UNIVERSITY OF VENICE

Industrial PhD. Program in Science and technology of bio and nanomaterials

XXXIV Cycle

Title: Development and validation of experimental workflows for a DNA foundry

Supervisor

Prof. Flavio Rizzolio

Co-supervisor

Dr. Davide De Lucrezia

Candidate

Sota Hirano

Explora Biotech srl

Academic Year 2020-2021

Table of Content

Acknowledgements

This work was supported by Explora Biotech srl and EU projects MIAMI (Grant Agreement No. 814645), ACDC (Grant Agreement No. 824060), LIAR (Grant Agreement No. 686585), TOPCAPI (Grant Agreement No. 720793), and DRIVE (Grant Agreement No. 645991).

I would like to express my deepest gratitude and affection to Dr Davide De Lucrezia for his invaluable guidance and support as Explora Biotech srl CEO and senior researcher in synthetic biology. Nothing in this project would have been possible without the support and nurturing of Davide.

I want to extend my sincere thanks to the University of Venice Ca' Foscari and my supervisor Dr Flavio Rizzolio for allowing me to apply and study for the Industrial PhD program in science and technology of bio and nanomaterials.

I cannot begin to express how thankful I am for every DNA foundry team member and Explora Biotech srl colleague. During this whole project, I could always rely on their help and their invaluable scientific insights.

I'm incredibly grateful for my family and the closest people I had during this journey. Without their constant love and support, I would not be in the position I am in today.

Abstract

Synthetic biology aims to exploit foundational technologies in the current expansion of biotechnology applications that make the design and manufacturing of engineered biological systems easier and more reliable. Instrumental to fulfil this vision is the decoupling between design and fabrication. Decoupling is defined as breaking down a complex task into simpler and independent ones such that the resulting work can eventually be recombined to produce a functioning whole. This is in stark contrast with the traditional setting where individual researchers needed to carry out different tasks ranging from DNA design over assembly to quality control. Thus, decoupling enables a team of complementary experts to leverage individual specializations to achieve better outcomes. By decoupling design from fabrication, we empower researchers to design complex constructs irrespective of and independently from the manufacturing technique, thus unlocking the full potential of synthetic biology. A typical example is DNA construct assembly, where a large number of standard plasmid architectures (e.g. SEVA, MoClo, GoldenGate, RFC10) and related assembly techniques actually hinder researchers' design capabilities rather than enable them. In this Ph.D. project, I focused on developing and validating DNA fabrication workflows capable of handling multiple assembly techniques independent from the design input. The set of workflows developed allows researchers to submit their DNA design without any constraints related to fabrication techniques. The work presented here represents the foundational technology to enable a truly automated DNA foundry for synthetic biology.

Disclaimer

Some sensitive business and experimental information are not included since these data would pose a risk to the company if released to a competitor or the general public.

The appendix section at the end of this document was created to include the working principles of basic molecular biology techniques used routinely in this project.

Introduction

Recombinant DNA (rDNA) consists of circular DNA molecules, called “plasmids” [1], achieved artificially through the assembly of DNA sequences from different origins inserted into a host organism to add new properties and functions useful for any downstream applications. The rDNA technology found many applications for industrial uses in the last decades providing solutions to agricultural, environmental, human health field applications. The critical goal corresponds to enabling host organisms (chassis) to produce products in large amounts that it does not usually make. Ten years after the first rDNA creation in 1972, a biotech company, Genentech, released the world’s first genetically engineered human drug: synthetic insulin [2]. By the first half of the 1980s, several DNA-based enabling technologies were discovered, including the polymerase chain reaction in 1983 by Kary Mullis [3], making the whole process to assemble (or “clone”) rDNA faster and more reliable. With the development of more advanced techniques and technologies for genetic engineering applications, the DNA molecule became a potentially disruptive technology that opened new frontiers for therapeutic and pharmacological applications (AAV [4], DNA-based vaccines [5][6], gene targeting plasmids [7] and more). However, despite genetic engineering bringing huge individual successes in several biotechnological industries and biological research projects [8][9][10], this scientific field remained either too expensive or too unpredictable due to the complexity of biological systems. In order to allow the emergence of a new, multidisciplinary area of research called "synthetic biology", three foundations needed to be applied to complex biological systems. These foundational concepts were standardization, abstraction, and decoupling [11][12][13][14].

Standardization involves the definition, interchangeability and reliability of fundamental biological parts, to provide a common set of elements to be used among different laboratories, exemplified by the registry of standard biological parts [10]. Moreover, the standardization aids in setting common vocabularies (Synthetic biology open language (SBOL) [16]), assembly standards (request for comments (RFC), modular cloning (MoClo) [17] etc.), and communities (international genetically engineered machine (iGEM)) for researchers from different disciplines [18].

Abstraction includes defining a level-based hierarchy of complex systems such as long DNA sequences divided into modules in which each level exchanges a limited set of information to other levels.

Decoupling permits the division of complex systems into simpler units decomposing complicated problems into simpler ones and separating the design from manufacturing. A handy analogy to visualize both the goal and approach of synthetic biology is the computer engineering hierarchy. Every component within the same hierarchy level is embedded in a more complex element in a higher hierarchy. The final design and its behavior are composed in a bottom-up fashion which can be seen at the highest level. [19]

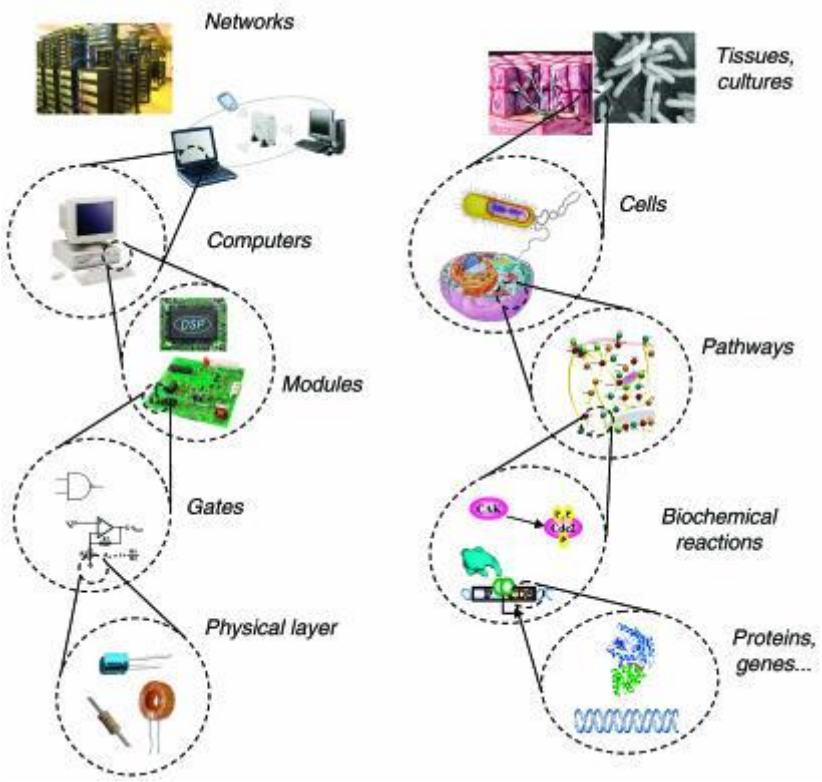

Figure I: Computer engineering inspired synthetic biology hierarchy.

In 2010, the year ended with the first crucial synthetic biology milestone: synthesis of a complete working bacterial genome at the J. Craig Venter Institute (JCVI) [20]. This research showed the possibility of synthesizing megabase-sized DNA sequences, achieving one of the century's biggest challenges. However, the synthesis of a large genome was not enough to deliver other ambitions of this field. Even after the creation of significant active communities and standards, in 2010 the paper “Five Hard Truths for Synthetic Biology” [21] was published, which showed how the lack of progress on engineering ambitions still made this science lacking reliability and automation. In 2016 “Cello” was published, an end-to-end computer-aided algorithm to design logic circuits in *E. coli* [22] which provided a remarkable solution in addition to other works released in that decade.

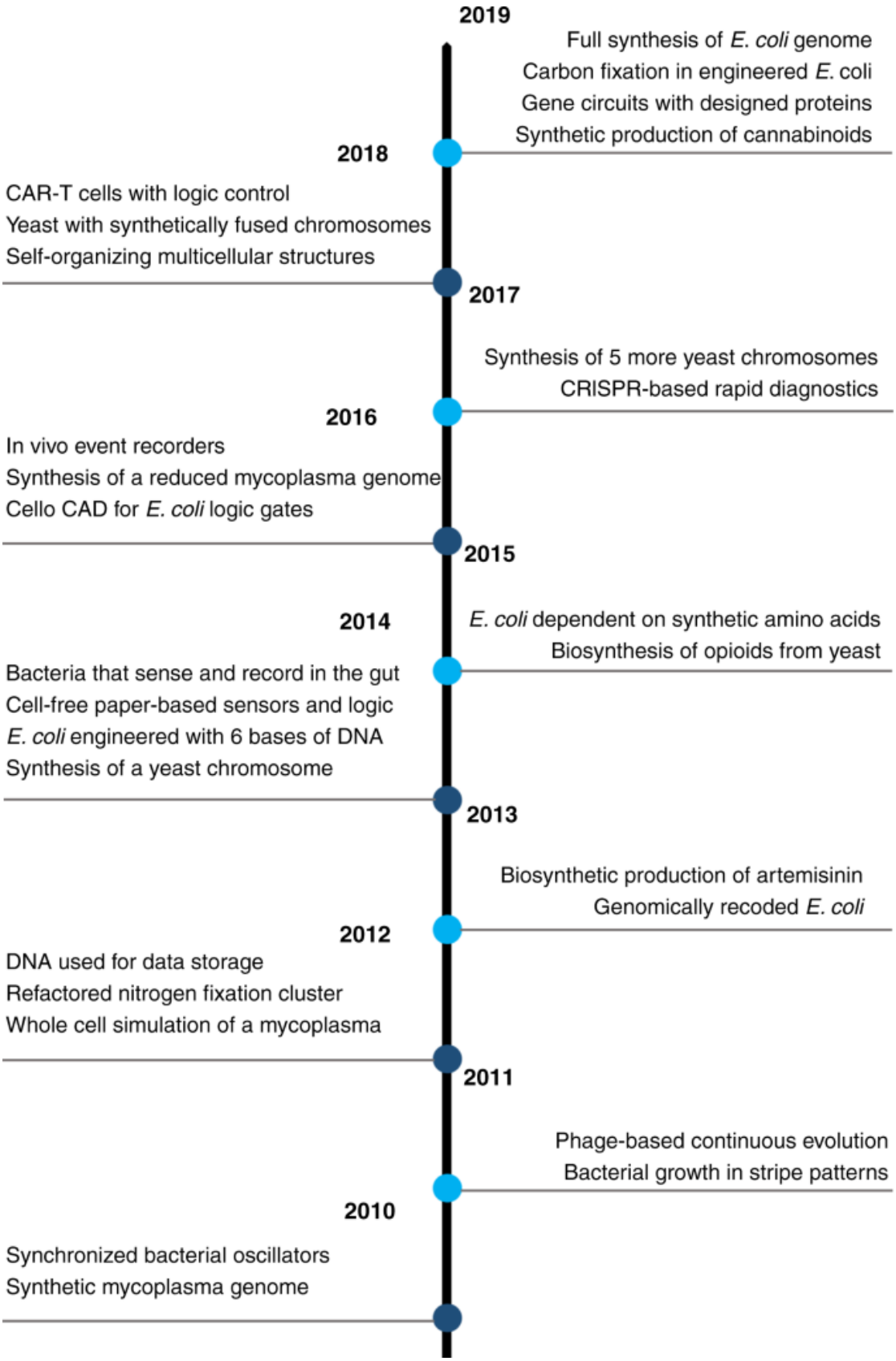

Figure II: Synthetic biology key milestones [23]

At the start of 2020, over 30 countries including UK [24], the Australian Council of Learned Academies (ACOLA) report [25], USA [26] implemented national strategies related to "bioeconomy", where the growth of synthetic biology capabilities is identified as critical to scientific and economic competitiveness. Biofoundries can acquire this goal. A biofoundry is a facility that combines synthetic biology with automation engineering to generate biological solutions for both academic and commercial purposes. These solutions help to generate and make more robust the Design-Build-Test-Learn (DBTL) approach to biological engineering [28].

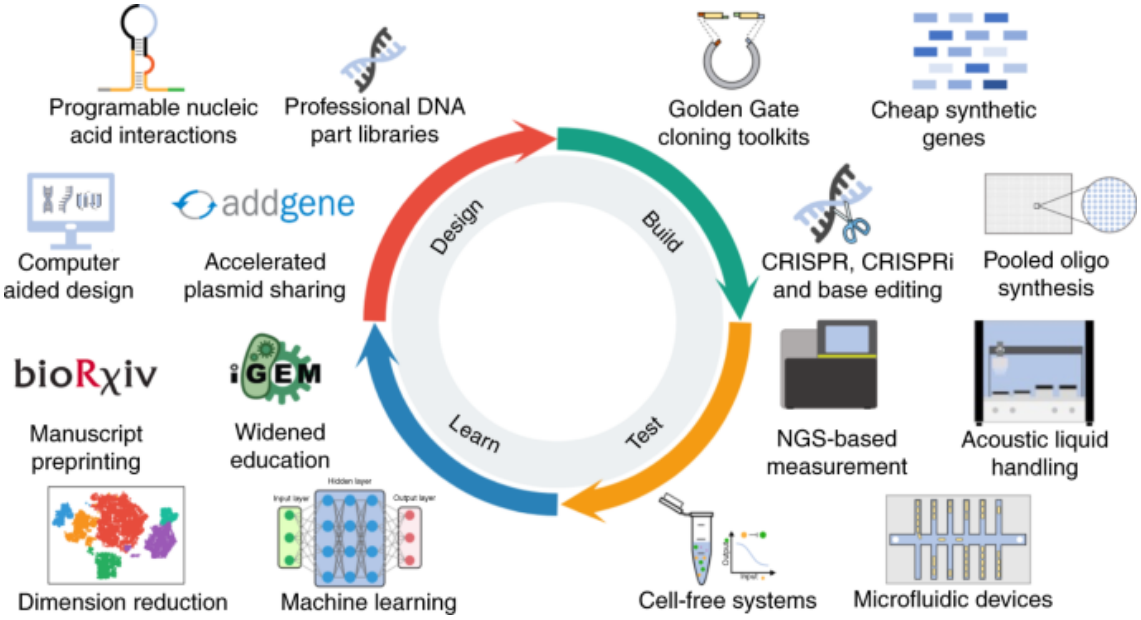

Figure III: DBTL cycle with key synthetic biology technologies that help to accelerate each phase of the cycle. [23]

Several biofoundries have already been built around the world and a Global Biofoundries Alliance (GBA) for non-commercial biofoundries was launched in 2019 with 16 members, already having grown to 27 members in 2020 [28]. Through a large number of experiments and repetitive, standardized tasks, biofoundries can increase the design and throughput space for biological engineering [29].

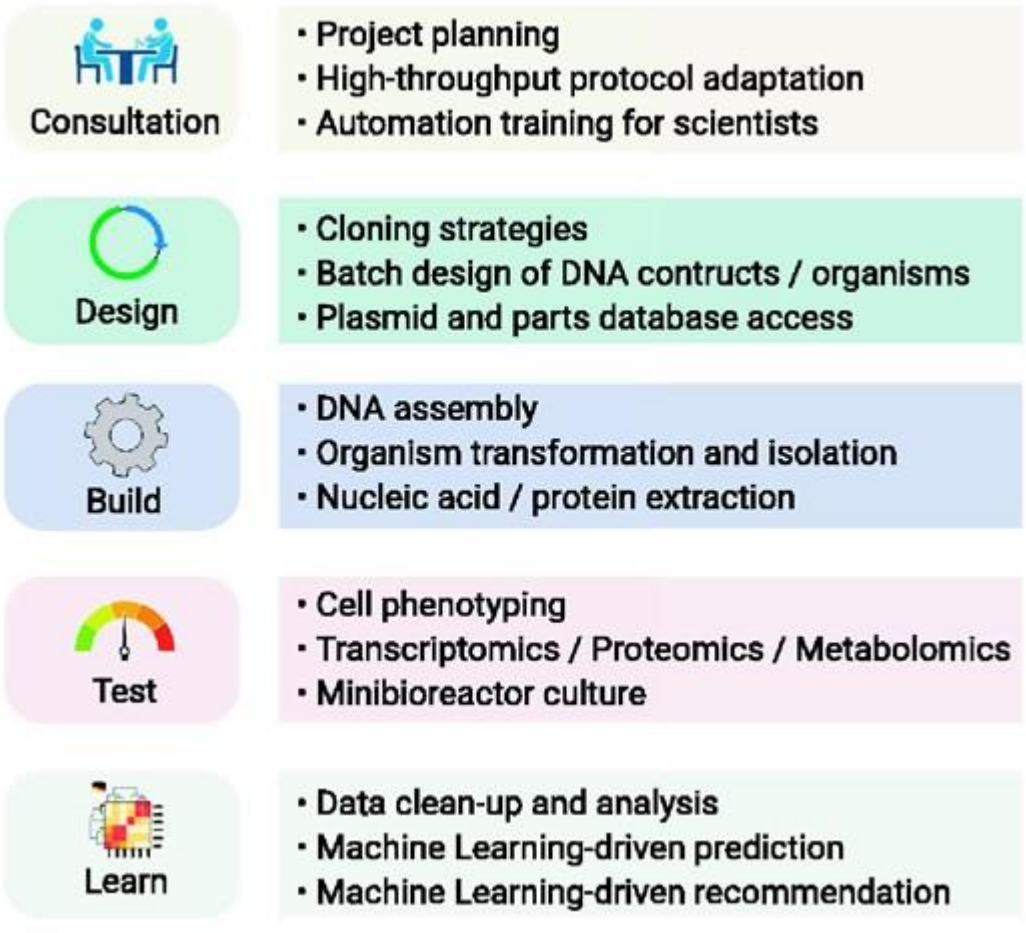

Figure IV: Service offered from a given biofoundry to cover a specific area of scientific research.

The synthesis and assembly of DNA oligonucleotides into ultramers, linear dsDNA fragments, genetic circuits and even entire genomes have become the most valuable techniques for the DBTL and repeat cycles, bringing either innovation or enabling major progress in synthetic biology. The development of a cheaper, faster, more robust manufacturing process and delivery of larger variety of synthetic DNA became crucial for the exploration of more complex biological hypotheses [30]. This critical fabrication process is carried out by DNA foundries, which are structures specialized in covering the “build” phase of the cycle, providing both DNA synthesis and assembly. The work described here exploited the synthetic biology basis to establish a new DNA foundry for industrial and commercial applications. This goal was achieved through the definition of a set of standards and procedures to be formalized and executed under the most stringent ISO-9001 quality controls; definition of project design approaches and synthesis techniques to decouple any initial DNA design from the fabrication; design of internal workflows and quality check criteria; development of digital infrastructure to store experimental data and a framework to assign, monitor and coordinate tasks between multiple operators and researchers.

In the middle of 2021, the synthesis and cloning success rate reached 98 % independently by the starting DNA designs, thereby establishing the first cloning agnostic DNA foundry in Europe. Moreover, our DNA foundry is now equipped to manufacture and assemble any DNA construct up to 42 kbp and has synthesized more than 1.5 million base pairs for customers located both in the EU and US.

Abbreviations

A = Adenine
AAV = Adeno-associated virus
AGE = Agarose gel electrophoresis
BB = Bio brick
BoM = Bill of materials
bp = Base pair
C = Cytosine
cPCR = Colony polymerase chain reaction
DBTL = Design-Build-Test-Learn
ddNTPs = Dideoxynucleotides
DNA = Deoxyribonucleic acid
dNTP(s) = Deoxynucleotide triphosphate(s)
dsDNA = Double stranded DNA
EF = Eurofins
EU = European union
G = Guanine
GBA = Global Biofoundries Alliance
HF = High fidelity
HPLC = High performance liquid chromatography
iGEM = International genetically engineered machine
ISO = International organization for standardization
KF = Klenow fragment
MoClo = Modular cloning
NEB = New England biolabs
NGS = Next generation sequencing
ORI = Origin of replication
o.n. = Overnight
PCR = Polymerase chain reaction
PEG = Polyethylene glycol
QC = Quality check
RBS = Ribosome binding site

rDNA = Recombinant deoxyribonucleic acid

RFC = Request for comments

RNA = Ribonucleic acid

SBOL = Synthetic biology open language

SOP = Standard operating procedures

ssDNA = Single strand DNA

T = Thymine

TAE = Tris-acetate-EDTA

TAT = Turnaround time

US = United states

UV = Ultraviolet

Aims of the thesis

The work carried out in this industrial Ph.D. project is divided into three sections described in the following chapters.

