Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

Maria Wilhelmsen Hoff

Establishment and characterization of artificial promoter and 5´UTR sequences in *Escherichia coli* and *Vibrio natriegens*

Master's thesis in Biotechnology

Supervisor: Rahmi Lale Co-supervisor: Lisa Tietze

July 2020

NTNU
Norwegian University of
Science and Technology

Maria Wilhelmsen Hoff

Establishment and characterization of artificial promoter and 5'UTR sequences in *Escherichia coli* and *Vibrio natriegens*

Master's thesis in Biotechnology
Supervisor: Rahmi Lale Co-supervisor: Lisa Tietze
July 2020

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

NTNU
Norwegian University of
Science and Technology

Acknowledgements

This project was performed at the Department of Biotechnology and Food Science (IBT) at the Norwegian University of Science and Technology (NTNU) in Trondheim.

This thesis concludes 5 years of studying at NTNU to achieve a bachelor's in cellular and molecular biology and a Master of Science in Biotechnology. These have been the most challenging, memorable and educational years so far, which I am very grateful to have been able to experience.

During the work for this thesis I was a member of the PhotoSynLab research group, which greatly increased the learning outcome and enjoyability of the process of becoming an independent student working in the lab. First of all, I would like to thank my supervisor Associate Professor Rahmi Lale for including me into the research group and providing valuable advices and knowledge throughout the year. I would also like to express my gratitude to PhD candidate and co-supervisor Liza Tietze who contributed significantly to my academic growth and development and for always being available with advice and demonstration both in the lab and during the writing process which greatly increased my understanding of synthetic biology.

Throughout this last year, I am grateful for the companionship and advice the members of the group provided, in particular Maxime Fages-Lartaud whom I was lucky enough to share a laboratory cubicle with and Che Fai Alex Wong who was to be found in the laboratory at most hours of the day. I would also like to thank Postdoc Swapnil Vilas Bhujbal for the opportunity to include imaging by confocal scanning laser microscopy in the thesis.

My time in Trondheim would not have been near as enjoyable without my friends and classmates, and the student community at Volvox and Alkymisten and Studentersamfundet. I wish to thank my family and Knut Ørke for always believing in me and for patiently listening to me drone on about bacteria and biology which they had no background to understand. Last but not least I am forever grateful to my grandmother who has supported me and provided advice and comfort in challenging times. I would not have achieved the goals I have today without her.

Oslo, July 2020

Maria Wilhelmsen Hoff

Abstract

Predictable and consistent regulation of protein production is crucial when designing synthetic genetic circuits and for industrial scale productions. To achieve this, it is necessary to identify or design promoter and 5′UTR sequences which produce proteins at a predictable time-points.

The goal of this thesis was to establish and characterize artificial constitutive promoter and 5′UTR sequences in *Vibrio natriegens* and *Escherichia coli*, and to compare the functionality of these sequences in the two organisms. The establishment and use of the non-model organism *Vibrio natriegens* for molecular biology research are still in the early days and methods for working with the organism had to be established for the PhotoSynLab group laboratory environment previous to the establishment of the promoter and 5′UTR libraries. In this thesis, the Gene Expression Engineering method was demonstrated to work well in *Vibrio natriegens* resulting in functional artificial constitutive promoter and 5′UTR sequences.
The advantages working with *Vibrio natriegens* was demonstrated and attributed to the rapid growth time, which is important both in research where time is invaluable and large-scale biotechnology industry both economically and environmentally.
PhD candidate Liza Tietze performed and established several methods making the establishment of *Vibrio natriegens* possible. Several other members of the laboratory group contributed by attempting to use the organism for other research goals.
The constructed pACYC-sfGFP 200N SD plasmid contained an artificial promoter and 5′UTR sequence of 200 random nucleotides containing a Shine-Dalgarno sequence (GGAG). The establishment of 36 unique functional promoter and 5′UTR sequences was identified and characterized by fluorescence measurement of the green fluorescent protein and sequence analysis. Comparison of normalized GFP expression levels in *Escherichia coli* and *Vibrio natriegens* showed that the protein production was variable and consistently higher in *Escherichia coli*. In *Escherichia coli* the majority of the promoter and 5′UTR's were activated during the late exponential to stationary growth phase. Further experiments with a larger dataset are needed to draw definite conclusions on the sequence analysis result and growth experiments for *Vibrio natriegens*. The sequence analysis performed *in silico* by the online software BPROM and Improbizer resulted in inconsistent predictions. The possibility of different functionality of the gene transcription systems of the two organisms for the process of transcription initiation were indicated.

Sammendrag

Forutsigbar og konsekvent regulering av protein produksjon er avgjørende når syntetiske genetiske kretser designes, og ved stor-skala industriell produksjon. For å oppnå dette er det nødvendig å identifisere eller designe promoter og 5′UTR sekvenser som fører til forutsigbar protein-produksjon.

Målet ved denne masteroppgaven var å etablere og karakterisere kunstige ikke-induserbare promoter og 5′UTR sekvenser som førte til kontinuerlig gen-uttrykkelse i *Vibrio natriegens* og *Escherichia coli,* og sammenligne funksjonaliteten av disse sekvensene i de to organismene. Etableringen og bruken av *Vibrio natrigens,* som ikke er en modellorganisme for molekylærbiologi er fortsatt i tidlige faser. Grunnet dette måtte metoder for bruk av organismen etableres for PhotoSynLab gruppens laboratorie-miljø før etableringen av promoter og 5′UTR bibliotekene. I denne avhandlingen er Gene Expression Engineering (GeneEE) metoden demonstrert i *Vibrio natriegens,* og resulterte i funksjonelle kunstige promoter og 5′UTR sekvenser.
Fordelene ved å bruke *Vibrio natriegens* ble demonstrert og tilskrevet den korte generasjonstiden som er essensiell både ved forskning hvor tid er mangelvare og ved stor-skala bioteknologisk industri, økonomisk og miljømessig sett.
PhD kandidat Lisa Tietze utførte og etablerte flere metoder som muliggjorde etableringen av *Vibrio natriegens.* Flere medlemmer av laboratoriegruppen bidro ved å forsøke å bruke organismen til andre eksperimentelle mål.
Det konstruerte plasmidet pACYC-200N SD inneholdt en kunstig promoter og 5′UTR DNA sekvens bestående av 200 tilfeldige nukleotider med en Shine-Dalgarno-sekvens (GGAG). Etableringen av 36 unike og funksjonelle promoter og 5′UTR sekvenser var identifisert og karakterisert ved måling av grønt fluorescerende protein (GFP) og sekvens-analyse. Sammenligning av normaliserte GFP-uttrykkelsesnivåer I *Escherichia coli* og *Vibrio natriegens* viste at protein-produksjonen var variable og konsekvent høyere I *Escherichia coli.* I *Escherichia coli* var hoveddelen av promoter og 5′UTR sekvensene aktivert under sen eksponentiell til stasjonær vekstfase. Det er nødvendig å utføre flere eksperimenter med større datasett for å trekke konklusjoner ved sekvens-analysen, og vekst-eksperimenter for *Vibrio natriegens.* Sekvens-analysen som ble utført *in silico* av den nettbaserte softwaren BRPOM og Improbizer førte til inkonsekvente resultater. Mulige ulikheter i de to organismenes transkripsjonssystemer og ved transkripsjons-initiering ble synliggjort.

Table of Contents

1. Introduction

1.1 Goal and structure of the thesis

The goal of the thesis is to establish and characterize artificial constitutive promoter and 5′-UTR libraries in *Escherichia coli (E. coli)* and *Vibrio natriegens (V. natriegens)*. The characterization is performed by analyzing DNA sequences of the promoter and 5′-UTR and measuring the fluorescence levels of the green fluorescent protein (GFP). First, I will provide background information on the current knowledge on the non-model organism *V. natriegens* followed its utilization in synthetic biology and industry. Then I will detail the relevant knowledge on bacterial promoters, consensus sequences and the role of sigma-factors and 5′-UTR in transcription initiation and translation. The introduction will be concluded by a definition of artificial promoters and outline the current utilization of genetic elements in *V. natriegens* before providing background on the methods utilized in this thesis. The material and methods are separated by organism as these are organism specific and includes additional sections for the establishment of *V. natriegens* for the PhotoSynLab group laboratory. In the results the fluorescence measurements are discussed separately for the two organisms before comparing fluorescence levels and includes imaging by confocal laser scanning microscopy to visualize the GFP expression in *V. natriegens.* The results are concluded by results from sequence analysis by the tools BPROM and Improbizer and a comparison of the output from these tools. The discussion follows the sectioning of the results, and I will attempt to provide an explanation on the functionality of the promoter and 5′-UTR sequences in the two organisms.

1.2 *Vibrio natriegens*

Vibrio natriegens (*V. natriegens*, strain ATCC 14048) is a unicellular gram-negative rod-shaped marine gamma-proteobacterium, earlier classified within both the genus *Beneckea* (1) and *Pseudomonas* (2), before finally recognised as belonging to the genus *Vibrio*. The bacterium *Vibrio cholerae* is found within the genus *Pseudomonas. Vibrio cholerae* is a well-known pathogen, which illustrates the diversity found within the genus (1). *V. natriegens* is facultatively anaerobic (1), first isolated from salt marsh mud in Georgia (3) requires Na^+ (~1,5 %) for growth and has an optimum growth temperature at 37℃ and is non-pathogenic (4). The organism is not yet established as a model organism in molecular biology and biotechnology (5) but has received a surge of interest during the last years due to its discovered applicability, which will be detailed in the next paragraphs (5 ,6).

The organism's rapid generation time, first reported in 1961 by R.G Eagon as 9.8 minutes (4), is the lowest doubling time known of any non-pathogenic bacterium. In recent studies, the doubling time has been measured as low as 9.4 minutes in complex medium (Brain heart infusion broth with sodium, BHIN). In addition to the rapid growth, *V. natriegens* has a versatile metabolism and a biomass-specific substrate consumption rate which is unusually high. Those are some of the characteristics illustrating a high potential in industrial scale fermentations (7).

One of the first publications describing use of genetic tools in the organism was published in 2016 (6) detailed the successful introduction of DNA through bacterial conjugation and transformation of chemical and electrocompetent cells. Protein production of green fluorescent protein (GFP) was demonstrated in addition to the possibility of protein secretion. Another highly relevant finding for use in synthetic biology was that assembly by Gibson cloning, a standard cloning method produced visible colonies after 6 hours and the possibility of DNA isolation after only 3 hours of growth in culture due to the organisms short doubling time (6).

1.2.1 Development of engineering methods and genetic tools in *V. natriegens*

In a preprint made available in 2016, the complete genome of *V. natriegens* was made available, revealing that it was comprised of two chromosomes with a total genome size of around 5.17 Mb and 11 rRNA operons was annotated. Both genome size and the number of rRNA operons exceeded *E. coli*, and there have been indications that rRNA operon abundance is a contributing factor to rapid growth rate (8). Further analysis by genome sequencing and

quantitative PCR lead to the discovery of two prophage regions, VNP1 and VNP2 in the first chromosome. The discovery revealed implications on growth as the prophages were induced spontaneously during cultivation which lead to increased cell lysis. The presence of prophage regions in bacterial genome is common and results in possible economic loss in industrial fermentation processes. Therefore, prophage-free strains were generated which outcompeted the WT in growth experiments. Additionally, the prophage-free strains were more robust towards hypo-osmotic stress and DNA damage, further increasing the industrial applicability of *V. natriegens*. Although the possibilities for improved function in a laboratory environment are promising, the strains have not yet been used as protein producers (9).

As exemplified in the previous section, several favourable characteristics are portraying *V. natriegens* as a promising host for biotechnology applications. However, the functionality of genetic tools in the organism is still being investigated and is there is yet a substantial amount of research required to establish it as a possible model organism (5).

But is there a need to establish a new model organism in biotechnology when *E. coli* could fill this niche? Except for outperforming *E. coli* growth rates, the research on *V. natriegens* is still in the early days. However, although *E. coli* is a well-established gene expression host for most purposes there are challenges with heterologous expression of recombinant proteins or membrane-bound protein complexes (10, 11). There are indications of a superior ability in *V. natriegens* to express functional membrane-bound protein complexes without the need for co-expression of genes or exogenous tools (11, 12).

1.2.2 An example of industrial applicability

Metabolic engineering is pursued to optimize *V. natriegens* as a production host in industrial scale operations (6, 13). Unlike other *Vibrio* species where competence is achieved after induction by environmental factors, *V. natriegens* does not have natural competence. Due to this, efficient methods for making cells competent and transformation must be developed for the organism to compare to established microbial model organisms (13, 14, 15).

The ability of *V. natriegens* to achieve natural competence by the means of genome editing was investigated by Dalia *et al.*, through ectopic expression of the TfoX competence regulator (13). The competence regulator TfoX was found to be induced by chitin in competent *Vibrio* species (16), and ectopic expression resulted in natural transformation of *V. natriegens* (13). The study also achieved to increase the natural accumulation of a precursor of an industrially

relevant polymer Bioplastic precursor poly-*β*-hydroxybutyrate (PHB) (13). PHB is an essential compound in of the production of a biodegradable bioplastic which, due to its similarities to petroleum derived plastics have been produced for the last 30 years (17). Successful rounds of scarless multiplex genomic editing with natural transformation resulted in a ~100-fold increase of PHB production by some genotypes. A ~3.3-fold increase was also achieved by overexpression of a locus related to PHB expression through promoter swaps to regulate several relevant operons by sequential rounds of co-selection (13). This stands as an example of the ability of *V. natriegens* to be optimized for the production of industrially valuable compounds.

1.2.3 Synthetic biology applications of *V. natriegens*

To further establish if *V. natriegens* is suitable as a bacterial model for experimental biotechnology and molecular biology, DNA elements must be introducible with ease and efficiency (18, 5). So far, in addition to the achieved natural competence detailed in the previous section (13), DNA elements have been successfully introduced in *V. natriegens* by most traditionally established methods with promising efficiency. Weinstock *et al.* obtained a transformation efficiency of 10^6-10^7 CFU/ µg plasmid DNA (colony forming units) through electroporation, and transformation efficiency of 10^5-10^6 CFU/ µg plasmid DNA with chemical competent cells. Additionally, plasmid delivery with bacterial conjugation was efficient, while Lee *et al.* engineered a reporter strain and inhibited chromosomal GFP expression with CRISPRi/ dCas9 demonstrating successful targeted gene editing (6, 8). Also, plasmids were successfully and efficiently constructed from multiple fragments by Gibson assembly (6).

Further investigation on the functionality of standard genetic parts and tools used for synthetic biology applications was characterized in *V. natriegens* by Tschirhart *et al.* In the publication, meant as a reference for further experimental work with *V. natriegens*, it was stated that most tested plasmid origins, resistance markers, promoters and ribosomal binding sites were proven as functional (5). The methods used for transformation was similar to those in the publication from Weinstock *et al.*, illustrating the collaborative work done to fully realize the use of this organism in the research community. The plasmids easily transformed into *V. natriegens* without selection included the origin of replications pMB1, ColE1, pUC and p15A, also demonstrated by Weinstock *et al.* Additionally, p15A and pUC were successfully replicated and maintained together, an important factor for synthetic biology work. In the case of

synthetic creation of genetic circuits which depend on multiple plasmids, compatibility of the origin of replication is crucial. Plasmid copy numbers were highest when cultivated at 37℃, decreasing when grown over longer periods such as overnight. There were explainable deviations in copy numbers from *E. coli*, both in the negative and positive direction, however plasmid copy number did not correlate with plasmid maintenance which remained unexplained. Suggested antibiotics for selection were ampicillin (100 µg/ml), kanamycin (250 µg/ml), chloramphenicol (6 µg/ml) and tetracycline (10 µg/ml), while resistance towards gentamycin was proven as high as 30 µg/ml (5).

1.3 Bacterial promoters and consensus sequences

One of the central applications of synthetic biology is the optimization of protein or product production through control of gene regulation. The regulation part of gene expression targeted by researchers is most commonly transcription and tools engineered for regulation of promoters, terminators and transcription regulators.

As the processes of genetic regulation are most studied and described in the model organism *E.coli*, this section will detail the genetic elements from this origin, before looking into how genetic elements in *V. natriegens* are structured, in particular promoters. The challenge attempted by many biologists within the field of synthetic biology, is to create a genetic sequence resulting in the ability to regulate protein production in with a predictable outcome. One of the central aspects is to achieve this without disturbing the biochemical functions of the system as a whole (19).

The well characterized process of gene expression involves several steps leading to the ultimate production of proteins. To obtain a flow of the information encoded in DNA, RNA is synthesized. The RNA copy can in some cases be the ultimate product but usually it carries the genetic information transcribed into messenger RNA (19) which is translated into proteins. The essential part of the process is the regulation of gene expression; how, where and when genes are transcribed and translated into proteins (20). The regulation of transcription can lead to more efficient use of resources by the cell, given that the production of certain proteins would increase fitness in the particular conditions. Another highly relevant aspect is possible downregulation of less utilized enzymes or the production of undesired compounds (metabolites) (19). Prokaryotic promoters are by definition “the sequence to which RNA polymerase binds” (20). The promoter DNA sequence, of 100-1000 bp of length (21) placed in front of the untranslated region, upstream of the genes or operon (20) and facilitates the

production of mRNA (19). RNA polymerase (RNAP) is a holoenzyme constructed of two components which synthesises RNA, that in bacteria are the sigma subunit and the core enzyme composed of five subunits (α, α, β, β' and ω) (20).

For the bacterial RNAP to recognize, bind to and initiate transcription the promoter includes several consensus sequences at defined regions found at the core in most characterised *E. coli* promoters (19, 20). Among the best characterised are the promoters recognized by the *E. coli* sigma-70 (σ^{70}) subunit of RNAP (19). The consensus sequences are found at –35 (TTGACA) and –10 (TATAAT) bp from the first transcribed base, at the +1 position of the promoter's end (19-22).

The strength of a promoter depends partly on the similarity to the consensus sequences (20) and the number of bp separating the two sequences at –10 and –35 (22) which for *E. coli* is 17 bp optimum (19). Only a few genes include the specific promoter consensus sequences described, but the –10 and –35 regions differ by only 3-4 bp before recognition by the sigma subunit of RNAP is unlikely (20, 23). Mutations affecting the strength of transcription are found in the promoter regions (22). The actual transcription of DNA into RNA begins at the genes transcription start site (TSS) located at the 3′ end of the promoter. The TSS is designated +1 and from there the mRNA is produced until transcriptional terminator (24, 20). In the case of strong promoters, some have an additional recognition element –40 to –60 upstream of the TSS which is rich in A+T and in contact with the α-subunit of σ^{70}. Transcription termination occurs after a signal, either intrinsically by the RNAP interacting with a G+C rich stem-loop secondary structure followed by U-residues in the mRNA. The second possible signal for termination in *E. coli* is by the protein Rho which binds to a conserved region in RNA and forces the transcription complex to dissociate (24).

1.3.1 The role and variability of σ-factors

In prokaryotes, also within species the RNAP multidomain σ -factors are variable in structure and ability to recognize different promoter sequences initiating transcription (23). The ability to recognize specific DNA regions is due to the interchangeability of σ-factors as the consensus sequence of the promoter varies with the σ-factors used by RNAP (21, 25). σ-factors originating from prokaryotes are classified in two families by comparison to the σ-factors found in *E. coli* and through shared homology. The σ-factors in *E. coli* are the previously mentioned σ^{70} which initiated transcription for normal growth purposes and the

σ^{54}, responsible for responses to environmental changes. The σ^{54} also requires enhancer proteins and use of ATP to initiate the process of transcription. The activity and expression of σ-factors are regulated particularly at the post-translational level (23). Through examination of σ-factors in other bacteria, it is evident that a large number of alternative σ-factor groups exist both within and among species which are used to induce gene expression in concert. Several conserved regions or sub-domains are shared by the different σ-factor groups each interacting with RNAP and defined promoter elements (23, 25).

A comparison of *V. natriegens* and *E. coli* rRNA operon P1 and P2 promoter sequences found that *V. natriegens* share consensus with *E. coli* at the –10 and –35 positions. The distance between the consensus regions differed by only 1 bp from *E. coli*, defined as 16 bp. In addition, as in *E. coli* an A+T rich sequence was found at the –40-position upstream. Another interesting observation was a sequence similarity from the –10 position to the transcription start site (26). The sequence in question, the ‘’discriminator’’ (19) is subject to mutations leading to transcription initiation regulation as previously mentioned (22, 26).

1.3.2 The 5′-untranslated region (5′-UTR)

The 5′-untranslated region (5′-UTR) is located between the +1 transcription start site (TSS) and the translational start site of the mRNA with a 3′-UTR in the opposite end, as shown in **figure 1.1**. These stretches of mRNA are not translated into protein but contain regulatory regions (20). The regulatory regions of the 5′-UTR and the promoter controls gene expression at different levels. While the promoter controls transcription, an established fact has been that the 5′-UTR regulates translation, thereby the rate of translation initiation, efficiency and mRNA lifetime (27, 28).

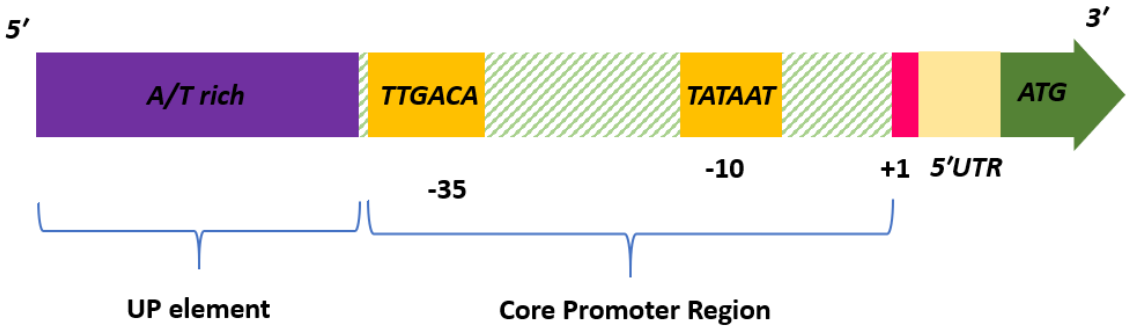


Figure 1.1: Schematic view of the elements of a standard prokaryotic promoter. Two hexamers with conserved motifs are located at the -10 and -35 position (yellow boxes) upstream of the Transcription Start Site, +1 (pink box). The conserved motifs are RNAP binding regions. An Adenine/ Thymine rich upstream element (UP element, purple box) can be present. The UP element is believed to interact with the RNAP and increase transcription rate. The shaded areas are sequences of spacer DNA. The 5′UTR indicated by a yellow box and the coding sequence is indicated by ATG (green arrow).

Figure 1.1 was adapted from figure 1 in (29).

The role of the promoter at the transcription level is often highlighted, while the function of the 5′-UTR at the translational level is often omitted in literature by comparison. Although these genetic elements are often created simultaneously, the 5′-UTR and promoter are not one component. In the world of synthetic biology focusing on engineering of promoters and 5′-UTR, the regulatory sequences are often separated in core promoters, RBS and 5′-UTR (24, 30, 31). New discoveries are indicating that we may be moving towards a shift in the established dogma towards a more holistic focus on the promoter and 5′-UTR role in transcription and translation. Previously acknowledged to function independently, a publication focusing on transcription and translation initiation in *E. coli* found that creation of

junctions coupling the promoter, 5′-UTR and gene of interest could improve protein synthesis (31). Indications that the 5′-UTR could have a role in transcription as well as in translation have surfaced as *E. coli* UTR-mutants resulted in increased expression at both levels (32).

It is *now* believed that interactions between promoters and the 5′-UTR could lead to variations in gene expression (19). By including synthesised sequences beyond the TSS of the promoter, regulation of translation initiation and mRNA lifetime are observed in combinations of promoter and 5′-UTR. Possibly altered promoter strength have been observed in specific promoter and 5′-UTR combinations. Such findings indicate a more complex and intertwined relationship between the promoter and 5′-UTR than previously acknowledged (32, 33).
In prokaryotes, gene expression is regulated first at the step of transcription initiation. During transcription initiation, promoter sequences are recognized by σ^{70}-factors in the RNAP. The promoter controls transcription of an operon consisting of ORFs with sequences transcribed into the ribosomal-binding sites (RBS). Variation in the RNAPs activity have been linked to differences in promoter sequences. Downstream of the coding sequence in the operon is the transcription terminator. After the production of the mRNA transcript by the RNAP, the RBS is bound by ribosomes during translation (19).

The ribosome, a complex of rRNA (ribosomal RNA) and proteins synthesises the protein product during translation. The ribosome has catalytic function and translates the mRNA transcript to proteins by reading its nucleotide composition, which often begins during transcription. Ribosomes consists of two subunits, one small (30S) and one large (70S) which assemble on a start codon. The translation continues until a stop codon is encountered (19-22, 34). The ribosomal-binding site (RBS) or Shine-Dalgarno (SD) sequence is located in the 5′-UTR of prokaryotes. A complementary sequence is found at the 3′-UTR of the 16s ribosomal RNA in the small subunit, the anti-Shine-Dalgarno sequence. The RBS indicates which start codon (AUG normally) that begins the coding sequence, the first AUG located downstream of the RBS (20).

The inconsistent function of RBS in translation initiation of coding sequences poses a challenge for synthetic biology engineering and effects the quantification of protein production (31). While some translation initiation rates can be predicted accurately, there are variations resulting in large differences in protein synthesis rates and protein abundance. As multiple ribosomes bind to mRNA consecutively during consistent translation initiation, the mRNA is unable to refold as fast as the binding of the ribosome itself occurs. In a publication

by Borujeni *et al.* (2020) implications of kinetics and energetics for mRNA folding/ refolding together with the rate of ribosome binding lead to the reported variations in translation initiation. The resulting rate-predictions of protein synthesis was altered over 1000-fold (34).

As outlined above, the rate limiting step in protein synthesis is translational initiation (34). During this step, RBS hybridization of the SD with the anti-SD sequence assist correct positioning of the ribosome (35). In *E. coli*, the RBS is a purine rich sequence with the most frequent motif 5′-*AGGAG*-3′, with a 7 bp optimum distance from the start codon (20, 35). Efforts to regulate translation initiation through creation of RBS libraries have been made, which will be exemplified in the following section (36).

1.4 Development and utilization of artificial promoters

As promoter sequences in one organism are found to be functionally divergent in other organisms it would be valuable to find and provide the host organism with consensus promoters when expressing cloned exogeneous genes (20). Increased knowledge of the mechanisms of protein production enabling engineering of microorganisms for optimising gene expression is central to synthetic biology development (36). Continued focus on translation initiation (34, 35, 36) have led to the possibility of realizing use of microorganisms not only as microscopic factories, but to improve the understanding of how we can reduce our environmental footprint (36, 37). The standard toolbox for synthetic biology includes (but are not limited to) RBS libraries (36, 38), inducible or constitutive promoters and engineered 5′-UTR regions (5, 20, 39, 40). As the utilization of these genetic elements progress, the importance of a holistic approach for sustainable and efficient expression is materializing (24).

Researches focusing on the synthesis of genetic elements for gene expression control are finding that the elements function are highly variable due to context dependency. A struggle to validate the use of created genetic elements across research communities are emerging. In many cases, the predicted output in the form of the amount of transcript produced is inconsistent. This was identified by Mutalic *et al.* (2013) to partly be caused by interactions between the elements involved in translation initiation and the genes downstream (31).

Promoters have an extensive role in gene expression through control of transcription in response to stimuli, determining if transcription is turned on or off and at what rate (5, 35).

Construction of synthetic promoter DNA sequences are proving useful for expression of biosynthetic genes as natural promoters often lack the ability to function properly in an engineered system. Due to the independence from endogenous transcription regulation which could interfere with the hosts gene transcription, use of synthetic promoter sequences have been proven advantageous (40, 35). Synthetic sequences can include synthetically engineered promoters with wildtype sequences (29, 36, 39). However, there are challenges with inconsistent expression and responses to environmental conditions (40, 41).

