth curves from the curcumin production experiment at 25°C in M9 minimal salt medium, with *L. lactis* LMG19460 harboring pTRKH3\_SLP\_4CL\_DCS\_CURS1 (open symbols) or the negative control pTRKH3 (closed symbols). Both were grown in the presence of three different ferulic acid concentrations: 0 mM (circles), 2mM (squares) or 4 mM (triangles).

Considering the 4 mM ferulic acid condition, *L. lactis* LMG19460 harboring pTRKH3\_SLP\_4CL\_DCS\_CURS1 seems much more resistant to 4 mM ferulic acid concentration (1.6% ethanol) than the negative control with only the pTRKH3. Even in the remaining ferulic acid conditions (0 and 2 mM), the clone with the curcumin producing genes achieved always higher specific growth rates, than the negative control counterpart. For both clones, the specific growth rates and the maximum ODs achieved decreased with increasing ethanol concentrations, as expected.



**Figure 5.50.** Chromatograms obtained for ferulic acid detection after 3 days of growth, in *L. lactis* LMG19460 harboring A) pTRKH3\_SLP\_4CL\_DCS\_CURS1 or B) pTRKH3.

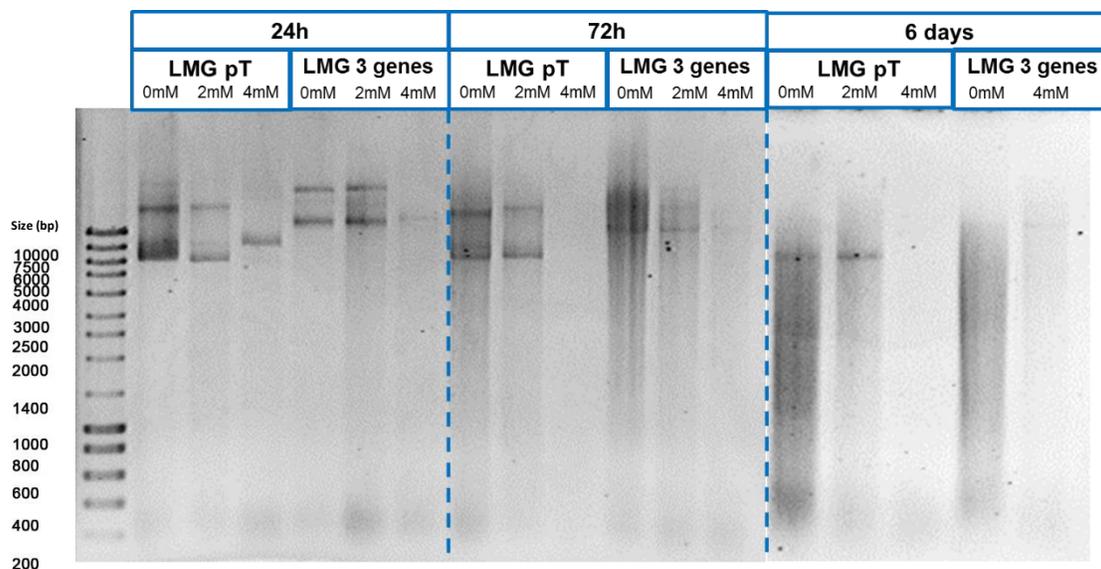
The HPLC samples analysis had shown a very similar pattern to what was obtained in the previous growth experiment, with no curcumin being detected. Concerning the ferulic acid detection, at time 0h the amount was above the detection limit, while after 1 day, 3 days and 6 days, other peaks appeared together again with the second peak with a higher retention time. The chromatograms after 3 days of growth are represented in Figure 5.50 and clearly shows that the ferulic acid peak is much smaller than the second peak detected at 4.587 min, for *L. lactis* LMG19460 harboring pTRKH3\_SLP\_4CL\_DCS\_CURS1 and also for the strain with the wild-type pTRKH3. Again, it seems that *L. lactis* LMG19460 is consuming ferulic acid and producing an unidentified metabolite.

In order to find a plausible explanation for the absence of the production of curcumin by *L. lactis* LMG19460 pTRKH3\_SLP\_4CL\_DCS\_CURS1, the pDNA or gene level was first investigated, to check if the prolonged growth or even the ferulic acid concentration (and consequently the ethanol concentration) could be decreasing the pDNA quality and consequently the curcumin production.

The impact of growth time and of the different ferulic acid concentrations in the pDNA quantities (Table 5.6) and quality (Figure 5.51) was assessed by pDNA purification from a fixed volume of cells.

**Table 5.6.** pDNA concentrations and specific yields after purification of 3 mL samples of *L. lactis* LMG19460 pTRKH3 and *L. lactis* LMG19460 pTRKH3\_SLP\_4CL\_DCS\_CURS1, grown in the presence of 0, 2 or 4 mM ferulic acid. The samples were taken after 24h, 3 days or 6 days of growth.

Sample	pDNA concentration (ng/μL)	pDNA specific yield (ng/μL/OD)
24h, <i>L. lactis</i> LMG19460 pTRKH3, 0 mM ferulic acid	182	122
2 mM ferulic acid	46	34
4 mM ferulic acid	73	133
24h, <i>L. lactis</i> LMG19460 pTRKH3_SLP_4CL_DCS_CURS1, 0 mM ferulic acid	51	29
2 mM ferulic acid	120	85
4 mM ferulic acid	72	63
3 days, <i>L. lactis</i> LMG19460 pTRKH3, 0 mM ferulic acid	280	199
2 mM ferulic acid	133	98
4 mM ferulic acid	52	83
3 days, <i>L. lactis</i> LMG19460 pTRKH3_SLP_4CL_DCS_CURS1, 0 mM ferulic acid	501	305
2 mM ferulic acid	154	105
4 mM ferulic acid	89	68
6 days, <i>L. lactis</i> LMG19460 pTRKH3, 0 mM ferulic acid	340	258
2 mM ferulic acid	145	111
4 mM ferulic acid	46	46
6 days, <i>L. lactis</i> LMG19460 pTRKH3_SLP_4CL_DCS_CURS1, 0 mM ferulic acid	470	305
2 mM ferulic acid	-	-
4 mM ferulic acid	84	68



**Figure 5.51.** Agarose gel with 10  $\mu$ L pDNA samples purified from *L. lactis* LMG19460 pTRKH3 and *L. lactis* LMG19460 pTRKH3\_SLP\_4CL\_DCS\_CURS1, grown in the presence of 0, 2 or 4 mM ferulic acid. The samples were taken after 24h, 3 days or 6 days of growth.

From the concentrations alone, it is visible that, as time progresses, increased ferulic acid concentrations in the medium seems to be responsible for a decrease in pDNA concentration and specific yield. Analyzing the agarose gel (Figure 5.51) it is clear that after 24h of growth all pDNA samples showed good quality, but after 72h (3 days) that good quality only remains for 0 and 2 mM ferulic acid conditions, for both clones. After 6 days, the pDNA quality drastically decreased in all conditions. These results showed that, considering the pDNA alone, at least it should be possible to detect curcumin production after 24h or 3 days, in the 2mM ferulic acid condition, which was not confirmed. This means that the problem should be in another level beyond the gene level.

The next approach was to quantify the mRNA from the *CURS1* gene present in the same samples (Table 5.7). It was chosen the *CURS1* gene instead the more upstream ones, to guarantee that all the genes in the operon were being correctly transcribed, but of course does not gives information about the translation/protein level.

The results (Table 5.7.) indicated that *CURS1* gene was being transcribed in the *L. lactis* LMG19460 pTRKH3\_SLP\_4CL\_DCS\_CURS1 clone, but not in the *L. lactis* LMG19460 pTRKH3, as expected, since lower CP values means more cDNA (and consequently more mRNA) being detected. A negative control without cDNA was also analyzed and no *CURS1* was detected. These results indicate that the problem is not in the transcription level, meaning that something should be impairing the translation level (protein folding or protein stability).

**Table 5.7.** *CURS1* mRNA quantification by real-time quantitative PCR. The Crossing Point (CP) values, as well as the melting temperature (T<sub>m</sub>) values are presented).

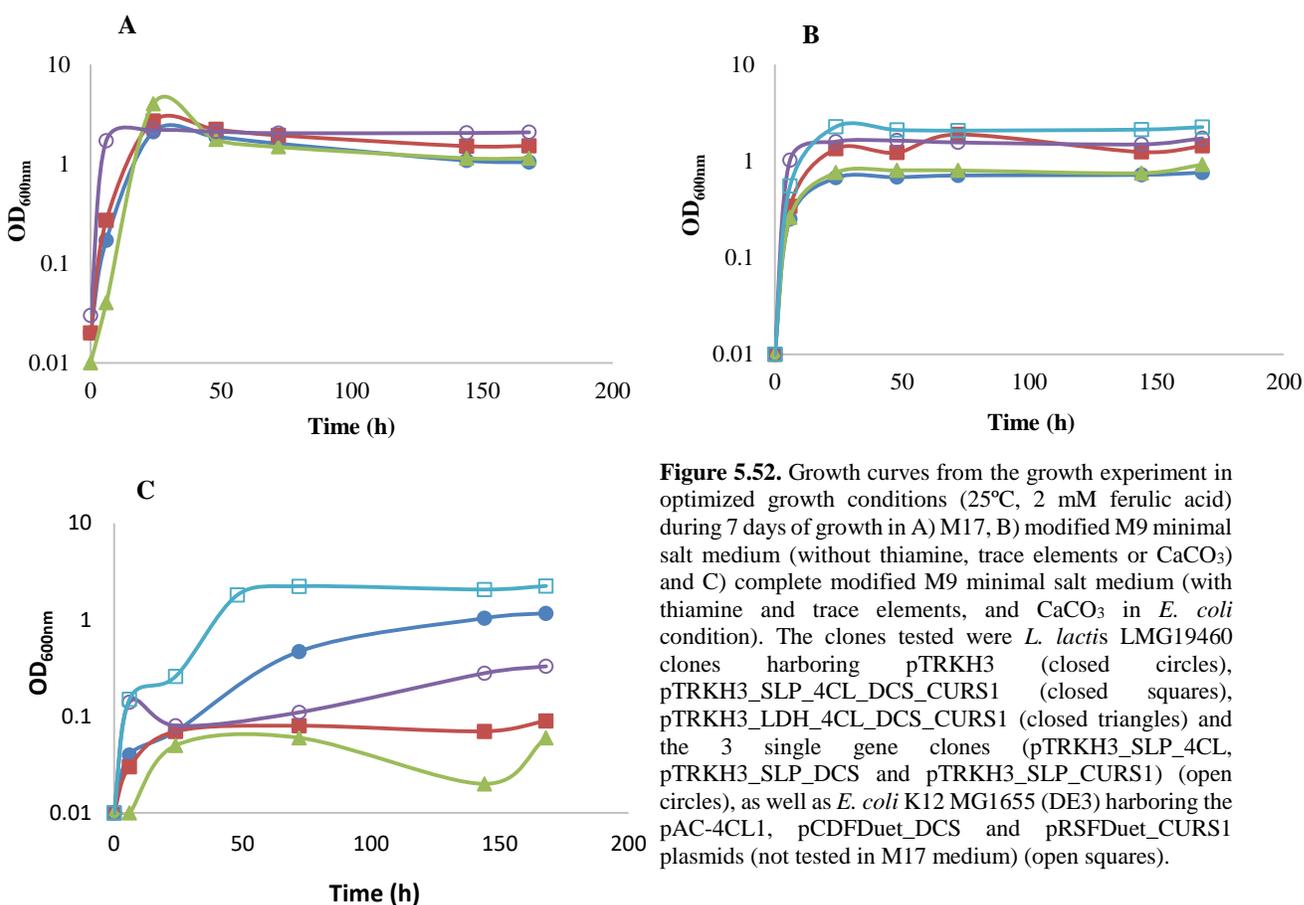
	<b>CP</b>	<b>T<sub>m</sub></b>
<i>L. lactis</i> LMG19460 pTRKH3, 48h	32.0	83.8
<i>L. lactis</i> LMG19460 pTRKH3, 72h	29.9	83.7
<i>L. lactis</i> LMG19460 pTRKH3_SLP_4CL_DCS_CURS1, 48h	14.5	83.6
<i>L. lactis</i> LMG19460 pTRKH3_SLP_4CL_DCS_CURS1, 72h	13.3	83.7
Negative control	30.2	83.9

#### 5.4.3.4. Induction of curcumin production using the optimized growth conditions

A growth experiment using the optimized growth conditions (25°C, induction with 2 mM ferulic acid) was performed as described in the Materials and Methods section 5.3.5, with the only difference that the *L. lactis* LMG19460 inoculum was grown in 3 different media to test if any could increase curcumin production. The first medium was M17 supplemented with glucose, erythromycin and ferulic acid, being chosen because it is the most widely used medium for growing LAB. The second medium was the modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>) with the same supplementation, since this medium was already described as appropriate for protein production<sup>203</sup>. And the third medium was the complete modified M9 minimal salt medium (with thiamine, trace elements, erythromycin and ferulic acid). For the *L. lactis* LMG19460 the CaCO<sub>3</sub> (for pH control) was not supplemented to the media, since it was already previously tested that it completely inhibited the LAB growth.

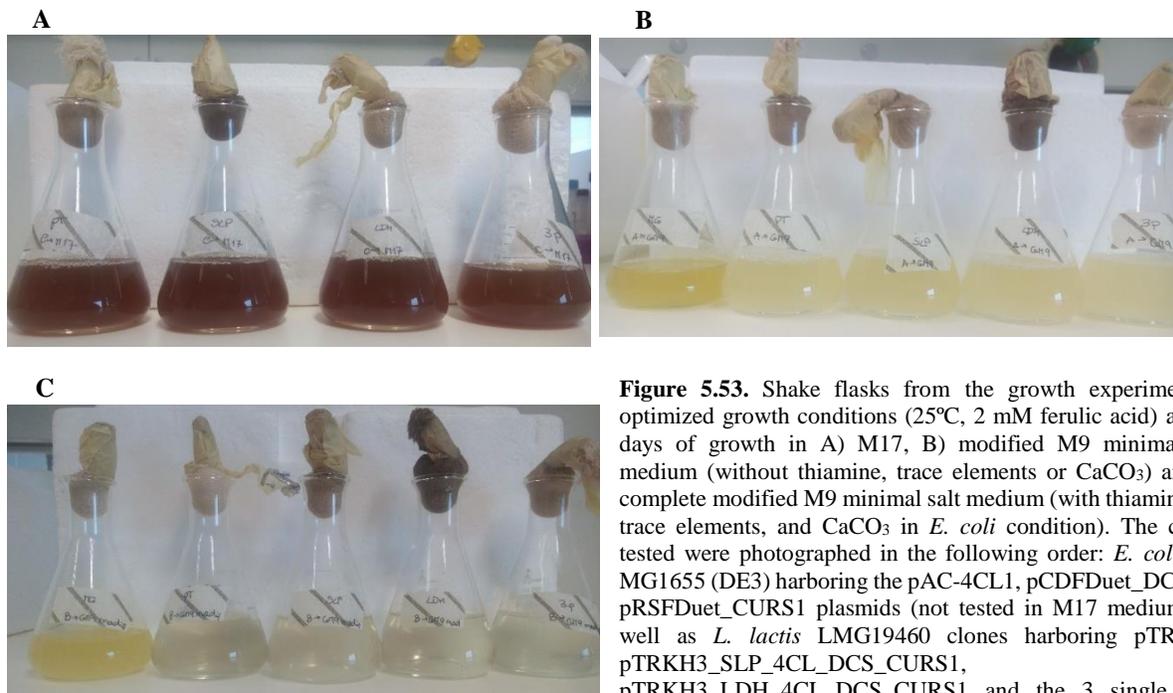
The *L. lactis* LMG19460 clones tested in this experiment were the negative control pTRKH3, pTRKH3\_SLP\_4CL\_DCS\_CURS1, pTRKH3\_LDH\_4CL\_DCS\_CURS1 and the strain transformed with the 3 single gene clones (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1). In parallel, the positive control *E. coli* K12 MG1655 (DE3) harboring the pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1 was grown in modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>) and in complete modified M9 minimal salt medium (with thiamine, trace elements, CaCO<sub>3</sub>, erythromycin and ferulic acid) as described with more details in Materials and Methods, section 5.3.5. For *E. coli* did not make sense to grown in M17 medium, since is a medium optimized for LAB. The growth curves of all the strains in the different media can be consulted in Figure 5.52 and the visual appearance of the shake flasks are depicted on Figure 5.53.