Chapter 1: “DNAmate level-based vectors” describes the first manufacturing project executed by the DNA foundry. This part comprises the design of the manufacturing process, the formalization of an SOP set to describe each experimental step and the first version of the manufacturing workflow of DNA constructs, and a reproducibility test through the execution of a second analogous project.

Chapter 2: “Decoupling from starting design and combinatorial library” describes the identification of a DNA assembly technique to decouple the manufacturing from any initial design of plasmids. Moreover, using the updated cloning workflow, it was tried to set up another workflow to synthesize a combinatorial DNA library with 3 NNN groups.

Chapter 3: “Production scale-up” describes an overview of executed projects between 2020 and the first half of 2021, formalizing the whole infrastructure and framework for our DNA foundry and final evaluation of synthesis outcomes.

Chapter 1: DNAmate level-based-vectors

In the very early stage of DNA foundry, the first project consisted of the synthesis of a set of biological parts (promoters and ribosome binding sites (RBS)) to be subcloned into an expression vector properly designed for in vivo characterization. The manufacturing project was split into three stages, exploiting the hierarchy-based abstraction system of synthetic biology: the synthesis of level 0 plasmids, which each of them carry a promoter to be characterized; the synthesis of level 1 plasmid coding for a functional gene for in vivo characterization of the biological part and the definition of a set of standard operating procedures (SOPs) to be reproduced with RBS level 0 and 1 plasmids. Each plasmid was properly designed to carry type IIs restriction enzyme recognition sites to cut and open the level 0 vector to release the insert DNA fragment. Once the level 0 plasmids were obtained, these constructs were used to generate level 1 plasmids carrying all previous DNA parts in a predefined arrangement. The progress from one level to another was carried out by executing DNAmate as an assembly technique. DNAmate is a Type IIs assembling system, like Golden Gate or MoClo, which allows single-pot directional assembly of up to 8 DNA fragments. Each fragment had a specific prefix and suffix, harboring a type IIS recognition sequence and a unique restriction site. This design allowed the combination of both techniques, since it offered the simultaneous cloning of multiple Golden Gate fragments and the possibility to manipulate the final construct at will, using unique restriction sites. The downside of DNAmate was that it left short, undesired sequences (“scars”) between two adjacent fragments. This limitation was solved by carefully designing the scars to not interfere with the construct functions.

Figure 1.1: DNAmate assembly mechanism overview.

The plasmid for level 0 was designed to have an ampicillin resistance gene, the DNA fragment flanked by hinges compatible with DNAmate assembly and BsaI, BsmBI RE recognition sites. The high copy number ORI of the backbone was derived from pUC plasmids. For this purpose, pDLX1A100 (https://doulix.com/biomodules/QEOCL7W/?q=pdlx_1A1) was chosen, which had been previously designed by Explora Biotech srl to be compatible with DNAmate assembly. The inserts consisted of 10 different promoters from - 35 region to the base - 1 to characterize different promoter's activity without including the first base of the RBS. This architecture permitted the evaluation of the promoter's strengths, scrambling different RBS as downstream applications.

Promoter name	Promoter sequence
BbaJ23119	TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGC
BbaJ23110	TTTACGGCTAGCTCAGTCCTAGGTACAATGCTAGC
BbaJ23114	TTTATGGCTAGCTCAGTCCTAGGTACAATGCTAGC
BbaJ23100	TTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGC
pbla	TTCAAATATGTATCCGCTCATGAGACAATAACCCT
pTacI promoter	TTGACAATTAATCATCGGCTCGTATAATGTGTGG
placUV5	TTTACACTTTATGCTTCCGGCTCGTATAATGTGTGG
lacIQ	GTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCC
pTrc	TTGACAATTAATCATCCGGCTCGTATAATGTGTGGA
T7 promoter	TAATACGACTCACTATA

Table 1.1: List of promoters to be subcloned into pDLX1A100 Level 0 plasmids. The highlighted blue nucleotides are part of the region -35, while the green nucleotides stand for the -10 region and the last nucleotides, indicate the -1 bases.

The BbaJ231XX promoter series was taken from the iGEM registry of biological parts with the properties shown in table 1.2.

Name	**Sequence**[a]	**Strength** [b]
J23119	ttgacagctagctcagtcctaggtataatgctag c	reference
J23114	ttt**t**atggctagctcagtcctaggta**ca**atgctag c	0.10
J23110	ttt**t**ac**g**gctagctcagtcctaggta**ca**atgctag c	0.33
J23100	ttgac**g**gctagctcagtcctaggta**c**a**g**tgctag c	1

*Table 1.2: List of promoter series from iGEM registry of biological parts. a) the sequence of individual promoters is shown in black and red. The black nucleotides are part of the consensus promoter sequence (BBa_J23119) among all promoters, while the **bold red** nucleotides highlight the differences between the individual promoters and the consensus sequence. b) The relative strengths of these promoters were measured by Chris Anderson and the 2006 Berkeley iGEM team.*

The sequence of the bla promoter was taken from: “Plasmid vector pBR322 and its special-purpose derivatives--a review.”[31]

>Pbla

TTCAAATATGTATCCGCTCATGAGACAATAACCCT

Both sequences of pTacI and lacUV5 promoters were taken from: “The tac promoter: A functional hybrid derived from the trp and lac promoters”[32]

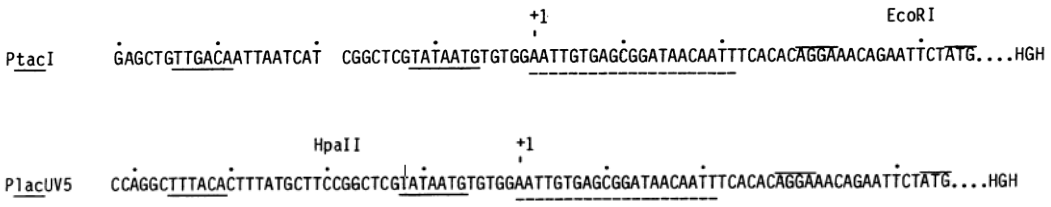

Figure 1.2: the -35 sequence and the Pribnow box sequence of the promoters were underlined. Dots indicate every tenth nucleotide. The transcription start sites are indicated with + 1.

Furthermore, these promoters were analyzed on “ Spacing of the -10 and -35 Regions in the tac Promoter” [33].

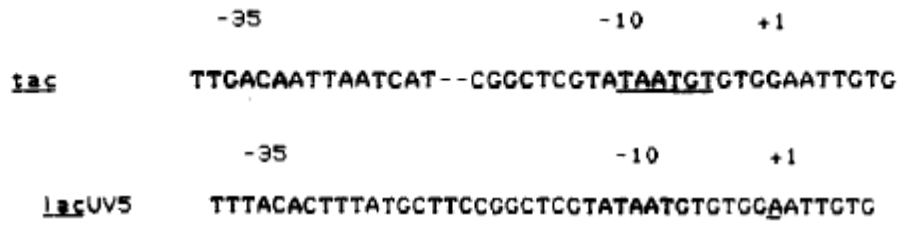

Figure 1.3: the -35, -10, and +1 sites identification for pTacI and lacUV5 promoters.

According to the previous studies, the regions between - 35 and + 1 were extracted.

>pTacI

TTGACAATTAATCATCGGCTCGTATAATGTGTGG

>placUV5

TTTACACTTTATGCTTCCGGCTCGTATAATGTGTGG

The sequence of the LacIq promoter was taken from: “From adjacent activation in *Escherichia coli* and DNA cyclization to eukaryotic enhancers: the elements of a puzzle” [34] and “DNA sequence for a low- level promoter of the lac repressor gene and an ‘up’ promoter mutation” [35].

> lacIQ

GTGCAAAACCTTTCGCGGTATGGCATGATAGCGCCC

The sequence of the Trc promoter was taken from: “Spacing of the -10 and -35 Regions in the tac Promoter”. [33]

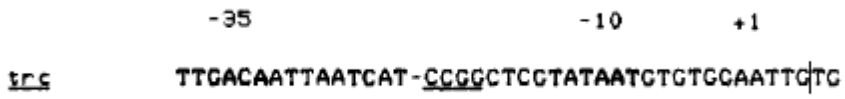

Figure 1.4: the -35, -10 and +1 sites identification for trc promoter.

> pTrc

TTGACAATTAATCATCCGGCTCGTATAATGTGTGGA

The sequence of the T7 promoter was identified according to: “Gene Position More Strongly Influences Cell-Free Protein Expression from Operons than T7 transcriptional Promoter Strength” [36].

The promoters' sequences lengths were within a range of 17 bp - 36 bp. As Level 0 cloning techniques mutagenic PCR was used. As level 1 backbone pDLX6K200 (https://doulix.com/biomodules/653CG7P/?q=6k2) was chosen. This backbone has a medium copy-number ORI from pET plasmids and harbors a kanamycin resistance gene.

The level 1 gene was assembled as follows:

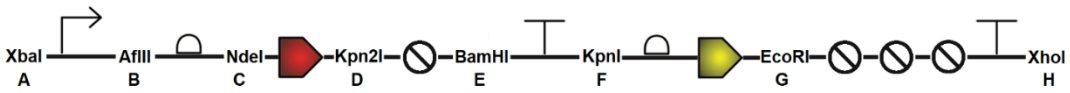

Figure 1.5: Representation in SBOL (Synthetic Biology open language) of Level 1 assembled insert. Under each Type II RE site, there is a letter indicating the respective Hinges. The variable element was the promoter between the hinges A and B.

Once the level 0 and 1 promoter libraries had been generated, the cloning process was analyzed to identify all occurred non-compliances and potential issues which could cause downstream manufacturing failures. At the end of this project, the first set of SOPs and a manufacturing workflow were generated to be reproduced for RBS library synthesis.

1.1 Materials and methods

1.1.1 Reagents and general supplies

All the restriction enzymes, Q5® High-Fidelity DNA Polymerase and *E. coli* NEB stable competent cells were provided from New England Biolabs; Wizard® SV Gel and PCR Clean-Up System Wizard® Plus SV Minipreps DNA Purification System were purchased from Promega; molecular biology water, LB, agar, NaOH, Phusion DNA polymerase, T4 DNA Ligase, PEG4000 and glycerol were provided by Sigma-Aldrich; Isopropanol, Ethanol, 50X TAE buffer was purchased from Thermo Fisher; E.Z.N.A.® Plasmid DNA Mini Kit I was purchased from Omega Biotek; Phusion polymerase was purchased from Agilent.

1.1.2 Primer design and providers

All the primers were provided by external suppliers (Eurofins Genomics, Ebersberg, GER and Genewiz, Leipzig, GER) with salt-free oligo qualities. The primers were designed using the online tool IDT oligo analyzer (https://eu.idtdna.com/pages/tools/oligoanalyzer) with the following parameters: oligo concentration 0.5 µM, Mg++ concentration 1.5 mM, dNTPs concentration 0.2 mM. The annealing temperature was kept within a range of 55 °C – 59 °C, and the secondary structure melting temperature was < 50 °C. Primers for mutagenic PCR were designed to keep the same criteria as the normal primers at the annealing region between primer-to-template and primer-to-primer overlapping zone.

Primer name	Sequence
pTac_RV	ATTATACGAGCCGATGATTAATTGTCAATCTAGAGAGACCGCTGAGGGC
pTac_FW	ATTAATCATCGGCTCGTATAATGTGTGGCTTAAGGAGACGCTGCAGCTG
placUV5_RV	TATACGAGCCGGAAGCATAAAGTGTAAATCTAGAGAGACCGCTGAGGG C
placUV5_FW	TTATGCTTCCGGCTCGTATAATGTGTGGCTTAAGGAGACGCTGCAGCTG
Pbla_RV	TTGTCTCATGAGCGGATACATATTTGAATCTAGAGAGACCGCTGAGGGC
Pbla_FW	ATGTATCCGCTCATGAGACAATAACCCTCTTAAGGAGACGCTGCAGCTG
PlacIQ_RV	TCATGCCATACCGCGAAAGGTTTTGCACTCTAGAGAGACCGCTGAGGGC
PlacIQ_FW	CCTTTCGCGGTATGGCATGATAGCGCCCCTTAAGGAGACGCTGCAGCTG
PJ23119_RV	TTATACCTAGGACTGAGCTAGCTGTCAATCTAGAGAGACCGCTGAGGGC
PJ23119_FW	CTAGCTCAGTCCTAGGTATAATGCTAGCCTTAAGGAGACGCTGCAGCTG
PJ23114_RV	GTACCTAGGACTGAGCTAGCCATAAATCTAGAGAGACCGCTGAGGGCC G
PJ23114_FW	CTAGCTCAGTCCTAGGTACAATGCTAGCCTTAAGGAGACGCTGCAGCTG
PJ23110_RV	GTACCTAGGACTGAGCTAGCCGTAAATCTAGAGAGACCGCTGAGGGCC G
PJ23110_FW	CTAGCTCAGTCCTAGGTACAATGCTAGCCTTAAGGAGACGCTGCAGCTG
PJ23100_RV	GTACCTAGGACTGAGCTAGCCGTCAATCTAGAGAGACCGCTGAGGGCC G
PJ23100_FW	CTAGCTCAGTCCTAGGTACAGTGCTAGCCTTAAGGAGACGCTGCAGCTG
PT7_RV	TATAGTGAGTCGTATTATCTAGAGAGACCGCTGAGGGCCG
PT7_FW	TAATACGACTCACTATACTTAAGGAGACGCTGCAGCTG
pTrc _FW	TTAATCATCCGGCTCGTATAATGTGTGGACTTAAGGAGACGCTGCAGCT G
pTrc _RV	ATTATACGAGCCGGATGATTAATTGTCAATCTAGAGAGACCGCTGAGGG C
BBa_B0030_RV	TTTCTCCTCTTTAATCTTAAGGAGACCGCTGAGGGCCG
BBa_B0030 + TAC_FW	TTAAGATTAAAGAGGAGAAATACCATATGGAGACGCTGCAGCTGG
BBa_B0031_RV	GGTTTCCTGTGTGACTTAAGGAGACCGCTGAGGGCCG
BBa_B0031 + TAC_FW	CTTAAGTCACACAGGAAACCTACCATATGGAGACGCTGCAGCTGG

BBa_B0032_RV	CTTTCCTGTGTGACTTAAGGAGACCGCTGAGGGCCG
BBa_B0032 + TAC_FW	CCTTAAGTCACACAGGAAAGTACCATATGGAGACGCTGCAGCTGG
BBa_B0034_RV	TTTCTCCTCTTTCTTAAGGAGACCGCTGAGGGCCG
BBa_B0034 + TAC_FW	TCCTTAAGAAAGAGGAGAAATACCATATGGAGACGCTGCAGCTGG
BBa_J61100_RV	TGTCCCCTCTTTCTTAAGGAGACCGCTGAGGGCCG
BBa_J61100 + TAC_FW	TCCTTAAGAAAGAGGGGACATACCATATGGAGACGCTGCAGCTGG
BBa_J61101_RV	GGTCCTGTCTTTCTTAAGGAGACCGCTGAGGGCCG
BBa_J61101 + TAC_FW	TCCTTAAGAAAGACAGGACCTACCATATGGAGACGCTGCAGCTGG
BBa_Z0261_RV	ATTGATTTCTCCTATTGATTCTTAAGGAGACCGCTGAGGGCCG
BBa_Z0261 + TAC_FW	AATCAATAGGAGAAATCAATTACCATATGGAGACGCTGCAGCTGG

Table 1.1.1: Primer list used for mutagenic PCR cloning processes.

1.1.3 Mutagenic PCR and promoter list

All mutagenic PCRs were executed using Phusion polymerase. A 50 µL mix contained 1X HF buffer, 0.2 mM each dNTPs, 0.5 µM of each primer, 0.02 U/µL Phusion, ~ 2 ng/µL plasmid DNA template. The reaction ran on a thermal cycler with the following protocol: 98°C for 2 min for initial denaturation, 35 cycles of 98 °C for 10 s, annealing temperature for 10 s and 72 °C for 20 s/kb and a final extension step at 72 °C for 10 min. PCR fragments were treated with 0.4 U/µl of DpnI for at least 1 h at 37 °C to remove the starting DNA template. The annealing temperature was evaluated according to the following table:

Promoter	**Tm primer RV**	**Tm primer FW**	**Tm primer FW/RV**	**Length bp**
BbaJ23100	59.1	66.7	56.8	2954
BbaJ23110	66.7	59.1	56.8	2954
BbaJ23114	59.1	66.7	56.8	2954
BbaJ23119	59.1	66.7	56.6	2954
pbla	60.6	59.1	23.9	2954
lacIQ	60.6	59.1	64.7	2955
placUV5	60.6	59.1	60.2	2955
T7_promoter	66.7	59.1	45.8	2936
tac promoter	60.6	59.1	56.7	2953
pTrc	60.3	59.4	58.7	2954
BBa_B0030+TAC	65.9	64.1	61.2	2937
BBa_B0031+TAC	65.9	64.1	59.3	2936
BBa_B0032+TAC	65.9	64.1	62.7	2935
BBa_B0034+TAC	65.9	64.1	59.1	2934
BBa_J61100+TAC	65.9	64.1	60.6	2934

BBa_J61101+TAC	65.9	64.1	59.3	2934
BBa_Z0261+TAC	65.9	64.1	61.6	2942

Table 1.1.2: Melting temperature evaluation table for each primer combination for mutagenic PCR.

As general annealing temperature condition, 60 °C were used with an elongation time of 1 min and 30 seconds.

1.1.4 Ribosome binding site sequences

Seven most used RBS were taken from iGEM registry of biological parts used in a prokaryotic system in an *E. coli* chassis.

RBS	**Description**
BBa_B0030	RBS.1 (strong): modified from R. Weiss
BBa_B0031	RBS.2 (weak): derivative of BBa_0030
BBa_B0032	RBS.3 (medium): derivative of BBa_0030
BBa_B0034	RBS (Elowitz 1999): defines RBS efficiency
BBa_J61100	Ribosome Binding Site Family Member
BBa_J61101	Ribosome Binding Site Family Member
BBa_Z0261	Strong T7.2 RBS

Table 1.1.3: RBS list used for mutagenic PCR cloning processes.

The BBa_B0030 series is a family of RBSs highly characterized by Team Warsaw (2010) and Madras (2016). According to their characterization, BBa_B0034 was considered as a reference promoter where its relative translational strength was set to be 1.0. Known as James Anderson's RBS series, parts J61100-J61150 were a family of similar ribosome binding site basic parts identified from a saturation mutagenic library. All RBS sequences from Anderson's library were weaker than BBa_B0030 series (Team Warsaw 2010). The BBa_Z0261 consists of a strong RBS.

RBS	**Sequence**
BBa_B0030	ATTAAAGAGGAGAAA
BBa_B0031	TCACACAGGAAACC
BBa_B0032	TCACACAGGAAAG
BBa_B0034	AAAGAGGAGAAA
BBa_J61100	AAAGAGGGGACA
BBa_J61101	AAAGACAGGACC
BBa_Z0261	AATCAATAGGAGAAATCAAT

Table 1.1.4: Summary of RBS sequences used for subcloning project.

The primers for mutagenic PCR were designed to add a spacer between the RBS and the ATG start codon. The spacer consisted of TAC CAT sequence.

1.1.5 PCR reaction cleanup

An equal volume of membrane binding solution was added into each PCR reaction tube. SV columns provided by the OMEGA kit were placed into a collection tube for each PCR reaction. The PCR product and binding solution mix were added into each SV column and incubated for 1 minute at room temperature. After centrifugation at 15000 x g, the filtrate was discarded, and 700 µl of membrane wash solution were added to the column. After another centrifugation at 15000 x g for 1 minute and subsequent removal of the filtrate, this washing step was repeated with 500 µl of the same wash solution. The membranes in the columns were dried through centrifugation at 15000 x g for 1 minute, and the collection tubes were discarded. Each column was placed into a 1.5 mL microcentrifuge tube, and 100 µl of pre heated molecular biology grade water at 50 °C were added into each tube. After the last centrifugation at 15000 x g, the DNA solution was recovered and used for downstream processes.