To fully explain the novelty of the promoter and 5′-UTR sequences utilized in this thesis, it is crucial to define what separates the established synthetic regulatory sequences. Publications referring to use of synthetic randomized sequences often includes standardized genetic parts, which consist in part of known DNA sequences. These sequences originate from natural originated constitutive promoters found in well-characterized model organisms as *E. coli*. When creating synthesized promoter libraries, more often only the flanking regions are mutated while known consensus motifs are kept, in an attempt to regulate promoter activity (29).

The different types of synthetic promoter sequences are categorized by the method utilized creating the libraries. These include hybrid regulatory sequences, mutated sequences, semi-artificial sequences and artificial sequences (29, 36), defined in **table 1.1** below. Synthetic promoter sequences currently utilized normally does not have a randomized 5′-UTR sequence and consist of the binding site for the σ-factors and the 17-bp spacer (40, 42). In a pending publication, Lale *et. al* demonstrated that inclusion of artificial 5′-UTR sequences in an artificial promoter system permits modulation of expression in response to environmental cues (39).

Table 1.1: Categories and definitions of synthetic randomized promoter sequences. The methods utilized include targeted randomization of either/ or the promoter sequence, 5′UTR and gene. Table 1 defines artificial sequences as the only case where no natural occurring sequences are included in either the promoter or 5′UTR (29, 36, 39). The table is inspired by figure 1. in an unpublished review on synthetic biology by PhD candidate Lisa Tietze.

Category name	Category of random nucleotide sequence defined.
Hybrid	A hybrid promoter is created by assembly of multiple previously characterized parts. These parts can originate from promoters, fusion of an operon with a promoter, exchange of core promoters and motifs, or different combinations of promoter and 5′UTR sequences.
Mutagenesis	A mutated promoter sequence originates from characterized promoter sequences, as natural promoters. To create a synthetic promoter, the flanking regions are mutated by introducing single nucleotide mutations. The method used to obtain mutation is error-prone PCR, producing large variations of promoter sequences.
Semi-artificial	A semi-artificial promoter sequence is created by maintaining core parts of known regulatory sequences, while mutating the surrounding regions or flanking regions. The method used is called saturation mutagenesis and results in randomized flanking regions and a core promoter with known consensus motifs. In semi-artificial promoter sequences randomization is introduced to a larger extent than in mutated promoter sequences.
Artificial	An artificial promoter sequence does not contain known or previously characterized sequences and does not occur naturally. Artificial promoters are regulatory sequences consisting completely of random nucleotides. The ArtPromU inserts in this thesis consist of artificial promoter and 5′UTR sequences with 200 random nucleotides and a Shine-Dalgarno sequence.

1.4.1 Promoters and genetic elements in *V. natriegens*

The strength of promoter sequences is influenced by the ability of RNA polymerase to bind. Promoter strength is crucial for the efficient *in vivo* and *in vitro* gene expression and promoters contains several binding sites ensuring proper transcription in varying environmental conditions. This will ultimately lead to higher amounts of protein production (43, 24). In *V. natriegens*, the native ribosomal promoter P1 results in high expression levels of proteins and results in some of the highest measured protein production in *E. coli* as well as in *V. natriegens* (5). To optimize gene expression, regulation of transcription is crucial for economic and viable protein production (28). Inducible promoters demonstrated to function in *V. natriegens* are the IPTG-inducible promoters ptac (5), *lacUV5* and *trc* (6), the arabinose-inducible promoter P_{BAD} (5, 6), the anhydrotetracycline inducible promoter P_{Tet} (5), the temperature inducible phage λ *pR* promoter (6) and additionally light inducibility was possible through the P_{Dawn} system (5). Relevant for synthetic biology research and the possibility of computational analyse, multiple well-established constitutive promoters were tested, including promoters from the Anderson library (5, 44). The efficiency and functionality of both inducible and constitutive promoters were variable compared to *E. coli*. Several of the promoters resulted in significantly lower responses in *V. natriegens*. Most of the inducible promoters were leaky compared to in *E. coli*, which could not be explained. This could be caused differences in the organism's transcription and translation processes. The ribosomal binding site strengths were similar in both organisms (5).

The challenges observed by Tschirhart *et. al* (2019) with lack of transferability utilizing genetic parts optimized for use in *E. coli* is far from unique. As synthetic promoters often include part of natural promoters from model organisms, the functionality across bacterial species have been observed highly variable in several studies (5, 6). An additional factor which have been less focused on is the potential of upstream sequences beyond the promoter to contribute to this decrease in functionality. By acknowledging the impact of the genetic context, it could be valuable to extend the definition of regulatory sequences beyond the core promoter (40, 37).

Context-dependency of synthetic promoters are now being increasingly examined, a previously ignored factor resulting in variable functionality. Context-dependency can be explained by a simple statement: As the functionality of a promoter sequence in the organism which it originates from naturally are influenced by interactions with regulatory proteins, extracting the sequence from this environment probably results in loss of activity-

predictability. In order to bypass this challenge, researchers have increased the use of randomized sequences as part of the solution, while maintaining core elements of natural promoters. This may work for use in model organisms, but less for non-model organisms where the conserved motifs could differ considerably. This may be an argument for excluding previously characterized sequences completely, and rather focus on the identification of functional artificial promoter consisting of randomized nucleotides alone (29, 37, 39).

As the results reported are showing promising trends for use in molecular and synthetic biology experimental work as well as for industrial scale production, one could argue that there is a value of continued research into the establishment of *V. natriegens* as a potential microbial model organism (5, 6, 13). The findings also support the need for more established promoter sequences for the continued development of *V. natriegens*, which potentially should include increased focus on artificially promoters (5, 6, 39).

1.5 The GeneEE method

The method utilized in this thesis to establish artificial constitutive promoter libraries with 5′-UTR in *E. coli* and *V. natriegens* is a newly developed Gene Expression Engineering (GeneEE) method by Lale *et al.* The preprint is currently available at bioRxiv (39). The goal of GeneEE is the generation of novel and gene-specific artificial promoters and 5′-UTR (ArtPromU) that recruits the host's own native transcription and translation machinery.

Promoter and 5′-UTR sequences were constructed of 200 random nucleotides in two different plasmid DNA libraries, with or without an additional SD sequence (GGAG), believed necessary for ribosome recruitment for translational initiation of *E. coli* chromosomal genes. (As there is N^4 possible combinations of the four nucleotides A, T, G or C the authors postulated that 1 in 325 DNA segments within 200 random nucleotides probably would lead to transcription initiation). The calculation was based on an estimate that eight nucleotides could be sufficient for initiation of transcription as the σ^{70}-factor of *E. coli* requires interaction with six nucleotides.

The random 200N ArtPromU inserts were placed directly upstream of the genes coding sequence. The wildtype promoter and 5'-UTR were removed from the plasmid. Immediately after the ArtPromU insert followed the gene of interest translational start site (ATG). For confirmation and characterization of the ArtPromU inserts functionality, the inserts were cloned in front of a fluorescence reporter gene (mCherry) for measurements of the protein production levels form each clone, which resulted in variable expression levels. DNA and RNA sequencing were performed for experimental analyse of the DNA sequences resulting from ArtPromU inserts which localized multiple transcription start sites (TSS). Downstream of the TSS from +1 to +25, promoter motif analysis identified unique motifs and led to the detection of multiple σ-factor motifs.

The versatile applicability of the method was demonstrated by successful introduction of the ArtPromU (-SD) inserts to seven microorganisms (*Pseudomonas putida, Thermus thermophilus, Corynebacterium glutamicum, Streptomyces albus, S. lividans* and the yeast *Saccharomyces cerevisiae*) (39). Recently, the master thesis of Andreas Lykke (2019, PSL NTNU) described successful introduction and characterization in *Bacillus subtilis* (46).

1.6 Experimental design

The goal of this thesis to establish and characterize ArtPromU sequences in *E. coli* and *V. natriegens* utilizing the GeneEE method described in the previous section. Additionally, the use of the organism *V. natriegens* was established for the laboratory by the work done for the thesis. This next section will describe the technicalities of the method, concluded by the workflow for promoter characterization based on DNA sequence and reporter gene expression levels. The process is detailed in **figure 1.2** on the following page. The inserts were assembled into the plasmid by Golden Gate Assembly, with simultaneously restriction digest and ligation (45). Golden gate requires a plasmid for ligation of the insert to create a new plasmid (45). A plasmid backbone was created by eliminating the plasmids native promoter and 5′-UTR. In addition, any BsaI restriction sites were removed from the plasmid backbone. A backbone was created by primer design and PCR introducing BsaI restriction sites at the 5′-ends. The forward primer (appendix) was designed to bind to the reporter gene start to exclude the promoter and 5'-UTR, while the reverse primer hybridizing upstream of the native promoter. To avoid the backbone from re-ligating unto itself, uncomplimentary sticky ends were designed (45). The resistance gene and OriT on the backbone were kept for selection on antibiotic plates.

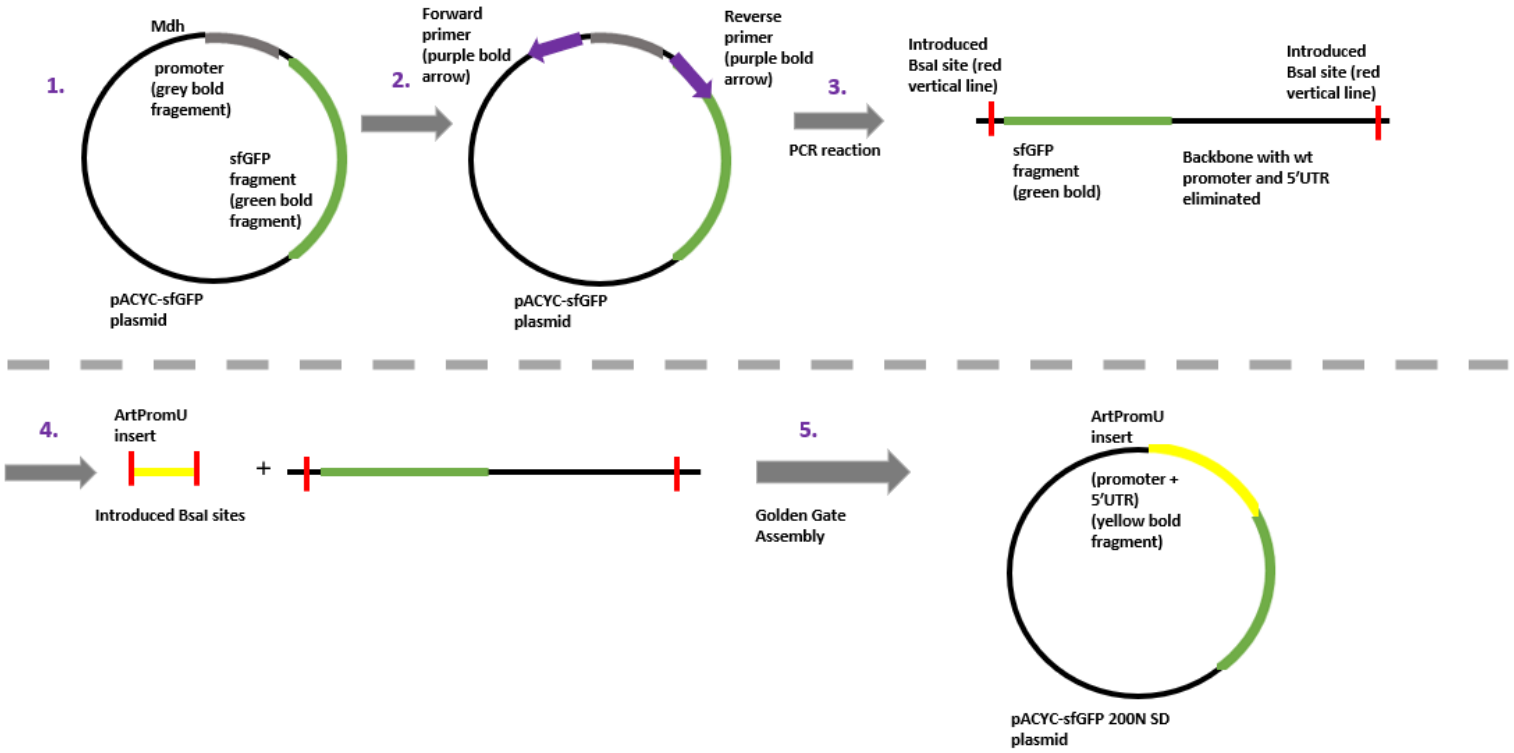


Figure 1.2: Overview of the workflow creating the pACYC-sfGFP 200N SD plasmid. In order to characterize the artificial promoter and 5′UTR sequences, the ArtPromU insert was ligated in front of the sfGFP reporter gene. 1. Simplified version of the original pACYC-sfGFP plasmid with sfGFP fragment and mdh promoter. 2. The forward primer binding to the reporter gene start to exclude the promoter and 5'-UTR, and the reverse primer hybridizes upstream of the native promoter. 3. The native promoter and 5′UTR are excluded by the primer design in a PCR reaction. BsaI sites are introduced and a linearized backbone is the product. 4. The ArtPromU insert with BsaI sites are ligated with the backbone product by Golden Gate Assembly. 5. Ligation of the fragments in correct position results in removal of the BsaI sites, resulting in the hybrid plasmid pACYC-sfGFP 200N SD. The workflow overview was inspired by figure 2. in chapter 9 of DNA cloning and Assembly (45).

The ArtPromU inserts ends were created complementary to the backbone sticky ends by BsaI restriction digestion. The inserts consisted of a BioBricks (47) prefix and a BsaI site in front of the 200N +SD sequence (211 nt) (39), followed by a second BsaI site and a BioBricks (47) suffix. To avoid changes in the random nucleotide sequence of the promoter and *5'UTR*, cloning was performed without introduction of scars in the inserts. If a method that introduced scars to the nucleotide sequence was utilized, the insert would not be completely randomized after the cloning process (39). The ArtPromU inserts were ligated directly in front of the reporter genes translational start site (ATG) on the backbone plasmid, resulting in a closed circular hybrid plasmid. The ligation resulted in removal of the BsaI sites and prevented continued cuts, as illustrated in **figure 1.3**. After Golden Gate Assembly (described in the methods) the reactions were transformed into *E. coli*.

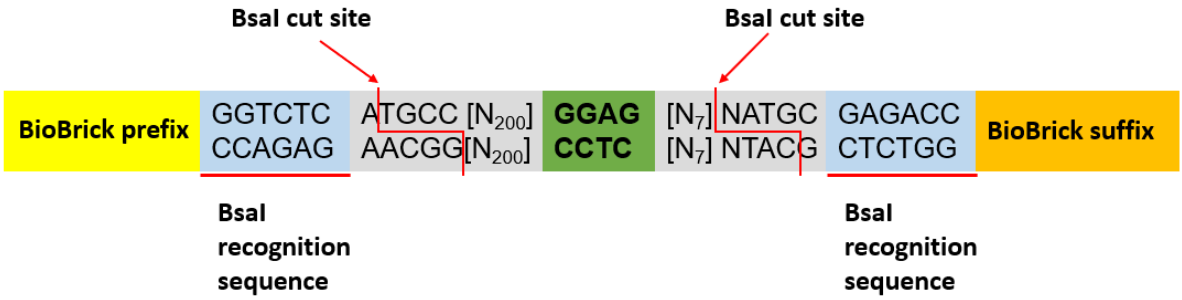


Figure 1.3: The GeneEE nucleotide sequence for restriction cloning with BsaI sites. Flanking the insert sequence are the BioBrick prefix – and suffix in yellow and orange boxes (sequence not detailed).

1. After the BioBrick prefix is a BsaI recognition sequence. 2. Cut sites for BsaI indicated by red figure and arrows in grey box, followed by the 200N random nucleotide sequence (N $[_{200}]$), light blue box. 3. Shine-Dalgarno sequence in bold (**GGAG**) located towards the end of the random nucleotide sequence, green box. 4. Residual 200N sequence and the gene start (ATG) followed by BsaI cut sites, grey box. 5. BsaI recognition sequence in grey box, followed by BioBrick suffix. Figure 1.3 was adapted from figure 1. in supplementary online material from the article by Lale *et. al* (39).

As the promoter inserts were ligated directly upstream of a sfGFP-reporter gene, the appearing colonies with functional promoter and 5′UTR insert sequences were expected to display green fluorescence under UV light. After confirmation of correct expected product length by colony PCR and agarose gel electrophoresis the plates were scraped, plasmid DNA purified, and promoter libraries transformed into *V. natriegens*. When transformants appeared, the *V. natriegens* colonies were used to inoculate 96-well plates which were screened for sfGFP-expression to identify functional promoter and 5′UTR's. By replica plating of the 96-well plates, clones expressing sfGFP in the range of low to medium to high were picked and further re-introduced into *E. coli*. Characterization of relative expression was performed by comparing time dependent expression levels of clones in *V. natriegens* to *E. coli*. The promoters were sent for sequencing and the characterization was concluded by sequence analysis for promoter motifs. Further experimental details are provided in the materials and methods section.

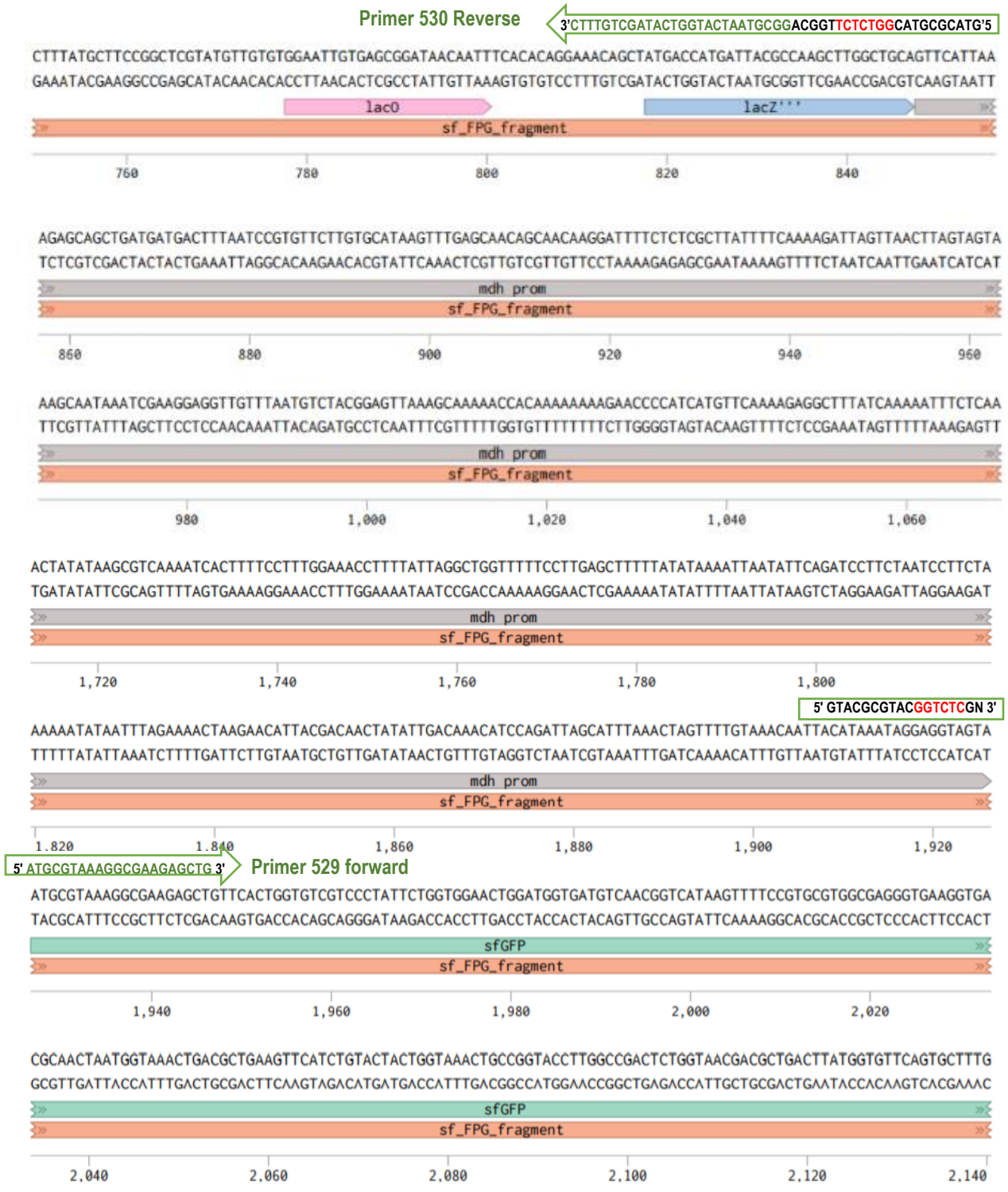


Figure 1.1: Overview of the sequence for the area eliminated from the pACYC-sfGFP backbone plasmid by PCR. The primers (529 forward and 530 reverse direction) are complementary to the plasmid. The BsaI restriction sites on the 5' end of the primers is coloured red. The primers have uncomplimentary overhangs around the restriction sites to avoid creation of primer dimers (sequence coloured black). Arrowheads indicate the direction of the primers. The area withheld from the PCR product resulting from the forward/ reverse primers 529 and 530 are approximately from 860 bp to 1,900 bp. The BsaI restriction enzyme cuts outside its restriction site, creating sticky ends.

2. Materials and methods

The materials and methods are divided in separate main sections for the two organisms due to different growth condition requirements. In this chapter I will first present the methods for growing, cloning and transforming *E. coli* and *V. natriegens*. Section 2.2 includes the protocols, antibiotics and plasmids tested to establish the growth conditions and transformation protocol for *V. natriegens*. In the second section, the experimental design for fluorescence measurements to characterise promoter activity in *E. coli* and *V. natriegens* are detailed. This chapter is concluded by a section on image acquisition with confocal microscopy for visualisation of sfGFP expression in *V. natriegens* and the methods for analysis of the promoter and 5′-UTR sequences.

2.1 Materials and methods in *E. coli*

The artificial promoter library and 5′-UTR was cloned with Golden Gate Assembly PCR and transformed into *E. coli*. After confirmation of expected product lengths by colony PCR and agarose gel electrophoresis, *V. natriegens* was transformed with the pACYC-sfGFP 200N SD promoter and 5′-UTR libraries. Plasmids were sent for DNA sequencing to confirm correct assembly of promoter and 5′-UTR sequences.

2.1.1 Media and growth conditions

Chemical competent *E. coli* cells was prepared based on the protocol by R, Green *et al.*, edited by Lisa Tietze, in section 2.1.2. The media recipes are described in table 2.1, and the protocol for heat-shock transformation is described in section 2.1.3 (48).

Table 2.1: Components and concentrations for media utilized preparing chemical competent cells and heat-shock transformation in *E. coli.* The media solutions were adjusted to the correct volume with dH_2O before sterilization.

yB Media (500 ml)	**Comments**
2.5g yeast extract	
10g tryptone	Adjust the pH to 7.6 with KOH
0.38g KCl	After autoclavation, add autoclaved $MgSO_4$ (17 ml)
SOC solution (1 L)	
20g tryptone	
5g yeast extract	
0.584g NaCl	
0.186 KCl	Autoclaved, stored at 4℃ in the dark
TfBI solution (500 ml)	
1.47g potassium acetate	
4.95 g $MnCl_2$	
6.05 RbCl	
0.74g $CaCl_2$	Adjust the pH to 5.8 with acetic acid (throw away if overshot).
75 ml glycerol (99.95 % bi-distilled)	Filter sterilized, stored at 4℃
TfBII solution (100 ml)	
100 mM MOPS (10 ml, 209.26 g/mol, free acid)	
1.10g $CaCl_2$	
0.12g RbCl	Adjust the pH to 7.0 with KOH
15 ml glycerol (99.95 % bi-distilled)	Autoclaved, stored at 4℃ in the dark

2.1.2 Protocol for preparation of chemical competent cells

The afternoon before preparing competent cells, SOC media (20 ml) in a sterile 125-ml baffled Erlenmeyer flask was inoculated with a colony of DH5a-cells, grown at 37℃ overnight in a shaking incubator (225 rpm). The following morning, the overnight culture was transferred to pre-warmed yB-media (300 ml, 37℃) separated in two 250-ml baffled Erlenmeyer flasks. Growth was continued until OD_{600} of ~0.4-0.6, after 1.5-2 hours. From this step, all solutions and cultures were kept on ice. The cells were chilled on ice for 5 minutes before transfer to six chilled 50 ml Falcon tubes, and centrifuged for 10 minutes (4000 rpm, 4℃).

After centrifugation the supernatant was discarded, and the cells were resuspended in TfBI (15 ml). The centrifugation step was repeated before the cells were resuspended gently in TfBII (1 ml). The competent cells were aliquoted (100 µl) by pipetting into sterile 1.5 ml Eppendorf tubes and stored at -80℃. The tubes had been frozen beforehand to ensure flash-freezing of the cells.

2.1.3 Heat-shock transformation

An aliquot of competent cells was thawed on ice for 5-10 minutes, and 1 µl of plasmid DNA was added to the cells and gently mixed. The cells and DNA were incubated on ice for 30 minutes. After incubation, the cells were heat-shocked for 45 seconds at 42℃, and incubated on ice for 3 minutes. LB medium (900 µl) was added to the cells before 60 minutes of incubation in a shaker (37℃, 225 rpm). After incubation, 100 µl of cells were plated on selective media (LB with 1,5 % agar). The plates were incubated overnight at 37℃.

After successful transformation, a single colony was inoculated ON with 4 ml LB media and antibiotics. For the pACYC-plasmid constructs, 4 µl of chloramphenicol from 25 mg/ml stocks were added. 500 µl of culture was used to prepare glycerol stocks of plasmids, before DNA plasmid isolation. Plasmids were isolated with the QIAprep Spin Miniprep Kit. Isolated plasmids were stored at -20℃ (49).

2.1.4 Golden Gate Assembly

A modified version of the plasmid pACYC184 in figure 2.1, the pACYC-sfGFP plasmid in figure 2.2 was the vector used for all cloning procedures. Plasmids were constructed *in silico* with the software Benchling. The plasmid pACYC-sfGFP contains a chloramphenicol antibiotic resistance gene in *E. coli* and *V. natriegens*. The native promoter in front of the sfGFP gene is the mdh promoter, and the origin of replication is the low copy number oriT p15A. The reporter gene superfolder-GFP expresses fluorescence constitutively in both organisms. The superfolder version of the GFP protein was reported to be more resistant to denaturation due to chemical or thermal conditions, and robust against misfolding (50).

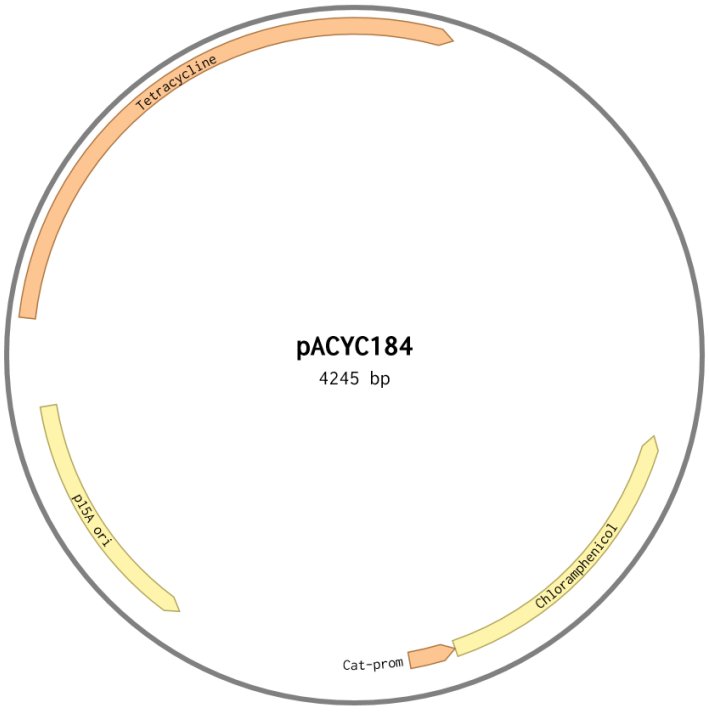


Figure 2.1: The pACYC184 plasmid (4245 bp) with the native Cat-promoter, a chloramphenicol antibiotic resistance gene and the p15A origin of replication.

