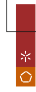
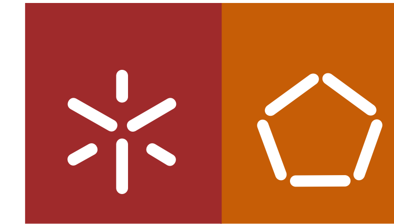


Carla Isabel Pereira Magalhães **The role of the Restriction/Modification system of *Clostridium pasteurianum* on its electro-transformation**

UMinho | 2013

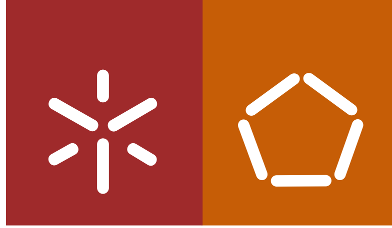


Universidade do Minho
Escola de Engenharia

Carla Isabel Pereira Magalhães

The role of the Restriction/Modification system of *Clostridium pasteurianum* on its electro-transformation

outubro de 2013



Universidade do Minho
Escola de Engenharia

Carla Isabel Pereira Magalhães

The role of the Restriction/Modification system of *Clostridium pasteurianum* on its electro-transformation

Dissertação de Mestrado
Mestrado em Bioengenharia

Trabalho efetuado sob a orientação do
Professor Doutor Manuel José Magalhães Gomes Mota
e coorientação do
Doutor Leonardus Dorothea Kluskens

outubro de 2013

DECLARAÇÃO

Nome: Carla Isabel Pereira Magalhães

Endereço eletrónico: c.pereira.mg@gmail.com; pg19019@alunos.uminho.pt

Número do Bilhete de Identidade: 13808152

Título da dissertação:

The role of the Restriction/Modification system of *Clostridium pasteurianum* on its electro-transformation

Orientador:

Professor Doutor Manuel José Magalhães Gomes Mota

Co-Orientador:

Doutor Leonardus Dorothea Kluskens

Ano de conclusão: 2013

Designação Ramo de Conhecimento do Mestrado:

Ciências Biológicas

DE ACORDO COM A LEGISLAÇÃO EM VIGOR, NÃO É PERMITIDA A REPRODUÇÃO DE QUALQUER PARTE DESTA DISSERTAÇÃO DE MESTRADO

Universidade do Minho, _____ de outubro de 2013

Assinatura:

____

AGRADECIMENTOS

Gostaria de agradecer ao professor Manuel Mota pelo apoio demonstrado desde o início do trabalho, mesmo depois da reviravolta de tema que este sofreu.

À Doutora Lígia Rodrigues e ao Doutor Leon Kluskens, que foram das pessoas mais importantes para a motivação e desenvolvimento crítico ao longo de todo o trabalho. De forma particular gostava de agradecer à Doutora Lígia pela prontidão com que me recebeu no seu projeto e pela oportunidade de estudar um organismo com tanto carisma. Ao Doutor Leon por toda a atenção, disponibilidade e paciência que dedicou ao trabalho bem como à minha pessoa. A sua orientação atenta, sugestões e conselhos fizeram com que todos os contratempos fossem ultrapassados de forma mas tranquila. É bem mais fácil trabalhar quando se é autorizado a cometer erros, antes de mais porque somos humanos e depois porque é uma forma excelente de aprender. Um sincero obrigado a ambos.

À professora Madalena Alves e a todos os elementos do LBA, de forma especial à Sónia e à Raquel, que me fizeram sentir bem-vinda e apoiada em qualquer que fosse o meu dilema. Aos meninos da PBMS, Ilha2, onde a carga de trabalhos e as frustrações encontradas foram maiores, obrigada pelo bom ambiente e apoio mútuo, estes são sem dúvida fatores importantes no desenvolvimento da investigação científica, um trabalho com tantos altos e baixos. Um especial obrigado ao Franklin que apesar de estar sempre sobrecarregado com o próprio trabalho cede, sem qualquer reticência, algum do seu tempo para esclarecer qualquer tipo de dúvida que vá surgindo a pessoas menos experientes. Um outro obrigado sincero à Gracinha pela partilha das frustrações e vitórias.

Aos meninos do mestrado em Bioengenharia, especialmente aos "Só me apetece ganir" pelas magníficas pausas para o café, onde descarregámos as frustrações de quem está agora a começar e onde nos era permitido dizer as maiores barbaridades e que só nós compreendemos.

A todos os amigos que acompanharam esta viagem direta ou indiretamente e que me apoiaram incondicionalmente fosse qual fosse o estado de espírito. Um obrigado carinhoso à Carlinha que apesar de estar na mesma condição e com dilemas semelhantes não deixava de me pôr para cima e de me dizer que tudo ia correr bem.

À pessoa mais carinhosa do mundo, o Clip, obrigada por todo o apoio e valorização. Também és a pessoa que me mantém acima da linha de água.

Nada disto seria possível sem o apoio incondicional da família. Espero que todo o esforço da parte deles para que chegasse a este momento seja realmente recompensado. Eu sei que não foi fácil. À Patrícia, que apesar de viver comigo, nem sempre recebeu a merecida atenção da irmã.

Como não poderia deixar de ser, a todos aqueles que mesmo não mencionados fizeram desta minha jornada um fardo menos pesado.

ABSTRACT

Clostridium pasteurianum is a Gram-positive and anaerobic bacterium with a great biotechnological potential. It is one of the few microorganisms capable of hydrolyzing glycerol to produce solvents as ethanol and butanol, which have a wide applicability in the market as biofuels.

The development of a genetic system for this microorganism would increase its application opportunities since gene overexpression or inactivation could improve their solventogenic characteristics. Its genetic information is already known but this organism has a particular resistance to transformation. This resistance can be explained by a very efficient restriction system that does not allow the entrance of non-methylated DNA or DNA with a methylation pattern different from it. Therefore, foreign DNA must be correctly methylated prior to transformation. For this purpose, a specific methyltransferase is needed to transfer methyl groups to a certain nucleotide of a specific sequence.

The goal of this thesis was to create a genetic system in *C. pasteurianum* that allows genome modification and foreign protein expression, ultimately improving *C. pasteurianum* DSM 525 transformation.

Preliminary simple electro-transformations in which the parameters to make competent cells and the electroporation conditions were altered, did not result in positive results.

Being aware of the possibility of a restriction system presence in this organism, experiments with M.*Msp*I methylated DNA were performed, however they demonstrated the inability of this methyltransferase to improve the microorganism transformation.

The presence of restriction enzymes was confirmed when a characterization of the restriction system of *C. pasteurianum* was performed using *Msp*I methylated and non-methylated DNA. The presence of a discrete digestion pattern was detected, and M.*Msp*I methylation could not protect the foreign DNA from *C. pasteurianum* restriction action.

The polyamine spermidine, with known affinity for negatively charged DNA, showed to be efficient against *C. pasteurianum* crude extract digestion action, however not sufficiently to facilitate this microorganism electro-transformation.

By accessing the genome information, the Restriction/Modification (R/M) systems of this microorganism were analyzed. The GATC type IIP R/M system was chosen in order to verify the restriction and methylation enzymes activity with the same target sequence. Three genes, one REase (DpnII) and two MTases (Dam and MdpnII) were cloned in pETduet-1, followed by overproduction in BL21 (DE3).

The codon usage of the host and original organism were not compatible, and the protein production in tRNAs provider strains was tested. Protein production was detected, however was not possible to re-confirm their presence.

The common protein folding problems were analyzed using a disulfide bond enhancer strain. Nevertheless, the production problem may not be related to this, since no different protein over-production was detected.

Restriction reactions with the REase *Bst*UI and *C. pasteurianum* crude extract, using DNA methylated by M.*Sss*I (^{m5}CG), were developed and showed that the REase responsible for hindering foreign DNA entering *C. pasteurianum* recognizes the sequence 5'-CGCG- 3'.

In a second analysis of the *C. pasteurianum* genome a methyltransferase-encoding gene was identified that may be involved in methylating the sequence 5'-CGCG- 3'. The *in silico* analysis was performed and its codon usage was also improved to be compatible with *E. coli*.

In this work, the reasons for *C. pasteurianum's* recalcitrance to transformation were identified, the knowledge about its R/M systems was extended, and a proposal to efficiently transform this bacterium was provided.

RESUMO

Clostridium pasteurianum DSM 525 é uma bactéria Gram-positiva anaeróbia com um elevado potencial biotecnológico. Este é um dos poucos microrganismos capaz de hidrolisar glicerol para produzir solventes como etanol e butanol, que têm uma grande aplicabilidade no mercado.

O desenvolvimento de um sistema genético para este organismo permitiria aumentar as suas oportunidades de aplicação sendo que a sobre-expressão ou inativação de um determinado gene pode melhorar as suas características solventogénicas. A sua informação genética já é conhecida, mas este microrganismo apresenta uma particular resistência à transformação. Esta resistência pode ser explicada pela presença de um eficiente sistema de restrição que não permite a entrada de DNA não metilado ou DNA metilado de forma diferente da própria bactéria. Desta forma, o DNA estranho deve ser corretamente metilado antes da transformação. Para que isto seja possível é necessária a presença de uma metilase específica para transferir grupos metilo para um determinado nucleótido de uma sequência específica.

O objetivo desta tese foi criar um sistema genético em *C. pasteurianum* que permitisse modificações no genoma e a expressão de proteínas heterólogas, ou seja, que permitisse melhorar a transformação de *C. pasteurianum* DSM 525.

Foram realizadas transformações preliminares simples com parâmetros que diferem na forma de obter células competentes e nas condições de eletroporação, contudo os resultados obtidos não foram positivos.

Tendo conhecimento da possibilidade da presença de um sistema de restrição neste organismo, foram realizadas experiências com DNA metilado pela enzima M.*Msp*I, sendo que estas demonstraram a incapacidade da metiltransferase para melhorar a transformação deste microrganismo.

Foi confirmada a presença de enzimas de restrição aquando da caracterização do sistema de restrição de *C. pasteurianum* usando DNA não metilado ou metilado pela enzima M.*Msp*I. Foi detetada a presença de um padrão de digestão distinto, verificando-se que a enzima M.*Msp*I não consegue proteger o DNA estranho da ação de restrição de *C. pasteurianum*.

A poliamina espermidina, com conhecida afinidade por DNA negativamente carregado, mostrou ser eficiente contra a ação de digestão do extrato cru de *C. pasteurianum*, contudo não o suficiente para facilitar a electro-transformação deste microrganismo.

Tendo acesso ao genoma, foi então analisado o sistema de Restrição e Modificação (R/M) deste microrganismo. Foi escolhido o sistema R/M tipo IIP GATC para verificar a atividade de enzimas de restrição e metilação com a mesma sequência de reconhecimento. Foram clonados três genes no vetor pETduet-1, uma enzima de restrição (REase – DpnII) e duas metiltransferases (MTases – Dam and Mdpn), seguindo-se a produção em BL21 (DE3). O conjunto de codões usados pelo hospedeiro e pelo organismo de origem não eram compatíveis, foi então testada a produção proteica em estirpes fornecedoras de tRNAs. Foi observada produção proteica contudo não foi possível re-avaliar a sua presença. Foram analisados problemas de enrolamento (do inglês *folding*), usando uma estirpe que facilita a formação de pontes dissulfito. No entanto, o problema na produção não deve estar associado ao enrolamento proteico sendo que não foi detectada produção proteica nestas condições.

Foram desenvolvidas reações de restrição com a REase *Bst*UI e extrato cru de *C. pasteurianum* usando DNA metilado pela enzima M.*Sss*I (^{m5}CG) e foi mostrado que a REase responsável pelo impedimento da entrada de DNA em *C. pasteurianum* reconhece a sequência 5'-CGCG- 3'.

Numa segunda análise ao genoma de *C. pasteurianum* DSM 525 foi identificado um gene que codifica uma metiltransferase que pode estar envolvida na metilação da sequência 5'-CGCG- 3'. Foi feita a análise *in silico* e o tipo de codões usados foi melhorado para ser compatível com *E. coli*.

Neste trabalho, foram identificadas as razões para a resistência deste microrganismo à transformação, foi consolidado o conhecimento sobre o seu sistema de R/M e foi proposta uma metodologia para transformar de forma eficiente esta bactéria.

TABLE OF CONTENTS

AGRADECIMENTOS iii
ABSTRACT v
RESUMO vii
LIST OF FIGURES xii
LIST OF TABLES xiv
ABBREVIATIONS xv
1. LITERATURE REVIEW 3
1.1. *Clostridium* Genus 3
1.2. *Clostridium pasteurianum* in Biofuel Industry 4
1.3. Transformation in *Clostridium* 6
1.3.1. Bacterial transformation 7
1.3.2. Electroporation 7
1.3.3. *Clostridium* transformation 9
1.3.4. R/M Systems 10
1.3.4.1. Types of restriction enzymes 10
1.3.4.2. Type II 12
1.3.5. *Clostridium* R/M systems 13
1.4. ClosTron system in *Clostridium* transformation 17
1.5. Aim 21
2. MATERIALS AND METHODS 25
2.1. Bacterial strains and plasmids 25
2.2. Genes 26
2.3. Primers 26
2.4. Culture media 27
2.5. Nucleic acids methods 27
2.5.1. Plasmid isolation of *E. coli* and DNA purification 27
2.5.2. Genomic DNA extraction from *C. pasteurianum* DSM 525 28
2.5.3. DNA sample analysis by agarose gel electrophoresis and DNA concentration determination 28
2.5.4. Gene Amplification (Polymerase Chain Reaction - PCR) 28
2.5.5. Colony PCR 29
2.5.6. DNA Sequencing 30
2.6. Cloning 30

2.6.1. Double digestion 30
2.6.2. Dephosphorylation of restriction fragments 31
2.6.3. Ligation of DNA fragments 31
2.6.4. Introduction and maintenance in a host organism 31
2.7. *C. pasteurianum* crude extract....32
2.8. DNA modification....32
2.8.1. Plasmid enzymatic cleavage....32
2.8.2. DNA methylation....32
2.8.3. DNA condensation by spermidine 33
2.8.4. *C. pasteurianum* restriction endonuclease assay....33
2.9. Bacterial cell transformation 33
2.9.1. Electro-transformations *of C. pasteurianum* DSM 525....33
2.9.1.1. *C. pasteurianum* electrocompetent cells....34
2.9.1.2. *C. pasteurianum* electro-transformation....34
2.9.2. *E. coli* transformation....35
2.9.2.1. Chemically competent *E. coli* cells 35
2.9.2.2. Heat shock transformation of *E. coli*....35
2.10. Protein Production and analysis 36
2.10.1. Protein Expression 36
2.10.2. Cell rupture....36
2.10.3. SDS-PAGE analysis 36
2.10.3.1. Coomassie Blue staining....37
2.10.3.2. Silver Staining 37
2.10.4. Protein purification....38
2.10.5. Western Blottting....39
3. RESULTS AND DISCUSSION....43
3.1. Glycerol as *C. pasteurianum* carbon source....43
3.2. Preliminary electro-transformations of *C. pasteurianum*....44
3.3. M.*Msp*I role on *C. pasteurianum* DSM 525 transformation....45
3.4. Crude extract of *C. pasteurianum* exhibits restriction activity....46
3.5. DNA condensation by the polyamine spermidine 47
3.6. *dpnII*, *dam* and *mdpn* genes as a *C. pasteurianum* R/M System 49
3.7. Cloning....52
3.7.1. *dpnII*, *dam* and *mdpn* genes amplification....52
3.7.2. Vector – inserts construction....53

3.8. Protein production 56
3.8.1. BL21 (DE3) production 57
3.8.2. Improvement of the conditions for protein production 57
3.8.2.1. Low temperatures of protein production 57
3.8.2.2. Codon Usage Analysis 58
3.8.2.3. BL21 (DE3) RIL and Rosetta (DE3) pLysS production 59
3.8.2.4. Enzymatic experiments 61
3.8.2.5. Origami (DE3) production 61
3.9. *dcm* gene importance in *C. pasteurianum* DSM 525 R/M system 62
3.10. *C. pasteurianum* DSM 525 REase identification 62
4. CONCLUDING REMARKS AND FUTURE PERSPECTIVES 69
5. REFERENCES 73
6. APPENDIX 83
I. Media and solutions recipe 83
II. Gene amplification sequence by PCR 84
III. SDS-PAGE protein production results 85
IV. Multiple sequence alignment 86
V. Primer list designed to further *dcm* gene amplification 87

LIST OF FIGURES

Chapter 1

Figure 1.1. US biodiesel production and its impact on crude glycerol prices. 5

Figure 1.2. Sequence of events occurring in the bilayer membrane when subjected to an electric field as electroporation. 8

Figure 1.3. Specific methylation ceases restriction endonucleases action. 14

Figure 1.4. Schematic representation of mutant generation using the ClosTron system. 19

Chapter 2

Figure 2.1. pET-24d+ (a) and pETduet (b) vector maps. 30

Figure 2.2. Simplified protein purification process using the His-Pur™ Ni-NTA columns with the GFP protein. 38

Chapter 3

Figure 3.1. *C. pasteurianum* growth using glucose (20 g/L) and glycerol (20 g/L) as a carbon source. 43

Figure 3.2. *C. pasteurianum* crude extract characterization.. 47

Figure 3.3. Spermidine condensation protects DNA of *C. pasteurianum* crude extract restriction. 48

Figure 3.4. Genomic organization of the genes of GATC R/M system of *C. pasteurianum* DSM 525 are presented as cluster on this microorganism genome.. 51

Figure 3.5. Gene amplification by PCR. 52

Figure 3.6. pET24d(+) enzyme digestions.. 53

Figure 3.7. Vectors used to replace pET24d(+). 54

Figure 3.8. Colony PCR results from the transformation of JM109 *E. coli* with the vector pETduet-1 ligated to genes *dam*, *dpnII* or *mdpn*. 55

Figure 3.9. Length analysis of plasmids extracted from the cells transformed with the ligation pETduet-1 and the genes *mdpn* (pCM001), *dpnII* (pCM002) and *dam* (pCM003).. 56

Figure 3.10. SDS-PAGE analysis of the CFE after BL21 (DE3) production of the enzymes Dam (a) and DpnII (b) at 37°C. 57

Figure 3.11. SDS-PAGE analysis of BL21 (DE3) RIL CFE production of the enzymes Dam (a) DpnII (b) and Mdpn (c) at 37°C.. 59

Figure 3.12. SDS-PAGE gel of His-tag purified CFE of the BL21 (DE3) RIL DpnII (a) and Mdpn (b) 4.5 h sample production. 60

Figure 3.13. Western blotting result of the 4.5 h sample of Mdpn production in BL21 (DE3) RIL. 61

Figure 3.14. Restriction reaction of ^{m5}CG methylated and non-methylated pMTL007C-E2::Cpa-*spo0A*-666a with the enzymes *Dpn*II, *Bst*UI and *C. pasteurianum* crude extract.. 63

Figure 3.15. Restriction activity comparison between *C. pasteurianum* (Cp) crude extract (C. Ext) and the enzyme *Bst*UI. 64

Chapter 6

Figure 6.1. SDS-PAGE analysis of BL21 (DE3) CFE production of the enzymes Dam (a) Mdpn (b) and DpnII (c) at 16 ◦C.. 85

Figure 6.2. SDS-PAGE analysis of Rosetta (DE3) pLysS CFE and TE production of the enzymes Dam (a) DpnII (b) and Mdpn (c) at 37 ◦C. 85

Figure 6.3. SDS-PAGE analysis of Origami (DE3) CFE and TE production of the enzymes Dam (a) DpnII (b) and Mdpn (c) at 37 ◦C.. 85

Figure 6.4. Alignment of the *dcm* gene with codon usage optimized (opC) and the conventional *dcm* gene (dcm) from *C. pasteurianum*. 87

LIST OF TABLES

Chapter 1

Table 1.1.General properties of the four restriction enzyme types. 13
Table 1.2. *Clostridium* REases identified. 13
Table 1.3. Type II R/M Systems present in *Clostridium* listed in REBASE 16

Chapter 2

Table 2.1. Bacterial Strains and plasmids. 25
Table 2.2. List of primers used. 26
Table 2.3. PCR componentes and temperature program conditions to gene amplification with Kapa HiFi Polymerase. 29
Table 2.4. PCR components and temperature program conditions to gene amplification with KapaTaq DNA polymerase. 29
Table 2.5. *C. pasteurianum* electro-transformation parameters. 35

Chapter 3

Table 3.1. *C. pasteurianum* preliminary electrotransformation parameters. 45
Table 3.2. *C. pasteurianum* electrotransformation parameters using *Msp*I methylated DNA. 46
Table 3.3. *C. pasteurianum* electrotransformation parameters using spermidine-condensed DNA. 49
Table 3.4. Genes and their corresponding proteins from *C. pasteurianum* DSM 525 genome related with type IIP R/M system. 50

Chapter 6

Table 6.1. Recipe for mRCM media 83
Table 6.2. Recipe for T68 basal media. 83
Table 6.3. Acrylamide/Bis-acrylamide gels constitution.. 83
Table 6.4. List of primers to use in *dcm* cloning process.. 87

ABBREVIATIONS

ABE fermentation – Acetone - Butanol - Etanol fermentation

AdoMet – S-adenosyl-L-methionine

Amp^R – Ampicilin Resistance

APS – Ammonium persulfate

ATP – Adenosine Triphosphate

BSA – Bovine Serum Albumin

C. – *Clostridium*

°C – Degree Celsius

$CaCl_2$ – Calcium Chloride

CAI – Codon Adaptation Index

CAGR – Compound Annual Growth Rate

CFE – Cell Free Extract

Cm^R – Chloramphenicol / Thiamphenicol Resistance

Da – Dalton(s)

DNA – Deoxyribonucleic Acid

DNP – DNA polymerase

dNTP – Deoxyribonucleotide triphosphate

E. – *Escherichia*

EDTA – Ethylenediaminetetraacetic acid

Em^R – Erythromycin Resistance

et al. – And others (*et alii*)

F – Farad

FRT - FLP recognition target sites

H_2 – Hydrogen

His-Tag - Histidine tag (six histidines in a row)

IEP – Intron Encoded Protein

IMAC – Immobilized metal affinity chromatograthy

Inc. – Incorporation

IPTG - Isopropyl β-D-1-thiogalactopyranoside

Kan^R – Kanamycin Resistance

L - Liter

M – Molar

m – Mili (10^{-3})

MCS – Multiple Cloning Site

min – Minute (s)

mRCM – Modified Reinforced Clostridial Medium

ms – milliseconds

MTase - Methyltransferase

μ – micro (10^{-6})

n – Nano (10^{-9})

N_2 – Nitrogen

NEB – New Englands Biolabs

NTA – Nickel-charged nitrilotriacetic acid

O_2 – Oxygen

OC – Open circular plasmid form

O.D. – Optical Density

o/n - overnight

Ω – Ohm

PCR – Polymerase Chain Reaction

PDO – Propanediol

pg – pictograms

RA – Restriction alleviation

RAM – Retrotransposition Activated Marker

REase – Restriction Endonuclease

RNA – Ribonucleic Acid

RNP – Ribonuclear Protein

rpm – Rotations per minute

rRNA – Ribossomal Ribonucleic Acid

RT – Room Temperature

s – Second (s)

SAM – S-adenosyl-L-methionine

SDS-PAGE – Sodium dodecyl sulfate - Polyacrylamide gel electrophoresis

SOB – Super optimal broth

SOC – Super optimal broth for catabolite repression

TAE – Tris-Acetate-EDTA Buffer

TBST – Tris-buffered saline with tween

Tm – Melting temperature

V – Volts

v – Volume

v/v – Volume per volume

WT – Wild type

w/v – Weight per volume

Chapter 1

Literature Review

1. LITERATURE REVIEW

1.1. *Clostridium* Genus

The genus *Clostridium* comprises a diverse group of bacteria that share several features; they are Gram-positive bacteria, strictly anaerobic, form endospores, present a rod-shaped morphology and are unable to reduce sulfate. The associated wide range of metabolic and physiological diversity makes it one of the largest group in the prokaryote kingdom with more than 200 known species (Dürre, 2007). The anaerobic bacteria became subject of scientific research when in 1861, the famous microbiologist Louis Pasteur, discovered that life is possible without oxygen. Although, some of the *Clostridium* genus on health effects were much earlier described by the Greek physician Hippocrates (460-377 BC) when he was analyzing the lockjaw and gas gangrene diseases.

The classification of microorganisms has long been based solely on physiological characteristics, which has led to some wrong classifications. Currently, 16S rRNA gene sequence analysis allows for a much more sophisticated taxonomic classification and has triggered the reclassification of some species, formerly classified as Clostridia. The phylum Firmicutes is structured into the class Clostridia, the order Clostridiales and several families, among them the Clostridiaceae. This family consists of 16 genera, one of them being Clostridium (Stackebrandt, 2004). The *Clostridium* designation was introduced in the literature by Trécul and Prazmowski between 1865 and 1880. It was the cells morphology that gave them the name, from the Greek kloster (κλωστήρ) or small spindle (Dürre, 2007).

Currently, the public reputation of clostridia is ill-famed; it is usually associated with severe diseases. However, only a few members of this genus are pathogens that cause debilitating and life-threatening intoxications, like botulism (*Clostridium botulinum*), tetanus (*Clostridium tetani*), gangrene (*Clostridium perfringens*), antibiotic associated diarrhea (*Clostridium difficile*) and food poisoning (*C. perfringens*). The vast majority of clostridia are entirely benign and some possess properties and attributes of great benefit that can have an important role in several research areas.

In bioremediation, some species are capable of dehalogenation of insecticides as tetrachloromethane or reduction of nitroaromatics such as trinitrotoluene (TNT) to the corresponding amino compounds (*Clostridium acetobutylicum*, *Clostridium pasteurianum*, *Clostridium bifermentans*, *Clostridium sordellii* and *Clostridium sporogenes*) (Dürre, 2007).

Within the medical field, namely in cancer therapy, a new and recent medical application uses recombinant clostridial spores to fight cancer (Umer et al., 2012), using the limitation of spores to only germinate in anaerobic or hypoxic environments such as the tumor microenvironment. Therefore, recombinant spores (where a gene product is able to convert a harmless pro-drug into a cytotoxic drug) will grow and provide specific targeting and multiplication at the tumor without any harmful effect to the neighboring healthy cells.

Clostridia, and more specifically the solventogenic species, are also applied in the industrial biofuels field. They are useful due to their capacity to produce organic solvents such as acetone (Liou et al., 2005), ethanol (NG et

al., 1981), and 1-butanol (Taconi et al., 2009), by fermentation. This application is increasing tremendously in an attempt to reduce the dependency of petroleum-based fuels and the environmental concerns associated with their use (Yazdani and Gonzalez, 2007).

In conclusion, *Clostridium* is one of the largest bacterial genera and belonging to this genus are important pathogens, but also organisms with impressive biotechnological and medical applications. Because of these huge capabilities they are considered industrial avant-garde microorganisms with a lot of metabolic potential that still awaits elucidation and commercial exploitation (Dürre, 2007).

1.2. *Clostridium pasteurianum* in Biofuel Industry

Owing to the enormous development of science and technology in the last two centuries, the worldwide population lifestyle has changed drastically, becoming much more comfortable and consumerist. The great amount of energy required, as well as other products, resulted in the exploitation of our fossil fuels (e.g., petroleum, coal and natural gases). However, as the third Newton's law states, every action has a reaction. With the advantages of progress came also the environmental problems, such as air pollution, climate changes, global warming and global economic problems. The steadily decreasing availability of fossil fuels is also considered a concerning element. These issues have raised the incentive to paradigm-changing ideas and to create cleaner technologies with the same energy capacity (Sakuragi et al., 2011).

The introduction of biofuels as an alternative has been analyzed and proved to have great chances to reduce the dependence on fossil fuels, thus they are considered a promising approach in this area. Bioethanol (produced through the microbial fermentation of sugars derived from corn, sugarcane or sugar beet (Gray et al., 2006)), biodiesel (produced by the transesterification of vegetable oils or animal fats with an alcohol to produce esters (Marchetti et al., 2007)), and biobutanol (a fermentation product of anaerobic bacteria (Dürre, 2008)) are some examples of biofuels that can replace fossil fuels. Biobutanol has been gaining more attention than other biofuels and is expected to contribute significantly to the industrial needs in the very near future. Biobutanol is very interesting because it presents better properties than ethanol, such as lower corrosiveness, lower vapor pressure, lower water solubility, higher heating value and much higher energy content. Besides its application as a biofuel, butanol is also an important bulk chemical with a wide range of industrial uses. This solvent was, until the middle of the last century, mainly produced by ABE (Acetone-Butanol-Etanol) fermentation using anaerobic microorganisms of *Clostridium* species. However, with the declining price of crude oil and the increase of carbon source prices, it began to be produced almost exclusively by petrochemical processes. Nowadays, mainly due to the oil crisis, the interest in using biotechnology for butanol production has been renewed (Dürre, 2008). It is important to notice that even the fossil fuels based companies, such as BP or DuPont, have shown interest in developing alternative fermentative butanol processes (Lee et al., 2008).

Although it has been proved that these alternatives are secure, renewable and environmentally safe, the production of biofuels also presents some issues, such as low and nonspecific productivity, high cost, and low

solvent tolerance of microorganisms. Therefore, their economic viability is a concern (Sakuragi et al., 2011). The alternatives to continue the proliferation and co-integration with fossil fuels of biofuels are reducing the production cost of and increasing the revenues from industrial biofuels. To achieve these goals the use of biorefineries has been proposed (Kamm and Kamm, 2007). In this case the objective is to use the same feedstock for the co-production of a higher value, small-market chemical along with the biofuel(s). The revenues of both products will improve the economics of biofuel production. To economically improve this process, the biorefinery should consider the use of by-products or waste-streams of the biofuel as feedstock (Yazdani and Gonzalez, 2007). Since the higher investment of an industrial fermentation process is associated with feedstock, this is the principal target for cost reduction.

The biodiesel industry is facing a big expansion in the market. From this process about 10% (w/w) glycerol is obtained as waste-stream (Johnson and Taconi, 2007; da Silva et al., 2009; Yang et al., 2012). As biodiesel is produced in great amounts, more than the current needs, there is a stock accumulation of glycerol, resulting in a substantial fall in its price (Figure 1.1), (Yazdani and Gonzalez, 2007). In 2011, 1,995.5 kilo tons of glycerol were globally produced and in 2018 this value is expected to reach 3,060.4 kilo tons with a CAGR (Compound Annual Growth Rate) of 6.3% from 2012 to 2018 (http://www.transparencymarketresearch.com).

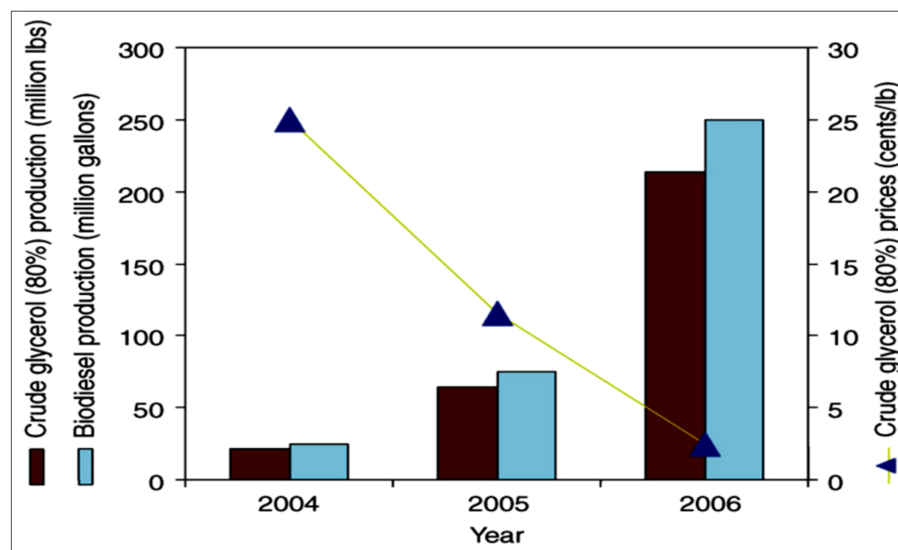


Figure 1.1. US biodiesel production and its impact on crude glycerol prices. In two years the glycerol price presented a 10-fold decrease. From (Yazdani and Gonzalez, 2007).