The analysis of the growth curves showed that *L. lactis* LMG19460 clones were able to achieve higher ODs in the M17 medium, while having the lower ones in complete modified M9 minimal salt medium (with thiamine and trace elements). *E. coli* K12 MG1655 (DE3) harboring the pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1 always achieved higher ODs when compared with the *L. lactis* clones.



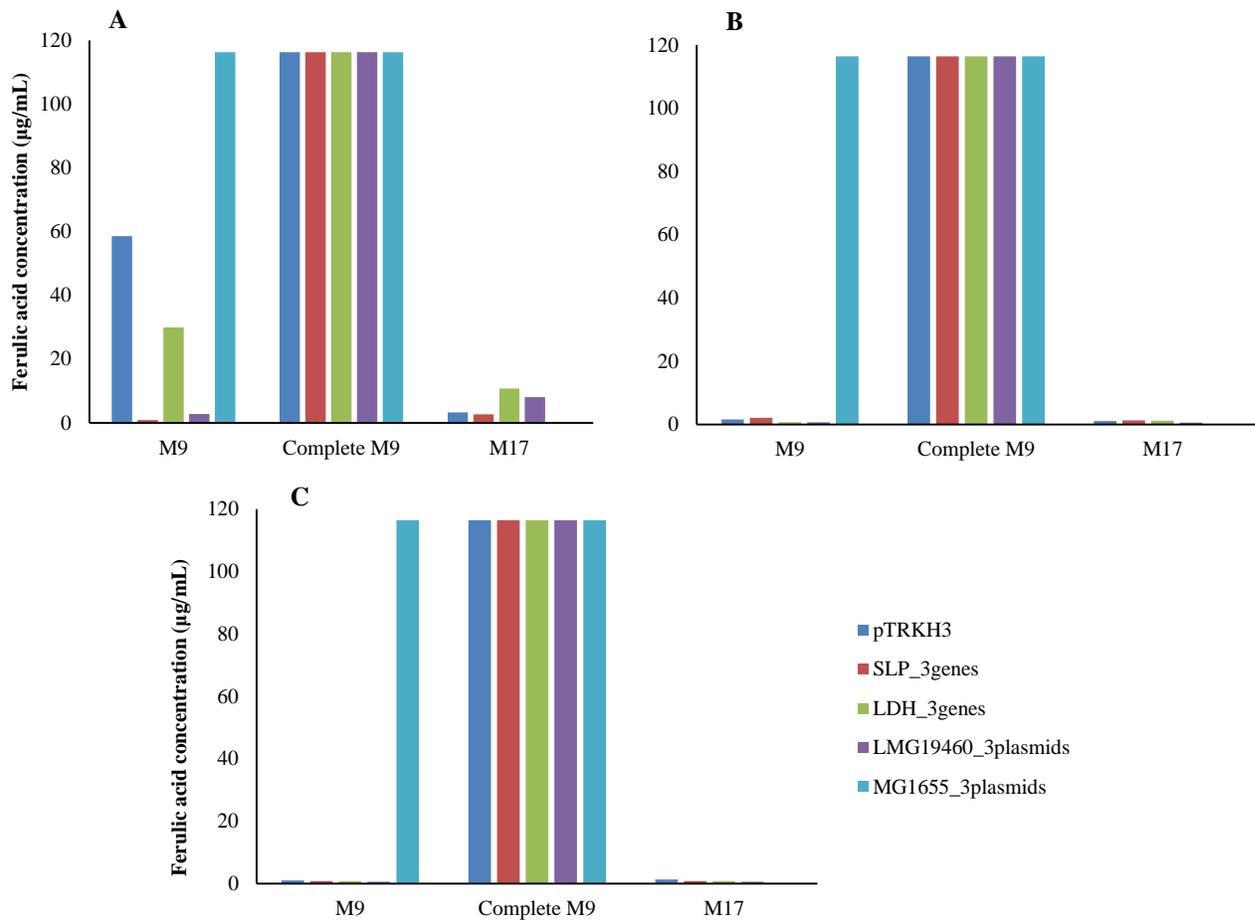
**Figure 5.52.** Growth curves from the growth experiment in optimized growth conditions (25°C, 2 mM ferulic acid) during 7 days of growth in A) M17, B) modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>) and C) complete modified M9 minimal salt medium (with thiamine and trace elements, and CaCO<sub>3</sub>) in *E. coli* condition). The clones tested were *L. lactis* LMG19460 clones harboring pTRKH3 (closed circles), pTRKH3\_SLP\_4CL\_DCS\_CURS1 (closed squares), pTRKH3\_LDH\_4CL\_DCS\_CURS1 (closed triangles) and the 3 single gene clones (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1) (open circles), as well as *E. coli* K12 MG1655 (DE3) harboring the pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1 plasmids (not tested in M17 medium) (open squares).

While solely looking at the growth curves does not give any indication about curcumin production, the analysis of the photographs taken to the shake flasks were more elucidative (Figure 5.53). It is clearly visible a strong yellow color in both shake flasks containing *E. coli* K12 MG1655 (DE3) with the pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1 plasmids, indicating that at least these two conditions should have detectable curcumin production.



**Figure 5.53.** Shake flasks from the growth experiment in optimized growth conditions (25°C, 2 mM ferulic acid) after 7 days of growth in A) M17, B) modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>) and C) complete modified M9 minimal salt medium (with thiamine and trace elements, and CaCO<sub>3</sub> in *E. coli* condition). The clones tested were photographed in the following order: *E. coli* K12 MG1655 (DE3) harboring the pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1 plasmids (not tested in M17 medium), as well as *L. lactis* LMG19460 clones harboring pTRKH3, pTRKH3\_SLP\_4CL\_DCS\_CURS1, pTRKH3\_LDH\_4CL\_DCS\_CURS1 and the 3 single gene clones (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1).

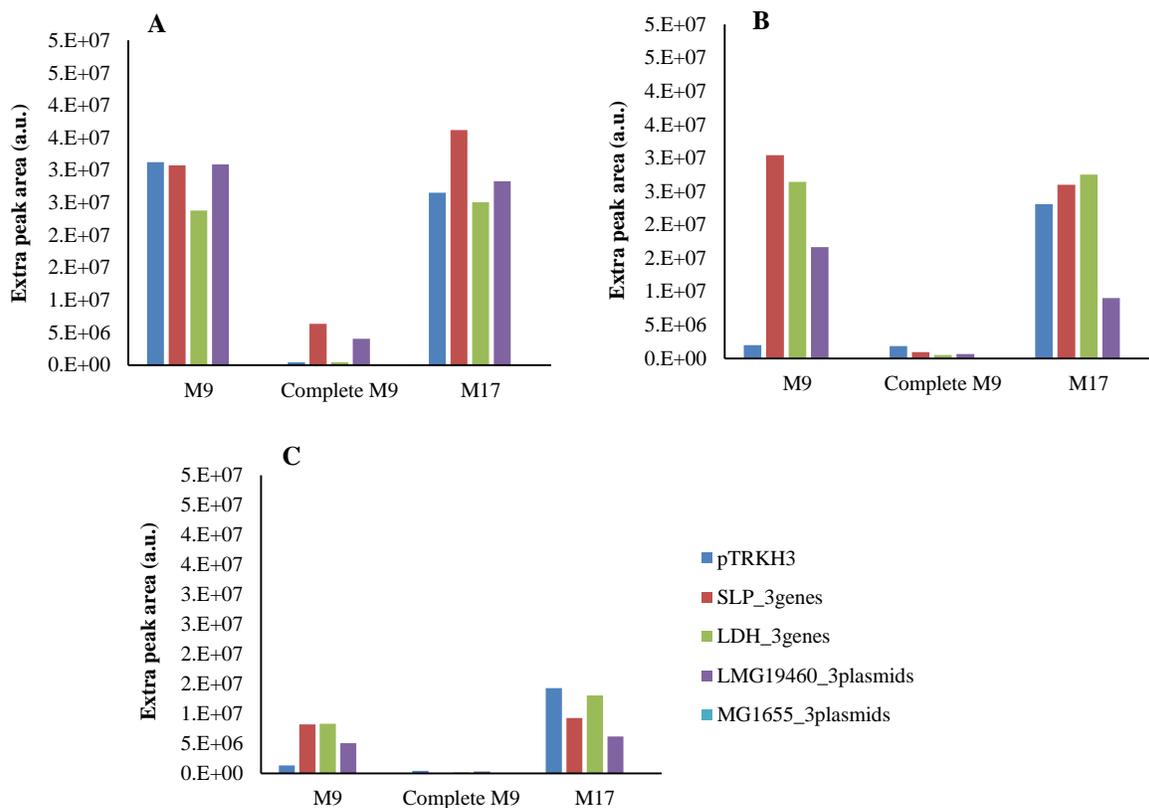
The next step was to extract and quantify the curcumin, as well as the ferulic acid, of each growth sample, using HPLC. Starting with the analysis of the ferulic acid quantification (Figure 5.54) it is immediately noticeable that in complete modified M9 minimal salt medium (with thiamine, trace elements, CaCO<sub>3</sub>, erythromycin and ferulic acid) the ferulic acid concentration never decreased along the 7 days of growth, remaining above the HPLC equipment detection limit (200 µg/mL or 1.03 mM) for all the analyzed samples. The y axis was deliberately truncated by the 120 µg/mL level to make it easier to visualize the smaller concentrations. These results are coherent with the *L. lactis* growth curve results, where all the strains registered a small or even inexistent growth in this media, with only *E. coli* MG1655 having a normal growth curve. This means that since the *L. lactis* cells were not growing, they were not consuming the ferulic acid, at least at a detectable rate. For *L. lactis* samples, in the remaining media tested (M17 and modified M9 minimal salt medium, without thiamine, trace elements or CaCO<sub>3</sub>), the ferulic acid concentrations decreased after 24h and were almost exhausted after 3 days of growth, being in agreement with the higher specific growth rates registered in these medium. In the *E. coli* MG1655 specific case, this strain was not tested in M17 media, since it is specific for LAB growth, but in the modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>) and in the complete M9 minimal salt medium (with thiamine and trace elements, and CaCO<sub>3</sub>) the ferulic acid concentration never decreased, remaining above the detection limit during the 7 days.



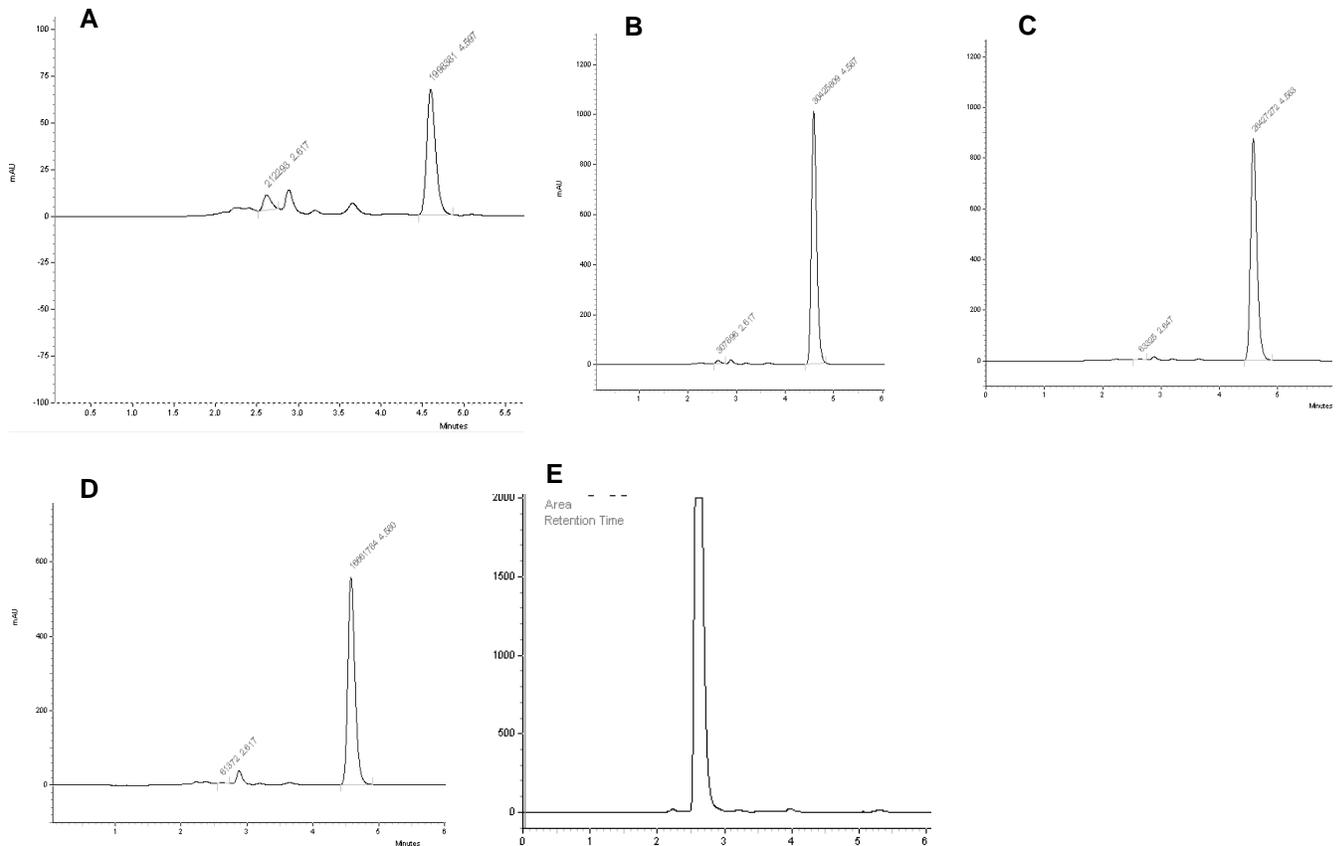
**Figure 5.54.** Ferulic acid quantification by HPLC of the samples from the third growth experiment, after A) 24h of growth, B) 72h of growth and C) 7 days of growth. The data for the 3 media used (M9, complete M9 and M17) are presented. The strains tested were *L. lactis* LMG19460 harboring the pTRKH3 plasmid (dark blue columns), *L. lactis* LMG19460 harboring the pTRKH3\_SLP\_4CL\_DCS\_CURS1 (red columns), *L. lactis* LMG19460 harboring the pTRKH3\_LDH\_4CL\_DCS\_CURS1 (green columns), *L. lactis* LMG19460 with the 3 single-gene plasmids (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1) (purple columns) and *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1) (light blue columns). All the samples were the columns reached the top of the graph means that it exceeded the equipment detection limit. The yy axis was deliberately truncated by the 120 µg/mL level to make it easier to visualize the smaller concentrations.