The pACYC-sfGFP plasmid

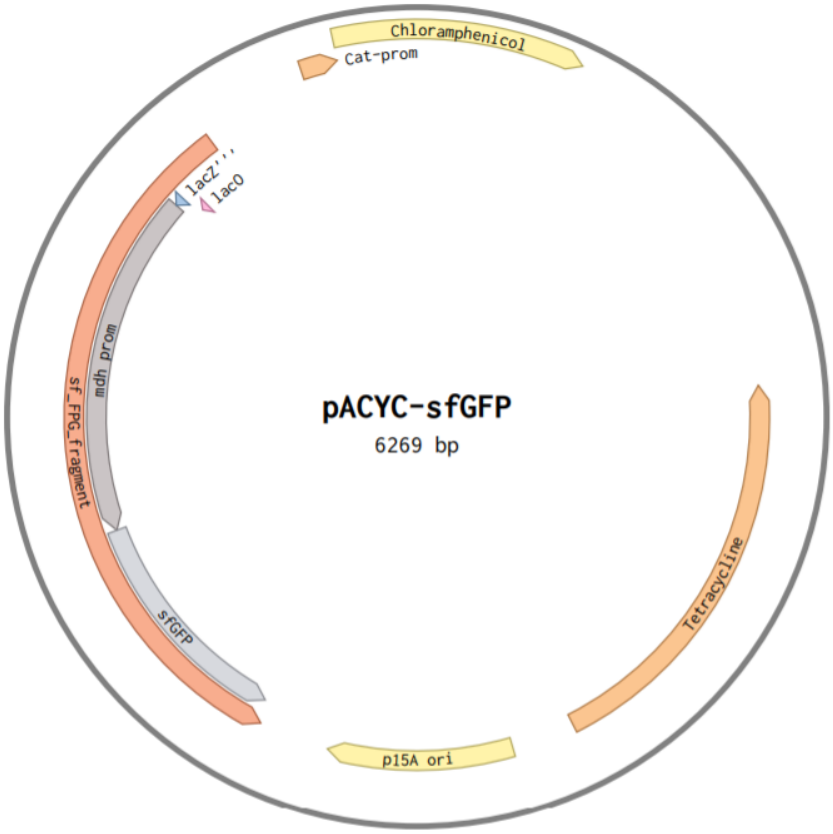


Figure 2.2: The pACYC-sfGFP plasmid (6269 bp) with the native mdh promoter, a chloramphenicol antibiotic resistance gene, the p15A origin of replication and the sf-GFP reporter gene.

Two constructs were prepared by PCR with the pACYC-sfGFP plasmid before assembly by Golden Gate cloning. A backbone where the native mdh promoter and 5′ UTR was eliminated and a closed circular backbone plasmid. The resistance gene and oriT was maintained in the backbone, but the size was reduced for the ability to identify mutants on gel without sequencing. The ends of the backbone had to be complementary to the insert to avoid self-ligation of the backbone. The 200N SD insert sequence was placed in front of the translation start site of the sfGFP reporter gene.

2.1.5 Creating the pACYC-sfGFP backbone and 200N SD insert

The DNA backbone was created by Takara CloneAmp^{TM} HiFi PCR due to low concentration of the pACYC-sfGFP plasmid (template DNA). The primers used for the PCR were ALY529 forward and ALY530 reverse.

The components used is displayed in table 2.2, and the PCR thermocycler program is detailed in table 2.3.

Table 2.2: Components and amounts for 1 reaction of Takara CloneAmpTM HiFi PCR for elimination of the mdh promoter to create a linearized pACYC-sfGFP backbone. The primers sequences are found in the appendix.

Component	Amount
CloneAmp^{TM} HiFi PCR premix	12.5 µl
10 µM forward primer	0.5 µl
10 µM reverse primer	0.5 µl
Template DNA	1 µl (<100 pg)
dH_2O	Fill to 25 µl

Table 2.3: Takara CloneAmpTM HiFi PCR thermocycler programme for amplification of the pACYC-sfGFP backbone without the mdh promoter. Step 2-4 (denaturation, annealing and extension) was repeated for 30 cycles. The expected product size was approximately 4600 bp.

Step	Time	Temperature	Cycle
1. Initial denaturation	20s	98℃	
2. Denaturation	10s	98℃	
3. Annealing	15s	55℃	Repeat step 2-4 30X
4. Extension	25s (5s/kb)	72℃	
5. Final extension	60s	72℃	
Hold		4℃	

The 200N SD random artificial promoter library insert was created with a PCR reaction displayed in table 2.4 based on the standard NEB Q5 High-Fidelity DNA polymerase protocol. The primers used for the PCR reaction was the BB prefix forward primer and the BB suffix reverse primer. The 200N SD R library (diluted 1:100) was the template DNA for the insert. The PCR thermocycler programme is displayed in table 2.5 (51).

Table 2.4: Components and amounts for one reaction of Q5 polymerase PCR for amplification of the 200N SD insert. The primers sequences are found in the appendix.

Component	Amount
5x Q5 Reaction buffer	10 µl
10 mM dNTPs	1 µl
10 µM Forward primer	2,5 µl
10 µM Reverse primer	2,5 µl
Template DNA	1 µl
Q5 High-Fidelity DNA polymerase	0,5 µl
dH_2O	32,5 µl

Table 2.3: Q5 polymerase PCR thermocycler programme for amplification of the 200N SD insert. Step 2-4 (denaturation, annealing and extension) was repeated for 7 cycles. The expected product size was approximately 250 bp.

Step	Time	Temperature	Cycle
1. Initial denaturation	30s	98℃	
2. Denaturation	10s	98℃	
3. Annealing	30s	71℃	Repeat step 2-4 7X
4. Extension	20s	72℃	
5. Final extension	1 min	72℃	
Hold		4℃	

DpnI digestion of the backbone fragment

After the backbone PCR, the reaction was digested with the DpnI restriction enzyme, for elimination of the template DNA. The restriction enzyme cleaved methylated DNA, while the unmethylated PCR product remained intact. For thorough elimination of the template DNA the PCR reaction in table 2.6 was incubated with the enzyme at 37℃ overnight (52).

Table 2.6: Components and amounts used for DpnI digestion for removal of the DNA template used in the PCR reaction to create the pACYC-sfGFP backbone.

Component	Amount
PCR reaction	25 µl
DpnI	1 µl
5X CutSmart buffer	5 µl
dH_2O	20 µl

2.1.6 Creating a closed backbone plasmid

A closed circular backbone plasmid with the native mdh promoter eliminated was prepared by phosphorylation and blunt end ligation. The PCR product was incubated with T4 Polynucleotide Kinase for 30 minutes at 37℃, before heat inactivation of the enzyme for 20 minutes at 65℃ detailed in table 2.7 (53).

Table 2.7: Components and amounts used for the phosphorylation of the linearized pACYC-sfGFP backbone by the enzyme T4 polynucleotide kinase (PNK).

Component	Amount
pACYC-sfGFP backbone (PCR product, DpnI treated)	8.75 µl (142,1 ng/µl)
T4 ligase buffer	1 µl
T4 PNK	0.25 µl

After phosphorylation, the fragment was blunt end ligated by addition of T4 ligase to the reaction displayed in table 2.8. The reaction was incubated for 2 hours at room temperature followed by heat inactivation for 10 minutes at 65 ℃. The reaction was cooled on ice before heat shock transformation into *E. coli*.

Table 2.8: Components and amounts used for blunt-end ligation of the phosphorylated pACYC-sfGFP backbone by the enzyme T4 DNA ligase.

Component	Amount
T4 ligase buffer	2 µl
T4 DNA ligase	1 µl
dH_2O	7 µl

Golden Gate Assembly of pACYC-sfGFP 200N SD

The pACYC-sfGFP 200N SD plasmids with artificial promoters and 5′UTR was cloned with the Golden Gate Assembly method. Before cloning, the backbone PCR product was purified using the QIAquick PCR purification kit. The concentrations of the backbone and 200N SD insert in table 2.9 were measured with the BioNordika NanoDrop (54, 55).

Table 2.9: Concentrations of the PCR products pACYC-sfGFP backbone and 200 N SD insert in nanograms per microliter measured with the BioNordika NanoDrop.

Construct	Concentration
pACYC-sfGFP backbone (PCR purified)	142,7 ng/µl
200N SD promoter insert	15 ng/µl

The one-pot restriction and ligation cloning followed a protocol based on the publication from V. Potapov *et al.* The components were kept on ice during pipetting in a sterile environment to avoid any contamination. In table 2.10, the components and amounts for two 10 µl reactions are displayed (56).

Table 2.10: Components and amounts for two 10 µl Golden Gate reactions for heat-shock transformation in *E. coli*. The components were kept sterile and on ice at all times.

Component	Amount
T4 ligation buffer	2 µl
pACYC-sfGFP Backbone	0.52 µl (75 ng :142.7 ng/µl)
200N SD insert	5 µl (75 ng : 15 ng/µl)
T4 ligase (500 U)	1.25 µl
BsaI HF-v2 (15 U)	0.75 µl
dH_2O	To 20 µl

The reaction was incubated in a thermocycler for approximately 8 hours, until completion of the programme in table 2.11. Directly after the reaction was used for heat shock transformation into *E. coli* with 10 µl of the reaction per aliquot of competent cells. After incubation, the cells were plated on chloramphenicol 25 plates. As a negative control, a transformation with pACYC-sfGFP backbone (0.26 µl) was plated, with the same amount used for one 10 µl transformation.

Table 2.11: PCR thermocycler program for the Golden Gate cloning. The PCR reaction was used directly for heat-shock transformation upon completion of programme.

Temperature	Time	Cycle
37℃	5 minutes	
16℃	5 minutes	45X
65℃	10 minutes	Heat inactivation

2.1.6 Colony PCR

To confirm the expected lengths of PCR products, colony PCR was performed after transformation. The master-mix was aliquoted into PCR tubes and the DNA template was acquired by gently touching a colony with a pipette tip before submerging it into the aliquoted master-mix in table 2.12. The thermocycler settings in table 2.13 were based on the standard protocol for *Taq* polymerase, and the annealing temperature was calculated by adding the primer sequences to the NEB Tm calculator (57, 58).

Table 2.12: Components and amounts for one reaction of *Taq* polymerase PCR for identification of mutants and confirmation of expected product lengths. The primers sequences are found in the appendix.

Component	Amount
10x Standard *Taq* reaction buffer	1 µl
10 mM dNTPs	0.2 µl
10 µM forward primer	0.2 µl
10 µM reverse primer	0.2 µl
Template DNA	Solid colony
Taq DNA polymerase	0.05 µl
dH_2O	8.4 µl

Table 2.13: *Taq* polymerase PCR thermocycler programme for amplification of the target regions. Step 2-4 (denaturation, annealing and extension) was repeated for 15 cycles.

Step	Time	Temperature	Cycle
1. Initial denaturation	10 min	95℃	
2. Denaturation	15s	95℃	
3. Annealing	15s	45-68℃	Repeat step 2-4 15X
4. Extension	1 min/kb	68℃	
5. Final extension	5 min	68℃	
Hold		4℃	

2.1.7 Agarose gel electrophoresis

For confirmation of correct PCR product lengths, the samples were analysed by agarose gel electrophoresis (0,8% agarose). For each sample, 1-2 µl was added to 5 µl of loading dye and 7 µl of a 1 kb DNA ladder. The gel was run at 90 V for 40-55 minutes depending on the size difference of the product and WT. After the gel was run to completion, the result was imaged by the gel doc system Image lab by Bio-Rad. The components used in the agarose gel electrophoresis is displayed in table 2.14.

Table 2.14: Recipes for components used for agarose gel electrophoresis prepared by the lab engineer; Tris-EDTA (TAE) buffer, Agarose solution (0,8 %), Loading dye and 1 kb DNA ladder.

50x Tris-acetate-EDTA (TAE) buffer (1L)	**Comments**
242g Tris	
57.1 ml acetic acid	
100 mL Ethylenediaminetetraacetic acid (EDTA, 0.5M)	dH_2O to 1L
Agarose solution (0,8 %)	
3.2g Agarose (Cambrex)	
400 ml 1x TAE solution	Microwaved for 5 minutes
20 µl GelRed	
Loading dye	
20 µl NEB Gel Loading Dye, purple (6X)	
100 µl dH_2O	
1 kb DNA ladder	
10 µl NEB Gel Loading Dye, purple (6X)	
47 µl dH_2O	

2.2 Materials and methods in *Vibrio natriegens*

Part of the work done for the thesis was establishing the use of *V. natriegens* as an organism in the lab, as it was previously not used by any of the lab groups members. The experiments were performed under close advisement and often in parallel with Ph.D. candidate and co-supervisor Lisa Tietze. The goal was to establish which media and growth conditions that produced competent cells with high transformation efficiency, in addition to identify working plasmids.

All methods described in the *E. coli* section are applicable to this section unless changes are mentioned in the sections below.

2.2.1 Media and growth conditions

Four protocols for preparation of competent cells, the media in table 2.15 and electroporation protocols with electroporation buffers in table 2.16 were tested. Several plasmids and antibiotics of different concentrations were tested.

After preparing competent cells, aliquots were either used directly for transformation or transferred to a -80℃ freezer for storage.

Table 2.15: Components and concentrations for media used for preparing electrocompetent cells with protocol 1-4 and electroporation transformation in *V. natriegens*. The media solutions were adjusted to the correct volume with dH_2O before sterilization.

LBN media (1L)	**Comments**
10g tryptone	
5g yeast extract	
10g NaCl	
Add 1,5% agar for plates	Autoclave
LB3 media (1L)	
10g tryptone	
5g peptone	
30g NaCl	Autoclave
BHIN media (1L)	
37g Brain hearth infusion broth (Oxoid)	
15g NaCl	Autoclave
V2 salts (250 ml)	
2.98g NaCl	
0.078g KCl	
1.18g $MgCl \cdot 6H_2O$	Filter sterilize

Table 2.16: Components and concentrations for electroporation buffers used for preparing electrocompetent cells with protocol 1-4 and electroporation transformation in *V. natriegens.* The media solutions were adjusted to the correct volume with dH_2O before sterilization.

Glycerol 10% (10 ml)	**Comments**
1 ml glycerol (99.95 %, bi-distilled)	
9 ml dH_2O	Filter sterilize
Sorbitol 1M (100 ml)	
18.2g sorbitol	Autoclave
Sucrose buffer (50 ml)	
2.32g sucrose	
0.061g K_2HPO_4	Filter sterilize

The following four protocols were tested for preparation of electrocompetent cells and electroporation. This section is concluded by the protocol which were concluded to be most time efficient and provided a sufficient amount of cells, with the most effective media composition.

1 Modified *P. putida* protocol for preparation of electrocompetent cells and electroporation

LBN media (5 ml) was inoculated with a colony of *V. natriegens* in a 13-ml tube and placed in a shaking incubator at 30℃ (225 rpm). The following day LBN media (10 ml) in a 125 ml Erlenmeyer flask was inoculated with 200 µl of overnight culture and incubation at 30℃ (225 rpm) until an OD_{600} of ~0.4 (1-1.5 hours). After incubation the cells were kept on ice at all times. The culture was divided in Eppendorf tubes (1 ml) and centrifuged for 1 minute (12,000 rpm, 4℃). The supernatant was discarded carefully, and the pellet resuspended in ice cold glycerol (10%). The wash was repeated once at 12,000 rpm, before a last wash at 6,000 rpm. After the final wash, the cells were resuspended in 50 µl glycerol (10%).

Electroporation protocol

An aliquot of cells was transferred to a chilled electroporation cuvette (2 mm) on ice and 1-2 µl (<100 ng) of plasmid DNA was added to the cells and mixed gently. The cuvette was wiped dry and placed into the electroporator. A protocol with 800V, 25 µF and 200 Ω was set for one pulse. Immediately after electroporation, the cells were recovered in LBN media (1 ml) and incubated for 45 minutes at 30℃ (225 rpm). Following incubation, 100 µl of cells were plated on selective agar plates and incubated overnight at 30℃ (59).

2 H, Lee. Protocol for preparation of electrocompetent cells and electroporation

LB3 media (3 ml) was inoculated with a colony of *V. natriegens* in a 13-ml tube and placed in a shaking incubator at 37℃ (225 rpm). The following day 600 µl overnight culture was transferred to an Eppendorf tube and pelleted at 13000 rpm for 1 minute. The supernatant was discarded, and the cells were resuspended in fresh LB3 media (600 µl). 60 ml of LB3 was inoculated with the cells, transferred to an Erlenmeyer flask (250 ml) before incubation at 37℃ (225 rpm) until an OD_{600} of ~0.4 (approximately 1 hour). During the incubation time, sterile electroporation cuvettes and 10 ml of sorbitol (1M) was chilled on ice. During the preparation, the cells were kept on ice at all times.

After incubation the culture was transferred to 50 ml tubes, centrifuged for 5 minutes (7,830 rpm, 4℃). The supernatant was decanted carefully and the pellet resuspended in 1 ml of sorbitol (1M) before transfer to an Eppendorf tube. The cells were pelleted at 7,830 rpm for 1 minute, and the sorbitol wash was repeated 2 additional times. The final pellet was resuspended in 300 µl of sorbitol and divided in 50 µl aliquots.

Electroporation protocol

An aliquot of 50 µl was transferred to a chilled electroporation cuvette (2mm) on ice and plasmid DNA (<100 ng) was added to the cells and gently mixed. The cuvette was wiped dry and placed into the electroporator. A protocol with 400V, 25 µF and 1000 Ω was set for one pulse. Immediately after electroporation, the cells were recovered in LB3 media (1 ml) and incubated for 45 minutes at 37℃ (225 rpm). Following incubation, 100 µl of cells were plated on selective agar plates and incubated overnight at 37℃ (60).

3 Gibson paper protocol for preparation of electrocompetent cells and electroporation

BHIN + V2 media (5 ml) was inoculated with a colony of *V. natriegens* in a 13-ml tube and placed in a shaking incubator at 37℃ (225 rpm). On the following day BHIN media (250 ml) split in two 250-ml Erlenmeyer flasks was inoculated with 2,5 ml of overnight culture and incubated at 37℃ (225 rpm) until an OD_{600} of ~0.5 (approximately 1 hour). The cultures were transferred to two chilled 250-ml centrifuge bottles and kept on ice for 15 minutes. The cells were centrifuged for 20 min (6,500 rpm, 4℃). The supernatant was decanted carefully and the pellets were resuspended in electroporation buffer (5 ml). The cells were transferred to 50-ml tubes and filled with additional electroporation buffer (30 ml). The cells were centrifuged for 15 minutes (6,750 rpm, 4℃) and the supernatant was decanted with a pipette. The wash was repeated two times, and after the final wash the cells were resuspended in residual electroporation buffer and the volume was adjusted until an OD_{600} of 16 was reached.

Electroporation protocol

An aliquot of 50 µl was transferred to a chilled electroporation cuvette (2mm) on ice and plasmid DNA (<100 ng) was added to the cells and gently mixed. The cuvette was wiped dry and placed into the electroporator. A protocol with 800V, 25 µF and 200 Ω was set for one pulse. Immediately after electroporation, the cells were recovered in BHIN + V2 media (500 µl) and incubated for 1.5 hours at 37℃ (225 rpm). Following incubation, 100 µl of cells were plated on selective agar plates and incubated overnight at 37℃ (6).

4 Protocol for preparation of electrocompetent cells and electroporation

BHIN media (5 ml) was inoculated with a colony of *V. natriegens* in a 13-ml tube and placed in a shaking incubator at 37℃ (225 rpm). The following morning two 500 ml Erlenmeyer flasks with BHIN media (125 ml) was inoculated with the overnight culture (2,5 ml) and incubated until an OD_{600} of ~0.5 at 37℃ (225 rpm, ~1 hour). After incubation the culture was transferred to centrifuge bottles (250 ml) and centrifuged for 15 minutes (4,500 g, 4℃). The supernatant was decanted and the pellet resuspended in 20 ml ice cold sorbitol (1M). The cells were transferred to chilled 50 ml tubes and kept on ice. The wash was repeated twice and after the final wash the pellets were resuspended in 250 µl and aliquoted in 50 µl into Eppendorf tubes.

Electroporation protocol

An aliquot of 50 µl was transferred to a chilled electroporation cuvette (2mm) on ice and plasmid DNA (<500 ng) was added to the cells and gently mixed. The cuvette was wiped dry and placed into the electroporator. A protocol with 800V, 25 µF and 200 Ω was set for one pulse. Immediately after electroporation, the cells were recovered in BHIN media (5 ml) and incubated for 45 minutes at 37℃ (225 rpm). Following incubation, 100 µl of cells were plated on selective agar plates and incubated overnight at 37℃ (9).

2.2.2 Electroporation protocol in *V. natriegens*

When preparing competent *V. natriegens* cells, a modified version of protocol number 1 was used for electro-competency and electroporation. The media used was BHIN+V2 salts, incubation temperatures were 37℃ and the transformed cells were plated on LBN agar plates with chloramphenicol 12.5 µg/ml.

2.2.3 Colony PCR

To confirm the expected lengths of PCR products in *V. natriegens*, colony PCR was performed after transformation. The PCR was performed as described in section 2.1.5, with an additional step of freeze-thawing at -20℃ once for ~20 minutes before thermocycling to achieve visible bands after agarose gel electrophoresis.

2.2.4 Plasmids tested in *V. natriegens*

For *V. natriegens*, the plasmids in table 2.17 were tested with the electroporation protocols in section 2.2.1. All origins of replication were previously reported to result in successful transformants of *V. natriegens.*

Table 2.17: Plasmids tested in *V. natriegens* with the different origin of replication (OriT) of p15A, RK2, pmB1, pBL1 and coIE1 and resistance genes for gentamycin, chloramphenicol, kanamycin or ampicillin tested at various concentrations.

Plasmid	OriT	Size	Resistance gene	Antibiotic concentrations tested (µg/ml)
PJQ200	p15A	5555 bp	Gentamycin	Gen30, Gen35, Gen40, Gen50
pACYC184	p15A	4245 bp	Chloramphenicol	Chl5, Chl12.5, Chl15
pVB1-kan	RK2	6719 bp	Kanamycin	Kan100, Kan200, Kan300, Kan500
pVB1	RK2	6764 bp	Ampicillin	Amp50, Amp100
yGRA_pTpA	pmB1	8218 bp	Ampicillin	Amp50, Amp100
PXMJ19	pBL1	6601 bp	Chloramphenicol	Chl5
PHT10	coIE1	8651 bp	Ampicillin	Amp50, Amp100

2.2.5 Image acquisition by confocal microscopy

To visualise the expression of sfGFP in populations of *V. natriegens* cells, image acquisition by confocal microscopy was performed by PhD Swapnil Vilas Bhujbal, a postdoctoral researcher at the Department of Biotechnology, NTNU.

For image acquisition, an inverted microscope (Axio Observer from Zeiss, 2.3.64.0) with an air objective (20x, NA 0.8). A GFP filter was used to inspect the viability and sfGFP expression of 2 μl of immobilized cells. The images were later processed with the Zeiss image analysis software (2.3.64.0). The samples prepared for imaging included a population with high-expressing cells identified during the fluorescence screening, a moderately expressing population, a population transformed with the positive control plasmid, pACYC-sfGFP and a population transformed with the negative control plasmid pACYC.

2.2.6 Experimental setup screening for successful mutants

After successful transformation of the pACYC-sfGFP 200N SD library into *V. natriegens*, the transformants were screened for GFP expression. Colonies were picked in a flow cabinet with sterile toothpicks into Costar® flat bottom transparent 96-well plates. The wells were filled with 200 µl BHIN+V2 media and chloramphenicol 5 µg/ml. Controls were inoculated from glycerol stocks of plasmids transformed into *V. natriegens*. The pACYC-sfGFP plasmid was used as the positive control and the pACYC184 plasmid was used as the negative control.

The plates were placed in a shaking incubator at 37℃ and 725 rpm for 2-2.5 hours until visible growth. The OD_{600} and absorbance was measured utilizing a Tekan infinite M200 Pro plate reader. Before measurement, the colonies were transferred by a 96-well array replica plater to agar plates. The absorbance spectra for sfGFP was set to 485 (excitation wavelength) to 520 (emission wavelength) with gain set at 50.

After measurement, replica plated agar plates were incubated overnight at 37℃. The following day, colonies expressing from mid-low-high was picked with a pipette tip and inoculated in a 13-ml tube with BHIN+V2 media (4 ml) and chloramphenicol (12.5 µg/ml). The morning after, 500 µl of culture from each inoculation was added to 500 µl of glycerol (40%) in a cryogenic vial and stored at -80℃. The remaining culture was MP to isolate the plasmid for transformation into *E. coli*.

	1	2	3	4	5	6	7	8	9	10	11	12
A	Blank	Pos.	Neg.									
B												
C												
D												
E												
F												
G												
H												

Figure 2.3: Experimental setup of a Costar ® 96-well plate screening for GFP expressing transformants in *V. natriegens*. The plate was filled with 200 µl BHIN+V2 media and chloramphenicol 5 µg/ml in each well. The pACYC-sfGFP plasmid as positive control and the pACYC184 as the negative control. Absorbance was read at OD_{600} and fluorescence at 485-520 nm.

2.3 Time dependent screening of promoters in *E. coli* and *V. natriegens*

The aim of time dependent screening of promoters was to observe at which growth phase promoter activity began in the two organisms, in particular if expression was detectable during log-phase or later.

After pre-screening for promoters expressing in a range from mid-low-high in *V. natriegens* as described in section 2.2.6, the isolated plasmids were transformed into *E. coli* by heat-shock transformation. The following day, 13-ml tubes with LB media (4 ml) and chloramphenicol (25 µg/ml) was inoculated with one colony from each transformation. The morning after, glycerol stocks were prepared as in section 2.2.6 and five plasmids were isolated for sequencing. The sequencing results was used to confirm that the correct plasmids were transformed back into *E. coli*.

In order to characterize and compare the GFP expression in *V. natriegens* and *E. coli*, the GFP expression over a period of ~6 hours were performed as described in section 2.2.6. Separate 96-well plates with triplicates from the glycerol stocks of mid-low-high expression in *V. natriegens* and the corresponding clones in *E. coli* were incubated at 37℃ and 725 rpm. 96-well plates with *E. coli* clones were inoculated and grown overnight. The following morning, 10 μl of the overnight culture from each well was used to inoculate 190 μl of media on a fresh 96-well plate. After two hours of growth, measurements of time dependent expression could be started.

The plates with *V. natriegens* were measured every 20-minutes starting from 20 minutes before visible growth was expected, while the time intervals were 30 minutes for *E. coli.* Measurements were continued until the GFP expression declined.

2.4 Sequence analysis

The clones picked for characterization of sfGFP expression in *V. natriegens* and *E. coli* were sent for sequencing at Eurofins Genomics. The samples with plasmids isolated by miniprep from *V. natriegens* was prepared as presented in table 2.18. Six random plasmids were isolated from *E. coli* to confirm transformation of the correct plasmids.

Table 2.18: Components and amounts of samples prepared for sequencing.

Component	**Amount**
Plasmid	>400 ng
Primer (ALY533 rv)	2,5 µl (10 µM)
dH_2O	To 10 µl total volume

When the sequence results arrived, the sequences were first aligned with the plasmid pACYC-sfGFP 200N SD in Benchling to confirm promoter and 5′UTR sequences with complete, random 200N SD inserts and the following ATG start codon in front of the sfGFP reporter gene. The promoter and 5 ′UTR insert sequences were then analysed with the online tools BPROM and Improbizer. Output from the analysis is in the results sections 3.5 (table 3.4) and 3.6 (table 3.5-3.9). The goal of the sequence analysis was to identify matches to expected consensus sequences at -35 and -10 elements and used to characterize the promoter activity together with the measured GFP expression levels.

The BRPOM tool located promoters within the insert sequences and presented scores from 0-100 for how close to consensus located possible -35 and -10 elements was. The Improbizer tool was utilized to located motifs in the sequence possible not detected by BPROM. Improbizer used the input sequences to generate profiles, or averages of sub-sequences identified as possible motifs. These profiles were determined by the algorithm as appearing more often in the sequences than expected by chance. For the Improbizer search, the number of motifs was set to two, and the initial motifs size as eight (61, 62).