Crude glycerol as a by-product in biodiesel production is considered a hazardous waste (it contains other impurities such as methanol, ash/salts and residual fatty acids), even though it can be used to produce a large array of products such as ethanol (Jarvis et al., 1997), 1,3-propanediol (1,3-PDO) (Biebl et al., 1992; Yazdani and Gonzalez, 2007), succinic acid (Lee et al., 2001), hydrogen (Ito et al., 2005) and butanol (Biebl, 2001). Glycerol can be highly competitive with the common fermentation sugars, not only due to the lower price, but also because it contains highly reduced carbon atoms, enabling higher fuel yield and highly reduced chemicals (Yazdani and Gonzalez, 2007). Taconi and collaborators defended that glycerol will become a versatile building block chemical for

the production of high value compounds within an integrated biorefinery (Taconi et al., 2009). The advantages of crude glycerol recycling are huge: not only will the discard of a hazardous waste be hindered, but also a cheaper and good carbon source for many fermentation processes will be obtained, fomenting industrial biofuels production.

As mentioned before, a variety of microorganisms can metabolize glycerol, but in general, Clostridia are preferred. These particular anaerobic bacteria are generally not pathogenic, only need little vitamin supplementation and minimal nutrient supply to grow and produce a variety of products (Taconi et al., 2009) such as 1,3-PDO, lactate, acetate, butyrate, ethanol, butanol and even hydrogen (H_2) (Dabrock et al., 1992). Great research and investment efforts are currently being conducted regarding the potential use of solventogenic clostridia for the development of biotech processes towards the production of biofuels, with special attention to biobutanol production. The most studied butanol producing clostridia are *C. acetobutylicum* and *Clostridium beijerinckii* but these do not have the capacity to use glycerol as a sole carbon source, only metabolize glycerol in the presence of glucose (Vasconcelos et al., 1994). *C. butyricum* can ferment glycerol as sole carbon source but it does not express a functional butanol biosynthetic pathway (Heyndrickx et al., 1991). *C. pasteurianum* is a solventogenic species capable of metabolizing glycerol as a sole carbon source tolerating high concentrations (Biebl, 2001; Dabrock et al., 1992; Nakas et al., 1983) and produces higher amounts of PDO, butanol, and ethanol with just trace amounts of acetate and butyrate, compared to the former mentioned clostridia. The product yield obtained is extremely high, from 75% to 90%, with very small carbon loss by CO_2, meaning that most of the substrate is converted to product. Recent studies also demonstrated that this anaerobic bacterium can efficiently grow and obtain the same products when biodiesel-derived crude glycerol is used (Ahn et al., 2011; Jensen et al., 2012a; Taconi et al., 2009; Venkataramanan et al., 2012). The main disadvantage associated to *C. pasteurianum* crude glycerol bioconversion is the slow process kinetics (Khanna et al., 2012).

Therefore, the butanol production by *C. pasteurianum* using biodiesel waste streams containing glycerol as a carbon source is of particular interest, and has garnered special attention unlike the previous biobutanol workhorses *C. acetobutylicum* and *C. beijerinckii*. The lack of knowledge on its genetics and metabolism explains the scarce investment in the use of this species at an industrial level, until now. Since the *C. pasteurianum* DSM 525 genome has been recently published (Rappert et al., 2013), it will certainly be deeply investigated and it is expected that the microorganism's capacities will be greatly enhanced.

1.3. Transformation in *Clostridium*

As previously mentioned, *Clostridium* is one of the largest bacterial genera with medical, biotechnological and environmental relevance. Despite the broad range of applications the basic biology of this genus is not well known when compared with their aerobic counterparts of the *Bacillus* genus (Heap et al., 2009). The better way to recognize and counterattack some pathogens action or exploit biotechnological or environmental properties for better applications, is to understand and thereafter modify essential biological processes at the molecular level. Thus, the need for tools that enable the genetic manipulation of this microorganism becomes clear.

1.3.1. Bacterial transformation

Genetic manipulation of microorganisms started with Frederick Griffith´s investigation (1928), when the development of a vaccine for pneumonia caused by *Streptococcus pneumoniae* led him to discover that biological material ("a transforming principle") from a pathogenic *Streptococcus* strain could transform a non-pathogenic *Streptococcus* strain, making it pathogenic (Griffith, 1928). The explanation was given by Avery and collaborators in the 1940´s. They demonstrated that, after isolation of the biological material, DNA molecules from dead pathogenic strains could be transferred to non-pathogenic strains causing a heritable change (transformation). Two other natural processes of bacterial genetic modification were later discovered: the conjugation process, which is the transfer of DNA from one cell to another through direct contact, and the transduction process, defined as DNA transfection by bacteriophage infection (Prescott et al., 2005). Comparative analyses of prokaryotic genomes show that acquisition of genetic material through lateral gene transfer has been a major driving force in the evolution of these organisms (Johnsborg et al., 2007). Transformation is then defined as the uptake and expression of foreign DNA by cells. Specifically, transformation of bacteria involves DNA adsorption to the cell surface followed by uptake across the wall-membrane complex into the cytoplasm (Prescott et al., 2005).

Even though gene acquisition through natural transformation has contributed significantly to the adaptation and ecological diversification of several bacterial species, it is still a rare event for most bacteria. In modern bacterial genetics, these natural processes are of limited value, because ideally, DNA should be easily transferable into and stably maintained in any given bacterial species (Aune and Aachmann, 2010). Therefore, it was necessary to develop efficient artificial systems to transfer DNA, and currently, artificial transformation of many organisms is commonly done in many laboratories. The first report on the introduction of exogenous DNA into *Escherichia coli* was demonstrated by Mandel and Higa in 1970. They treated a suspension of *E. coli* cells (which are not naturally competent) and bacteriophage λ DNA with a solution of $CaCl_2$ at 0°C and observed cell transformation without using the phage particle. Since then, many other processes were developed for transformation of bacteria but, with the two main methods for introduction of exogenous DNA into bacterial cells being chemical transformation and electroporation (Aune and Aachmann, 2010).

1.3.2. Electroporation

Electroporation or electropermeabilisation is the term applied to describe the appearance of pores in artificial or cellular membranes due to an elevated transmembrane voltage caused by an applied electric field (Neumann et al., 1982). When the transmembrane voltage exceeds a certain threshold (normally 0.2 to 1 V), the rearrangement of the molecular structure of the membrane occurs, leading to the formation of pores that allow the entry of a range of molecules. These can be drugs, antibodies, DNA and plasmids among others. This process can be divided in six steps.

1. Electric pulses are applied (on a microsecond to millisecond time) that produce the elevated transmembrane potential resulting in an electric field across the membrane around 10^8 V/m, for a 5 nm thickness membrane.

2. The membrane is charged due to the ion flow.

3. A rapid and localized rearrangement of the molecular structure of the membrane occurs.

4. The pore formation takes place; these are then coated with water molecules (hydrophilic pores).

5. The transport of ions and molecules through these pores is increased.

6. After the electric field removal, the cell membrane recovers and returns to its normal state (Chen et al., 2006).

Figure 1.2 illustrates the sequence of events occurring at the inter-membrane level when an electric field, as electroporation, is applied.

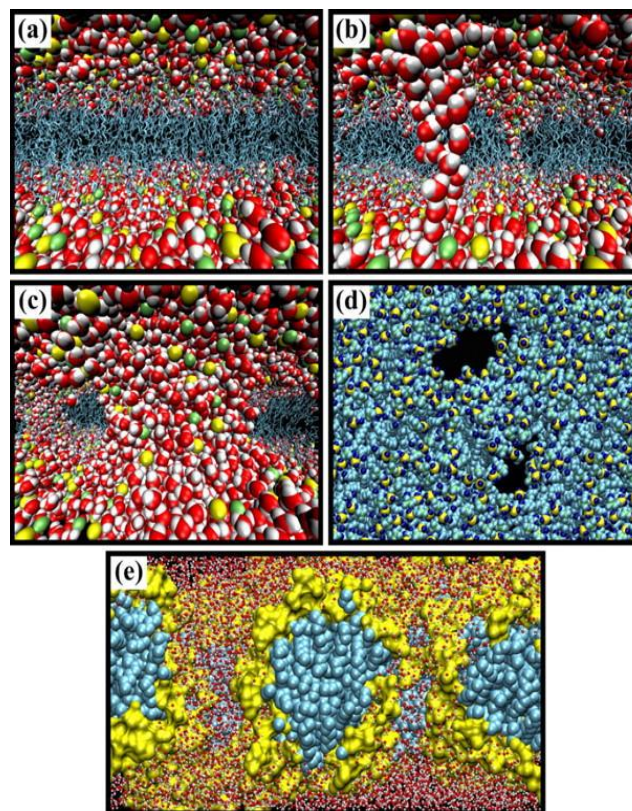


Figure 1.2. Sequence of events occurring in the bilayer membrane when subjected to an electric field as electroporation. (a) Bilayer at equilibrium. (b) Formation of water wires at the initial stage of the electroporation process when the bilayer is subjected to a transverse electric field. (c) Formation at a later stage of large pores stabilized by lipid headgroups. Topology of water pores. (d) Top view (e) side view. In the first three panels, water molecules (O, red; H, white), lipid phosphate (yellow), and nitrogen (green) atoms are represented by van der Waals radii, and the acyl chains (cyan) by stick representation. In the last panel, the hydrophilic lipid head group (yellow) and the hydrophobic acyl chains (cyan) are represented by van der Waals spheres to underline the topology and nature of the water pores. Due to the use of perspective views, atoms in the front appear bigger than those in the back. From (Tarek, 2005).

This method has been used for a large number of bacteria, including previously recalcitrant strains, resistant to transformation. Today, the electroporation of whole cells is a common method practiced in many laboratories

because of its high efficiency and broad range of bacteria that can be transformed. However, this process is not evenly applied and does not have the same efficiency for all types of bacteria, which is due to differences in the bacterial membrane and cell wall structure. Hence, electroporation usually needs to be optimized for each specific cell type (Aune and Aachmann, 2010).

Since many bacterial strains used in laboratories are not naturally transformable, methods have been developed to allow DNA entrance. A transformation protocol usually follows three basic steps: the "preparation step", where the cell is made competent, i.e. prepared to receive DNA; the "shock step", where the cells are exposed to DNA by a non-lethal procedure, the artificial method *per se* that allows DNA to permeate into the cell; and the "recovery step", where the cells are placed under optimal conditions (rich medium, optimal temperature) to allow restoring the damaged membrane and wall.

Bacterial transformation is a very inefficient procedure because it depends on many factors: bacterial strain, medium composition, growth phase, the competence procedure, and DNA size and concentration (Hanahan and Bloom, 1996). Actively growing cells in early exponential phase are more susceptible to transformation. The transformation process also differs between Gram-negative and Gram-positive bacteria (Chen and Dubnau, 2004). The former have an outer membrane and a low percentage of peptidoglycan (the component that confers rigidity and works as the scaffold of the cell wall), while the latter do not have an outer membrane, although they present a thicker peptidoglycan layer. Because of this higher percentage of peptidoglycan, the Gram-positive bacteria are generally more difficult to transform. In general, smaller molecules transform more efficiently than larger ones and with the increase of DNA concentration the dying rate of bacteria increases. The number of transformants rises until the saturation concentration is reached, decreasing slowly after that point (Szostková and Horáková, 1998). A poorly performed procedure can result in cells with reduced competence to take up DNA. The competency of a stock of competent cells is determined by calculating how many colonies (transformants) are produced per microgram of DNA added. An excellent preparation of competent cells will give approximately 10^8 colonies/ µg and a poor preparation will yield about 10^4 colonies/ µg or less.

1.3.3. *Clostridium* transformation

The *Clostridium* genus' importance, as previously discussed, led to the development of many studies at the genetic and molecular level. To achieve this knowledge the introduction of recombinant DNA into viable clostridia cells is required. Regardless, as the *Clostridium* genus consists of strictly anaerobic Gram-positive bacteria (which makes the transformation process more laborious and time consuming), only a few successful protocols to transform *Clostridia* have been developed (Young et al., 1999). Processes such as protoplast transformation are being used in *C. acetobutylicum* (Lin and Blaschek, 1984; Reid et al., 1983). The conjugation process was used to insert DNA into *C. difficile* (Herbert et al., 2003; Purdy et al., 2002). Conjugative transposons, Tn*916* or Tn*1545*, have been used to insert mutations in a wide range of bacteria, including some clostridia (Young et al., 1999). The electro-transformation is the most widely applied method to introduce foreign DNA into *Clostridium* cells: *C.*

acetobutylicum (González-Pajuelo et al., 2005; Kuit et al., 2012; Mermelstein et al., 1992; Nakotte et al., 1998; Oultram et al., 1988; Tyurin et al., 2000), *C. cellulolyticum* (Jennert et al., 2000; Tardif et al., 2001), *C. beijerinckii* (Birrer et al., 1994), *C. thermocellum* (Olson and Lynd, 2012; Tyurin et al., 2004), *C. perfringens* (Chen et al., 1996), and these are just some examples of the use of this technique in *Clostridium*. Electroporation has been applied as an alternative to the previously mentioned methods, since these are technically cumbersome, time consuming and can present inconsistent results when compared to the quick and simple electroporation technique known to provide reproducible results.

1.3.4. R/M Systems

The way *Clostridium* is transformed, i.e. the shock step, although extremely important, is not the critical step on its transformation process, as these cells present a highly aggressive defense system against foreign DNA that protects the host cell from DNA not recognized as their own.

All the bacteria have naturally developed resistance mechanisms against invading nucleic acids (virus DNA or RNA) that hinder the cells infection and dead, preserving genetic isolation. Bacterial cells have developed resources to prevent bacteriophage (phage that only infects bacteria) adsorption and cell entry, to cut foreign DNA, and sometimes also commit “suicide” if it results in the elimination of the invader and prevention of the infection of the remaining cells (Labrie et al., 2010). The DNA cutting resources are very common.

The work of Bertani and Weigle (Bertani and Weigle, 1953), Luria and Human (Luria and Human, 1952) and Arber (Arber, 1965) brought evidence of the existence of bacterial enzymes that cleave foreign DNA. These enzymes, and others subsequently found, were grouped as Restriction systems, distinctive to each bacterium. Their main functions are to defend their host against extraneous DNA, the maintenance of species identity and allow speciation (Jeltsch, 2003), and modulate the frequency of genetic variation (Arber, 2000). These systems have been classified as an “immune system” seconded to eliminate invading DNA without adequate modification by methylation (Nikolajewa et al., 2005). A classical restriction system, constituted by restriction endonucleases (REases) is normally associated with methyltransferases (MTases), producing the Restriction and Modification (R/M) systems. REase recognize and cleave a specific DNA sequence, and MTase hinder the DNA cleavage, by modifying adenosyl or cytosyl residues through the transference of methyl groups from the donor S-adenosyl-L-methionine (SAM) within an also specific DNA sequence (Bickle and Krüger, 1993). These methylated bases can be 5-methylcytosine (^{m5}C), N4-methylcytosine (^{m4}C) or N6-methyladenine (^{m6}A). The methylation pattern of an organism works as a bar code that represents the unique identity of each strain (Jeltsch, 2003).

1.3.4.1. Types of restriction enzymes

The REases are commonly distinguished into four types (I to IV) based on enzyme composition, requirements and mode of action (Pingoud et al., 2005; Roberts et al., 2003), as described below. The R/M system enzymes belong just to I, II and III type. Between these four types, type II is the most known, developed and with the higher

application, as will be further described. They are nominated according to the name of the bacterium of origin, with the first letter corresponding to the genus, the first and second epitope letters, and the first letter/number to the strain. As an example, the enzyme *Eco*RI was isolated from *E. coli* strain RY13 and was the first (I) enzyme of this strain to be analyzed. Restriction enzymes can be distinguished by adding the letter R to the name (R.*Eco*RI) and the methyltransferases adding the letter M (M.*Eco*RI). The genes are nominated, italicized, as *ecoRIR* and *ecoRIM*, respectively, to restriction enzymes and methyltransferases.

Type I R/M system is constituted by multi-subunit enzymes functioning as a single protein complex. This system generally comprises two HsdM subunits (*hsd* for host specificity for DNA), which modify DNA by methylation (^{m6}A), two HsdR subunits, responsible for the restriction of the target sequence and one HsdS subunit, involved in the recognition of the target sequence. When in contact with non-methylated DNA the entire complex acts as REase, requiring ATP and Mg^{2+} for activity. Usually, two asymmetrical bi-partite target sites are recognized, but the cleavage site is away from the recognition sequences, approximately half-way between two sites. The complex acts as MTase when in contact with hemimethylated DNA (just one strand is methylated in a double stranded sequence as it happens after DNA replication of a completely methylated sequence). This complex requires the presence of the methyl group donor AdoMet (S-adenosyl-L-methionine) and is 100% active just with the HsdM and HsdS subunits. Some typical type I R/M system examples are the enzymes *Eco*AI, *Eco*R124I, and *Sty*SBLI, representing the subtypes IA, IB, IC and ID, respectively (Bickle and Krüger, 1993; Pingoud et al., 2005; Roberts et al., 2003).

Furthermore, it is believed that the type I R/M system is related with the restriction alleviation (RA) process responsible for the maintenance (non-cleavage by REases) of the own cell DNA when this is not yet methylated. Moreover, it was discovered that in case of damage, the restriction is alleviated, being a good opportunity for the entry of foreign DNA and for cell mutation (Keatch et al., 2004; Makovets et al., 2003).

Type III R/M system encompasses enzymes constituted by two subunits with modification and recognition activity (Mod), and two subunits responsible for restriction (Res). Cleavage (Res subunit) requires the presence of all the subunits, ATP hydrolysis, Mg^{2+} as cofactor and AdoMet as stimulator. Two asymmetrical sequences are recognized in inverse orientation (head to head arranged) and the cleavage occurs close to one recognition site. The modification by methylation (Mod subunit) can occur independently of the Res subunit, creating an ^{m6}A base. Some common examples are the enzymes *Eco*P1I and *Eco*P15I (Pingoud et al., 2005; Roberts et al., 2003).

Enzymes belonging to type IV R/M system only recognize and cleave non-specific, variable, and modified DNA (methylated, hydroxymethylated and glucosyl-hydroxymethylated bases) and have, unlike the other restriction systems, no methyltransferases. For successful cleavage REases require GTP and Mg^{2+}. The recognition sequence is not well defined, but for example, the *Mcr*BC enzyme, cleaves ^{m5}C-modified DNA. The *Mcr*B subunit is responsible for recognition and the *Mcr*C subunit is responsible for cleavage (Pingoud et al., 2005; Roberts et al., 2003).

1.3.4.2. Type II

The type II R/M system is a large and diverse R/M group, and due to its characteristics the enzymes are commonly used in genetic engineering processes. They are differentiated from the other R/M system enzymes by their simplified subunit organization, the generation of a defined restriction pattern, the recognition of specific targets and the cleavage within or close to the recognition sequence. REases can be monomers, dimers or tetramers, and normally require the presence of Mg^{2+} as cofactor. Their action is, saving some exceptions, methyltransferase-dependent. MTases are monomers that require the presence of the methyl group donor AdoMet to directly transfer to the DNA double strand (^{m4}C, ^{m5}C or ^{m6}A). The diversity of the type II R/M system has led to the development of subdivisions based on the target sequence or on the enzyme structure. These may lead to an overlap of classifications, as one enzyme can be part of more than one subdivision (Pingoud et al., 2005; Roberts et al., 2003).

Type II-classified enzymes recognize asymmetrical sequences (type IIA), that cleave both sides of the recognition sequence (type IIB), enzymes that require just one polypeptide to have both cleavage and methylation activity (type IIC), that require two copies of the recognition sequence (types E and F), enzymes that require the presence of AdoMet as a stimulator to REase or MTase activity (type IIG), type II enzymes structurally similar to type I (type IIH), enzymes that only cleave methylated DNA (type IIM), homodimeric enzymes that cleave just one DNA strand shifted to the recognition sequence (type IIS), heterodimeric enzymes (type IIT) and the orthodox restriction enzymes (type IIP).

The most well-known type II subtype is the IIP. This contains the orthodox type II enzymes, homodimers recognizers of fixed 4 to 8 bp palindromic (symmetric) sequences, which cleave the DNA within the sequence or immediately adjacent to it in both strands, and producers of 3´-hydroxyl and 5´-phosphate ends. These enzymes generate either ‘blunt’ or ‘sticky’ ends. All these features together have led to an increase in type IIP enzymes-related research and development, and have increased their biotechnology and molecular biology applications. They are especially useful in gene analysis and cloning work. The existence of symmetric sequences is considered advantageous since it is energetically more economical for the cell. The speciation of two subunits is energetically more favorable, recognizing, each one, the half symmetrical sequence, than the speciation of a larger protein to recognize an entire sequence (Nikolajewa et al., 2005). A good example is the enzyme *Eco*RI that recognizes the sequence

Table 1.1 presents a compilation of the general properties of the four restriction enzyme systems.

Table 1.1.General properties of the four restriction enzyme types.

R/M System	Restriction		Methylation	
	Machinery	Cleavage site	Machinery	Cleavage site
Type I	R_2M_2S	Variable	M_2S	^{m6}A
Type II	R	Fixed	M	^{m6}A, ^{m5}C, ^{m4}C
Type III	R_2M_2	Variable	M_2	^{m6}A
Type IV	R	Variable		

Note: R- restriction subunit, M- methylation subunit, S- sequence recognition subunit. From (Suzuki, 2012).

In August 2013, 13,508 type II restriction enzymes, 16,716 type I restriction enzymes, 4,256 type III restriction enzymes, 4,747 type IV restriction enzymes, and 33,030 methyltransferases were listed on REBASE – The Restriction Enzyme Database (http://rebase.neb.com/rebase/rebms.html).

1.3.5. *Clostridium* R/M systems

The highly aggressive defense system found in *Clostridium* consists of restriction endonucleases that digest foreign DNA, not methylated like their own. When the first attempts were made to insert exogenous DNA into microorganisms of this genus, no positive clones with extra DNA insertions were obtained (Mermelstein and Papoutsakis, 1993). Experiments using crude extract (cell content after membrane disruption) showed a specific digestion pattern of the DNA placed in contact with it. The existence of a pattern means that restriction endonucleases with a specific recognition sequence (type IIP) are present. Many *Clostridium* restriction enzymes have been identified and these are summarized in Table 1.2.

Table 1.2. *Clostridium* REases identified.

Organism	REase	Restriction site	Source
C. acetobutylicum	*Cac*I	5'-|GATC- 3'	(Azeddoug et al., 1989)
	*Cac*II	5'-GCN|NGC- 3'	(Azeddoug and Reysset, 1991)
	*Cac*824I	5'-GCNGC - 3'	(Mermelstein et al., 1992)
C. bifermentans	*Cbi*I	5'-GCN|NGC- 3'	(Murakami et al., 1991)
C. botulinum	*Cbo*I	5'-C|CGG- 3'	(Davis et al., 2000)
C. cellulolyticum	*Cce*I	5'-C|CGG - 3'	(Jennert et al., 2000)
C. coccoides	*Cco*I	5'-GCC|GGC- 3'	(Kanematsu and Ozawa, 1991)
C. cylindrosporum	*Ccy*I	5'-|GATC- 3'	(Wen et al., 1989)
C. difficile	*Cdi*I	CATCG (-1/-1)	(Roberts et al., 2007)
	*Cdi*CD6I	5'-GGNCC- 3'	(Purdy et al., 2002)
	*Cdi*CD6II	5'-GATC- 3'	(Purdy et al., 2002)
C. pasteurianum	*Cpa*I	5'-GATC- 3'	(Roberts et al., 2007)
	*Cpa*1150I	5'-CGCG- 3'	(Roberts et al., 2007)
	*Cpa*AI	5'-CGCG- 3'	(Richards et al., 1988)
C. sporogenes	*Csp*45I	5'-TT|CGAA- 3'	(Schoenfeld et al., 1989)
C. thermocellum	*Cth*I	5'-TGATCA- 3'	(Choi et al., 1987)
	*Cth*II	5'-CC|WGG- 3'	(Choi and Yoo, 1990)

N- Any base, W- A or T. (|) Cleavage site identified.

All this REases activity cease when methylated DNA was used. As Mermelstein stated "(...) if the nature of the restriction system is known, transforming DNA may be protected by methylation (...)" (Mermelstein and Papoutsakis, 1993). Cleavage can be completely hindered by the cognate methyltransferase if this mimics the cell intern methylation pattern. It can be affected if the methylation is not specific, i.e. if the REase is sensitive to other kind of methylation the cleavage is "impaired" (http://rebase.neb.com/rebase/rebms.html). However, the cleavage cannot be hindered at all if the REase is methylation-insensitive. Figure 1.3 represents a scheme of what occurs with methylated and non-methylated foreign DNA in a cell with an active restriction system.

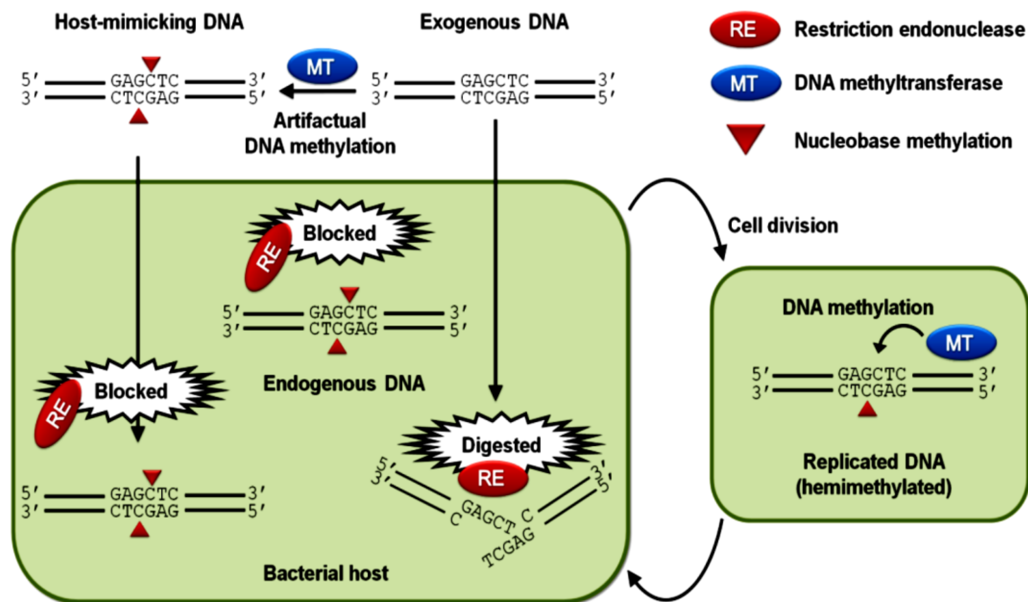


Figure 1.3. Specific methylation ceases restriction endonucleases action. The transformation process is hindered in a cell with a very active restriction system since the cell REases (restriction endonucleases) cleave the DNA, not methylated as their own. If the foreign DNA is methylated as is their own cellular DNA (mimic), by one specific methyltransferase, the REases will not cleave this DNA and it will be inserted in the genetic information of this cell. Future generations will also have this extra information. From (Suzuki, 2012).

Positive clostridia transformations were achieved by *in vitro* or *in vivo* DNA methylation, i.e. respectively, using a purified MTase or one MTase gene inserted into a plasmid that in turn is inserted into a bacterium (e.g. the best known bacterial host - *E. coli*).

The most well-known *in vivo* strategy was developed by Mermelstein and collaborators in 1993, pAN1 and pAN2 plasmids were used to methylate the sequences 5'-G^{m5}CNGC - 3' and 5'-GG^{m5}CC - 3', recognized and digested by REases from *C. acetobutylicum*. These plasmids had the Φ3TI MTase gene (*φ3tI*) from *Bacillus subtillis* phage Φ3T inserted (Mermelstein and Papoutsakis, 1993). All the successful *C. acetobutylicum* transformations were performed using the plasmid pAN1 (*φ3tI*, cloramphenicol resistance) (Dong et al., 2010; González-Pajuelo et al., 2005; Mermelstein and Papoutsakis, 1993; Nakotte et al., 1998; Shao et al., 2007; Tyurin et al., 2000) or pAN2 (*φ3tI*, tetracycline resistance) (Dong et al., 2010; Kuit et al., 2012). In the *in vivo* system, the pAN1 (or similar vector) bacterial cell is transformed with the DNA intended to methylate and the modification occurs inside the cell.

After modification the DNA is extracted and transformed into the bacterium with the correspondent active restriction system.

The *in vitro* approach is simple and effective but limited because the commercial MTases usually lack each cell's specific methylation pattern. This process was used to transform *Clostridium* strains as *C. cellulolyticum*, where only the mixture of the DNA intended to methylate with the enzyme in the enzyme medium is needed, along with the optimal temperature requested (Jennert et al., 2000; Tardif et al., 2001). This need led to the identification, production and purification of these organism's MTases to further use them as a modification strategy required to transform *Clostridium* cells (Davis et al., 2000).

At the moment (August, 2013) the REBASE lists 151 *Clostridium* strains. From these 34 have, identified or predicted, at least one type II R/M system with REase and the cognate(s) MTase(s) (presented in Table 1.3). The study of these R/M systems, specially the MTases will allow the transformation of recalcitrant microorganisms by specifically methylated DNA. Thus, their applications will be exploited and may be useful to biotechnological applications.