The analysis of the Figure 5.55 results gives a plausible explanation for the *L. lactis* LMG19460 ferulic acid patterns. In the ferulic acid chromatogram (Figure 5.50), besides the ferulic acid peak (average retention time of 2.59 min) it appeared a second peak with a higher retention time (retention time of 4.59 min). After 24h of growth, everytime the ferulic acid peak was high, this second peak was low and the opposite also appeared. This could only mean that *L. lactis* LMG19460, but not *E. coli* MG1655 were converting the ferulic acid in an unknown metabolite, which were detectable with the same HPLC conditions than ferulic acid. After 3 days and 7 days of growth, the area of the second peak decreased in all the samples, which could indicate metabolite degradation or consumption (Figure 5.55). These

areas could not be converted to concentrations, since it was not determined the identity of the unknown metabolite in order to be possible to calculate a calibration curve. There are some reports that LAB, mainly from the *Lactobacillus* genera, are able to metabolize ferulic acid during growth, converting it into 4-vinylguaiacol (4-VG) and hydroferulic acid (HFA)<sup>204</sup>. The same study has shown that the ferulic acid degradation occurred simultaneously with 4-VG appearance, while HFA formation started in the mid-exponential phase<sup>204</sup>. The chromatograms obtained after 3 days of growth for the strains grown in the modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>) are represented in Figure 5.56, where is visible that *L. lactis* LMG19460, but not *E. coli* MG1655 were producing the unknown metabolite, while the ferulic acid peak decreases.

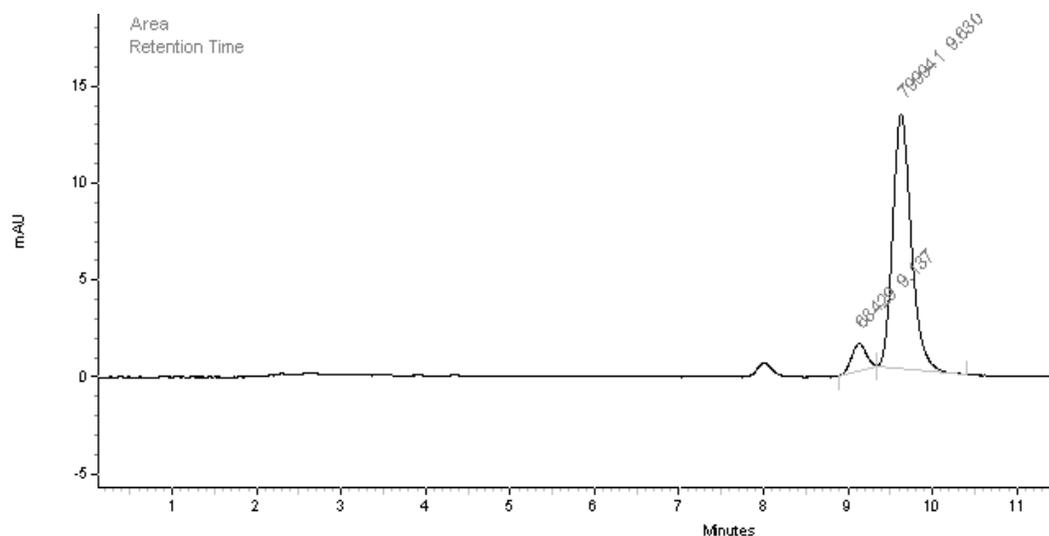


**Figure 5.55.** Areas (a.u.) of the extra peak that appeared in the ferulic acid chromatogram from HPLC analysis of the samples from the third growth experiment, after A) 24h of growth, B) 72h of growth and C) 7 days of growth. The data for the 3 media used (M9, complete M9 and M17) are presented. The strains tested were *L. lactis* LMG19460 harboring the pTRKH3 plasmid (dark blue columns), *L. lactis* LMG19460 harboring the pTRKH3\_SLP\_4CL\_DCS\_CURS1 (red columns), *L. lactis* LMG19460 harboring the pTRKH3\_LDH\_4CL\_DCS\_CURS1 (green columns), *L. lactis* LMG19460 with the 3 single-gene plasmids (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1) (purple columns) and *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1) (light blue columns).



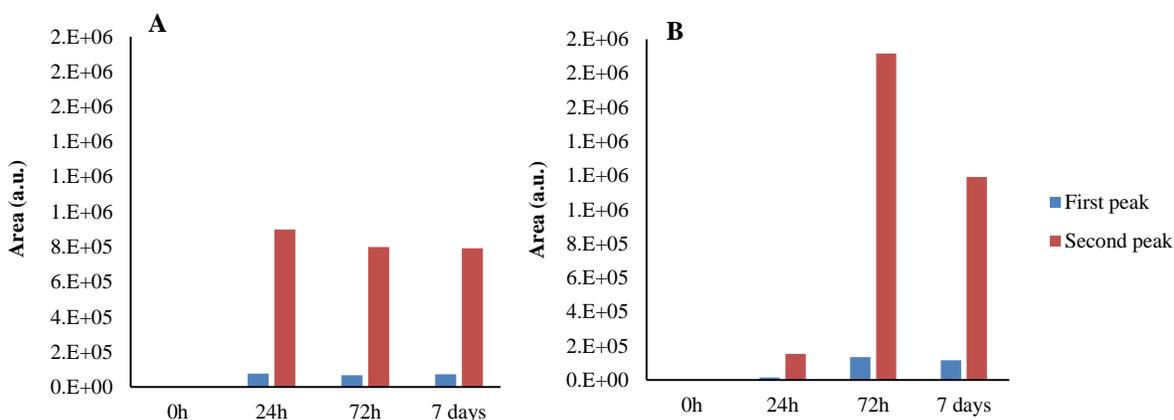
**Figure 5.56.** Chromatograms for ferulic acid obtained after 3 days of growth in modified M9 minimal salt medium (without thiamine, trace elements or  $\text{CaCO}_3$ ). The strains tested were A) *L. lactis* LMG19460 harboring the pTRKH3 plasmid, B) *L. lactis* LMG19460 harboring the pTRKH3\_SLP\_4CL\_DCS\_CURS1, C) *L. lactis* LMG19460 harboring the pTRKH3\_LDH\_4CL\_DCS\_CURS1, D) *L. lactis* LMG19460 with the 3 single-gene plasmids (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1) and E) *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1).

Nevertheless, the most important is to evaluate if the ferulic acid consumption is being also converted in curcumin production. In the *L. lactis* LMG19460 samples it was not detected any curcumin production, meaning that the ferulic acid consumption was exclusively due to the conversion into the unknown metabolite mentioned above. Only the *E. coli* MG1655 samples registered peaks with the curcumin HPLC method. In the curcumin chromatogram it always appeared two different peaks, one smaller peak in the correct retention time (9.13 min) and a bigger peak with a higher retention time (9.63 min) (Figure 5.57).



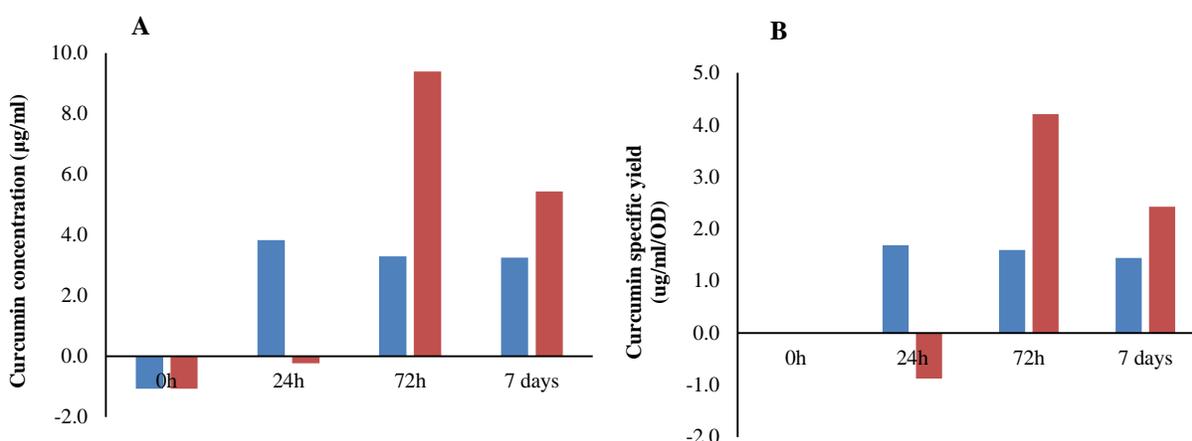
**Figure 5.57.** Curcumin chromatogram for *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1), obtained after 3 days of growth in modified M9 minimal salt medium (without thiamine, trace elements or CaCO<sub>3</sub>).

Both peaks areas are represented in Figure 5.58, where it is visible that the second peak had a much bigger area than the peak that corresponds to curcumin. One can hypothesize that this second bigger peak is other curcuminoid, such as demethoxycurcumin or bisdemethoxycurcumin, but further studies are necessary.



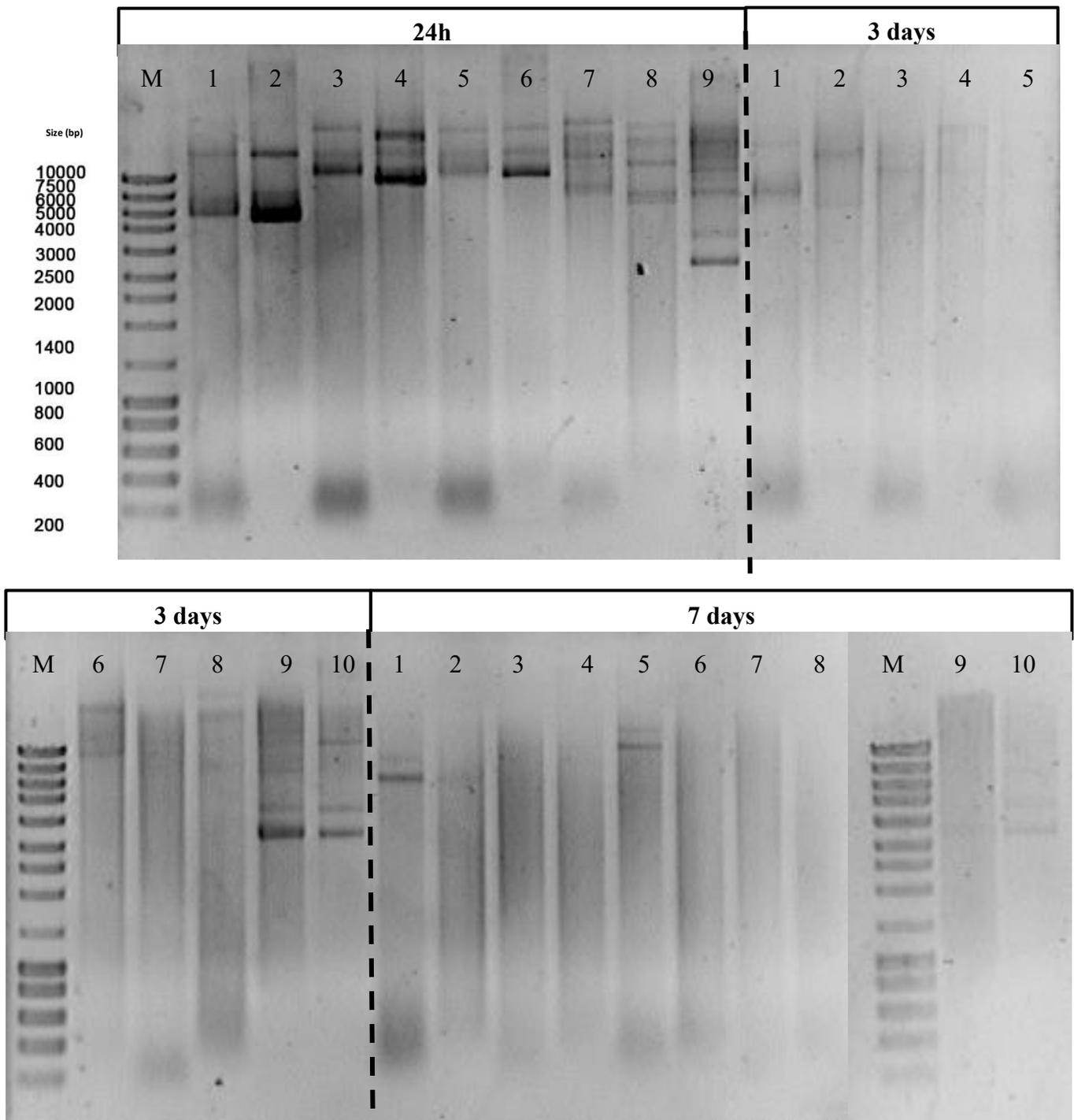
**Figure 5.58.** Areas (a.u.) of the curcumin peaks that appeared in the HPLC chromatogram of the *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1), after 0h, 24h, 72h and 7 days of growth, with A) M9 medium and B) complete M9 medium. The area values for the 2 different peaks (first peak in blue columns and second peak in red columns) that appeared in the chromatogram are presented.

Assuming that both peaks are curcuminoids, when the area (a.u) were converted to curcumin concentration ( $\mu\text{g/mL}$ ) values, the first peak only had negative values, while the second peak was the only one with interpretable positive values (Figure 5.59). At the beginning of the growth (0h), the MG1655 harboring the 3 plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1) did not have positive curcumin values as expected, but in the modified M9 minimal salt medium (without thiamine, trace elements or  $\text{CaCO}_3$ ) the curcumin concentration started to increase after 24h, decreasing slightly along the remaining 7 days of growth. With the complete modified M9 minimal salt medium (with thiamine, trace elements,  $\text{CaCO}_3$ , erythromycin and ferulic acid), the curcumin concentration started to be detectable later, only after 3 days of growth, but reached higher concentrations when compared with the first media. After 7 days of growth in the complete M9 media, the curcumin concentration also decreased. These results showed that after some days of growth, curcumin suffers some degree of degradation and/or consumption.



**Figure 5.59.** Curcumin A) concentration and B) specific yield after analysis of the second peak of the curcumin HPLC chromatogram of the *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1) from the third growth experiment, after 0h, 24h, 72h and 7 days of growth, with M9 (blue columns) and complete M9 medium (red columns).