3. Results

3.1 Establishing how to work with *V. natriegens* in the laboratory environment

Final protocol, growth conditions and antibiotic concentrations

3.1.1 Establishing use of the organism *Vibrio natriegens*

V. natriegens is not a model organism for molecular biology, however multiple publications have established use of genetic elements in the bacterium. Several plasmids have been reported to function in *V. natriegens*. When establishing which plasmid that would be used for Golden Gate cloning our experience was that only one plasmid was compatible in *V. natriegens.* As use of the organism was not previously established in the laboratory group, I tested several protocols, strains and plasmids for preparation of competent cells with sufficient transformation efficiency. Since the work done for this thesis included thorough testing of protocols, media combinations and plasmids to establish work with *V. natriegens* as a new host for molecular biology for the laboratory group PhotoSynLab, a brief section summarizing this is included in the results.

3.1.2 Plasmids

To perform cloning of the promoter libraries, the first hindrance was the ability of *V. natriegens* to maintain and replicate plasmids. A combination of several origins of replications, copy numbers, antibiotic resistances and concentrations was tested. To date we found a single plasmid, pACYC184 to be transformable with sufficient efficiency and selection consistency. Table 3.1 summarizes the results of transformation attempts of multiple different plasmids in *V. natriegens*. Several attempts were made with different batches of competent cells, and the results are the conclusion of transformations where the control plasmid pACYC184 was successfully introduced.

3.1.3 Results of protocol and media composition testing

A modified P. putida protocol for preparation of electrocompetent cells and electroporation with *V. natriegens*

BHIN media + V2 salts (5 ml) was inoculated with a colony of *V. natriegens* in a 13-ml tube and placed in a shaking incubator at 37℃ (225 rpm). The following day BHIN media + V2 salts (10 ml) in a 125 ml Erlenmeyer flask was inoculated with 200 µl of overnight culture and incubated at 37℃ (225 rpm) until an OD_{600} of ~0.4. Control of OD_{600} often due to rapid growth rate, perform first measurement after 45 minutes. After incubation the cells were kept on ice at all times. The culture was divided in chilled Eppendorf tubes (1 ml) and centrifuged for 1 minute (12,000 rpm, 4℃). The supernatant was discarded carefully, and the pellet resuspended in a sucrose electroporation buffer. The wash was repeated once at 12,000 rpm, before a last wash at 6,000 rpm. After the final wash, the cells were resuspended in electroporation buffer (50 µl).

Electroporation protocol

An aliquot of cells was transferred to a chilled electroporation cuvette (2 mm) on ice and 1-2 µl (<100 ng) of plasmid DNA was added to the cells and mixed gently. The cuvette was wiped dry and placed into the electroporator. A protocol with 700V, 25 µF and 200 Ω was set for one pulse. Immediately after electroporation, the cells were recovered in BHIN media + V2 salts (1 ml) and incubated for 45 minutes at 37℃ (225 rpm). Following incubation, 100 µl of cells were plated on chloramphenicol *5* µg/ml LBN agar plates and incubated overnight at 37℃.

All media recipes are in the material and methods chapter.

Table 3.1: The plasmids with different OriT, size, resistance gene and antibiotic concentrations tested (µg/ml) for transformation of electrocompetent *V. natriegens* cells. The results of the transformations are described in the last column, indicating successful introduction of the plasmid pACYC184. Attempts were repeated multiple times. *Chloramphenicol 5 µg/ml was chosen as the working concentration for agar plates while chloramphenicol 12,5 µg/ml was chosen as the concentration for liquid media.

Plasmid	OriT	Size	Resistance	Antibiotic (µg/ml, final concentration)	Result of transformation
PJQ200	p15A	5555 bp	Gentamycin	Gen 30/35/40/50	No growth
pACYC184	p15A	4245 bp	Chloramphenicol	Chl 5*/12.5*/15	Consistent successful transformants (used as control plasmid)
pVB1-kan	RK2	6719 bp	Kanamycin	Kan 100/200/300/500	Inconsistent/ background growth on antibiotic plates
pVB1	RK2	6764 bp	Ampicillin	Amp 50/100	No growth
yGRA_pTpA	pmB1	8218 bp	Ampicillin	Amp 50/100	No growth
PXMJ19	pBL1	6601 bp	Chloramphenicol	Chl 5	No growth
PHT10	coIE1	8651 bp	Ampicillin	Amp 50/100	No growth

3.2 Golden Gate Cloning

The plasmid pACYC-sfGFP 200N SD was assembled by Golden Gate cloning and transformed into *E. coli*. Plasmid libraries were isolated from the transformants in order to establish the plasmid libraries in *V. natriegens*, which was screened for functional promoter and 5′UTR's. The plasmid pACYC-sfGFP was used as the backbone for the plasmid containing the 200N SD promoter and 5′UTR. The reporter gene superfold-GFP was included for characterization and comparison of the promoter strength by GFP expression in *E. coli* and *V. natriegens*.

Before the assembly by Golden Gate cloning, two separate PRC reactions produced the backbone and insert. The backbone was amplified with Takara CloneAmp^{TM} HiFi PCR, and the insert was amplified with Q5 High-Fidelity DNA polymerase, as detailed in the materials and methods section 2.1.4. The backbone PCR resulted in a concentration of 142,5 ng/µl, and the 200N SD promoter and 5′UTR insert PCR resulted in a concentration of 15 ng/µl. To confirm that the PCR products were of correct lengths, agarose gel electrophoresis was run with the WT pACYC-sfGFP plasmid as a control to confirm that the native promoter was removed. Successful removal of the promoter would result in visible differences on the gel when comparing the WT plasmid to the backbone. Confirmation of expected product lengths of the backbone and insert are displayed in figures 3.1 and 3.2. Due to poor visibility of the control plasmid in figure 3.2, the gel was run twice. Due to the supercoiled nature of the control plasmid pACYC-sfGFP, the band ran lower than expected.

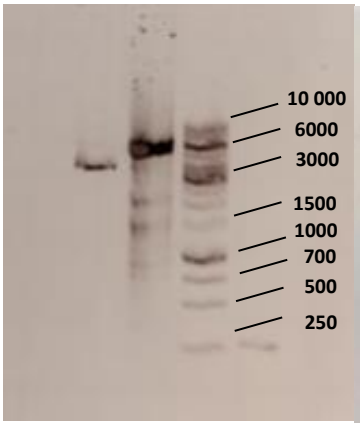


Figure 3.1: Gel electrophoresis of PCR for pACYC-sfGFP backbone and 200N SD promoter and 5′UTR insert in *E. coli*. The control (supercoiled) plasmid pACYC-sfGFP with an expected length of 6269 bp is in lane 1, and the pACYC-sfGFP backbone with an expected length of ~4600 bp is in lane 2, a 1Kb ladder is in lane 3, and the insert with the expected length of ~250 bp in lane 4.

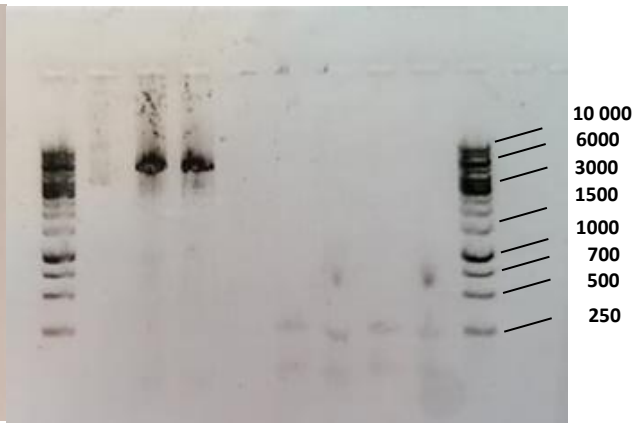


Figure 3.2: Gel electrophoresis of PCR for pACYC-sfGFP backbone and 200N SD promoter and 5′UTR insert in *E. coli*. The control plasmid, pACYC-sfGFP is not visible in lane 2. The pACYC-sfGFP backbone with an expected length of ~4600 bp is in lane 3 and 4, the inserts with the expected length of ~250 bp in lane 5-8. A 1Kb ladder is in lane 1and 9.

The cloning reactions were scaled up for use in experiments for a planned publication. The Golden Gate Assembly reactions transformed into *E. coli* resulted in 10,000 clones per plate. The total amount of clones produced for ten plates were approximately 10,000 clones. Colonies were visible under UV light, emitting faint green fluorescence. To confirm that the 200N SD promoter and 5′UTR inserts had been successfully ligated into the plasmid, Q5 PCR samples of ten clones was prepared for sequencing and gel electrophoresis for confirmation of expected lengths were performed. In **figure 3.3**, a difference of the promoter region length of the WT plasmid containing the native mdh promoter and the pACYC-sfGFP 200N SD plasmid containing the 200N SD promoter and 5′UTR is visible. To minimize the risk of contamination by the WT plasmid pACYC-sfGFP with the native mdh promoter the backbone PCR was treated with DpnI. For this transformation, the number of colonies resulting from the transformation with the backbone was 221 clones, which is sufficiently low. This indicates that the DpnI treatment was almost complete and successful.

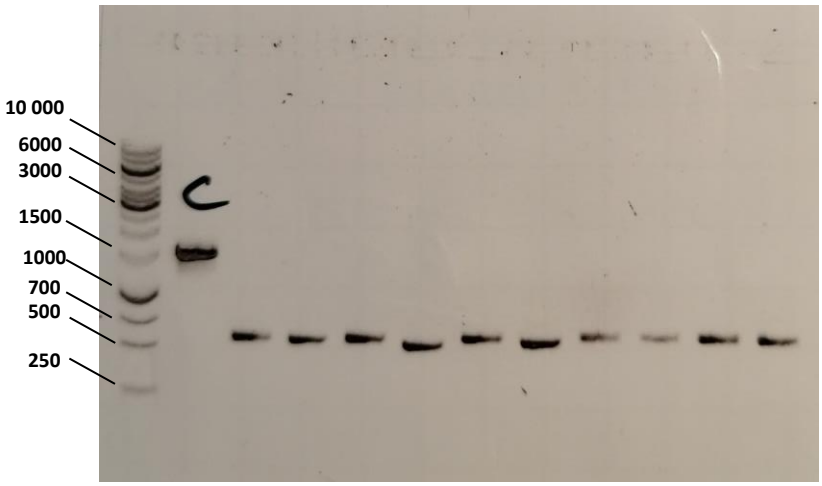


Figure 3.3: Gel electrophoresis of a colony PCR to confirm ligation of 200N SD promoter and 5′UTR insert into the plasmid pACYC-sfGFP by Golden Gate Assembly. A 1KB ladder is in the first lane, the WT plasmid pACYC-sfGFP is in the second lane with expected length of 1424 bp. In the following lanes 3-12 the promoter sequence area in ten clones with the pACYC-sfGFP 200N SD plasmid libraries are displayed, with an expected length of 549 bp. The primers used for the pACYC-sfGFP 200N SD clones in *E. coli* were ALY532 seq forward and ALY533 seq reverse. The colony PCR was performed with Q5, and the fragments were sent for sequencing to confirm correct ligation of the 200N SD promoter and 5′UTR fragments sequences.

The creation of the pACYC-sfGFP 200N SD plasmid with a GFP reporter gene and chloramphenicol was done by PCR amplification of the backbone and promoter and 5′UTR inserts, before Golden Gate Assembly of the constructs. Before the Golden Gate Assembly, the expected lengths shown in **figure 3.1 – 3.2** in the results section indicate successful amplification of both constructs. After correct fragment lengths were indicated by agarose gel electrophoresis, the backbone and insert were assembled by Golden Gate cloning. Primers had to be designed to amplify the correct part of the plasmid, for confirmation of correct assembly and positioning of the promoter and 5′UTR. Primers were designed *in silico* with the software Benchling, as well construction of the plasmid (pACYC-sfGFP 200N SD), shown in **figure 3.4**. The fragments displayed in **figure 3.3** with the expected product length of 549 bp amplified stretches of the backbone upstream of the promoter and 5′UTR, and the sfGFP gene downstream of the promoter and 5′UTR. Sequencing was necessary to confirm correct orientation of the promoter and 5′UTR in the backbone, which was detailed in tables 3.4 and 3.9 of the results. Here the backbone fragments upstream and downstream of the promoter and 5′UTR were excluded.

A plasmid map of the final product, the pACYC-sfGFP 200N SD plasmid is displayed in **figure 3.4**. The plasmid map was generated in the online software tool Benchling.

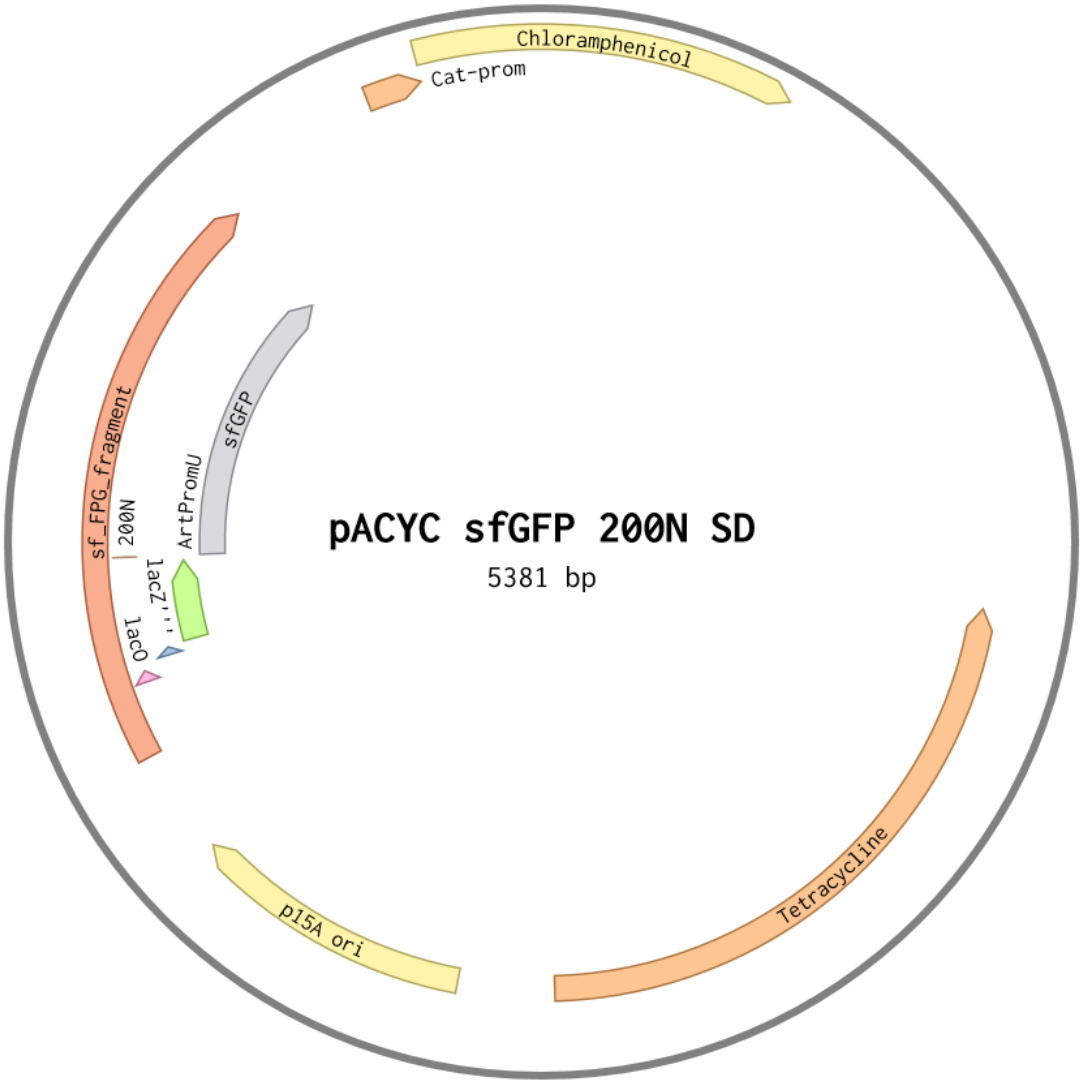


Figure 3.4: Plasmid map of the pACYC-sfGFP 200N SD plasmid of 5381 bp. The 200N SD promoter and 5′UTR insert is indicated as neon-green fragment placed directly in front of the sfGFP reporter gene, a grey labelled fragment. The origin of replication is p15A, and the plasmid has a chloramphenicol resistance gene under control of the cat promoter.

3.2.2 Creation of a negative control plasmid for GFP-expression

In order to confirm that any measured fluorescence originated from the 200N SD promoter and 5′UTR insert driving expression of the sfGFP gene and not from the backbone with the sfGFP gene, a control plasmid was created. The control plasmid did not contain a promoter and 5′UTR and was created with phosphorylation and blunt-end ligation. The plasmid was transformed into *E. coli* and *V. natriegens* and imaged by confocal laser scanning microscopy in *V. natriegens*. To confirm complete removal of the native mdh promoter, Q5 PCR and gel electrophoresis of the promoter region were performed with the primers ALY532 seq forward and ALY533 seq reverse. In **figure 3.4** three samples of the promoter region in the pACYC-sfGFP plasmid backbone are displayed with the expected lengths of 371 bp. The samples were sent for sequencing to confirm complete removal of the native promoter region.

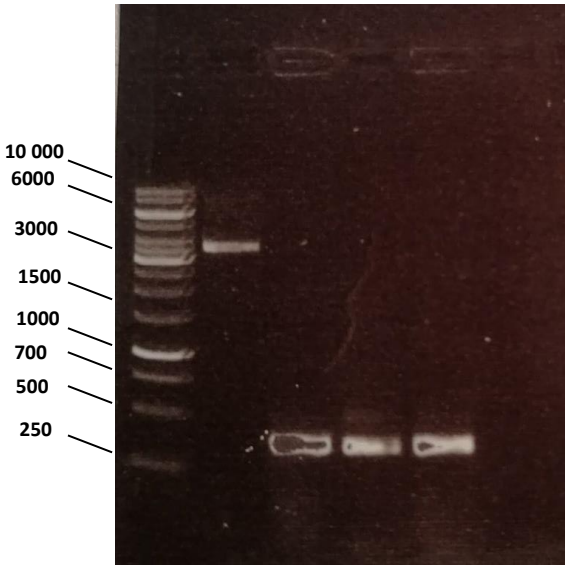


Figure 3.5: Gel electrophoresis of Q5 PCR reactions to confirm removal of the native promoter region in a closed pACYC-sfGFP backbone plasmid. Gel electrophoresis was run to confirm successful amplification. A 1Kb ladder is in lane one, a control plasmid for the PCR run, pACYC184 with expected product length of 4245 bp is in lane two and in lane 3-5 three samples of the promoter region in the pACYC-sfGFP plasmid backbone are displayed with expected lengths of 371 bp.

3.2.3 The pACYC-sfGFP backbone plasmid

In **figure 3.6** the plasmid map of the negative control pACYC-sfGFP backbone plasmid is displayed. The backbone plasmid was created without a promoter and 5′UTR to ensure that the backbone with the sfGFP gene did not result in any background fluorescence. The plasmid was created as described in the materials and methods section 2.1.4 and clones were screened for fluorescence compared with the positive control plasmid pACYC-sfGFP and the negative control plasmid pACYC184. The closed circular pACYC-sfGFP backbone did not emit any fluorescence.

The pACYC-sfGFP backbone plasmid

Figure 3.6: Plasmid map of the pACYC-sfGFP backbone plasmid of 5216 bp. The sfGFP reporter gene is the grey labelled fragment. The origin of replication is p15A, and the plasmid has a chloramphenicol resistance gene under control of the cat promoter.

3.3 Measurement of GFP expression in *V. natriegens* and *E. coli*

After a successful transformation of the pACYC-sfGFP 200N SD and 5′UTR random promoter library in *V. natriegens*, colonies were screened for GFP expression in a 96-well plate setup, detailed in the material and methods section 2.2.6. A small-scale pre-screening showed that about 10% of the unique *V. natriegens* clones contained functional promoter and 5′UTR sequences. However, when the setup was scaled up, there was a complete lack of expression. After several attempts and troubleshooting of the experimental setup, we discovered that fluorescence in *V. natriegens* had to be measured the day after transformation, otherwise the GFP expression diminished.

The screening for *V. natriegens* clones emitting GFP fluorescence was performed to identify functional 200N SD promoter and 5′UTR sequences in front of the sfGFP gene. The later screen resulted in multiple plasmids with 200N SD promoter and 5′UTR sequences, which was re-screened in triplicates to confirm fluorescence. Several plasmids picked from the screening were confirmed to emit fluorescence consistently in the second screening. A total of 36 plasmids were chosen for further screening, which emitted fluorescence in a low to medium to high level. This was important for the comparison of expression levels from the identical 36 plasmids in *E. coli* and illustrates the ability to pick the desired protein production level.

V. natriegens colonies from clones with functional 200N SD promoter and 5′UTR sequences did not appear green either when growing on selective agar, or under UV light. This was an indication of lower GFP expression levels compared to in *E. coli*, which appeared as green both as colonies and under UV light. This was confirmed by the GFP screening in section 3.3.1 below. The plasmids were confirmed to contain unique 200N SD promoter and 5′UTR sequences by sequencing. After the sequencing results arrived, the 36 isolated plasmids were transformed into *E. coli* for characterization and comparison of expression levels in the two organisms. When all plasmids had been transformed into *E. coli*, a time-dependent screening of *V. natriegens* and *E. coli* in triplicates was performed. Time-dependent screening was performed to investigate at which growth phase promoter activity and subsequently GFP expression began, and to characterize the development over time. The results from time-dependent screening are detailed in the following section.

3.3.1 Time-dependent expression in *V. natriegens* and *E. coli*

The goal of performing time-dependent measurement of GFP expression was to investigate at which growth phase promoter activity and subsequently GFP expression began, and to characterize the development over time. From earlier measurements, GFP expression was expected to begin after approximately 2-3 hours in *V. natriegens*, and approximately 4-5 hours in *E. coli* due to the different doubling times of the organisms. After earlier observations during measurements beyond 3 hours during the pre-screening, the different promoter and 5′UTR were expected to become active at different time points. The earliest activity was expected to occur during the exponential (logarithmic) growth phase, while other promoter and 5′UTR would begin expressing near or in the stationary growth phase. In the next section, **figures 3.13 – 3.16** show comparisons of GFP expression from identical plasmids in *V. natriegens* and *E. coli* at the final measurement, after 350 and 250 minutes.

The results of the measurements are displayed in figures **3.7-3.9** for *E. coli*, and figures **3.10-3.12** in *V. natriegens*. To improve the visibility of promoter activity as a function of GFP expression levels, the graphs are divided into low, medium and high GFP expression levels or fluorescence intensity. The measurements from the negative control (pACYC184 in *E. coli* or *V. natriegens*) are included in each graph. In figure 3.9 plasmids with low levels of GFP expression in *E. coli* are displayed. The GFP expression levels at the final measurement after 250 minutes varies for the plasmids in figure 3.9 with low GFP expression levels from 114 (EC1. 1G) to 330 (EC4. 4B). In figure 3.10 plasmids with medium levels of GFP expression levels are displayed, from 325 (EC4. 4H) to 679 (EC3. 6A). In figure 3.11 plasmids with high levels of GFP expression are displayed, from 782 (EC2. 10H) to 1791 (3. 6B).

For *E. coli*, all plasmids expressed above the negative control at the first measurement, when the OD_{600} is on average ~0.380 in the early exponential phase. An increase in GFP expression can be first observed from the plasmids EC3. 5D and EC4. 4B after 25 minutes when the OD_{600} is on average ~0.460. After 75 minutes and an OD_{600} average of ~0.550, the majority of the curves indicate an increase of promoter activity. After this time-point, a steady increase is evident for most plasmids indicated by steep curves with a logarithmic pattern until around 200 minutes when the OD_{600} have reached and average of ~0.800. After 20 minutes, the curves visibly flatten out indicating that a plateau will be reached. There are some potential outliers from the general trends. In figure 3.9 the plasmids EC3. 4B, EC1. 1G, EC3. 5D and EC1. 6B hits a plateau already around 175 minutes.

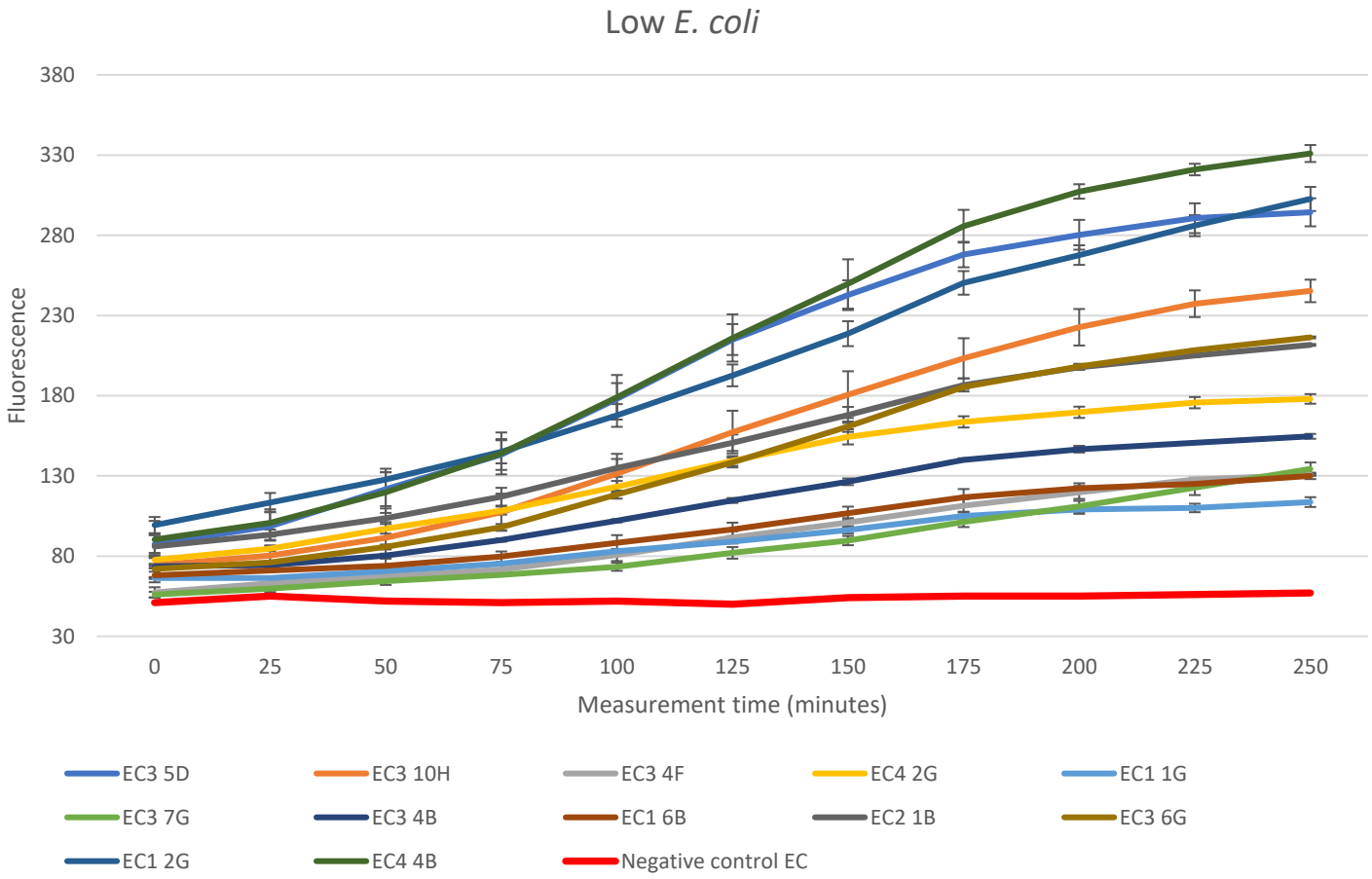


Figure 3.7: Fluorescence measurement results from the time dependent screening of GFP expression in *E. coli* resulting in low fluorescence levels. The measurements were continued for approximately 5 hours in 25-minute intervals. The plasmids were transformed into *E. coli* after isolation from *V. natriegens*. The measurements were performed in triplicates of clones and the graphs represent the calculated averages from each triplicate with standard deviations. The plasmid identities are named after the promoter and 5′UTR sequences identified by the screening of the pACYC-sfGFP 200N SD plasmid libraries in *V. natriegens*. The negative control plasmid (pACYC184) endpoint measurement was 57.