Table 1.3. Type II R/M Systems present in *Clostridium* listed in REBASE

Clostridium strain	R/M Systems Name		Recognition site (5'– 3')
	REase	MTase	
C. acetobutylicum ATCC 824	*Cac*824I	M.*Cac*824I	5'-GCNGC- 3'
C. acetobutylicum ABKn8	*Cac*8I	M.*Cac*8I	5'-GCN(|)NGC- 3'
C. acetobutylicum DSM 1731	*Cac*1731ORF1526P	M.*Cac*1731ORF1526P	5'-GCNGC- 3'
C. acetobutylicum EA 2018	*Cac*2018ORF1517P	M.*Cac*2018ORF1517P	5'-GCNGC-3'
C. acidurici 9a	*Cac*9aORF8000P	M.*Cac*9aORF8000P	5'-GGNCC- 3'
C. botulinum	*Cbo*I	M.*Cbo*I	5'-C(|^{m5}) CGG- 3'
C. botulinum A ATCC 19397	*Cbo*AORF2811P	M.*Cbo*AORF2811P	5'-GCWGC- 3'
C. botulinum A Hall	*Cbo*AHORF2744P	M.*Cbo*AHORF2744P	5'-GCWGC- 3'
C. botulinum B Eklund 17B (NRP)	*Cbo*17ORFCP	M.*Cbo*17ORFCP	5'-GATC- 3'
C. botulinum B1 Okra	*Cbo*B1ORF2548P	M.*Cbo*B1ORF2548P	5'-CCGG- 3'
C. botulinum Bf	*Cbo*BfORF47P	M.*Cbo*BfORF47P	5'-GGWCC- 3'
C. botulinum CFSAN001628	*Cbo*16ORFCP	M.*Cbo*16ORFCP	5'-CCGG- 3'
C. botulinum E1 'BoNT E Beluga'	*Cbo*E1ORF1092P	M.*Cbo*E1ORF1092P	5'-GCNGC- 3'
C. botulinum E3 Alaska E43	*Cbo*E3ORF2487P	M.*Cbo*E3ORF2487P	5'-GRCGYC- 3'
C. botulinum F 230613	*Cbo*F2ORF2114P	M.*Cbo*F2ORF2114P	5'-GCSGC- 3'
C. butyricum 5521	*Cbu*5521ORF1956P	M.*Cbu*5521ORF1956P	5'-GATC- 3'
C. cellulolyticum H10	*Cce*I	M.*Cce*I	5'- CCGG- 3'
	*Cce*ORF2549BP	M.*Cce*ORF2549P	5'-GCGC- 3'
	*Cce*ORF2549AP		
C. cf. *saccharolyticum* K10	*Ccf*10ORF12590P	M.*Ccf*10ORF12590P	5'-GATC- 3'
C. difficile CD3	*Cdi*I	M.*Cdi*I	5'-CATC|G (-1/-1)- 3'
			5'-TGGCCA-3'
C. difficile CD6	*Cdi*CD6I	M.*Cdi*CD6I	5'-GGNCC- 3'
	*Cdi*CD6II	M.*Cdi*CD6II	5'-G(^{m6})ATC- 3'
C. hathewayi WAL-18680	*Cha*18680ORF3698P	M.*Cha*18680ORF3698P	5'-GCWGC- 3'
C. kluyveri DSM 555	*Ckl*ORF2671P	M.*Ckl*ORF2671P	5'-CCGG- 3'
C. kluyveri NBRC 12016	*Ckl*AORF2367P	M.*Ckl*AORF2367P	5'-CCGG- 3'
C.leptum DSM 753	*Cle*DORF2452P	M.*Cle*DORF2452P	5'- GGATG- 3'
	*Cle*DORF3217P	M.*Cle*DORF3217P	5'-GGNCC- 3'
C. nexile DSM 1787	*Cne*1787ORF858P	M.*Cne*1787ORF858P	5'-GGNCC- 3'
C. pasteurianum DSM 525	*Cpa*525ORF11401P	M1.*Cpa*525ORF11401P	5'-GATC- 3'
		M2.*Cpa*525ORF11401P	
	*Cpa*525ORF2340P	M.*Cpa*525ORF2340P	5'-CGCG- 3'
C. perfringens ATCC 13124	*Cpe*AIIP	M.*Cpe*AII	5'-GGW(^{m5})CC- 3'
	*Cpe*AIIIP	M.*Cpe*AIII	5'-GAT(^{m5})C- 3'
C. saccharolyticum WM1	*Csa*WM1ORF1186P	M.*Csa*WM1ORF1186P	5'-CAATTG- 3'
C. saccharoperbutylacetonicum N1-4 (HMT)	*Csa*1ORF40510P	M1.*Csa*1ORF40510P	5'-GATC- 3'
		M2.*Csa*1ORF40510P	
C. tetani E88	*Cte*EORF387P	M.*Cte*EORF387P	5'-GATC- 3'
C. thermocellum ATCC 27405	*Cth*ORF1513P	M.*Cth*ORF1513P	5'-GATC- 3'
	*Cth*ORF1749P	M.*Cth*V	5'-G(^{m5})CWGC- 3'
	*Cth*ORF2320P	M.*Cth*VI	5'-GG(^{m5})CC- 3'
	*Cth*ORF2470P	M.*Cth*ORF2470P	5'-GATC- 3'
C. thermocellum LQ8	*Cth*VORFCP	M.*Cth*VORFCP	5'-GATC- 3'
	*Cth*VORF2275P	M.*Cth*VORF2275P	5'-GATC- 3'
C. thermocellum JW20	*Cth*20ORF2874P	M.*Cth*20ORF2874P	5'-GATC- 3'

N- Any base, W- A or T, R- A or G, Y- C or T, S- G or C. (|) Cleavage site identified. (^{m5} or ^{m6}) methylation site identified. (|^{m5}) cleavage and methylation site identified.

1.4. ClosTron system in *Clostridium* transformation

To date only a few *C. pasteurianum* mutants were obtained, which have been obtained only by chemical mutagenesis (Daldal, 1985; Jensen et al., 2012b; Malaviya et al., 2012), i.e. no new DNA entered into cells, a random DNA change happened in the bacterium own DNA.

In order to be able to study the metabolic engineering (e.g. metabolic pathway manipulation), specific DNA alterations are required and for this efficient genetic engineering tools are needed.

The common practice to obtain mutants is by classical recombination-based procedures. Clostridia mutants have been mostly obtained by homologous recombination. Some clostridia mutated genes were reported in few species, notably in *C. acetobutylicum*, *C. beijerinckii, C. perfringens* and even in *C. difficile* (Heap et al., 2007), but they are inherently unstable and especially mutants from homologous recombination are difficult to isolate (Heap et al., 2010). Although gene transfer to the *Clostridium* genus is already possible, Young considered that mutated genes are most readily isolated and characterized if they are generated by insertional mutagenesis (Young et al., 1999). The normally applied genetic methodologies are the gene modification systems through "Knock-out" and "knock-in" (Dong et al., 2010; Heap et al., 2009; Kuehne et al., 2011).

Based on the need for a reliable and efficient mutagenesis system, and given the existence of inadequate procedures for gene insertion/inactivation, Nigel Minton and collaborators developed the ClosTron (Heap et al., 2007). This is a mutagenesis system with molecular tools necessary to study and exploit the wide physiological range of the genus *Clostridium*. The construction system was based on the studies of the Lambowitz laboratory and on the use of group II intron from the *ltrB* gene of the bacterium *Lactococcus lactis* (Ll.ltrB) (Karberg et al., 2001). Bacterial group II introns are a newly characterized type of mobile elements, they are catalytic RNAs that excise themselves from RNA transcripts via a lariat intermediate and insert themselves into a new distal target site. They possess two important characteristics that enable their use for the directed construction of stable mutants. First, via an RNA-mediated "retrohoming" mechanism, these elements propagate to their specific site and establish connection mainly by base-pairing between the excised RNA intron (that can be rationally re-programmed) and the target DNA site. Second, the target specificity between target site and the excised RNA is dependent of the presence of a multifunctional intron encoded protein (IEP) which can be provided transiently during mutagenesis and subsequently removed to ensure the stability of the strain. The *ltrA* gene, that encodes the IEP, is normally within the Ll.LtrB intron to allow its permanent mobility. Knowing that, Lambowitz and coworkers, moved this gene from inside Ll.LtrB intron to another location in the plasmid and after the mutagenesis procedure, the plasmid will be lost and the excised RNA will not be moved again (Lambowitz & Zimmerly 2004). The retargeted introns resulting from this system were nominated "TargeTrons".

Some mutagen processes were developed with the TargeTron, but it was lacking a selection method to obtain mutants with successful insertion, as only 10% of the population was correctly mutated. To overcome this limitation, Zhong and coworkers (Zhong et al., 2003) introduced into the group II intron an antibiotic resistance gene interrupted by a self-splicing group I intron, nominated retrotransposition-activated marker (RAM). All three

elements were organized in such way that just only after successful target insertion, the group I intron can splice out and the antibiotic resistance gene is restored. Therefore, the correct mutants will be resistant to the antibiotic resulting from the RAM gene (Heap et al., 2007).

As previously mentioned, the ClosTron system applied the TargeTron system as a specific tool to genetically modify genes in clostridia, and increased some potentialities, further explained. Therefore, the ClosTron plasmids pMTL007 were developed. The ClosTron RAM is based on the ermB gene (confers resistance to erythromycin, Em, following plating on thiamphenicol). The inactivation of the ermB gene is accomplished by the insertion of a small region of DNA encompassing a group I intron (from the td phage) inside the gene. This group I introns splice-out from mRNA, but this splicing is also orientation-dependent, the transcription to mRNA is not sufficient to allow splicing. Just when the opposite DNA strand encompassing the group II intron region is transcribed, the ltrA protein binds to the transcript, leading to the formation of a ribonuclear protein (RNP) complex. Now the group I intron has the correct orientation and will self-splice. Then the RNP recognizes the target DNA and binds to specific regions within the chromosome. As the ltrA protein has endonuclease and reverse transcriptase activity, nicks the DNA target, inserts the RNA and synthesizes the complementary DNA strand. Host nucleases degrade the inserted RNA, one DNA polymerase (DNP) synthesizes the opposite DNA strand, and host ligases seal the gaps, finishing the process. Now the ermB gene is present in the target gene and will have functional activity conferring Em resistance (Kuehne and Minton, 2012). An illustration of the ClosTron process is provided in Figure 1.4.

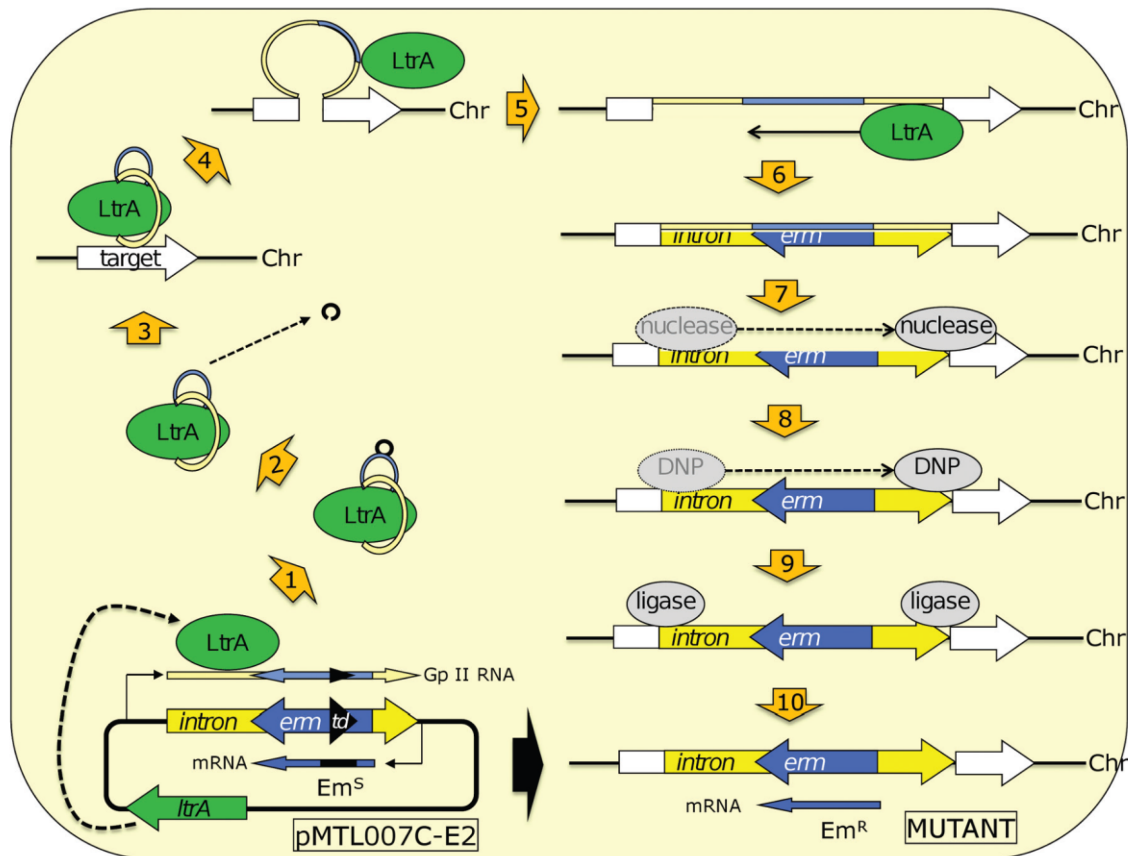


Figure 1.4. Schematic representation of mutant generation using the ClosTron system. pMTL007C-E2, the ClosTron plasmid, present the Group II intron (yellow) encompassed by the *emrB* gene (blue) itself encompassed by *td* group I intron (black). Also present, out of group II intron, is the *ltrA* gene (green) to IEP coding. The transcription of *ermB* gene occurs, the group I intron even in RNA transcript do not excise because is with incorrect orientation, therefore cell continues erythromycin sensitive (Em^s). Just when the opposite DNA strand encompassing the group II intron region is transcribed the *td* sequence acquires a correct orientation and can now spliced out. Meanwhile LtrA is translated and binds to the transcript forming the ribonuclear protein (RNP) complex (1). After the complex formation occurs the group I intron splice out (2). Take place the recognition of and ligation to the DNA target sequence by the correspondent RNA sequence encompassed into the group II intron (3). LtrA nicks the DNA target (4) and insert the RNA into the chromosome (5). It is synthesized the complementary DNA strand by the reverse transcriptase activity of *L*trA (6). The inserted RNA is degraded by host nucleases (7) and the opposite DNA strand is synthesized by the DNA polymerase (DNP) (8). The gaps are sealed by host ligases (9) and the process is finalized when is obtained a functional *ermB* gene leading to mutant cells erythromycin resistant's (Em^R) (10). From (Kuehne and Minton, 2012).

The ClosTron technology has been largely used and has presented positive results. However, following the first report of this mutagenic system their developers have been refining it with extra properties. In 2009, John Heap (Heap et al., 2009) presented "a modular system for *Clostridium* shuttle plasmids". They specified, designed and constructed a standardized modular system for *Clostridium* – *E. coli* shuttle plasmids using existing Gram-negative and Gram-positive replicons and selectable markers. The Modular ClosTron plasmids were developed, consisting of a broad range of components that are allowed to associate so that the best combinations are formed, taking into account the target microorganism and the final objective. In 2010, the same group (Heap et al., 2010) demonstrated the possibility of constructing mutants with multiple insertions. They studied the possibility to recycle the marker by using FLP recombinase. *ermB* RAM is flanked by repeated FLP recognition target (FRT) sites that

when in contact with FLP recombinase will be excised making the mutated cells lose the marker gene. Marker loss is confirmed, and the same cells are again transformed with a second plasmid with the same marker gene to mutate a second DNA target. This process allows at least three mutations. To complement all this work, the same research group, created a free access website (www.clostron.com) with a computer algorithm that predicts the necessary alterations for intron retargeting and allows the creation of the most suitable vector for the process using various components combinations (selectable marker gene, *E. coli* and *Clostridium* replicons). Members of the group (Kuehne and Minton, 2012) summarized the entire process to generate a ClosTron mutant in eight easy steps:

1 – Choose the most suitable ClosTron vector with the best replicons and resistance marker according the microorganism specie.

2 – Use the online algorithm to predict the required re-targeted region of the group II intron for the gene to be inactivated.

3 – Transfer the retargeted plasmid into clostridia host by conjugation or transformation.

4 – Select the transconjugants with the appropriated antibiotic and isolate three independent integrants.

5 – Make a preliminary screen of positive mutants by PCR using appropriate primers.

6 – Confirm PCR results by sequencing or southern blot.

7 – Analyze some mutants phenotypically.

8 – Complementation of mutant phenotype using pMTL8000 vector.

In 2007, Heap and collaborators presented the ClosTron as a "universal gene knock-out system for the genus *Clostridium"*. The procedure was characterized as "highly efficient and reproducible" that "should revolutionize functional genomic studies in clostridia". The constant improvements led to a great expansion of the system as it minimized the labor-intensity and maximized the accessibility of the mutagenesis method. Additionally, in 2012, Kuehne and Minton attested that ClosTron technology was broadly widespread by the research community being classified as a mature, robust and rapid mutagenesis system that improves basic understanding and addresses applied challenges, revolutionizing clostridia molecular biology (Heap et al., 2007, 2009, 2010; Kuehne and Minton, 2012).

This mutagenesis system has been successfully combined with *in vivo* (pAN2 plasmid) (Cooksley et al., 2012; Dong et al., 2010; Kuit et al., 2012; Lehmann et al., 2012a, 2012b) or *in vitro* (M.*Msp*I) (Cui et al., 2012) systems to produce mutant strains.

In the current thesis work plan, the use of the ClosTron system will be useful, since the obtained transformants will have a deleted gene with a visible phenotype leading to an easy confirmation of the results.

1.5. Aim

The goal of this thesis is to create a Genetic System in *C. pasteurianum* that allows genome modification and foreign protein expression, i.e. improving *C. pasteurianum* transformation. The development of such system would facilitate their metabolic manipulation, increasing the competitiveness of this organism in biotechnological processes.

Chapter 2

Materials and Methods

2. MATERIALS AND METHODS

2.1. Bacterial strains and plasmids

All bacterial strains and plasmids used or derived from this study are listed in Table 2.1.

Table 2.1. Bacterial Strains and plasmids.

Strains/Plasmid	Genotype/Relevant Characteristics[a]	Source/Reference
Clostridium pasteurianum		
DSM 525 (ATCC 6013)	Wild type	DSMZ
Escherichia coli		
JM109	endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB^{+} 8(lac proAB) e14- [F' traD36 proAB^{+} lacI^{q} lacZΔM15] hsdR17(r_K^- m_K^+)	Nzytech
BL21 (DE3)	F^{-} ompT gal dcm lon hsdSB(r_B^- m_B^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	Nzytech
BL21-CodonPlus® (DE3)RIL	F^{-} ompT hsdS(r_B^- m_B^-) dcm^{+}Tet^{R} gal λ(DE3) endA Hte [argU ileY leuW Cm^{R}]	Stratagene
Rosetta (DE3)pLysS	F^{-} ompT hsdS$_B$(R_B^- m_B^-) gal dcm λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5]) pLysSRARE (Cm^{R})	Novagen
Origami™ (DE3)	Δ(ara-leu)7697 ΔlacX74 ΔphoA PvuII phoR araD139 ahpC galE galK rpsLF′[lac+ lacIq pro] (DE3)gor522::Tn10 trxB (Kan^{R}, Str^{R}, Tet^{R})	Merck Millipore
Plasmids		
pMTL007C-E2::Cpa-*spo0A*-666a	*E. coli* - *C. pasteurianum* shuttle vector (Cm^{R}; ColE1 ori; pCB102 ori, FRT-flanked *ermB* RAM). pMTL007C-E2 plasmid re-targeted to the intron target site at position 666 of the *spo0A* gene of *C. pasteurianum* DSM 525 in the antisense orientation	(Gallardo, 2013)
pET24d(+)	Kan^{R}	Novagen
pETduet-1	Amp^{R}	Novagen
pCOLAduet-1	Kan^{R}	Novagen
pCM001	Amp^{R}, pETduet with the *mdpn* gene	This study
pCM002	Amp^{R}, pETduet with the *dpn* gene	This study
pCM003	Amp^{R}, pETduet with the *dam* gene	This study

[a]Abbreviations: Cm^{R} - chloramphenicol resistance, Kan^{R} – kanamycin resistance, Amp^{R} – ampicillin resistance, Str^{R} – Streptomycin resistance, Tet^{R} – Tetracycline resistance, *ermB* – confers resistance to erythromycin, ColE1 ori – *E. coli* origin of replication, pCB102 ori – clostridial origin of replication.

All the plasmids were maintained and replicated in *E. coli* strains using the respective selective marker (antibiotic). Concentrations of antibiotics used in selective medium were as follows: chloramphenicol (Nzytech) 25 µg/mL, kanamycin (Applichem) 50 µg/mL and ampicillin (Applichem) 100 µg/mL for *E. coli* strains; and thiamphenicol (Sigma) 15 µg/mL for *C. pasteurianum* DSM 525.

2.2. Genes

The genes of interest were selected, as a Type II R/M system, from *C. pasteurianum* DSM 525 genome obtained by pyro-sequencing technology (Life sequencing) (Gallardo, 2013).

They were named based on the enzymes' predicted functionality; Type II site-specific deoxyribonuclease (GenBank: ELP59087.1) has similar predicted function to *Dpn*II enzymes (*dpnII* gene – 861 bp), DNA methylase N-4/N-6 domain-containing protein (GenBank: ELP59090.1) it is predicted to be the complementar methyltransferase of *Dpn*II endonuclease (*mdpn* gene – 813 bp), Type II R/M system modification methyltransferase (GenBank: ELP59088.1) it is similar to DNA adenine methyltransferases (*dam* gene – 909 bp).

2.3. Primers

Primers for the amplification of the specific genes were designed using the Vector NTI® Software (Life Technologies – Invitrogen). The primers forward (equal to template) and reverse (complementary to template, 5´to 3´) were designed calculating the nearest melting temperature (Tm ºC) between them, the GC percentage and verifying the possibility of dimers and hairpins formation. At the 5´- end of the sequence that pairs with the gene sequence, enzyme restriction sites were added. These restriction sites have to be part of the MCS (multiple cloning site) of the respective vector and the chosen enzymes cannot cut the gene sequence. For the genes *mdpn*, *dam*, and *dpnII*, the enzymes *Nco*I and *Bam*HI were chosen. A 6xHis-tag was also added to allow future enzyme recovery (upstream of the protein), along with an enterokinase cleavage site to remove the His-tag. In Table 2.2 the primer list is shown, with the respective sequence additions highlighted and the associated melting temperature.

Table 2.2. List of primers used. These primers were synthesized by Invitrogen.

Primer	Sequence (5´ → 3´)	Tm ºC
M.dpnIIBCp.F	GCGGCCCATGGGTCACCATCACCATCACCATGATGACGATGACAAA**GTGGAATATATAAAGAAGGATAA**	52
M.dpnIIBCp.R	AAGGTGGATCC**TTACTCTCCACTACTTTTATC**	52
DamCp.F	GCGGCCCATGGGTCACCATCACCATCACCATGATGACGATGACAAA**GTGGAGAGTAAGATAAAACC**	52
DamCp.R	TTTAAGGATCC**TTATTTATCATAATTAGTTATTAGC**	52
DpnII.F	GCGGCCCATGGGTCACCATCACCATCACCATGATGACGATGACAAA**ATGAAGAATAGAAATTTTGAT**	52
DpnII.R	TAATTGGATCC**TTATACGAACACTTTATCC**	52
T7 Promoter	TAATACGACTCACTATAGGG	45
T7 Terminator	GCTAGTTATTGCTCAGCGG	45

Yellow: restriction sites, *Nco*I (C|CATGG) and *Bam*HI (G|GATCC)
Green: 6xHis-tag sequence;
Cccc Grey: Enterokinase cleavage site (Asp-Asp-Asp-Asp-Lys|);
Bold: gene sequence to amplify.

2.4. Culture media

C. pasteurianum DSM 525 was grown in 500 mL bottles containing 100 mL of anaerobic mRCM (modified Reinforced Clostridial Medium - 6I, Table 6.1) at 37°C, without agitation (Dong et al., 2010; Kuehne et al., 2011). For solid media, 1.5% (w/v) of agar was added to mRCM media.

Both solid and liquid media were boiled in order to remove the oxygen, with the liquid media being sparged with nitrogen (N_2) during its cool down period, following sterilization (120°C, 20 min). The culture agar medium was supplemented with 15 µg/ml of thiamphenicol for plasmid selection and maintenance. All cell manipulation procedures were performed in anaerobic atmosphere of H_2:N_2 (5:95%, v/v) in an anaerobic chamber (Coy Vinyl Anaerobic Chamber - Coy Laboratory Products, Inc). In all the procedures the cells were grown until an Optical Density at 600 nm ($O.D._{600}$) between 0.6 and 1.0. As a carbon source glucose or glycerol was used. The medium was maintained at pH 6.5. The inoculum was performed by adding 10 mL from a pre-inoculum to 100 mL mRCM medium bottle. Growth of *C. pasteurianum* was determined by measuring the $O.D._{600}$ in 1 mL cuvettes with 1 cm thickness (Frilabo), using a spectrophotometer (DR 2800 ™- HACH) and sterile medium as blank sample.

Escherichia coli was grown overnight in LB medium (Luria Bertani medium - Nzytech) at 37°C under 200 rpm agitation. Solid plates were obtained by adding 1% (w/v) of agar (Nzytech) to LB broth. The medium was supplemented with antibiotic for selection when needed. *E. coli* growth was determined by measuring the $O.D_{.600}$ in a 96-well plate (Orange Scientific) using a Sunrise 96 well Microplate Reader (Tecan) and sterile medium as blank sample.

2.5. Nucleic acids methods

2.5.1. Plasmid isolation of *E. coli* and DNA purification

Plasmid isolation was performed using the GenElute™ Plasmid Miniprep kit (Sigma-Aldrich). Plasmids were isolated from recombinant *E. coli* using 5 mL LB cultures grown overnight, as described above in 2.4. Cells were collected by centrifugation and lysed by an alkaline-SDS procedure. The DNA was then adsorbed in a silica column followed by contaminants removal. Plasmids were then eluted with elution buffer (Tris-EDTA buffer, pH 8) and its quality was assessed (2.5.3).

To purify DNA fragments from agarose gel and DNA cleaning the GRS PCR & Gel Band Purification Kit (Grisp) was used. The DNA is adsorbed to a silica membrane, followed by washing out all the contaminants such as nucleotides, primers, enzymes, mineral oils, PCR additives, detergents and dyes.

DNA was also purified by precipitation using 0.1 volume of 3 M sodium acetate and 2.5 volume of ice cold absolute ethanol on ice for 15 min. It was then washed with 70% ethanol, dried at room temperature, hydrated with autoclaved deionized water and stored at -20°C.

2.5.2. Genomic DNA extraction from *C. pasteurianum* DSM 525

Total genomic DNA was isolated by phenol/chloroform extraction. Tubes with 4 mL of culture (OD_{600}= 0.7) were centrifuged and cell pellet was resuspended in 250 µL of lysis buffer (0.05 M EDTA, 0.1 M NaCl, and 10 mg/mL lysozyme, pH 8.0) for 30 min at 37°C (until visible lysis occurred). Then, 30 µL of 15% sarkozyl was added as detergent for membrane disruption, and the solution was placed on ice for 5 min. The mixture was extracted once with phenol/chloroform (500 mL, 1:1), thoroughly mixed, and the organic phenol-chloroform was separated from the aqueous layer by centrifugation at 14000 rpm for 5 min. The top aqueous layer was collected and chloroform was added (250 µL), followed by thoroughly mixing. After centrifugation at 14000 rpm for 5 min, the upper, aqueous phase was again carefully removed. The DNA was then precipitated with 3 M sodium acetate and absolute ethanol as mentioned in 2.5.1. After isolation the genomic DNA was stored at -20°C.

2.5.3. DNA sample analysis by agarose gel electrophoresis and DNA concentration determination

In order to confirm its integrity and purity, DNA was analyzed by running an agarose gel. This process is based on the separation of DNA molecules according the length in an agarose matrix. An electric field is applied, leading to the movement of the molecules with negative charge toward the cathode. The smaller the molecule, the greater the displacement in the gel. The agarose gel, 0.8 - 1% (w/v), was prepared in TAE 1X (Tris-Acetate-EDTA Buffer), stained with Sybr Safe (Invitrogen) 0.0001% (v/v), and generally run at 90 volts for 1 hour. Gel Loading Dye Blue, 6x (New England Biolabs) was added to the samples at 1x final concentration. A 1-kb DNA ladder (Solis, New England Biolabs or Nzytech), 100 bp ladder (Solis) and λ-*Hin*dIII were used as a marker. The results were visualized using the ChemiDoc™ XRS+ System (BioRad) and the Image Lab™ 4.0 Software (BioRad).

DNA concentration was usually determined using the microsample spectrophotometer NanoDrop 1000 (Thermo Fisher Scientific), but also by running a sample in an agarose gel and analyzed by the Image Lab™ 4.0 Software.

2.5.4. Gene Amplification (Polymerase Chain Reaction - PCR)

In order to obtain a tangible amount of the genes in study, a PCR (MyCycler™ Thermal Cycler, BioRad) was performed using the primers listed in Table 2.2. The PCR technique is based on exponential synthesis of DNA fragments by using two specific single stranded primers and dNTPs (Deoxyribonucleotides triphosphate - Finnzymes). Each primer pairs with one DNA sequence strand in opposite directions leading to two new double stranded DNA fragment formation. This process is performed using a thermostable DNA polymerase. It is composed by cycles, and each cycle is divided in three steps. In the first step, the double strand is denatured since a high temperature (+/- 95°C) is applied. In the second step, the primers anneal with the respective strand at a specific annealing temperature. In the third step, the elongation of the primers occurs by the action of the DNA polymerase at 72°C. As a result, after each PCR cycle the number of copies of the amplifying fragment doubles.

The genes *dpnII*, *dam* and *mdpn* were directly amplified from the *C. pasteurianum* genome. For the amplification of the fragments the Kapa HiFi Polymerase (Kapa Biosystems) was used as it provides high efficiency and 3' - 5' proofreading activity. For gene amplification the components and temperature program conditions listed in Table 2.3 were used. Each gene sequence amplified by PCR is in 6II, Table 6.4.

Table 2.3. PCR componentes and temperature program conditions to gene amplification with Kapa HiFi Polymerase.

PCR Components	PCR Temperature program		
50 - 250 ng DNA template			
1x Kapa HiFi Buffer	Initial denaturing	2 min at 95°C	
200 µM dNTPs	Denaturing	20 sec at 98°C	30x
0.5 µM Forward Primer	Annealing	30 sec at 52°C	30x
0.5 µM Reverse Primer	Elongation	1.30 min at 72°C	30x
1 U Kapa HiFi polymerase	Final Elongation	5 min at 72°C	
Ultrapure water (Finalize)	Cooling	4°C	
25 µL			

52°C to genes *dpnII*, *dam*, and *mdpn*.

2.5.5. Colony PCR

The colony PCR technique was used as screening strategy to identify positive clones with correct plasmid insertion. Colony PCR is a normal PCR that uses total cell lysate as a template. This lysate is obtained by extending the initial denaturing PCR step to 10 min. Each colony is scraped from the agar plate and directly suspended in a microtube with 10 µL of ultrapure water and the other PCR components are then added to each microtube. The respective forward primer of each gene and the T7 terminator primer or both T7 promoter and T7 terminator primers (in vectors pET, T7 promoter pairs upstream the insertion and the T7 terminator pairs downstream the insertion) were used. With this combination of primers is expected to obtain the amplification of a specific size sequence, making sure that the gene was positively and correctly inserted into the vector. The PCR usually ran with the KapaTaq DNA polymerase (Kapa Biosystems) and with the components and temperature program listed in Table 2.4.

Table 2.4. PCR components and temperature program conditions to gene amplification with KapaTaq DNA polymerase.

PCR Components	PCR Temperature program		
50 - 250 ng DNA template			
1x Kapa Taq Buffer	Initial denaturing	10 min at 95°C	
200 µM dNTPs	Denaturing	30 sec at 95°C	30x
0.4 µM Forward Primer	Annealing	30 sec at 45°C[1] or 52°C[2]	30x
0.4 µM Reverse Primer	Elongation	1.30 min at 72°C	30x
0.4 U KapaTaq polymerase	Final Elongation	5 min at 72°C	
Ultrapure water (Finalize)	Cooling	4°C	
20 µL			

[1] 45°C when T7 primers were used; [2] 52°C when gene forward primer and T7 terminator were used.

2.5.6. DNA Sequencing

DNA sequencing was used after the cloning process (2.6), after confirmation of gene insertion into the vector. This kind of technique is important since although there was insertion of the gene in the vector, undesired mutations may have occurred as insertion or deletion of nucleotides that will alter the constitution of the desired protein. Sequencing was performed by Macrogen (Amsterdam, Netherlands) using universal (T7 promoter or terminator) or custom primers (gene specific). The samples were prepared according to the company's procedures (10 µL final volume with 5 µL of 100 ng/µL of plasmid and 5 µL of 5 µM of sequencing primer). After receiving, the sequence results were aligned and compared with the predicted sequences using the AlignX software (Invitrogen).

2.6. Cloning

To clone the selected genes the vectors pET-24d(+) (5307 bp) and pETduet-1 (5420 bp) were used. These plasmids belong to a system of plasmids (pET System) that is considered the most powerful developed for the cloning and expression of recombinant proteins in *E. coli* (Novagen, 2006). pET-24d(+) carries a N-terminal T7•Tag® sequence plus an optimal C-terminal His•Tag® sequence. The sequence is numbered by pBR322 convention so the T7 expression is reversed on the map and present kanamycin resistance gene (Figure 2.1 (a)). pETduet-1 encodes two multiple cloning sites (MCS), each preceded by a T7 promoter, *lac* operator and ribosome binding site (rbs). It also presents a pBR322-derived ColE1 replicon, *lacI* gene and ampicillin resistance (Figure 2.1 (b)).