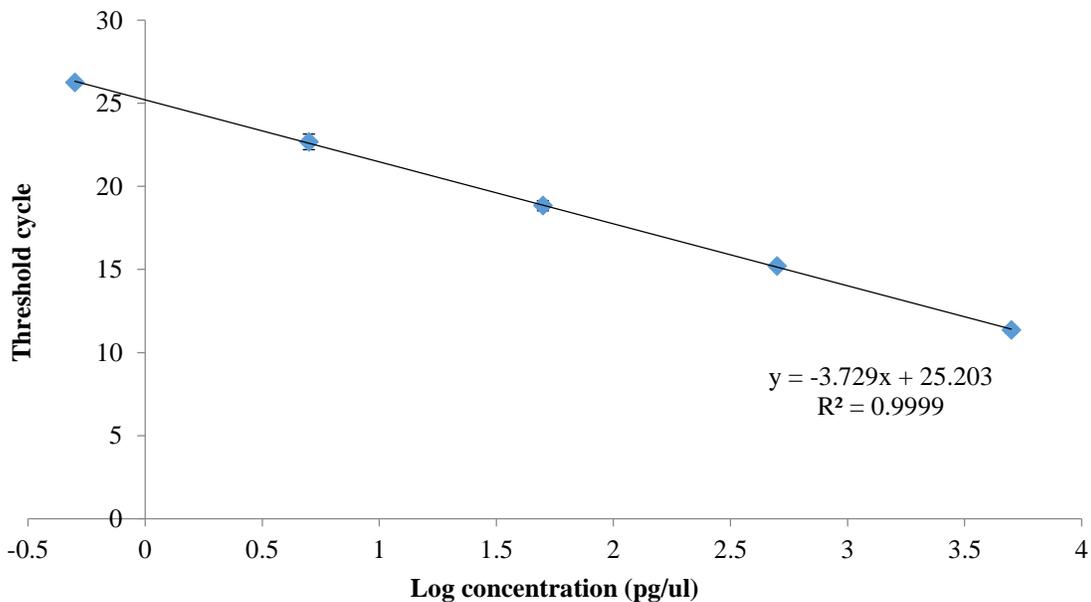
The gene and transcription levels were analyzed in search for an explanation to the absence of curcumin production by the *L. lactis* LMG19460 strains. The pDNA quality after 24 hours, 3 days and 7 days of growth can be visualized in Figure 5.60. It is clear that as time progresses, the pDNA quality decreases drastically, mainly in the *L. lactis* LMG19460 strains. After 24 hours of growth all plasmids from all the strains are visible, presenting good quality with a predominance of the supercoiled isoforms. After 3 days of growth, only the *E. coli* strains had plasmids visible in the agarose gel, and after 7 days all plasmids have been degraded. Since *E. coli* have good quality plasmids for a longer period of time, that could be one reason behind why it is able to produce curcumin, unlike *L. lactis* LMG19460 which plasmids started to degrade after 24 hours of growth, most probably due to presence of native endonucleases. Actually, *E. coli* had the maximum curcumin concentrations after 3 days of growth in complete M9 medium (Figure 5.59).



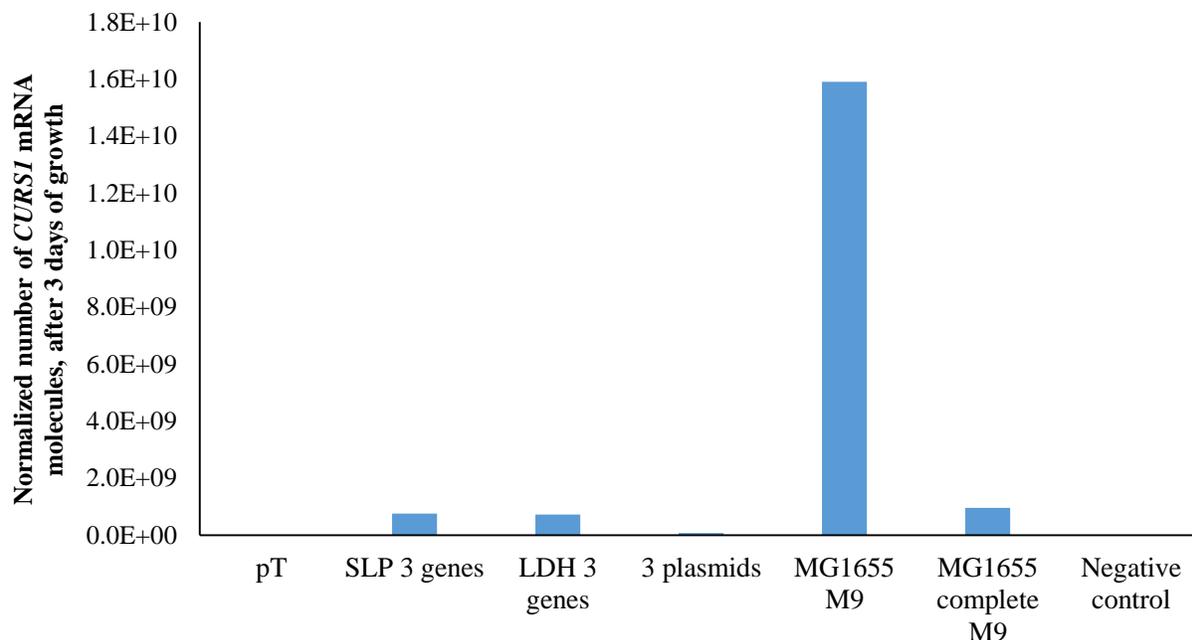
**Figure 5.60.** Agarose gel with 1 µg pDNA samples purified after 24h, 3 days and 7 days of growth. Lanes: **M**- Nzyladder III; **1**- *L. lactis* LMG19460 harboring the pTRKH3 plasmid, grown in modified M9 medium or **2**- M17 medium; **3**- *L. lactis* LMG19460 harboring the pTRKH3\_SLP\_4CL\_DCS\_CURS1, grown in modified M9 medium or **4**- M17 medium; **5**- *L. lactis* LMG19460 harboring the pTRKH3\_LDH\_4CL\_DCS\_CURS1, grown in modified M9 medium or **6**- M17 medium; **7**- *L. lactis* LMG19460 with the 3 single-gene plasmids (pTRKH3\_SLP\_4CL, pTRKH3\_SLP\_DCS and pTRKH3\_SLP\_CURS1), grown in modified M9 medium or **8**- M17 medium; **9**- *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1), grown in modified M9 medium or **10**- complete M9 medium.

The next approach was to quantify the mRNA from the *CURS1* gene present in the samples after 3 days of growth in modified M9 medium (and also the complete M9 medium for the *E. coli* strain). The standard curve for the *CURS1* mRNA is represented in the Figure 5.61, together with the respective equation.

The mRNA quantification results (Figure 5.62), normalized using the negative control values, indicated that *CURS1* gene was being transcribed in high quantity in the *E. coli* MG1655 harboring the 3 single-gene plasmids (pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1), grown in modified M9 medium, but in lower quantity in the same strain grown in the complete M9 medium. This is inconsistent with the curcumin quantification results where the highest curcumin concentrations were obtained for the *E. coli* strain grown in the complete M9 medium. The *L. lactis* LMG19460 harboring the pTRKH3\_SLP\_4CL\_DCS\_CURS1 and the pTRKH3\_LDH\_4CL\_DCS\_CURS1 also had detectable *CURS1* mRNA amounts, showing again that the problem does not seem to be in the transcription level, but instead something should be impairing the translation level (protein folding or protein stability). A negative control without cDNA was also analyzed and no *CURS1* was detected.



**Figure 5.61.** *CURS1* mRNA standard curve and respective equation.



**Figure 5.62.** Number of *CURS1* mRNA molecules per cell, quantified by real-time quantitative PCR, after 3 days of growth.

#### 5.4. Conclusions and future work

The main goal in this chapter was to engineer an artificial curcumin pathway in LAB, which was partially achieved. The molecular cloning steps to construct the necessary clones with the three genes of the curcumin pathway (*4CL*, *DCS* and *CURS1*) in operon were successfully achieved. The *L. lactis* LMG19460 strain was also successfully transformed with all the necessary plasmids. However, the *L. lactis* clones failed to produce curcumin, except the positive control (*E. coli* MG1655 harboring pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1). The reason why they failed in the curcumin production could be due to several different reasons. Starting by the production step, during growth, it is tempting to think that since the positive control *E. coli* strain produced reasonable amounts of curcumin, the problem should not be in this step. One could think that since *L. lactis* strains are Gram-positive bacteria, this could affect the ferulic acid uptake due to differences in the cell wall characteristics and transport mechanisms. However, the ferulic acid chromatogram showed that these *L. lactis* strains were efficiently consuming the ferulic acid, although converting it into an unknown metabolism, instead of directing it for curcumin production.

Here the problem could be in the plasmid copy number, which if it is too low, is not leading to sufficient expression of the three necessary proteins, preventing the production of curcumin. Therefore, it would be beneficial to construct and test the three genes curcumin pathway cloned in the high PCN mutant pTRKH3-b, to check if the problem is due to a low PCN. Also, the *L. lactis* strains plasmid quality starts decreasing after 24 hours of growth, probably not giving time for a quantifiable amount of curcumin to be produced. The reason behind this rapid plasmid degradation is the

presence of native endonucleases in this strain, being of utmost importance to achieve the *nth* gene knockout pursued in the Chapter 4.

The genes transcription seems to be occurring at some degree, as the *CURS1* mRNA quantification experiments showed, but stronger promoters could be tested as future work, in order to increase transcription efficiency and consequently the mRNAs amounts and ultimately the curcumin production.

The problem could also be in the translation step. Actually, the genes sequences were codon-optimized for *E. coli* and not for *L. lactis*, which could eventually interfere with the translation. As a future work, the sequences should be codon-optimized for *L. lactis*, in order to confirm if it could be the reason why bacteria are not producing curcumin. Also important as a future work is to engineer a *L. lactis* LMG19460 strain that is theoretically more efficient in redirecting the metabolism for pDNA production, for example by performing the knockout of the *pgi* gene.

Other potentially critical procedure is the curcumin extraction, which could be less effective in Gram-positive bacteria than in Gram-negative, due to the differences in the peptidoglycan thickness. If we consider that the *L. lactis* cells are producing curcumin, but are not secreting it efficiently to the extracellular medium, together with an inefficient extraction, could impair the curcumin detection. This theory is not incompatible with the ferulic acid results, since it is not possible to confirm if the ferulic acid being detected originates from the extracellular and/or intracellular medium. The decrease in the ferulic acid concentration along the 7 days of growth that is detected in the HPLC analysis could be due to extracellular ferulic acid detection only, since the cells are able to uptake it from the extracellular medium. After the uptake, the cells could be metabolizing it to curcumin, but if the extraction process is inefficient, the curcumin will not be detected. An additional extraction procedure was tested, consisting in a previous incubation with lysozyme, but the outcome was identical (results not shown). Therefore it is advisable to optimize further the extraction process.

Once the curcumin could be successfully produced by *L. lactis* LMG19460, the goal is to achieve the proof-of-concept that the LAB with the synthetic curcumin pathway are cytotoxic for cancer cells. To mimic *in vivo* gastrointestinal conditions<sup>206</sup>, Transwell cell culture chambers will be used<sup>207</sup> to co-culture human colorectal cancer cells (Caco-2) and LAB cells, separated by a porous membrane that mimics the intestinal wall. This *in situ* production of curcumin by LAB directly on the Caco-2 cells, will allow to attest if the engineered LAB can produce curcumin in a way that is cytotoxic to cancer cells. Bioimaging techniques, such as optical, fluorescence and confocal microscopy would be used to evaluate cancer cell morphology. Cytotoxicity and apoptosis assays would be conducted to evaluate the efficiency of the engineered LAB.

The anti-cancer properties of curcumin, along its ability to prevent cancer, having minimal adverse effects and being non-toxic for normal cells, makes it a promising alternative to the current chemotherapeutic drugs. The consideration of the chemical properties of this compound, namely its stability, degradation products and metabolism inside human body, revealed several interesting points to be considered in the future, while optimizing a way to be feasible to produce curcumin *in vivo*. Several characteristics were shown that give LAB the potential of being the ideal delivery vectors of curcumin for GI cancer therapy.



# CHAPTER 6

**Concluding remarks and perspectives for further research**

*The important thing is to never stop questioning.* – **Albert Einstein**

## 6.1. General conclusions and future work

This thesis aimed to develop a highly efficient Lactic Acid Bacteria (LAB)/artificial plasmid platform for the production of pharmaceutical-grade plasmid DNA and recombinant proteins, for use in DNA and mucosal vaccination. To accomplish this goal, the thesis tried to answer to three main challenges: 1) the low copy number of the available LAB plasmids; 2) the presence of a high level of endonuclease activity in the majority of the strains; and 3) the implementation of a highly efficient LAB/high copy number plasmid platform.

Concerning the first challenge, the goal was to increase the plasmid copy number (PCN) of pTRKH3, a shuttle cloning vector able to replicate both in Gram-negative and Gram-positive hosts. The Gram-positive replicon was the pAM $\beta$ 1, which is constituted by a *repDE* operon under the control of the P<sub>DE</sub> promoter, where the *repE* gene codes for the rate-limiting replication initiation protein essential for the unidirectional theta-type replication mechanism. The *repDE* Ribosome Binding Site (RBS) of the pTRKH3 plasmid was engineered by site-directed mutagenesis, generating the high PCN pTRKH3-b mutant, which after 10.5 hours of growth achieved an average of 215 copies of plasmid per chromosome, corresponding to a 3.5 fold increase when compared to the non-modified pTRKH3 (62 copies per chromosome). The highest PCN obtained with pTRKH3-b seems to be due to a combination of a stronger RBS sequence, an mRNA secondary structure that promotes the ribosome binding and an ideal intermediate amount of transcriptional repressors (CopF and CT-RNA) present in *L. lactis* LMG19460 cells. The presence of a duplicated region that added an additional RBS sequence and one new alternative *in frame* start codon should also be contributing for the high PCN value of the pTRKH3-b mutant. The pTRKH3-b PCN value is significantly higher than the highest previously reported for the same replicon, which was around 100 copies per *Bacillus subtilis* cell, contributing to bring LAB closer to be an industrially profitable and food/pharmaceutical-grade platform for the production of i) pDNA to be used in DNA vaccination or ii) heterologous proteins to be used as antigens in mucosal vaccination. As a future work, it should be interesting to detect and quantify the amount of protein RepE effectively produced by each pTRKH3 mutant, to verify if it is proportional to the differences in the PCN. Also, the quantification of each transcriptional repressor could shed some light on the mechanism behind the high PCN value for pTRKH3-b.

Still related with the first challenge, other goal was to use Gibson Assembly to engineer a minimal plasmid with improved BioBrick parts (including the high PCN replicon), to increase the PCN. The assembled plasmids were based on the already available *E. coli*/Gram-positive shuttle plasmid pTRKH3 and may contain the following BioBrick compatible parts: an *E. coli* cassette (useful for the initial steps of molecular cloning), a Gram-positive cassette and two protein expression cassettes, one for prokaryotic applications such as using LAB for producing proteins or metabolites, and the other when the goal is the expression in eukaryotic cells as needed in DNA vaccination. Both the *E. coli* and the Gram-positive cassettes (with type and high PCN mutant b pAM $\beta$ 1 replicon) were successfully assembled and the resulting plasmids used to transform *L. lactis* LMG19460<sup>111</sup>. The PCN of the assembled plasmids were unexpectedly lower than the parental pTRKH3 and pTRKH3-b<sup>111</sup>, probably due to problems of unwanted recombination events hosted by the RecA protein present in this non-modified strain. It was also attempted the assembly of a LAB protein expression cassette with the three genes of the curcumin synthetic pathway under the control of the SLP promoter, in order to increase plasmid size, in the hope to minimize recombination, but the assembly

did not work. Further work is needed to accomplish the assembly of the LAB protein expression cassette and transform the *L. lactis* LMG19460 strain, in order to determine the influence of the increased size in the PCN value, recombination phenomenon and ultimately in the curcumin production. With the same goal of increasing the PCN value and decrease the recombination probability, several alternative BioBricks could be tested, such as change the entire theta-type Gram-positive origin of replication for one based on the rolling-circle replication mechanism, which are described as having higher PCN values. Other alternative is to maintain the pAM $\beta$ 1 origin of replication, but removing the truncated *copF* or the *CT-RNA* genes, coding for transcriptional repressors, in order to increase the PCN. Also, a *recA* knockout (using a recombineering or CRISPR/Cas9 system (described in Chapter 4)) should be pursued, with the goal of decreasing the recombination events over the exogenous pDNA. An eukaryotic protein expression cassette, coding for a vaccine antigen, should also be assembled to the plasmid already containing the *E. coli* and Gram-positive cassettes, for a future application in DNA vaccination using LAB as food-grade live delivery vectors.