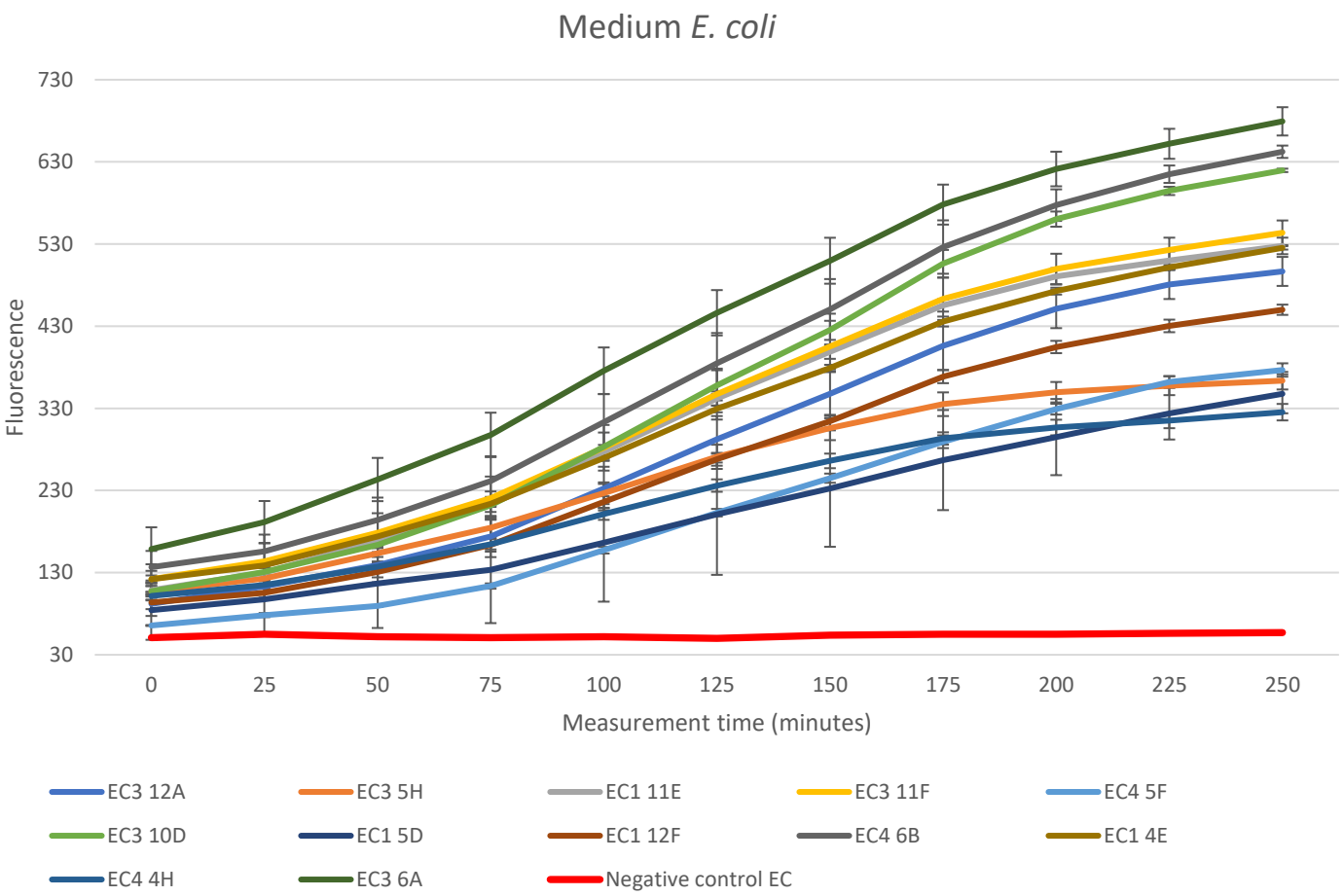


Figure 3.8: Fluorescence measurement results from the time dependent screening of GFP expression in *E. coli* resulting in medium fluorescence levels. The measurements were continued for approximately 5 hours in 25-minute intervals. The plasmids were transformed into *E. coli* after isolation from *V. natriegens*. The measurements were performed in triplicates of clones and the graphs represent the calculated averages from each triplicate with standard deviations. The plasmid identities are named after the promoter and 5′UTR sequences identified by the screening of the pACYC-sfGFP 200N SD plasmid libraries in *V. natriegens*. The negative control plasmid (pACYC184) endpoint measurement was 57.

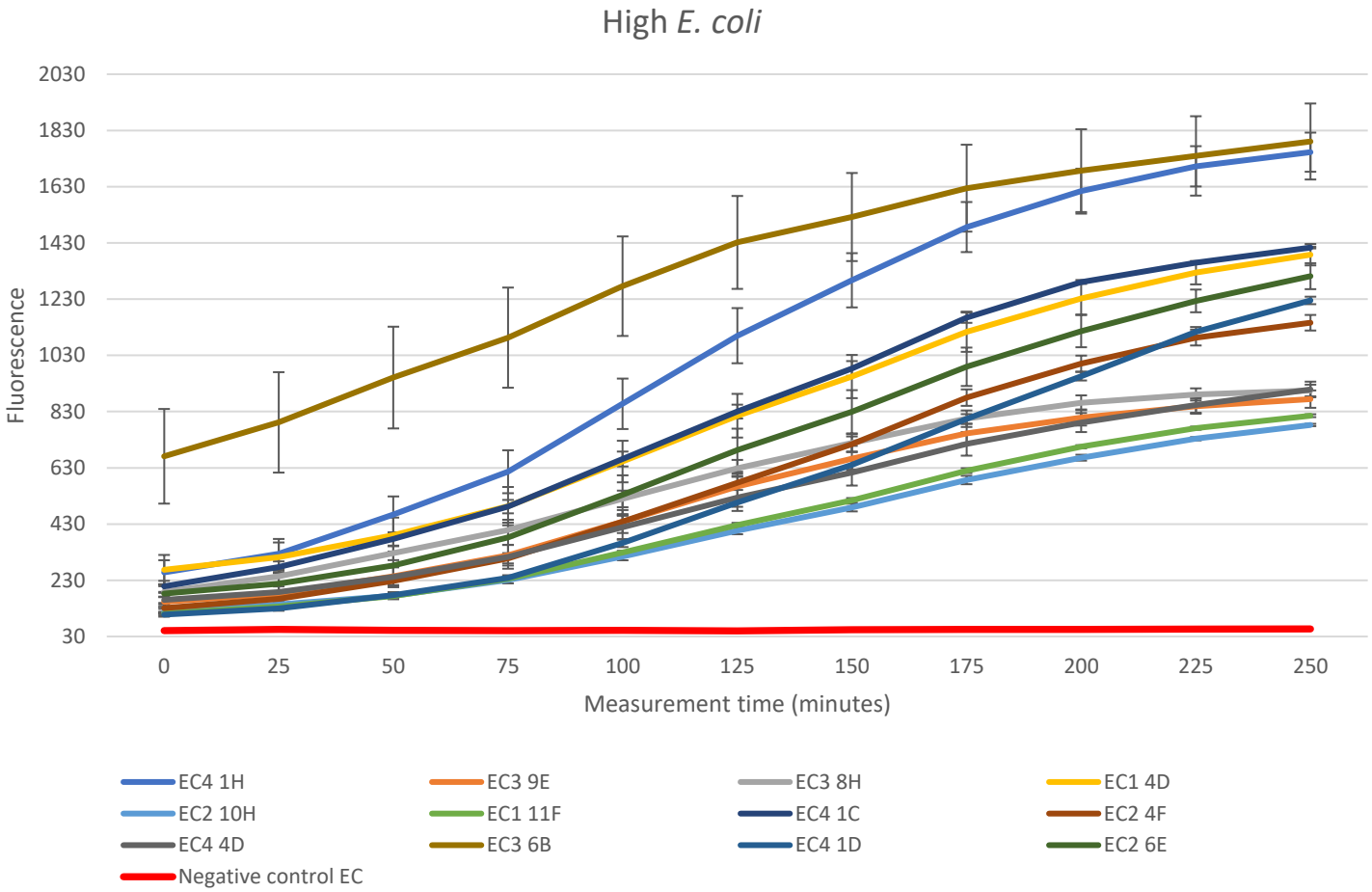


Figure 3.9: Fluorescence measurement results from the time dependent screening of GFP expression in *E. coli* resulting in high fluorescence levels. The measurements were continued for approximately 5 hours in 25-minute intervals. The plasmids were transformed into *E. coli* after isolation from *V. natriegens*. The measurements were performed in triplicates of clones and the graphs represent the calculated averages from each triplicate with standard deviations. The plasmid identities are named after the promoter and 5′UTR sequences identified by the screening of the pACYC-sfGFP 200N SD plasmid libraries in *V. natriegens*. The negative control plasmid (pACYC184) endpoint measurement was 57.

The measurements of GFP expression in *V. natriegens* continued at approximately 20-minute intervals for 350 minutes. Measurements were started 2,5 hours after incubation. As shown in figure 3.10, the majority of promoter activity and expression above background begun after 75 minutes, when the OD_{600} was above 0.5500. Earlier measurements showed a clear indication of GFP expression lacking until similar OD_{600} levels were reached. In **figures 3.10– 3.12** the curves produced by the time-dependent screening in *V. natriegens* at first sight potentially looks to have more variation in promoter activity. However, this is most likely an artefact resulting from the overall lower GFP expression levels which highlights small differences in expression between measurements. In figure 3.10 the plasmids with low GFP expression levels are displayed, from 65 (VN1. 1G) to 73 (VN4. 2G and VN3. 9E) at the final timepoint after 350 minutes of measurement. In figure 3.11 are the plasmids with medium GFP expression levels from 75 (VN2. 1B) to 88 (VN3. 10D) displayed, and in figure 3.12 are the plasmids with high GFP expression from 88 (VN4. 1D) to 130 (VN4. 1H).
For *V. natriegens*, all plasmids except VN4. 4H (figure 3.10) expressed above the negative control after ~65 minutes of measurements, when the OD_{600} ~0.500. The GFP expression seem to increase from most plasmids until 100 minutes for most plasmids at an OD_{500} of ~0.600. The curves in figure 3.10 and 3.11 cannot be described as having a logarithmic shape between those timepoints, but the shape is clearly steeper indicating an increase in fluorescence.

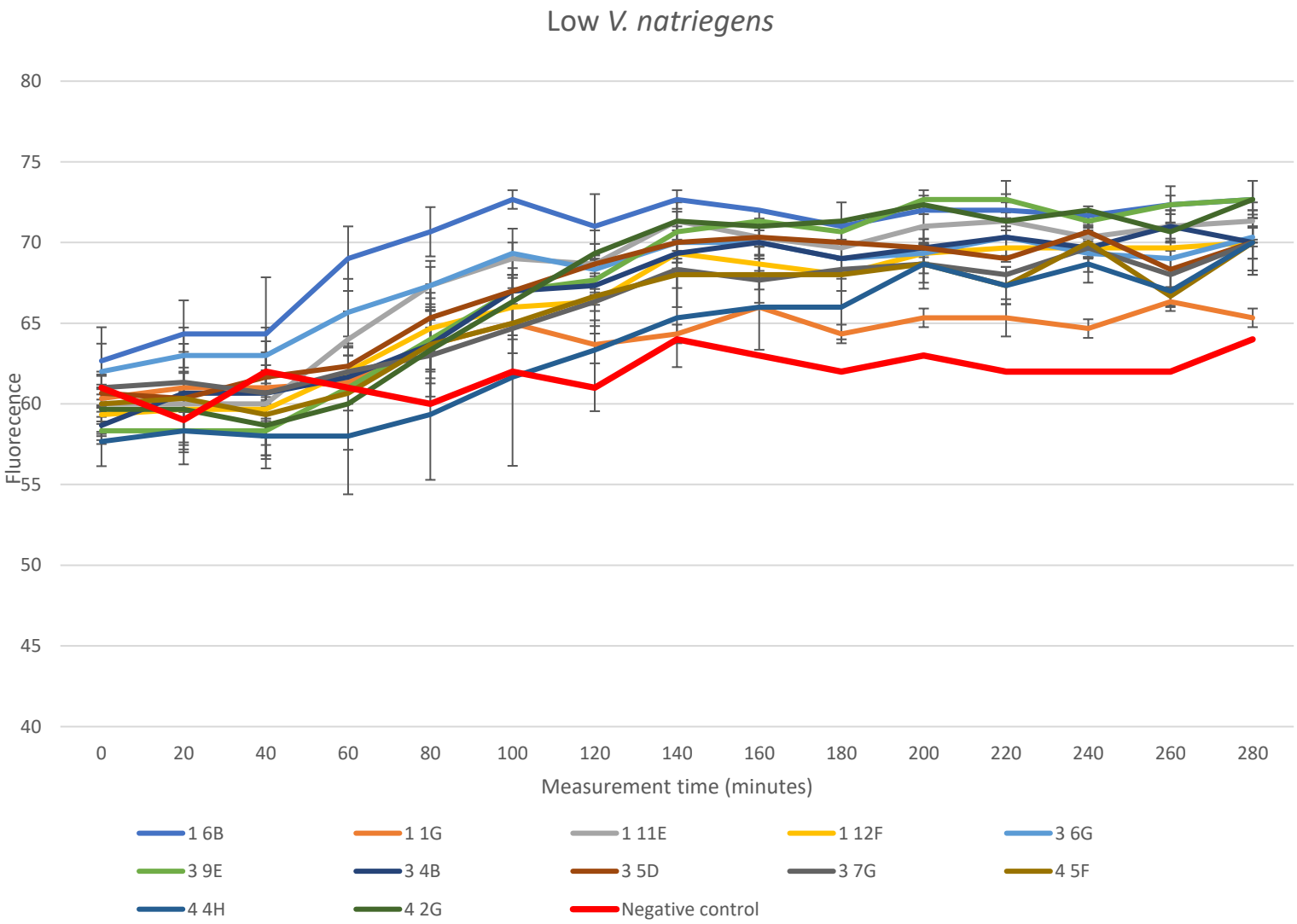


Figure 3.10: Fluorescence measurement results from the time dependent screening of GFP expression in *V. natriegens* resulting in low fluorescence levels. The measurements were continued for approximately 5 hours in 20-minute intervals. The measurements were performed in triplicates of clones and the graphs represent the calculated averages from each triplicate with standard deviations. The plasmid identities are named after the promoter and 5′UTR sequences identified by the screening of the pACYC-sfGFP 200N SD plasmid libraries in *V. natriegens*. The negative control plasmid (pACYC184) endpoint measurement was 64.

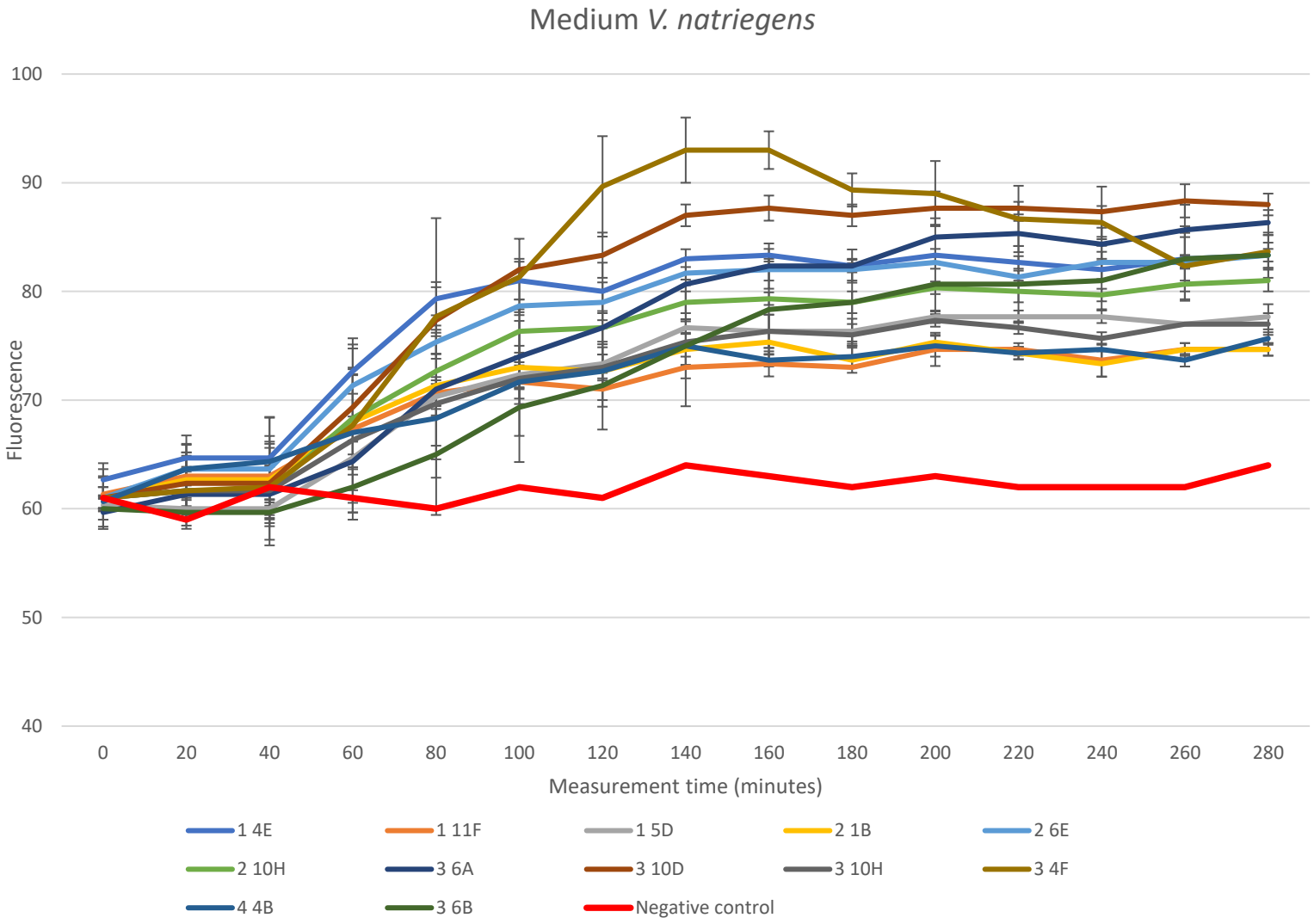


Figure 3.11: Fluorescence measurement results from the time dependent screening of GFP expression in *V. natriegens* resulting in medium fluorescence levels. The measurements were continued for approximately 5 hours in 20-minute intervals. The measurements were performed in triplicates of clones and the graphs represent the calculated averages from each triplicate with standard deviations. The plasmid identities are named after the promoter and 5′UTR sequences identified by the screening of the pACYC-sfGFP 200N SD plasmid libraries in *V. natriegens*. The negative control plasmid (pACYC184) endpoint measurement was 64.

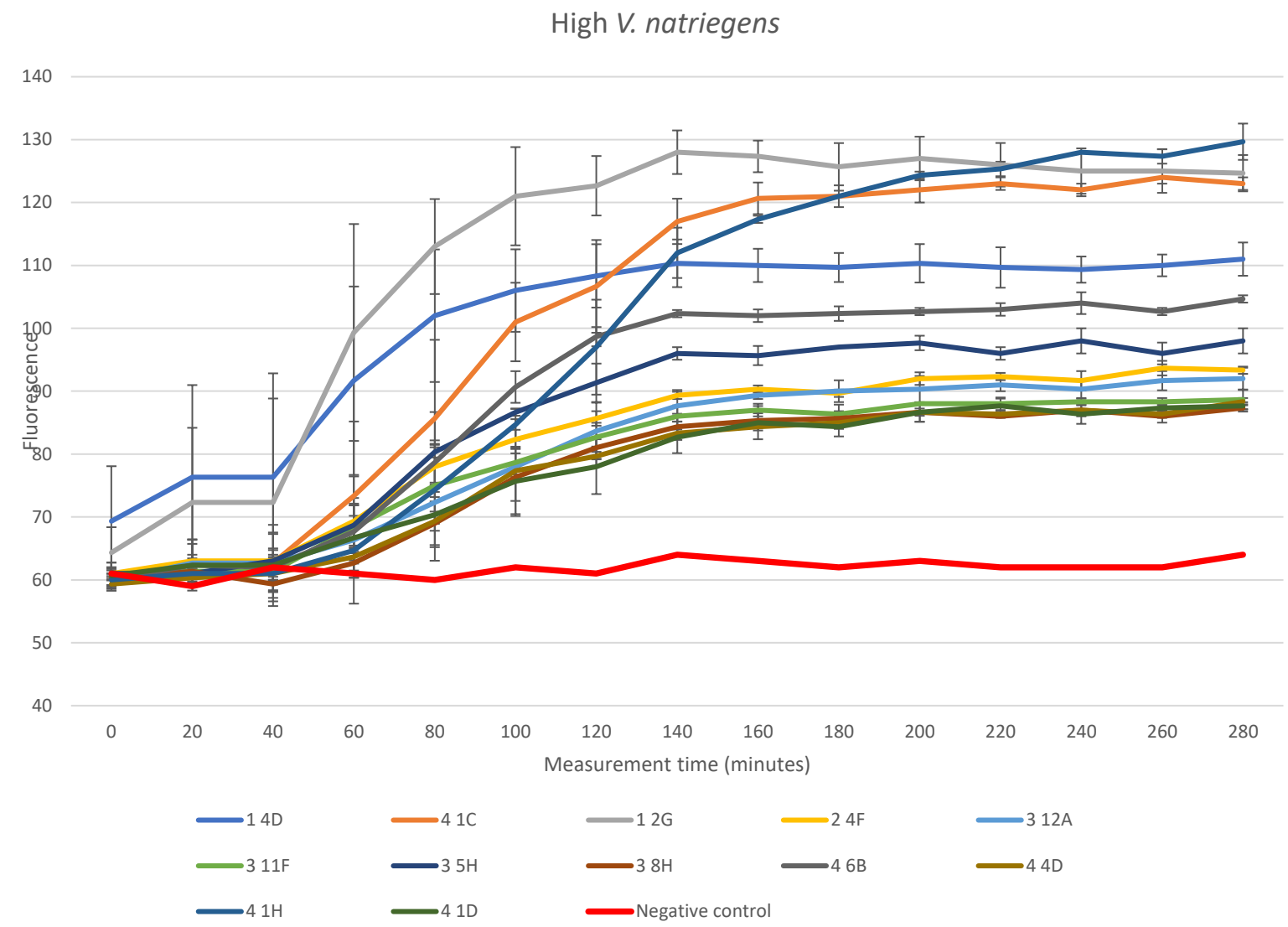


Figure 3.12: Fluorescence measurement results from the time dependent screening of GFP expression in *V. natriegens* resulting in high fluorescence levels. The measurements were continued for approximately 5 hours in 20-minute intervals. The measurements were performed in triplicates of clones and the graphs represent the calculated averages from each triplicate with standard deviations. The plasmid identities are named after the promoter and 5′UTR sequences identified by the screening of the pACYC-sfGFP 200N SD plasmid libraries in *V. natriegens*. The negative control plasmid (pACYC184) endpoint measurement was 64.

3.3.2 Comparison of GFP expression levels in *V. natriegens* and *E. coli*

In this section, the difference in GFP expression levels resulting from promoter activity is highlighted in tables 3.2 – 3.3 and graphically displayed in **figures 3.13 – 3.16**. Pairs of identical plasmids with unique promoter sequences was transformed into both *V. natriegens* and *E. coli*. The 200N SD promoter and 5′UTR sequences were confirmed as unique and complete by sequence analysis. The results of the sequence analysis conclude this chapter. The measurements detailed in this section were the last of the measurements from the time-dependent screening, when the GFP expression was highest overall, as displayed in **figures 3.7-3.12**

In **figure 3.13 – 3.16**, all fluorescence values have been normalized, by dividing the raw fluorescence measurements by the OD_{600} absorbance measurements belonging to each particular well on the 96-well plate. In practice, the resulting values in tables 3.2 and 3.3 below are well-specific normalized GFP expression levels. To highlight the difference in promoter activity characterized by GFP expression levels from identical plasmids in the two organisms *V. natrigens* and *E. coli*, table 3.2 and **figure 3.13** was created. In this particular snapshot, measurements from 9 plasmids which in *V. natriegens* led to low-mid-high GFP expression are displayed. As the expression levels in *V. natriegens* are markedly lower than in *E. coli*, the tables should be studied together with the graphs. In table 3.2 and **figure 3.13**, the GFP expression in *E. coli* do not follow the same pattern as in *V. natriegens.* At the lower range, the plasmid 1. 12F in *V. natriegens* has a normalized fluorescence expression value of 102 (VN negative control is 93). In *E. coli*, the measurement is in the middle range for the organism with a normalized GFP expression value at 535 (EC negative control is 72). At high range, the plasmid 1. 2G in *V. natriegens* with a normalized GFP expression value of 183. In *E. coli*, the measurement is in the low-middle range for the organism with a normalized GFP expression value at 365. In table 3.3 and **figures 3.14 – 3.16**, similar trends are shown in which promoters with a high activity and resulting normalized GFP expression levels are low in *V. natriegens* but high in *E. coli*, and vice versa. This will be further highlighted in the discussion section.

Table 3.2: Normalized GFP expression levels from plasmid pairs with identical 200N SD promoter and 5′UTR sequences transformed into *V. natriegens* and *E. coli*. The measurements are performed in triplicates with calculated standard deviations in the last row. The results are from the final time-dependent measurement at 350 and 250 minutes. Plasmid identities are marked in purple and bold for *V. natriegens*, and green for *E. coli*. Normalized negative control values for the plasmid pACYC184 is displayed in the first rows. The data is organized after low-mid-high GFP expression in *V. natriegens*, where the promoters in plasmids 1. 12F, 3. 7G and 4. 4H resulted in low expression, while 1. 4D, 2. 4F and 3.6B resulted in middle expression levels and 4. 1C, 1. 2G and 4. 1H resulted in high expression levels. The promoter sequences resulted in variable expression levels in *E. coli*.

The positive control (pACYC-sfGFP) values were 926 for *V. natriegens* and 3573 in *E. coli*.