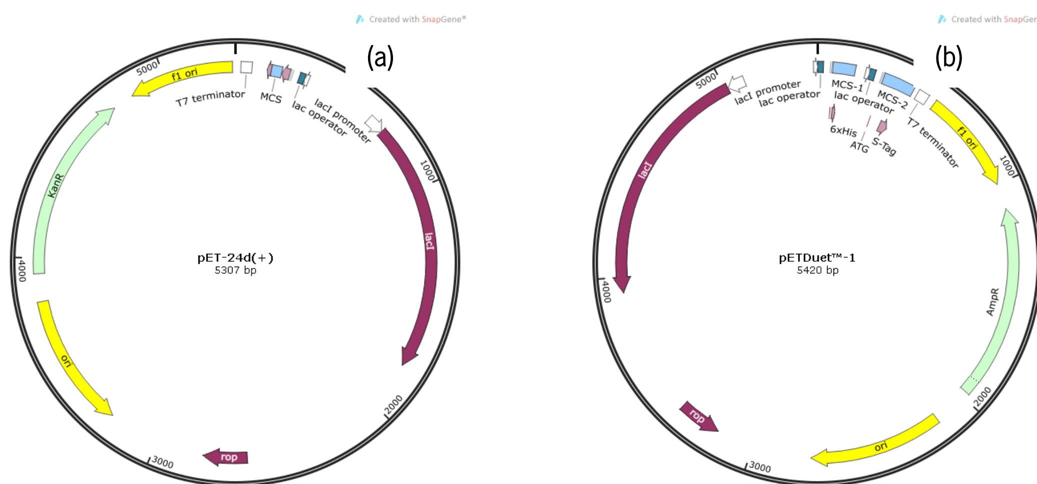


Figure 2.1. pET-24d+ (a) and pETduet (b) vector maps.

2.6.1. Double digestion

To insert the gene into the plasmid, a plasmid and respective insert (each gene amplified by PCR) digestion with both enzymes previously chosen was made. The genes *dpnII*, *dam* and *mdpn* were inserted into a vector using the enzymes *Nco*I-HF and *Bam*HI-HF (New England Biolabs). For the double reaction of vector and inserts with

*Bam*HI and *Nco*I, digestion was initiated with *Bam*HI-HF, and only after confirmation of correct digestion the second enzyme, *Nco*I-HF, was added. One µg of plasmid and 1 µg of insert were digested in NEBuffer 4, BSA 1 µg/mL, and 10 U of *Bam*HI HF in 50 µL final volume at 37°C for 6 h. Afterwards, the correct digestion of the plasmid was confirmed by agarose gel electrophoresis, and, to the same microtube, 10 U of *Nco*I-HF was added. The reaction ran overnight at the same conditions.

2.6.2. Dephosphorylation of restriction fragments

To prevent reattachment of plasmid ends after digestion with one restriction enzyme the enzyme antarctic phosphatase was used, according to the manufacturer's instructions (New England Biolabs). This enzyme has the capability to remove phosphate groups formed at 5´end of DNA molecules (digested plasmids), hindering the reattachment of the ends. To the overnight plasmid double digestion reaction 5 U of antarctic phosphatase and the antarctic phosphatase buffer were added in 60 µL final volume and the reaction was incubated at 37°C for 1 h. Then the enzyme was inactivated at 65°C for 5 min. Subsequently, the digestions were confirmed by running an agarose gel and the Image Lab 4.0 Software was used to analyze the DNA concentration. The DNA was washed using the DNA cleaning the GRS PCR & Gel Band Purification Kit (Grisp) followed by plasmid-insert ligation procedure.

2.6.3. Ligation of DNA fragments

Plasmids and inserts were ligated using the enzyme T4 DNA Ligase, according to the manufacturer's instructions (New England Biolabs), in an incubation step at RT (room temperature) for 1 h. This enzyme, ATP dependent, catalyzes the formation of phosphodiester bonds between juxtaposed 5´phosphate and 3´hydroxyl ends in double stranded DNA. To 50 ng of plasmid the concentration of insert added was calculated using the formula:

$$\text{Insert (ng)}=\frac{[\text{gene length (bp)}]}{[\text{insert length (bp)}]}*\text{vector (ng)}$$

in a 1:3 proportion (vector:insert). After DNA ligation, T4 DNA ligase was inactivated at 65°C for 10 min. Besides plasmid-insert ligation a "ligation" with the empty vector was also performed. This procedure is useful to prevent the vector from back ligating. The ligation was then used to transform competent *E. coli* cells.

2.6.4. Introduction and maintenance in a host organism

As it was intended to make a direct transformation using a ligation product, the easiest method is the transformation by heat shock. The high salt concentration in a ligation reaction could lead to an electrical discharge when applying an electric field to the competent cells, reducing the cell viability. Another DNA cleaning step could be included, but this might result in loss of product. Thus, commercial chemo-competent JM109 *E. coli* cells (Nzytech) were transformed with the ligation products using the heat-shock bacterial transformation procedure. To 50 µL of

freshly thawed competent cells 10 µL of ligation mixture was added (aprox. 25 ng of plasmid). The remaining procedure was performed as mentioned in 2.9.2.2.

To confirm gene insertion a plasmid extraction and digestion from transformants was carried out. One colony was scraped from the plate and suspended in 5 mL of LB broth with antibiotic and allowed to grow overnight. Then, a plasmid extraction and linearization was done, using the size to verify whether the gene was or was not inserted into the vector.

2.7. *C. pasteurianum* crude extract

Crude extracts of *C. pasteurianum* were prepared from a 100 mL early exponential phase (DO_{600} = 0.6) culture. The extraction was based on the method of Azeddoug & Reysset (Azeddoug and Reysset, 1991; Jennert et al., 2000). From the initial 100 mL culture, 20 mL were placed on ice, the remaining was discarded. From this moment the cells were always kept on ice and the buffers in use are maintained refrigerated. The pellet was washed twice, each time resuspended with 10 and 5 mL of T68 basal medium (6I, Table 6.2) and then centrifuged (4°C, 5000 rpm, 10 min) and carefully resuspended in 250 µL of lysis buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0, and 10 mM β-mercaptoethanol). The cells were disrupted by sonication at 4°C (3 times 20 sec sonication ON, and 10 sec sonication OFF) and were allowed to rest for 10 min on ice. The lysate was centrifuged (RT, 5000 rpm, 5 min) to obtain the supernatant, supplemented with 50% (v/v) glycerol and stored at -20°C.

C. pasteurianum crude extract was also prepared using the previous protocol but with wash buffer (Tris-HCl 0.1 M, pH 7.5, and NaCl 0.5 M) adjusting the supernatant with 20% (v/v), final concentration, of glycerol.

2.8. DNA modification

2.8.1. Plasmid enzymatic cleavage

To verify plasmid digestion pattern of some enzymes or just to make a plasmid linearization, digestions were performed with commercial restriction enzymes. Linearizations were performed according to the respective enzyme manufacturer's instructions, normally using 100 ng to 200 ng of plasmid and 1 µL of enzyme in 10 µL final volume at 37°C for 2 h. The digestions of methylated DNA (M.*Sss*I) (2.8.2) were performed after methylation reactions, at the same conditions but in a final volume of 20 µL.

2.8.2. DNA methylation

In vitro methylation of pMTL007C-E2::Cpa-*spo0A*-666a was carried out using the commercial methyltransferases M.*Msp*I and M.*Sss*I (New England Biolabs), according to the manufacturer's instructions. M.*Msp*I methylates the sequence CCGG on the external cytosine (^{m5}C) and M.*Sss*I methylates the cytosine residue of CG sequence (^{m5}C). To the electro-transformation 5 U of M.*Msp*I were used to methylate 10 µg of plasmid in a final volume of 200 µL in 80 µM of SAM at 37°C for 1 h. For the restriction assays 2.5 to 25 U were required, at the same conditions, using 10 µL as final volume. After methylation the enzyme was then inactivated by heat incubation

at 65°C for 15 min, followed by ethanol precipitation of the DNA (see 2.5.1). The precipitation step was not necessary for restriction assays.

Normally, 4 M.*Sss*I U were used to methylate 100 ng to 300 ng of plasmid in a final volume of 10 µL at 37°C for 1 h. The enzyme reaction was then inactivated at 65°C for 20 min.

2.8.3. DNA condensation by spermidine

The protocol to condense the plasmid DNA by the polyamine spermidine was the same as used by Vasu and colleagues (Vasu et al., 2012). Five hundred ng of plasmid were mixed with spermidine (Applichem), to reach a final concentration of 50 mM, 25 mM, 10 mM, 5 mM, 2.5 mM, 1 mM and 0.5 mM in 10 mM Tris-HCl (pH 7.4). Samples were kept on ice for 15 min before the addition of 1 µL of crude extract and the digestion buffer, 1 µL of NEBuffer 2. The digestion reaction, with a final volume of 10 µL, was held at 37°C for 2 h and the results were analyzed by agarose gel electrophoresis.

2.8.4. *C. pasteurianum* restriction endonuclease assay

The crude extract activity was tested by incubating the plasmid pMTL007C-E2::Cpa-*spo0A*-666a with crude extract of *C. pasteurianum*. An amount of 350-400 ng of plasmid was incubated with 1 µL of crude extract at 37°C for 2 h in NEBuffer 2 and BSA in a final volume of 10 µL (Jennert et al., 2000). The products were analyzed by agarose gel electrophoresis.

2.9. Bacterial cell transformation

2.9.1. Electro-transformations *of C. pasteurianum* DSM 525

In *C. pasteurianum* transformations several parameters were evaluated, such as carbon source (glucose or glycerol), time of cell harvesting (early or late exponential phase), electroporation buffer (with or without magnesium chloride - $MgCl_2$), cell quantity used to transform (300 µL or 570 µL), plasmid quantity (ng to µg), electroporation equipment conditions (voltage - volts, resistance – ohm, and capacitance - microfarad), quantity of recuperation medium (1.5 mL or 3 mL) and recovery time (4 or 16 h). The use of methylated (methyltransferases) and condensed (spermidine) DNA was also tested. In all electroporation experiments three controls were used.

- Non-competent and non-electroporated cells without plasmid placed in a plate without antibiotic. The cells, without any interference, must grow; this control is used to compare with the other cells that passed by competence and electroporation processes. It is also used to be sure that there were no problems with oxygen during incubation.
- Competent and electroporated cells without plasmid placed in a plate without antibiotic. It is used to analyze the quality of competent cells. They must recover and grow after the electric shock as a good indicator about the cell viability after electroporation.

- Competent and electroporated cells without plasmid placed in a plate with antibiotic. This is the negative control; if the antibiotic concentration is correct, the cells cannot grow as they do not have the plasmid that confers such resistance.

2.9.1.1. *C. pasteurianum* electrocompetent cells

C. pasteurianum cells were grown in 100 mL mRCM medium at 37°C until early or late exponential phase ($O.D._{600}$ = 0.6 - 1). After reaching the correct O.D., the culture was kept on ice for 20-30 min. Once inside the anaerobic chamber the culture was divided to two pre-chilled (-20°C) 50 mL sterile tubes (45 mL each) and the remaining was maintained in ice to use later on plating as control. From this moment the cells were always maintained on ice. The tubes were then centrifuged at 4°C and 5000 rpm for 10 min. The supernatant was carefully discarded and the pellet resuspended in 5 or 10 mL of pre-chilled electroporation buffer (5 mM sodium phosphate, 270 mM sucrose, 1mM $MgCl_2$) without vortexing. The cells were centrifuged as before. Supernantant was discarded and the cells were resuspended in 1.5 or 2 mL electroporation buffer. At this point the cells are ready for electroporation and are stored on ice.

2.9.1.2. *C. pasteurianum* electro-transformation

To perform electroporation of *C. pasteurianum* cells plasmid DNA, pMTL007C-E2::Cpa-*spoOA*-666a, was added to the competent cells. Inside the anaerobic chamber electrocompetent cells and different quantities of DNA were placed in pre-chilled electroporation cuvettes (0.2 or 0.4 cm Gene Pulser/MicroPulser Electroporation Cuvettes, BioRad). These cuvettes, permanently on ice, were taken out from the anaerobic chamber and immediately electroporated using different electroporation conditions (V-volts, µF-microfarad, Ω-ohm) in the Gene Pulser Xcell Microbial System (Bio-Rad). Immediately after electroporation the mixture in the cuvettes was suspended in 1 mL of pre-warmed (37°C) anaerobic recovery medium (mRCM), taken from 10 mL serum bottles containing 1.5 or 3 mL medium each. The mixture was put back into serum bottles with a sterile 2.5 mL syringe and was allowed to recover for 3 or 16 h at 37°C.

After recovery, the serum bottles were taken to the anaerobic chamber and cell plating was performed. From each bottle 200 µL of culture were seeded in pre-reduced RCM-agar plates supplemented with 15 µg/mL thiamphenicol. The remaining volume was concentrated by centrifugation at room temperature. The supernatant was discarded and the pellet was resuspended in 200 µL medium and plated the same way. After seeding the plates, they were placed inside an anaerobic box (AnaeroPack Rectangular Jar, Mitsubishi Gas Chemical Company, Inc.) which was placed at 37°C until visible growth occurred.

The electroporation parameters analyzed along the work are shown in Table 2.5.

Table 2.5. *C. pasteurianum* electro-transformation parameters.

Carbon source	Growth phase ($O.D._{600}$nm)	Electroporation Buffer	Transformed cells (mL)	Plasmid (ng)	Plasmid conditions	Electroporation conditions	Recuperation medium (mL)	Recovery time (h)
Glucose Glycerol	Early to Late Exp. Phase	With or without MgCl2	300 - 570	0.1 to 4	Methylated Condensed	2/4 mm cuvette; 1.25/2 kV, 25 µf, 100 Ω	1.5 to 3	3 to 16

2.9.2. E. coli transformation

Since *E. coli* cells are naturally non-competent, the cells need to be subjected to a process to make them competent. The heat- shock is a common *E. coli* transformation method.

2.9.2.1. Chemically competent *E. coli* cells

E. coli cells were made competent using $CaCl_2$. The positive charge of cations is responsible for condensation of naked DNA by shielding the negative phosphate groups, hence making a smaller DNA packet for cell uptake. The cations also seem to cancel out repulsive forces between DNA and the outer membrane and thereby facilitate DNA-membrane interactions (Weston et al., 1981). Cells were grown in 5 mL LB broth overnight at 200 rpm and 37°C. The initial culture was diluted (1:100) in fresh medium and was grown until an $O.D._{600}$ of 0.5 (about 3 h) under the same conditions. The culture was centrifuged at 5000 rpm and 4°C for 10 min, the pellet was resuspended in ice-cold 0.1 M $CaCl_2$ in half the culture volume and was maintained on ice for 30 min. After a second centrifugation the pellet was resuspended in 0.1 M $CaCl_2$ in 1/10 culture volume. The cells were immediately used or frozen at with 10% glycerol at -80°C.

2.9.2.2. Heat shock transformation of *E. coli*

All the *E. coli* transformations were performed using a heat-shock procedure. The exposure to sudden high temperatures creates a pressure difference between cell inside and outside inducing pores formation. To 50 µL of freshly thawed chemically competent cells 10 µL DNA were added in case of ligation, or 0.5 µg to 1 µg of normal plasmid. The mixture was then incubated on ice for 30 min. Thereafter it was exposed to heat shock at 42°C for 40 sec to induce DNA uptake and was again incubated on ice for 2 min. It was allowed to recover with the addition of 950 µL of SOC medium (SOB (Nzytech) with 20 mM of glucose) and was incubated on a rotatory shaker at 200 rpm, and 37°C for 1 h. From each sample 10 µL, 100 µL and the remaining volume after concentration were plated. The sample concentration was performed at the maximum velocity in a microcentrifuge (ScanSpeed mini centrifuge - Labogene) for 2 min, 800 µL of medium were discarded and the pellet was carefully resuspended with the remaining supernatant. The samples were plated in LB - agar medium with the appropriate antibiotic concentration. The plates were then incubated overnight at 37°C.

2.10. Protein Production and analysis

2.10.1. Protein Expression

To produce the desired proteins, the transformation of the cells *E. coli* BL21 (DE3) with the vectors correctly inserted was needed. This *E. coli* strain is successfully used for high level protein expression with an easy induction. Transformed plasmids containing T7 promoter driven expression (e.g. pET) are repressed until IPTG (Isopropyl β-D-1-thiogalactopyranoside, Nzytech) induction of T7 RNA polymerase from a lac promoter. They express efficiently heterologous proteins since they lack the lon protease and the OmpT outer membrane protease that degrade foreign proteins. Chemocompetent BL21 (DE3) cells were used and the transformation was performed as described in 2.9.2.2, using 50 to 150 ng of plasmid construction, or empty vector (control) and 50 µL of competent cells. A BL21 (DE3) colony with the construction and a colony with the empty vector were scraped and incubated in 5 mL LB medium with antibiotic overnight at 200 rpm and 37°C. Fresh medium (50 mL) was inoculated (1:100) with an overnight culture until an $O.D._{600}$ of 0.6. Then, 2 mL were collected as no-induction control, and 1 mM IPTG (final concentration) was added to the remaining cultures. Subsequently, the cultures were placed at 16°C or 37°C and samples were collected at 0, 1, 2, 4 and 20 hours (overnight) after induction. The collected samples were centrifuged at maximum velocity (13000 rpm) in a microcentrifuge for two min and stored at -20°C to posterior cell rupture and analysis.

2.10.2. Cell rupture

To obtain the proteins produced in 2.10 it was necessary to burst the cells and release the internal content. To achieve this effect the physical method sonication was used. The cell pellet was thawed on ice and suspended with 100 µL of ice-cold 20 mM Tris-HCl. From this moment the cells were always maintained on ice to prevent proteolysis. The sonication was performed by pulsing the samples 15 sec, 3 times, with 10 sec intervals using the sounder 630-0422 with 40% amplification. Thirty µL of TE (total extract) were collected and stored and the remaining TE was centrifuged at 15000 rpm and 4°C for 15 min. The supernatant or the CFE (cell free extract) was taken and stored. The TE and the CFE of each sample were analyzed by polyacrylamide gel electrophoresis in order to confirm the protein presence and size in insoluble or soluble form, respectively.

2.10.3. SDS-PAGE analysis

Each protein extract was analyzed by SDS-PAGE (Sodium dodecyl sulfate – polyacrylamide gel electrophoresis), using a discontinuous denaturing buffer system (Laemmli, 1970). This technique is based on protein separation, according to their size, when an electric field is applied. A discontinuous system is used, where a large-pore stacking gel on top of a small-pore resolving gel is used. The large pores of the stacking gel facilitates protein movement, but the small pores of resolving gel slow down the protein movement, which firstly causes protein condensation, making the separation of the proteins start at the same time, which improves overall

resolution. The pore size of the resolving gel can be controlled by adjusting the acrylamide/bis-acrylamide concentration; the higher the protein size, the lower the acrylamide/bis-acrylamide concentration should be.

The samples were treated with Laemmli sample buffer (2x sample buffer: 4% SDS, 20% glycerol, 10% β-mercaptoethanol, 0.004% bromophenol blue and 0.125 M Tris HCl, pH approx. 6.8). The sample is mixed with the buffer (1:1) and boiled for 5 min at 95°C. β-mercaptoethanol is used to reduce the intra- and inter-molecular disulfide bonds. SDS is a detergent important to protein denaturation and to mask the protein intrinsic charge as negative, which provides a protein separation exclusively by size. The dye bromophenol blue marks the forward running and facilitates sample manipulation because it becomes stained. The glycerol increases the sample density, facilitating their entry into the well. Routinely the Low Molecular Weight (LMW) Protein Marker (Nzytech) was used, the gels were of 0.75 mm with 13% acrylamide/bis-acrylamide (BioRad) solution (the constitution of the gels is listed in 6I, Table 6.3). The electrophoresis was performed in Tris/Glycine Buffer 1X (25 mM Tris, 192 mM glycine, and 0.1% SDS pH 8.3 – BioRad) using the system Mini-PROTEAN Tetra system (BioRad) at 100 volts for 2 h 15 min.

After electrophoretic separation the gels were stained by Coomassie Blue or silver staining depending on the lower or higher desired sensitivity, respectively.

2.10.3.1. Coomassie Blue staining

The polyacrylamide gel was placed in a staining solution (Coomassie blue 0.25% (w/v), methanol 50% (v/v) and acetic acid (10% (v/v)) for 15-30 min. The solution was removed and the destaining buffer was added (methanol 25% (v/v) and acetic acid 5% (v/v)) until bands start to become visible and the background is lost. The gel was then preserved in ultrapure water.

2.10.3.2. Silver Staining

When using silver staining, the gel was placed in fixation solution (ethanol 50% (v/v), acetic acid 12% (v/v), formaldehyde 0.05% (v/v)) for 2 h at RT to protein impregnation and to ions and detergents removal. The solution was removed and the rinsing solution was added (ethanol 20% (v/v)) for 20 min, renewing the solution three times. The sensitizing solution substituted the last one (sodium thiosulfate 0.02% (w/v)) for 2 min and was rinsed three times with ultrapure water (3x1 min). The water was discarded and gel was placed in ice-cold staining solution (silver nitrate 0.2% (w/v), formaldehyde 0.076% (v/v)) in the dark for 20 min. The gel was rinsed again with abundant water to remove all free ions. The water was removed again and the gel was placed in reducing solution (formaldehyde 0.05% (v/v), sodium carbonate 6% (w/v), sodium thiosulfate 0.0004% (w/v)) until bands appear with the desired staining. The reaction was stopped by the addition of stop solution (acetic acid 12% (v/v)).The gels were preserved in ultrapure water.

2.10.4. Protein purification

To separate the desired proteins from other contaminants protein purification was performed using the 0.2 mL HisPur™ Ni-NTA spin columns (ThermoScientific). These columns enable effective immobilized metal affinity chromatography (IMAC) on purification of polyhistidine-tagged proteins from a soluble protein extract. They are constituted by nickel-charged nitrilotriacetic acid (NTA) chelate immobilized onto 6% crosslinked agarose resin. These resins are required due to the presence of four metal binding sites on the chelate, allowing for a high-binding capacity and low metal leaching. It is recommended to use buffers with imidazole, because of the structural similarity to histidine residues of the 6x His-tag. At lower imidazole concentrations, nonspecific, low affinity binding of background proteins is prevented, while 6xHis-tagged proteins still bind strongly to the Ni-NTA matrix. At higher concentrations, imidazole can also bind to the nickel ions and disrupt the binding of dispersed histidine residues in non-tagged background proteins.

The column was equilibrated with 2 mL equilibration buffer (20 mM sodium phosphate, 300 mM sodium chloride with 10 mM imidazole; pH 7.4) and was centrifuged at 700 g for 2 min to remove the buffer. Meanwhile the 2 mL of protein sample were also mixed with the same proportion of equilibration buffer, and were passed through the column, while being mixed for 30 min. As the sample volume was greater than the column volume, several applications (0.5 mL) were performed until the entire sample had been processed. After each application the column was centrifuged at 700 g for 2 min and the flow-through was always recovered. Once the sample column insertion was finished, the column was washed 3 times with wash buffer (20 mM sodium phosphate, 300 mM sodium chloride and 25 mM imidazole, pH 7.4), the flow-through were also recovered. The His-tagged proteins were eluted from the resin by adding 0.5 mL of elution buffer with increasing imidazole concentrations (50 mM, 100 mM, 250 mM and 350 mM) to verify the better imidazole concentration to elute the proteins. The simplified process is shown in Figure 2.2. Each collected fraction was analyzed by SDS-PAGE as mentioned in 2.10.3.

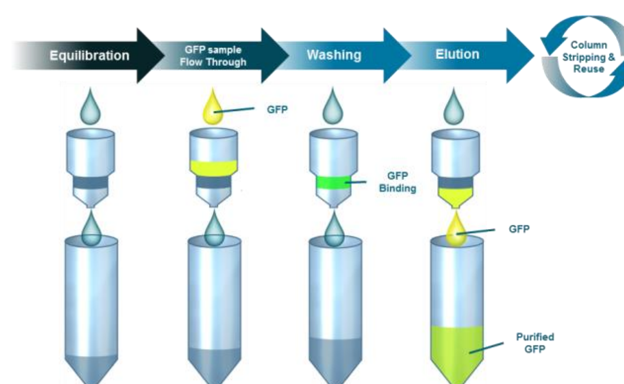


Figure 2.2. Simplified protein purification process using the His-Pur™ Ni-NTA columns with the GFP protein. First the column is equilibrated, then the also equilibrated protein sample is passed through the column. Several wash steps are done and at last the protein is eluted using higher imidazole concentrations. From (http://www.yeastern.com/Products.php?pkid=12&pttype=69).

2.10.5.Western Blottting

Western blotting, or immunoblotting, is a common technique used to detect specific proteins by the use of a specific antibody. A protein-antibody complex is formed, that is later identified by several detection methods. To perform the Western blot technique the Mini Trans-Blot® Electrophoretic Transfer Cell and the PVDF-Transfer membrane (Thermo Scientific) were used. After SDS-PAGE the polyacrylamide gel was incubated for 5 min in 1x blotting buffer (2 M glycine and 250 mM tris) and the membrane was pre-wet in 100% methanol for 5 min and rinsed with water. The system was mounted in a wet-system and a voltage of 30 V was applied for 1 hour for membrane protein immobilization. After transference the membrane was washed under shaking conditions in TBST (10 mM Tris-HCl, 150 mM NaCl, 0.05% Tween-20 and 2% BSA) for 30 min. Subsequently, the membrane was washed in blocking buffer (TBST with 2% BSA) for 30 min to prevent non-specific antibodies binding to non-protein areas. The incubation was performed with the primary Anti-His Antibody (GE Healthcare) for 30 min to achieve specific his-tag protein detection. Then the membrane was washed 3x10 min with TBST to remove unbound antibodies. Finally, the Clarity™ Western ECL Substrate (BioRad) was added for 5 min for protein detection. The membrane was washed one last time with TBST and visualized with the ChemiDoc™ XRS+ System (BioRad).

Chapter 3

Results and Discussion

3. RESULTS AND DISCUSSION

3.1. Glycerol as *C. pasteurianum* carbon source

The use of glycerol, particularly, crude glycerol (biodiesel waste), as a carbon source for anaerobic bacteria such as *C. pasteurianum*, has been discussed and analyzed in the last few years. This bacterium is capable of metabolizing glycerol, producing a variety of compounds such as 1,3-PDO, ethanol and butanol, which are of great value to the biofuels field (Ahn et al., 2011; Biebl, 2001; Dabrock et al., 1992; Jensen et al., 2012a; Nakas et al., 1983; Taconi et al., 2009; Venkataramanan et al., 2012). The tolerance of *C. pasteurianum* to high glycerol concentrations is remarkable when compared to the tolerance to glucose. Thus, the product yields resulting from the *C. pasteurianum* growth in glycerol are also greatly enhanced (Dabrock et al., 1992).

C. pasteurianum growth, using glucose and glycerol as a carbon source, was briefly analyzed and the results are presented in Figure 3.1. To determine the better carbon source for *C. pasteurianum* growth, glucose (20 g/L) and glycerol (20 g/L) were added to the mRCM.

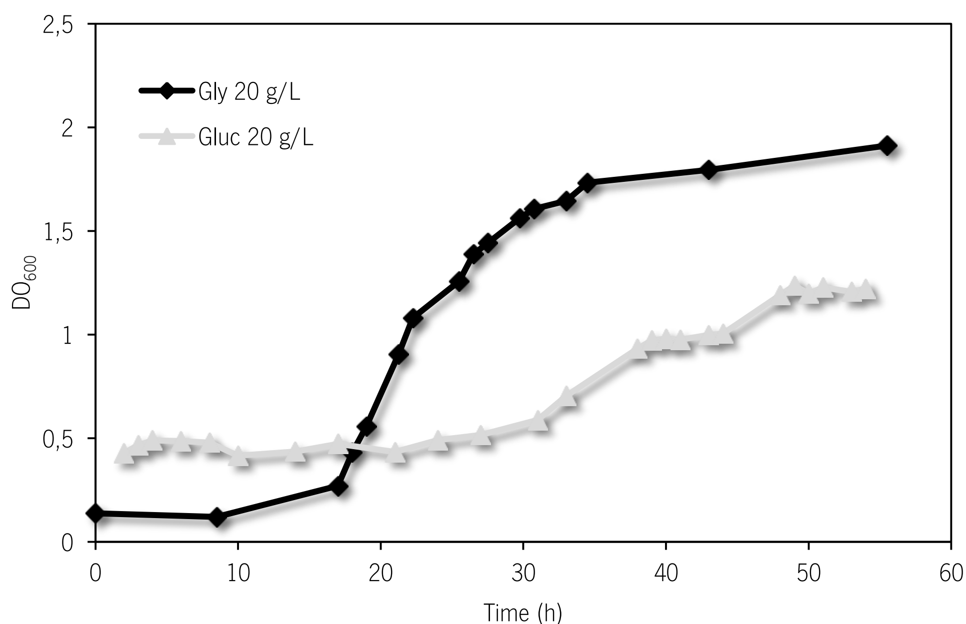


Figure 3.1. ***C. pasteurianum*** **growth using glucose (20 g/L) and glycerol (20 g/L) as a carbon source.** *C. pasteurianum* growth was analyzed for 55.5 h in the experiments conducted with glycerol and for 54 h in those conducted with glucose.

The data shown suggest that glycerol is a more effective carbon source. Visually, the growth phases can be extrapolated; with glycerol the lag phase (0 h to 10-12 h) is shorter and more defined than with glucose (0 h to 25-30 h). Analyzing the exponential growth phase, it can be suggested that with glycerol the cells grow much more ($O.D._{600}$ – 1.7-1.8) and faster (35 h to 40 h) than with glucose ($O.D._{600}$ – 1.2; 50 h). Therefore, the use of glycerol as carbon source can be advantageous to grow *C. pasteurianum*.

In addition to the positive effects described, the use of glycerol also decreases the contamination risk, since a limited number of organisms are able to use this component as a carbon source.

3.2. Preliminary electro-transformations of *C. pasteurianum*

To develop a *C. pasteurianum* transformation protocol some preliminary transformations were assayed, as described in Materials and Methods - 2.9.1. These experiments differ in the way of making competent cells and in the electroporation conditions. Parameters such as carbon source (glucose or glycerol), time of cell harvesting (early or late exponential phase), electroporation buffer (with or without $MgCl_2$), cell volume used to transform (300 µL or 570 µL), plasmid amount (0.1 to 5 µg), electroporation equipment conditions, quantity of recuperation medium (1.5 mL or 3 mL) and recovery time (4 or 16 h) were analyzed.

The mentioned parameters were selected based on parameters that affect the electro-transformation efficiency (Bio-Rad-Catalogue, 2000) and in parameters used to other *Clostridium* species transformation. In electroporation, for most bacteria species, the highest transformation yields are obtained when cells are harvested in early to mid-log growth, the $O.D._{600}$ used in the current work, 0.7 and 1.1, is within this parameter (Jennert et al., 2000; Nakotte et al., 1998; Olson and Lynd, 2012). The electroporation medium must be preferentially constituted by non-ionic solutions, since their presence may result in sample arcing; however, the presence of ions benefits the transformation. Therefore, the best electroporation buffer must be prepared in such a way that benefits the transformation without arcing. To test this, electroporation buffer with and without $MgCl_2$ (Mg^{2+} and Cl^-) was used. The transformation efficiency increases with the cell concentration within the range of 10^9 to at least 3×10^{10} cells/mL. The DNA concentration to be used depends on the final objective; the use of a high DNA concentration leads to high transformation frequency (transformant/survivor), while low DNA concentrations and high cell concentration leads to high transformation efficiency (transformants/µg DNA). These parameters were tested using 300 µL or 570 µL of cells with a range of plasmid concentration (100 ng to 1000 ng). The field strength, 6,25 kV/cm on the first experiment and 5 kV/cm on the second and third experiments are in accordance with those used in bacterial transformation (Chen et al., 1996; Tardif et al., 2001). The recovery time is a highly undefined parameter, on literature, after the shock step, can be observed recover times of 0 h to 16 h (Nakotte et al., 1998; Olson and Lynd, 2012; Oultram et al., 1988; Tardif et al., 2001; Tyurin et al., 2004). To test the importance of this parameter, recovery times of 3 or 16 hours were used. The analyzed parameters and the changed conditions are summarized in Table 3.1.