The second challenge included the use of the Optflux software to predict the best candidate genes to knockout for biomass and pDNA production increase. The genes predicted by genomic analysis were *araT*, *nucA*, *yjhF*, *gltB*, *nrdF*, *adhE*, *argh*, *argg* and *purH*, which intervened in several different pathways, such as glycolysis/gluconeogenesis and purine metabolism. Other relevant genes to perform knockout were chosen solely based on literature and previous studies, such as *nth*, *recA*, *pgi*, *htrA* and *pyk*. The gene chosen for the first knockout was the *nth* gene, which codes for an endonuclease III responsible for the degradation of exogenous DNA by non-specific digestion, meaning that its knockout will allow *L. lactis* LMG19460 transformation with pDNA, resulting in higher plasmid yields and better plasmid quality. After several attempts to use random mutagenesis and the Datsenko & Wanner recombineering strategy, without success, the CRISPR/Cas9 system was chosen as the most suitable to perform the knockouts in this strain. Until now, the appropriate plasmid based on pKCCas9dO, a one-step system originally designed for *Streptomyces*, was successfully designed and constructed. To the best of our knowledge this is the only one-step plasmid designed for LAB, without necessity of helper plasmids. The first constructed plasmid was the pKCCas9dO\_nth\_apra containing the apramycin resistance gene, which was not considerate appropriate for *L. lactis* LMG19460 selection, due to its high natural resistance to this antibiotic. A second plasmid was designed, pKCCas9dO\_nth\_ery, with the erythromycin resistance gene instead of the apramycin one. The successful construction of this is yet to be confirmed, as well as its ability to transform *L. lactis* LMG19460 after assembly. Several problems were considered as could be contributing to this, such as the fact that the *Streptomyces* origin of replication pSG5 could not function in *L. lactis*, being the reason behind why some work was done towards the assembly of the necessary CRISPR/Cas9 components (*ScoCas9* gene, sgRNA and the two homologous arms) with the pTRKH3 backbone, still without success. An additional problem addressed at the end of the Chapter 4 is that the *ScoCas9* under the control of the *tipA* promoter was not induced (even after several attempts) by thioestrepton, perhaps because it is not inducible by this antibiotic in our LAB strain. This failure does not allow to knockout the *nth* gene, and further work is necessary to obtain a *L. lactis* LMG19460 strain that does not degrade plasmids. Future work suggestions include the replacement of the *tipA* promoter for a nisin A inducible promoter. The sequence of the *ScoCas9* itself was codon-optimized for *Streptomyces* and as a future work it should be optimized for the codon usage bias of *L. lactis* to try to increase its expression in this strain. If in the future, the *Anth* strain could be engineered, it should be transformed with a known

plasmid, such as pTRKH3, to ascertain if the new PCN increased, when compared with the PCN in the wild-type strain. The quality of the pDNA, namely the increase of the supercoiled fraction, should also be evaluated.

The third challenge is at the cell factory level and has as main goal to implement a highly efficient LAB/high copy number plasmid platform using a curcumin synthetic pathway. Curcumin is the yellow pigment from *Curcuma longa* and completed clinical trials have shown that curcumin is useful for the chemoprevention of colon cancer in humans, being safe for normal cells and with minimal adverse effects, making it a good complement to the current chemotherapeutic agents. The biggest problem with curcumin is its slow cellular uptake, low systemic bioavailability and the fact that repetitive oral doses of the molecule are required to achieve an effective therapeutic concentration inside the cells. The possibility of synthesizing curcumin *in situ*, i.e. in the human gastrointestinal tract, by LAB represents a putative powerful and innovative solution for gastrointestinal cancer therapy. The molecular cloning steps to construct the necessary clones with the three genes of the curcumin pathway (*4CL*, *DCS* and *CURS1*) in operon were successfully achieved and the *L. lactis* LMG19460 strain was also successfully transformed with all the necessary plasmids. However, the *L. lactis* clones failed to produce curcumin, although the positive control (*E. coli* MG1655 harboring pAC-4CL1, pCDFDuet\_DCS and pRSFDuet\_CURS1) did not. The reason why they failed to produce curcumin could be due to several different motives. One problem could be the plasmid copy number, which if it is too low, does not lead to sufficient expression of the three necessary proteins, preventing the production of curcumin. Here, it would be beneficial to construct and test the three genes of the curcumin pathway cloned in the high PCN mutant pTRKH3-b, to check if the problem is due to a low PCN. Also, the plasmid quality in the *L. lactis* LMG19460 strains starts to decrease after 24 hours of growth, probably not giving enough time for the curcumin to be produced in a quantifiable amount. This does not occur in the *E. coli* strain, meaning that in *L. lactis* the reason behind this rapid plasmid degradation is the presence of native endonucleases, being of utmost importance to achieve the *nth* gene knockout pursued in the Chapter 4. At the transcription level, it seems to be working at some level in *L. lactis*, but is clearly more efficient in the *E. coli* strain. Stronger promoters could be tested as future work, in order to increase transcription efficiency and consequently the mRNAs amounts and ultimately the curcumin production. The genes sequences were codon-optimized for *E. coli* and not for *L. lactis*, which could interfere with the translation. As a future work, the sequences should be codon-optimized for *L. lactis*, in order to confirm if it could be the reason why bacteria are not producing curcumin. Also important as a future work is to engineer a *L. lactis* LMG19460 strain that is theoretically more efficient in redirecting the metabolism for pDNA production, for example by performing the knockout of the *pgi* gene. Once the curcumin could be successfully produced by *L. lactis* LMG19460, the goal is to achieve the proof-of-concept that the LAB with the synthetic curcumin pathway are cytotoxic for cancer cells. To mimic *in vivo* gastrointestinal conditions, transwell cell culture chambers should be used to co-culture human colorectal cancer cells (Caco-2) and LAB cells, separated by a porous membrane that mimics the intestinal wall.

Although only the first challenge was successfully solved, this thesis contributed with important insight and results to bring LAB several steps closer to the goal of using it as a safe, profitable and pharmaceutical-grade alternative to *E. coli* for DNA and mucosal vaccination.

# **CHAPTER 7**

## **References**

1. Wood BJB, Holzapfel WH (1995) The genera of lactic acid bacteria: edited by B.J.B. Wood and W.H. Holzapfel. Springer Science & Business Media, New York
2. Schleifer K, Kraus J, Dvorak C, Kilpper-Bälz R, Collins M, Fischer W (1985) Transfer of *Streptococcus lactis* and Related Streptococci to the Genus *Lactococcus* gen. nov. Systematic and Applied Microbiology 6:183–195. doi: 10.1016/s0723-2020(85)80052-7
3. Bolotin A, Wincker P, Mauger S, Jaillon O, Malarne K, Weissenbach J, Ehrlich SD, Sorokin A (2001) The Complete Genome Sequence of the Lactic Acid Bacterium *Lactococcus lactis* ssp. *lactis* IL1403. Genome Research 11:731–753. doi: 10.1101/gr.1697r
4. Silva I, Duarte S, Moreira LM, Monteiro GA (2017) Draft Genome Sequence of the Plasmid-Free *Lactococcus lactis* subsp. *lactis* strain LMG 19460. Genome Announcements. doi: 10.1128/genomea.00210-17
5. D’Souza R, Pandeya DR, Hong ST (2012) Review: *Lactococcus Lactis*: An efficient Gram positive cell factory for the production and secretion of recombinant protein. Biomedical Research 23(1):1-7
6. Bermúdez-Humarán LG, Kharrat P, Chatel J-M, Langella P (2011) Lactococci and lactobacilli as mucosal delivery vectors for therapeutic proteins and DNA vaccines. Microbial Cell Factories 10. doi: 10.1186/1475-2859-10-s1-s4
7. Morello E, Bermúdez-Humarán L, Llull D, Solé V, Miraglio N, Langella P, Poquet I (2007) *Lactococcus lactis*, an Efficient Cell Factory for Recombinant Protein Production and Secretion. Journal of Molecular Microbiology and Biotechnology 14:48–58. doi: 10.1159/000106082
8. Le Loir Y, Azevedo V, Oliveira SC, Freitas DA, Miyoshi A, Bermúdez-Humarán LG, Nouaille S, Ribeiro L, Leclercq S, Gabriel J, Guimaraes V, Oliveira M, Charlier C, Gautier M, Langella P (2005) Protein secretion in *Lactococcus lactis*: an efficient way to increase the overall heterologous protein production. Microbial Cell Factories 4(1):2. doi: 10.1186/1475-2859-4-2
9. Reyes-Sandoval A, Ertl H (2001) DNA Vaccines. Current Molecular Medicine 1:217–243. doi: 10.2174/1566524013363898
10. Stern AM, Markel H (2005) The History Of Vaccines And Immunization: Familiar Patterns, New Challenges. Health Affairs 24:611–621. doi: 10.1377/hlthaff.24.3.611
11. Kutzler MA, Weiner DB (2008) DNA vaccines: ready for prime time? Nature Reviews Genetics 9:776–788. doi: 10.1038/nrg2432
12. Gurunathan S, Klinman DM, Seder RA (2000) DNA vaccines: immunology, application, and optimization. Annual review of immunology 18(1):927-974. doi: 10.1146/annurev.immunol.18.1.927
13. Pereira VB, Zurita-Turk M, Saraiva TDL, Castro CPD, Souza BM, Agresti PM, Lima FA, Pfeiffer VN, Azevedo MSP, Rocha CS, Pontes DS, Azevedo V, Miyoshi A (2014) DNA Vaccines Approach: From Concepts to Applications. World Journal of Vaccines 04:50–71. doi: 10.4236/wjv.2014.42008
14. Gardlík R, Pálffy R, Hodosy J, Lukács J, Turna J, Celec P (2005) Vectors and delivery systems in gene therapy. Medical Science Monitor 11(4):RA110-RA121.
15. Luo D, Saltzman WM (2000) Synthetic DNA delivery systems. Nature biotechnology 18(1):33-37. doi: 10.1038/71889
16. Williams DA (2003) MEDICINE: Gene Therapy--New Challenges Ahead. Science 302:400–401. doi: 10.1126/science.1091258
17. Wolff J, Malone R, Williams P, Chong W, Acsadi G, Jani A, Felgner P (1990) Direct gene transfer into mouse muscle in vivo. Science 247:1465–1468. doi: 10.1126/science.1690918
18. Donnelly J, Berry K, Ulmer JB (2003) Technical and regulatory hurdles for DNA vaccines. International Journal for Parasitology 33:457–467. doi: 10.1016/s0020-7519(03)00056-0
19. Kemp M (2015) DNA VACCINE TECHNOLOGY – A Vaccine Breakthrough That Could Change Lives & Enable Vaccine Development Programs. Drug Development & Delivery.

20. Prather KJ, Sagar S, Murphy J, Chartrain M (2003) Industrial scale production of plasmid DNA for vaccine and gene therapy: plasmid design, production, and purification. *Enzyme and Microbial Technology* 33:865–883. doi: 10.1016/s0141-0229(03)00205-9
21. Gram GJ, Fomsgaard A, Thorn M, Madsen SM, Glenting J (2007) Immunological analysis of a *Lactococcus lactis*-based DNA vaccine expressing HIV gp120. *Genetic vaccines and therapy* 5(1):3. doi: 10.1186/1479-0556-5-3
22. Yu Q, Zhu L, Kang H, Yang Q (2013) Mucosal *Lactobacillus* vectored vaccines. *Human Vaccines & Immunotherapeutics* 9:805–807. doi: 10.4161/hv.23302
23. Lin I, Van T, Smooker P (2015) Live-Attenuated Bacterial Vectors: Tools for Vaccine and Therapeutic Agent Delivery. *Vaccines* 3:940–972. doi: 10.3390/vaccines3040940
24. Pereira VB, Saraiva TDL, Souza BM, Zurita-Turk M, Azevedo MSP, Castro CPD, Mancha-Agresti P, Santos JSCD, Santos ACG, Faria AMC, Leclercq S, Azevedo V, Miyoshi A (2014) Development of a new DNA vaccine based on mycobacterial ESAT-6 antigen delivered by recombinant invasive *Lactococcus lactis* FnBPA. *Applied Microbiology and Biotechnology* 99:1817–1826. doi: 10.1007/s00253-014-6285-3
25. Pereira V, Cunha VD, Preisser T, Souza B, Turk M, Castro CD, Azevedo M, Miyoshi A (2017) *Lactococcus lactis* carrying a DNA vaccine coding for the ESAT-6 antigen increases IL-17 cytokine secretion and boosts the BCG vaccine immune response. *Journal of Applied Microbiology* 122:1657–1662. doi: 10.1111/jam.13449
26. Souza BM, Preisser TM, Pereira VB, Zurita-Turk M, Castro CPD, Cunha VPD, Oliveira RPD, Gomes-Santos AC, Faria AMCD, Machado DCC, Chatel J-M, Azevedo VADC, Langella P, Miyoshi A (2016) *Lactococcus lactis* carrying the pValac eukaryotic expression vector coding for IL-4 reduces chemically-induced intestinal inflammation by increasing the levels of IL-10-producing regulatory cells. *Microbial Cell Factories*. doi: 10.1186/s12934-016-0548-x
27. Carvalho RD, Breyner N, Menezes-Garcia Z, Rodrigues NM, Lemos L, Maioli TU, Souza DDG, Carmona D, Faria AMCD, Langella P, Chatel J-M, Bermúdez-Humarán LG, Figueiredo HCP, Azevedo V, Azevedo MSD (2017) Secretion of biologically active pancreatitis-associated protein I (PAP) by genetically modified dairy *Lactococcus lactis* NZ9000 in the prevention of intestinal mucositis. *Microbial Cell Factories*. doi: 10.1186/s12934-017-0624-x
28. Gomes-Santos AC, Oliveira RPD, Moreira TG, Castro-Junior AB, Horta BC, Lemos L, Almeida LAD, Rezende RM, Cara DC, Oliveira SC, Azevedo VAC, Miyoshi A, Faria AMC (2017) Hsp65-Producing *Lactococcus lactis* Prevents Inflammatory Intestinal Disease in Mice by IL-10- and TLR2-Dependent Pathways. *Frontiers in Immunology*. doi: 10.3389/fimmu.2017.00030
29. Yagnik B, Sharma D, Padh H, Desai P (2016) Immunization with r-*Lactococcus lactis* expressing outer membrane protein A of *Shigella dysenteriae* type-1: evaluation of oral and intranasal route of administration. *Journal of Applied Microbiology* 122:493–505. doi: 10.1111/jam.13353
30. Beck BR, Lee SH, Kim D, Park JH, Lee HK, Kwon S-S, Lee KH, Lee JI, Song SK (2017) A *Lactococcus lactis* BFE920 feed vaccine expressing a fusion protein composed of the OmpA and FlgD antigens from *Edwardsiella tarda* was significantly better at protecting olive flounder (*Paralichthys olivaceus*) from edwardsiellosis than single antigen vaccines. *Fish & Shellfish Immunology* 68:19–28. doi: 10.1016/j.fsi.2017.07.004
31. Kim D, Beck BR, Lee SM, Jeon J, Lee DW, Lee JI, Song SK (2016) Pellet feed adsorbed with the recombinant *Lactococcus lactis* BFE920 expressing SiMA antigen induced strong recall vaccine effects against *Streptococcus iniae* infection in olive flounder (*Paralichthys olivaceus*). *Fish & Shellfish Immunology* 55:374–383. doi: 10.1016/j.fsi.2016.06.010
32. Sun N, Zhang R, Duan G, Peng X, Wang C, Fan Q, Chen S, Xi Y (2017) An engineered food-grade *Lactococcus lactis* strain for production and delivery of heat-labile enterotoxin B subunit to mucosal sites. *BMC Biotechnology*. doi: 10.1186/s12896-017-0345-6