Plasmid identity	Fluorescence/ absorbance	Standard deviation
Negative control VN	**93**	**0**
Negative control EC	72	0
VN1 12F	**102**	**2**
EC1 12F	535	18
VN3 7G	**98**	**0**
EC3 7G	181	6
VN4 4H	**100**	**3**
EC4 4H	380	15
VN1 4D	**128**	**2**
EC1 4D	1580	21
VN2 4F	**135**	**2**
EC2 4F	1368	31
VN3 6B	**124**	**2**
EC3 6B	2219	123
VN4 1C	**182**	**1**
EC4 1C	1684	30
VN1 2G	**183**	**3**
EC1 2G	365	7
VN4 1H	**182**	**3**
EC4 1H	2082	111

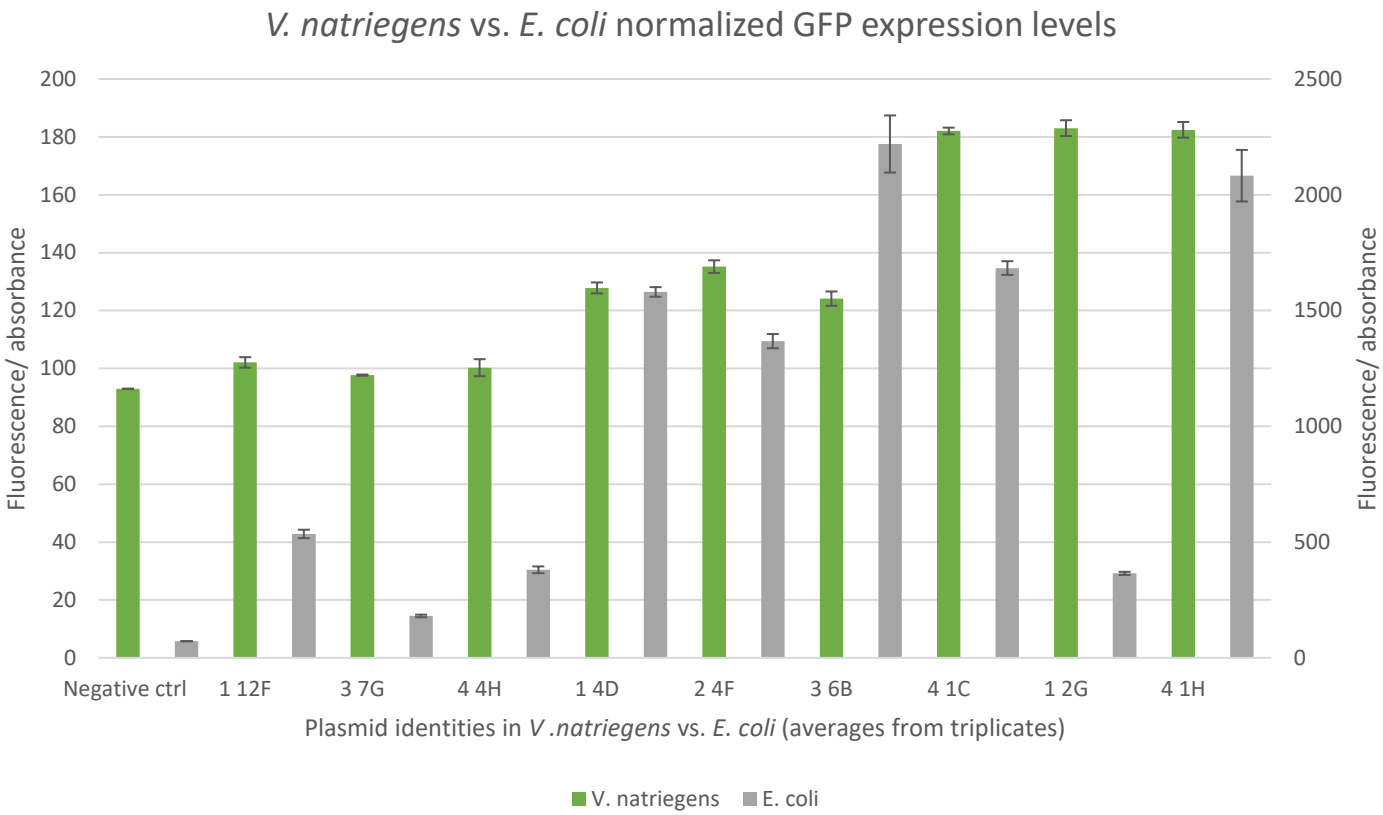


Figure 3.13: Normalized GFP expression levels from plasmid pairs with identical 200N SD promoter and 5′UTR transformed into *V. natriegens* and *E. coli*. The measurements are performed in triplicates with calculated standard deviations in the last row. The results are from the final time-dependent measurement at 350 and 250 minutes. Plasmid identities are marked in purple for *V. natriegens*, and green shaded for *E. coli* displayed to the right in the figure. Normalized negative control values for the plasmid pACYC184 is displayed in the first rows. The data is organized after low-mid-high GFP expression in *V. natriegens*, where the promoters in plasmids 1. 12F, 3. 7G and 4. 4H resulted in low expression, while 1. 4D, 2. 4F and 3.6B resulted in middle expression levels and 4. 1C, 1. 2G and 4. 1H resulted in high expression levels. The promoter sequences resulted in variable expression levels in *E. coli*.

The positive control (pACYC-sfGFP) values were 926 for *V. natriegens* and 3573 in *E. coli*.

Table 3.3: Normalized GFP expression levels from plasmid pairs with identical 200N SD promoter and 5′UTR transformed into *V. natriegens* and *E. coli*. The measurements are performed in triplicates with calculated standard deviations in the third rows. The results are from the final time-dependent measurement at 350 and 250 minutes. Normalized negative control values for the plasmid pACYC184 is displayed in the first rows. The total number of unique plasmids is 36. The positive control (pACYC-sfGFP) values were 926 for *V. natriegens* and 3573 in *E. coli*.

1. Plasmid identity	Fluorescence/ absorbance	Standard deviation	**2. Plasmid identity**	Fluorescence/ absorbance	Standard deviation	**3. Plasmid identity**	Fluorescence/ absorbance	Standard deviation
Negative control VN	**93**	**0**	**Negative control VN**	**93**	**0**	**Negative control VN**	**93**	**0**
Negative control EC	72	0	Negative control EC	72	0	Negative control EC	72	0
VN3 8H	**122**	**2**	**VN1 11F**	**108**	**0**	**VN4 2G**	**102**	**2**
EC3 8H	1184	28	EC1 11F	979	8	EC4 2G	209	3
VN4 1D	**128**	**2**	**VN4 4D**	**124**	**1**	**VN1 12F**	**102**	**2**
EC4 1D	1580	21	EC4 4D	1094	50	EC1 12F	535	18
VN3 9E	**109**	**0**	**VN2 10H**	**118**	**0**	**VN1 6B**	**105**	**1**
EC3 9E	1027	29	EC2 10H	998	23	EC1 6B	149	3
VN3 10D	**127**	**1**	**VN1 11E**	**105**	**1**	**VN3 6A**	**127**	**1**
EC3 10D	742	5	EC1 11E	644	27	EC3 6A	795	9
VN3 5D	**99**	**1**	**VN1 5D**	**114**	**1**	**VN1 4D**	**161**	**1**
EC3 5D	354	6	EC1 5D	453	17	EC1 4D	1627	45
VN4 4B	**104**	**0**	**VN3 11F**	**127**	**2**	**VN2 1B**	**111**	**2**
EC4 4B	394	9	EC3 11F	651	16	EC2 1B	252	5
VN3 7G	**98**	**0**	**VN3 4F**	**118**	**1**	**VN2 4F**	**135**	**2**
EC3 7G	181	6	EC3 4F	177	2	EC2 4F	1368	31
VN3 4B	**105**	**2**	**VN4 1C**	**182**	**1**	**VN3 10H**	**111**	**3**
EC3 4B	180	5	EC4 1C	1684	30	EC3 10H	312	5
VN1 4E	**122**	**1**	**VN4 4H**	**100**	**3**	**VN1 1G**	**97**	**2**
EC1 4E	620	7	EC4 4H	380	15	EC1 1G	130	1
VN4 1H	**182**	**3**	**VN3 5H**	**137**	**2**	**VN1 2G**	**183**	**3**
EC4 1H	2082	111	EC3 5H	442	17	EC1 2G	365	7
VN3 6G	**101**	**2**	**VN4 6B**	**150**	**3**	**VN4 5F**	**99**	**1**
EC3 6G	257	1	EC4 6B	782	15	EC4 5F	529	11
VN2 6E	**120**	**0**	**VN3 6B**	**124**	**2**	**VN3 12A**	**135**	**3**
EC2 6E	1564	22	EC3 6B	2219	123	EC3 12A	665	19

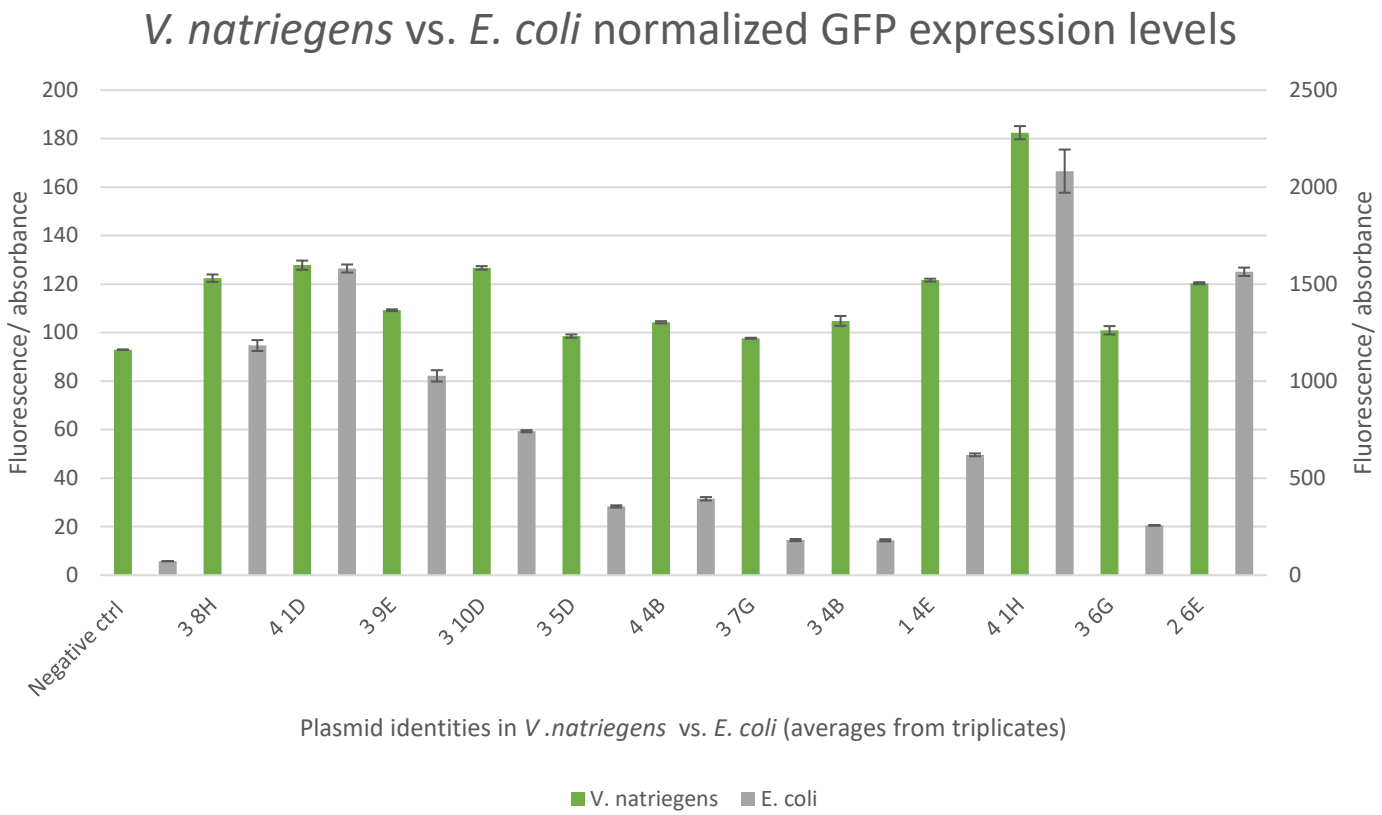


Figure 3.14: Figure 3.14 show the constructs 3. 8H to 2. 6E from the table section in table 3.3.

Normalized GFP expression levels from plasmid pairs with identical 200N SD promoter and 5′UTR transformed into *V. natriegens* and *E. coli*. The measurements are performed in triplicates with calculated standard deviations. The results are from the final time-dependent measurement at 350 and 250 minutes. Plasmid identities are marked in green for *V. natriegens*, and grey for *E. coli*. Normalized negative control values for the plasmid pACYC184 is displayed in the first column.

The positive control (pACYC-sfGFP) values were 926 for *V. natriegens* and 3573 in *E. coli*.

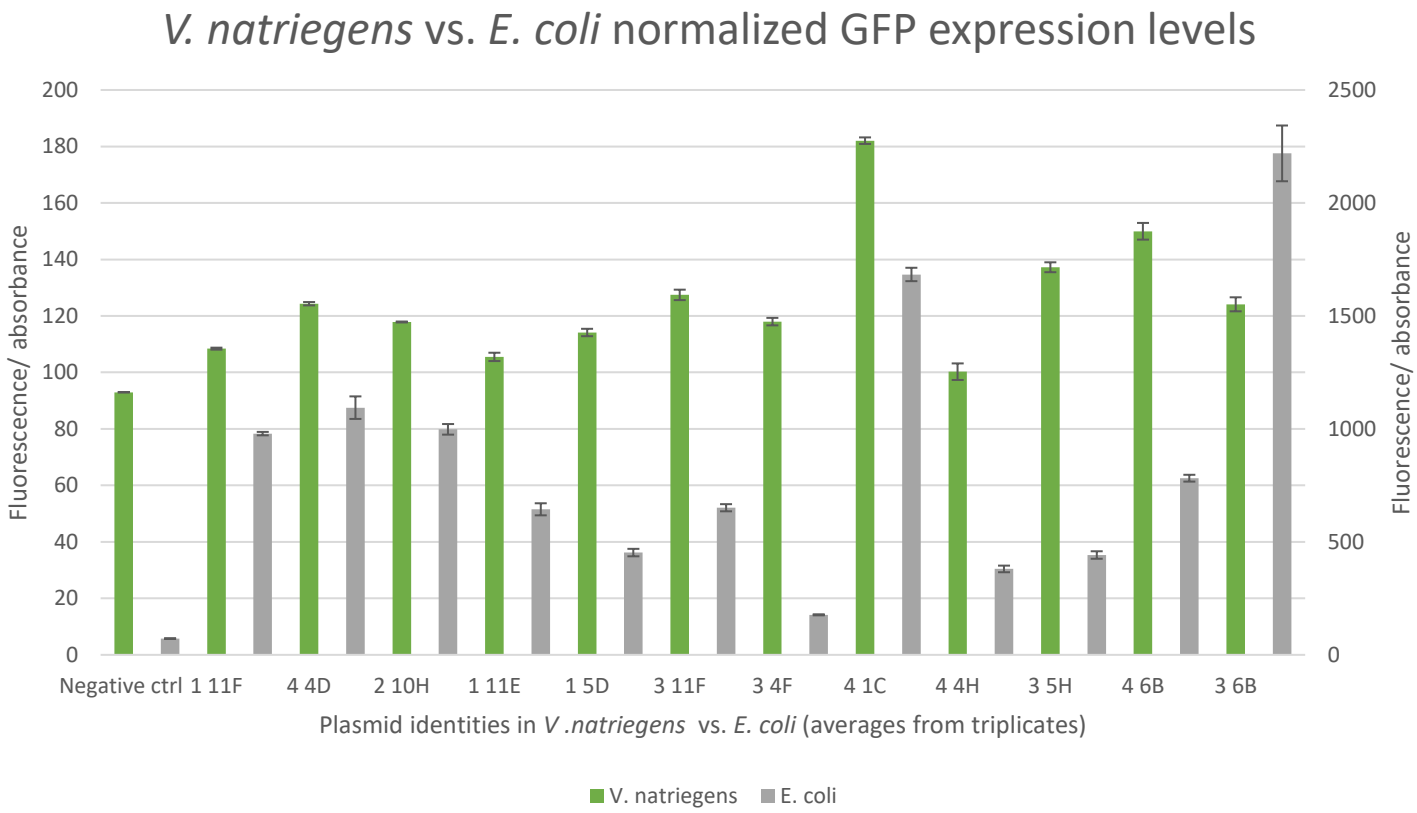


Figure 3.15: Figure 3.15 show the constructs 1. 11F to 3. 6B from the table section in table 3.3.

Normalized GFP expression levels from plasmid pairs with identical 200N SD promoter and 5′UTR transformed into *V. natriegens* and *E. coli*. The measurements are performed in triplicates with calculated standard deviations. The results are from the final time-dependent measurement at 350 and 250 minutes. Plasmid identities are marked in green for *V. natriegens*, and grey for *E. coli*. Normalized negative control values for the plasmid pACYC184 is displayed in the first column.

The positive control (pACYC-sfGFP) values were 926 for *V. natriegens* and 3573 in *E. coli*.

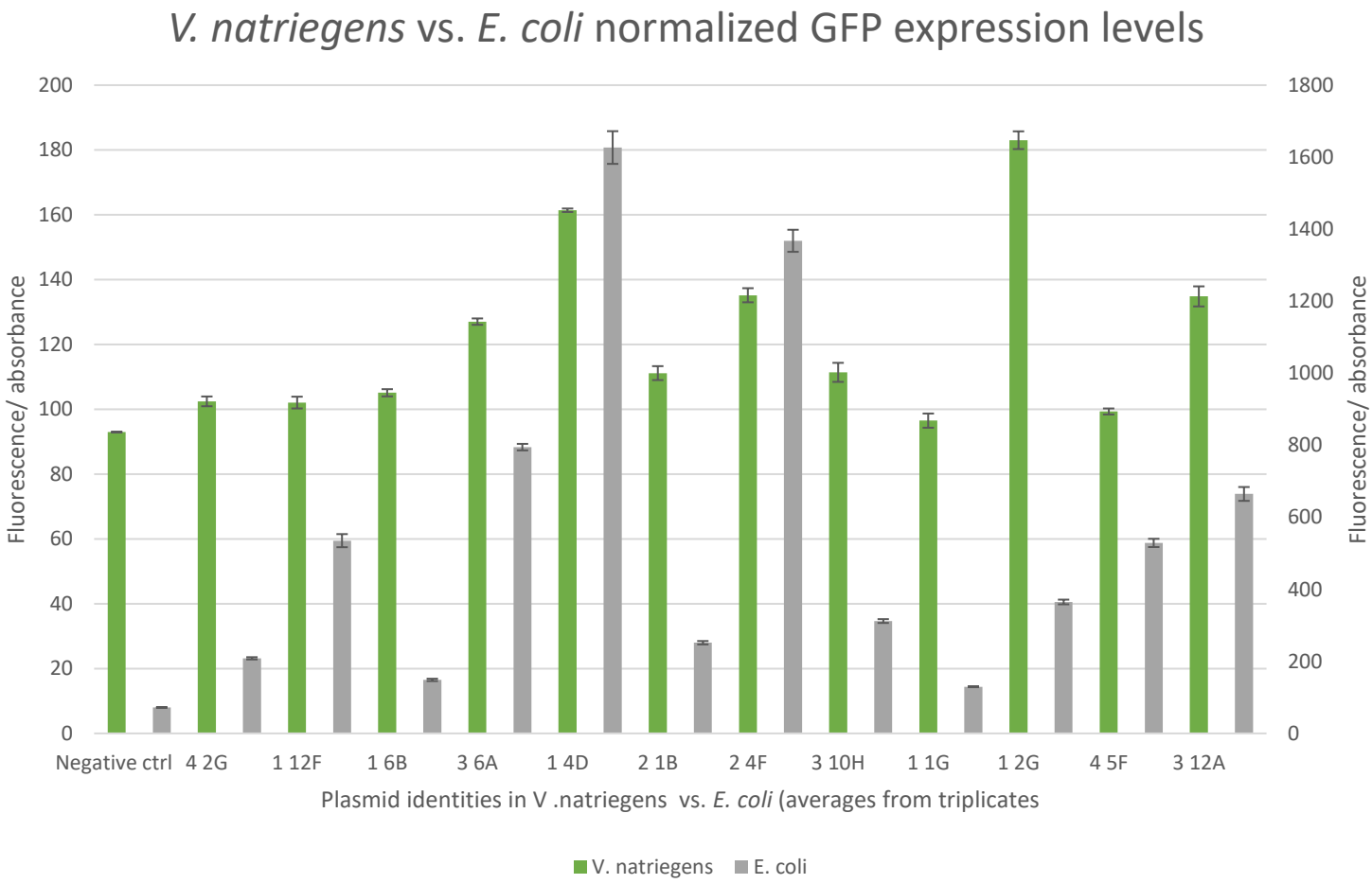


Figure 3.16: Figure 3.16 show the constructs 4. 2G to 3. 12 A from the third section in table 3.3.

Normalized GFP expression levels from plasmid pairs with identical 200N SD promoter and 5′UTR transformed into *V. natriegens* and *E. coli*. The measurements are performed in triplicates with calculated standard deviations. The results are from the final time-dependent measurement at 350 and 250 minutes. Plasmid identities are marked in green for *V. natriegens*, and grey for *E. coli*. Normalized negative control values for the plasmid pACYC184 is displayed in the first column.

The positive control (pACYC-sfGFP) values were 926 for *V. natriegens* and 3573 in *E. coli*.

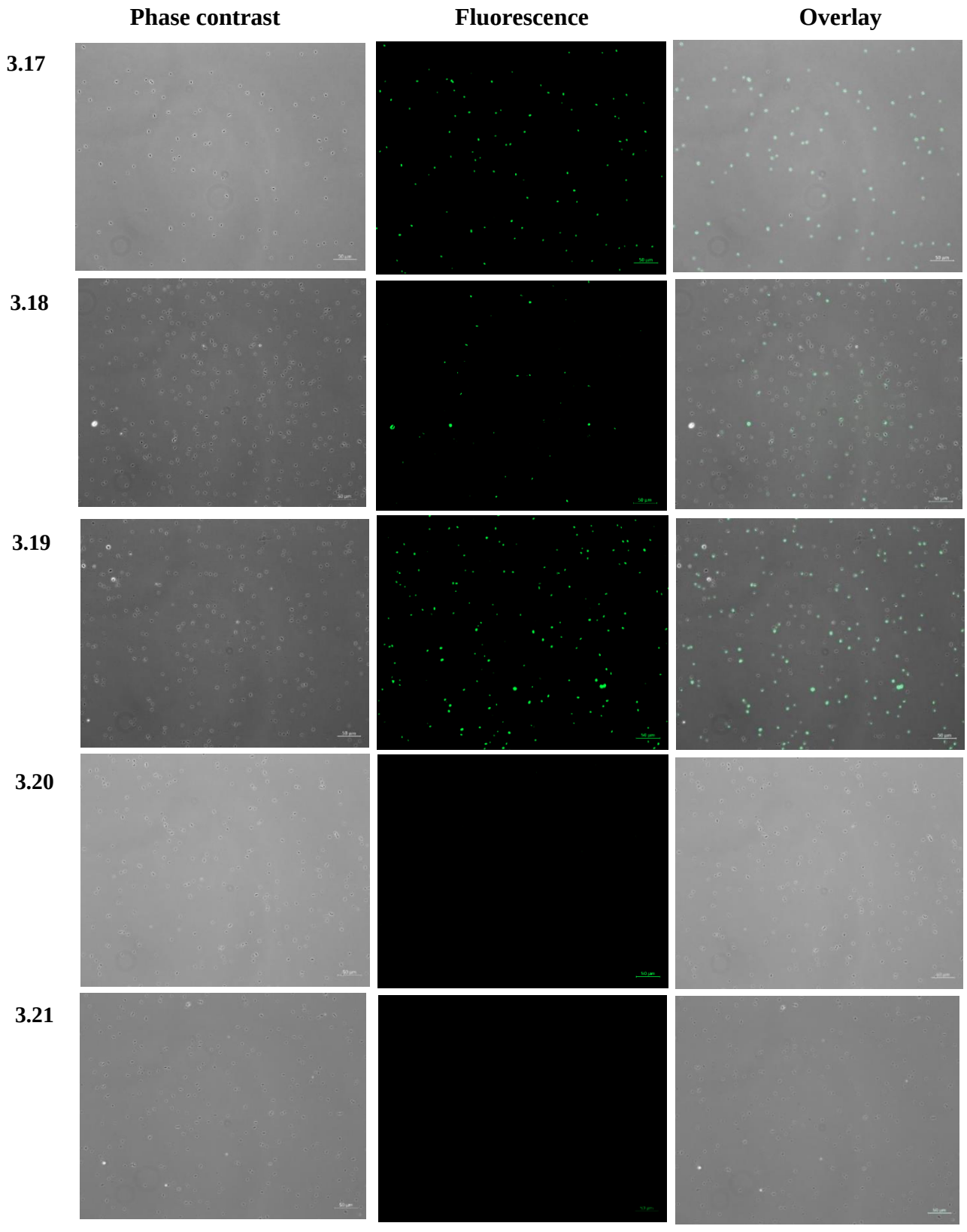
Student t-test

A student t-test was performed to show if there was a statistically significant difference of the different plasmids within the two organisms. The student t-test was performed in Excel on the average GFP expression values resulting from the normalized endpoint fluorescence measurements displayed in table 3.3. The plasmid identities measured in *V. natriegens* were tested against the plasmid identities measured in *E. coli*. The variation was expected to originate from the methodology and of the two organisms and led to equal variance between the two datasets. The resulting p-values were calculated by a two-tailed t-test, assuming equal variance but expecting higher expression from the plasmids in *E. coli*, thereby two-tailed. A p-value of 2,90238E-09 indicated that there was a large probability of a statistically significant difference.

3.4 Confocal microscopy of GFP expression in *V. natriegens*

Different microscopy methods were used to visualize the varying GFP expression levels in populations of *V. natriegens* cells transformed with the pACYC-sfGFP 200N SD promoter and 5′UTR libraries. imaging was performed with confocal laser scanning microscopy and phase contrast, displayed in **figures 3.13 - 3.17**. The result of the microscopy were three different types of images; phase contrast visualized all cells while confocal laser scanning microscopy visualized fluorescence and finally an overlay of the phase contrast with GFP expression. The overlay showed a visualization of which cells that expressed GFP, and the proportion of non-expressing cells.

Images were captured of a high-expressing population (figure 3.17), as well as population displaying lower expression (figure 3.18). As a positive control for GFP expression, a population of cells transformed with the pACYC-sfGFP plasmid (figure 3.19) were imaged. In figure 3.17 and figure 3.18, overlay of the high expression population show that the fraction of cells expressing GFP in this population is much higher compared to the lower expressing population. As negative controls of GFP expression, a population of cells transformed with the pACYC184 plasmid (figure 3.20) and a WT *V. natriegens* population of cells (figure 3.21) were imaged. The resulting images show no GFP expression in the negative control population, confirming the suitability as a control for the GFP expression screening.



Figures 3.17-3.21: Images of different *V. natriegens* populations as described in the text above. Imaging were performed by confocal laser scanning microscopy and produced phase contrast, GFP expression and an overlay. The scale bar is 50 µm.

3.5 Sequence analysis of promoter sequences in *V. natriegens* with BPROM

To further characterize the 200N SD promoter and 5′UTR sequences, plasmids were isolated and sent for sequencing. When the results arrived, the sequences were aligned in the software Benchling, with the pACYC-sfGFP 200N SD plasmid constructed *in silico*. A successful cloning was expected to result in a sequence that aligned completely with the backbone, 200N SD promoter and 5′UTR insert and ATG start codon at the sfGFP gene start. Additionally, this was an important step to discover any plasmid duplicates as results of poorly mixed 200N SD promoter and 5′UTR DNA libraries before transformation. As described in **appendix E**, all promoter and 5′UTR sequences were unique and distinct from the native mdh promoter and 5′UTR.

The tool BPROM was used to search for promoter motifs in the input, which was the 200N SD promoter and 5′UTR sequence isolated from the sequencing files from Eurofins Genomics. Table 3.5 show the output from sequence analysis in BPROM, with the position and score of potential promoter motifs and the position of the transcription start site. The total length of the input 200N SD promoter and 5′UTR inserts are also displayed, showing some variation in the length from 199 bp to 215 bp. All sequences contain an intact Shine-Dalgarno motif, -GGAG-, and most sequences have a complete ATG start codon at the end. The tool BPROM promoter predicter software was able to identify motifs in -10 and -35 positions for all sequences, and only one promoter was identified per sequence.

The search could potentially identify multiple core motifs for one promoter and 5′UTR sequence input. In addition to identifying possible motifs, the tool scored the strength if the located motifs. Scores from -6 to 100 were given, indicating how well the motif matched probable consensus sequences where 100 indicated the best located match. The scores varied for the -10 position from 21 for the promoter 1. 6B to 88 for the promoter 2. 1B. For the motif identified at the -35 position the score varied from -6 for the promoter 1. 3G to 66 for the promoter 4. 4B. The highest expressing promoter in *V. natriegens*, 1. 2G (183) that in *E. coli* is in the middle range (365) has a score of 63 at the -10 position and -6 for the -35 position. This was unexpected due to the low score at the -35 position. The lowest expressing promoter in *V. natriegens*, 1. 1G (97) that in *E. coli* is also low (130) has a score of 67 at the -10 position and 32 for the -35 position. This was also unexpected, as the scores indicated that these should be high producers. For the lowest expressing promoter in *E. coli* was 1. 1G (130) as described above for *V. natriegens*. The highest expressing promoter in *E. coli* was 3. 6B (2219) that in *V. natriegens* are in the middle range (124) has a score of 63 at the -10 position

and -4 (second lowest) for the -35 position. This indicates large discrepancies between the expected production levels from the sequence analysis and the results from the fluorescence measurements.