As a vector, the plasmid pMTL007C-E2::Cpa-*spo0A*-666a (9,033 bp) was used. This is a construct of the plasmid pMTL007C-E2 (from the ClosTron system developed by Minton and coworkers (Kuehne et al., 2011)) retargeted to the intron target site at position 666a of the *spo0A* gene of *C. pasteurianum*. The *spo0A* gene is a main regulator of sporulation in *Bacillus* and *Clostridium* (Zhao et al., 2002). This vector can be used as a proof of principle; the cells acquiring this plasmid will have the gene *spo0A* deleted and would be phenotypically asporogenous.

Table 3.1. *C. pasteurianum* preliminary electrotransformation parameters. Each experiment was performed in duplicated. The principal parameters changed are highlighted at bold.

Carbon source	Growth phase (O.D.$_{600}$ nm)	Electroporation Buffer	Transformed cells (mL)	Plasmid (ng)	Electroporation conditions	Recuperation medium (mL)	Recovery time (h)
Glucose 20 g/L	**0.7**	* 100 Mm Sodium Phosphate Buffer * 284 Mm Sucrose *** 1mM Magnesium Chloride ($MgCl_2$)**	**300**	100 g, 500 ng, 1000 ng	**2 mm** cuvette; 1.25 kV, 25 µf, 100 Ω	**1.5**	**16**
Glucose 20 g/L	**1.1**	* 100 Mm Sodium Phosphate Buffer * 284 Mm Sucrose	**570**	100 g, 500 ng, 1000 ng	**4 mm** cuvette; 2 kV, 25 µf, 100 Ω	**3**	**3**
Glycerol 20 g/L **Glycerol 40 g/L**	1.354 (20 g/L) 0.422 (40 g/L)	* 100 Mm Sodium Phosphate Buffer * 284 Mm Sucrose	570	100 g, 500 ng, 1000 ng	4 mm cuvette; 2 kV, 25 µf, 100 Ω	3	3

No positive results were obtained with these electro-transformation assays. The applied conditions could be not sufficient to create pores in this Gram-positive bacterium, and consequently there is no possibility for the DNA to enter into the cells, implying that more aggressive methods should be used to eliminate the physical barrier (cell wall). During the electric treatment the cells were exposed to oxygen for a short period of time (1 min, approximately); being a strict anaerobe this hostility may have triggered *C. pasteurianum* cells to sporulate. In the spore form, the DNA can neither enter the cells nor transform them, which could explain the inefficiency of the transformation.

3.3. M.*Msp*I role on *C. pasteurianum* DSM 525 transformation

The host restriction systems have been highlighted as a major barrier in the transformation of *Clostridium* species (Mermelstein and Papoutsakis, 1993). This barrier is composed by a group of specialized restriction enzymes responsible for the protection of cells against foreign DNA. As a *Clostridium* member, *C. pasteurianum* may also have such a defense system. These restriction enzymes (REases) are normally associated with other enzymes, the methyltransferases (MTases). The role of the latter is the addition of methyl groups to the same sequences recognized by the restriction enzymes, thereby preventing cleavage of host DNA. Transformations in *Clostridium* have only been achieved when methylated DNA (Mermelstein and Papoutsakis, 1993) is used or if the restriction system is inactivated (Dong et al., 2010).

A common methyltransferase used to hinder the endonucleases restriction of exogenous DNA is the enzyme M.*Msp*I. This enzyme recognizes the sequence 5'-CCGG- 3' and methylates the first cytosine residue. The plasmid under study was methylated as mentioned in 2.8.2 and used as vector in the *C. pasteurianum* transformation. The transformation parameters are described in Table 3.2.

Table 3.2. *C. pasteurianum* electrotransformation parameters using *Msp*I methylated DNA Each experiment was performed in duplicate. The principal parameters changed are highlighted at bold and underlined.

Carbon source	Growth phase (O.D.$_{600}$ nm)	Electroporation Buffer	Transformed cells (mL)	Plasmid (ng)	Electroporation conditions	Recuperation medium (mL)	Recovery time (h)
Glycerol 20 g/L	0.7	* 100 Mm Sodium Phosphate Buffer * 284 Mm Sucrose	570	**500 ng, 1000 ng, 4000 ng** <u>MspI methylated</u> and non-methylated vector	4 mm cuvette; 2 kV, 25 µf, 100 Ω	3	16

Although a high vector quantity was used to improve the transformation probability, no positive results were obtained with this experiment. This can be explained by either the electroporation protocol inefficiency or by plasmid cleavage, i.e. M.*Msp*I methylation could not protect the plasmid from *C. pasteurianum* restriction system.

3.4. Crude extract of *C. pasteurianum* exhibits restriction activity

Since methylation with a common, commercially available, methyltransferase did not result in successful transformation, it was decided to examine in detail what the main restriction-modification system of *C. pasteurianum* is. As previously mentioned, there was a suspicion that a restriction system exists, to prove this suspicion a characterization of the restriction system of *C. pasteurianum* was performed. Crude extracts were prepared based on the method described by Azeddoug and Reysset (as mentioned in 2.7).

Plasmid pMTL007C-E2::Cpa-*spo0A*-666a (Figure 3.2 (a), lane 1) was incubated with crude extract as indicated in 2.8.4. Experiments with non-methylated (lane 4) and M.*Msp*I methylated DNA were performed. The action of different enzyme units was also analyzed (lane 6 – 2.5 U and lane 8 – 5 U). As it can be observed, with and without methylation, a discrete DNA digestion pattern is present, i.e. not a smear. This means that the *C. pasteurianum* crude extract has an active restriction system, i.e. REases digest the foreign DNA, since they don't recognize it as its own.

Lanes 5, 7 and 9 of Figure 3.2 are digestion reactions without reaction buffer and BSA. As no digestion occurs without these components they seem to be indispensable for the reaction to occur.

The methyltransferase quantity required to protect the DNA and hinder the restriction was questioned, therefore restriction reactions using DNA methylated with 2.5, 12.5 and 25 methyltransferase U were performed (Figure 3.2 (b)). The DNA cleavage by the enzymes present in the crude extract was not hindered even in the presence of high amounts of methyltransferase.

These results indicate that the commonly used *Msp*I methyltransferase does not protect the foreign DNA from the restriction system of *C. pasteurianum.* Therefore, it can be concluded that *C. pasteurianum* has a distinct restriction system as compared to other clostridia.

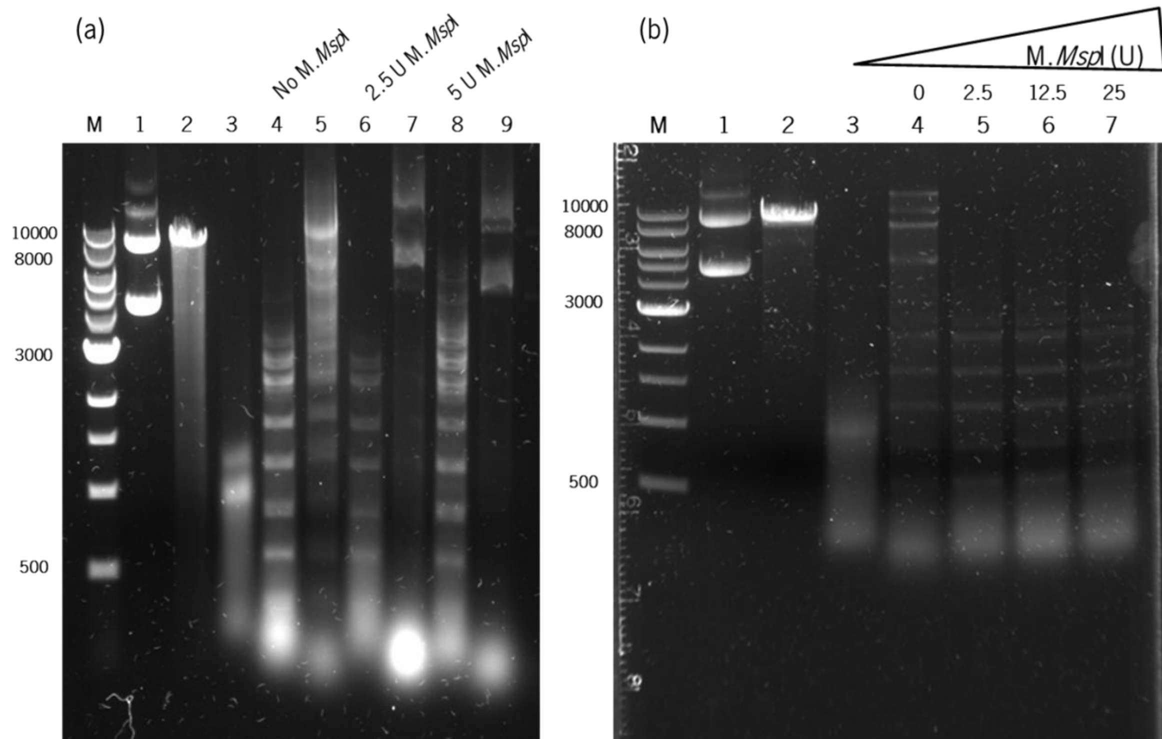


Figure 3.2. *C. pasteurianum* crude extract characterization. (a) Analysis of crude extract activity against M.*Msp*I methylated and non-methylated DNA. M. 1kb ladder (bp) – NEB; 1, pMTL007C-E2::Cpa-*spo0A*-666a; 2, *Hind*III linearized plasmid; 3, *C. pasteurianum* crude extract; 4, Plasmid digested by *C. pasteurianum* crude extract with BSA and NEBuffer 2; 5, Plasmid digested by *C. pasteurianum* crude extract without reaction buffer and BSA; 6, M.*Msp*I (2.5 U) methylated plasmid digested by *C. pasteurianum* crude extract with BSA and NEBuffer 2; 7, M.*Msp*I (2.5 U) methylated plasmid digested by *C. pasteurianum* crude extract without reaction buffer and BSA; 8, M.*Msp*I (5 U) methylated plasmid digested by *C. pasteurianum* crude extract with BSA and NEBuffer 2; 9, M.*Msp*I (5 U) methylated plasmid digested by *C. pasteurianum* crude extract without reaction buffer and BSA. (b) Analysis of methyltransferase quantity required to protect the plasmid. M, 1kb ladder (bp) – NEB; 1, pMTL007C-E2::Cpa-*spo0A*-666a; 2, *Hind*III linearized plasmid; 3, *C. pasteurianum* crude extract; 4, Plasmid digested by *C. pasteurianum* crude extract with BSA and NEBuffer 2; 5, 6, and 7 M.*Msp*I methylated plasmid digested by *C. pasteurianum* crude extract with BSA and NEBuffer 2 with 2.5 U, 12.5 U and 25 U of enzyme, respectively.

3.5. DNA condensation by the polyamine spermidine

It is known that besides methylation other methods are capable of DNA protection against REases action. One example that has been reported includes the use of polyamines.

Polyamines, like spermine and spermidine, are largely present in nature as amines with a huge biological importance, since numerous microorganisms require them for growth. They are normal cell constituents that occur intracellularly at high concentrations. Because of their cationic structure they present great affinity for nucleic acids and their interaction in nucleic acids biosynthesis and *in vitro* metabolism has been shown (Tabor and Tabor, 1976). Many studies have been done to analyze this special interaction between nucleic acids and polyamines and it has been observed that polyamines can modulate nucleic acids interaction with other ligands (Pingoud et al., 1984). These polyamines are positively charged, spermidine (3+) and spermine (4+), and interact with DNA in such a way that they neutralize about 90% of its charge, leading to DNA condensation. Many cations have been studied

as condensers, but theses amines are the most studied and those that provided the clearest findings. It was shown that DNA condensation by polyamines stabilizes the DNA double helix and leads to its precipitation (Keatch et al., 2004; Pelta et al., 1996; Raspaud et al., 1998; Tabor and Tabor, 1976; Vasu et al., 2012). Besides condensing and precipitating the DNA, these amines provide a switch mechanism to regulate the activity of restriction enzymes. At low concentrations (lower than 5 mM) the cleavage by endonucleases is increased. Concentrations higher than 5 mM inhibit the activity of the endonucleases. Enzymes such as DNase, restriction endonucleases as *Eco*KI, *Eco*RI, *Kpn*I, among others, had their activity suppressed in the presence of polyamine coating. It is believed that the *in vitro* DNA coating with some polyamines mimic protection reactions *in vivo* (Conrad and Topal, 1989; Keatch et al., 2004; Kuosmanen and Pösö, 1985; Pingoud, 1985; Pingoud et al., 1984; Vasu et al., 2012).

Thus, the polyamines capacity to limit the restriction action of *C. pasteurianum* crude extract endonucleases was evaluated, with the ultimate goal of improving bacterium transformation. Efforts were concentrated on spermidine, rather than spermine, because spermidine, unlike spermine, is an essential component of many, if not all, prokaryotic organisms, its intracellular concentration in *E. coli* accounts for 4.7 pmol/g wet weight (Pingoud, 1985).

Restriction reactions with the *C. pasteurianum* crude extract and condensed plasmid pMTL007C-E2::Cpa-*spo0A*-666a by different spermidine concentrations were performed. Figure 3.3 illustrates the restriction results.

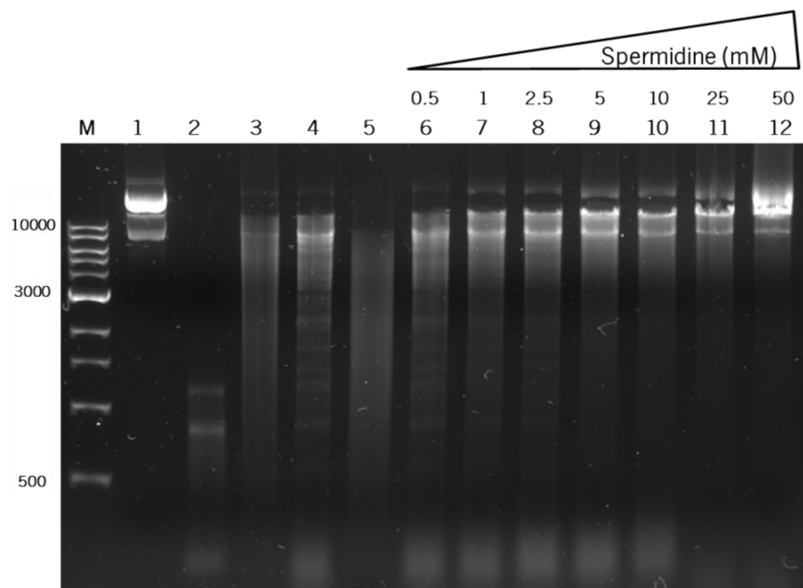


Figure 3.3. Spermidine condensation protects DNA of *C. pasteurianum* crude extract restriction. *C. pasteurianum* crude extract digestion of pMTL007C-E2::Cpa-*spo0A*-666a condensed with different spermidine concentrations. M, 1kb ladder (bp) – NEB; 1, pMTL0007C-E2::Cpa-*spo0A*-666a; 2, *C. pasteurianum* crude extract; 3, Plasmid condensed by 10 mM Spermidine with no extract; 4, *C. pasteurianum* crude extract digestion of plasmid; 5, *Hind*III linearized plasmid; 6 to 12, *C. pasteurianum* crude extract digestions of spermidine condensed plasmid, 0.5 mM, 1 mM, 2.5 mM, 5 mM, 10 mM, 25 mM and 50 mM, respectively.

As can be observed, as the spermidine concentration increases, the plasmid digestion decreases, i.e. the band pattern is reduced. The spermidine effect can also be detected in the open circular plasmid (superior) band, as spermidine concentration increases this band became more condensed and smaller running more in the gel.

Similarly to what was advocated by Pingoud in 1984, at concentrations above 5 mM the restriction prevention is more evident.

Based on these results, two transformation experiments with spermidine-condensed DNA were performed. It was expected that the DNA condensation increased the contact time between the DNA and cells, creating an opportunity for DNA entrance without its cleavage, since in case of damage (e.g. the electric pulse), the restriction system action can be alleviated (Keatch et al., 2004; Makovets et al., 2003).

The used conditions are presented on Table 3.3.

Table 3.3. *C. pasteurianum* electrotransformation parameters using spermidine-condensed DNA. In the first experiment low DNA and high spermidine concentration was used. In the second experiment a higher DNA and a lower spermidine concentration was used. Each experiment was performed in duplicate. The principal parameters changed are highlighted at bold and underlined.

Carbon source	Growth phase (O.D.$_{600}$nm)	Electroporation Buffer	Transformed cells (mL)	Plasmid (ng)	Electroporation conditions	Recuperation medium (mL)	Recovery time (h)
Glycerol 20 g/L	0.5	* 100 Mm Sodium Phosphate Buffer * 284 Mm Sucrose	300	**500 g, 1000 ng** Spermidine condensed (10 mM and 50 mM)	4 mm cuvette; 2 kV, 25 µf, 100 Ω	3	16
Glycerol 20 g/L	0.5	* 100 Mm Sodium Phosphate Buffer * 284 Mm Sucrose	300	**1000 g, 2000 ng** Spermidine condensed (5 mM and 10 mM)	4 mm cuvette; 2 kV, 25 µf, 100 Ω	3	16

Initially, low DNA (500 ng and 1000 ng) and high spermidine concentrations (10 mM and 50 mM) were used, and no positive results were achieved. This can be explained by the extremely rigid structure in which the DNA is completely entrapped and hindered to enter in contact with the cell. To overcome this possibility, a second experiment with higher DNA concentration (1000 ng and 2000 ng) and lower spermidine concentration (5 mM and 10 mM) was performed. These conditions should continue to protect the DNA while giving it more movement freedom. Unfortunately, the second experiment did not result in transformants either.

Thus, it can be suggested that the *C. pasteurianum* restriction system is extremely efficient. Most likely, when protected by the spermidine coat, the DNA is not digested, but as soon as it loses this protection it may be immediately eliminated.

3.6. *dpnII*, *dam* and *mdpn* genes as a *C. pasteurianum* R/M System

The previously obtained results directed the further work to the study of the *C. pasteurianum* R/M systems. Privileged access to this organism genome sequence (Gallardo, 2013) facilitated the identification of the genes of interest. The entire genome was evaluated for the presence of type IIP REase and MTase genes, preferentially belonging to the same R/M system, i.e. recognizing the same target sequence. In Table 3.4, the *C. pasteurianum* DSM 525 gene products expected to be related with type IIP R/M system are listed.

Table 3.4. Genes and their corresponding proteins from the *C. pasteurianum* DSM 525 genome related with type IIP R/M system.

Similar to gene	Description
ogt	methylated-DNA–protein-cysteine methyltransferase%2C constitutive
bspRIM	modification methylase *Bsp*RI
-	methyltransferase type 11 domain protein
-	methyltransferase type 12 domain protein
-	S-adenosylmethionine-dependent methyltransferase
bstV	N-6 DNA Methylase family protein
mraW	S-adenosyl-methyltransferase *Mra*W
-	type II restriction-modification enzyme domain protein
-	type IIS restriction enzyme R and M domain protein
dpnB	type-2 restriction enzyme *Dpn*II
Dam	DNA adenine methylase family protein
dpnA	modification methylase *Dpn*IIB
ycgJ	methyltransferase family protein

The proteins of greatest interest are highlighted in gray.

The genes with a more detailed functional description were preferred to the others. Four genes with the preferential characteristics were highlighted, with only three of them being ultimately selected for the current work as they constitute a type II R/M system. The selected genes can be identified as the GATC R/M system of *C. pasteurianum* (http://rebase.neb.com/rebase/rebase.html). It is constituted by one endonuclease and two methyltransferases. The endonuclease gene is similar to the gene *dpnB* that encodes the type-2 restriction enzyme *Dpn*II from *Streptococcus pneumoniae* (R.*Dpn*II – EC: 3.1.21.4). This homodimer protein recognizes the double stranded unmethylated sequence GATC, cleaving it before G nucleotide. A protein-protein BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) identifies this protein as belonging to DpnII superfamily, having 78% identity with type II restriction enzyme [Eubacterium sp. CAG:252]. The first MTase gene has sequence similarity to the gene *dam* that codes for a protein belonging to the DNA adenine methylase family. These proteins specifically methylate the amino group at the C-6 position of adenines in DNA. The dam methylation is associated to the recognition of the sequence GATC. The protein-protein BLAST performed identified it as belonging to MethyltransfD12 superfamily, presenting 58% identitity with type2 restriction enzyme MjaIII domain protein [*Clostridium difficille*]. The second MTase gene is similar to *dpnA* gene, also from *S. pneumoniae*, that encodes the modification methyltransferase *Dpn*IIB (M.*Dpn*IIB – EC: 2.1.1.72). Also a homodimer, this protein recognizes the sequence GATC and causes specific methylation on the A nucleotide on one or both DNA strands, thus protecting the DNA from cleavage by the *Dpn*II endonuclease. The protein-protein BLAST identifies this protein has belonnging to N6-N4_Mtase superfamily, presenting 66% identity with the modification methylase LlaDCH1B [*Clostridium saccharoperbutylacetonicum*].

All three enzymes genes are located in *C. pasteurianum* DSM 525 genome as a cluster, as presented in Figure 3.4, in the following order: *dpnII*, *dam* and *mdpn* genes.

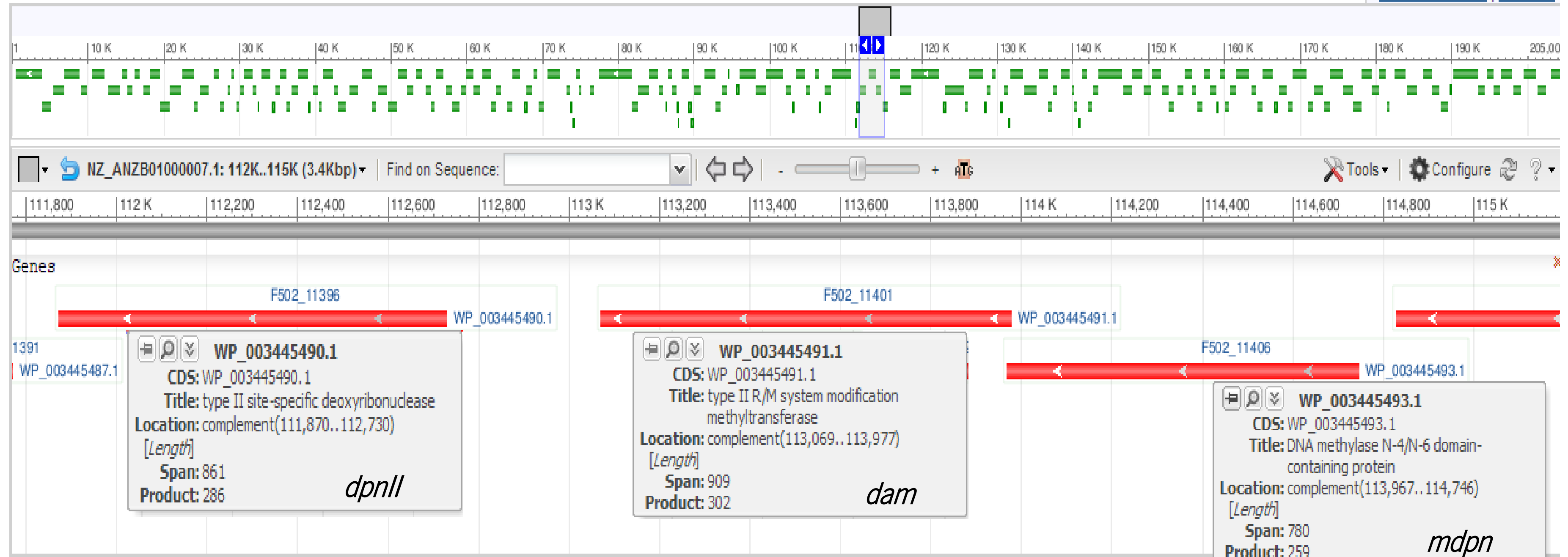


Figure 3.4. Genomic organization of the genes of GATC R/M system of *C. pasteurianum* DSM 525 are presented as cluster on this microorganism genome. Type II site-specific deoxyribonuclease is the *dpnII* gene. Type II R/M system modification methyltransferase if the *dam* gene. DNA methylase N-4/N-6 domain-containing protein is the *mdpn* gene.

Based on this information, it was decided to clone these genes in a suitable vector, produce and purify them in an *E. coli* strain, and test their activity as restriction enzyme and methyltransferases, in order to examine their role in the *C. pasteurianum* R/M system.

3.7. Cloning

3.7.1. *dpnII*, *dam* and *mdpn* genes amplification

The amplification of the genes involved in the GATC R/M system was performed using *C. pasteurianum* genomic DNA as a template, using gene-specific primers, according to the protocol described in 2.5.4. To amplify the genes from *C. pasteurianum*, different reaction volumes (25 µL - Figure 3.5 (b) and 50 µL - Figure 3.5 (a)), genomic DNA concentrations (200 ng, and the dilutions 1:10, 1:100, and 1:1000), and annealing temperatures (52°C, 55°C, and 60°C) were analyzed when using genomic DNA of 700 ng/µL.

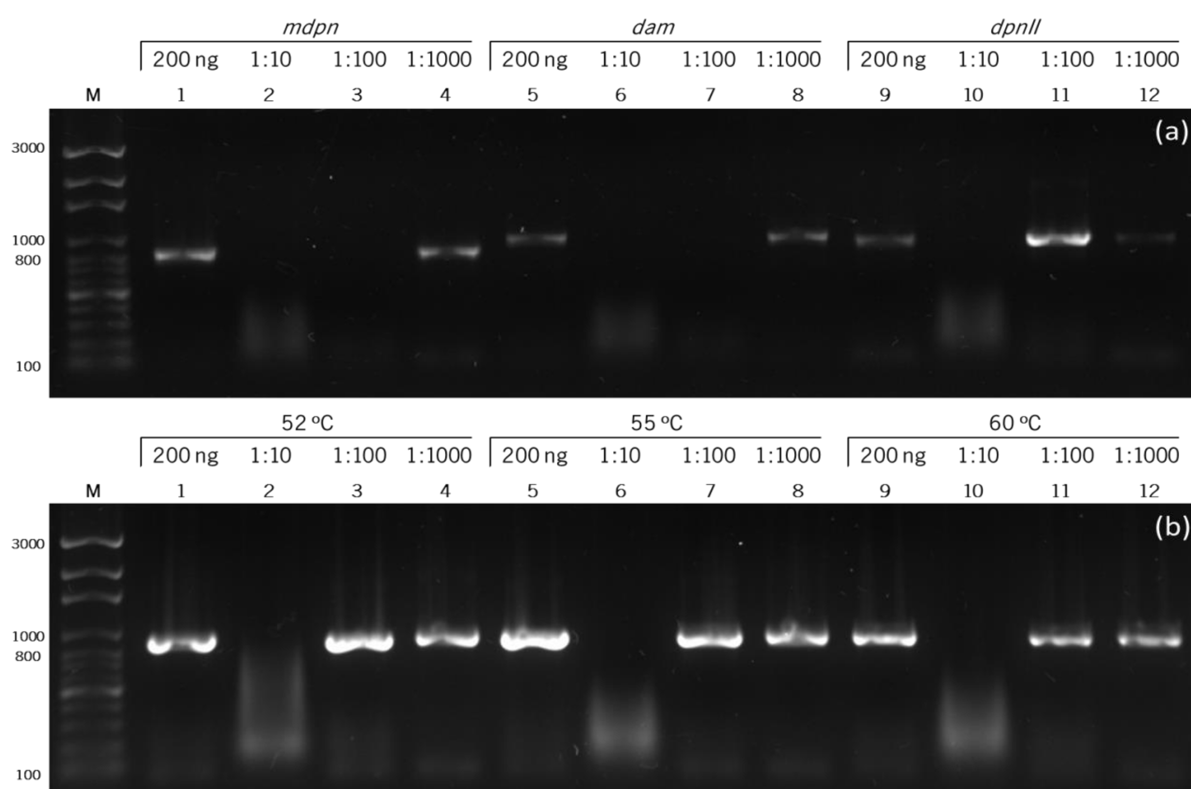


Figure 3.5. Gene amplification by PCR. (a) In 50 µL and at 52°C, genomic DNA concentrations were analyzed for the genes *mdpn*, *dam* and *dpnII*. M, 100 bp ladder (Solis); 1, 200 ng, *mdpn*; 2, 1:10, *mdpn*; 3, 1:100, *mdpn*; 4, 1:1000, *mdpn*; 5, 200 ng, *dam*; 6, 1:10, *dam*; 7, 1:100, *dam*; 8, 1:1000, *dam*; 9, 200 ng, *dpnII*; 10, 1:10, *dpnII*; 11, 1:100, *dpnII*; 12, 1:1000, *dpnII*. **(b)** In 25 µL, genomic DNA concentrations, and annealing temperatures were analyzed for the *dpnII* gene. M, 100 bp ladder (Solis); 1, 200 ng, 52°C; 2, 1:10, 52°C; 3, 1:100, 52°C; 4, 1:1000, 52°C; 5, 200 ng, 55°C; 6, 1:10, 55°C; 7, 1:100, 55°C; 8, 1:1000, 55°C; 9, 200 ng, 60°C; 10, 1:10, 60°C; 11, 1:100, 60°C; 12, 1:1000, 60°C.

As can be observed in Figure 3.5 (a), in 50 µL and at 52°C all genes were amplified (*mdpn* - 837 bp, *dam* – 966 bp and *dpnII* – 918 bp) with genomic DNA concentration of 0.7 ng/µL and 4 ng/µL except for *dpnII* that was

also amplified with 7 ng/μL. As can be observed in Figure 3.5 (b), results for PCR amplification in a reaction volume of 25 μL, were better than with 50 μL reaction volume, since the obtained bands are more intense. For the gene *dpnII*, good amplifications (918 bp) were obtained at 52°C or 55°C, being the best genomic DNA concentration between 0.1 ng/μL and 10 ng/μL. The protocol with the best conditions (25 μL of reaction volume, 52°C, and 1:100 genomic DNA) was repeated with remaining genes *dam* (966 bp) and *mdpn* (837 bp).

3.7.2. Vector – inserts construction

The selected vector to clone the *C. pasteurianum* genes was the plasmid pET24d(+). Several digestions with the selected restriction enzymes (*Nco*I and *Bam*HI) were performed but the plasmid was never correctly digested by the enzyme *Bam*HI, as can be observed in Figure 3.6 (lane 2). Numerous simultaneous and sequential double digestions of vector and inserts were performed. The addition of extra components such as Mg^{2+} and BSA and longer digestions were also attempted. The plasmid extraction procedure was also verified, changing extraction kits or conducting a 3M sodium acetate and absolute ethanol precipitation. After each digestion, the ligation and transformation processes were performed and the colonies analyzed by colony PCR using the Kapa-Taq DNA polymerase and the primers T7 promoter and T7 terminator. Nevertheless, none of these changes resulted in colonies with insertion (*data not shown*). Since there was no adequate explanation for this digestion inefficiency, the plasmid sequence was analyzed by sequencing. The sequencing results showed no differences from the original sequence, so the reason for the reported digestion inefficiency remained unknown.

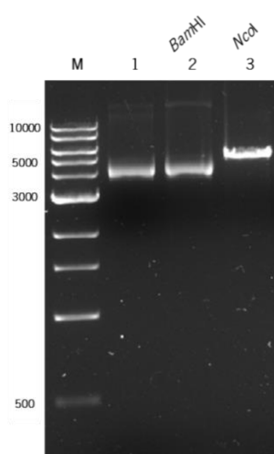


Figure 3.6. pET24d(+) enzyme digestions. Linearization's with the selected enzymes (*Bam*HI and *Nco*I) were done to verify their digestion activity in pET24d(+). M, 1kb ladder (bp) – NEB; 1, non-linearized pET24d(+); 2, *Bam*HI digested pET24d(+); 3, *Nco*I digested pET24d(+).