33. Szczepankowska AK, Szatraj K, Salański P, Rózga A, Górecki RK, Bardowski JK (2017) Recombinant *Lactococcus lactis* Expressing Haemagglutinin from a Polish Avian H5N1 Isolate and Its Immunological Effect in Preliminary Animal Trials. *BioMed Research International* 2017:1–8. doi: 10.1155/2017/6747482
34. Ma C, Zhang L, Gao M, Ma D (2017) Construction of *Lactococcus lactis* expressing secreted and anchored *Eimeria tenella* 3-1E protein and comparison of protective immunity against homologous challenge. *Experimental Parasitology* 178:14–20. doi: 10.1016/j.exppara.2017.05.001
35. Ma D, Gao M, Dalloul RA, Ge J, Ma C, Li J (2013) Protective effects of oral immunization with live *Lactococcus lactis* expressing *Eimeria tenella* 3-1E protein. *Parasitology Research* 112:4161–4167. doi: 10.1007/s00436-013-3607-9
36. Mao R, Wu D, Hu S, Zhou K, Wang M, Wang Y (2017) Secretory expression and surface display of a new and biologically active single-chain insulin (SCI-59) analog by lactic acid bacteria. *Applied Microbiology and Biotechnology* 101:3259–3271. doi: 10.1007/s00253-017-8125-8
37. Daniel C, Titecat M, Poiret S, Cayet D, Boutillier D, Simonet M, Sirard J-C, Lemaître N, Sebbane F (2016) Characterization of the protective immune response to *Yersinia pseudotuberculosis* infection in mice vaccinated with an LcrV-secreting strain of *Lactococcus lactis*. *Vaccine* 34:5762–5767. doi: 10.1016/j.vaccine.2016.09.060
38. Mao R, Zhou K, Han Z, Wang Y (2016) Subtilisin QK-2: secretory expression in *Lactococcus lactis* and surface display onto gram-positive enhancer matrix (GEM) particles. *Microbial Cell Factories*. doi: 10.1186/s12934-016-0478-7
39. Zadavec P, Marečková L, Petroková H, Hodnik V, Nanut MP, Anderluh G, Štrukelj B, Malý P, Berlec A (2016) Development of Recombinant *Lactococcus lactis* Displaying Albumin-Binding Domain Variants against Shiga Toxin 1 B Subunit. *Plos One*. doi: 10.1371/journal.pone.0162625
40. Shigemori S, Ihara M, Sato T, Yamamoto Y, Nigar S, Ogita T, Shimosato T (2016) Secretion of an immunoreactive single-chain variable fragment antibody against mouse interleukin 6 by *Lactococcus lactis*. *Applied Microbiology and Biotechnology* 101:341–349. doi: 10.1007/s00253-016-7907-8
41. Liu S, Li Y, Deng B, Xu Z (2016) Recombinant *Lactococcus lactis* expressing porcine insulin-like growth factor I ameliorates DSS-induced colitis in mice. *BMC Biotechnology*. doi: 10.1186/s12896-016-0255-z
42. Zhang Q, Ai C (2016) Development of House Dust Mite Vaccine. *Vaccine Design Methods in Molecular Biology* 739–751. doi: 10.1007/978-1-4939-3387-7\_42
43. Zhang B, Li A, Zuo F, Yu R, Zeng Z, Ma H, Chen S (2016) Recombinant *Lactococcus lactis* NZ9000 secretes a bioactive kisspeptin that inhibits proliferation and migration of human colon carcinoma HT-29 cells. *Microbial Cell Factories*. doi: 10.1186/s12934-016-0506-7
44. Breyner NM, Michon C, Sousa CSD, Boas PBV, Chain F, Azevedo VA, Langella P, Chatel JM (2017) Microbial Anti-Inflammatory Molecule (MAM) from *Faecalibacterium prausnitzii* Shows a Protective Effect on DNBS and DSS-Induced Colitis Model in Mice through Inhibition of NF-κB Pathway. *Frontiers in Microbiology*. doi: 10.3389/fmicb.2017.00114
45. Kasareňo K, Szczepankowska A, Kwiatkowska-Patzer B, Lipkowski AW, Gadamski R, Sulejczak D, Łachwa M, Biały M, Bardowski J (2016) Effect of recombinant *Lactococcus lactis* producing myelin peptides on neuroimmunological changes in rats with experimental allergic encephalomyelitis. *Folia Neuropathologica* 3:249–258. doi: 10.5114/fn.2016.62534
46. Kobińska PA, Olech B, Książek M, Derlatka K, Adamska I, Majewski PM, Jagusztyn-Krynicka EK, Wszyńska AK (2016) Cell Wall Anchoring of the *Campylobacter* Antigens to *Lactococcus lactis*. *Frontiers in Microbiology*. doi: 10.3389/fmicb.2016.00165
47. Almeida JF, Breyner NM, Mahi M, Ahmed B, Benbouziane B, Boas PCV, Miyoshi A, Azevedo V, Langella P, Bermúdez-Humarán LG, Chatel J-M (2016) Expression of fibronectin binding protein A (FnBPA) from *Staphylococcus aureus* at the cell surface of *Lactococcus lactis* improves its immunomodulatory properties when used as protein delivery vector. *Vaccine* 34:1312–1318. doi: 10.1016/j.vaccine.2016.01.022

48. Durmaz E, Hu Y, Aroian RV, Klaenhammer TR (2015) Intracellular and Extracellular Expression of *Bacillus thuringiensis* Crystal Protein Cry5B in *Lactococcus lactis* for Use as an Anthelmintic. *Applied and Environmental Microbiology* 82:1286–1294. doi: 10.1128/aem.02365-15
49. Seegers JFML, Meijer WJJ, Venema G, Bron S, Zhao AC, Khan SA (1995) Structural and functional analysis of the single-strand origin of replication from the lactococcal plasmid pWV01. *MGG Molecular & General Genetics* 249:43–50. doi: 10.1007/bf00290234
50. Bryksin AV, Matsumura I (2010) Rational Design of a Plasmid Origin That Replicates Efficiently in Both Gram-Positive and Gram-Negative Bacteria. *PLoS ONE*. doi: 10.1371/journal.pone.0013244
51. Gasson MJ, de Vos W (2012) *Genetics and Biotechnology of Lactic Acid Bacteria*. Springer Science & Business Media, New York
52. Leenhouts KJ, Kok J, Venema G (1991) Lactococcal plasmid pWV01 as an integration vector for lactococci. *Applied and environmental microbiology* 57(9):2562-2567.
53. Leenhouts KJ, Tolner B, Bron S, Kok J, Venema G, Seegers JF (1991) Nucleotide sequence and characterization of the broad-host-range lactococcal plasmid pWV01. *Plasmid* 26:55–66. doi: 10.1016/0147-619x(91)90036-v
54. Yagnik B, Sharma D, Padh H, Desai P (2017) Dual recombinant *Lactococcus lactis* for enhanced delivery of DNA vaccine reporter plasmid pPERDBY. *Microbiology and Immunology* 61:123–129. doi: 10.1111/1348-0421.12473
55. Guimarães V, Innocentini S, Chatel J-M, Lefèvre F, Langella P, Azevedo V, Miyoshi A (2009) A new plasmid vector for DNA delivery using lactococci. *Genetic Vaccines and Therapy* 7:4. doi: 10.1186/1479-0556-7-4
56. Vos WM (1987) Gene cloning and expression in lactic streptococci. *FEMS Microbiology Letters* 46:281–295. doi: 10.1111/j.1574-6968.1987.tb02466.x
57. Tauer C, Heintz S, Egger E, Heiss S, Grabherr R (2014) Tuning constitutive recombinant gene expression in *Lactobacillus plantarum*. *Microbial Cell Factories* 13:150. doi: 10.1186/preaccept-1225362755140327
58. MoBiTec (2017) Plasmids for intracellular expression and secretion. [https://www.mobitec.com/cms/products/bio/04\\_vector\\_sys/1\\_lactis\\_nice\\_nisin\\_expression.html](https://www.mobitec.com/cms/products/bio/04_vector_sys/1_lactis_nice_nisin_expression.html). Accessed November 7<sup>th</sup>, 2017
59. Frelet-Barrand A, Boutigny S, Moyet L, Deniaud A, Seigneurin-Berny D, Salvi D, Bernaudat F, Richaud P, Pebay-Peyroula E, Joyard J, Rolland N (2010) *Lactococcus lactis*, an Alternative System for Functional Expression of Peripheral and Intrinsic *Arabidopsis* Membrane Proteins. *PLoS ONE*. doi: 10.1371/journal.pone.0008746
60. Bruand C, Chatelier EL, Ehrlich SD, Janniere L (1993) A fourth class of theta-replicating plasmids: the pAM beta 1 family from gram-positive bacteria. *Proceedings of the National Academy of Sciences* 90:11668–11672. doi: 10.1073/pnas.90.24.11668
61. Chatelier EL, Ehrlich SD, Jannière L (1994) The pAM $\beta$ 1 CopF repressor regulates plasmid copy number by controlling transcription of the repE gene. *Molecular Microbiology* 14:463–471. doi: 10.1111/j.1365-2958.1994.tb02181.x
62. Lertcanawanichakul M (2007) Construction of Plasmid Vector for Expression of Bacteriocin N15-Encoding Gene and Effect of Engineered Bacteria on *Enterococcus faecalis*. *Current Microbiology* 54:108–112. doi: 10.1007/s00284-006-0186-3
63. Mancha-Agresti P, Drummond MM, Carmo FLRD, Santos MM, Santos JSCD, Venanzi F, Chatel J-M, Leclercq SY, Azevedo V (2017) A New Broad Range Plasmid for DNA Delivery in Eukaryotic Cells Using Lactic Acid Bacteria: In Vitro and In Vivo Assays. *Molecular Therapy - Methods & Clinical Development* 4:83–91. doi: 10.1016/j.omtm.2016.12.005
64. Pérez-Arellano I, Zúñiga M, Pérez-Martínez Gaspar (2001) Construction of Compatible Wide-Host-Range Shuttle Vectors for Lactic Acid Bacteria and *Escherichia coli*. *Plasmid* 46:106–116. doi: 10.1006/plas.2001.1531

65. O' Sullivan DJ, Klaenhammer TR (1993) High- and low-copy-number *Lactococcus* shuttle cloning vectors with features for clone screening. *Gene* 137:227–231. doi: 10.1016/0378-1119(93)90011-q
66. Que Y-A, Haefliger J-A, Francioli P, Moreillon P (2000) Expression of *Staphylococcus aureus* Clumping Factor A in *Lactococcus lactis* subsp. *cremoris* Using a New Shuttle Vector. *Infection and Immunity* 68:3516–3522. doi: 10.1128/iai.68.6.3516-3522.2000
67. Williams J (2013) Vector Design for Improved DNA Vaccine Efficacy, Safety and Production. *Vaccines* 1:225–249. doi: 10.3390/vaccines1030225
68. Gonçalves GAL, Bower DM, Prazeres DMF, Monteiro GA, Prather KLJ (2011) Rational engineering of *Escherichia coli* strains for plasmid biopharmaceutical manufacturing. *Biotechnology Journal* 7:251–261. doi: 10.1002/biot.201100062
69. Dixon D, Darveau R (2005) Lipopolysaccharide Heterogeneity: Innate Host Responses to Bacterial Modification of Lipid A Structure. *Journal of Dental Research* 84:584–595. doi: 10.1177/154405910508400702
70. Pontes DS, Azevedo MSPD, Chatel J-M, Langella P, Azevedo V, Miyoshi A (2011) *Lactococcus lactis* as a live vector: Heterologous protein production and DNA delivery systems. *Protein Expression and Purification* 79:165–175. doi: 10.1016/j.pep.2011.06.005
71. Solar GD, Espinosa M (2002) Plasmid copy number control: an ever-growing story. *Molecular Microbiology* 37:492–500. doi: 10.1046/j.1365-2958.2000.02005.x
72. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual* (Ed. 2). Cold Spring Harbor Laboratory Press, New York.
73. Kiewiet R, Kok J, Seegers JF, Venema G, Bron S (1993) The mode of replication is a major factor in segregational plasmid instability in *Lactococcus lactis*. *Applied and Environmental Microbiology* 59:358–364.
74. Emond E, Lavallee R, Drolet G, Moineau S, Lapointe G (2001) Molecular Characterization of a Theta Replication Plasmid and Its Use for Development of a Two-Component Food-Grade Cloning System for *Lactococcus lactis*. *Applied and Environmental Microbiology* 67:1700–1709. doi: 10.1128/aem.67.4.1700-1709.2001
75. Chatelier E, Ehrlich SD, Janni re L (1996) Countertranscript-driven attenuation system of the pAM $\beta$ 1 repE gene. *Molecular Microbiology* 20:1099–1112. doi: 10.1111/j.1365-2958.1996.tb02550.x
76. Brantl S (2015) Antisense-RNA mediated control of plasmid replication – pIP501 revisited. *Plasmid* 78:4–16. doi: 10.1016/j.plasmid.2014.07.004
77. Levin-Karp A, Barenholz U, Bareia T, Dayagi M, Zelcbuch L, Antonovsky N, Noor E, Milo R (2013) Quantifying Translational Coupling in *E. coli* Synthetic Operons Using RBS Modulation and Fluorescent Reporters. *ACS Synthetic Biology* 2:327–336. doi: 10.1021/sb400002n
78. Maier T, G uell M, Serrano L (2009) Correlation of mRNA and protein in complex biological samples. *FEBS Letters* 583:3966–3973. doi: 10.1016/j.febslet.2009.10.036
79. Man JCD, Rogosa M, Sharpe ME (1960) A Medium For The Cultivation Of Lactobacilli. *Journal of Applied Bacteriology* 23:130–135. doi: 10.1111/j.1365-2672.1960.tb00188.x
80. Jones SE, Versalovic J (2009) Probiotic *Lactobacillus reuteri* biofilms produce antimicrobial and anti-inflammatory factors. *BMC Microbiology* 9:35. doi: 10.1186/1471-2180-9-35
81. Terzaghi BE, Sandine WE (1975) Improved medium for lactic streptococci and their bacteriophages. *Applied Microbiology* 29:807–13.
82. Registry of Standard Biological Parts - Ribosome Binding Sites/Prokaryotic/Constitutive/Anderson. iGEM. [http://parts.igem.org/Ribosome\\_Binding\\_Sites/Prokaryotic/Constitutive/Anderson](http://parts.igem.org/Ribosome_Binding_Sites/Prokaryotic/Constitutive/Anderson). Accessed October 17<sup>th</sup>, 2015