1.1.6 DNA quantification using Nanodrop 2000

The DNA concentration was measured by a nanodrop spectrophotometer measuring the absorbance at 260 nm as a triplicate for each same sample. The ratio of absorbance at 260 and 280 nm was used to check the DNA purity. A ratio of ~1.8 was accepted for this assessment. The 260/230 nm ratio was used as a secondary measure of DNA purity to be within a range of 2.0 and 2.2. For each measurement, 1.5 µL of DNA sample were used.

1.1.7 Chemically competent *E. coli* cells preparation

E. coli NEB Stable stored in glycerol stocks were grown overnight in 5 ml of LB, then reinoculated in 50 ml of LB to a starting dilution of 1:100. Cells were grown up to OD600 = 0.5, then chilled on ice for 10 min and harvested by centrifugation at 5000 g for 10 min at4 °C. The pellet was resuspended in 15 ml of transformation buffer (10 mM Tris-HCl, pH 7.0, 50 mM CaCl2), chilled on ice for 15 min, and spun down again at 5000 g for 10 min at 4°C. The pellet was resuspended in 4 ml of transformation buffer and 20% glycerol. Cells were stored at -80°C.

1.1.8 Transformation of *E. coli* cells

One aliquot of chemically competent *E. coli* was thawed on ice and incubated for 30 min with the DNA to be transformed. Heat shock was carried out for 90 seconds at 42 °C, placed on ice for 2 min, then 800 µl of NZY were added to the cells. The culture was incubated at 37 °C with shaking at 220 rpm for 1 h.

1.1.9 Colony screening PCR (cPCR)

50 µL of sterile LB medium were pipetted into each PCR tube, and each colony grown on LB agar with the respective antibiotics was picked up through an inoculation needle into LB aliquot under the hood. The Paq 5000 polymerase PCR reaction was assembled. A 10 µL mix contained 1X Dream Taq buffer, 0.2 mM each dNTPs, 0.5 µM of each primer, 0.02 U/µl Paq5000, 1 µl of colony template. The positive samples were selected through agarose gel electrophoresis (AGE), and the DNA bands were visualized under UV light. The positive colonies were inoculated in 6 ml LB with the respective antibiotics and then incubated o.n. at 37 °C at 220 rpm in a 13 ml inoculation tube.

1.1.10 Glycerol stock preparation and plasmid miniprep

For each sample, 1 ml glycerol stock was prepared, containing 30 % glycerol and 70 % o.n. inoculum and then stored at – 80 °C. The rest of the o.n. inoculum was centrifuged at 15000 x g for 5 minutes at room temperature and the supernatant culture medium was discarded. Onto the pellet, 250 µL of solution I were added (provided with RNAse and previously stored at + 4 °C). Each pellet was resuspended through the vortex, and the cell suspension was transferred into a distinctly labelled 2 ml microcentrifuge tube. 250 µL of solution II containing the lysis buffer were added to each tube. The tubes were gently rotated several times to obtain a clear lysate. At this step, it was strictly avoided a vigorous sample agitation and incubations longer than 5 minutes. On each lysate 350 µL Solution III containing the alkaline lysis neutralization buffer were added and immediately inverted several times until a flocculent white precipitate was formed. The tubes were centrifuged at 15000 x g for 10 minutes, and the

HiBind DNA Mini Column provided from the kit was placed into a 2 ml collection tube for each sample. 100 µl 3 M NaOH were added into each HiBind DNA Mini Column and centrifuged at 15000 x g for 1 minute. The filtrate was discarded, and the columns were placed back into the respective 2 ml collection tube. The supernatants from the neutralized lysate were added into each column and centrifuged at 15000 x g for 1 minute. After discarding the filtrate, 500 µL of HBC buffer were added into each column and centrifuged at 15000 x g. Once the filtrate was discarded, this process was repeated twice with 700 µL of DNA wash buffer. Once the filter in each column was dried via centrifugation, the collection tubes were discarded, and each column was transferred into a fresh, nuclease-free 1.5 mL tube. One hundred microliters of pre-heated molecular biology grade water at 50 °C were added into each column and centrifuged at 15000 x g to elute the DNA from the binding matrix.

1.1.11 Sequencing QC

All the sequencing services were outsourced (Eurofins Genomics, Ebersberg, GER and Genewiz, Leipzig, GER), and the samples were prepared according to the respective submission guidelines.

1.1.12 DNAmate (Explora Biotech)

Each assembly process was executed using an equimolar ratio of starting plasmid mixture with 150 fmol for each plasmid sample. Twenty microliters of final total volume reaction were assembled, containing the DNA mix, 1 X T4 Ligase Buffer, 1 U T4 DNA Ligase, 10 % PEG 4000 and type IIs RE enzymes without exceeding the 10 % of total reaction volume. The reaction was run on a thermal

cycler with the following protocol: 25 cycles of 37 °C for 5 minutes, 25 °C for 10 minutes, and a final step at 80 °C for 10 min.

1.2 Results

1.2.1 Mutagenic PCR for level 0 plasmids

Due to the delivery delay for the pTrc primer set, the cloning process proceeded with the other nine promoters.

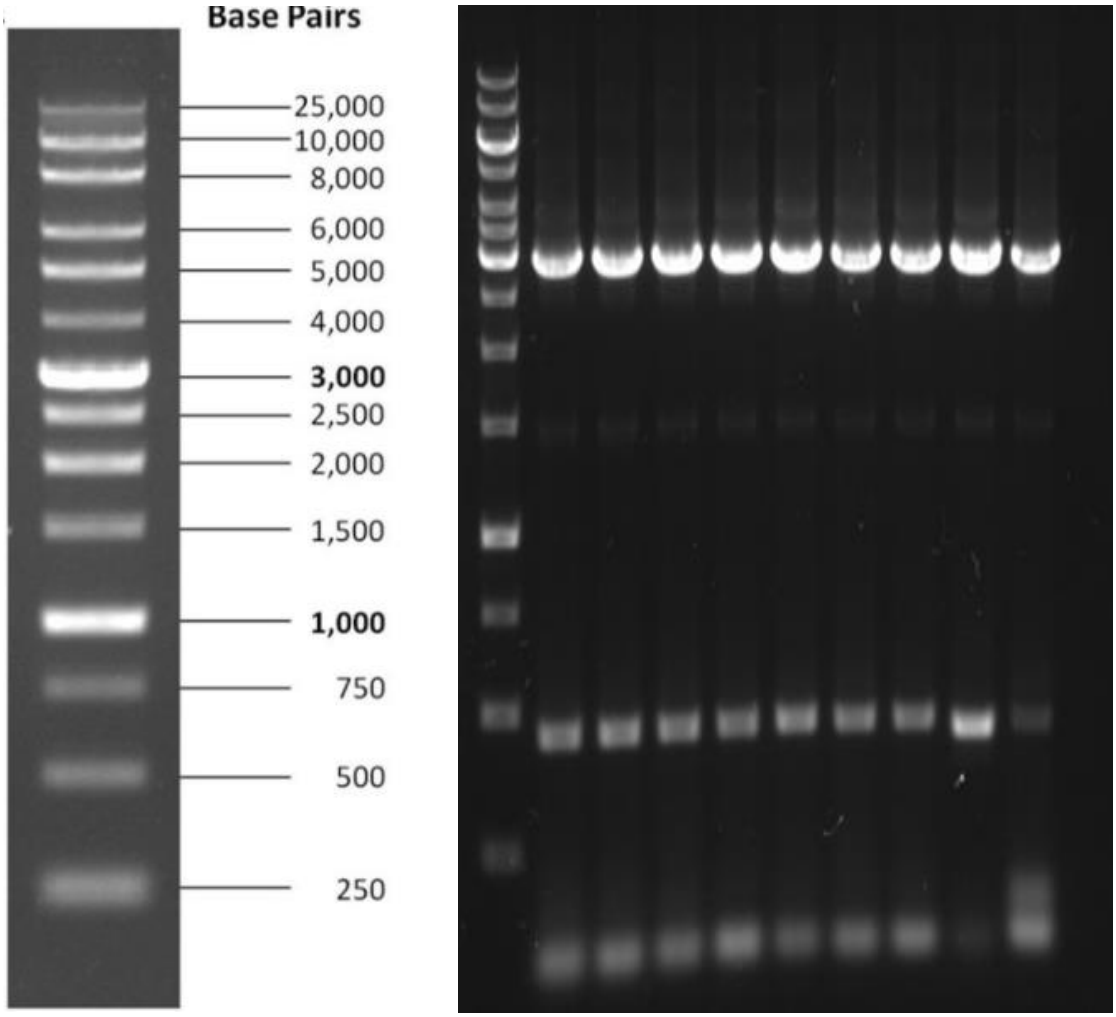

Figure 1.2.1: The marker gene ruler used for AGE (left). The AGE of mutagenic PCR reactions (Right).

The main products were seen as the most intense bands around the expected length (2950 bp). The by-products around 750 bp band were ignored. After the DpnI digestion and cleanup, the PCR reactions were used to transform *E. coli* NEB stable competent cells. After the o.n. incubation of

transformation plates, the number of colonies-per-mutagenic PCR was higher than 800.

Eight colonies from each plate were picked and analyzed through cPCR QC, and the compliant samples were purified through plasmid miniprep and verified via Sanger sequencing.

Promoter	**cPCR positive sample**	**Sequencing result**
BbaJ23100	6	Correct mutagenesis
BbaJ23110	8	Correct mutagenesis
BbaJ23114	8	Correct mutagenesis
BbaJ23119	5	Correct mutagenesis
pbla	8	Correct mutagenesis
lacIQ	8	Correct mutagenesis
placUV5	8	Correct mutagenesis
T7 promoter	8	Deletion on BsaI site
tac promoter	3	Poor sequencing quality

Table 1.2.1: overview of mutagenic PCR QC results.

The positive level 0 samples which host promoters BbaJ23100, BbaJ23110, BbaJ23114, BbaJ23119, pbla, lacIQ, placUV5 were accepted for the downstream level 1 assembly. Two cPCR QC compliant plasmid samples and the failed sequencing sample for T7 promoter were purified to execute the enzymatic digestion through the combination of Eco31I (BsaI) and BpiI restriction enzymes.

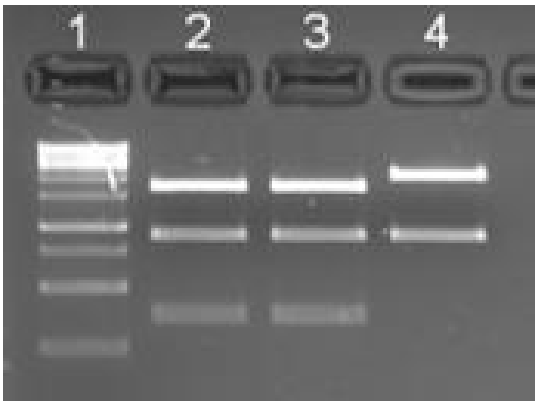

Figure 1.2.2: The BsaI and BpiI digestions AGE of level 0 plasmids with T7 promoters. The lane 4 is the sample lacking the BsaI recognition site. Lane 2 and 3 are the other two positive samples from cPCR QC.

After the digestion, the expected bands corresponded to 360 bp, 884 bp, 1692 bp, and all of them were verified according to 1.2.2 (lane 2 and 3). The failed T7 promoter level 0 plasmid, which lacks one BsaI site from table 1.2.1 was used as negative control. One of the two positive samples was sequence verified. For the ptac promoter, 14 more colonies were screened through cPCR QC, and 12 of them passed. Three ptac level 0 plasmids from positive colonies were analyzed via sequencing, and all of them contained several deletions at different places along the promoter region. The primer set for ptac was re-synthesized by EF changing the purification protocol from salt-free to HPLC to reduce the oligo manufacturing-related issues. The same cloning workflow was executed two more times. However, the cloning process issue still occurred at the sequencing QC phase with different mutation types deactivating, most of the time, the BsaI recognition site. The cloning workflow was re-executed using Q5 polymerase, which has a lower error rate (5.3E-7 [37]) compared to Phusion DNA polymerase (4.4E-7 according to manufacturer technical sheet). This cloning round had 19 out of 24 positive cPCR QC colonies, of which 3 out of 3 level 0 plasmids were sequence-verified. The pTrc was subcloned

correctly at the first cloning round using the mutagenic PCR approach with Phusion polymerase. The storing process was designed to preserve both digital and physical versions of the obtained samples. Both DNA sequence maps and the given sample ID were digitally stored among different cloud-based platforms. For each physical sample (plasmid preps and the respective glycerol stock) a specific label was attached, containing the following information: internal project ID; digital lab-book ID; strain name (if glycerol stock); sample name; plasmid amount (if plasmid prep); the sample form (solution or glycerol 30 %); the unique sample ID and barcode linked to the internal database. The whole cloning process took 13 working days for the best-case scenarios.

1.2.2 DNAmate for level 1 plasmids

DNAmate consisted of an octa-molecular assembly which involved seven parts from level 0 plasmids illustrated in Figure 1.1 and pDLX6A200 as destination vector.

Level 1 Promoter	**Total colonies**	**Screened colonies**	**cPCR QC compliant**	**Sequencing verified**
T7	48	6	1	1
BbaJ23100	25	4	1	1
BbaJ23110	81	48	1	1
BbaJ23114	99	33	1	1
BbaJ23119	50	75	4	0
pbla	60	11	3	3
lacIQ	83	11	1	1
placUV5	107	4	1	1
pTrc	82	84	0	0
ptac	26	26	2	1

Table 1.2.2: Cloning workflow overview for each level 1 promoter synthesis

According to table 1.2.2, with the first cloning attempt it was possible to obtain all the level 1 plasmids, except for BbaJ23119 and pTrc. The pTrc could be obtained with 26 more screened colonies where only one was both cPCR and sequencing verified. BbaJ23119 required 62 more screened colonies with only one being both cPCR and sequence-verified. For the level 1 plasmids, the whole cloning process took 15 working days for the best-case scenario.

1.2.3 Standard operating procedure for the cloning process

The cloning workflow executed to generate level 0 and level 1 plasmids, was divided into three main phases. The first phase was the manufacturing step of the starting DNA fragments. The second phase consisted of the assembly step and the transformation. The next phase included the quality-check phase, including the cPCR QC, (giving information about assembly outcomes) and the sequencing QC.

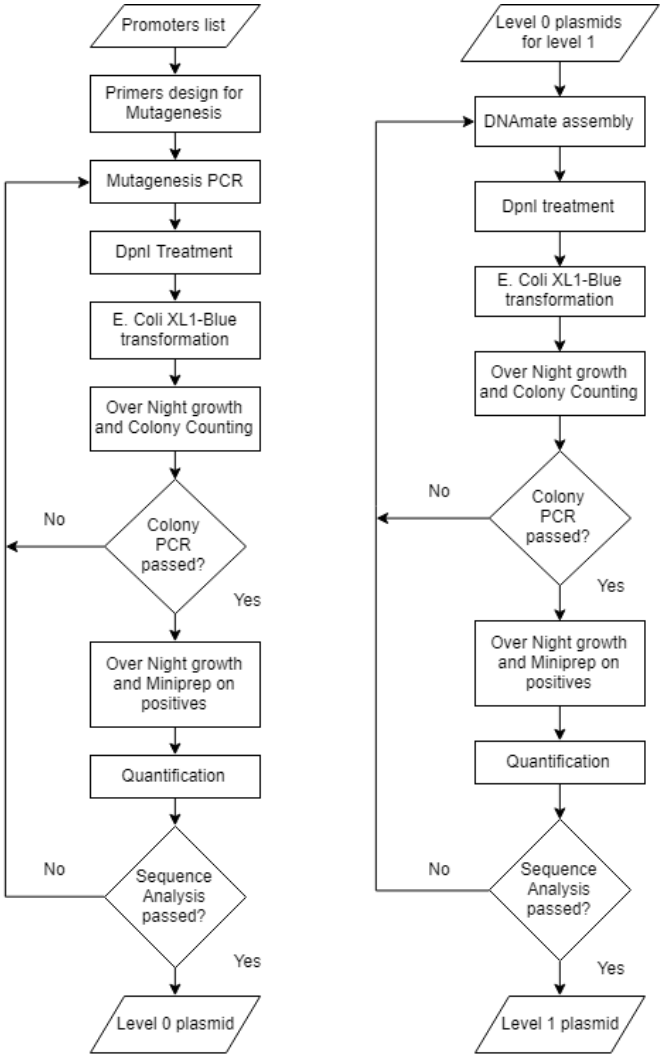

Figure 1.2.3: Schematic illustration of the assembly workflow for level 0 and level 1 plasmids.

Once the key phases of an assembly/cloning workflow were defined, every single step was formalized through the definition and release of a procedure that every single operator in the laboratory could reproduce. Every single SOP was designed to be executed independently regardless of the cloning context. In this case study, every SOP was compatible with both mutagenic PCR and DNAmate assembly techniques. At the end of this first cloning project execution, the following SOPs were designed, tested and recorded at the quality department: Preparation of RbCl ice-competent *E. coli* cells; Preparation of agar plates; Plasmid Transformation using RbCl competent *E. coli* Cells; Preparation of Antibiotics 1000X Solutions; Electrophoresis on Agarose Gel and Visualization with UV; Restriction Enzyme QC; Polymerase Chain Reaction(PCR); Restriction Digestion; DNA Ligation Reaction with 3 to 1 ratio; Oligo and Plasmids Reception and Storage Procedures; PCR Purification PROMEGA; Preparation of Phusion Green HF-CG Master Mix 2X; Sterilization using autoclave; DNAmate assembling reaction; Sequencing analysis; Preparation of culture media; Gel acquisition with Image Lab Software; inoculum and glycerol stock preparation; Sequencing QC; DNAmate plasmids nomenclature; Preparation of Paq5000 Master Mix 5X; Plasmid Extraction with OMEGA bio-tek Kit; Production of TAE solution; Production of 80% glycerol solution.

1.2.4 Workflow and SOP reproducibility test for RBS level 0 and 1 production

All level 0 and level 1 plasmids for RBS were obtained with the first workflow execution.

1.2.5 Workflow validation and analysis of weak points

The designed SOPs were validated once the cloning procedure was reproduced with the RBS cloning case. However, the workflow couldn't cover any assemblies either without DNAmate or inserts longer than 60 bp. This aspect made this workflow plasmid architecture-dependent without the complete decoupling between design and fabrication. Another weak aspect was the low assembly success rate, high production costs, and time which made the Golden Gate-based technique incompatible with any production scale-up scenario.

Chapter 2: Decoupling from starting design and combinatorial library

From chapter 1, with a DNAmate-based project it was possible to design and set SOPs and cloning workflow to manufacture DNA constructs. However, the assembly technique was plasmid architecture-dependent making the workflow not decoupled from the DNA design. To overcome this main obstacle, a restriction-ligation-free assembly technique was chosen. The Gibson Assembly® method was a cloning procedure that allowed the cloning of two or more fragments without the need for restriction enzyme digestion or compatible restriction sites. Instead, user-defined overlapping ends are incorporated into the fragments to allow the seamless joining of adjacent fragments [38]. This cloning method required overlapping regions between each DNA fragment with a length from 20 bp to 80 bp. The success of the cloning depended on several factors like secondary structures generated by each fragment, GC-content in overlapping regions, melting temperatures, and the presence of repetitive sequences in each sequence.

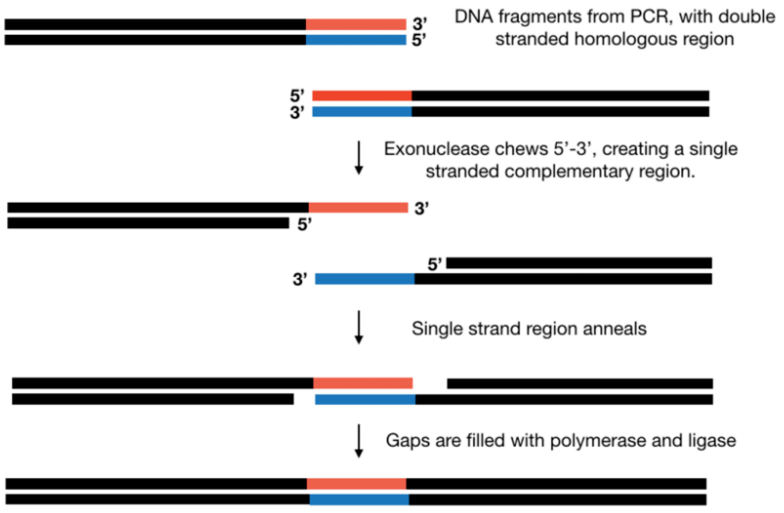

Figure 2.1: Schematic representation of key steps in a Gibson assembly 1-step reaction from SGI DNA Gibson assembly guidelines V2.