The inconclusive results obtained with the vector pET24d(+) incited the use of other vectors for comparison purposes, pETduet-1, 5,420 bp (Figure 3.7 (a)) and pCOLAduet-1, 3,719 bp (Figure 3.7 (b)). These plasmids have two multiple cloning sites (MCS), where the first (MCS-1) is identical to that of pET24d(+). Consequently, the same restriction enzymes can be applied, maintaining the cloning process.

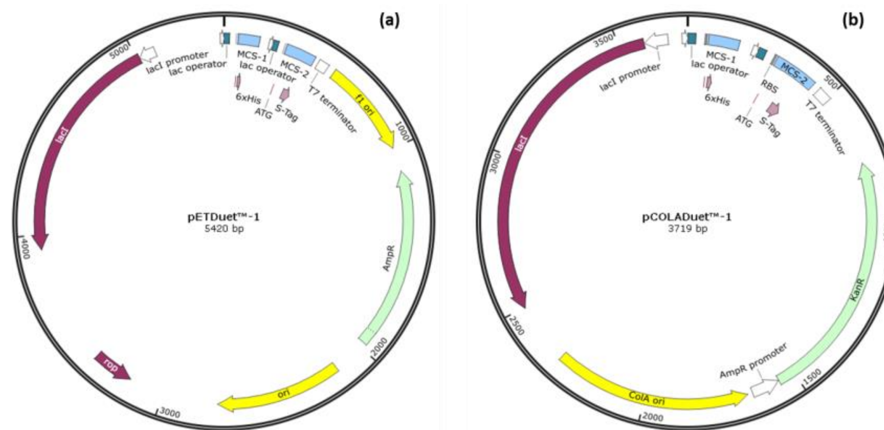


Figure 3.7. Vectors used to replace pET24d(+). (a) pETduet-1; (b) pCOLAduet-1.

A new digestion experiment was designed: the plasmids pETduet-1 and pCOLAduet-1 were sequentially digested in a 50 µL final reaction volume with extra BSA, approximately 1 µg of plasmid and 20 U of *Bam*HI. After 6 hours of individual digestion, the enzyme *Nco*I was added to the reaction and this was held overnight. The *dpnII*, *mdpn*, and *dam* genes were subjected to the same conditions. To hinder the re-attachment of the vector ends only once digested, reducing the background of wild type colonies, the enzyme antarctic phosphatase was used. This enzyme removes phosphate groups from the linearized plasmid ends that are responsible for providing linking between bases. The reaction was performed as described in 2.6.2. Only after plasmid dephosphorylation the DNA was cleaned (2.5.1) and the concentration determined to proceed to ligation (2.6.3). Throughout the entire process an empty vector was used as a negative control. If the double digestion is effective the plasmid ends cannot ligate since they are not complementary. If the digestion is 100% effective, colonies are not expected on this plate.

The transformation was made using commercial chemo-competent JM109 *E.coli* cells in a heat shock experiment (2.9.2.2). Only when the plasmid pETduet-1 was used colonies appeared on plate. Visually, the results with pETduet-1 were promising, i.e. the cells transformed with empty vector presented lower colonies number than the cells transformed with the ligations products. To confirm this result a colony PCR was carried out.

A first colony PCR was done using the T7 promoter and T7 terminator primers, but the results were not as expected, as the insertions were not amplified (*data not shown*). In a second assay, the forward primer of each gene and the T7 terminator were used. The use of these primers allows the confirmation of gene correct insertion into the vector, since the first is gene specific and the second vector specific. Three colonies with the empty vector and 10 colonies with each gene insertion were picked and subjected to colony PCR conditions (2.5.5), although this time the super-efficient Kapa-Hifi polymerase was used, along with the adequate conditions (2.5.4). The results are presented on Figure 3.8.

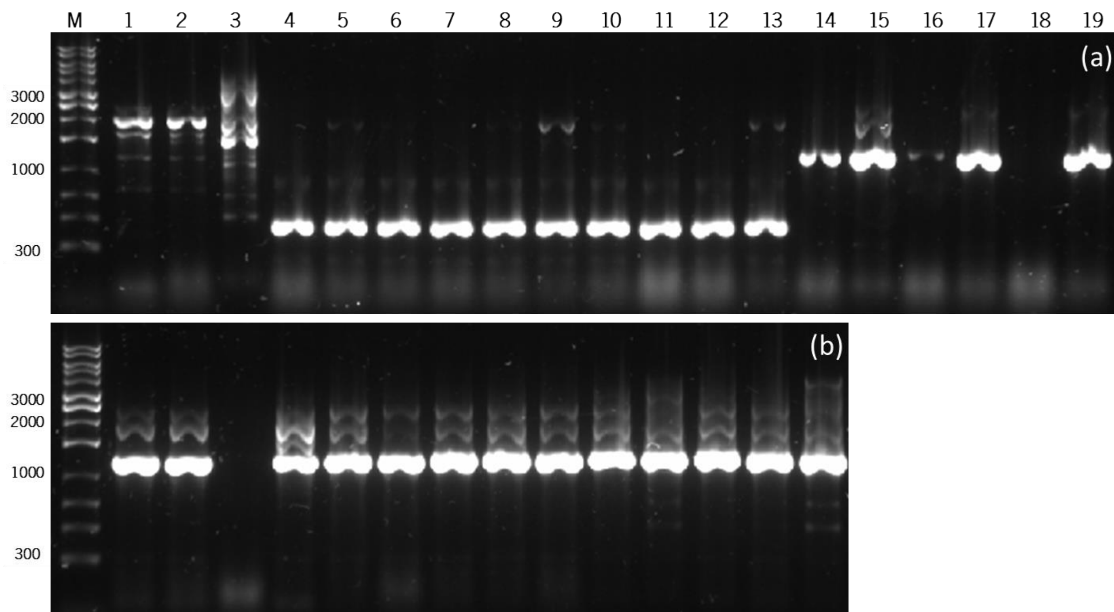


Figure 3.8. Colony PCR results from the transformation of JM109 *E. coli* with the vector pETduet-1 ligated to genes *dam*, *dpnII* or *mdpn*. The primers forward of each gene and T7 terminator were used with an annealing temperature of 52°C. **(a)** M, 1 kb ladder (bp) – Solis; 1-3, colonies transformed with the empty vector; 4-13, colonies transformed with pETduet-*mdpn*; 14-19, colonies transformed with pETduet-*dpnII*; **(b)** M, 1 kb ladder (bp) – Solis; 1-4 colonies transformed with pETduet-*dpnII*; 5-14, colonies transformed with pETduet-*dam*.

The expected amplifications have 1,197 bp (*mdpn*), 1,277 bp (*dpnII*) and 1,326 bp (*dam*). These amplification lengths result from the distance between each gene beginning (*Nco*I restriction site) and the T7 terminator primer complementary sequence (localized after the second MCS). Analyzing the agarose gel image (Figure 3.8) we can observe the correct insert amplification in almost all colonies with the gene *dpnII*, and all colonies with the *dam* genes as they present the expected length. In contrast, the *mdpn* gene was not correctly amplified. Some non-specific PCR products were detected in the control lanes (1-3), and in a most evident form in the colonies transformed with *dam* gene. These are most likely the result of complementarity between the gene-specific primers and the vector.

Gene insertion was subsequently confirmed by the extraction and digestion of the plasmids that presented a positive result in colony PCR. The plasmids were linearized with the enzyme *Nco*I (Figure 3.9 (a)) and their length confirmed by comparison with the predicted value (Figure 3.9 (b)). Plasmids linearization was performed as indicated in 2.8.1. The new constructed vectors were denominated as pCM001 (pETduet-1-*mdpn*), pCM002 (pETduet-1-*dpnII*), and pCM003 (pETduet-1-*dam*).

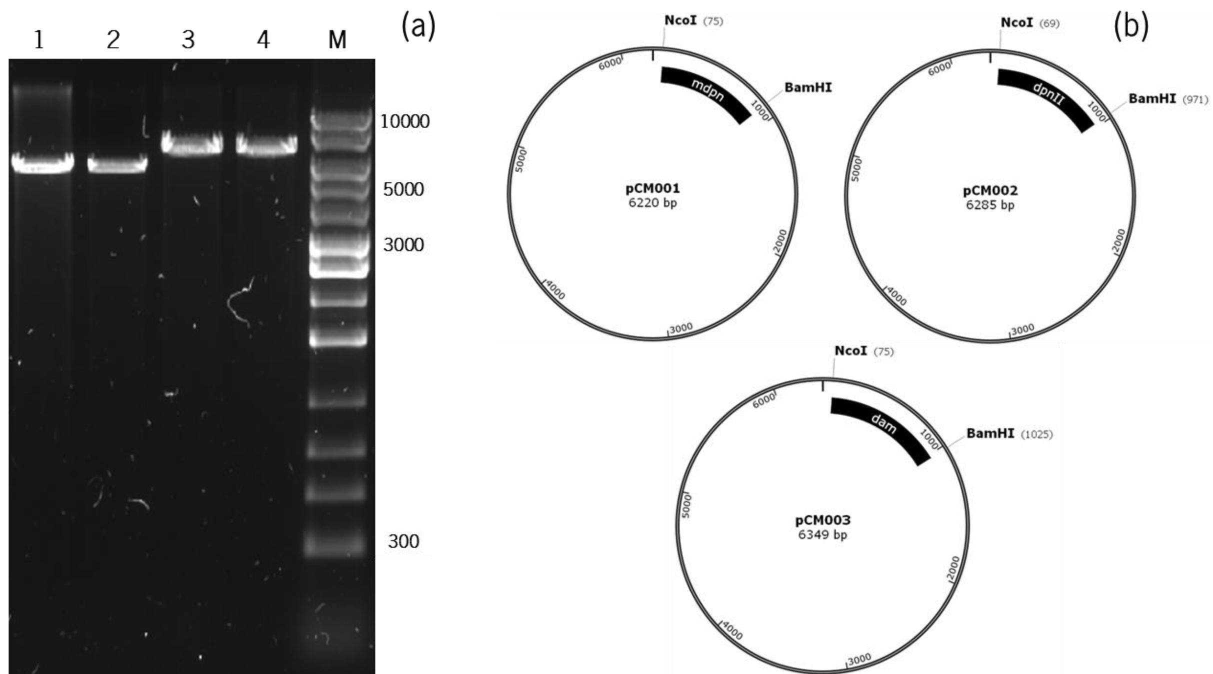


Figure 3.9. Length analysis of plasmids extracted from the cells transformed with the ligation pETduet-1 and the genes *mdpn* (pCM001), *dpnII* (pCM002) and *dam* (pCM003). (a) Plasmid restriction analysis with the enzyme *NcoI* at 37°C for 1.5 h. 1, pEtduet-1; 2, pCM001; 3, pCM002; 4, pCM003; M, 1kb ladder (bp) – Solis. **(b)** Constructs with the plasmid pETduet-1 and the genes *mdpn* (pCM001), *dpnII* (pCM002) and *dam* (pCM003).

In the Figure 3.9 (a) the differences between the empty vector and the inserted vectors length can be observed. The insertion of the genes *dpnII* (6,285 bp) and *dam* (6,349 bp) is evident since their bands are above 6000 bp. The *mdpn* gene was not inserted, the corresponding band size is lower than the inserted vectors and of the same size that the empty vector. These restriction tests confirmed the absence of *mdpn* gene amplification in the colony PCR. With regard to the *mdpn* – pETduet-1 insertion another cloning process was performed, exactly as the previously mentioned and this time positive insertions were obtained (*data not shown*). After the cloning step, the constructs were sequenced to determine if the nucleotide sequence was correct, i.e. do not present any difference when compared with the predicted construction (2.5.6).

3.8. Protein production

After sequencing confirmation, constructs were inserted into a high level protein expression *E. coli* strain (2.9.2), to DpnII (*dpnII*), Dam (*dam*), and Mdpn (*mdpn*) enzymes production.

The molecular weight of the proteins was determined, according to their amino acid constitution plus the histidine tail (His-tag), using the computational tool ExPASy, the Bioinformatics Resource Portal of SIB – Swiss Institute of Bioinformatics (Artimo et al., 2012). The calculated masses of 33,939.18 Da (daltons), 35,965.86 Da, and 31,821.71 Da correspond to the proteins DpnII, Dam and Mdpn, respectively.

3.8.1. BL21 (DE3) production

The common *E. coli* strain used as host for protein production is the strain BL21 (DE3). As mentioned in Materials and Methods, 2.9.2, this strain lacks specific proteases (lon protease and the OmpT outer membrane protease) responsible for preventing heterologous protein production. The protein expression was then performed as described in 2.10.1. A total extract (TE) and a cell free extract (CFE) (protein in soluble form) samples were analyzed at different times of induction, 0 h, 1 h, 2 h, 4 h and overnight (o/n). In Figure 3.10 the results of the CFE of the enzymes Dam (Figure 3.10 (a)) and DpnII (Figure 3.10 (b)) production are presented. The TE samples containing all the proteins present in the cell generated a highly dense gel, where bands can hardly be seen. As such, although the TE gels have been analyzed they are not presented herein.

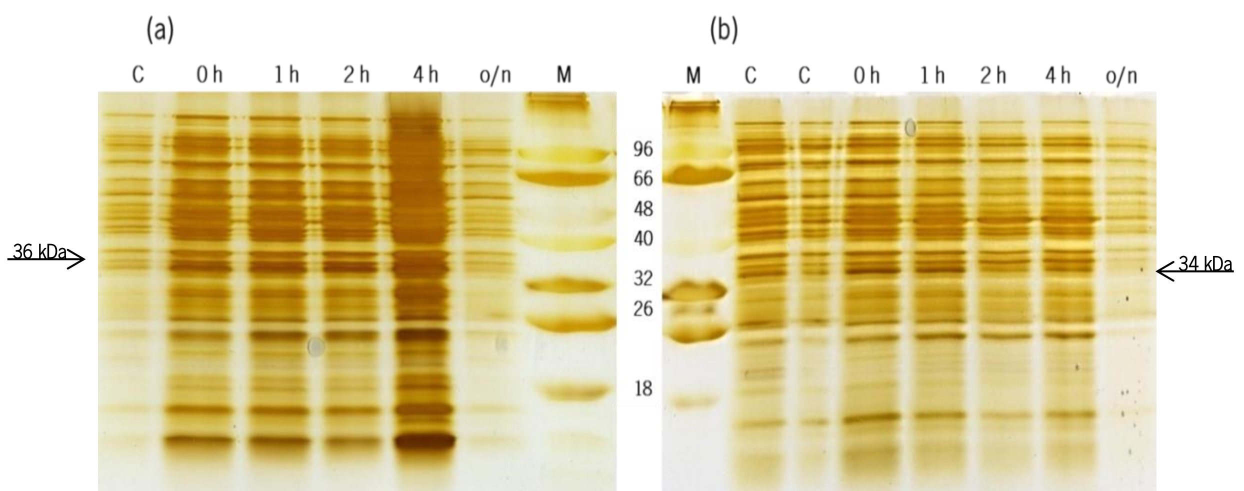


Figure 3.10. SDS-PAGE analysis of the CFE after BL21 (DE3) production of the enzymes Dam (a) and DpnII (b) at 37°C. M. LMW, kDa, C-control, empty pETduet o/n production, the samples were collected at 0 h, 1h, 2h, 4h and o/n. Gels were stained by silver staining. Arrows indicate the approximate molecular mass expected to the produced proteins.

The enzymes bands should be located between 32 and 40 kDa (DpnII may have 34 kDa, Dam may have 36 kDa). However, as can be observed in both images, when comparing the control with the induction times, no additional bands are presented, leading to the conclusion that the enzymes were not overproduced. The intensity difference observed in the control (C - a) can be explained by an inefficient cell rupture procedure. In the o/n procedure, this intensity difference may be due to cell death and consequent protein degradation, thus leading to lower protein production levels.

3.8.2. Improvement of the conditions for protein production

3.8.2.1. Low temperatures of protein production

In order to improve the protein production, the basic temperature for protein production, 37°C, was lowered to 16°C after IPTG induction. Contrary to the high protein aggregation observed at high rates, low temperatures can influence heterologous proteins solubility and folding, improving the protein yield (Vera et al., 2007). In this case,

lowering the temperature did not result in protein production, since no different results for the cells with the empty vector could be observed (6III, Figure 6.1).

3.8.2.2. Codon Usage Analysis

The bacterium *E. coli* is the most popular host for heterologous protein production as it is a simple and extensively studied microorganism with an easy cultivation process. Its genetics is well known and a lot of compatible tools are available that facilitate its usage. However, heterologous protein production can be complicated when genes display a non-canonical codon usage. The genetic code is composed by 64 codons, encoding 20 amino acids and 3 termination codons. This means that more than one codon codes for the same amino acid (except for methionine – AUG and tryptophan – UGG), meaning that different nucleotide sequences can code for the same protein sequence. The codon usage is the predilection shown by organisms to use synonymous codons to code the amino acids, each kind of organism has their own favorable codons. *Codon bias* is the organism (or genes) preference by a certain codon to encode an amino acid (Neme and Miramontes, 2005). Some studies have demonstrated that codon predilection reduces the isoacceptor tRNAs diversity, i.e. the tRNAs of the preferred codons are present in higher quantity contrary to the non-preferred codons (Andersson and Kurland, 1991). Currently, the reason of such predilection is still not clear, but it has become evident that codon bias affects recombinant protein production in organisms with a different codon preference (Gustafsson et al., 2004; Rosano and Ceccarelli, 2009).

As a universal host, the favorite *versus* non-favorite codons of the bacterium *E. coli* were investigated. The non-preferred codons are classified as low-usage and rare codons. Low usage codons are rarely or infrequently used, and rare codons are rarely or infrequently used but are also decoded by low abundant tRNAs. Chen and Texada (2006) studied the low and rare codons of *E. coli*, and of 30 low usage codons, 20 are rare codons. Seven codons, AGG, AGA, CGA, CUA, AUA, CCC and CGG (group I) are used with a frequency of <0.5 %, and thirteen, ACA, CCU, UCA, GGA, AGU, UCG, CCA, UCC, GGG, CUC, CUU, UCU and UUA (group II) are used with a frequency of >0.5%. This research group also demonstrated that all the rare codons from group I and the first 6 of group II are responsible for the translational problems observed in *E. coli* (Chen and Texada, 2006).

The *C. pasteurianum* genes under study were analyzed for the presence of *E. coli* rare codons and all of the previously presented rare codons were found to be present in the genes of this bacterium, some of them being the most commonly used codons to encode an amino acid. The GenScript Rare Codon Analysis Tool (http://www.genscript.com/cgi-bin/tools/rare_codon_analysis) was used to determine the distribution of codon usage frequency for each gene under analysis. This bioinformatic tool calculates the Codon Adaptation Index (CAI) that measures the deviation of a given protein coding gene sequences with respect to a reference set of genes, in this case, *E. coli*. CAI values range from 0 to 1; a CAI of 1.0 is considered to be ideal, while a CAI above 0.8 is rated as good for expression in the desired expression organism. The CAI values obtained for the genes studied in the current work were: 0.60 for *mdpn*, 0.67 for *dpnII*, and 0.61 for *dam*, i.e. CAI values much lower than the ones

recommended for heterologous protein expression in *E. coli*. The production of clostridial proteins in *E. coli* have been shown problematic because the different codon usage caused by a high AT content (Zdanovsky and Zdanovskaia, 2000).

3.8.2.3. BL21 (DE3) RIL and Rosetta (DE3) pLysS production

To overcome this codon usage difference between *C. pasteurianum* and *E. coli*, the vectors pCM001, pCM002 and pCM003 were inserted into two *E. coli* strains that supply extra tRNA copies, BL21 (DE3) RIL and Rosetta (DE3) pLysS. BL21 (DE3) RIL contains the copies AGA and AGG for arginine, AUA for isoleucine, and CUA for leucine. Rosetta (DE3) pLysS contains the same previous tRNA copies plus CCC for proline and GGA for glycine. The cells were made competent and transformed as described in 2.9.2, using chloramphenicol and ampicillin as selection agents for both strains. The production was performed as mentioned in 2.10.1. After induction with 1 mM IPTG the cells were maintained at 37◦C, and as previously mentioned, samples were collected at 0 h, 1 h, 2 h, 4 h and o/n. From these cells CFE samples of 0 h, 4.5 h and o/n production were simultaneously analyzed. An empty vector was used as negative control. No protein production was obtained with Rosetta (DE3) pLysS (the results are presented in 6III, Figure 6.2). The results of BL21 (DE3) RIL are presented in Figure 3.11.

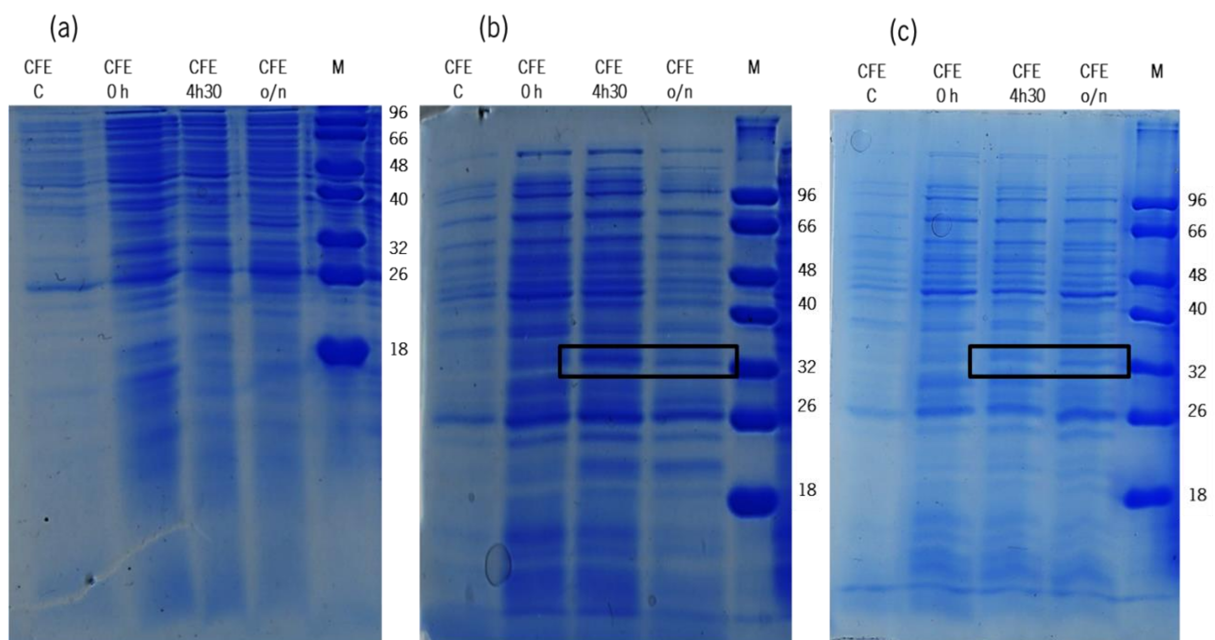


Figure 3.11. SDS-PAGE analysis of BL21 (DE3) RIL CFE production of the enzymes Dam (a) DpnII (b) and Mdpn (c) at 37◦C. M. LMW, kDa; C-control, empty pETduet-1 o/n production; the samples analyzed were of 0 h, 4h30 and overnight (o/n). The possible protein productions are highlighted by a black rectangle. All gels were stained by coomassie blue staining.

A possible production of enzymes DpnII (b) and Mdpn (c) was observed. In CFE 4.5 h and CFE o/n lanes of (b) and (c) images, some prominent bands, compared to the remaining commonly proteins produced by BL21 (DE3) RIL, were detected (highlighted by a black rectangle). At 26 kDa the enzyme chloramphenicol acetyltransferase is present (25,660 Da), providing chloramphenicol resistance.

To verify production, a His-purification of 4.5 h sample was done, since the enzymes have a histidine tail (His-tag) that establishes affinity to nickel. These experiments were performed in nickel columns, as mentioned in 2.10.4, and the results are presented in Figure 3.12.

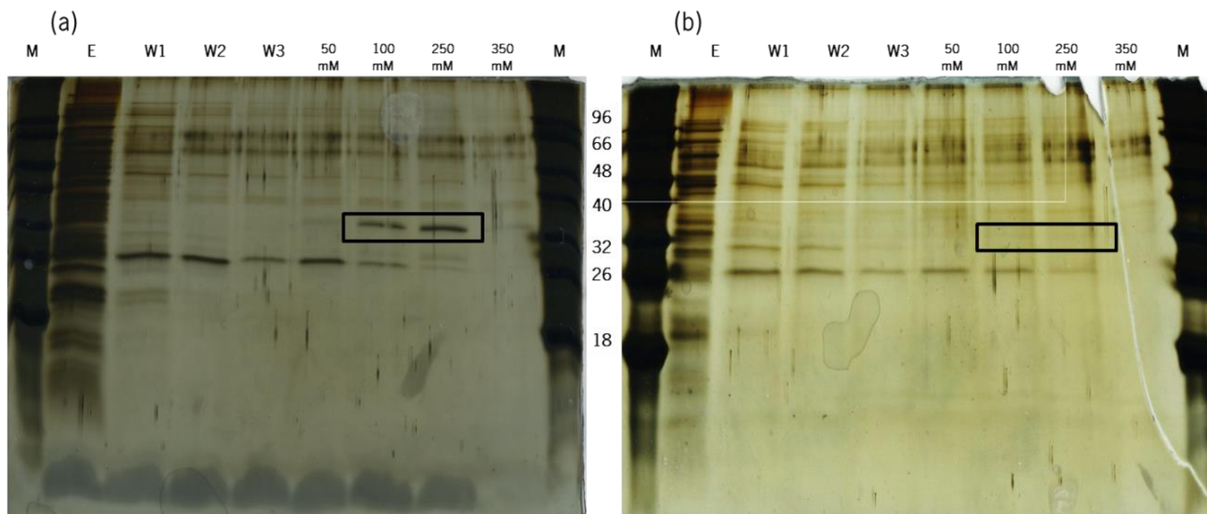


Figure 3.12. SDS-PAGE gel of His-tag purified CFE of the BL21 (DE3) RIL DpnII (a) and Mdpn (b) 4.5 h sample production. M. LMW, kDa; E, sample in equilibration buffer; W1-W3 wash 1 to 3; 50 mM imidazole elution; 100 mM imidazole elution; 250 mM imidazole elution; 350 mM imidazole elution. The possible protein productions are highlighted by a black rectangle. Gels were stained by coomassie blue and silver staining.

As can be observed, the purification of the overproduced enzymes was not 100% effective, since many bands are still appearing after the washing steps. This can be explained by column saturation, the total protein extract used to pass through each column (2 mL) may be excessive (each column has a 200 µL capacity). Bands with a molecular weight equal to the overproduced enzyme of interest eluted at every imidazole concentration. In Figure 3.12 (a), purification of DpnII (33,9 kDa), bands slightly above 32 kDa are present. This may be indicative of the protein presence, i.e. the protein may have been produced in this strain, although in low quantities. In Figure 3.12 (b), purification of Mdpn (31,8 kDa), hardly any visible bands were detected.

The low protein production levels observed for Mdpn incited us to use Western blotting techniques to identify its presence using the primary Anti-His Antibody. This highly specific technique has a minimal detection of 10 pg (picograms) of protein. The experiment was performed as mentioned in 2.10.5, and the results are presented in Figure 3.13. The 4.5 h Mdpn sample was purified in nickel columns, run on an SDS-PAGE gel, followed by Western blotting. As a positive control an 18 kDa protein was used with confirmed histidine tail presence.

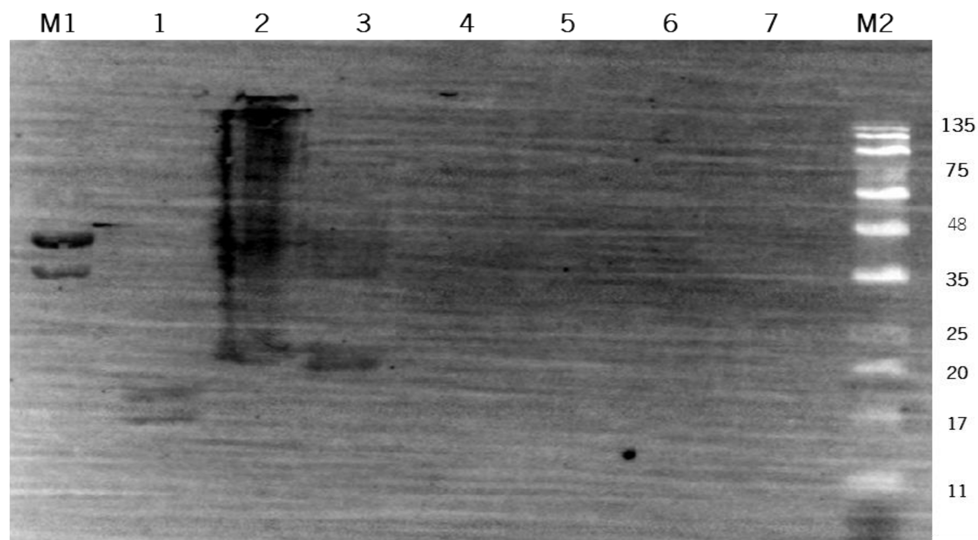


Figure 3.13. Western blotting result of the 4.5 h sample of Mdpn production in BL21 (DE3) RIL. M1, LMW, kDa; 1, positive control (aprox 18 kDa); 2, CFE without purification; 3, sample in equilibration buffer after column; 4, 10th wash; 5, 100 mM imidazole elution; 6, 250 mM imidazole elution; 7, 350 mM imidazole elution; M2, NZYColour Protein Marker II (kDa) – Nzytech.

The Western blotting technique was successful, as the positive control (lane 1) was detected. Some proteolysis appears to be present, since two bands were detected. With the CFE sample (lane 2), some reaction with the antibody was obtained, but it can be an unspecific reaction since a great amount of proteins is present. In lane 3 the same explanation can be considered, as the high protein amount is still maintained. After the washing step (also included), no specific binding is detected. If the His-tag was entrapped inside the protein it will not be in contact with the nickel in the column and therefore will not be purified. Or, the previously SDS-PAGE highlighted proteins may be only a common *E. coli* protein produced in higher quantities at that moment.

3.8.2.4. Enzymatic experiments

If the proteins are being produced, even in low quantities, they may have sufficient activity that can be tested by performing digestion reactions. Since only the GATC REase DpnII showed significant overexpression (Figure 3.12 a), only this enzyme´s activity was examined. The reactions were performed in 10 µL final volume with 100 ng to 200 ng of pMTL007C-E2::Cpa-*spo0A*-666a, in commercial 1x *Dpn*II buffer and 1 µL of *DpnII* production crude extract at 37°C for 2 hours. The obtained results (*data not shown*) were inconclusive relative to REase activity, since the DNA was methylated by a normal dam methylation process present in *E. coli* strains as JM109. This kind of methylation hypothetically hinders DpnII restriction.

3.8.2.5. Origami (DE3) production

In two of the three genes being studied more than one cysteine residue was detected (*mdpn* – 7 and *dam* – 4). These residues are important for the protein conformation since most are commonly covalently linked to other cysteine residues to form a disulfide bond. Origami (DE3) is an *E. coli* strain with mutations in both thioredoxin

reductase (*trxB*) and glutathione reductase (*gor*) genes, greatly enhancing the disulfide bonds and facilitating protein folding. The cells were made competent and transformed as described in 2.9.2 using kanamycin (15 µg/µL), tetracycline (12.5 µg/µL) and ampicillin (40 µg/µL) as selector agents. Although cysteine residues are present, the linkage between them does not seem to be the problem, as no different results were obtained for the cells with the empty vector (6III, Figure 6.3).