83. Chiaruttini C, Milet M (1993) Gene Organization, Primary Structure and RNA Processing Analysis of a Ribosomal RNA Operon in *Lactococcus lactis*. *Journal of Molecular Biology* 230:57–76. doi: 10.1006/jmbi.1993.1126
84. Eichenbaum Z, Federle MJ, Marra D, De Vos WM, Kuipers OP, Kleerebezem M, Scott JR (1998) Use of the lactococcal nisA promoter to regulate gene expression in gram-positive bacteria: comparison of induction level and promoter strength. *Applied and Environmental Microbiology* 64:2763-2769.
85. Novoradovsky A, Zhang V, Ghosh M, Hogrefe H, Sorge JA, Gaasterland T (2005) Computational principles of primer design for site directed mutagenesis. In: *Technical Proceedings of 2005 NSTI Nanotechnology Conference and Trade Show, Vol 1*, pp 532-535
86. Palomino MM, Allievi MC, Prado-Acosta M, Sanchez-Rivas C, Ruzal SM (2010) New method for electroporation of *Lactobacillus* species grown in high salt. *Journal of Microbiological Methods* 83:164–167. doi: 10.1016/j.mimet.2010.08.017
87. Skulj M, Okrslar V, Jalen S, Jevsevar S, Slanc P, Strukelj B, Menart V (2008) Improved determination of plasmid copy number using quantitative real-time PCR for monitoring fermentation processes. *Microbial Cell Factories* 7:6. doi: 10.1186/1475-2859-7-6
88. Borujeni AE, Channarasappa AS, Salis HM (2013) Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Research* 42:2646–2659. doi: 10.1093/nar/gkt1139
89. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology* 27:946–950. doi: 10.1038/nbt.1568
90. Roche (2011) RealTime ready RT-qPCR Assay Development and Qualification. Roche. [https://lifescience.roche.com/wcsstore/RASCatalogAssetStore/Articles/RTR\\_04.11.pdf](https://lifescience.roche.com/wcsstore/RASCatalogAssetStore/Articles/RTR_04.11.pdf). Accessed October 29<sup>th</sup>, 2016
91. Hospodsky D, Yamamoto N, Peccia J (2010) Accuracy, Precision, and Method Detection Limits of Quantitative PCR for Airborne Bacteria and Fungi. *Applied and Environmental Microbiology* 76:7004-7012. doi: 10.1128/AEM.01240-10
92. Hou Y, Zhang H, Miranda L, Lin S (2010) Serious Overestimation in Quantitative PCR by Circular (Supercoiled) Plasmid Standard: Microalgal pcna as the Model Gene. *PLoS ONE*. doi: 10.1371/journal.pone.0009545
93. Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-Sequence Determinants of Gene Expression in *Escherichia coli*. *Science* 324: 255-258. doi: 10.1126/science.1170160
94. Pflieger BF, Pitera DJ, Smolke CD, Keasling JD (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nature Biotechnology* 24:1027–1032. doi: 10.1038/nbt1226
95. Schümperli D, Mckenney K, Sobieski DA, Rosenberg M (1982) Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. *Cell* 30:865–871. doi: 10.1016/0092-8674(82)90291-4
96. Lereclus D, Ribier J, Klier A, Menou G, Lecadet MM (1984) A transposon-like structure related to the delta-endotoxin gene of *Bacillus thuringiensis*. *EMBO Journal* 3:2561. doi: 10.1002/j.1460-2075.1984.tb02174.x
97. Solovyev V, Salamov A (2011) Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In: Li RW (ed) *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*. Nova Science Publishers, pp 61-78
98. Softberry (2017) BPROM – Prediction of bacterial promoters. <http://www.softberry.com/berry.phtml?topic=bprom&group=programs&subgroup=gfindb>. Accessed November 15<sup>th</sup>, 2017
99. Synthetic Biology Project (2017) <http://www.synbioproject.org/>. Accessed October 30<sup>th</sup>, 2017
100. Abil Z, Xiong X, Zhao H (2014) Synthetic biology for therapeutic applications. *Molecular pharmaceutics*, 12(2):322-331. doi: 10.1021/mp500392q

101. Freemont P, Kitney R The UK Synthetic Biology Scene. APBN <http://www.asiabiootech.com/18/1805/18050039x.html>. Accessed November 5<sup>th</sup>, 2017
102. Registry of Standard Biological Parts – iGEM. Help: Synthetic Biology. [http://parts.igem.org/Help:Synthetic\\_Biology](http://parts.igem.org/Help:Synthetic_Biology). Accessed November 3<sup>rd</sup>, 2017
103. Shetty RP, Endy D, Knight TF (2008) Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering* 2:5. doi: 10.1186/1754-1611-2-5
104. BioBricks and Golden Gate Cloning. <https://sites.google.com/a/ncsu.edu/biobricks-and-golden-gate-cloning/assembly-standards-for-biobricks>. Accessed November 2<sup>nd</sup>, 2017
105. Anderson JC, Dueber JE, Leguia M, Wu GC, Goler JA, Arkin AP, Keasling JD (2010) BglBricks: A flexible standard for biological part assembly. *Journal of Biological Engineering* 4:1. doi: 10.1186/1754-1611-4-1
106. Shetty R, Lizarazo M, Rettberg R, Knight TF (2011) Assembly of BioBrick Standard Biological Parts Using Three Antibiotic Assembly. *Methods in Enzymology Synthetic Biology, Part B - Computer Aided Design and DNA Assembly* 311–326. doi: 10.1016/b978-0-12-385120-8.00013-9
107. Gibson D (2009) One-step enzymatic assembly of DNA molecules up to several hundred kilobases in size. *Protocol Exchange*. doi: 10.1038/nprot.2009.77
108. New England Biolabs (2017) NEBuilder Hi-Fi DNA Assembly Master Mix – Instruction Manual (version 1.3). <https://www.neb.com/-/media/catalog/datacards-or-manuals/manuale2621.pdf>. Accessed May 30<sup>th</sup>, 2016
109. Gibson DG (2011) Enzymatic Assembly of Overlapping DNA Fragments. *Methods in Enzymology Synthetic Biology, Part B - Computer Aided Design and DNA Assembly* 349–361. doi: 10.1016/b978-0-12-385120-8.00015-2
110. Hutchison CA, Chuang R-Y, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi ZQ, Richter RA, Strychalski EA, Sun L, Suzuki Y, Tsvetanova B, Wise KS, Smith HO, Glass JI (2016) Design and synthesis of a minimal bacterial genome. *Science* 351:aad6253. <http://dx.doi.org/10.1126/science.aad6253>
111. Gonçalves C (2017) Synthetic biology based assembly of new plasmids for Lactic Acid Bacteria. MSc thesis, Instituto Superior Técnico, Lisboa, Portugal.
112. Holo H, Nes IF (1989) High-frequency transformation, by electroporation, of *Lactococcus lactis* subsp. *cremoris* grown with glycine in osmotically stabilized media. *Applied and environmental microbiology* 55(12):3119-3123.
113. Thompson K, Collins MA (1996) Improvement in electroporation efficiency for *Lactobacillus plantarum* by the inclusion of high concentrations of glycine in the growth medium. *Journal of Microbiological Methods* 26:73–79. doi: 10.1016/0167-7012(96)00845-7
114. Davis MW (2017) APE – a plasmid editor. <http://biologylabs.utah.edu/jorgensen/wayned/ape/>
115. SnapGene® software (from GSL Biotech; available at [snapgene.com](http://snapgene.com)).
116. NEBuilder Assembly Tool. <http://nebuilder.neb.com/>. Accessed: April 3<sup>rd</sup>, 2017.
117. Platteeuw C, van Alen-Boerrigter I, van Schalkwijk S, de Vos WM (1996) Food-grade cloning and expression system for *Lactococcus lactis*. *Applied and Environmental Microbiology* 62(3):1008-1013.
118. Biswas I, Jha JK, Fromm N (2008) Shuttle expression plasmids for genetic studies in *Streptococcus mutans*. *Microbiology* 154:2275–2282. doi: 10.1099/mic.0.2008/019265-0
119. Katsuyama Y, Matsuzawa M, Funa N, Horinouchi S (2008) Production of curcuminoids by *Escherichia coli* carrying an artificial biosynthesis pathway. *Microbiology* 154:2620–2628. doi: 10.1099/mic.0.2008/018721-0
120. Katsuyama Y, Kita T, Funa N, Horinouchi S (2009) Curcuminoid Biosynthesis by Two Type III Polyketide Synthases in the Herb *Curcuma longa*. *Journal of Biological Chemistry* 284:11160–11170. doi: 10.1074/jbc.m900070200

121. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences* 97:6640–6645. doi: 10.1073/pnas.120163297
122. Hidalgo-Cantabrana C, O’Flaherty S, Barrangou R (2017) CRISPR-based engineering of next-generation lactic acid bacteria. *Current Opinion in Microbiology* 37:79–87. doi: 10.1016/j.mib.2017.05.015
123. Song X, Huang H, Xiong Z, Ai L, Yang S (2017) CRISPR-Cas9D10A Nickase-Assisted Genome Editing in *Lactobacillus casei*. *Applied and Environmental Microbiology* 83(22):e01259-17. doi: 10.1128/AEM.01259-17
124. Leenhouts K (1996) A general system for generating unlabelled gene replacements in bacterial chromosomes. *Molecular and General Genetics MGG* 253:217–224. doi: 10.1007/s004380050315
125. Law J, Buist G, Haandrikman A, Kok J, Venema G, Leenhouts K (1995) A system to generate chromosomal mutations in *Lactococcus lactis* which allows fast analysis of targeted genes. *Journal of Bacteriology* 177:7011–7018. doi: 10.1128/jb.177.24.7011-7018.1995
126. Yang P, Wang J, Qi Q (2015) Prophage recombinases-mediated genome engineering in *Lactobacillus plantarum*. *Microbial Cell Factories*. doi: 10.1186/s12934-015-0344-z
127. Pijkeren J-PV, Britton RA (2012) High efficiency recombineering in lactic acid bacteria. *Nucleic Acids Research*. doi: 10.1093/nar/gks147
128. Pijkeren JPV, Britton RA (2014) Precision genome engineering in lactic acid bacteria. *Microbial Cell Factories*. doi: 10.1186/1475-2859-13-s1-s10
129. Oh J-H, Van Pijkeren J-P (2014) CRISPR–Cas9-assisted recombineering in *Lactobacillus reuteri*. *Nucleic Acids Research*. doi: 10.1093/nar/gku623
130. Jang Y, Seo S, Kim S, Li L, Kim T, Kim SC, Jin Y, Han NS (2017) Elimination of the cryptic plasmid in *Leuconostoc citreum* by CRISPR/Cas9 system. *Journal of Biotechnology* 251:151-155. doi: 10.1016/j.jbiotec.2017.04.018
131. Pijkeren J-PV, Barrangou R (2017) Genome Editing of Food-Grade *Lactobacilli* To Develop Therapeutic Probiotics. *Microbiology Spectrum*. doi: 10.1128/microbiolspec.bad-0013-2016
132. Doudna JA, Charpentier E (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science* 346:1258096–1258096. doi: 10.1126/science.1258096
133. Makarova KS, Wolf YI, Alkhnbashi OS, Costa F, Shah SA, Saunders SJ, Barrangou R, Brouns SJJ, Charpentier E, Haft DH, Horvath P, Moineau S, Mojica FJM, Terns RM, Terns MP, White MF, Yakunin AF, Garrett RA, Oost J, Backofen R, Koonin EV (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology* 13(11):722-736. doi: 10.1038/nrmicro3569
134. Sander JD, Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology* 32(4):347-355. doi: 10.1038/nbt.2842
135. Huang H, Zheng G, Jiang W, Hu H, Lu Y (2015) One-step high-efficiency CRISPR/Cas9-mediated genome editing in *Streptomyces*. *Acta Biochimica et Biophysica Sinica* 47:231–243. doi: 10.1093/abbs/gmv007
136. Derkx PM, Janzen T, Sørensen KI, Christensen JE, Stuer-Lauridsen B, Johansen E (2014) The art of strain improvement of industrial lactic acid bacteria without the use of recombinant DNA technology. *Microbial Cell Factories*. doi: 10.1186/1475-2859-13-s1-s5
137. Cherepanov PP, Wackernagel W (1995) Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of FLP-catalyzed excision of the antibiotic-resistance determinant. *Gene* 158:9–14. doi: 10.1016/0378-1119(95)00193-a
138. Domselaar GHV, Stothard P, Shrivastava S, Cruz JA, Guo A, Dong X, Lu P, Szafron D, Greiner R, Wishart DS (2005) BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Research* 1(33)(Web Server issue):W455-9. doi: 10.1093/nar/gki593
139. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS (2011) PHAST: A Fast Phage Search Tool. *Nucleic Acids Research* 39(suppl 2):W347-W352. doi: 10.1093/nar/gkr485

140. Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, Nielsen J, Patil KR, Ferreira EC, Rocha M (2010) OptFlux: an open-source software platform for *in silico* metabolic engineering. BMC Systems Biology 4:45. doi: 10.1186/1752-0509-4-45
141. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2016) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research. doi: 10.1093/nar/gkw1092
142. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M (2015) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Research. doi: 10.1093/nar/gkv1070
143. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research 28(1):27-30. doi: 10.1093/nar/28.1.27
144. Kadam SR, Patil SS, Bastawde KB, Khire JM, Gokhale DV (2006) Strain improvement of *Lactobacillus delbrueckii* NCIM 2365 for lactic acid production. Process Biochemistry 41:120–126. doi: 10.1016/j.procbio.2005.06.007
145. Patel S, Goyal A (2010) Isolation, characterization and mutagenesis of exopolysaccharide synthesizing new strains of lactic acid bacteria. Internet Journal of Microbiology 8(1):3-4.
146. Sawitzke JA, Thomason LC, Bubunenko M, Li X, Costantino N, Court DL (2013) Recombineering: using drug cassettes to knock out genes in vivo. Methods in Enzymology 533:79-102. doi: 10.1016/B978-0-12-420067-8.00007-6
147. Prykhozhij S (2015) CRISPR MultiTargeter. <http://www.multicrispr.net/>. Accessed October 30<sup>th</sup>, 2017
148. Takano E, White J, Thompson CJ, Bibb MJ (1995) Construction of thiostrepton-inducible, high-copy-number expression vectors for use in *Streptomyces* spp. Gene 166:133–137. doi: 10.1016/0378-1119(95)00545-2
149. Schmitt-John T, Engels J (1992) Promoter constructions for efficient secretion expression in *Streptomyces lividans*. Applied Microbiology and Biotechnology. doi: 10.1007/bf00170190
150. *Lactococcus lactis* subsp. *lactis* strain LMG 19460, whole genome shotgun sequencing project, <https://www.ncbi.nlm.nih.gov/nuccore/MUBH01000000>. Accessed March 1<sup>st</sup>, 2018.
151. Kelleher P, Bottacini F, Mahony J, Kilcawley KN, Sinderen DV (2017) Comparative and functional genomics of the *Lactococcus lactis* taxon; insights into evolution and niche adaptation. BMC Genomics. doi: 10.1186/s12864-017-3650-5
152. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA (2003) The COG database: an updated version includes eukaryotes. BMC bioinformatics 4(1):41. doi: 10.1186/1471-2105-4-41
153. Milo R, Phillips R (2015) E-book – cell biology by the numbers. <http://book.bionumbers.org/>. Accessed March 1<sup>st</sup>, 2018.
154. Pe'er I, Felder CE, Man O, Silman I, Sussman JL, Beckmann JS (2004) Proteomic signatures: amino acid and oligopeptide compositions differentiate among phyla. PROTEINS: Structure, Function, and Bioinformatics 54(1):20-40. doi: 10.1002/prot.10559
155. Fan X, Xie L, Li W, Xie J (2014) Prophage-like elements present in *Mycobacterium* genomes. BMC Genomics 15:243. doi: 10.1186/1471-2164-15-243
156. Desiere F, Lucchini S, Canchaya C, Ventura M, Brüßow H (2002) Comparative genomics of phages and prophages in lactic acid bacteria. Lactic Acid Bacteria: Genetics, Metabolism and Applications 73–91. doi: 10.1007/978-94-017-2029-8\_5
157. Oliveira J, Mahony J, Lugli GA, Hanemaaijer L, Kouwen T, Ventura M, Sinderen DV (2016) Genome Sequences of Eight Prophages Isolated from *Lactococcus lactis* Dairy Strains. Genome Announcements. doi: 10.1128/genomea.00906-16
158. Duwat P, Ehrlich SD, Gruss A (1992) Use of degenerate primers for polymerase chain reaction cloning and sequencing of the *Lactococcus lactis* subsp. *lactis* *recA* gene. Applied and environmental microbiology 58(8):2674-2678.