The Gibson assembly 1-step reaction was tested in several projects, which involved the execution of more than 40 cloning projects to validate this assembly technique. Another game-changer in DNA foundry was the introduction of an automated platform for dsDNA synthesis to decouple synthesis limitations imposed by external synthetic DNA providers. Several other secondary internal research was executed to define additional potential services provided by the DNA foundry, which included the combinatorial DNA library synthesis and the identification of the best protocol to achieve the highest amount of transformants. The final product for the combinatorial assembly consisted of a 328 bp dsDNA synthesis fragment to be subcloned into a custom cloning vector where the insert contained 4 NNN sites. Each "N" site had to carry all the four nucleotides' combinations (A, T, C, G) to generate at least 70 % of 412 (16.8E6) plasmids combinations. This project was divided into several phases, which consisted of the synthesis of the dsDNA fragment, assembly into the destination vector using the cloning workflow with Gibson assembly, the cPCR, and sequencing QC success rate evaluation. The starting fragment was synthesized as ssDNA, and the conversion from ssDNA to dsDNA was done using the Klenow fragment (KF) reaction as an alternative to PCR amplification. DNA Polymerase I, Large (Klenow) Fragment is a proteolytic product of *E. coli* DNA Polymerase I, which retains polymerization and 3'→ 5' exonuclease activity but has lost 5'→ 3' exonuclease activity [39]. Klenow retains the polymerization fidelity of the holoenzyme without degrading 5' termini.

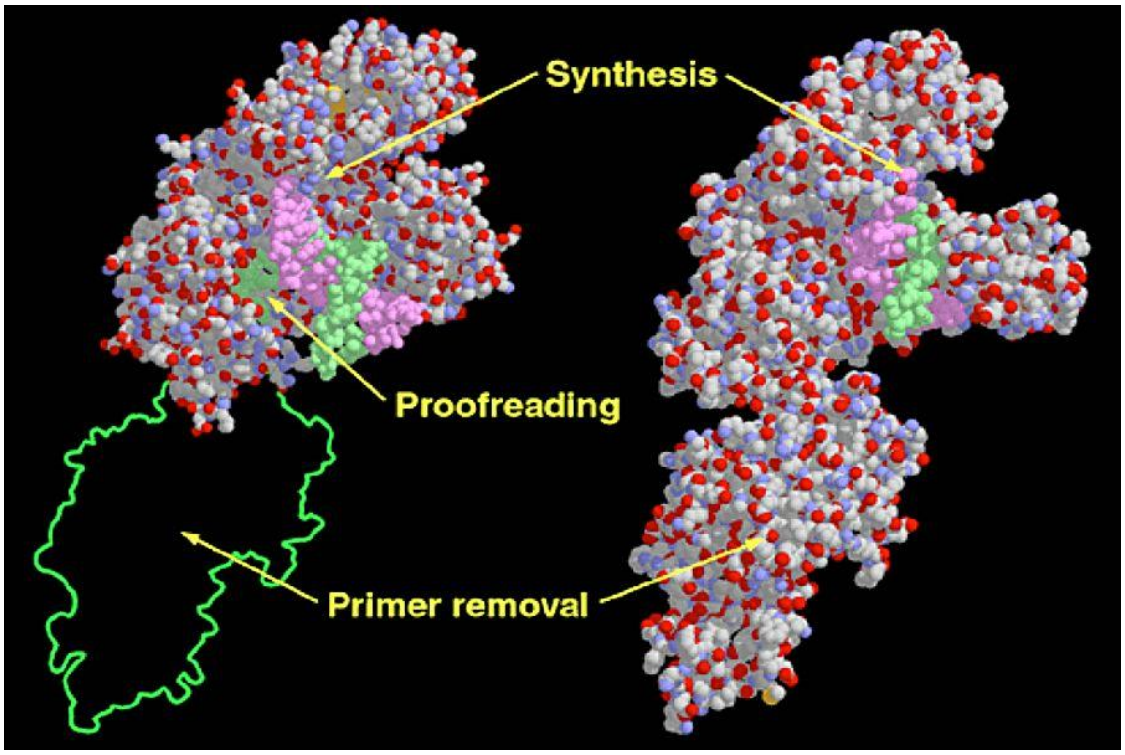

Figure 2.2: Klenow fragment on the left and DNA Polymerase I on the right.

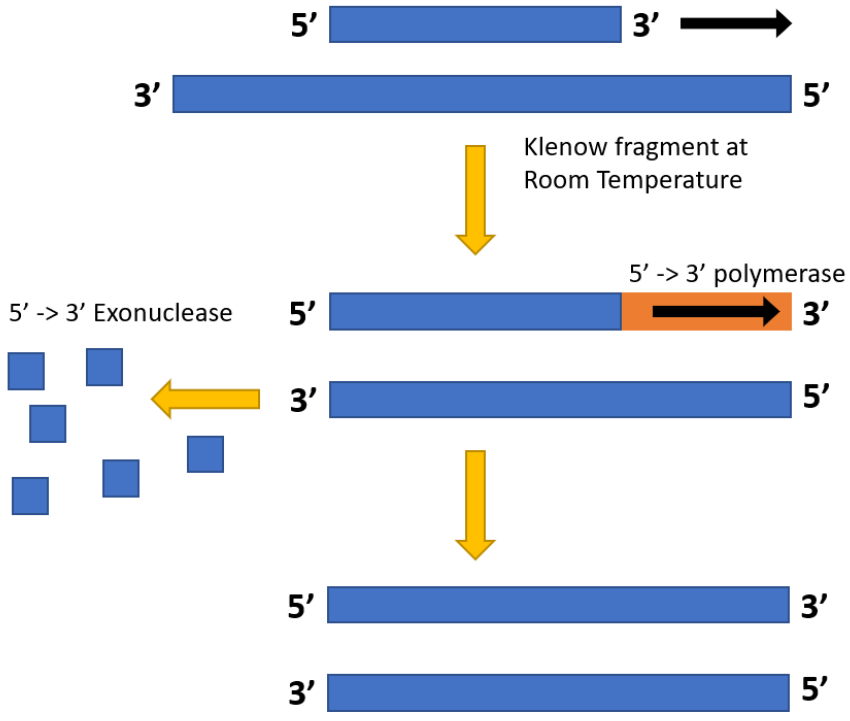

Figure 2.3: Schematic representation of KF polymerase and exonuclease activity with 2 complementary ssDNA with different size.

The PCR amplification technique was avoided due to the risk of exponentially amplifying biased nucleotide composition populations [40] to achieve an unbiased dsDNA from the ssDNA template containing the mutagenic nucleotide triplets. The Klenow fragment was used to generate dsDNA fragments through the 5'-3' polymerase activity starting with two ssDNA with a complementary overlapping region.

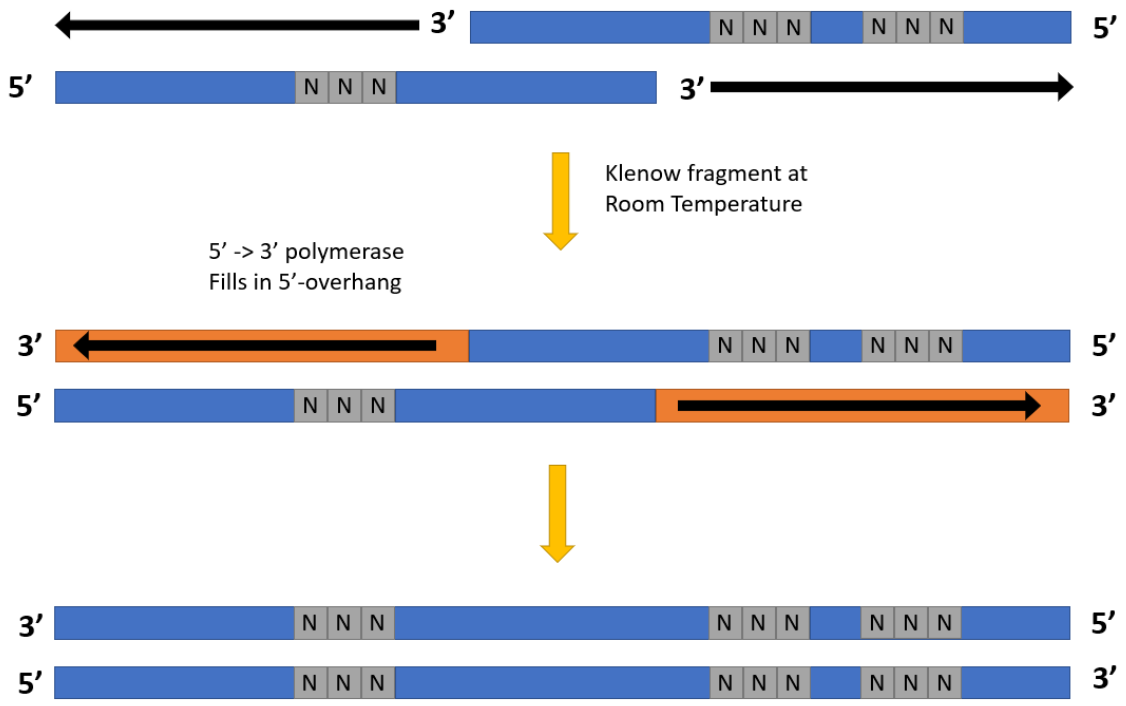

Figure 2.4: Schematic representation of KF isothermal reaction to generate the complementary DNA fragment of mutagenic library-containing ssDNA.

Once synthesized, the starting dsDNA fragments were used to execute a Gibson assembly into the collaborator's destination vector. Heat shock [41] and electroporation methods [42] were compared with several conditions, evaluating the total achieved colony numbers to identify the optimal technique for achieving the highest transformants number. At the end of this part, the workflow with Gibson assembly technique was validated in terms of decoupling the cloning method from the starting design and, moreover, a set of SOPs was drafted to generate a mutagenic library. However, due to the lack of an in-house NGS technique, it was not possible to evaluate the library coverage.

2.1 Materials and methods

2.1.1 Reagents and general supplies

All the restriction enzymes, Q5® High-Fidelity DNA Polymerase and *E. coli* NEB stable competent cells were provided from New England Biolabs; Wizard® SV Gel and PCR Clean-Up System Wizard® Plus SV Minipreps DNA Purification System were purchased from Promega; molecular biology water, LB, agar, NaOH, Phusion DNA polymerase and glycerol were provided by Sigma-Aldrich; Isopropanol, Ethanol, 50X TAE buffer was purchased from Thermo Fisher; E.Z.N.A.® Plasmid DNA Mini Kit I was purchased from Omega Biotek; Phusion polymerase was purchased from Agilent. The Gibson assembly 1-step reaction kit was provided by CODEX DNA.

2.1.2 Primer design and providers

All the primers were provided by external suppliers (Eurofins Genomics, Ebersberg, GER and Genewiz, Leipzig, GER) with salt-free oligo qualities. The primers were designed using the online tool IDT oligo analyzer (https://eu.idtdna.com/pages/tools/oligoanalyzer) with the following parameters: oligo concentration 0.5 µM, Mg++ concentration 1.5 mM, dNTPs concentration 0.2 mM. The annealing temperature was kept within a range of 55 °C – 59 °C, and the secondary structure melting temperature was < 50 °C. Primers for mutagenic PCR were designed to keep the same criteria as the normal primers at the annealing region between primer-to-template and primer-to-primer overlapping zone.

2.1.3 Synthesis fragments design and synthesis

The synthetic linear DNA fragments were provided by Eurofins, Twist, Explora Biotech. Each fragment was designed to carry 20 bp – 80 bp of the homologous overlapping region with the neighbor fragment at each end. Each end was properly designed to avoid stable secondary structures at 50 °C.

2.1.4 Q5 PCR reactions

All PCR was executed using Phusion polymerase. A 50 µL mix contained 1X Q5 reaction buffer, 0.2 mM each dNTPs, 0.5 µM of each primer, 0.02 U/µL Q5 polymerase, ~ 2 ng/µL plasmid DNA template. The reaction ran on a thermal cycler with the following protocol: 98°C for 2 min for initial denaturation, 35 cycles of 98 °C for 10 s, annealing temperature for 10 s and 72 °C for 20 s/kb and a final extension step at 72 °C for 10 min. PCR fragments were treated with 0.4 U/µl of DpnI for at least 1 h at 37 °C to remove the starting DNA template. Regardless of the PCR reaction, it was used at 57 °C as annealing temperature.

2.1.5 PCR reaction cleanup

An equal volume of Membrane binding solution was added into each PCR reaction tube. SV columns provided by the OMEGA kit were placed into a collection tube for each PCR reaction. The PCR product and binding solution mix were added into each SV column and incubated at 1 minute at room temperature. After centrifugation at 15000 x g, the filtrate was trashed, and 700 µl of Membrane wash solution was added to the column.

After another centrifugation at 15000 x g for 1 minute and once removed the filtrate, this washing step was repeated with 500 µl of the same wash solution. The membranes in the columns were dried through centrifugation at 15000 x g for 1 minute, and the collection tubes were discarded. Each column was placed into a 1.5 mL microcentrifuge tube, and 100 µl of pre heated molecular biology grade water at 50 °C were added into each tube. After the last centrifugation at 15000 x g, the DNA solution was recovered and used for downstream processes.

2.1.6 DNA quantification using Nanodrop 2000

The DNA concentration was measured by a nanodrop spectrophotometer measuring the absorbance at 260 nm as a triplicate for each same sample. The ratio of absorbance at 260 and 280 nm was used to check the DNA purity. A ratio of ~1.8 was accepted for this assessment. The 260/230 was used as a secondary measure of DNA purity to be within a range between 2.0 and 2.2. For each measurement, it was used 1.5 µL of DNA sample.

2.1.7 Chemically competent *E. coli* cells preparation

E. coli NEB Stable stored in glycerol stocks were grown overnight in 5 ml of LB, then reinoculated in 50 ml of LB to a starting dilution of 1:100. Cells were grown up to OD600 = 0.5, then chilled on ice for 10 min and harvested by centrifugation at 5000 g for 10 min at 4 °C. The pellet was resuspended in 15 ml of transformation buffer (10 mM Tris-HCl, pH 7.0, 50 mM CaCl2), chilled on ice for 15 min, and spun down again at 5000 g for 10 min at 4°C. The pellet was resuspended in 4 ml of transformation buffer and 20% glycerol. Cells were stored at -80°C.

2.1.8 Transformation of chemically *E. coli* cells

One aliquot of chemically competent *E. coli* was thawed on ice and incubated for 30 min with the DNA to be transformed. Heat shock was for 90 seconds at 42 °C, placed on ice for 2 min, then 800 µl of NZY were added to the cells. The culture was incubated at 37 °C with shaking at 220 rpm for 1 h. Each transformation reaction was plated in 2 different Petri dishes (100 µl and 900 µl of inoculation volumes) provided with the corresponding antibiotic resistance in solid LB agar medium.

2.1.9 Transformation of electro competent *E. coli* cells

Empty PCR tubes and Bio-Rad empty cuvettes for electroporation were placed in a 0°C cooler for small tubes. Commercial tubes of competent cells with 25 ul of E. coli SS320 electrocompetent and ElectroMAX™ Stbl4™ Competent Cells were placed on ice to be thawed. Bio-Rad electroporator device was plugged into the power supply. The pre-chilled PCR tubes were labeled, 1 ul of DNA assembly reaction was added and gently mixed with electrocompetent cells. The cell + DNA mixture was added carefully into each cuvette (without introducing bubbles), and they were placed on ice as soon as each of them was filled. 13 ml inoculation tube per reaction was prepared with 974 ul of SOC medium. The cuvettes were placed into the electroporation chamber and electroporated using 1.8 KV. Twenty-six microliters of electroporated cells were added into 13 ml tubes containing SOC medium. The whole electroporation step was strictly executed within 10 seconds, keeping the cuvette as cold as possible. Each inoculation tube was placed at 37°C in agitation at 250 rpm for 1 hour. The inoculation volume was diluted 1:10 six times, and each 100 µl of each dilution was

plated on Petri dishes provided with the corresponding antibiotic resistance in solid LB agar medium.

2.1.10 Colony screening PCR (cPCR)

50 µL of sterile LB medium were pipetted into each PCR tube, and each colony grown on LB agar with the respective antibiotics was picked up through an inoculation needle into LB aliquot under the hood. The Paq 5000 polymerase PCR reaction was assembled. A 10 µL mix contained 1X Dream Taq buffer, 0.2 mM each dNTPs, 0.5 µM of each primer, 0.02 U/µl Paq5000, 1 µl of colony template. The positive samples were selected through agarose gel electrophoresis (AGE), and the DNA bands were visualized under UV light. The positive colonies were inoculated in 6 ml LB with the respective antibiotics and then incubated o.n. at 37 °C at 220 rpm in a 13 ml inoculation tube.

2.1.11 Glycerol stock preparation and plasmid miniprep

For each sample, 1 ml glycerol stock was prepared, containing 30 % glycerol and 70 % o.n. inoculum and then stored at – 80 °C. The rest of the o.n. inoculum was centrifuged at 15000 x g for 5 minutes at room temperature and the supernatant culture medium was discarded. Onto the pellet, 250 µL of solution I were added (provided with RNAse and previously stored at + 4 °C). Each pellet was resuspended through the vortex, and the cell suspension was transferred into a distinctly labelled 2 ml microcentrifuge tube. 250 µL of solution II containing the lysis buffer were added to each tube. The tubes were gently rotated several times to obtain a clear lysate. At this step, it was strictly avoided a vigorous sample agitation and incubations longer than 5 minutes. On each lysate 350 µL Solution III

containing the alkaline lysis neutralization buffer were added and immediately inverted several times until a flocculent white precipitate was formed. The tubes were centrifuged at 15000 x g for 10 minutes, and the HiBind DNA Mini Column provided from the kit was placed into a 2 ml collection tube for each sample. 100 µl 3 M NaOH were added into each HiBind DNA Mini Column and centrifuged at 15000 x g for 1 minute. The filtrate was discarded, and the columns were placed back into the respective 2 ml collection tube. The supernatants from the neutralized lysate were added into each column and centrifuged at 15000 x g for 1 minute. After discarding the filtrate, 500 µL of HBC buffer were added into each column and centrifuged at 15000 x g. Once the filtrate was discarded, this process was repeated twice with 700 µL of DNA wash buffer. Once the filter in each column was dried via centrifugation, the collection tubes were discarded, and each column was transferred into a fresh, nuclease-free 1.5 mL tube. One hundred microliters of pre-heated molecular biology grade water at 50 °C were added into each column and centrifuged at 15000 x g to elute the DNA from the binding matrix.

2.1.12 Sequencing QC

All the sequencing services were outsourced (Eurofins Genomics, Ebersberg, GER and Genewiz, Leipzig, GER), and the samples were prepared according to the respective submission guidelines.

2.1.13 Gibson Assembly (CODEX DNA)

Each assembly process was executed using an equimolar ratio with 50 fmol for each DNA linear fragment. Ten microliters of the final total volume contained the DNA mix, 5 µl of Gibson assembly 1-step reaction enzyme mix. The incubation occurred at 50 °C for 1 hour.

2.1.14 Cloning projects with Gibson assembly

All the analyzed projects were taken from orders placed from European projects MIAMI and Syn4flav consortium as case studies. From Syn4flav EU project thirteen subcloning jobs where eight in pET-15b destination vector and five into pDLX1A100 standard vector were requested. From the MIAMI consortium was requested twenty-four cloning jobs into a customized destination vector. The cloning projects were referred to by a code: “S4F” for Syn4flav cloning projects and “MIA” for projects from the MIAMI consortium. As preliminary cloning outcomes, it was considered eight pET-15b based assemblies from Syn4flav as case studies taking into account the assembly complexity the assembly order (number of DNA fragments involved in an assembly including the backbone), the total colonies obtained after transformation, samples which passed cPCR QC and the sequencing QC. In a later stage, the same workflow with the same fragments design criteria, the enzyme was reproduced with the other projects. A custom project (CS_Project) with high assembly order was taken into account to identify manufacturing difficulties related to the assembly order increase.