3.9. *dcm* gene importance in *C. pasteurianum* DSM 525 R/M system

On April 2013, Pyne and coworkers (2013) reported a transformation protocol for *C. pasteurianum* ATCC 6013 (DSM 525). They were successful with an *in vivo* methylation approach using the *Fnu*DII methyltransferase gene into pFnuDIIM (Lunnen et al., 1988). *In vitro* approaches were effective against the crude extract action but were not useful in the transformation, i.e. transformants were not obtained when DNA was methylated *in vitro*. The MTase *Fnu*DII, recognizes the sequence 5'-CGCG- 3', the same sequence recognized and cleaved by the REase *Cpa*AI, R.*Fnu*DII isoschizomer, found in 1988 by Richards and collaborators (1988) in *C. pasteurianum* ATCC 6013. This enzyme seems, different from what was expected, to be the major barrier for DNA entry in this strain of *C. pasteurianum* (Richards et al., 1988).

This report, published in the course of this thesis (Pyne et al., 2013), led to some modifications of the priorities initially defined in the work plan. The publication of the *C. pasteurianum* whole genome (Rappert et al., 2013) was also a helpful tool, since other bioinformatic tools, as REBASE, could use the provided data in a more comprehensive form. Based on this, we searched for other *C. pasteurianum* methyltransferases that can be possibly involved in the bacterium enzymatic restriction system. Our sequencing results (Gallardo, 2013), the published genome information (Rappert et al., 2013) and the REBASE enzyme analysis were combined in order to find one CG methyltransferase in *C. pasteurianum* genome, capable of specific foreign DNA methylation. The enzyme gene closest to M.*Fnu*DII and *Cpa*AI was the gene of DNA-methyltransferase *Dcm* (GenBank: ELP61257.1), similar to the modification methyltransferase *Bsp*RI (*dcm* gene – 1131 bp). A protein-protein BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) identifies this protein as belonging to Cyt_C5_DNA_Methylase superfamily, presenting 54% identity with DNA-methyltransferase Dcm [*Clostridium* sp. BNL1100].

Following the identification of the gene, it was decided to have its codon usage optimized (see section 3.8.2.2) by IDT- Integrated DNA Technologies, Inc (Belgium). A comparison of the conventional and optimized version is shown in 6IV, Figure 6.4. Four primers were designed for sequence amplification, these are presented in 6V, Table 6.4. Due to complications in the synthesis process by IDT, further cloning and expression experiments could not be included in this work.

3.10. *C. pasteurianum* DSM 525 REase identification

Meanwhile, experiments with the commercial REases *Dpn*II (5'-|GATC- 3') and *Bst*UI (5'-CG|CG- 3'), the ^{m5}CG methyltransferase M.*Sss*I, and the *C. pasteurianum* crude extract were performed to verify their activity in the

plasmid pMTL007C-E2::Cpa-*spo0A*-666a. The enzyme *Dpn*II has the same restriction activity predicted for the enzyme DpnII from the GATC R/M system of *C. pasteurianum*. The enzyme *Bst*UI has the same restriction site of *Cpa*AI. The m5CG methyltransferase, M.*Sss*I, methylates the cytosine residue of the sequence CG, so every sequence with these two nucleotides followed will be methylated and, hypothetically, protected from restriction enzymes. Reactions with the restriction enzymes and the crude extract were performed with methylated and non-methylated (M.*Sss*I) plasmid. The results are illustrated in Figure 3.14.

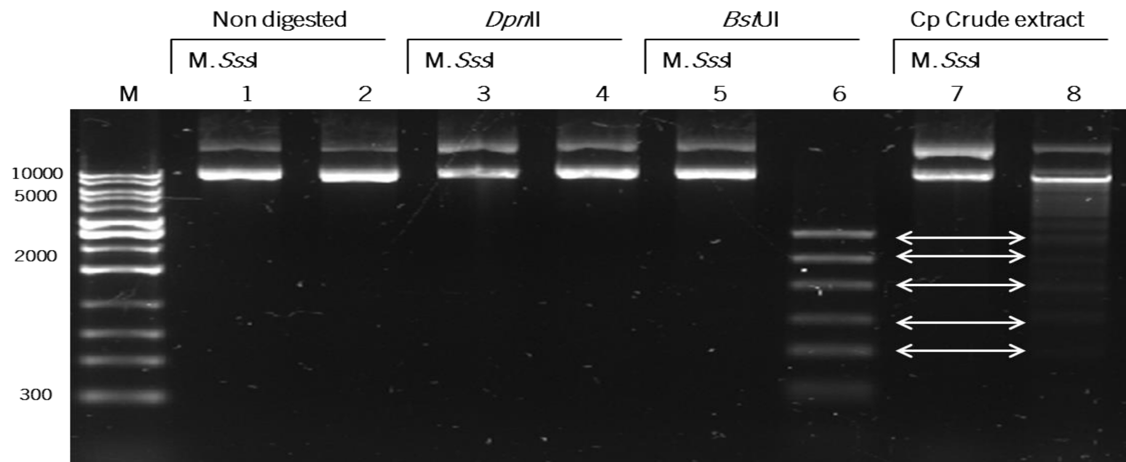


Figure 3.14. Restriction reaction of m5CG methylated and non-methylated pMTL007C-E2::Cpa-*spo0A*-666a with the enzymes *Dpn*II, *Bst*UI and *C. pasteurianum* crude extract. M, 1kb ladder (bp) – Solis; 1, M.*Sss*I plasmid non digested; 2, plasmid non digested; 3, M.*Sss*I plasmid reaction with *Dpn*II; 4, plasmid reaction with *Dpn*II; 5, M.*Sss*I plasmid reaction with *Bst*UI; 6, plasmid reaction with *Bst*UI; 7, M.*Sss*I plasmid reaction with *C. pasteurianum* (Cp) crude extract; 8, plasmid reaction with *C. pasteurianum* crude extract. Similar bands are compared by the white arrows.

With the commercial enzyme *Dpn*II (lanes 3 and 4) the plasmid should be cleaved into 14 fragments, even in M.*Sss*I methylated DNA, since this enzyme does not protect the sequence recognized by this REase. However, some *E. coli* strains have native methylation - dam methylation, methylating the sequence 5'-GATC- 3'. The JM109 *E. coli* strain, in which the plasmid pMTL007C-E2::Cpa-*spo0A*-666a was replicated, is an example of this kind of bacteria with natural methylation. Thus, the protected plasmid was not cleaved by this enzyme. The digestion of this plasmid replicated in a mutated *E. coli* strain (*E. coli* K12 ER2925), without this natural methylation, was performed, but the results were inconclusive (*data not shown*).

When non-methylated plasmid was used (lane 6), the enzyme *Bst*UI digested it. The plasmid has fourteen restriction sites for this enzyme, with only the bands of 2,532 bp, 1,778 bp, 1,289 bp, 969 bp, 582 bp, 581 bp, 360 bp, 336 bp, and 295 bp being visible. When methylated DNA was used (lane 5) the restriction was hindered, as expected.

The reaction with the crude extract (lanes 7 and 8) had an interesting result, the non-methylated DNA (lane 8) was digested as previously proved, but the m5CG methylated plasmid (lane 7) was entirely protected from the restriction action of *C. pasteurianum* crude extract. The protection of the sequences with the dinucleotide CG means

that the restriction sequence recognized by the enzyme responsible for the plasmid cleavage recognizes a sequence with this dinucleotide.

Analyzing the digestion pattern of the plasmid digested by *Bst*UI and crude extract bands of very similar length are present. The digestion pattern resulting from crude extract reaction is less profound, maybe because total extract was used, in contrast to pure *Bst*UI enzyme.

The similarities between these digestion patterns were further analyzed. Figure 3.15 presents the results.

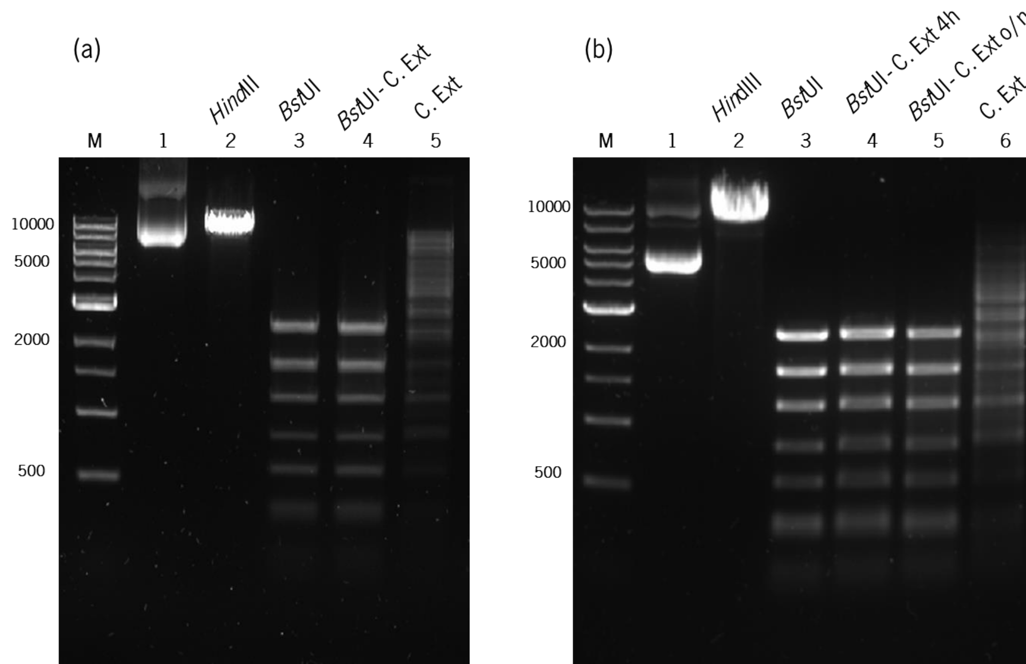


Figure 3.15. Restriction activity comparison between *C. pasteurianum* (Cp) crude extract (C. Ext) and the enzyme *Bst*UI. (a) Restriction comparison between *C. pasteurianum* crude extract and the enzyme *Bst*UI, for 2 h. M, 1kb ladder (bp) – NEB; 1, pMTL007C-E2::Cpa-*spoOA*-666a non digested; 2, linearized plasmid (*Hind*III); 3, plasmid with *Bst*UI; 4, plasmid with *Bst*UI and *C. pasteurianum* crude extract; 5, plasmid with *C. pasteurianum* crude extract. **(b)** New reaction comparing the restriction activity between *C. pasteurianum* crude extract and the enzyme *Bst*UI. M, 1kb ladder (bp) – NEB; 1, pMTL007C-E2::Cpa-*spoOA*-666a non digested; 2, linearized plasmid (*Hind*III); 3, plasmid with *Bst*UI (2 h); 4, plasmid with *Bst*UI (2 h) and *C. pasteurianum* crude extract (2 h); 5, plasmid with *Bst*UI (2 h) and *C. pasteurianum* crude extract (o/n); 6, plasmid with *C. pasteurianum* crude extract.

In Figure 3.15 (a), individual and simultaneous reactions of the enzyme *Bst*UI and crude extract are presented. The reaction was performed at 37°C (not the optimal temperature indicated by the manufacturers, but it remains active) in the designated enzyme buffer (crude extract also presents activity in these conditions, lane 5). In lane 4 the simultaneous reaction of *Bst*UI and crude extract can be observed, while the digestion pattern observed is the same presented by the enzyme alone.

A second experiment was done (b) to prove these results. The plasmid was digested for 2 hours by the *Bst*UI (lane 3), after two hours the crude extract with fresh reaction buffer was added and the samples were collected 2 h (lane 4) later or overnight (5).

These results lead us to assume that the restriction resulting from *C. pasteurianum* DSM 525 crude extract is performed, apparently, by just one REase, since even after an overnight reaction no other enzyme possibly present on *C. pasteurianum* crude extract presented activity differentiating the digestion pattern. Or, if other enzymes are present, their restriction sites are not present on this vector and need to be tested with other vectors with different DNA sequences.

Concluding, *C. pasteurianum* DSM 525 crude extract seems to has just one REase responsible for foreign DNA cleavage and has the same restriction sequence of the commercial *Bst*UI 5'-CGCG- 3'. To perform a successful *C. pasteurianum* DSM 525 transformation, the focal point should be this enzyme's restriction activity.

Chapter 4

Concluding Remarks and Future Perspectives

4. CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The transformation of *C. pasteurianum*, a bacterium with a high resistance to transformation, was the main goal of this thesis. Genetic engineering of this solventogenic microorganism has become highly relevant due to its ability to metabolize crude glycerol to produce valuable chemicals as ethanol and butanol that can be successfully used as biofuels. Various strategies were discussed, but specific methylation of foreign DNA prior to transformation turned out to be an essential step to achieve the proposed goal.

Our results, combined with the recent publication of Pyne *et al.*, allowed us to identify the restriction modification system that *C. pasteurianum* DSM 525 most likely uses as a defense mechanism.

The work of Pyne and their co-workers reported not only the insertion of specific methylated DNA into *C. pasteurianum* with the *Fnu*DII methyltransferase, but also the improvement of the electro-transformation physical conditions. To a high-level electro-transformation, osmotic stabilizers such as sucrose, a membrane solubilization with ethanol, and the changes of pulse parameters that increase the exposure time were required. Therefore, these results suggested that the simple electro-transformation protocols used in this thesis could, if specific methylated DNA was used, lead to transformation; however the insertion of this group optimized electroporation conditions could enhance the transformation yield.

M.*Msp*I methylates the sequence 5'-CCGG- 3' on the first cytosine residue and *C. pasteurianum* REase recognizes and cleaves the sequence 5'-CGCG- 3', which may justify the inability of M.*Msp*I to protect pMTL007C-E2::Cpa-*spo0A*-666a from *C. pasteurianum* restriction system.

The spermidine DNA condensation was very effective in protecting the DNA from the REase present in *C. pasteurianum* crude extract. However, when the protection is lost, the enzyme most likely cleaves the foreign DNA. In the future this event could be analyzed more in detail, i.e. to discover what happens to this spermidine-condensed DNA when in contact with the cell.

GATC R/M system genes were cloned in pETduet-1 (pCM001, pCM002 and pCM003) and transformed in the common expression host *E. coli* BL21 (DE3), however no protein expression could be obtained since the codon usage of *C. pasteurianum* is not the same as the one used by *E. coli*. Transformations in BL21 (DE3) RIL and Rosetta (DE3) pLysS, *E. coli* strains with extra tRNAs corresponding to rare codons in *E. coli* were performed. Protein production was detected, however the results were not conclusive and require confirmation.

Protein production in *E.coli* Origami (DE3) was tested, in order to benefit protein folding, due to the establishment of disulfide bonds; however no difference was observed, without detection of protein production. The codon usage may have been influencing the recombinant protein production. Although the number of available tRNAs has been increased with strains BL21 (DE3) RIL and Rosetta (DE3) pLysS, this may not have been sufficient for the production to be improved. The best chance to correctly produce these enzymes would be when their codon usage is optimized for expression in *E. coli*.

Indeed, the production of this protein may have occurred at a very small scale. If this happened, enzymatic experiments with REase could be performed with the protein production extracts. However, these experiments have to be carried out with DNA replicated in a Dam- *E. coli* strain, like *E. coli* K12 ER 2925. If the predicted restriction site DpnII were correct (5'-GATC- 3'), then its activity would be suppressed by dam methylation, naturally performed on *E. coli* strains, not confirming the enzyme functionality.

Results presented by Pyne and co-workers showed that choosing the GATC R/M system was not the most advantageous option, since this system may have restriction and methylation activity, however the DpnII is not the principal defense system of *C. pasteurianum* against foreign DNA. If active, the methyltransferases would not protect the plasmid from the *C. pasteurianum* restriction system.

A new and more specific research on *C. pasteurianum* MTase genes was performed in order to find one CG methyltransferase capable of specific foreign DNA methylation. The enzyme gene closest to M.*Fnu*DII and *Cpa*AI was the gene of DNA-methyltransferase *Dcm* (GenBank: ELP61257.1), similar to the modification methyltransferase *Bsp*RI (*dcm* gene – 1131 bp, (Gallardo, 2013)). The *in silico* study of this gene was done, together with codon improvement for expression in *E. coli*. In the near future, the gene will be cloned and overexpressed in a pET-*E.coli* expression system. After production the activity on plasmid pMTL007C-E2::Cpa-*spo0A*-666a compared to *C. pasteurianum* crude extract will be examined. Using a functional enzyme, an *in vivo* methylation system will be developed; the enzyme will be produced within the host to specifically methylate the DNA with which *C. pasteurianum* is transformed. DNA methylated by this specific methyltransferase is expected to be easily inserted into *C. pasteurianum*. As a result, a huge number of opportunities for the application of this organism will be open.

A set of experiments with the enzymes M.*Sss*I (^{m5}CG), *Bst*UI (5'-CG|CG- 3') and *C. pasteurianum* crude extract demonstrated that this microorganism's principal defense system is an REase recognizing the sequence 5'-CGCG- 3'. It is therefore expected that the CG methyltransferase previously selected will be capable of methylating this sequence, thus hindering the action of the native REase.

In this work, more insight was given on the origin of *C. pasteurianum* DSM 525*'s* recalcitrance to transformation, the knowledge about R/M systems in *C. pasteurianum* DSM 525 was extended, and a proposal to efficiently transform this bacterium was done.

References

5. REFERENCES

Ahn, J.-H., Sang, B.-I., and Um, Y. (2011). Butanol production from thin stillage using *Clostridium pasteurianum.* Bioresour. Technol. *102*, 4934–4937.

Andersson, G., and Kurland, C. (1991). An extreme codon preference strategy: codon reassignment. Mol. Biol. Evol. *8*, 530–544.

Arber, W. (1965). Host specificity of DNA produced by *Escherichia coli* V. The role of Methionine in the production of host specificity. J. Mol. Biol. *11*, 247–256.

Arber, W. (2000). Genetic variation: molecular mechanisms and impact on microbial evolution. FEMS Microbiol. Rev. *24*, 1–7.

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., et al. (2012). ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. *40*, W597–603.

Aune, T.E.V., and Aachmann, F.L. (2010). Methodologies to increase the transformation efficiencies and the range of bacteria that can be transformed. Appl. Microbiol. Biotechnol. *85*, 1301–1313.

Azeddoug, H., and Reysset, G. (1991). Recognition sequence of a new methyl-specific restriction system from *Clostridium acetobutylicum* strain ABKn8. FEMS Microbiol. Lett. *78*, 153–156.

Azeddoug, H., Hubert, J., and Reysset, G. (1989). Characterization of a methyl-specific restriction system in *Clostridium acetobutylicum* strain N1-4081. FEMS Microbiol. Lett. *65*, 323–326.

Bertani, G., and Weigle, J. (1953). Host controlled variation in bacterial viruses. J. Bacteriol. *65*, 113–121.

Bickle, T. a, and Krüger, D.H. (1993). Biology of DNA restriction. Microbiol. Rev. *57*, 434–450.

Biebl, H. (2001). Fermentation of glycerol by *Clostridium pasteurianum*—batch and continuous culture studies. J. Ind. Microbiol. Biotechnol. *27*, 18–26.

Biebl, H., Marten, S., Hippe, H., and Deckwer, W. (1992). Glycerol conversion to 1, 3-propanediol by newly isolated clostridia. Appl. Microbiol. Biotechnol. *36*, 592–597.

Bio-Rad-Catalogue (2000). MicroPulserTM Electroporation Apparatus Operating Instructions and Applications Guide.

Birrer, G., Chesbro, W., and Zsigray, R. (1994). Electro-transformation of *Clostridium beijerinckii* NRRL B-592 with shuttle plasmid pHR106 and recombinant derivatives. Appl. Microbiol. Biotechnol. *41*, 32–38.

Chen, D., and Texada, D. (2006). Low-usage codons and rare codons of *Escherichia coli*. Gene Ther. Mol. Biol. *10*, 1–12.

Chen, I., and Dubnau, D. (2004). DNA uptake during bacterial transformation. Nat. Rev. Microbiol. *2*, 241–249.

Chen, C., Boucle, C., and Blaschek, H. (1996). Factors involved in the transformation of previously non-transformable *Clostridium perfringens* type B. FEMS Microbiol. Lett. *140*, 185–191.

Chen, C., Smye, S.W., Robinson, M.P., and Evans, J. a (2006). Membrane electroporation theories: a review. Med. Biol. Eng. Comput. *44*, 5–14.

Choi, K., and Yoo, O. (1990). A new restriction endonuclease, *Cth*II, from *C. thermocellum* ATCC 27405. Korean Biochem. J. *23*, 418–421.

Choi, K.D., Kim, K., and Yoo, O.J. (1987). A new restriction endonuclease from *Clostridium thermocellum*. Sanop Misaengmul Hakhoe Chi *15*, 352–355.

Conrad, M., and Topal, M.D. (1989). DNA and spermidine provide a switch mechanism to regulate the activity of restriction enzyme Nae I. Proc. Natl. Acad. Sci. U. S. A. *86*, 9707–9711.

Cooksley, C.M., Zhang, Y., Wang, H., Redl, S., Winzer, K., and Minton, N.P. (2012). Targeted mutagenesis of the *Clostridium acetobutylicum* acetone-butanol-ethanol fermentation pathway. Metab. Eng. *14*, 630–641.

Cui, G., Hong, W., Zhang, J., Li, W., Feng, Y., Liu, Y., and Cui, Q. (2012). Targeted gene engineering in *Clostridium cellulolyticum* H10 without methylation. J. Microbiol. Methods *89*, 201–208.

Dabrock, B., Bahl, H., and Gottschalk, G. (1992). Parameters Affecting Solvent Production by *Clostridium pasteurianum*. Appl. Environ. Microbiol. *58*, 1233–1239.

Daldal, F. (1985). Isolation of mutants of *Clostridium pasteurianum*. Arch. Microbiol. *142*, 93– 96.

Davis, T.O., Henderson, I., Brehm, J.K., and Minton, N.P. (2000). Development of a Transformation and Gene Reporter System for Group II , Non-Proteolytic *Clostridium botulinum* Type B Strains. *2*, 59–69.

Dong, H., Zhang, Y., Dai, Z., and Li, Y. (2010). Engineering clostridium strain to accept unmethylated DNA. PLoS One *5*, e9038.

Dürre, P. (2007). Clostridia. In ENCYCLOPEDIA OF LIFE SCIENCES, J. Willey, and Sons, eds. pp. 1–8.

Dürre, P. (2008). Fermentative butanol production: bulk chemical and biofuel. Ann. N. Y. Acad. Sci. *1125*, 353–362.

Gallardo, R. (2013). New strategies for the production of butanol and 1,3-propanediol from glycerol. PhD in Bioengineering (MIT-Portugal Program). University of Minho.

González-Pajuelo, M., Meynial-Salles, I., Mendes, F., Andrade, J.C., Vasconcelos, I., and Soucaille, P. (2005). Metabolic engineering of *Clostridium acetobutylicum* for the industrial production of 1, 3-propanediol from glycerol. Metab. Eng. *7*, 329–336.

Gray, K. a, Zhao, L., and Emptage, M. (2006). Bioethanol. Curr. Opin. Chem. Biol. *10*, 141–146.

Griffith, F. (1928). The significance of pneumococcal types. J. Hyg. (Lond). *27*, 113–159.

Gustafsson, C., Govindarajan, S., and Minshull, J. (2004). Codon bias and heterologous protein expression. Trends Biotechnol. *22*, 346–353.

Hanahan, D., and Bloom, F.R. (1996). Mechanisms of DNA transformation. In *Escherichia Coli* and *Salmonella* — Cellular and Molecular Biology, F.C. Neidhardt, ed. (Washington, D.C.: American Society for Microbiology), pp. 2449–2459.

Heap, J.T., Pennington, O.J., Cartman, S.T., Carter, G.P., and Minton, N.P. (2007). The ClosTron: a universal gene knock-out system for the genus *Clostridium.* J. Microbiol. Methods *70*, 452–464.

Heap, J.T., Pennington, O.J., Cartman, S.T., and Minton, N.P. (2009). A modular system for *Clostridium* shuttle plasmids. J. Microbiol. Methods *78*, 79–85.

Heap, J.T., Kuehne, S. a, Ehsaan, M., Cartman, S.T., Cooksley, C.M., Scott, J.C., and Minton, N.P. (2010). The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. J. Microbiol. Methods *80*, 49–55.

Herbert, M., O'Keeffe, T. a., Purdy, D., Elmore, M., and Minton, N.P. (2003). Gene transfer into *Clostridium difficile* CD630 and characterisation of its methylase genes. FEMS Microbiol. Lett. *229*, 103–110.

Heyndrickx, M., Vos, P. De, and Ley, J. De (1991). Fermentation char act e r i st ics of *Clostridium pasteurianum* LMG 3285 grown on glucose and mannitol. 52–58.

Ito, T., Nakashimada, Y., Senba, K., Matsui, T., and Nishio, N. (2005). Hydrogen and ethanol production from glycerol-containing wastes discharged after biodiesel manufacturing process. J. Biosci. Bioeng. *100*, 260–265.

Jarvis, G.N., Moore, E.R.B., and Thiele, J.H. (1997). Formate and ethanol are the major products of glycerol fermentation produced by a *Klebsiella planticola* strain isolated from red deer. J. Appl. Microbiol. *83*, 166–174.

Jeltsch, A. (2003). Maintenance of species identity and controlling speciation of bacteria: a new function for restriction/modification systems? Gene *317*, 13–16.

Jennert, K., Tardif, C., Young, D., and Young, M. (2000). Gene transfer to *Clostridium cellulolyticum* ATCC 35319. Microbiology *146*, 3071–3080.

Jensen, T.O., Kvist, T., Mikkelsen, M.J., Christensen, P.V., and Westermann, P. (2012a). Fermentation of crude glycerol from biodiesel production by *Clostridium pasteurianum*. J. Ind. Microbiol. Biotechnol. *39*, 709–717.

Jensen, T.O., Kvist, T., Mikkelsen, M.J., and Westermann, P. (2012b). Production of 1,3-PDO and butanol by a mutant strain of *Clostridium pasteurianum* with increased tolerance towards crude glycerol. AMB Express *2*, 44.

Johnsborg, O., Eldholm, V., and Håvarstein, L.S. (2007). Natural genetic transformation: prevalence, mechanisms and function. Res. Microbiol. *158*, 767–778.

Johnson, D., and Taconi, K. (2007). The glycerin glut: Options for the value-added conversion of crude glycerol resulting from biodiesel production. Environ. Prog. *26*, 338–348.

Kamm, B., and Kamm, M. (2007). Biorefineries–multi product processes. Adv Biochem Engin/Biotechnol *105*, 175–204.

Kanematsu, T., and Ozawa, O. (1991). New restriction enzyme *Cco*I having good pH and temp. stability - recognises base sequence of double chain DNA and cuts as specified position. (Japan).

Karberg, M., Guo, H., Zhong, J., Coon, R., Perutka, J., and Lambowitz, a M. (2001). Group II introns as controllable gene targeting vectors for genetic manipulation of bacteria. Nat. Biotechnol. *19*, 1162–1167.

Keatch, S. a, Su, T.-J., and Dryden, D.T.F. (2004). Alleviation of restriction by DNA condensation and non-specific DNA binding ligands. Nucleic Acids Res. *32*, 5841–5850.

Khanna, S., Jaiswal, S., Goyal, A., and Moholkar, V.S. (2012). Ultrasound enhanced bioconversion of glycerol by *Clostridium pasteurianum*: A mechanistic investigation. Chem. Eng. J. *200-202*, 416–425.

Kuehne, S., and Minton, N. (2012). ClosTron-mediated engineering of *Clostridium*. Bioengineered *3*, 247–254.

Kuehne, S.A., Heap, J.T., Cooksley, C.M., Cartman, S.T., and Minton, N.P. (2011). Clostron-mediated Engineering of *Clostridium*. In Strain Engineering Methods and Protocols, J.A. Williams, ed. (Totowa, NJ: Humana Press), pp. 389–407.

Kuit, W., Minton, N.P., López-contreras, A.M., and Eggink, G. (2012). Disruption of the acetate kinase (ack) gene of *Clostridium acetobutylicum* results in delayed acetate production. Appl. Microbiol Biotechnol. *94*, 729–741.

Kuosmanen, M., and Pösö, H. (1985). Inhibition of the activity of restriction endonucleases by spermidine and spermine. FEBS Lett. *179*, 17–20.

Labrie, S.J., Samson, J.E., and Moineau, S. (2010). Bacteriophage resistance mechanisms. Nat. Rev. Microbiol. *8*, 317–327.

Laemmli, U.K. (1970). Cleavage of Structural proteins during the assenbly of the head of Bacteriophage T4.pdf. Nature *227*, 680–685.

Lambowitz, A.M., and Zimmerly, S. (2004). Mobile group II introns. Annu. Rev. Genet. *38*, 1–35.

Lee, P.C., Lee, W.G., Lee, S.Y., and Chang, H.N. (2001). Succinic acid production with reduced by-product formation in the fermentation of *Anaerobiospirillum succiniciproducens* using glycerol as a carbon source. Biotechnol. Bioeng. *72*, 41–48.

Lee, S., Park, J., Jang, S., Nielsen, L., Kim, J., and Jung, K. (2008). Fermentative butanol production by clostridia. Biotechnol. Bioeng. *101*, 209–228.

Lehmann, D., Radomski, N., and Lütke-Eversloh, T. (2012a). New insights into the butyric acid metabolism of *Clostridium acetobutylicum.* Appl. Microbiol. Biotechnol. *96*, 1325–1339.

Lehmann, D., Hönicke, D., Ehrenreich, A., Schmidt, M., Weuster-Botz, D., Bahl, H., and Lütke-Eversloh, T. (2012b). Modifying the product pattern of *Clostridium acetobutylicum*: physiological effects of disrupting the acetate and acetone formation pathways. Appl. Microbiol. Biotechnol. *94*, 743–754.

Lin, Y.L., and Blaschek, H.P. (1984). Transformation of Heat-Treated *Clostridium acetobutylicum* Protoplasts with pUB110 Plasmid DNA. Appl. Environ. Microbiol. *48*, 737–742.

Liou, J.S.-C., Balkwill, D.L., Drake, G.R., and Tanner, R.S. (2005). *Clostridium carboxidivorans* sp. nov., a solvent-producing clostridium isolated from an agricultural settling lagoon, and reclassification of the acetogen *Clostridium scatologenes* strain SL1 as *Clostridium drakei* sp. nov. Int. J. Syst. Evol. Microbiol. *55*, 2085–2091.

Lunnen, K.D., Barsomian, J.M., Camp, R.R., Card, C.O., Chen, S.Z., Croft, R., Looney, M.C., Meda, M.M., Moran, L.S., and Nwankwo, D.O. (1988). Cloning type-II restriction and modification genes. Gene *74*, 25–32.

Luria, S., and Human, M. (1952). A nonhereditary, host-induced variation of bacterial viruses. J. Bacteriol. *64*, 557–569.

Makovets, S., Powell, L.M., Titheradge, A.J.B., Blakely, G.W., and Murray, N.E. (2003). Is modification sufficient to protect a bacterial chromosome from a resident restriction endonuclease? Mol. Microbiol. *51*, 135–147.

Malaviya, A., Jang, Y.-S., and Lee, S.Y. (2012). Continuous butanol production with reduced byproducts formation from glycerol by a hyper producing mutant of *Clostridium pasteurianum*. Appl. Microbiol. Biotechnol. *93*, 1485–1494.

Marchetti, J.M., Miguel, V.U., and Errazu, A.F. (2007). Possible methods for biodiesel production. Renew. Sustain. Energy Rev. *11*, 1300–1311.

Mermelstein, L., and Papoutsakis, E. (1993). In vivo methylation in *Escherichia coli* by the *Bacillus subtilis* phage Φ3T I methyltransferase to protect plasmids from restriction upon transformation of *Clostridium*. Appl. Environ. Microbiol *59*, 1077–1081.

Mermelstein, L., Welker, N., Bennett, G., and Papoutsakis, E. (1992). Expression of cloned homologous fermentative genes in *Clostridium acetobutylicum* ATCC 824. Biotechnology *10*, 190–195.

Murakami, M., Ozawa, O., Kanematsu, T., and Yamada, Y. (1991). Characterization of restriction endonuclease *CbI*I, an isoschizomer of *Asu*II, from *Clostridium bifermentans* strain B-4. Nucleic Acids Res. *19*, 3458.