159. Nomura M, Nakajima I, Matsuzaki M, Kimoto H, Suzuki I, Aso H (1997) The N-Terminal Sequence of *Lactococcus lactis* Phosphoglucose Isomerase Purified by Affinity Chromatography Differs from the Other Species. *Archives of Biochemistry and Biophysics* 341:315–320. doi: 10.1006/abbi.1997.9959
160. Gonçalves GAL, Prazeres DMF, Monteiro GA, Prather KLJ (2012) De novo creation of MG1655-derived *E. coli* strains specifically designed for plasmid DNA production. *Applied Microbiology and Biotechnology* 97:611–620. doi: 10.1007/s00253-012-4308-5
161. Poquet I, Saint V, Seznec E, Simoes N, Bolotin A, Gruss A (2000) HtrA is the unique surface housekeeping protease in *Lactococcus lactis* and is required for natural protein processing. *Molecular Microbiology* 35:1042–1051. doi: 10.1046/j.1365-2958.2000.01757.x
162. Llanos RM, Harris CJ, Hillier AJ, Davidson BE (1993) Identification of a novel operon in *Lactococcus lactis* encoding three enzymes for lactic acid synthesis: phosphofructokinase, pyruvate kinase, and lactate dehydrogenase. *Journal of Bacteriology* 175:2541–2551. doi: 10.1128/jb.175.9.2541-2551.1993
163. Holmes DJ, Caso JL, Thompson CJ (1993) Autogenous transcriptional activation of a thiostrepton-induced gene in *Streptomyces lividans*. *The EMBO Journal* 12(8):3183–3191. doi: 10.1002/j.1460-2075.1993.tb05987.x
164. Zhu D, Liu F, Xu H, Bai Y, Zhang X, Saris PEJ, Qiao M (2015) Isolation of strong constitutive promoters from *Lactococcus lactis* subsp. *lactis* N8. *FEMS microbiology letters* 362(16). doi: 10.1093/femsle/fnv107
165. Wright F, Bibb MJ (1992) Codon usage in the G+C-rich *Streptomyces* genome. *Gene* 113:55–65. doi: 10.1016/0378-1119(92)90669-g
166. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. *CA: a cancer journal for clinicians* 61(2):69–90. doi: 10.3322/caac.20107
167. Cancer Research UK, <http://www.cancerresearchuk.org/health-professional/cancer-statistics/worldwide-cancer>, Accessed February 22<sup>nd</sup>, 2018.
168. Lacy AM, García-Valdecasas JC, Delgado S, Castells A, Taurá P, Piqué JM, Visa J (2002) Laparoscopy-assisted colectomy versus open colectomy for treatment of non-metastatic colon cancer: a randomised trial. *The Lancet* 359(9325):2224–2229. doi: 10.1016/S0140-6736(02)09290-5
169. Roukos D (2000) Current status and future perspectives in gastric cancer management. *Cancer Treatment Reviews* 26:243–255. doi: 10.1053/ctrv.2000.0164
170. Gill S, Loprinzi CL, Sargent DJ, Thome SD, Alberts SR, Haller DG, Benedetti J, Francini G, Shepherd LE, Francois Seitz J, Labianca R, Chen W, Cha SS, Heldebrant MP, Goldberg RM (2004) Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: who benefits and by how much? *Journal of clinical oncology* 22(10):1797–1806. doi: 10.1200/JCO.2004.09.059
171. Martenson JA, Willett CG, Sargent DJ, Mailliard JA, Donohue JH, Gunderson LL, Thomas CR, Fisher B, Benson AB, Myerson R, Goldberg RM (2004) Phase III Study of Adjuvant Chemotherapy and Radiation Therapy Compared With Chemotherapy Alone in the Surgical Adjuvant Treatment of Colon Cancer: Results of Intergroup Protocol 0130. *Journal of Clinical Oncology* 22:3277–3283. doi: 10.1200/jco.2004.01.029
172. Shawver LK, Slamon D, Ullrich A (2002) Smart drugs: Tyrosine kinase inhibitors in cancer therapy. *Cancer Cell* 1:117–123. doi: 10.1016/s1535-6108(02)00039-9
173. Aggarwal BB, Sethi G, Baladandayuthapani V, Krishnan S, Shishodia S (2007) Targeting cell signaling pathways for drug discovery: An old lock needs a new key. *Journal of Cellular Biochemistry* 102:580–592. doi: 10.1002/jcb.21500
174. Ravindran J, Prasad S, Aggarwal BB (2009) Curcumin and Cancer Cells: How Many Ways Can Curry Kill Tumor Cells Selectively? *The AAPS Journal* 11:495–510. doi: 10.1208/s12248-009-9128-x
175. Metzler M, Pfeiffer E, Schulz SI, Dempe JS (2012) Curcumin uptake and metabolism. *BioFactors* 39:14–20. doi: 10.1002/biof.1042

176. Padhye S, Chavan D, Pandey S, Deshpande J, Swamy K, Sarkar F (2010) Perspectives on Chemopreventive and Therapeutic Potential of Curcumin Analogs in Medicinal Chemistry. *Mini-Reviews in Medicinal Chemistry* 10:372–387. doi: 10.2174/138955710791330891
177. Basnet P, Skalko-Basnet N (2011) Curcumin: An Anti-Inflammatory Molecule from a Curry Spice on the Path to Cancer Treatment. *Molecules* 16:4567–4598. doi: 10.3390/molecules16064567
178. Johnson JJ, Mukhtar H (2007) Curcumin for chemoprevention of colon cancer. *Cancer Letters* 255:170–181. doi: 10.1016/j.canlet.2007.03.005
179. Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. *Cell* 144(5):646–674. doi: 10.1016/j.cell.2011.02.013
180. Sa G, Das T (2008) Anti cancer effects of curcumin: cycle of life and death. *Cell Division* 3:14. doi: 10.1186/1747-1028-3-14
181. Hatcher H, Planalp R, Cho J, Torti FM, Torti SV (2008) Curcumin: From ancient medicine to current clinical trials. *Cellular and Molecular Life Sciences* 65:1631–1652. doi: 10.1007/s00018-008-7452-4
182. Sharma R, Gescher A, Steward W (2005) Curcumin: The story so far. *European Journal of Cancer* 41:1955–1968. doi: 10.1016/j.ejca.2005.05.009
183. Hoehle SI, Pfeiffer E, Solyom AM, Metzler M (2006) Metabolism of Curcuminoids in Tissue Slices and Subcellular Fractions from Rat Liver. *Journal of Agricultural and Food Chemistry* 54:756–764. doi: 10.1021/jf058146a
184. Usta M, Wortelboer HM, Vervoort J, Boersma MG, Rietjens IMCM, Bladeren PJV, Cnubben NHP (2007) Human Glutathione S-Transferase-Mediated Glutathione Conjugation of Curcumin and Efflux of These Conjugates in Caco-2 Cells. *Chemical Research in Toxicology* 20:1895–1902. doi: 10.1021/tx7002245
185. Anand P, Kunnumakkara AB, Newman RA, Aggarwal BB (2007) Bioavailability of Curcumin: Problems and Promises. *Molecular Pharmaceutics* 4:807–818. doi: 10.1021/mp700113r
186. Prather KLJ, Martin CH (2008) De novo biosynthetic pathways: rational design of microbial chemical factories. *Current Opinion in Biotechnology* 19:468–474. doi: 10.1016/j.copbio.2008.07.009
187. Yamazaki K, Tsunoda A, Sibusawa M, Tsunoda Y, Kusano M, Fukuchi K, Yamanaka M, Kushima M, Nomoto K, Morotomi M (2000) The effect of an oral administration of *Lactobacillus casei* strain *shirota* on azoxymethane-induced colonic aberrant crypt foci and colon cancer in the rat. *Oncology Reports*. doi: 10.3892/or.7.5.977
188. Lin C-C, Lu Y-P, Lou Y-R, Ho C-T, Newmark HH, Macdonald C, Singletary KW, Huang M-T (2001) Inhibition by dietary dibenzoylmethane of mammary gland proliferation, formation of DMBA–DNA adducts in mammary glands, and mammary tumorigenesis in Sencar mice. *Cancer Letters* 168:125–132. doi: 10.1016/s0304-3835(01)00506-7
189. Katsuyama Y, Kita T, Horinouchi S (2009) Identification and characterization of multiple curcumin synthases from the herb *Curcuma longa*. *FEBS Letters* 583:2799–2803. doi: 10.1016/j.febslet.2009.07.029
190. Katsuyama Y, Matsuzawa M, Funa N, Horinouchi S (2007) In vitro synthesis of curcuminoids by type III polyketide synthase from *Oryza sativa*. *Journal of Biological Chemistry* 282(52):37702–37709. doi: 10.1074/jbc.M707569200
191. Katsuyama Y, Hirose Y, Funa N, Ohnishi Y, Horinouchi S (2010) Precursor-Directed Biosynthesis of Curcumin Analogs in *Escherichia coli*. *Bioscience, Biotechnology, and Biochemistry* 74:641–645. doi: 10.1271/bbb.90866
192. Srinivasan M, Sudheer AR, Menon VP (2007) Ferulic acid: therapeutic potential through its antioxidant property. *Journal of clinical biochemistry and nutrition* 40(2):92–100. doi: 10.3164/jcbn.40.92
193. Rodrigues JL, Araújo RG, Prather KL, Kluskens LD, Rodrigues LR (2015) Production of curcuminoids from tyrosine by a metabolically engineered *Escherichia coli* using caffeic acid as an intermediate. *Biotechnology journal* 10(4):599–609. doi:10.1002/biot.201400637

194. Vos WMD, Hugenholtz J (2004) Engineering metabolic highways in Lactococci and other lactic acid bacteria. *Trends in Biotechnology* 22:72–79. doi: 10.1016/j.tibtech.2003.11.011
195. Hugenholtz J, Sybesma W, Groot MN, Wisselink W, Ladero V, Burgess K, Sinderen DV, Piard J-C, Eggink G, Smid EJ, Savoy G, Sesma F, Jansen T, Hols P, Kleerebezem M (2002) Metabolic engineering of lactic acid bacteria for the production of nutraceuticals. *Lactic Acid Bacteria: Genetics, Metabolism and Applications* 217–235. doi: 10.1007/978-94-017-2029-8\_13
196. Konings WN, Poolman B, Driessen AM (1992) Can the excretion of metabolites by bacteria be manipulated? *FEMS Microbiology Letters* 88:93–108. doi: 10.1111/j.1574-6968.1992.tb04959.x
197. Cooper GM, Hausman RE (2016) *The cell: a molecular approach*. Sinauer Associates, Inc., Publishers, Sunderland, MA, U.S.A.
198. Martín JF, Casqueiro J, Liras P (2005) Secretion systems for secondary metabolites: how producer cells send out messages of intercellular communication. *Current Opinion in Microbiology* 8:282–293. doi: 10.1016/j.mib.2005.04.009
199. Nielsen DR, Yoon S-H, Yuan CJ, Prather KLJ (2010) Metabolic engineering of acetoin and meso-2, 3-butanediol biosynthesis in *E. coli*. *Biotechnology Journal* 5:274–284. doi: 10.1002/biot.200900279
200. Grote A, Hiller K, Scheer M, Munch R, Nortemann B, Hempel DC, Jahn D (2005) JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Research*. doi: 10.1093/nar/gki376
201. Kim SF, Baek SJ, Pack MY (1991) Cloning and nucleotide sequence of the *Lactobacillus casei* lactate dehydrogenase gene. *Applied and environmental microbiology* 57(8):2413–2417.
202. Boot HJ, Kolen CP, Andreadaki FJ, Leer RJ, Pouwels PH (1996) The *Lactobacillus acidophilus* S-layer protein gene expression site comprises two consensus promoter sequences, one of which directs transcription of stable mRNA. *Journal of Bacteriology* 178:5388–5394. doi: 10.1128/jb.178.18.5388-5394.1996
203. Schotte L, Steidler L, Vandekerckhove J, Remaut E (2000) Secretion of biologically active murine interleukin-10 by *Lactococcus lactis*. *Enzyme and Microbial Technology* 27:761–765. doi: 10.1016/s0141-0229(00)00297-0
204. Knockaert D, Raes K, Wille C, Struijs K, Camp JV (2012) Metabolism of ferulic acid during growth of *Lactobacillus plantarum* and *Lactobacillus collinoides*. *Journal of the Science of Food and Agriculture* 92:2291–2296. doi: 10.1002/jsfa.5623
205. Gräslund S, Nordlund P, Weigelt J, Hallberg BM, Bray J, Gileadi O, Knapp S, Oppermann U, Arrowsmith C, Hui R, Ming J, dhe-Paganon S, Park HW, Savchenko A, Yee A, Edwards A, Vincentelli R, Cambillau C, Kim R, Kim SH, Rao Z, Shi Y, Terwilliger TC, Kim CY, Hung LW, Waldo GS, Peleg Y, Albeck S, Unger T, Dym O, *et al* (2008) Protein production and purification. *Nature Methods* 5:135–146. doi: 10.1038/nmeth.f.202
206. Béduneau A, Tempesta C, Fimbel S, Pellequer Y, Jannin V, Demarne F, Lamprecht A (2014) A tunable Caco-2/HT29-MTX co-culture model mimicking variable permeabilities of the human intestine obtained by an original seeding procedure. *European Journal of Pharmaceutics and Biopharmaceutics* 87:290–298. doi: 10.1016/j.ejpb.2014.03.017
207. Ferreira AS, Silva IN, Fernandes F, Pilkington R, Callaghan M, Mcclean S, Moreira LM (2014) The Tyrosine Kinase BceF and the Phosphotyrosine Phosphatase BceD of Burkholderia contaminants Are Required for Efficient Invasion and Epithelial Disruption of a Cystic Fibrosis Lung Epithelial Cell Line. *Infection and Immunity* 83:812–821. doi: 10.1128/iai.02713-14