2.1.15 In-house produced DNA fragments

The synthesis of double-stranded DNA fragments was performed, introducing an automated platform infrastructure at Explora biotech srl. At this stage of the manufacturing laboratory, the Gibson assembly method was validated as the standard technique, and the pilot projects for synthetic fragments analysis were mainly from private researchers and the Syn4flav consortium.

2.2 Results

2.2.1 Case study: pET-15b based assemblies

For this preliminary study, all the DNA fragment synthesis were outsourced. Once executed the Gibson assembly reaction, 1 µl of each reaction was used to transform *E. coli* NEB stable, competent cells. From this point, the whole cloning workflow was executed according to figure 1.2.3.

Assembly ID	**Assembly order**	**Colonies**	**Screened colonies**	**Passed Screenings**	**Sequenced**	**Obtained Plasmids**
S4F_01	5th	42	8	3	3	2
S4F_02	4th	51	8	3	3	0
S4F_03	4th	53	8	4	1	1
S4F_04	4th	32	15	1	2	0
S4F_05	4th	42	15	4	2	2
S4F_06	4th	62	15	2	2	1
S4F_07	4th	35	15	2	2	0
GA Background	--	253	--	--	1	--
DpnI Background	--	0	--	--	--	--

Table 2.2.1: Cloning outcome overview for pilot assemblies with Gibson assembly with outsourced synthesis fragments. As Gibson assembly background the DpnI digested backbone with the enzyme mix was used. As DpnI background, it was used the same sample but without the Gibson assembly mix.

The overall colony number was comparable regardless of the assembly complexity. However, only three out of eight cloning projects were successful. One colony from the Gibson assembly background was sequenced, showing a reclosure issue of the backbone. If the Gibson assembly reaction occurs with the backbone alone, the chew back process by exonucleases continues until the two single-stranded ends containing at least nine base pairs as homologous regions caused the backbone rearrangement. This phenomenon was seen in a 56 bp region before the 3' of the backbone and 112 bp after the 5' of the backbone. The transformants of DpnI treated backbone without Gibson assembly mix were zero, indicating the absence of any template used to generate the pET-15b backbone. All the failed plasmid sequencing results showed several mutation spots within the synthesized fragment, suggesting the poor quality of the starting synthetic fragments.

2.2.2 Gibson assembly-based workflow execution with other projects

The same workflow from the previous part was executed for the other assembly projects from Syn4flav, MIAMI, and the custom project.

Project ID	**Assemblies**	**Backbone**	**Assembly order**	**Achieved**	**Project completion**
S4F_Project_01	8	pET-15b	4°/5°	8	63 %
S4F_Project_02	5	pDLX1A100	2°	5	100 %
MIA_Project_01	24	Custom	3°	21	88 %
CS_Project_01	3	Custom	6°	2	67 %

Table 2.2.2: Project completion overview using Gibson assembly 1-step reaction and outsourced synthesis fragments.

For the projects S4F_Project_01 and CS_Project_01 for their higher assembly order, it was not possible to complete the projects with a project completion rate of 50 % and 67 %. Only the project S4F_Project_02, which had a low assembly order, was 100 % completed. The three failed assemblies from MIA_Project_01 were hypothesized to be caused by the toxicity of the inserts, which upon transformation led to a low growth rate of the colonies. None of the surviving colonies passed the sequencing QC even after an in-depth troubleshoot.

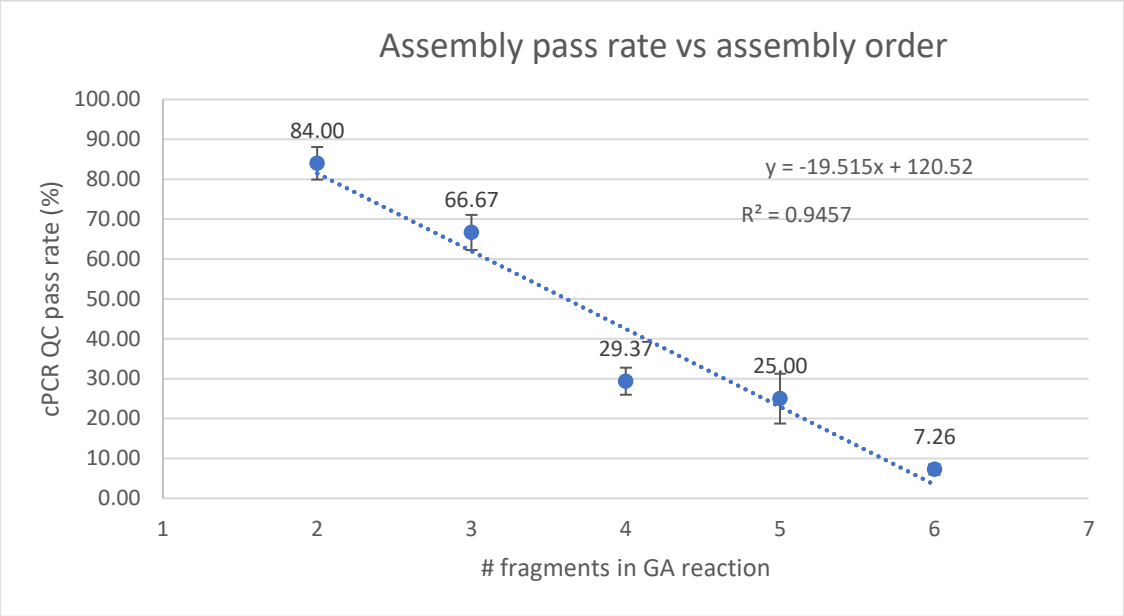

Table 2.2.3: Regression analysis of cPCR QC pass rate increasing the assembly order.

Increasing the assembly order caused two main obstacles: the larger number of DNA fragments involved for each reaction made the plasmid rearrangement rate higher; the longer DNA sequence in combination with DNA synthesis mutation rate made it harder to achieve sequence-verified samples. The plasmid rearrangement results were captured by cPCR QC colony screening and by comparing the correctly synthesized DNA lengths with the mutations seen in the sequencing data. The mutation rate was estimated to be 1 nucleotide each 10E3 bp. With the Gibson assembly technique, it was possible to synthesize plasmids decoupled from the design of the plasmids. However, the mutation rate was incompatible for a scaled-up synthesis scenario and the delivery time was still higher than 20 working days.

2.2.3 In-house produced linear DNA fragments

The synthesis of double-stranded DNA fragments was performed, introducing an automated platform at Explora biotech srl to solve the high mutation rate of synthetic DNA fragments and to produce the linear synthetic DNA in-house. The cloning workflow with Gibson assembly 1-step reaction was executed using the in-house produced starting DNA fragments to evaluate the quality of the fragment.

Project ID	**Total assemblies**	**Backbone**	**Assembly order**	**Project completion**
CS_Project_02	4	pUC19	2°	100 %
S4F_Project_03	2	pDLX1A100	2°	100 %
S4F_Project_04	10	pSEVA181	3°	100 %
S4F_Project_05	5	pDLX1A100	3°	100 %
CS_Project_03	5	Custom	4°	100 %
CS_Project_04	6	Custom	2° and 3°	100 %

Table 2.2.4: Projects analyzed from in-house produced fragments quality assessment.

All the assemblies at this preliminary stage were achieved and reviewed the sequencing data. The mutation rate changed from one nucleotide each 10E3 bp to 10E4 bp. As a result, the synthesis of in-house produced fragments took seven working days less than the external provider's fragments. With the faster synthesis time and the lower mutation rate, the whole manufacturing process moved from 20 – 23 working days down to 13 working days.

2.2.4 Workflow evolution and new SOPs implementation

The workflow in figure 1.2.3 evolved, setting Gibson assembly as the official assembly technique using the in-house produced fragments by these preliminary experiments. As a result, the following SOPs have been released: Gibson assembly; design and synthesis of DNA fragments for Gibson assembly. The implementation of the in-house synthesis of linear dsDNA and the scarless assembly technique decoupled DNA constructs by the starting design. Other minor changes were implemented on the previously released SOP versions (from three to two technical replica for each sample measurements; fixed DMSO concentration and polymerase set for each PCR reaction; fixed PCR reaction thermocycler conditions; less expensive material alternatives in some plasticwares-consuming steps). The next step was identifying synthesis limits of in-house produced fragments with the production scale-up.

2.2.5 Case study: streamline of mutagenic library synthesis project

Since the template included a high overall GC content (64.2 %), the PCR reactions were performed with the following conditions: Q5 polymerase (sample 1), Q5 with DMSO 3% (sample 2), Q5 with GC-enhancer buffer (sample 3), and Q5 with GC-enhancer buffers with DMSO 3% (sample 4). The same thermal cycler condition was used for all reactions.

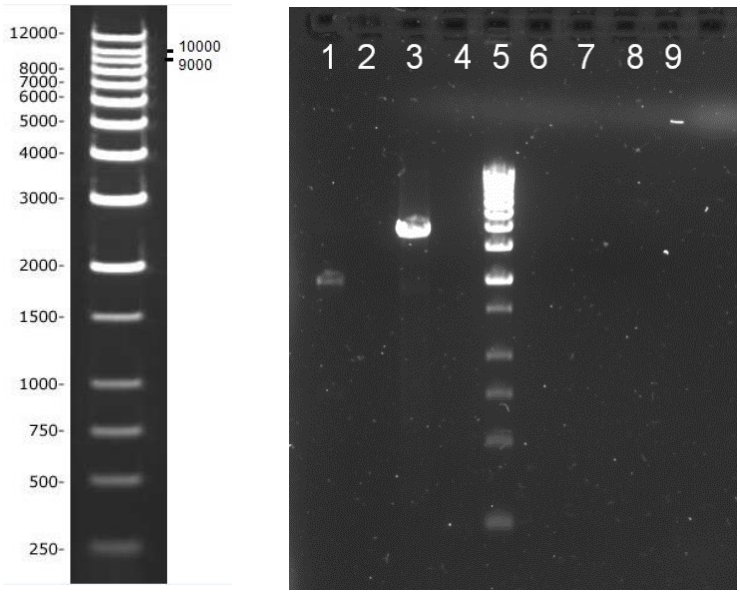

Figure 2.2.1: The marker gene ruler used for AGE (left). The AGE of mutagenic PCR reactions (Right). The lane 1, 3, 6 and 8 corresponds to the sample from 1 to 4. The lane 2, 4, 7 and 9 corresponded to each PCR master mixes without template as negative controls.

The expected band was 3817 bp, and only the reaction with Q5 and DMSO 3% worked adequately. Sample 2 was treated with DpnI enzyme, purified, and quantified with nanodrop according to the SOPs.

Sample name	**A260**	**260/280**	**260/230**	**Conc. (ng/uL)**	**MW (KDa)**	**µM (fmoles/µL)**
m.b. water	-0.02 (± 0.05)	0.52 (± 0.04)	0.33 (± 0.05)	-0.75 (± 0.93)	--	--
Custom Backbone (Sample 2)	1,17 (± 0.01)	1.94 (± 0.01)	1.14 (± 0.02)	58.75 (± 0.52)	2358.72	24.90

Table 2.2.5: nucleic acid quantification nanodrop data of PCR reaction. The measurements were performed on the same sample three times.

The insert was divided into two parts (F1 and F2), and the respective ssDNA fragments were synthesized and resuspended.

Sample name	KDa (ssDNA)	Amount	Added mb water (ul)	Final concentrartion (ng/ul)
ssDNA_F1	61,38	120 ug	400,00	300
ssDNA_F2	61,94	110 ug	366,67	300

Table 2.2.6: ssDNA fragments synthesized amount and the resuspension volume.

2.5 µM of each ssDNA was used as primers and templates for PCR reaction using Q5 polymerase (ssDNA-based PCR). 1 ng of ssDNA templates was used for classic PCR using Q5 polymerase including two oligonucleotides as primers in the reaction. Were used in 32.5 micromolar concentration of each ssDNA fragment was used for the Klenow reaction described in Figure 2.4. Each reaction was executed in triplicate to have enough yield for downstream applications. Each triplicate was merged before the purification phase and eluted in 50 µl m.b. water.

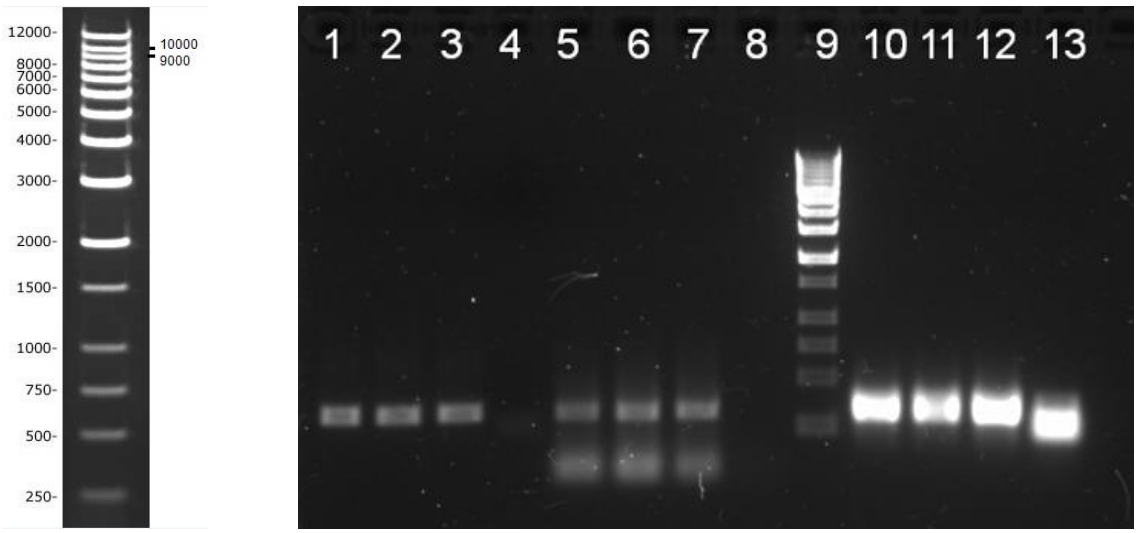

Figure 2.2.2: The marker gene ruler used for AGE (left). The lanes 1,2,3 correspond to ssDNA-based PCR samples. The lanes 5,6,7 corresponds to classic PCR reactions. The lane 10, 11, 12 corresponds to Klenow reaction. The lane 4 was the negative control containing only the starting ssDNA templates. The lane 8 was the negative control containing the PCR reaction without any primer. The lane 13 was the negative control containing only the ssDNA mix used for Klenow reaction. The starting template lengths corresponded to 200 bp and the expected band was 328 bp.

Sample name	A260	A260 % intra-CV	260/280	260/230	Conc. (ng/uL)	MW (KDa)	µM (pmoles/µL)
Mb water	0,01	47,14	-18,58	4,11	0,45	--	--
Symm_PCR_Insert	0,62	2,06	1,89	1,21	30,85	202,77	0,15
PCR_Insert	0,75	2,56	1,92	1,54	37,30	202,77	0,18
Klen_PCR_Insert	4,17	0,93	1,73	1,96	208,55	202,77	1,03

Table 2.2.7: nanodrop data of symmetric PCR, classic PCR and Klenow reaction

The classic PCR approach was discarded since it has generated by-products. The ssDNA-based PCR was discarded due to the low yield, considering the material loss during any purification phase. The Klenow reaction was accepted for both AGE QC compliance and the material yield compatibility for downstream purposes. The purified dsDNA generated through Klenow reaction was used for 3 Gibson assembly assemblies changing the ratio between the destination vector and the purified insert as follows: 1:1, 1:3, and 1:5 ratios. The backbone amount was kept fixed at 50 fmoles. The NEB stable chemically competent cells were transformed through heat shock, and the colonies were counted.

Sample	**Plated volume**	**Counts**
empty	--	0
NZY transformation media	1 ml	0
Negative ctrl	1 ml	0
Positive control	900 ul	6012
	100 ul	668
Ratio 1:1	900 µl	19036
	100 µl	2204
Ratio 1:3	900 µl	2016
	100 µl	186
Ratio 1:5	900 µl	40896
	100 µl	3995

Table 2.2.8: Colony count data after transformation through heat shock

According to the results in table 2.2.8, the highest colony amount was reached with a 1:5 ratio. However, 4.1E4 as the total colony number was insufficient to cover 70% of 16.8E6 plasmids combinations. Therefore, the electroporation was performed using the 1:5 ratio Gibson assembly reaction to increase the number of transformants. The reaction was dialyzed to remove all the salts. The following conditions were used for the electroporation: 1.8 KV, 0.1 cm cuvette, and SS320 from Lucigen Corporation as electro competent cells. The transformation experiment was repeated seven times, and the cell number was estimated, counting the total colonies on the plate in the last two dilutions (10E-3; 10E-4).

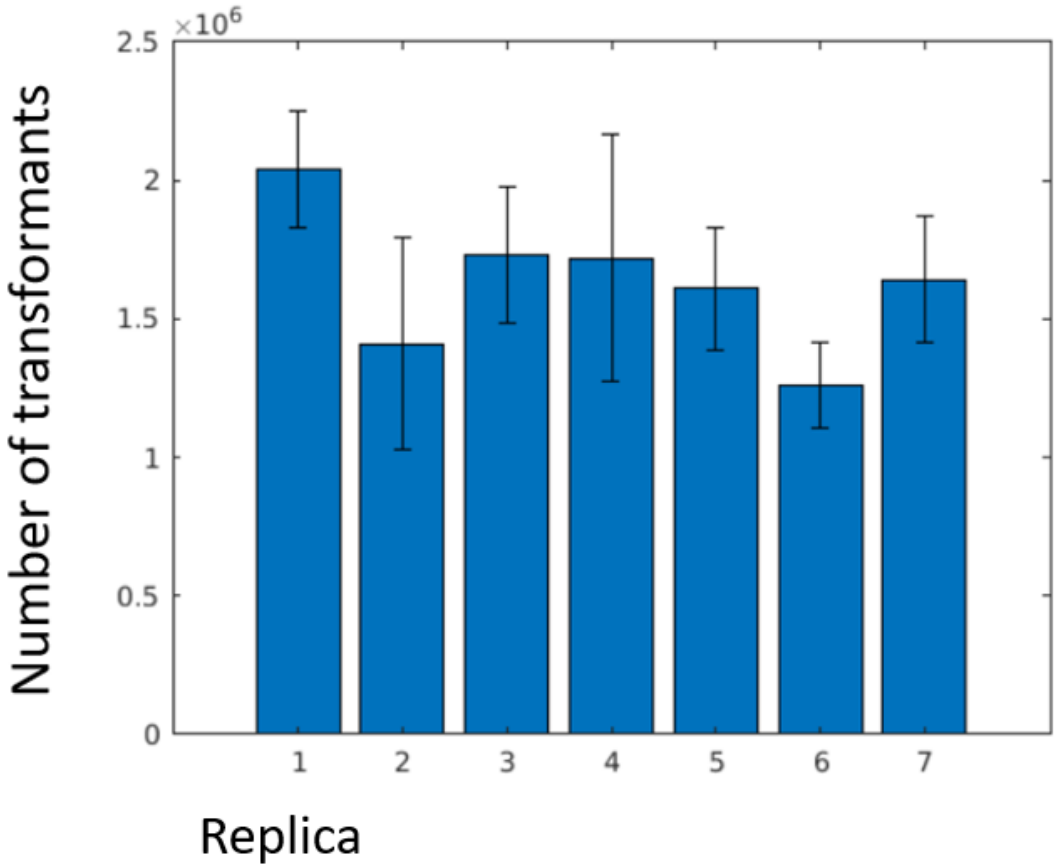

Figure 2.2.3: Colony count data of cells transformed through electroporation.

The colony amount was two orders higher than the colony number obtained with the heat shock, and it was compatible with the purpose of this project. The colony screening QC gave a 100 % assembly pass ratio on 20 colonies.