Nakas, J.P., Schaedle, M., Parkinson, C.M., Coonley, C.E., and Tanenbaum, S.W. (1983). System development for linked-fermentation production of solvents from algal biomass. Appl. Environ. Microbiol. *46*, 1017–1023.

Nakotte, S., Schaffer, S., Böhringer, M., and Dürre, P. (1998). Electroporation of, plasmid isolation from and plasmid conservation in *Clostridium acetobutylicum* DSM 792. Appl. Microbiol. Biotechnol. *50*, 564–567.

Neme, A., and Miramontes, P. (2005). Biological domain identification based in codon usage by means of rule and tree induction. In Computational Methods in Systems Biology, pp. 221–224.

Neumann, E., Schaefer-Ridder, M., Wang, Y., and Hofschneider, P. (1982). Gene transfer into mouse lyoma cells by electroporation in high electric fields. EMBO … *1*, 841–845.

NG, T.K., Ben-Bassat, A., and Zeikus, J.G. (1981). Ethanol Production by Thermophilic Bacteria: Fermentation of Cellulosic Substrates by Cocultures of *Clostridium thermocellum* and *Clostridium thermohydrosulfuricum*. Appl. Environ. Microbiol. *41*, 1337–1343.

Nikolajewa, S., Beyer, A., Friedel, M., Hollunder, J., and Wilhelm, T. (2005). Common patterns in type II restriction enzyme binding sites. Nucleic Acids Res. *33*, 2726–2733.

Novagen (2006). pETSystem Manual.

Olson, D.G., and Lynd, L.R. (2012). Transformation of *Clostridium thermocellum* by electroporation. (Elsevier Inc.).

Oultram, J., Loughlin, M., Swinfield, T.-J., Brehm, J., Thompson, D., and Minton, N. (1988). Introduction of plasmids into whole cells of *Clostridium acetobutylicum* by electroporation. FEMS Microbiol. Lett. *56*, 83–88.

Pelta, J., Livolant, F., and Sikorav, J.L. (1996). DNA aggregation induced by polyamines and cobalthexamine. J. Biol. Chem. *271*, 5656–5662.

Pingoud, A. (1985). Spermidine increases the accuracy of type II restriction endonucleases. Eur. J. Biochem. *109*, 105–109.

Pingoud, A., Urbanke, C., Alves, J., Ehbrecht, H.J., Zabeau, M., and Gualerzi, C. (1984). Effect of polyamines and basic proteins on cleavage of DNA by restriction endonucleases. Biochemistry *23*, 5697–5703.

Pingoud, A., Fuxreiter, M., Pingoud, V., and Wende, W. (2005). Type II restriction endonucleases: structure and mechanism. Cell. Mol. Life Sci. *62*, 685–707.

Prescott, L.M., Harley, J.P., and Klein, D.A. (2005). Microbiology: Recombinant DNA technology (The McGraw-Hill Companies, Inc).

Purdy, D., O'Keeffe, T. a T., Elmore, M., Herbert, M., McLeod, A., Bokori-Brown, M., Ostrowski, A., and Minton, N.P. (2002). Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier. Mol. Microbiol. *46*, 439–452.

Pyne, M.E., Moo-Young, M., Chung, D. a, and Chou, C.P. (2013). Development of an electrotransformation protocol for genetic manipulation of *Clostridium pasteurianum*. Biotechnol. Biofuels *6*, 50.

Rappert, S., Song, L., Sabra, W., Wang, W., and Zeng, A.-P. (2013). Draft genome sequence of type strain *Clostridium pasteurianum* DSM 525 (ATCC 6013), a promising producer of chemicals and fuels. Genome Announc. *1*, e00232–12.

Raspaud, E., Olvera de la Cruz, M., Sikorav, J.L., and Livolant, F. (1998). Precipitation of DNA by polyamines: a polyelectrolyte behavior. Biophys. J. *74*, 381–393.

Reid, S.J., Allcock, E.R., Jones, D.T., and Woods, D.R. (1983). Transformation of *Clostridium acetobutylicum* Protoplasts with Bacteriophage DNA. Appl. Environ. Microbiol. *45*, 305–307.

Richards, D., Linnett, P., Oultram, J., and Young, M. (1988). Restriction endonucleases in *Clostridium pasteurianum* ATCC 6013 and *C. thermohydrosulfuricum* DSM 568. J. Gen. Microbiol. *134*, 3151–3157.

Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A., Bickle, T., Bitinaite, J., Blumenthal, R., Degtyarev, S., Dryden, D., Dybvig, K., et al. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. Nucleic Acids Res. *31*, 1805–1812.

Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2007). REBASE–enzymes and genes for DNA restriction and modification. Nucleic Acids Res. *35*, D269–D270.

Rosano, G.L., and Ceccarelli, E. a (2009). Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted *Escherichia coli* strain. Microb. Cell Fact. *8*, 41.

Sakuragi, H., Kuroda, K., and Ueda, M. (2011). Molecular breeding of advanced microorganisms for biofuel production. J. Biomed. Biotechnol. *2011*, 416931.

Schoenfeld, T., Mead, D., and Fiandt, M. (1989). Purification and characterization of an isoschizomer of *Asu*II from *Clostridium sporogenes.* Nucleic Acids Res. *17*, 4417.

Shao, L., Hu, S., Yang, Y., Gu, Y., Chen, J., Yang, Y., and Jiang, W. (2007). Targeted gene disruption by use of a group II intron (targetron) vector in *Clostridium acetobutylicum*. 963–965.

Da Silva, G.P., Mack, M., and Contiero, J. (2009). Glycerol: a promising and abundant carbon source for industrial microbiology. Biotechnol. Adv. *27*, 30–39.

Stackebrandt, E. (2004). The phylogeny and classification of anaerobic bacteria. In Strict and Facultative Anaerobes. Medical and Environmental Aspects, P. Nakano, M. M. Zuber, ed. (Wymondham: Horizon Bioscience), pp. 1–25.

Suzuki, H. (2012). Host-Mimicking Strategies in DNA Methylation for Improved Bacterial Transformation. In Methylation - From DNA, RNA and Histones to Diseases and Treatment, A. Dricu, ed. (InTech, Chapters), pp. 219–236.

Szostková, M., and Horáková, D. (1998). The effect of plasmid DNA sizes and other factors on electrotransformation of *Escherichia coli* JM109. Bioelectrochemistry Bioenerg. *47*, 319–323.

Tabor, C.W., and Tabor, H. (1976). 1,4-Diaminobutane (putrescine), spermidine, and spermine. Annu. Rev. Biochem. *45*, 285–306.

Taconi, K.A., Venkataramanan, K.P., and Johnson, D.T. (2009). Growth and Solvent Production by *Clostridium pasteurianum* TM ATCC 6013 Utilizing Biodiesel-Derived Crude Glycerol as the Sole Carbon Source. Environ. Prog. Sustain. Energy *28*, 100–110.

Tardif, C., Maamar, H., Balfin, M., and Belaich, J. (2001). Electrotransformation studies in *Clostridium cellulolyticum*. J. Ind. Microbiol. Biotechnol. *27*, 271–274.

Tarek, M. (2005). Membrane electroporation: a molecular dynamics simulation. Biophys. J. *88*, 4045–4053.

Di Tommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., Taly, J.-F., and Notredame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. Nucleic Acids Res. *39*, W13–17.

Tyurin, M., Padda, R., Huang, K., Wardwell, S., Caprette, D., and Bennett, G.N. (2000). Electrotransformation of *Clostridium acetobutylicum* ATCC 824 using high-voltage radio frequency modulated square pulses. 220–227.

Tyurin, M., Desai, S., and Lynd, L. (2004). Electrotransformation of *Clostridium thermocellum*. Appl. Environ. Microbiol. *70*, 883–890.

Umer, B., Good, D., Anné, J., Duan, W., and Wei, M.Q. (2012). Clostridial spores for cancer therapy: targeting solid tumour microenvironment. J. Toxicol. *2012*.

Vasconcelos, I., Girbal, L., and Soucaille, P. (1994). Regulation of carbon and electron flow in *Clostridium acetobutylicum* grown in chemostat culture at neutral pH on mixtures of glucose and glycerol. J. Bacteriol. *176*, 1443–1450.

Vasu, K., Nagamalleswari, E., and Nagaraja, V. (2012). Promiscuous restriction is a cellular defense strategy that confers fitness advantage to bacteria. Proc. Natl. Acad. Sci. U. S. A. *109*, E1287–93.

Venkataramanan, K.P., Boatman, J.J., Kurniawan, Y., Taconi, K. a, Bothun, G.D., and Scholz, C. (2012). Impact of impurities in biodiesel-derived crude glycerol on the fermentation by *Clostridium pasteurianum* ATCC 6013. Appl. Microbiol. Biotechnol. *93*, 1325–1335.

Vera, A., González-Montalbán, N., Arís, A., and Villaverde, A. (2007). The conformational quality of insoluble recombinant proteins is enhanced at low growth temperatures. Biotechnol. Bioeng. *96*, 1101–1106.

Wen, T.N., Tung, P.H., and Chen, C.S. (1989). Purification and some properties of a restriction endonuclease *CcyI* from *Clostridium cylindrosporum.* Bot. Bull. Acad. Sin. *30*, 91–98.

Weston, A., Brown, M.G., Perkins, H.R., Saunders, J.R., and Humphreys, G.O. (1981). Transformation of *Escherichia coli* with plasmid deoxyribonucleic acid: calcium-induced binding of deoxyribonucleic acid to whole cells and to isolated membrane fractions. J. Bacteriol. *145*, 780–787.

Yang, F., Hanna, M. a, and Sun, R. (2012). Value-added uses for crude glycerol–a byproduct of biodiesel production. Biotechnol. Biofuels *5*, 13.

Yazdani, S.S., and Gonzalez, R. (2007). Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry. Curr. Opin. Biotechnol. *18*, 213–219.

Young, D.I., Evans, V.J., Jefferies, J.R., Jennert, K.C.B., Phillips, Z.E. V, Ravagnani, A., and Young, M. (1999). 6 Genetic Methods in Clostridia. In Methods in Microbiology, Genetic Methods for Diverse Prokaryotes, Margaret C M Smith, and R.E.S.B. T, eds. (Academic Press), pp. 191–207.

Zdanovsky, A., and Zdanovskaia, M. (2000). Simple and efficient method for heterologous expression of clostridial proteins. Appl. Environ. Microbiol. *66*, 3166–3173.

Zhao, H., Msadek, T., Zapf, J., and Madhusudan, H. (2002). JA and Varughese, KI (2002) DNA complexed structure of the key transcription factor initiating development in sporulating bacteria. Structure *10*, 1041–1050.

Zhong, J., Karberg, M., and Lambowitz, A.M. (2003). Targeted and random bacterial gene disruption using a group II intron (targetron) vector containing a retrotransposition-activated selectable marker. Nucleic Acids Res. *31*, 1656–1664.

Appendix

6. APPENDIX

I. Media and solutions recipe

Table 6.1. Recipe for mRCM media. The reagents were dissolved in distilled water.

Reagent	Amount (g/L)
Beef extract	10
Peptone	10
Sodium Chloride	5
Glucose/Glycerol	20/20 or 40
Yeast Extract	3
Sodium Acetate	3
Soluble Starch	1
L-cysteine HCl.H_2O	0.56

Table 6.2. Recipe for T68 basal media. The reagents were dissolved in distilled water.

Reagent	Amount (g/L)
Potassium dihydrogen phosphate	0.5
Ammonium acetate	2
Magnesium sulfate heptahydrate	0.3
Iron sulfate heptahydrate	0.01
Cysteine-HCl	0.5
Biotin	0.001
Thiamin	0.001
p-aminobenzoic acid	0.001
Glucose 1% (w/v) (or Glycerol)	10

Table 6.3. Acrylamide/Bis-acrylamide gel constitution. The recipe is performed to 2 gels.

	Stacking Gel	Resolving Gel
Reagent	4 %	13 %
1.5 M or Tris-HCl pH 8.3	—	2.8 mL
0.5 M Tris-HCl pH 6.8	2.5 mL	—
40 % acryl/bis-acrylamide	1.0 mL	3.72 mL
Ultrapure water	6.35 mL	4.34 mL
10 % SDS	100 µL	110 µL
10% APS*	50 µL	40 µL
TEMED	10 µL	10 µL

* APS - Ammonium persulfate

II. Gene amplification sequence by PCR

At bold are highlighted the enzyme restriction sequences (*Nco*I - 5´...CCATGG...3´ and *Bam*HI - 5´...GGATCC...3´).

At red are highlighted the start (GTG or ATG) and stop codons (TAA).

a. *mdpn* PCR sequence

5´...GCGGC**CCATGG**GTCACCATCACCATCACCATGATGACGATGACAAAGTGGAATATATAAAGAAGGATAATAA
TTATAAGTTATTACTAGGTGACTGTATTAAAGAACTAAAAAAAATTCAAAGTAAATCAGTAGATATGATATTTGC
TGATCCACCCTATAAGCTTAGTAATGATGGAATAACATGTAAGTCAGGGAAAATGGCAAGTGTAAATAAAGGAAG
TTGGGATAAATCTTTAGGACCCGAGCTTGATTATAAGTTTAATATGAAATGGCTAAAAGCTTGTGATAGAGTTTT
AAAAGATGATGGAACTATATGGATATCAGGAACATATCATATTATTCATTTAATTGCATTTGCATTACAGAAAAT
GGATTATTACATAATAAATGAAATTACTTGGGTTAAGCCTAATGCCGCTCCTAATATGGGATGTAGATGTTTTAC
AGCAAGTCAAGAAACAGTACTTTGGTTAAAAAAGACAAAAAAAGCAAAACATATATTTAACTATGAACAAATGAA
GGAAATGAATGGTGGAAAACAGATGAGAAGTGTATGGGAAATACCAACTACCCCTAAAAAAGAAAAGAAAAATGG
TTATCATCCTACTCAAAAACCAGAAGCATTATTACAAAGATGCATTAGATCTAGTACAAATGAAAGAAATGTAAT
TTTAGATCCATTTTGTGGCTCTGGAACAACAGGTGTTATAGCTATAAAAAACAAAAGAAAATTTATAGGAATAGA
TATAAATCAAGAATATTTAGATTTATCTAAAAAGAGATTTGATTTATACTTACAGGAGGATAAAAGTAGTGGAGA
GTAA**GGATCC**ACCTT...3´

b. *dpnII* PCR sequence

5´...GCGG**CCCATGG**GTCACCATCACCATCACCATGATGACGATGACAAAATGAAGAATAGAAATTTTGATGAATG
GTTAGACAAGTTTAGAATAAGTATTTCTAGGTATGATTATTATATTAATTTCAATAAAGTAGTTAGCAATGTAGA
GAAAATTAAAGTTGAATTAAATATTTTAAATTCATTGATAGGTTCAAAGAATATAGAAAATGATTTTGAAAATAT
CATTAGTAAGTATCCTGAGACCTTAAAGTGTATTCCAACATTATTGGCTGTACGCCAAAGCGAGATTTATGCACA
AGATAAAGATGGTGCATTTATGTATAATTTCAATGAAATGAATTTTGGCATCGAACAATATAGTGTATTTATGAA
AAAGACAGGATTATTTGATTTGATATCTAATCACCTTGTAAATAATTTAGTTGACTATGCTTTAGGTACAGAGAC
TGGTTTGGACTCAAACGGAAGAAAGAACCGTGGTGGTCATCAGATGGAAGATTTGGTAGAAGAATATATTAAGAA
GTCAAATGTTAAAGAGTATTATAAAGAAATGTATTTAGCTGACATAGAAACAAAATGGGATATTGATTTATCTGC
TTTATCCAATAATGGAAAAGCAAGGAAAAGATTTGATTTTGTTGTAAAAACAGATAGCATGATATATGCAATTGA
AACCAACTTTTATGGTGGGACAGGCGGTGGCTCTAAGCTGAATGAAACAGCTAGAAGTTATAAAATGCTTTCTCA
AGAATCTGATACAGTAGATGGATTTACTTTTGTATGGTTAACAGATGGAACGGCATGGCGAAGTGCAAGAGGAAA
CCTTAGAGAAACATTTGATGTAATGGACCACATTTATAGCATTGATGATATGGAAAATGGCATTATGGATAAAGT
GTTCGTATAA**GGATCC**AATTA...3´

c. *dam* PCR sequence

5´...GCGG**CCCATGG**GTCACCATCACCATCACCATGATGACGATGACAAAGTGGAGAGTAAGATAAAACCTTTTTT
AAAGTGGGCAGGCGGTAAAACACAATTGTTACAACAAATTATTGAAAATTTACCGGAGAATATTGAACAGATAAG
AAAATACGTAGAACCATTTGTTGGCGCAGGAGCTGTATTTTTATGTTTAGCAAGTAATAATTGTTTTGATGAATA
TATTATTAACGATATTAATCATAAATTAATTAATTTATATATAGTAATGAGAGATAATTGTGATGAATTAATAGA
AGAATTAAGAATTCTTAAGGAATTATATTTATCATTAGAATCAATTGAAAAAAAAGAGGAGTTTTACTACAAGAT
AAGAGACGAGTTTAATGAAGAAAACAGTAATAGTATACGAATGGGAGCTTTATTTGTATTTTTAAATAAGACTTG
TTTTAATGGTCTCTATAGAGAAAATTCTAAAGGCAAATTTAATGTACCTTTTGGAAAGCATATATCTCCCGGTAT
TTATCAAGAAAATGAAATAAGGGATATATCACAAGTCCTTAATACTAAAAACAGTAATGGTGAATTAAAAGTAAA
AATATTAAATACATCGTTTGAAAATATAAATGAATATATAGATAATAATACATTTGTTTATTTTGATCCACCATA
TAGACCAGTTACATTGGGCGGATTTAATTCATATAGTAAGAGTGGTTTTAATGACGATAGCCAGATAAAACTAAG
AGATTTTTATGATGATATGGATAAAAAAGGCGCTAAACTAATGTTAAGTAATTCTGATCCTAGGATATTAGATAA
GGATGATGATTTTTTTGATTTGCTTTATAAAAAATTCTCAATAAAAAGAGTACGTGCAAGTAGGATGATAAATAG
CAAAGGAAGTGGTAGAGGTGCTATATCTGAACTGCTAATAACTAATTATGATAAATAA**GGATCC**TTAAA...3´

III. SDS-PAGE protein production results

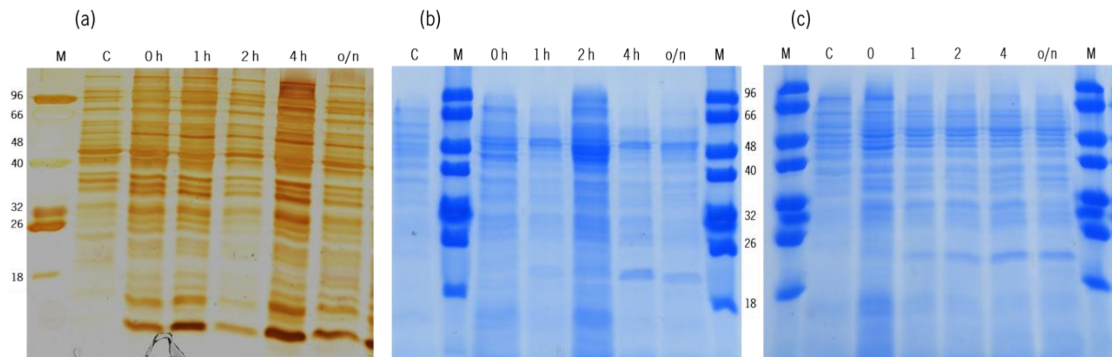


Figure 6.1. SDS-PAGE analysis of BL21 (DE3) CFE production of the enzymes Dam (a) Mdpn (b) and DpnII (c) at 16 °C. M. LMW, kDa; C- control, empty pETduet o/n production; the samples were collected at 0 h, 1h, 2h, 4h and overnight (o/n).

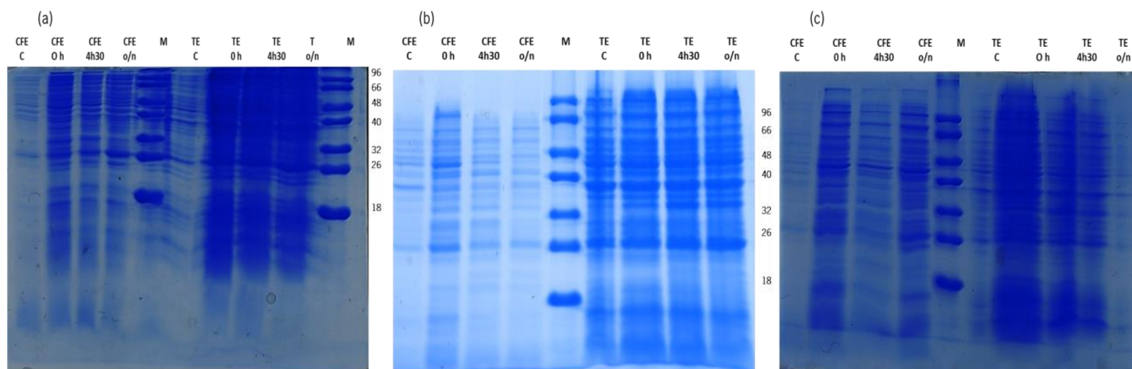


Figure 6.2. SDS-PAGE analysis of Rosetta (DE3) pLysS CFE and TE production of the enzymes Dam (a) DpnII (b) and Mdpn (c) at 37 °C. M. LMW, kDa; C- control, empty pETduet o/n production; the samples analyzed were of 0 h, 4h30 and overnight (o/n).

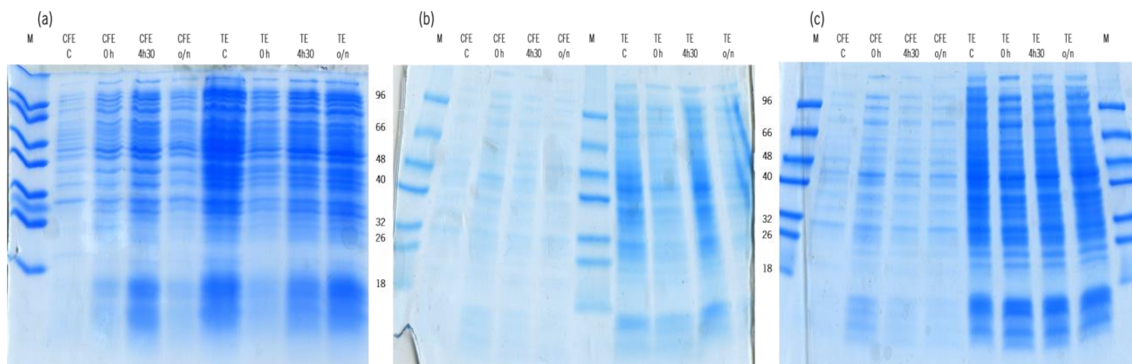


Figure 6.3. SDS-PAGE analysis of Origami (DE3) CFE and TE production of the enzymes Dam (a) DpnII (b) and Mdpn (c) at 37 °C. M. LMW, kDa; C- control, empty pETduet o/n production; the samples analyzed were of 0 h, 4h30 and overnight (o/n).

IV. Multiple sequence alignment

```
opC             TTGGAGCAGCTGTCCATTTTTAATAAAGTTGACGATTTTCAGGAACAACAGACCGACGAC 60
dcm             TTGGAACAATTATCAATTTTTAATAAAGTAGATGACTTCCAAGAACAACAAACTGATGAT 60
                *****.**. *.**.**************:** ** ** **.********.** ** **

opC             CGTGAACTTTCAATCGAAGAGATCAACAAATTCATCAACGAGCATAAACTGACGGAGCGC 120
dcm             AGGGAATTAAGCATAGAAGAAATTAATAAATTTATTAATGAACACAAACTTACCGAAAGA 120
                .* *** *:: .**.*****.** ** ***** ** ** **.** ***** ** **..*.

opC             ATTGACATTATTAACACGGAGAATGCGTCCAAACGTCGTTTCACCATCTTATCTCTGTTC 180
dcm             ATTGATATTATAAACACTGAGAATGCTTCAAAAAGAAGATTTACAATTTTATCTCTATTT 180
                ***** *****:***** ******** **.***.*:.*:** **.** ********.**

opC             TCGGGGTGCGGTGGTTTAGATCTGGGCTTTAAAGGTGGCTTCACGTACCTGCACGGCAAC 240
dcm             TCAGGTTGTGGCGGTTTGGACTTAGGTTTCAAGGGTGGATTTACATATTTACATGGGAAT 240
                **.** ** ** *****.**  *.** ** **.*****.** **.**  *.** ** **

opC             TATGAACGCAATAACTTCGATATTATTTGGGCGAATGAAATCAACTCCCAGGCGGTGGAA 300
dcm             TATGAAAGAAATAACTTTGATATTATTTGGGCTAATGAGATAAATAGTCAAGCTGTTGAG 300
                ******.*.******** ************** *****.**.** :  **.** ** **.

opC             ACGTACCGTAGCTATTTCGGGAATCACATCGTATGCGAAGACATTAACAACATCCGTGAT 360
dcm             ACATATAGGAGTTACTTTGGAAATCATATTGTTTGTGAAGATATAAATAATATTAGAGAT 360
                **.** .* ** ** ** **.***** ** **:** ***** **:** ** ** .*:***

opC             GATGAGTTTCCGCAGGCTGATATCATCATTGGGGGTTTCCCATGCCAAGATTTTTCACTG 420
dcm             GATGAATTCCCACAAGCAGATATTATAATAGGTGGATTCCCTTGTCAAGATTTTAGCTTA 420
                *****.** **.**.**:***** **.**:** **:*****:** *********: . *.

opC             GCGGGCAAGAAACAGGGTCTGAACGTTGAACGCGGTCGCCTGTACCTCCAGATGAAACGC 480
dcm             GCTGGTAAGAAGCAAGGTCTTAATGTAGAAAGAGGAAGACTATATTTACAAATGAAAAGA 480
                ** ** *****.**.***** ** **:***.*.**:.*.**.**  *.**.******.*.

opC             GCAATCGATGCAGTCAAACCAGTGGCCTTTATCGCCGAGAACGTCCGTAATCTCATGGTG 540
dcm             GCAATAGATGCTGTTAAACCAGTGGCATTTATAGCTGAAAATGTTAGAAATTTAATGGTG 540
                *****.*****:** ***********.*****.** **.** ** .*:*** *.******

opC             ATGGGTAATGGGGTGGTTCTGAAAACTATCATTGACGATTTTAAGCAGAGCGGCTACAAT 600
dcm             ATGGGAAATGGAGTTGTTCTAAAGACAATAATAGATGATTTTAAACAAAGTGGATATAAT 600
                *****:*****.** *****.**.**:**.**:** ********.**.** **.** ***

opC             GTCTATTTCCACCTGTACAATGCGGCAAATTATGGCGTTCCGCAGAATCGCGAACGCGTT 660
dcm             GTTTATTTTCATTTATATAATGCAGCAAATTATGGAGTACCTCAAAATAGAGAAAGGGTA 660
                ** ***** **  *.** *****.***********.**:** **.***.*.***.* **:

opC             ATCATCTATGGCATCCGTGAGGACTTAAATAACATCCCGTTTATTCCTCTGGAAACCCAT 720
dcm             ATTATTTATGGTATAAGAGAAGACCTTAATAATATACCATTTATACCATTAGAGACTCAT 720
                ** ** ***** **..*:**.*** *:***** **.**.*****:**: *.**.** ***

opC             TCACTGTATAATTGGGTGACGGCGAGCGAAGCAATTGATGATCTGTGGGATAAACTGGAC 780
dcm             TCTTTATATAATTGGGTGACGGCAAGTGAAGCTATAGATGATTTATGGGATAAACTAGAT 780
                **: *.****************.** *****:**:****** *.**********.**

opC             ACTAACATCCCAAATCATTCGCGCCGCGATTACTCGAAGGCGAAATTCTATGAGGGTAAA 840
dcm             ACTAATATTCCAAATCATTCAAGGCGAGATTATTCAAAAGCAAAGTTTTATGAAGGAAAA 840
                ***** ** ***********..* **.***** **.**.**.**.** *****.**:***

opC             CGTACACAAGGCAATATTCGTATCCAGAGTGATAAGGTGGCCCCGACTATCCGCGCTGAA 900
dcm             AGAACTCAAGGGAACATAAGGATACAAAGTGATAAGGTAGCACCAACTATAAGAGCAGAG 900
                .*:**:***** ** **:.* **.**.***********.**.**.*****..*.**:**.

opC             CACCACGGGAATATTGAAGGTCATTACCGGACCTACGGAGATGAATCGGATCTGTCTAAC 960
dcm             CATCATGGAAATATTGAAGGGCATTATAGGACATATGGTGATGAATCAGACTTATCAAAT 960
                ** ** **.*********** *****  .****.** **:********.**  *.**:**
```

```
opC        TGGCGTCGTCTTAGCGTCCGCGAGTGTGCGCGCATCCAGACCTTCCCGGACGACTTCATT 1020
dcm        TGGAGAAGGTTATCAGTAAGAGAATGTGCTAGGATACAAACATTTCCTGATGATTTCATT 1020
           ***.*:.*  *:: .**..*.**.***** .* **.**.**.** ** ** ** ******

opC        TTTCAAAGCGCAGCCTCTTCAGCGTACAAACAGGTTGGCAACGCCGTGCCGCCGGTACTG 1080
dcm        TTTCAATCTGCGGCATCATCAGCATATAAGCAAGTTGGTAATGCTGTACCACCAGTTTTA 1080
           ******:  **.**.**:*****.** **.**.***** ** ** **.**.**.**: *.

opC        GCCTGGAACATCGCGCGTGCGCTGTTCCTGAGCCTGATTCGCATTAAATAA 1131
dcm        GCATGGAATATAGCAAGAGCACTATTTTTATCATTAATAAGAATTAAGTAG 1131
           **.***** **.**..*:**.**.**  *.: . *.**:.*.*****.**.
```

Figure 6.4. Alignment of the *dcm* gene with codon usage optimized (opC) and the conventional *dcm* gene (dcm) from *C. pasteurianum*. Alignment was performed by T-coffee (Di Tommaso et al., 2011). ***** indicate nucleotide similarity; **.** replacement of A for C or G; **:** replacement of A for T or T for A; **space** replacement of G for T, of T for C, of C for G or of T for G.

V. Primer list designed to further *dcm* gene amplification

The plasmid pETduet-1 was selected as the vector and the primers were already designed. In these primers the enzymes *Nco*I and *Eco*RI restriction sequences were inserted. A 6xHis-tag was also added to allow future enzyme recovery (up- or downstream of the protein), along with an enterokinase cleavage site to remove the His-tag or a sequence (spacer), in order to give flexibility to the His-tag. In Table 6.4 the primer list is shown, with the respective sequence additions highlighted and the associated melting temperature.

Table 6.4. List of primer to use in *dcm* cloning process. These primers were synthesized by Invitrogen.

Primer	Sequence (5´ → 3´)	Tm ºC
Dcm1.F	GCGGCCCATGGGTCACCATCACCATCACCATGATGACGATGACAAA**TTGGAGCAGCTGTCCATT**	48
Dcm1.R	AAGGTGAATTC**TTATTTAATGCGAATCAGGC**	48
Dcm2.F	GCGGCCCATGGCC**TTGGAGCAGCTGTCCATTT**	48
Dcm2.R	AAGGTGAATTCTTAATGGTGGTCATCGTCATCGCCGCTGCC**TTTAATGCGAATCAG**	48

Yellow: restriction sites, *Nco*I (C|CATGG) and *Eco*RI (G|AATTC)
Green: 6xHis-tag sequence;
Grey: Enterokinase cleavage site (Asp-Asp-Asp-Asp-Lys|)
Blue: sequence (Gly-Ser-Gly linker) for flexibility of Histag;
Bold: gene sequence to amplify.