The ribosome binding site, or Shine-Dalgarno sequence in the 5′-UTR with the motif 5′-AGGAG-3′ as outlined in the introduction, has an optimum distance from the start codon. This optimum distance is 7 bp in *E. coli*, and the artificial promoter and 5′-UTR used in this thesis is designed to contain a 7 bp spacer. For the analysed sequences in **table 3.5**, spacing between the SD-sequence and ATG gene start varied between 7bp and 4bp, and only 16 of 37 promoter sequences had a spacer of the intended 7bp. In table 3.3, the normalized endpoint measurements of normalized GFP expression levels could highlight the effect of the spacer length. For the constructs with 7 bp spacing, the expression in *E. coli* and *V. natriegens* was mainly consistently medium high to high, including some if the highest producers in *E. coli*. The highest normalized endpoint value in *E. coli* was by 3. 6B (2219) and in *V. natriegens* by 1. 2G (183). Table 3.3.1 below summarizes the expression levels for these 16 constructs. The plasmid 3. 6G had only a 2 bp spacer, and the expression level was 257 in *E. coli* and 101 in *V. natriegens*. For the promoters and 5′UTR in plasmid 4. 1D (table 3.3.1), the spacer had an additional ATG start codon, while in 2. 4F ATG was mutated to ATTG at the gene start, which did not affect the expression level.

Table 3.3.1: Normalized endpoint fluorescence expression levels for 16 (of 36 total) *E. coli* and *V. natriegens* promoter and 5′UTR sequences with the optimum 7 bp spacing between the RBS and the ATG gene start. The promoter and 5′UTR sequences include the highest GFP producer in *E. coli* (3. 6B, 2219) and in *V. natriegens* (1. 2G, 183). In the first column, the plasmid identity of the promoter and 5′UTR are displayed, followed by the normalized expression level in *E. coli* in the second column and the normalized expression in *V. natriegens* in the third column. Standard deviations are displayed in table 3.3.

Plasmid identity	**Normalized endpoint expression level in *E. coli***	**Normalized endpoint expression level in *V. natriegens***
4. 1D	1580	128
1. 4E	620	122
3. 6A	795	127
3. 5H	442	137
3. 7G	181	98
2. 6E	1564	120
3. 6B	2219	124
4. 1H	2082	182
2. 10H	998	118
1. 11E	644	105
1. 5D	453	114
4. 4H	380	100
3. 11F	651	127
3. 10H	312	111
1. 4D	1627	161
1. 2G	365	183

3.5 Sequence analysis of 200N SD promoter and 5′UTR sequences in *V. natriegens*: BPROM results

Table 3.5: 200N SD promoter and 5′UTR sequence analysis output from the online software BPROM, a bacterial promoter predictor. The input was the 200N + SD and 5′UTR promoter sequence. The identity of the plasmid pairs with identical promoters in *V. natriegens* and *E. coli* is given in the first column, the promoter sequence with the sequences identified by the BPROM as possible -10 and -35 elements in the second column. In the third and fourth column is the position and score for consensus matches, which is scored from 0-100 where higher scores indicate stronger matches. In the last row, the position of the transcription start site (TSS) is given, as well as the total length of the promoter sequence.

	Bprom: SD sequence marked and any deviations from the ATG start. Motifs are marked in bold, SD-sequence is underlined and TSS positions are green.	**Position/ score**		**Pos./ Length**
Identity V.n/ E.c		**-10**	**-35**	**TSS**
3. 8H/ 10	ATATTTTATCGGATGTATGGTTCAATTAGAAAAATATATAGTG**TTGTTA**CGTCGGTTTTGCTC**TCTTAT AAT**CGTATG**A**ATTAGATGGACTTATTGGGCCTTTGCCCTAATTTGCATAAATTCGAGTATGAGCTATG GGATCTTAAGATGGTTTCAAATTCGCGTTTACTTTTGCAGGTTGGGCGTTTGATGGGGAGTATTATG	64/ 79	44/ 45	79/ 204
4. 1D/ 29	CTTGTGCGCTAGTATGCTACCCACTGATTCGCG**TTGATG**TTCACAAGTGGA**AGGTAGAGT**ATACAT**A** CCCTCATAAATTTGGGTAGGTCGGATTCAAGTTTGTGCCTAGACCTGTATGATTACCTAGTGGGTTAA CTTTGAGTGTATGGGGTAACCATTCACGCTTAAACTAGTCTAAACGGTGAGGTTCGCTGTCGTGCGG GAGAAATGGTATG	52/ 48	34/ 52	67/ 215
3. 9E/ 7	ATCGTCGCGTTGGGTGCATGATCCATAGGAAATTCGTTAATCATCGCGGTAGGCAGAAGCCCAGGG GCGAGAAGCTAGGTTAATCACATGTGGTGACGCGAAGTGATATGTAACGGAGGGGGCGTTGTACCT GTAAGGGGGG**ATGATC**AAGGGTGCCCTACATG**CCGTAGGAT**GTAATGT**G**CTTATAAGGGGGGGAG CGGTATG	165/ 43	143/ 6	180/ 204
3. 10D/ 32	CTATGTTGCTCTAAAAGGGGGGGGTTCTATTCGCATCGAACGTAGGCCTATCCACATCATCTTGCA TTACGGTTATGAATGACGGTGGCAATTATT**TTGAGC**CCCGTACGACCAATTACA**TGGTAATGT**TTGT ATTGAGGCCAATCGCCGAGTTTAGCTGTGAAATTTATGGCATGGTAAAGAGACTTTGGGGAGAAAA ATG	122/ 49	98/ 21	137/ 203
3. 5D/ 1	TTTTTTCGATTGTTCCTTATAGGAAATCCGATGTCACATAAGTATGGTAGGTATGTCGCTGTATTTTTG TGTTGCTTAGCGGGACTAGTGTTGTACTTGTGTCGTCTGGTCCGACGCTGGATCGTGCAGAAGTAGA GGG**TGTACG**CAGTAGGCCAACAAG**TGGTAACGT**TTTTTCTTCGGTTTGTCGGATGAGGAGAAGAAT G	161/ 45	140/ 4	176/ 203
4. 4B/ 28	GCGCATGTGGGCTTTTTGGGACTATTTTTATATGTGGGGTTCTCTT**TTGACA**AGCGGGTTTGT**TGGTG TAAT**ACAGGTA**G**GAAACACGAGGTGAATAGCTGGTCCGGAACAACATTGATTCCAGTTTGGTCGAT GCATAGGATCGAAGAACACAACTGCTTCTGTATTTCGTTACTTTAGGGAAATTAGGGTGGGAGAAAA TTG	64/ 56	47/ 66	80/ 204
3. 7G/ 20	CTGAGGGAGTTGTGTCGTGTGGGTATATAATGAGTGTTTTACGGTTCGATCTTGTCGGCGTCCTGGG GAGCCGTTTAAGAGGCGGGGGTTATTTTGACTTACTTGCATGTAGTGCTGCACGTAGTGGTGGGTTC GACATGAAGTGTTTC**GTGACT**TTACGTTGTCA**GTGTATCAT**TAGGAT**C**GCATATCTGTCATGTGGAT GGAGCGATCGGATG	167/ 66	150/ 28	182/ 215

3. 4B/ 30	CTGGTAAAGTTAATGGTATTGAGCTCCGGGTCGGT**CTGAAA**CTGTAACGTATTT**AGGTATGTT**GTTC GTGTGTGGTCTTTTCGAAAATACTATACGACAATATACCTGGGGAGGATCGTGCGATGTTACACCTG GCGGAGGGTGATACTGGCGTAGGGATTGAGATCATAAGGAAGATTGAGGCGGTGAAGAGGGAGG GGGATG	55/ 53	36/ 25	70/ 204
1. 4E/ 41	CATCGAATACTATTGTAAATACGGTATTTTTGTTGGATATTTGTTTAATAAGTAT**GTGCTG**AGAACGA TGCTTTC**TCGTACTAT**TTTTATTCTTTGCAGATTCACTCTGAATGCCTCCGTCGCCGGTAGACAGTCGG TTAAGTGACGACCATGCATAGGGGCGGGGGTGGCTGTGCATTATTGTGATTGCGGGCAGGTTAGG GAGAGTAGTGATG	76/ 50	56/ 14	91/ 215
4. 1H/ 5	AATCGTAGTGGACGGTTTAAATGGCACAGGCGT**TTTATT**TTTTGTATATC**GGATATACT**CTAGCAAT GTGAATTTCTTTTATAAACTGTATGCAAAACATTGCTGTCGAAGTGTGTGATTATTGCATTAAGTGTT GTTAATGCGGCGCAGGGCGGTCGTTCAGAGACGACTAATATAGCAGATGCGTTGGGAGGGGTTATA TG	51/ 69	34/ 34	66/ 203
3. 6G/ 39	TTAAGCCTTATACCGTGGCTTGG**TTCACC**AAAGAGCTTGAGT**TTCTATTAT**GTTTTTTGCGGATCCTG ATAGACCAACGAAGGTTGGAGTGGGCGTAATGAGTTATATGATAGAGTAGCAACGAGTCGTCAGCC AGCCGTTTGAAGCGAGTTTAGGGAAGGGGGATTTGCTGTTGGTCGGAAAGAGTGAAGGAGGTATG	43/ 67	24/ 24	62/ 199
2. 6E/ 35	CGATATACAGAGGTCGCTTTTACATATTTTGAGCC**TTGAGA**GCGTACTAGATCA**ATGTAAGGT**GGTT GAACTACTATGTTATGTGGATTGTTACGTATTTGGGAATGAAGGGTTGGGTGTGTAATCGGTTGGTG GTTACGCTGTGGCTTTGGCACTTACGGACTTGTTAGGCGTGGTGGGAATGGATCTGATAGGGTTCTT GGAGTGTGGCGATG	55/ 35	36/ 43	70/ 215
1. 11F/ 21	AACACATTTACCAGCTTTATGGTCGCTCTTTTCGCTGCGGAGATCTTGTGTTT**TTGGTA**ACGGGCTCA CGTTTC**AGGTAATGTG**GTTAG**T**GTTTCTGTATTATCTCGCACCGGTCCCTAGGCCAAGACCGTCTTCT GGGTCGGACAACTTATAAATAGGCTAGATTCCATGTACAACAAAGCATGATACTGGAGGAGAGGGA TG	75/ 44	54/ 30	90/ 204
4. 4D/ 25	TATCTGGTATGAATGCGTTCATTTCGGATGGGGGGGACCAACGTACTAGTTCCAGCAATTTAAATGT AGTAGTGTGTGCACCCAATTTCCTGCGACGGCAAGAGCTGATAGTCATAAACGATCAGAATTATGTA GTTTCAGGTTGGGTA**TTGTCG**ACGCAAACAAGA**GGGTACTCT**GAGGCT**A**GAAGGAGGGGAGTGAG ATG	168/ 51	150/ 47	183/ 202
2. 10H/ 38	TTTTTGCTTTATAGGATTGGTGCGGGGGAGTGCTCTATCTGTTCTTTTGGGTTAAAGAGGTGCCTTAT CGTTTTTTGGCAGAAGTGTTGATGATGTGGGTAGGTC**TTGTGG**TACAGATCCTGAAA**TTATACAGT**G GCGGATTTATTCAGTTTTGTACCGCTGAGGGTGGTTTAAATCAGCTTTCGAGCGGGAGACTTATCAT G	126/ 35	106/ 24	141/ 203
1. 11E/ 13	CTCTGATAGTCCGGTTGCTTATTACGACTATGCTGATTAGAGATTCGGAGTAGTTACTGCTATGTATG TGAGGGTAGTCA**TTTCAG**TGTCGAGAAGTGTGCACG**ATTTAGGTT**TGATATTCATTCACCGGACGTT TGGCGCGAATTGTGGATTTTACGGCCGGGGTATCTGTAGGTTGACCGCTTGTTGGGAGCGCCCGAA TG	105/ 28	81/ 30	120/ 203
1. 5D/ 33	ATCGTGATATCAGTGATTGTCATT**TTGTTA**GAGTCCTACATTAGGACTG**TGATAGTCT**TGATTGATGG TAGTTTTTTGTCTTGTCTTGTGTAGTGTATGGTTACTTCGGATTTAATCCTTAGGTTTTGAAGTGCAGG GGAAGGTCAAGCTGCTGAATAATGAAACTAAAGTGAGGGACCTTTGTTGGTCTTTAAGGGAGTCCT CGGATG	50/ 41	25/ 45	65/ 209
3. 11F/ 15	TGGTTATGTGTCGAAACTATGTGATTTTGCTTTTTAGGGGCATTTATGGTAATAGTGGTCTTGTTCGG TCTGCTCATTTTGGGTTATG**TAGATA**CGTTACCGTGGCT**TTGTGTAAT**CATTGCCCCTAGTGTCCTTG ATGCTATCGGCTTATGGGCGATCAGGGTTGTACCGGTAAATTGGTGAAAGGAGTGGCATAGGTTGG AGACTCGCAATG	108/ 49	89/ 21	123/ 214
3. 4F / 3	GATATAGTTCATTCGCGTGTTAAGAAAAATCGGAACAACTGTTAAATTAGGATCAGTAAAAAACTGG TGCGGTTTAGTCTGCTGTAGTGTGCCGAGGGCTCTTATATAGTACCGCCGATTCCGCGCTGATAGCC CGT**TTCTCA**ATTGATAATGGCTT**TGTTTGAAT**GCCCTGGAGTAGATGTGACAGCCAGCAGGAGTGTT ATG	158/ 38	138/ 33	173/ 204

4. 1C/ 22	ATAATGATTGAAGTGGCTCAGTAC**TTGTCA**ATTCTAAACGGCCGGACA**TGGTATGGT**TAGGGTTAAT AGTCGTTTAGCTTCATCCTCTATTACGGATGGGTGTTGAAAGGGGCTATAGGTTCGGGGTATTGTCT GTTGGGGGATTAGTTTAGATGGGTAATTGCATTGGTTGTGGGCAATGTGGCCGTGGACGGAGAAA GAGG	49/ 54	25/ 53	64/ 204
4. 4H/ 43	TCCGGTTATTAATAGAGTGAGTCTTTTGTCAATATGGGTTATTGGACTGGAATCAATGTGGGGAGTT TATCGGATGCTCGGGCACTTGGAGTGTGAGTTAGCAAGTAGTTGTCACTTTGATCGGGTGCGGTT**TT GTCA**GCTCGGTCGTGGAAA**TTTTATTTT**GATATGTCGCATACATTGCTATCGGAGTCTTATGGAGTCG TAGGATG	154/ 52	133/ 53	169/ 209
3. 5H/ 12	CTTTGTAGGCATTCTCGTTATTAAGATAAGAGTCCAGTCATTAATGTGCTCTTCGGTCTGTC**TTCATT**G CTGTCTGTGGTT**AATTAATAT**TCGCGCTATGTTGTTTCTTTTTTGTACCGGGTAGTGGAAGGCCTATC GTTTTGGACCGTATGAGTGCTGTATCAGCTTATAGCTGTAGTGTGTGGAGTTTAGGTAGGGTGAGG AGCGTTAGTATG	82/ 53	63/ 33	97/ 215
4. 6B/ 40	TGTTTACTCTCTAGTTGTTGGTAAGTAGTCTTTAATTCTAGTTT**TTTATT**CTTGCGTTAATGTTAGTT**TT GTATAAT**CACACTTCGTGGAAGATCGCACGTTCGCCGTCCAGGAATAAGGACGTGTGGTCGGGTTTT GGTAGAATTGTGTGGATTCTTATGGTAGGATGTCAACTTTCAGATTGCGAGTCCGGGAGTCCAATG	69/ 87	45/ 34	84/ 204
3. 6B/ 26	GTGCATCATTAGCCCGTTTCGGTCCTGCAGAGTATACTATGGATCGATCCGGGCCTGGGAGGTTTTT CATGTCGTCACTCGCGTAAGGTTTGCAGCGCATACAAGAGAAGTGACTTGGGTAATTATGAGTGAA AGG**TCTATG**GTCAGGTAGGGTTG**TTGTATAGT**GCTTATGTGATCATAGTGTGACTTTTATGAGGAGC ATTAGTATG	157/ 63	137/ -4	172/ 209
4. 2G/ 6	TGTTGGTATCCATGAGTTTACTGATGAAGCTTAGTACTTCCGGGTATTTCATGTTCACACAGAGTGTA GAGTTACTATGGCAGTTCTGTTATT**TTGACG**GTGTGAGGGCGTT**TGTTATCGT**GCTAGTAAAATCGT CTAATGTTGTTGTGGAAGCTTGACATAGTCATGTTAATCATGAAGGTAGTGGCTGGGAGTAGGATG	113/ 51	94/ 60	128/ 201
1. 12F/ 36	TAACATGCTAGGATGGTGAGAAGGAACTACCATTG**TTATCG**ACCCGGGTTACCATCTTT**AGGTAATA T**CCCGACATCGCTTGATAAATAAATAGATCGGTGTTTGTCACTACTAGTGCCAGGTGCACTGTGTTG GTGTAAACCGGGTGTACTCCAAGAAATGGGTAGAGTCAATTTCCATATCCAAGATAGAGGGAGTGA GATG	60/ 68	36/ 24	75/ 204
1. 6B/ 31	ACGTCACTAGATTGCCGTATGTTATTCATTTTTCTAGCAGGATAATGCAACCAATAAGTAAACCAACC **TTGTGG**ATGCTGAAAAG**TGCAAATCT**TGCATGCACAACTTTGCAGAGTGTTGACACATGCAAAGAG GGGAGTCGGATTTTGGAAATGTCGTCGAAACCACGCGCGATATATGGCGACAGGACGTGGAGGGG GATG	86/ 21	69/ 24	101/ 203
3. 6A/ 19	CTTTACATGTTATGCTTTGCGTTGTTGAACAGTTGTTAATTGGGCGAACCGACTATGAGAATCTGTTT CTTTTGTAATTATT**TTTACA**AGGGCGTTTTCGGAT**AGATAAACT**TTTGTATGGAGGTTTGGGCGACTT GCATACTTACGTAGGGTACTATTTGTAAACTTAGTAGACGTAATTATGCGAGATCGTGATATTTGGGG AGTTGTGTTATG	104/ 60	83/ 47	119/ 215
1. 4D/ 18	GATGTTACGTGCTTTTTTGCGGTAG**TTGCCT**AGTGATTCAGGGCA**GTTTATATT**GTGCTAACATCTTT CATTCCAATCTTGTAATGTACACATATTTTCCTGATTATGACGGAGGTCCTGTAACACGGCTGGGAGG TATGCGATTGGATGCGCGCCAATTTGAATGTTTTTTTGTTTTCGGGGGGGATAGGAGGTGTTAGATG	46/ 63	26/ 56	61/ 203
2. 1B/ 34	CGGTTTAGGTTTAGTAACAG**TTTACG**AAACTGTTGTTTGTG**AGTTATAAT**GCGGACTGTCTTGACGT AGTGAGTGCTAGATATTGGAACTGGCATGCTGGCCCAATGAACTGGATACACCTGAGGCCGTAAGA GCTGTCGGGTCGTATGGCTGAGGCGTTTAAGAGATGCGTTTTGTTGAGTAGTAAAGCAATGGAGAG CTATG	42/ 88	21/ 41	61/ 204
2. 4F/ 23	TCTATTTTGACGCTTTACTTTCTTTATGGTCGAATTACGAAATGCCTTTTCACTTTAGCAATGACGGTA CCAGTCACGTCATCGCGT**TTGTCA**TCTGAGCGTTTATT**ATCTATTCT**CTGTTTTAAAGTACGAGGAAT AAGTGGGGATAGACTGACGGAGACGGCACTACTGATGGGCCAGGAGAAGAGCGTGAGGAGAGAT AAG	108/ 52	88/ 53	123/ 204
3. 10H/ 2	CTTGTTAGGATTGGTAGTATTGATGGAATAGTCACGTTGATCTG**TTTAAA**CAAGCTGTAAGG**GGCTA GCTT**CGTCTTATTCGGTGGATGAACTAGTGTGCTGACGTGTGGGTTTACTTGGTTAGAGTCATAGCC	63/ 30	45/ 41	78/ 215

	TGAGGCTTGGCGGCTAGAGCGTCTGGTGGAGTGTGGAGTGAGAGAGGTGAATGCAATGGATAAGAATGGAGTCAGGGTATG			
1. 1G/ 16	GTTCATCCTGTCGTCGTTATAACGGTTATTCGTTCGCTTTTTTCCTGATAGTGTTAGAATAAATCTAAAGACAAAGT**TTGGAA**AACCGGTAATGTT**TTCTATTAT**AGTCTTG**A**CCACCTCAGGTTTGTTGACGGTAGGAAGCCACCGTTTTTCAGGGAACTGTCGGATGGGGCTTGTGGGCAGGGACGGGCGGGAGACGAATG	97/ 67	78/ 32	113/ 203
1. 2G/ 42	CTGAAATTGGGTAAATTTATGCAGTGTTTATTTAGGATTTATCTAACG**TAGCTC**TTTTTGATCCCTGGG**GGATATGAT**TATGAA**G**TTGATATGTGTACGAGGGTTATTTGATGTTATTTCCGGTTTTAGAACAGTGCGTGTGGTGCGGGTATATGTATGTGATCAACGTTTCAAGGATGTATGAGGAGGGCACACGCCAGGGGAGGGCTGAAATG	70/ 63	49/ -6	85/ 215
4. 5F/ 4	AATTAGGGACGCTTGGGCGGTCGAAATCGTAGGACACAAGTGGCTTTATAACTCT**GTGTCA**TAGCTCAGTTCG**GGGTATGAT**CGTGGA**G**CAGCTAGCTGCGTGTCCTCCAGACAGTCAGAGGGGTGTAAGGTGCCATGTGGATGAGGATGGCTGCTCCTCGGGGCGGTAATGTATCTGAATGAATTCTTGTGGGGAGGTTTATG	74/ 75	56/ 20	89/ 204
3. 12 A/ 9	GCGTTGTTATATAGAGTCTCAATCATAGGTGCATGGTTCTGAAGTTGGAGTACTTATTTAGATCTATCACCATTTTCGGGAGAGTTTTGG**TTGCGG**CACTGATTGTATGTTGG**TGTGAAGAT**TCACT**G**TTTCTTTTAGGGACTGTGGGTAAGTAGTGCGGAGTGTTGTCGTTGGTAGGAGGGGTTGTTTGTTGTTGGAGGGAGTGGATG	114/ 34	91/ 32	129/ 209

3.6 Sequence analysis of the 200N SD promoter and 5′UTR sequences in *V. natriegens* with Improbizer

The online tool Improbizer was utilized as an additional method for possible detection of consensus sequences for the -35 and -10 positions in the unique promoter and 5′UTR sequences . In table 3.10 the results from Improbizer runs are detailed. The output is color-coded, where darker colors indicate stronger matches to a probable consensus sequence. Improbizer works by identifying sequences unlikely to occur by chance. As Improbizer could analyse maximum 20 sequences at a time, the tool was run two times. For the first run, with sequences 3. 8H to 4. 1C Improbizer identified sequences similar to TCTTTTA at the -35 position, and TGGAGGGAGA for the -10 position. For the second run, with sequences 4. 4H to 3. 12A sequences similar to CAGTCATAGCT for the -35 position and GTGGGGAG for the -10 position. The motifs located by the Improbizer tool indicated that the -35 position tends to be A/T- rich while the -10 positions tends to be G-rich.

In table 3.10 it is indicated that several pattern matches to possible consensus motifs are located. The patterns located for the -35 and -10 positions by Improbizer were created by identifying one initial pattern, and through multiple repeated pattern-searches (iterations) the software produced an average of the resulting patterns. These patterns are represented by the sequences in the previous paragraph. Improbizer automatically adjusted the sizes of expected motifs from eight to variable lengths.

Improbizer produced scores from 1-0 for indicating the strength of a match to the motif, with scores close to 1 for individual nucleotides in the motif indicating higher chances of a conserved position. If the score is 0, there is no indication that the nucleotide in the motif will occur in that position. At the following page, table 3.6-3.7 displays the scores generated for the profiles generated in the 20 first sequences run in Improbizer, and table 3.8-3.9 for the last 16 sequences. The results are concluded with a comparison of the output from the two methods in section 3.7.

Table 3.6-3.7: Identified motifs, or profiles by the tool Improbizer for the first 20 200N SD promoter and 5′UTR sequences in **table 3.9**. The placement of the motifs in the sequences is given in the top first column with an sd-value. The motifs are scored from how improbable they are to have occurred by chance at the rates found in the sequences. Individual nucleotides in the generated profile given in the top second column are scored from 1 to 0 indicating chances of a conserved position. If the score is 1, the position is probably conserved.

10.0547 @ 51.55 sd 24.92			**Generated profile: TCTTTTA**				
Nucleotide	**Score for nucleotide position in profile at -35**						
A	0.052	0.003	0.322	0.126	0.251	0.052	0.547
C	0.177	0.990	0.003	0.177	0.027	0.003	0.326
G	0.076	0.003	0.102	0.150	0.102	0.098	0.003
T	0.694	0.003	0.573	0.547	0.620	0.847	0.124

10.9645 @ 185.16 sd 14.58					**Generated profile: TGGAGGGAGA**					
Nucleotide	**Score for nucleotide position in profile at -10**									
A	0.077	0.257	0.077	0.449	0.276	0.028	0.003	0.769	0.177	0.400
C	0.250	0.048	0.049	0.124	0.150	0.003	0.003	0.151	0.125	0.152
G	0.322	0.350	0.548	0.128	0.500	0.965	0.990	0.003	0.646	0.250
T	0.351	0.326	0.326	0.300	0.074	0.003	0.003	0.076	0.052	0.198

Table 3.8-3.9: Identified motifs, or profiles by the tool Improbizer for the last 17 200N SD promoter and 5′UTR sequences in **table 3.10**. The placement of the motifs in the sequences is given in the top first column with an sd-value. The motifs are scored from how improbable they are to have occurred by chance at the rates found in the sequences. Individual nucleotides in the generated profile given in the top second column are scored from 1 to 0 indicating chances of a conserved position. If the score is 1, the position is probably conserved.