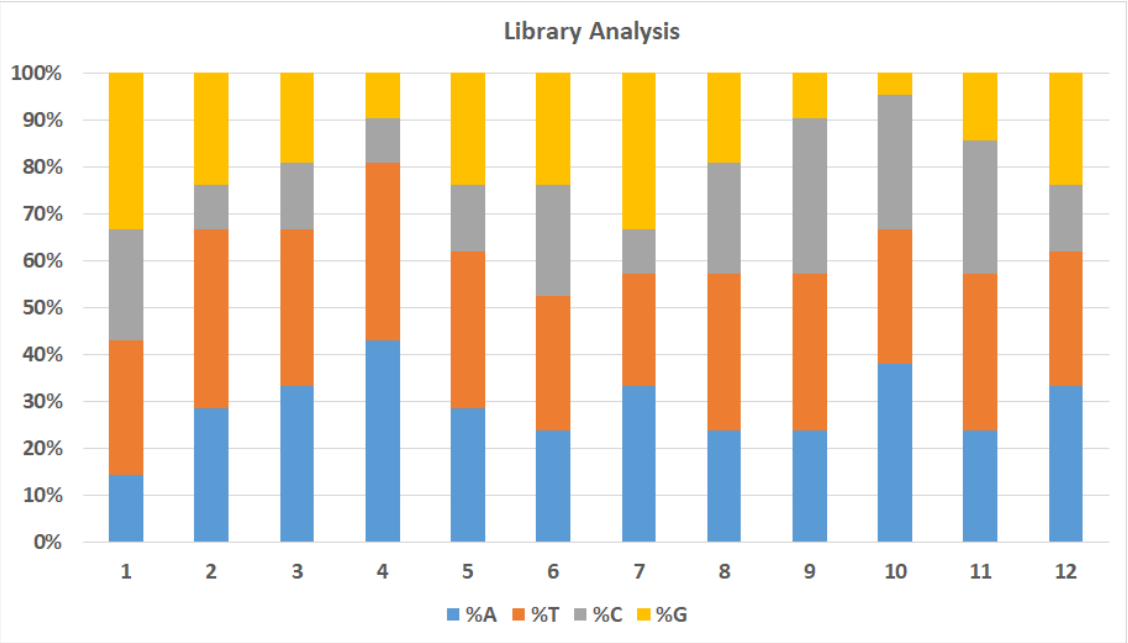

Figure 2.2.4: Nucleotide percentage of occupation on each "N" position of the final plasmid. The data was estimated analyzing Sanger sequencing data of 40 samples.

Through 40 samples analyzed with the sanger sequencing, the nucleotide occupation on each "N" position on the final constructs was verified. The collaborator accepted the achieved quality check data and results for its downstream processes. Both experimental workflow and the synthesis approaches were accepted as the manufacturing method for degenerated libraries. However, optimizing this workflow requires integrating an NGS sequencing technique to estimate the library coverage.

Chapter 3: Decoupling from starting design and combinatorial library

Compared to DNA fragments produced by external providers, the DNA fragments built in-house dramatically reduced the project failures related to unwanted mutations of starting fragments. After one year of implementing Gibson assembly and in-house produced fragments, 109 cloning projects were executed involving the synthesis of 243 synthesis fragments. However, 29 fragment synthesis failed for the following reasons: empty products, smears, by-products, multiple bands in AGE without the expected band. These fragments were analyzed to determine manufacturability limitations depending on the fragments' sequences. Other aspects were still required for a manufacturing laboratory with a higher synthesis volume of DNA constructs which included: defined quality check scoring systems, a system to assess the fragments' recyclability; troubleshooting plans; mitigation plans. A quantitative scoring system was required to set a predefined group of rules to identify non-compliance with fixed conditionals. In 2020, 99 projects required the recycling of already-produced plasmid portions. During any project design step, it was crucial to implement an automatized system to aid the operator in analyzing, align and identifying any reusable sequences. The previous project's non-compliances made it possible to identify the main reasons for any cloning workflow execution failure designing plans depending on the issue severity grade and type. A set of mitigation plans was formalized to prevent future manufacturing projects' potential issues by a more significant amount of design scenarios and outcomes. By the start of 2021, the definitive workflow and infrastructure were implemented, and the success rate of cloning projects was used as the key performance indicator for manufacturing workflow and infrastructure. The number of customized

projects per year was considered to evaluate the decoupling from starting plasmid designs.

3.1 Materials and methods

3.1.1 Reagents and general supplies

All the restriction enzymes, Q5® High-Fidelity DNA Polymerase, *E. coli* NEB stable competent cells were provided from New England Biolabs; Wizard® SV Gel and PCR Clean-Up System Wizard® Plus SV Minipreps DNA Purification System were purchased from Promega; molecular biology water, LB, agar, NaOH, Phusion DNA polymerase, Glycerol were provided by Sigma-Aldrich; Isopropanol, Ethanol, 50X TAE buffer was purchased from Thermo Fisher; E.Z.N.A.® Plasmid DNA Mini Kit I was purchased from Omega Biotek; Phusion polymerase was purchased from Agilent. The Gibson assembly 1-step reaction kit was provided by CODEX DNA.

3.1.2 In-house produced fragments quality check

The produced fragments were checked through qualitative and quantitative quality check steps. The qualitative quality checks were performed through AGE loading 5 µl of product in the agarose gel 1 % with 1 µl of loading dye with 120 V for 30 minutes. The following quality check scoring system was used to identify compliant and non-compliant synthesis fragments.

Synthesis fragment AGE result	Score	Gel description
Empty or smear	0	Empty well or smear
Compliant	1.00	One intense band of the expected product
MB_main/ byprod	0.5	Expected product and by product with the same band intensity
MB_main	0.75	Expected product with high band intensity and by-product with low band intensity
Low_yield	0.75	Expected product with low yield
MB_byprod or Non-compliant	0	Multiple bands of unwanted products or an unwanted product

Table 3.1.1: Scoring system of produced synthesis fragments AGE QC assessments. A = expected product with high yield; a = expected product with low yield; B = by product with high yield; b = by product with low yield; Ø = empty; SS = smear.

The quantitative QC was assessed by TECAN spectrophotometer measuring the absorbance at 260 nm as a duplicate for each same sample. The ratio of absorbance at 260 and 280 nm was used to check the DNA purity. A ratio of ~1.8 was accepted for this assessment. The 260/230 was used as a secondary measure of DNA purity to be within a range between 2.0 and 2.2. For each measurement, 2 µL of DNA sample were used. The same scoring criteria were used for any other PCR AGE QC analysis.

3.1.3 Synthesis fragment analysis

The synthesized fragments were analyzed using Geneious Prime. The GC content was analyzed as: GC% = Count(G + C)/Count(A + T + G + C) * 100 [%]. The repeats were analyzed using the repeat finder plugin with a minimum repeat length of 14 bp and maximum mismatches of 0 %. The GC content graph function analyzed the GC distribution along each sequence with the sliding window size of 9 bp to identify local high/low GC% clusters.

3.1.4 Recyclability analysis

Since all produced DNA molecule sequences were digitally stored, it was necessary to find an automatized system to find any recyclable DNA sequence portion given an entire DNA construct as input. This part was implemented exploiting the pre-existing annotation function in Geneious software, integrating three different sequence databases.

3.1.5 Troubleshoot integration into manufacturing workflow

In 2020 projects outcomes, 79 out of 109 cloning projects required at least one interruption of the regular execution of manufacturing workflow caused by any trouble entity. The lack of formalized approach to solving these issues has been caused by avoidable project delivery delays and extra synthesis costs in 2019. Therefore, the troubleshoot severity grade was divided into three tiers.

Troubleshoot Tier	**Delay on workflow**	**Responsible approval required**	**Internal research required**
T1	0 – 2 wds	N	N
T2	5 - 10 wds	Y	N
T3	> 10 wds	Y	Y

Table 3.1.2: Troubleshoot tiers classification depending on trouble severity.

Tier 1 includes all the issues that can be solved by just increasing the sample population to be analysed at the critical QC step without any approval from the unit responsible. Tier 2 covers any scenario which involves the re-execution of the whole workflow and includes either the correction of achieved sample mutations through PCR-based techniques or assembly technique and condition optimization. Finally, the T3 tier requires an in-depth analysis of the whole cloning project involving the knowledge of multiple researchers to design a tailored research project. Each troubleshoot entity was designed a dedicated set of conditional workflows to solve most of the troublesome scenarios.

3.2 Results

3.2.1 Synthesized fragments overview

After the implementation of the manufacturing workflow, 243 synthesis fragments were synthesized. However, the synthesis of 29 fragments failed, obtaining a score of 0.

ID#	Length (bp)	Repeats	GC%	AGE score	Note
001	805	none	47.10%	1	Compliant
002	987	none	50.50%	1	Compliant
003	1274	1x 14 bp direct repeat	75.20%	1	Compliant
004	536	none	52.40%	1	Compliant
005	451	none	50.60%	1	Compliant
006	706	none	54.20%	1	Compliant
007	611	none	53.40%	1	Compliant
008	1314	none	38.60%	1	Compliant
009	1528	none	50.90%	1	Compliant
010	1120	1x 15 bp inverted repeat	51.20%	0.75	MB_main
011	1194	none	49.80%	1	Compliant
012	920	none	51.80%	1	Compliant
013	863	none	38.90%	1	Compliant
014	550	none	50.50%	1	Compliant
015	842	none	51.20%	1	Compliant
016	1996	none	50.50%	1	Compliant
017	2726	none	51.60%	1	Compliant
018	498	1x 90 bp direct repeat	53.80%	0	MB_byprod
019	1348	none	47.20%	1	Compliant

020	884	none	52.00%	1	Compliant
021	1460	1x 20 bp direct repeat	56.60%	0.5	MB_main/byprod
022	1580	1x 20 bp direct repeat 1x 14 bp direct repeat	49.70%	0.5	MB_main/byprod
023	1180	none	48.40%	1	Compliant
024	1600	1x 20 bp direct repeat 1x 101 bp direct repeat	49.10%	0	MB_byprod
025	2261	none	54.10%	0	Empty
026	1220	none	53.90%	1	Compliant
027	760	none	52.10%	1	Compliant
028	1361	none	55.70%	0.75	Low_yield
029	635	none	55.40%	1	Compliant
030	1532	none	58.20%	0	Empty
031	940	none	59.60%	1	Compliant
032	1395	1x 19 bp direct repeat 1x 14 bp inverted repeats	40.90%	0.5	MB_main/byprod
033	874	none	46.60%	1	Compliant
034	1550	1x 23 bp inverted repeat	46.90%	0.5	MB_main/byprod
035	752	1x 14 bp inverted repeat 1x 23 bp inverted repeat	47.50%	1	Compliant
036	269	1x 23 bp inverted repeat	52.00%	0.75	Low_yield
037	250	1x 14 bp inverted repeat 1x 18 bp inverted repeat	45.60%	1	Compliant
038	1440	none	48.40%	0.75	MB_main
039	1600	1x 15 bp direct repeat	50.50%	0.75	MB_main
040	595	none	59.30%	1	Compliant
041	1741	none	51.80%	1	Compliant
042	1651	none	50.30%	1	Compliant
043	1031	none	50.50%	0.75	MB_main
044	1013	none	50.90%	0.75	MB_main
045	960	none	54.40%	1	Compliant

046	1161	none	50.00%	1	Compliant
047	801	none	55.10%	1	Compliant
048	1354	none	50.50%	1	Compliant
049	1354	none	50.40%	1	Compliant
050	1320	none	50.80%	1	Compliant
051	632	none	56.50%	0.75	MB_main
052	1300	1x 14 bp direct repeat	57.20%	0.5	MB_main/byprod
053	565	1x 14 bp direct repeat	64.80%	0.75	MB_main
054	1449	none	71.80%	0.75	Low_yield
055	2209	2x 14 bp direct repeat	69.10%	0.5	MB_main/byprod
056	1213	2x 14 bp inverted repeat	37.00%	1	Compliant
057	1074	2x 14 bp inverted repeat	51.90%	1	Compliant
058	1047	1x 15 bp inverted repeat 1x 20 bp direct repeat	44.30%	1	Compliant
059	1791	none	67.40%	0.5	MB_main/byprod
060	504	none	57.50%	0.75	MB_main
061	1424	none	57.00%	1	Compliant
062	1015	none	66.20%	1	Compliant
063	1340	none	38.00%	1	Compliant
064	1760	none	55.20%	0	Empty
065	411	none	47.00%	1	Compliant
066	1375	2x 15 bp direct repeat	48.40%	0.75	Low_yield
067	964	none	46.50%	1	Compliant
068	964	none	42.40%	1	Compliant
069	964	none	46.20%	1	Compliant
070	964	none	49.10%	1	Compliant
071	964	none	49.40%	1	Compliant
072	925	none	46.40%	1	Compliant
073	925	none	49.20%	0.75	MB_main
074	925	none	49.00%	1	Compliant

075	925	none	49.10%	0.75	MB_main
076	925	none	60.90%	0.75	MB_main
077	925	none	50.60%	1	Compliant
078	925	none	50.10%	1	Compliant
079	964	none	51.00%	1	Compliant
080	964	none	50.20%	1	Compliant
081	964	none	40.00%	1	Compliant
082	796	1x 14 bp inverted repeat	60.80%	0.75	Low_yield
083	398	1x 14 bp inverted repeat 1x 23 bp inverted repeat	49.70%	0.75	Low_yield
084	474	1x 14 bp inverted repeat 1x 23 bp inverted repeat	48.70%	0	Empty
085	399	1x 23 bp inverted repeat	47.90%	0	Non-compliant
086	809	none	51.20%	1	Compliant
087	955	none	51.60%	1	Compliant
088	1335	2x 14 bp inverted repeat	38.00%	0	MB_byprod
089	1278	2x 14 bp inverted repeat	52.60%	0.75	MB_main
090	1235	2x 14 bp inverted repeat	52.90%	0.5	MB_main/byprod
091	1196	2x 14 bp inverted repeat	51.40%	0.5	MB_main/byprod
092	796	1x 14 bp inverted repeat	49.50%	0.75	MB_main
093	1215	none	56.00%	1	Compliant
094	795	none	55.50%	1	Compliant
095	1427	1x 15 bp inverted repeat 1x 20 bp direct repeat	45.30%	0.75	MB_main
096	1472	1x 15 bp inverted repeat 1x 14 bp direct repeat	51.60%	1	Compliant
097	934	1x 14 bp direct repeat	50.60%	1	Compliant
098	1218	1x 41 bp direct repeat 1x 20 bp direct repeat	42.90%	0	MB_byprod
099	1035	1x 41 bp direct repeat	42.60%	0	MB_byprod
100	1209	1x 41 bp direct repeat	49.60%	0	MB_byprod

101	689	none	36.10%	1	Compliant
102	1115	1x 43 bp direct repeat	35.40%	0	Non-compliant
103	881	none	72.40%	0.75	Low_yield
104	696	none	66.20%	1	Compliant
105	966	none	71.80%	0	Non-compliant
106	1093	none	56.40%	1	Compliant
107	1111	none	47.90%	0.5	MB_main/byprod
108	1093	none	47.60%	0	Empty
109	1119	none	68.70%	0	MB_byprod
110	1549	none	47.80%	1	Compliant
111	1320	1x 14 bp direct repeat	53.20%	1	Compliant
112	1081	none	50.70%	1	Compliant
113	1081	none	50.50%	1	Compliant
114	1632	none	51.40%	1	Compliant
115	1280	1x 16 bp direct repeat	67.30%	0.5	MB_main/byprod
116	1331	1x 20 bp direct repeat 1x 80 bp direct repeat	44.40%	0	Non-compliant
117	535	none	50.50%	1	Compliant
118	685	none	54.00%	1	Compliant
119	1814	1x 18 bp direct repeat	49.30%	1	Compliant
120	1266	none	55.40%	0.75	Low_yield
121	1290	none	51.60%	1	Compliant
122	1260	none	52.20%	0.75	Low_yield
123	1260	none	51.90%	0	Non-compliant
124	597	none	43.20%	1	Compliant
125	1299	none	50.60%	1	Compliant
126	763	none	51.90%	1	Compliant
127	850	none	52.10%	1	Compliant
128	1287	none	49.70%	1	Compliant
129	1299	none	51.00%	1	Compliant

130	2836	1x 14 bp direct repeat 1x 18 bp inverted repeat	68.90%	0	Empty
131	813	none	65.90%	0.75	MB_main
132	1140	1x 15 bp inverted repeat	42.50%	0.75	MB_main
133	960	1x 15 bp inverted repeat	42.70%	1	Compliant
134	1266	1x 14 direct repeat	52.80%	1	Compliant
135	725	none	53.20%	1	Compliant
136	885	none	51.30%	1	Compliant
137	1292	none	51.70%	1	Compliant
138	885	none	51.40%	0.75	MB_main
139	1265	none	52.50%	1	Compliant
140	1292	none	51.50%	0.75	MB_main
141	969	1x 23 bp inverted repeat	48.90%	0.5	MB_main/byprod
142	1467	1x 14 bp direct repeat	54.10%	1	Compliant
143	1081	none	54.10%	1	Compliant
144	1207	1x 23 bp inverted repeat	47.90%	0	Non-compliant
145	1368	none	51.80%	0.75	MB_main
146	930	none	51.20%	1	Compliant
147	1460	none	51.20%	1	Compliant
148	540	1x 20 bp direct repeat 1x 14 bp inverted repeat	48.70%	0.75	MB_main
149	951	1x 14 bp inverted repeat	42.50%	1	Compliant
150	1330	1x 14 bp inverted repeat	51.10%	1	Compliant
151	870	1x 20 bp direct repeat 1x 14 bp inverted repeat	51.50%	0.75	MB_main
152	1790	1x 14 bp inverted repeat 1x 14 bp direct repeat	54.10%	1	Compliant
153	1570	1x 20 bp direct repeat 1x 14 bp inverted repeat	53.20%	1	Compliant
154	1590	1x 14 bp direct repeat	53.00%	1	Compliant
155	264	1x 14 bp inverted repeat	54.50%	0.5	MB_main/byprod

156	1760	1x 14 bp inverted repeat 1x 14 bp direct repeat	51.80%	1	Compliant
157	1000	1x 14 bp inverted repeat 1x 14 bp direct repeat	51.60%	1	Compliant
158	1220	1x 14 bp inverted repeat 1x 14 bp direct repeat	47.30%	1	Compliant
159	1460	1x 14 bp inverted repeat 1x 14 bp direct repeat	51.40%	0.75	MB_main
160	1680	1x 14 bp inverted repeat 1x 14 bp direct repeat	51.00%	1	Compliant
161	1630	1x 14 bp inverted repeat 1x 14 bp direct repeat	51.30%	1	Compliant
162	1410	1x 14 bp inverted repeat 1x 14 bp direct repeat	51.00%	1	Compliant
163	1760	1x 14 bp inverted repeat 1x 14 bp direct repeat	37.10%	1	Compliant
164	1770	1x 14 bp direct repeat	48.80%	1	Compliant
165	1266	1x 14 bp direct repeat	62.80%	1	Compliant
166	1030	1x 14 bp direct repeat	41.50%	1	Compliant
167	1860	1x 78 bp inverted repeat	42.70%	0.5	MB_main/byprod
168	680	1x 14 bp inverted repeat	49.60%	1	Compliant
169	1180	1x 14 bp inverted repeat	48.80%	1	Compliant
170	2650	1x 14 bp inverted repeat 1x 15 bp direct repeat	67.40%	0.75	MB_main
171	1130	1x 14 bp inverted repeat 1x 14 bp direct repeat	64.40%	1	Compliant
172	430	1x 14 bp inverted repeat	51.20%	0.5	MB_main/byprod
173	380	1x 14 bp inverted repeat	41.30%	1	Compliant
174	508	1x 14 bp inverted repeat	47.40%	1	Compliant
175	630	1x 14 bp inverted repeat	54.30%	1	Compliant
176	1834	1x 14 bp inverted repeat	52.20%	0.75	MB_main
177	1549	1x 14 bp inverted repeat	53.90%	0	Empty

178	1495	1x 14 bp inverted repeat	55.90%	1	Compliant
179	1558	1x 14 bp inverted repeat	50.20%	1	Compliant
180	1567	1x 14 bp inverted repeat	49.00%	1	Compliant
181	664	1x 14 bp inverted repeat	62.80%	1	Compliant
182	946	1x 14 bp inverted repeat 1x 14 bp direct repeat	71.20%	0	Non-compliant
183	1030	1x 14 bp inverted repeat	68.20%	1	Compliant
184	985	1x 14 bp inverted repeat	62.00%	1	Compliant
185	1102	1x 14 bp inverted repeat 1x 14 bp direct repeat	67.10%	0.75	Low_yield
186	590	1x 14 bp inverted repeat	62.20%	1	Compliant
187	889	1x 14 bp inverted repeat	60.90%	1	Compliant
188	1390	1x 14 bp inverted repeat	65.80%	1	Compliant
189	1240	1x 14 bp inverted repeat	61.50%	1	Compliant
190	470	1x 14 bp inverted repeat	56.60%	1	Compliant
191	1276	1x 14 bp inverted repeat	63.10%	1	Compliant
192	1001	1x 14 bp inverted repeat 1x 18 bp direct repeat	72.40%	0	Non-compliant
193	1450	1x 14 bp inverted repeat	68.70%	1	Compliant
194	350	1x 14 bp inverted repeat	64.30%	1	Compliant
195	1741	1x 14 bp inverted repeat	54.70%	1	Compliant
196	1678	1x 14 bp inverted repeat	51.30%	1	Compliant
197	810	1x 14 bp inverted repeat	50.90%	1	Compliant
198	2104	1x 14 bp inverted repeat	46.30%	1	Compliant
199	1510	1x 14 bp inverted repeat	67.50%	0.75	MB_main
200	1252	1x 14 bp inverted repeat	64.60%	1	Compliant
201	1402	1x 14 bp inverted repeat	64.00%	1	Compliant
202	826	1x 14 bp inverted repeat	67.80%	0	Non-compliant
203	285	1x 14 bp inverted repeat	42.80%	1	Compliant
204	985	1x 14 bp inverted repeat	61.50%	0	MB_byprod

		1x 20 bp inverted repeat 1x 41 bp inverted repeat			
205	268	1x 14 bp inverted repeat	42.90%	1	Compliant
206	1044	1x 14 bp inverted repeat	61.60%	1	Compliant
207	743	1x 14 bp inverted repeat	55.30%	1	Compliant
208	832	1x 14 bp inverted repeat	55.00%	1	Compliant
209	872	1x 14 bp inverted repeat 1x 15 bp inverted repeat	54.10%	1	Compliant
210	1010	1x 78 bp inverted repeat 1x 21 bp direct repeat 2x 15 bp direct repeat 1x 20 bp direct repeat	47.80%	0	MB_byprod
211	430	1x 14 bp inverted repeat 1x 21 bp direct repeat 3x 15 bp direct repeat	47.00%	1	Compliant
212	430	1x 14 bp inverted repeat	50.70%	1	Compliant
213	1095	1x 14 bp inverted repeat 1x 18 bp inverted repeat	52.60%	1	Compliant
214	1740	1x 14 bp direct repeat	57.90%	1	Compliant
215	866	none	53.20%	1	Compliant
216	417	none	52.00%	1	Compliant
217	460	none	65.40%	1	Compliant
218	1262	none	51.60%	1	Compliant
219	1274	none	52.40%	0	Non-compliant
220	1271	none	53.00%	0.75	MB_main
221	885	none	52.40%	0.75	MB_main
222	903	none	49.60%	0.75	MB_main
223	1021	none	50.40%	1	Compliant
224	673	none	53.20%	1	Compliant
225	673	none	52.90%	1	Compliant
226	673	none	52.70%	1	Compliant

227	673	none	53.00%	1	Compliant
228	1388	none	50.20%	1	Compliant
229	1370	none	50.50%	1	Compliant
230	1390	none	35.70%	0.5	MB_main/byprod
231	1146	1x 14 bp direct repeat	54.20%	1	Compliant
232	799	none	49.40%	1	Compliant
233	805	none	49.60%	1	Compliant
234	1891	none	39.00%	1	Compliant
235	1287	1x 14 bp direct repeat	69.50%	0	MB_byprod
236	621	none	55.60%	1	Compliant
237	370	1x 18 bp direct repeat	47.80%	1	Compliant
238	742	none	38.00%	0.75	MB_main
239	1792	1x 144 bp direct repeat 1x 14 bp inverted repeat	54.50%	0	Non-compliant
240	500	1x 20 bp direct repeat 1x 14 bp inverted repeat	48.80%	1	Compliant
241	870	1x 14 bp inverted repeat	43.90%	1	Compliant
242	1113	1x 20 bp direct repeat 1x 14 bp inverted repeat 1x 23 bp inverted repeat	47.70%	0	Non-compliant

Table 3.2.1: Analysis overview of synthesized fragments in 2020.

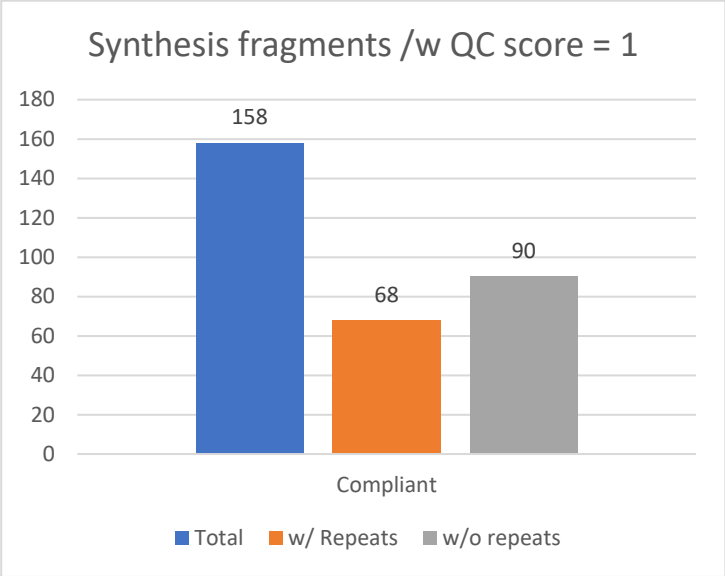

Figure 3.2.1: Compliant synthesis fragments overview of 2020 depending on the presence of repeats

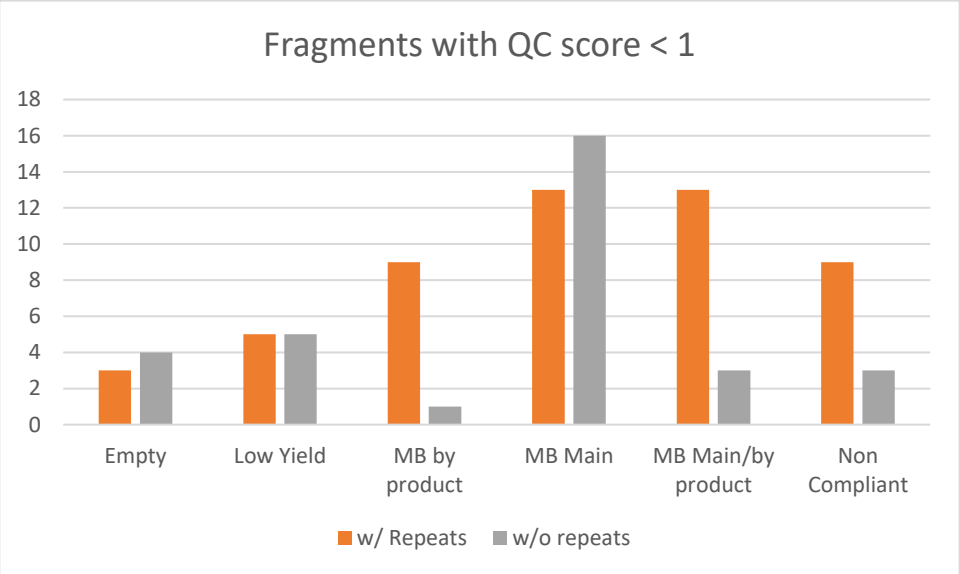

Figure 3.2.2: Synthesis fragments overview of 2020 w/ QC score < 1

According to figure 3.2.2, the presence of repeats within the fragment may increase the failure rate of the synthesis phase. All the synthesized

fragments were used for the downstream assembly applications recording the following effects:

Synthesis outcome	Effect on assembly cPCR QC	Effect on sequencing QC
Compliant	None	None
Low yield	None	None
MB main	Few assembly rearrangements	None
MB main/by product	Some false-positives issues and assembly rearrangements	Some unexpected assembly rearrangements
MB by product	Only non-compliances	Only non-compliances
Non-compliant	Only non-compliances	Only non-compliances

Table 3.2.2: Starting fragments QC effect on downstream assembly phases.

The synthesis score system was validated and used as the value to estimate the product's quality to be used for downstream processes. Even with the presence of by-products, projects involving fragments with a higher score than 0.5 did not cause relevant trouble during the assembly workflow execution. All the projects which involved the use of fragments with 0.5 as the score required a larger scale of colonies for cPCR QC since by-products interfered with the correct assembly of the constructs. All the failed synthesis fragments could not be used to assemble the final product. The synthesis fragments limitations evaluation set the following criteria: avoid repeats longer than 14 bp; GC% content within 20% and 80 %; fragment length range between 250 bp to 1800 bp.

3.2.2 Recyclability assessment outcomes

Before implementing the Geneious-based system at the end of 2020, the manual recyclability assessment took from 30 minutes to 2 hours for each synthesis project design. After integrating this system, each operator takes from 10 min to 20 min to execute this phase.

3.2.3 Troubleshoot tiers and mitigation plan integration on manufacturing process

The definition and integration of troubleshooting tiers made it possible to have a standardized procedure to capture and classify issues in any points of the workflow executed by any operator. In addition, releasing a guideline to solve several troubleshoot scenarios, decreased the delays related to project re-design from 3 extra working days to 1 working day.

3.2.4 Mitigation plan design

By the data achieved in 2020, a relationship was seen between the synthesis fragment quality and the sequences repeats. The mitigation plan to minimize synthesis failures consisted of introducing overlapping zones between the repeats. This step made the assembly execution possible to keep the homologous zones unique and dilute the sequence repetitions splitting the fragment into multiple ones. This key feature dramatically reduced the whole synthesis success rate in 2021 projects.

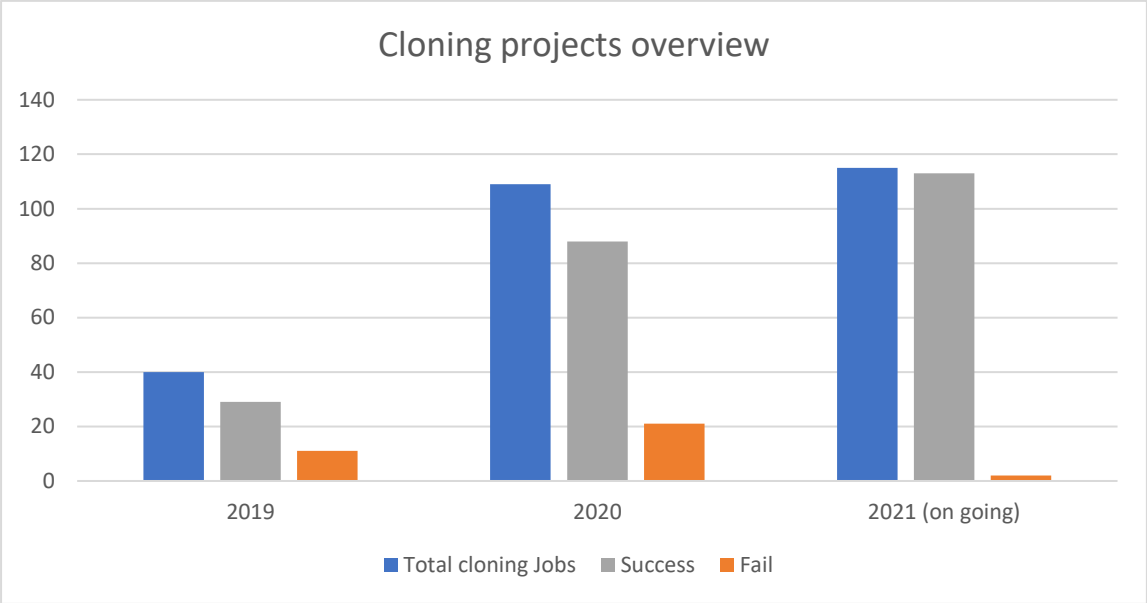

Figure 3.2.3: Cloning projects completion overview up to Aug 2021.

3.2.5 Infrastructure and framework

Since the synthesis and assembly requests escalated year by year, creating a framework to oversee the whole manufacturing facility was necessary. The following changes were introduced for 2021 projects. Browser-based lab-books: the project designer could generate the lab-book before the execution by an operator starting from a step-based template. Everything was link-based and made the task assignment more flexible. Experiment tracking software: any responsible could coordinate and review each experiment in the manufacturing facility. This change minimized issues related to overcrowded tasks scenarios. Internal knowledge base: any manufacturing and biological information was recorded in an internal article base to be used for more complex upcoming projects.

3.2.6 Cost evaluation

In the three years of workflow implementation, the bill of materials (BoM; the cost of reagents and materials required to complete a project without counting the personnel wages and salaries) to execute a cloning project changed. This parameter is generated taking in account the probability to achieve the correct plasmid according to the experimental data which defines the average number of samples to be analyzed and processed during the whole cloning pipeline. The following data represents the BoM estimation of a trimolecular subcloning project with a synthesis region length of 3000 bp without counting troubleshoots and mitigation plans (these factors are case dependent).

Year	Synthesis cost	Assembly QC cost	Sequencing QC cost
2019	1320.00 €	21.92 €	144.00 €
2020	1050.00 €	18.04 €	108.00 €
2021	840.00 €	7.87 €	54.00 €

Table 3.2.3: BoM estimation for a trimolecular assembly with a synthesis portion of 3000 bp using the techniques and conditions active per each year.

The cost reduction of the DNA synthesis between 2019 and 2020 was caused by the change from outsourced to in-house synthetic DNA production. The synthesis cost reduction from 2020 to 2021 mainly stemmed from the optimization of both DNA fragment design and experimental processes. Since the assembly pass rate increased and the mutation rate decreased with each protocol evolution, both respective quality check BoM estimations to achieve a compliant sample were smaller.

3.2.7 Manufacturing decoupling analysis

The manufacturing workflow evolved to achieve the full-decoupling from any starting plasmid design year by year. Considering as “standard assembly” each cloning job involved a standardized destination vector (pDLX, pSEVA, pUC, pET), according to figure 3.2.4, the total number of project-tailored assemblies covered the majority of projects compared to standard assemblies in 2021.

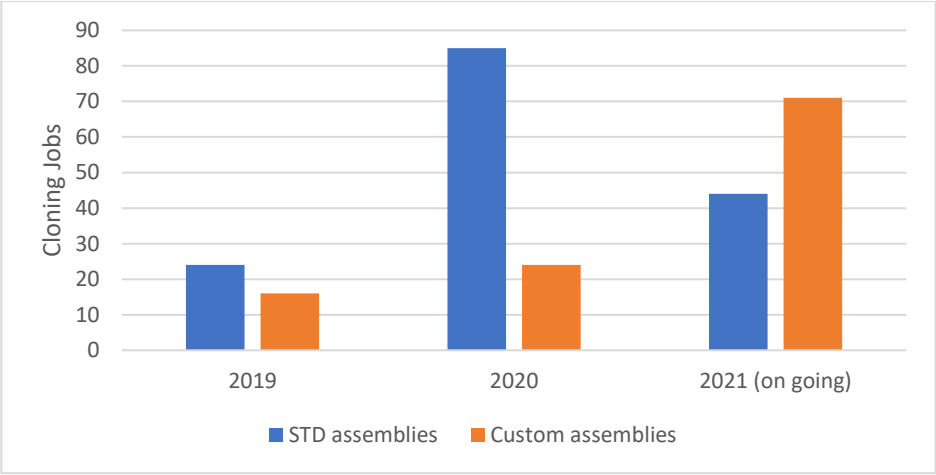

Figure 3.2.4: standard assemblies and custom assemblies project amount year-by-year.

Year	Total cloning Jobs	Success	Fail	Success rate (%)
2019	40	29	11	72.50
2020	109	88	21	80.73
2021 (till Aug)	115	113	2	98.26

Table 3.2.4: success rate of cloning projects sort by year

Even if the customized project covered the majority of cloning projects, the project success rate till the half of 2021 reached 98 %.

Conclusions

From the first DNAmate-based cloning project at Explora Biotech srl, it was possible to formalize a set of defined and approved SOPs to execute the whole experimental workflow to achieve the final product. However, this workflow was still DNA architecture-dependent, making it impossible to apply in every manufacturing scenario.

At the second part of the Industrial Ph.D., some cloning projects were executed using a scarless overlap-based system: the Gibson assembly. In addition, the starting fragment synthesis method was formalized to produce in-house manufactured linear dsDNA sequences. Integrating both changes into the manufacturing workflow made it possible to completely decouple the DNA architecture's starting design from the whole cloning workflow. Finally, this cloning workflow was tested and validated as a potentially compatible cloning approach to assemble degenerated mutagenic libraries.

The last step consisted of identifying all the limits of the linear dsDNA synthesis and cloning phase since a large amount of quality check data after the scale-up of cloning requests was collected. The evaluation and analysis of achieved data permitted optimizing the workflow at the experiments' design stage, reducing production days and total error rate. The following additional points were achieved for the DNA foundry: the definition and formalization of workflows to solve troubleshooting scenarios dependent on the issue severity; integration of several cloud-based platforms to track and assign any task; a set of standardized common vocabulary and syntaxes to identify any produced item in the lab.

The following objectives were achieved at the end of this industrial Ph.D. program:

- Manufacturing success rate of 98.26 %
- Competitive turnaround time of 16 working days
- 42 kbp as the largest produced construct size
- Customer retention rate almost 100 %

Future perspectives

The next step for this commercial full-working DNA foundry will be the conversion into a complete biofoundry. The following conditions need to be satisfied in order to achieve this ambitious goal:

- Formalize a dedicated PhD level team which aids the synthetic biology researchers in the design of their projects taking in account the knowledge from previous *in vivo* experiments and the manufacturability.
- Integrate scalable, fully automated robotic platforms to streamline the whole synthesis process in the build phase.
- Implement an infrastructure to cover the learn phase which includes fermenters for metabolic analysis and a dedicated unit for protein engineering.
- Build up a dedicated system to acquire, store and analyze any collected experimental data along the DBTL cycle.

Appendix

Polymerase chain reaction (PCR)

The PCR is a revolutionary, widely used technique in molecular biology based on the capability of DNA polymerase enzyme to make millions to billions of copies of a defined DNA region (amplicons). The amplified sequence is defined by a couple of short DNA molecules called "primer" where sequences are complementary to the amplicon ends. These primers can anneal on the template that carries the target sequence letting the polymerase recognise them and generate copies of the sequence of interests. [43][44].

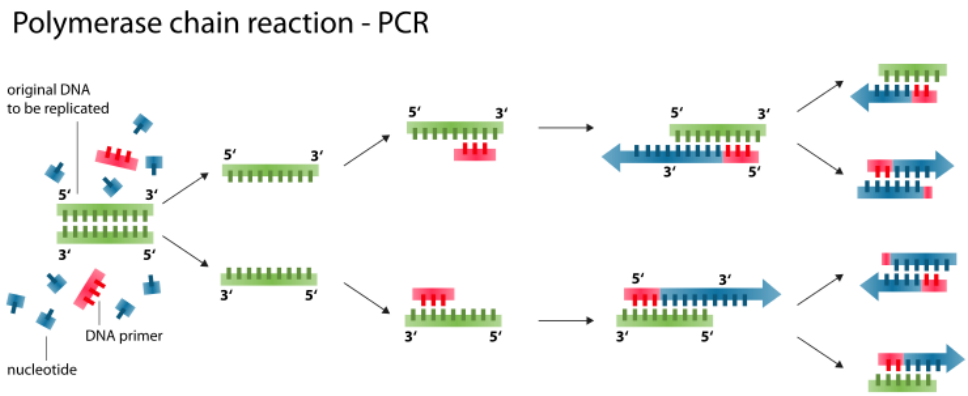

Figure A: graphical illustration of the PCR mechanism (image from https://www.genomeup.com/Glossary/pcr/)

PCR reaction relies on thermal cycling, which exposes reactants to repeated heating and cooling cycles composed of 3 main steps. The denaturation step is characterized by a high temperature that permits separating each DNA strand and avoiding any secondary structures. The annealing step corresponds to the optimal temperature at which each primer anneals to the complementary sequence. Finally, the extension phase is the optimal temperature for the

polymerase activity, with the time depending on the length of the amplicon of interest.

Mutagenic polymerase chain reaction (mPCR)

The mPCR is a variant of the classic polymerase chain reaction where mutations can be added to the staring template sequence exploiting the primer design. For this work, the mutagenic PCR was largely used through the bacterial DNA repair approach.

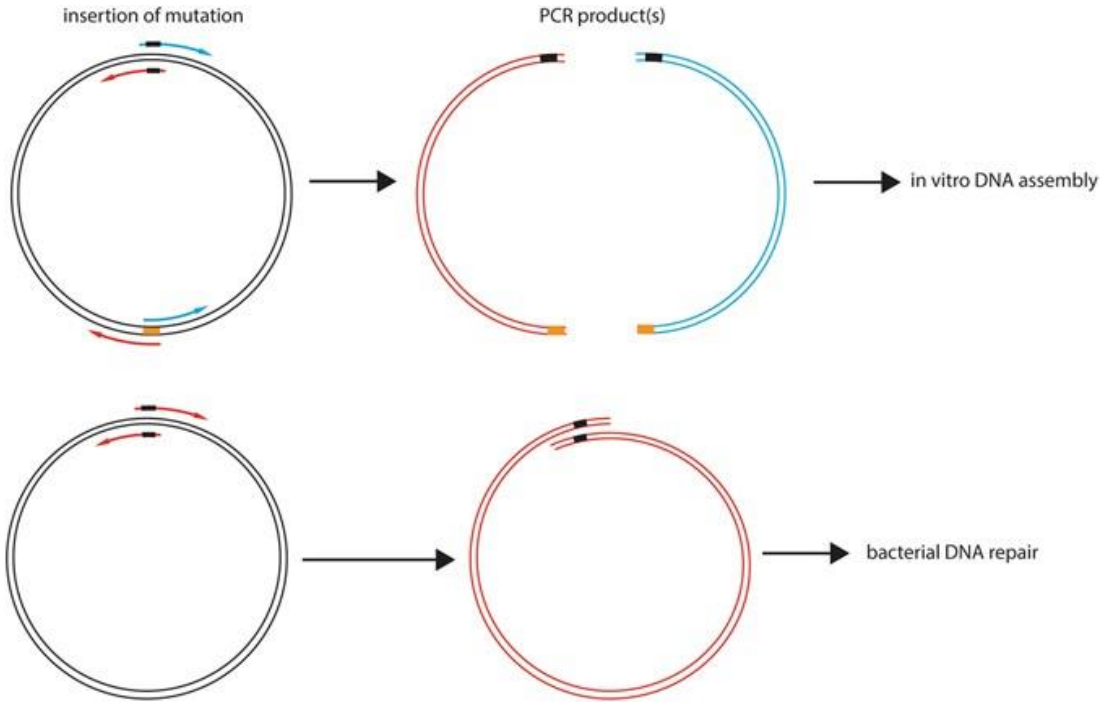

Figure B: Graphical illustration of the mutagenic PCR technique [45]

Colony PCR (cPCR)

This method is a variant of the classic PCR, which uses bacterial colonies as a DNA template, where the DNA is released from the cells after an initial boiling step [46]. This technique is primarily used as a screening technique since it permits the identification of the presence and absence of insert DNA in the destination vector.

Agarose gel electrophoresis (AGE)

This technique is a gel electrophoresis method that permits the separation of a mixed population of macromolecules such as DNA [47]. For this work, this method was used to identify DNA fragments lengths and the presence of any by-products through the comparison of markers containing known sizes of DNA fragments.

Transformation through heat shock

Heat shock is an essential technique in molecular biology in which foreign plasmids or assembly products are inserted into bacteria [48]. This technique is based on the application of a sudden increase in temperature, which creates pores in the plasma membrane of the bacteria. Afterwards, these pores allow for plasmid DNA to enter the bacterial cell.

Transformation through electroporation

Electroporation (or electropermeabilization) is a technique that applies an electrical field to cells to increase the cell membrane permeability. This process permits DNA to be introduced into the cell [49].

The plasmid preparation (or plasmid prep) is used to extract and purify plasmid DNA from transformed bacterial cells. The first phase of this process consists of the cells growth in liquid medium. After the centrifugation and the resuspension of the cell pellet with an isotonic solution, the release and the denaturation of DNA are allowed through the alkaline lysis of DNA. The lysed cells and released DNA mixture is neutralized, adding a buffer solution taking the pH back to the physiological condition. Several kits are commercially available to purify plasmid DNA. However, all of them aim to create conditions to make either DNA or other biomolecules precipitate, allowing the separation of plasmid DNA [50][51].

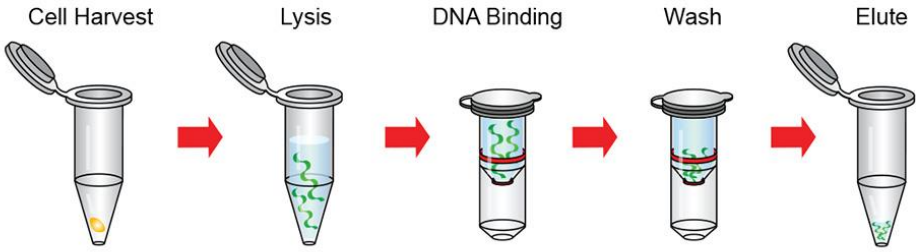

Figure C: graphical illustration of a column-based plasmid prep technique (image from https://www.genedirex.com/product/plasmid-miniprep-kit/)

Sanger sequencing

Sanger sequencing is a technique based on an in Vitro DNA replication mediated by DNA polymerase and the selective incorporation of chain-terminating dideoxynucleotide triphosphates (ddNTPs) [52][53]. The process consists of the execution of four reactions in parallel. All reactions share the presence of DNA polymerase, a primer, the template, and all four dNTPs for each nucleotide. Each reaction contains only one type of ddNTPs in lower concentration than dNTPs mix, causing the termination of replication extension of ssDNA when the ddNTP is included. This extension-termination approach generates populations of ssDNA with different lengths which the 3' ends terminate with the same type of ddNTP. The nucleotide type is checked depending on the ddNTP included at the end of the strand. The nucleotide position is identified through the strand length. The detection method of these two parameters changes by the Sanger sequencing variant.

DpnI digestion

Since plasmid DNA derived from plasmid preparation are commonly used in PCR reactions as templates and may interfere with the downstream transformation phase, the removal of these DNA molecules is crucial. DpnI is a restriction enzyme that recognizes and cuts the Gm6A^TC site, where N6-methyladenosine (m6A) is the most abundant transcriptional modification [54]. This short and common recognition site makes DpnI enzyme universal to cut specifically plasmids used as templates in PCR reaction keeping untouched DNA molecules generated during the PCR reaction for the lack of transcriptional modifications.

References

1. Carnes, A., & Williams, J. (2007). Plasmid DNA Manufacturing Technology. B Recent Patents on Biotechnology (Vol. 1, Issue 2, c. 151–166). Bentham Science Publishers Ltd. https://doi.org/10.2174/187220807780809436
2. Johnson, I. S. (1983). Human Insulin from Recombinant DNA Technology. B Science (Vol. 219, Issue 4585, c. 632–637). American Association for the Advancement of Science (AAAS). https://doi.org/10.1126/science.6337396.
3. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. (1986). Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. B Cold Spring Harbor Symposia on Quantitative Biology (Vol. 51, Issue 0, c. 263–273). Cold Spring Harbor Laboratory. https://doi.org/10.1101/sqb.1986.051.01.032
4. Naso, M. F., Tomkowicz, B., Perry, W. L., III, & Strohl, W. R. (2017). Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. B BioDrugs (Vol. 31, Issue 4, c. 317–334). Springer Science and Business Media LLC. https://doi.org/10.1007/s40259-017-0234-5
5. Liu, S., Wang, S., & Lu, S. (2016). DNA immunization as a technology platform for monoclonal antibody induction. B Emerging Microbes & Infections (Vol. 5, Issue 1, c. 1–12). Informa UK Limited. https://doi.org/10.1038/emi.2016.27
6. Ramos, H. R., & Ho, P. L. (2014). Developing Snake Antivenom Sera by Genetic Immunization: A Review. B Clinical Toxinology (c. 1–12). Springer Netherlands. https://doi.org/10.1007/978-94-007-6288-6_36-1

7. Lartigue, C., Duret, S., Garnier, M., & Renaudin, J. (2002). New plasmid vectors for specific gene targeting in Spiroplasma citri. B Plasmid (Vol. 48, Issue 2, c. 149–159). Elsevier BV. https://doi.org/10.1016/s0147-619x(02)00121-x
8. Goeddel, D. V., Kleid, D. G., Bolivar, F., Heyneker, H. L., Yansura, D. G., Crea, R., Hirose, T., Kraszewski, A., Itakura, K., & Riggs, A. D. (1979). Expression in Escherichia coli of chemically synthesized genes for human insulin. B Proceedings of the National Academy of Sciences (Vol. 76, Issue 1, c. 106–110). Proceedings of the National Academy of Sciences. https://doi.org/10.1073/pnas.76.1.106
9. Murray, A. W., & Szostak, J. W. (1983). Construction of artificial chromosomes in yeast. B Nature (Vol. 305, Issue 5931, c. 189–193). Springer Science and Business Media LLC. https://doi.org/10.1038/305189a0
10. Perlak, F. J., Deaton, R. W., Armstrong, T. A., Fuchs, R. L., Sims, S. R., Greenplate, J. T., & Fischhoff, D. A. (1990). Insect Resistant Cotton Plants. B Nature Biotechnology (Vol. 8, Issue 10, c. 939–943). Springer Science and Business Media LLC. https://doi.org/10.1038/nbt1090-939
11. Weiss, R., Homsy, G. E., & Knight, T. F., Jr. (2002). Toward in vivo Digital Circuits. B Natural Computing Series (c. 275–295). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-55606-7_14
12. Weiss, R., & Knight, T. F., Jr. ,. (2004). Cellular Computation and Communication Using Engineered Genetic Regulatory Networks. B Cellular Computing. Oxford University Press. https://doi.org/10.1093/oso/9780195155396.003.0012
13. Endy, D. (2005). Foundations for engineering biology. B Nature (Vol. 438, Issue 7067, c. 449–453). Springer Science and Business Media LLC. https://doi.org/10.1038/nature04342

14. Pleiss, J. (2006). The promise of synthetic biology. In Applied Microbiology and Biotechnology (Vol. 73, Issue 4, pp. 735–739). Springer Science and Business Media LLC. https://doi.org/10.1007/s00253-006-0664-3
15. Knight, T. (2002). DARPA BioComp Plasmid Distribution 1.00 of Standard Biobrick Components. http://hdl.handle.net/1721.1/21167
16. Rodriguez, C., Bartram, S., Ramasubramanian, A., Endy, D. (2009). Bbf rfc 16: Synthetic biology open language visual (sbolv) specification. Available online at: http://dspace.mit.edu/handle/1721.1/49523
17. Weber, E., Engler, C., Gruetzner, R., Werner, S., & Marillonnet, S. (2011). A Modular Cloning System for Standardized Assembly of Multigene Constructs. In J. Peccoud (Ed.), PLoS ONE (Vol. 6, Issue 2, p. e16765). Public Library of Science (PLoS). https://doi.org/10.1371/journal.pone.0016765
18. Müller, K. M., & Arndt, K. M. (2011). Standardization in Synthetic Biology. In Methods in Molecular Biology (pp. 23–43). Humana Press. https://doi.org/10.1007/978-1-61779-412-4_2
19. Andrianantoandro, E., Basu, S., Karig, D. K., & Weiss, R. (2006). Synthetic biology: new engineering rules for an emerging discipline. In Molecular Systems Biology (Vol. 2, Issue 1). EMBO. https://doi.org/10.1038/msb4100073
20. Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z.-Q., Segall-Shapiro, T. H., … Venter, J. C. (2010). Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. In Science (Vol. 329, Issue 5987, pp. 52–56). American Association for the Advancement of Science (AAAS). https://doi.org/10.1126/science.1190719

21. Kwok, R. (2010). Five hard truths for synthetic biology. In Nature (Vol. 463, Issue 7279, pp. 288–290). Springer Science and Business Media LLC. https://doi.org/10.1038/463288a
22. Nielsen, A. A. K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., Ross, D., Densmore, D., & Voigt, C. A. (2016). Genetic circuit design automation. In Science (Vol. 352, Issue 6281, pp. aac7341–aac7341). American Association for the Advancement of Science (AAAS). https://doi.org/10.1126/science.aac7341
23. Meng, F., & Ellis, T. (2020). The second decade of synthetic biology: 2010–2020. In Nature Communications (Vol. 11, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41467-020-19092-2
24. (2021). Bioeconomy strategy: 2018 to 2030 - GOV.UK. (n.d.). Available online at: https://www.gov.uk/government/publications/bioeconomy-strategy-2018-to-2030
25. Gray, P., Meek, S., Griffiths, P., Trapani, J., Small, I., Vickers, C., et al. (2018). Synthetic Biology in Australia: An Outlook to 2030, Australian Council of Learned Academies. Available online at: https://acola.org/hs3-synthetic-biology-australia/
26. (2020). Safeguarding the Bioeconomy. National Academies Press. https://doi.org/10.17226/25525
27. Holowko, M. B., Frow, E. K., Reid, J. C., Rourke, M., & Vickers, C. E. (2020). Building a biofoundry. In Synthetic Biology (Vol. 6, Issue 1). Oxford University Press (OUP). https://doi.org/10.1093/synbio/ysaa026
28. Hillson, N., Caddick, M., Cai, Y., Carrasco, J. A., Chang, M. W., Curach, N. C., Bell, D. J., Le Feuvre, R., Friedman, D. C., Fu, X., Gold, N. D., Herrgård, M. J., Holowko, M. B., Johnson, J. R., Johnson, R. A., Keasling,

J. D., Kitney, R. I., Kondo, A., Liu, C., … Freemont, P. S. (2019). Building a global alliance of biofoundries. In Nature Communications (Vol. 10, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41467-019-10079-2

29. (2019). Biofoundries: the new factories for genetic products. World Economic Forum. (n.d.). Available online at: https://www.weforum.org/agenda/2019/10/biofoundries-the-new-factories-for-genetic-products/
30. Hughes, R. A., & Ellington, A. D. (2017). Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. In Cold Spring Harbor Perspectives in Biology (Vol. 9, Issue 1, p. a023812). Cold Spring Harbor Laboratory. https://doi.org/10.1101/cshperspect.a023812
31. Balbás, P., Soberón, X., Merino, E., Zurita, M., Lomeli, H., Valle, F., Flores, N., & Bolivar, F. (1986). Plasmid vector pBR322 and its special-purpose derivatives — a review. In Gene (Vol. 50, Issues 1–3, pp. 3–40). Elsevier BV. https://doi.org/10.1016/0378-1119(86)90307-0
32. de Boer, H. A., Comstock, L. J., & Vasser, M. (1983). The tac promoter: a functional hybrid derived from the trp and lac promoters. In Proceedings of the National Academy of Sciences (Vol. 80, Issue 1, pp. 21–25). Proceedings of the National Academy of Sciences. https://doi.org/10.1073/pnas.80.1.21
33. Brosius, J., Erfle, M., & Storella, J. (1985). Spacing of the -10 and -35 regions in the tac promoter. Effect on its in vivo activity. In Journal of Biological Chemistry (Vol. 260, Issue 6, pp. 3539–3541). Elsevier BV. https://doi.org/10.1016/s0021-9258(19)83655-4
34. Amouyal, M. (2014). From adjacent activation in Escherichia coli and DNA cyclization to eukaryotic enhancers: the elements of a puzzle. In Frontiers in Genetics (Vol. 5). Frontiers Media SA. https://doi.org/10.3389/fgene.2014.00371

35. Calos, M. P. (1978). DNA sequence for a low-level promoter of the lac repressor gene and an ‘up’ promoter mutation. In Nature (Vol. 274, Issue 5673, pp. 762–765). Springer Science and Business Media LLC. https://doi.org/10.1038/274762a0
36. Chizzolini, F., Forlin, M., Cecchi, D., & Mansy, S. S. (2013). Gene Position More Strongly Influences Cell-Free Protein Expression from Operons than T7 Transcriptional Promoter Strength. In ACS Synthetic Biology (Vol. 3, Issue 6, pp. 363–371). American Chemical Society (ACS). https://doi.org/10.1021/sb4000977
37. Potapov, V., & Ong, J. L. (2017). Examining Sources of Error in PCR by Single-Molecule Sequencing. In R. Kalendar (Ed.), PLOS ONE (Vol. 12, Issue 1, p. e0169774). Public Library of Science (PLoS). https://doi.org/10.1371/journal.pone.0169774
38. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., III, & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. In Nature Methods (Vol. 6, Issue 5, pp. 343–345). Springer Science and Business Media LLC. https://doi.org/10.1038/nmeth.1318
39. Jacobsen, H., Klenow, H., & Overoaard-Hansen, K. (1974). The N-Terminal Amino-Acid Sequences of DNA Polymerase I from Escherichia coli and of the Large and the Small Fragments Obtained by a Limited Proteolysis. In European Journal of Biochemistry (Vol. 45, Issue 2, pp. 623–627). Wiley. https://doi.org/10.1111/j.1432-1033.1974.tb03588.x
40. Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., & Polz, M. F. (2005). PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample. In Applied and Environmental Microbiology (Vol. 71,

Issue 12, pp. 8966–8969). American Society for Microbiology. https://doi.org/10.1128/aem.71.12.8966-8969.2005

41. Froger, A., & Hall, J. E. (2007). Transformation of Plasmid DNA into *E. coli* Using the Heat Shock Method. In Journal of Visualized Experiments (Issue 6). MyJove Corporation. https://doi.org/10.3791/253
42. Potter, H. (2003). Transfection by Electroporation. In Current Protocols in Molecular Biology (Vol. 62, Issue 1). Wiley. https://doi.org/10.1002/0471142727.mb0903s62
43. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, N. (1985). Enzymatic Amplification of β-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. In Science (Vol. 230, Issue 4732, pp. 1350–1354). American Association for the Advancement of Science (AAAS). https://doi.org/10.1126/science.2999980
44. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., & Erlich, H. A. (1988). Primer-Directed Enzymatic Amplification of DNA with a Thermostable DNA Polymerase. In Science (Vol. 239, Issue 4839, pp. 487–491). American Association for the Advancement of Science (AAAS). https://doi.org/10.1126/science.2448875
45. Heydenreich, F. M., Miljuš, T., Jaussi, R., Benoit, R., Milić, D., & Veprintsev, D. B. (2017). High-throughput mutagenesis using a two-fragment PCR approach. In Scientific Reports (Vol. 7, Issue 1). Springer Science and Business Media LLC. https://doi.org/10.1038/s41598-017-07010-4
46. Bergkessel, M., & Guthrie, C. (2013). Colony PCR. In Methods in Enzymology (pp. 299–309). Elsevier. https://doi.org/10.1016/b978-0-12-418687-3.00025-2

47. Lee, P. Y., Costumbrado, J., Hsu, C.-Y., & Kim, Y. H. (2012). Agarose Gel Electrophoresis for the Separation of DNA Fragments. In Journal of Visualized Experiments (Issue 62). MyJove Corporation. https://doi.org/10.3791/3923
48. Froger, A., & Hall, J. E. (2007). Transformation of Plasmid DNA into *E. coli* Using the Heat Shock Method. In Journal of Visualized Experiments (Issue 6). MyJove Corporation. https://doi.org/10.3791/253
49. Gehl, J. (2003). Electroporation: theory and methods, perspectives for drug delivery, gene therapy and research. In Acta Physiologica Scandinavica (Vol. 177, Issue 4, pp. 437–447). Wiley. https://doi.org/10.1046/j.1365-201x.2003.01093.x
50. Zhang, S., & Cahalan, M. D. (2007). Purifying Plasmid DNA from Bacterial Colonies Using the Qiagen Miniprep Kit. In Journal of Visualized Experiments (Issue 6). MyJove Corporation. https://doi.org/10.3791/247
51. Koontz, L. (2013). Explanatory Chapter. In Methods in Enzymology (pp. 23–28). Elsevier. https://doi.org/10.1016/b978-0-12-418687-3.00002-1
52. Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. In Journal of Molecular Biology (Vol. 94, Issue 3, pp. 441–448). Elsevier BV. https://doi.org/10.1016/0022-2836(75)90213-2
53. Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. In Proceedings of the National Academy of Sciences (Vol. 74, Issue 12, pp. 5463–5467). Proceedings of the National Academy of Sciences. https://doi.org/10.1073/pnas.74.12.5463
54. Chen, X.-Y., Zhang, J., & Zhu, J.-S. (2019). The role of m6A RNA methylation in human cancer. In Molecular Cancer (Vol. 18, Issue 1).

Springer Science and Business Media LLC. https://doi.org/10.1186/s12943-019-1033-z