6.7581 @ 104.86 sd 44.17				**Generated profile: CAGTCATAGCT**							
Nucleotide	**Score for nucleotide position in profile at -35**										
A	0.120	0.352	0.292	0.352	0.061	0.932	0.061	0.699	0.120	0.003	0.003
C	0.410	0.236	0.003	0.003	0.932	0.003	0.178	0.120	0.120	0.643	0.120
G	0.236	0.060	0.469	0.120	0.003	0.062	0.061	0.061	0.758	0.061	0.176
T	0.234	0.352	0.236	0.525	0.003	0.003	0.700	0.120	0.003	0.292	0.701

9.0229 @ 198.82 sd 1.54			**Generated profile: GTGGGGAG**					
Nucleotide	**Score for nucleotide position in profile at -10**							
A	0.177	0.236	0.294	0.236	0.003	0.003	0.990	0.003
C	0.119	0.177	0.003	0.003	0.003	0.003	0.003	0.003
G	0.526	0.177	0.352	0.526	0.990	0.990	0.003	0.990
T	0.177	0.410	0.236	0.235	0.003	0.003	0.003	0.003

Sequence analysis of the 200N SD promoter and 5′UTR in *V. natriegens*:

Improbizer results

Table 3.10: 200N SD promoter and 5′UTR sequence analysis output from the online software tool Improbizer, which recognizes possible consensus sequences. The input was the 200N SD promoter and 5′UTR sequence. The identity of the plasmid pairs with identical 200N SD promoter and 5′UTR in *V. natriegens* and *E. coli* is given in the first column, the 200N SD promoter and 5′UTR with the sequences identified by the Improbizer as possible -10 and -35 elements in the second column. The sequences are color-coded, where darker colors indicate stronger matches to a probable consensus sequence. The strengths of the matches are coded from turquoise to grey to black/bold underlined for low to medium to high matches. The SD-sequence is marked in red. In the third column the total length of the promoter sequence is given. The motifs are

Identity	**Improbizer:** Promoter -35 and -10 approximate locations, Transcription Start Site.	Sequence length
3. 8H/ 10	ATATTTTATCGGATGTATGGT**TCAATTA**GAAAAATATATAGTGTTGTTACGTCGGTTTTGC**TCTCT TA**TAATCGTATGAATTAGATGGACTTATTGGGCCTTTGCCCTAATTTGCATAAATTCGAGTATGAG CTATGGGATCTTAAGATGGTTTCAAATTCGCGTTTACTTTTGCA**GGTTGGGCGT**TT**GATGGGGAG T**ATTATG	204
4. 1D/ 29	CTTGTGCGCTAGTATGCTAC**CCACTGA**TTCGCGTTGATGTTCACAAGTGGAAGGTAGAGTATACA TAC**CCTCATA**AATTTGGGTAGGTCGGATTCAAGTTTGTGCCTAGACCTGTATGATTACCTAGTGG GTTAACTTTGAGTGTATGGGGTAACCATTCACGCTTAAACTAGTC**TAAACGGTGA**GGTTCGCTGT C**GTGCGGGAG**AAATGGTATG	215
3. 9E/ 7	ATCGTCGCGTTGGGTGCATGATCCATAGGAAAT**TCGTTAATCATCGC**GGTAGGCAGAAGCCCAG GGGCGAGAAGCTAGGTTAATCACATGTGGTGACGCGAAGTGATATGTAACGGAGGGGGCGTTG TACCTGTAAGGGGGGATGATCAAGGGTGCCCTACATG**CCGTAGGATG**TAATGTGCTTATAA**GGG GGGGAGC**GGTATG	204
3. 10D/ 32	CTATGTTGCTCTAAAAGGGGGGGGT**TCTATTC**GCATCGAACGTAGGCCTAT**CCACATC**ATCTTG CATTACGGTTATGAATGACGGTGGCAATTATTTTGAGCCCCGTACGACCAATTACATGGTAATGT TTGTATTGAGGCCAATCGCCGAGTTTAGCTGTGAAATTTATGGCAT**GGTAAAGAGA**C**TTTGGGG AGA**AAAAATG	203
3. 5D/ 1	TTTTTTCGATTGTTC**CTTATA**GGAAATCCGATG**TCACATA**AGTATGGTAGGTATGTCGCTGTATTT TTGTGTTGCTTAGCGGGACTAGTGTTGTACTTGTGTCGTCTGGTCCGACGCTGGATCGTGCAGAA GTAGAGGGTGTACG**CAGTAGGCCA**ACAAGTGGTAACGTTTTTTCTTCGGTTTGTCG**GATGAGGA GA**AGAATG	203
4. 4B/ 28	GCGCATGTGG**GCTTTTT**GGGACTATTTTTATATGTGGGGT**TCTCTTT**TGACAAGCGGGTTTGTTG GTGTAATACAGGTAGGAAACACGAGGTGAATAGCTGGTCCGGAACAACATTGATTCCAGTTTGG TCGATGCATAGGATCGAAGAACACAACTGCTTCTGTATTTCGTTAC**TTTAGGGAAA**TTA**GGGTG GGAGA**AAAATTG	204
3. 7G/ 20	CTGAGGGAGTTGTGTCGTGTGGGTATATAATGAGTGTTTTACGGTTCGA**TCTTGTC**GGCGTCCTG GGGAG**CCGTTTA**AGAGGCGGGGGTTATTTTGACTTACTTGCATGTAGTGCTGCACGTAGTGGTG GGTTCGACATGAAGTGTTTCGTGACTTTACGTTGTCAGTGTAT**CATTAGGATC**GCATATCTGTCAT G**TGGATGGAGC**GATCGGATG	215
3. 4B/ 30	CTGGTAAAGTTAATGGTATTGAGCTCCGGGTCGG**TCTGAAA**CTGTAACGTATTTAGGTATGTTGT TCGTGTGTGG**TCTTTTC**GAAAATACTATACGACAATATACCTGGGGAGGATCGTGCGATGTTACA CCTGGCGGAGGGTGATACTGGCGTAGGGATTGAGAT**CATAAGGAAG**ATTGAGGCGGTG**AAGA GGGAGG**GGGATG	204
1. 4E/ 41	CATCGAATACTATTGTAAATACGGTATTTTTGTTGGATATTTGTTTAATAAGTATGTGCTGAGAAC GATGCTT**TCTCGTA**CTATTTTTAT**TCTTTGC**AGATTCACTCTGAATGCCTCCGTCGCCGGTAGACA	215

	GTCGGTTAAGTGACGACCATGCATAGGGGCGGGGGTGGCTGTGCATTATTGTGA**TTGCGGGCA GGTTAG**GGAG**A**GTAGTGATG	
4. 1H/ 5	AATCGTAGTGGACGGTTTAAATGGCACAGGCGTTTTATTTTTTGTATA**TCGGATA**TACTCTAGCA ATGTGAATT**TCTTTTA**TAAACTGTATGCAAAACATTGCTGTCGAAGTGTGTGATTATTGCATTAAG TGTTGTTAATGCGG**CGCAGGGCGG**TCGTTCAGAGACGACTAATATAGCAGATG**CGTTG**GGAG**G** GGTTATATG	203
3. 6G/ 39	TTAAG**CCTTATA**CCGTGGCTTGGTTCACCAAAGAGCTTGAGTT**TCTATTA**TGTTTTTTGCGGATCC TGATAGACCAACGAAGGTTGGAGTGGGCGTAATGAGTTATATGATAGAGTAGCAACGAGTCGT CAGCCAGCCGTTTGAAGCGAGTTTAGGGAAGGGGGATTTGCTGT**TGGTCGGAAA**GA**GTGAAGG AG**GTATG	199
2. 6E/ 35	CGATATACAGAGGTC**GCTTTTA**CATATTTTGAGCCTTGAGAGCGTACTAGA**TCAATGT**AAGGTGG TTGAACTACTATGTTATGTGGATTGTTACGTATTTGGGAATGAAGGGTTGGGTGTGTAATCGGTT GGTGGTTACGCTGTGGCTTTGGCA**CTTACGGACT**TGTTAGGCG**TGGTGGGAAT**GGATCTGATAG GGTTCTTGGAGTGTGGCGATG	215
1. 11F/ 21	AAC**ACATTTA**CCAGCTTTATGGTCGC**TCTTTTC**GCTGCGGAGATCTTGTGTTTTTGGTAACGGGCT CACGTTTCAGGTAATGTGGTTAGTGTTTCTGTATTATCTCGCACCGGTCCCTAGGCCAAGACCGTC TTCT**GGGTCGGACA**ACTTATAAATAGGCTAGATTCCATGTACAACAAAGCATGATA**CTGGA**GGA **GA**GGGATG	204
4. 4D/ 25	TA**TCTGGTA**TGAATGCGT**TCATTTC**GGATGGGGGGGGACCAACGTACTAGTTCCAGCAATTTAAAT GTAGTAGTGTGTGCACCCAATTTCCTGCGACGGCAAGAGCTGATAGTCATAAACGATCAGAATT ATGTAGTTTCAGGTTGGGTATTGTCGACGCAAAC**AAGAGGGTAC**TCTGAGGC**TAGAAGGAGG**G GAGTGAGATG	202
2. 10H/ 38	TTTTTGCTTTATAGGATTGGTGCGGGGGAGTGCTCTA**TCTGTTC**TTTTGGGTTAAAGAGGTG**CCT TATC**GTTTTTTGGCAGAAGTGTTGATGATGTGGGTAGGTCTTGTGGTACAGATCCTGAAATTATA CAGTGGCGGATTTATTCAGTTTTGTACCG**CTGAGGGTGG**TTTAAATCAGCTTTC**GAGCG**GGAG**A** CTTATCATG	203
1. 11E/ 13	C**TCTGATA**GTCCGGTTGCTTATTACGACTATGCTGATTAGAGATTCGGAGTAGTTACTGCTATGT ATGTGAGGGTAG**TCATTTC**AGTGTCGAGAAGTGTGCACGATTTAGGTTTGATATTCATTCACCGG ACGTTTGGCGCGAATTGTGGAT**TTTACGGCCG**GGGTATCTGTAGGTTGACCGCT**TGTTG**GGAG**C** GCCCGAATG	203
1. 5D/ 33	ATCGTGATATCAGTGATTG**TCATTTT**GTTAGAGTCCTACATTAGGACTGTGATAGTCTTGATTGAT GGTAGTTTTTTG**TCTTGTC**TTGTGTAGTGTATGGTTACTTCGGATTTAATCCTTAGGTTTTGAAGT GCAGGGGAAGGTCAAGCTGCTGAATAATGAAACTAAA**GTGAGGGACC**TTTGTTGGTCT**TTAAG GGAG**TCCTCGGATGC	209
3. 11F/ 15	TGGTTATGTGTCGAAACTATGTGATTTTGCTTTTTAGGG**GCATTTA**TGGTAATAGTGGTCTTGTTC GGTCTGC**TCATTTT**GGGTTATGTAGATACGTTACCGTGGCTTTGTGTAATCATTGCCCCTAGTGTC CTTGATGCTATCGGCTTATGGGCGATCAGGGTTGTACCGGTAAATTGG**TGAAAGGAGT**GGCAT**A GGTT**GGAG**A**CTCGCAATG	214
3. 4F / 3	GATATAGTTCATTCGCGTGTTAAGAAAAATCGGAACA**ACTGTTA**AATTAGGATCAGTAAAAAACT GGTGCGGTTTAGTCTGCTGTAGTGTGCCGAGGGC**TCTTATA**TAGTACCGCCGATTCCGCGCTGAT AGCCCGTTTCTCAATTGATAATGGCTTTGTTTGAAT**GCCCTGGAGT**AGATGTGACAGC**CAGCAGG AGT**GTTATG	204
4. 1C/ 22	ATAATGATTGAAGTGGCTCAGTACTTG**TCAATTC**TAAACGGCCGGACATGGTATGGTTAGGGTTA ATAG**TCGTTTA**GCTTCATCCTCTATTACGGATGGGTGTTGAAAGGGGCTATAGGTTCGGGGTATT GTCTGTTGGGGGATTAGTTTAGATGGGTAATTGCATTGG**TTGTGGGCA**ATGTGGCCG**TGGACG GAGA**AAGAGG	204
4. 4H/ 43	TCCGGTTATTAATAGAGTGAGTCTTTTGTCAATATGGGTTATTGGACTGGAATCAATGTGGGGAG TTTATCGGATGCTCGGGCACTTGGAGTGTGAGTTAGCAAGTAGTTGTCACTTTGATCGGGTGCG GTTTTGTCAGCTCGGTCGTGGAAATTTTATTTTGATATGTCG**CATACATTGCT**ATCGGAGTC**TTAT GGAG**TCGTAGGATG	209

3. 5H/ 12	CTTTGTAGGCATTCTCGTTATTAAGATAAGAGTCCAGTCATTAATGTGCTCTTCGGTCTG**TCTTCA TTGCT**GTCTGTGGTTAATTAATATTCGCGCTATGTTGTTTCTTTTTTGTACCGGGTAGTGGAAGGC CTATCGTTTTGGACCGTATGAGTGCTGTATCAGCTTATAGCTGTAGTGTGTGGAGTTTAGGTAGG **GTG**AGGAGCGTTAGTATG	215
4. 6B/ 40	TGTTTACTCTCTAGTTGTTGGTAAGTAGTCTTTAATTCTAGTTTTTTATTCTTGCGTTAATGTTAGTT TTGTA**TAATCACACTT**TCGTGGAAGATCGCACGTTCGCCGTCCAGGAATAAGGACGTGTGGTCG GGTTTTGGTAGAATTGTGTGGATTCTTATGGTAGGATGTCAACTTTCAGATTGCGAG**TCCGGGA G**TCCAATGC	204
3. 6B/ 26	GTGCATCATTAGCCCGTTTCGGTCCTGCAGAGTATACTATGGATCGATCCGGGCCTGGGAGGTTT TTCATGTCGTCACTCGCGTAAGGTTTGCAGCGCATACAAGAGAAGTGACTTGGGTAATTATGAGT GAAAGGTCTATGGTCAGGTAGGGTTGTTGTATAGTGCTTATG**TGATCATAGTG**TGACTTTT**ATGA GGAG**CATTAGTATG	209
4. 2G/ 6	TGTTGGTATCCATGAGTTTACTGATGAAGCTTAGTACTTCCGGGTATTTCATGTTCACACAGAGTG TAGAGTTACTATGGCAGTTCTGTTATTTTGACGGTGTGAGGGCGTTTGTTATCGTGCTAGTAAAA TCGTCTAATGTTGTTGTGGAAGC**TTGACATAGTC**ATGTTAATCATGAAGGTAGTG**GCTGGGAG**T AGGATG	201
1. 12F/ 36	TAACATGCTAGGATGGTGAGAAGGAACTACCATTGTTATCGACCCGGGTTACCATCTTTAGGTAA TATC**CCGACATCGCT**TGATAAATAAATAGATCGGTGTTTGTCACTACTAGTGCCAGGTGCACTGT GTTGGTGTAAACCGGGTGTACTCCAAGAAATGGGTAGAGTCAATTTCCATATCCAAGAT**AGAGG GAG**TGAGATG	204
1. 6B/ 31	ACGTCACTAGATTGCCGTATGTTATTCATTTTTCTAGCAGGATAATGCAACCAATAAGTAAACCAA CCTTGTGGATGCTGAAAAGTGCAAATCTTG**CATGCACAACT**TTGCAGAGTGTTGACACATGCAAA GAGGGGAGTCGGATTTTGGAAATGTCGTCGAAACCACGCGCGATATATGGCGACAGG**ACGTGG AG**GGGGATG	203
3. 6A/ 19	CTTTACATGTTATGCTTTGCGTTGTTGAACAGTTGTTAATTGGGCGAACCGACTATGAGAATCTGT TTCTTTTGTAATTATTTTTACAAGGGCGTTTTCGGATAGATAAACTTTTGTATGGAGGTTTGGGCG A**CTTGCATACTT**ACGTAGGGTACTATTTGTAAACTTAGTAGACGTAATTATGCGAGATCGTGATA **TTTGGGAG**TTGTGTTATG	215
1. 4D/ 18	GATGTTACGTGCTTTTTTGCGGTAGTTGCCTAGTGATTCAGGGCAGTTTATATTGTGCTAACATCT TTCATTCCAATCTTGTAATGTACACATATTTTCCTGATTATGACGGAGGTCCT**GTAACACGGCT**GG GAGGTATGCGATTGGATGCGCGCCAATTTGAATGTTTTTTTGTTTTCGGGGGG**GATAGGAG**GTG TTAGATG	203
2. 1B/ 34	CGGTTTAGGTTTAGTAACAGTTTACGAAACTGTTGTTTGTGAGTTATAATGCGGACTGTCTTGAC GTAGTGAGTGCTAGATATTGGAACTGGCATGCTGGCCCAATGAACTGGATACACCTGAGG**CCGT AAGAGCT**GTCGGGTCGTATGGCTGAGGCGTTTAAGAGATGCGTTTTGTTGAGTAGTAAAG**CAAT GGAG**AGCTATG	204
2. 4F/ 23	TCTATTTTGACGCTTTACTTTCTTTATGGTCGAATTACGAAATGCCTTTTCACTTTAGCAATGACGG TACCAGTC**ACGTCATCGCG**TTTGTCATCTGAGCGTTTATTATCTATTCTCTGTTTTAAAGTACGAG GAATAAGTGGGGATAGACTGACGGAGACGGCACTACTGATGGGCCAGGAGAAGAGC**GTGAGG AG**AGATAAG	204
3. 10H/ 2	CTTGTTAGGATTGGTAGTATTGATGGAATAGTCACGTTGATCTGTTTAAACAAGCTGTAAGGGGC TAGCTTCGTCTTATTCGGTGGATGAACTAGTGTGCTGACGTGTGGGTTTACTTGGTTA**GAGTCAT AGCC**TGAGGCTTGGCGGCTAGAGCGTCTGGTGGAGTGTGGAGTGAGAGAGGTGAATGCAATG GATAA**GAATGGAG**TCAGGGTATG	215
1. 1G/ 16	GTTCATCCTGTCGTCGTTATAACGGTTATTCGTTCGCTTTTTTCCTGATAGTGTTAGAATAAATCTA **AAGACAAAGTT**TGGAAAACCGGTAATGTTTTCTATTATAGTCTTGACCACCTCAGGTTTGTTGAC GGTAGGAAGCCACCGTTTTTCAGGGAACTGTCGGATGGGGCTTGTGGGCAGGGACG**GGCGGG AG**ACGAATG	203
1. 2G/ 42	CTGAAATTGGGTAAATTTATGCAGTGTTTATTTAGGATTTAT**CTAACGTAGCT**CTTTTTGATCCCT GGGGGATATGATTATGAAGTTGATATGTGTACGAGGGTTATTTGATGTTATTTCCGGTTTTAGAA	215

	CAGTGCGTGTGGTGCGGGTATATGTATGTGATCAACGTTTCAAGGATGTATGAGGAGGGCACAC GC**CAGGGGAG**GGCTGAAATG	
4. 5F/ 4	AATTAGGGACGCTTGGGCGGTCGAAATCGTAGGACACAAGTGGCTTTATAACTCT**GTGTCATAG CT**CAGTTCGGGGTATGATCGTGGAGCAGCTAGCTGCGTGTCCTCCAGACAGTCAGAGGGGTGTA AGGTGCCATGTGGATGAGGATGGCTGCTCCTCGGGGCGGTAATGTATCTGAATGAATTCTT**GTG GGGAG**GTTTATG	204
3. 12 A/ 9	GCGTTGTTATATAGAGTCT**CAATCATAGGT**GCATGGTTCTGAAGTTGGAGTACTTATTTAGATCT ATCACCATTTTCGGGAGAGTTTTGGTTGCGGCACTGATTGTATGTTGGTGTGAAGATTCACTGTT TCTTTTAGGGACTGTGGGTAAGTAGTGCGGAGTGTTGTCGTTGGTAGGAGGGGTTGTTTGTTGT T**GGAGGGAG**TGGATG	209

3. 7 Comparison of sequence results from BROM and Improbizer

To conclude the comparisons of the low to medium to high expression in *V. natriegens* with the resulting expression levels the promoter and 5′UTR produced when transformed back into *E. coli*, the output from the sequence analysis of the plasmids in table 3.2 was compared. These promoter and 5′UTR sequences was interesting as one of the overall goals of the thesis was to analyse why the promoter and 5′UTR libraries behaved differently in the two organisms. These results are discussed on the final part of the discussion attempting to explain the variabilities observed in GFP expression levels.

The following text presents a detailed comparison of the results of the sequence analysis by BPROM in table 3.5 and by Improbizer in table 3.6. The promoter and 5′UTR **3.7 G** resulted in low fluorescence levels in both organisms. BPROM indicated a strong match at the -10 position and low at the -35 position. Improbizer located four motifs, with a slight overlap for the last strong motif. The promoter and 5′UTR **4. 1H** resulted in high fluorescence levels in both organisms, which was unexpected due to high scores. BPROM indicated a medium strong match at the -10 position and weak at the -35 position. Improbizer located two weak and two strong motifs, with an almost complete overlap for one of the weak motifs.

The promoter and 5′UTR **4. 1C** resulted in high expression in both organisms, which was surprising. BPROM indicated a medium strong match at the -10 position and weak at the -35 position. Improbizer located four motifs, with a slight overlap for the first motif. The promoter and 5′UTR **4. 4H** resulted in low florescence in *V. natriegens* and low to medium fluorescence in *E. coli*. BPROM indicated medium strong matches at both the -10 and the -35 positions. Improbizer located two strong motifs with zero overlaps.

The promoter and 5′UTR **3. 6B** resulted in medium fluorescence in *V. natriegens* and the highest measured fluorescence in *E. coli*, which was unexpected as BPROM indicated a medium strong match at the -10 position, and the -35 position was low. Improbizer located one weak and one strong motif with zero overlaps. The promoter and 5′UTR **1. 12F** resulted in low fluorescence in *V. natriegens* and medium low fluorescence in *E. coli*. BPROM indicated a high match at the -10 position, but a low match at the -35 position. Improbizer located two strong motifs with zero overlaps. The promoter and 5′UTR **1. 4D** resulted in medium fluorescence in *V. natriegens* and high fluorescence in *E. coli*. BPROM indicated a high match at the -10 position and a medium high match at the -35 position. Improbizer located two strong motifs with zero overlaps. The promoter and 5′UTR **2. 4F** resulted in

medium fluorescence in *V. natriegens* and high fluorescence in *E. coli*. BPROM indicated high matches for the -10 and the -35 position for both organisms. Improbizer located two strong motifs with zero overlaps. The promoter and 5′UTR **1. 2G** resulted in high fluorescence in *V. natriegens* and low fluorescence in *E. coli*. BPROM indicated a high score at the -10 position and the lowest result at the -35 position. Improbizer located two strong motifs with a partial overlap for the first motif.

Discussion

4.1 Establishing *V. natriegens* in the PhotoSynLab laboratory environment

Final protocol, growth conditions and antibiotic concentrations

V. natriegens is not a model organism, and before the promoter and 5′UTR could be established in the organism, a protocol for making cells competent and transformation with plasmid DNA had to be identified. This first challenge was overcome after a labour-intensive period of trial and error, resulting in the protocol in the result section 3.1.3. Although the protocol resulted in a transformation efficiency high enough to accomplish the goal of the thesis, there is room for improvement. Conditions which would be interesting to investigate further is the media composition, especially the salts added to the BHIN, and to find an optimum combination tending to the organism's osmotic requirements as osmotic stress naturally could affect the competency and transformation efficiency. Another main goal when deciding on a protocol, was to find a balance between time-efficiency (time spent preparing competent cells) and the amounts of competent cells produced for each batch. The final protocol was not the protocol which resulted in the highest amounts of competent cells, but the time spent creating each batch was halved. For other research purposes than this thesis which focused on a rather small-scale screening, a protocol focusing on producing a large amount of competent cells in one batch would probably be recommended.

When transforming competent cells with plasmid DNA, it was crucial to keep the cells chilled continuously before electroporation, and to recover the cells immediately after the pulse was administered. It would be interesting to test the effect of higher voltage administered during electroporation, as a voltage of up to 1800 V have been recommended (6). A protocol for chemically competent cells and electroporation (6) was tested by PhD candidate Maxime Fages-Lartaud (PhotoSynLab) with several different plasmids and selection markers which was unsuccessful. Lower transformation efficiency was reported for chemically competent cells (6). During the pre-screening, both temperatures of 30℃ and 37℃ were tested. There was no evident change in growth time, but GFP expression seemed to begin earlier with 37℃ as the incubator temperature. In a publication by Tschirhart *et. al* (2019) which performed extensive test of plasmid compatibility in *V. natriegens* it was observed that the highest plasmid copy numbers were found when growing *V. natriegens* at 37℃ compared to 30℃ (5). This fits well with the observations made during the experiments performed for this thesis

of detecting GFP expression at an earlier time-point. Several publications working with *V. natriegens* have reported use of incubation temperatures at f 37°C when introducing plasmid DNA (5-7, 11). This was expected as the organism's optimum temperature was reported to be 37°C (4). As for the antibiotic final concentrations utilized in liquid media, the chloramphenicol concentration was increased from 5 µg/ml to 12,5 µg/ml to increase the plasmid maintenance in the bacterium after testing the effect of higher concentrations after discovering the recommendations made by Weinstock *et. al* (6). Up to twice the concentration had been suggested (6). For solid media, 5 µg/ml was utilized. Other factors which could lead to higher transformation efficiency is the DNA concentrations added to the competent cells. Due to difficulties of establishing plasmids in *V. natriegens,* chloramphenicol was the only antibiotic which was thoroughly tested. However, during the testing of different plasmids, we did observe that the antibiotic gentamicin resulted in inconsistent results, also for untransformed cells (WT control) which supports findings of resistance up to 30 µg/ml (5). When testing use of Gentamicin as a possible selection marker, background growth of WT *V. natriegens* competent cells resulted in background growth at 30 µg/ml. Possible background growth was also observed for Kanamycin up top 250 µg/ml (30). As this was higher that the reported resistances from research groups working with *V. natriegens*, it was decided to exclude selection markers Gentamicin and Kanamycin from further testing (5, 30). Another observation made during the laboratory work worth mentioning as a possible source of error for these inconsistent results was that antibiotic stocks which had been stored for longer periods in time declined drastically in strength. This observation was made when receiving a new stock of chloramphenicol which, through several rounds of tolerance testing resulted in changing the recommended working concentration to half of the previously utilized concentration. This was not accounted for while testing Gentamicin and Kanamycin and could be a potential factor affecting the results of antibiotic tolerance testing in either laboratory performing the experiments.

Established use of a plasmid and consistent maintenance

One of the essential steps were the identification of a plasmid which would work as the backbone for the Golden Gate assembly amplification of the plasmids with promoter and 5′UTR. The plasmids containing the promoter and 5′UTR also had to function in *V. natriegens,* with both a suitable antibiotic resistance gene and a reporter gene for the quantification of promoter and 5′UTR productivity. Although several recent publications have reported the successful maintenance of plasmids with different origins of replications (5, 6), the plasmid pACYC184 with the OriT p15A was the only plasmid we were able to get consistent results with after several round of experiments. Several members of the lab-group attempted to establish use of plasmids in the organisms without success. The reason for the failure to maintain plasmids with different OriT's are unclear and unexpected as several publications claim to have successful replication of different plasmids in *V. natriegens.* As shown in table 3.1 of the results section, several plasmids with identical OriT's but different resistance genes were tested, in different strains and with different protocols where transformations resulted in successful maintenance of the positive control plasmid (pACYC184). The plasmid PJQ200 also had the low copy OriT p15A. One of the reasons for unsuccessful use of this plasmid was that the resistance gene was Gentamycin, as previously described as not suitable for use with *V. natriegens,* in our experience. More through testing of OriT's and resistance genes would be recommended to obtain the possibility of establishing synthetic genetic circuits (6).

Throughout testing different plasmids, it was expected to identify a cause of the unsuccessful maintenance of other plasmids than pACYC184. After several round of laborious and intensive testing of the different protocols and growth conditions detailed in the results section 3.1, it was decided to discontinue testing due to time restrictions. One of the major reasons for this was that it was unclear how time consuming the process of establishing and screening the promoter and 5′UTR libraries would be in *V. natriegens* as the organisms was not established in our laboratory. This put a sense of urgency to finding a compatible plasmid to proceed with. The only possible reason for the faliure to reproduce the findings from several publications stating successful transformation of multiple plasmids (5-7, 11), was that the plasmids could impose a metabolic strain too large due to higher copy numbers. This was probably observed for the first promoter and 5′UTR with GFP as a reporter gene, which lead to no growth in *V. natriegens* after transformation during plasmid compatibility testing of the

plasmid pACYC-GFP. When the promoter and 5′UTR fragment was changed to contain the superfolder-GFP by PhD Liza Tietze, the transformation was successful.

A highly relevant observation made by Weinstock *et. al* (2017) is that transformation efficiencies in *V. natriegens* was markedly reduced when the plasmid utilized was isolated from *E. coli* and not sourced from *V. natriegens*. The research group also utilized the plasmid pACYC184 and a probable cause for this effect was differences in plasmid modifications by the host organisms as methylation patterns (6, supplementary materials). They were able to prove introduction of plasmids containing several different OriT's and selection markers Throughout the testing, plasmids were isolated from *E. coli* before transforming into *V. natriegens*. This was not discovered as a possible source of error until after the end of laboratory experiments. This would be worthwhile exploring working with *V. natriegens* as a possible cause for inconsistent transformation efficiency results (63). For this thesis, it was necessary to use plasmids sourced from *E. coli* as Golden Gate Assembly by electroporation would result in poor transformation efficiency due to the relaxed nature of the plasmids. However, when miniprepping the plasmid libraries from *E. coli* the plasmids were supercoiled which made it possible to obtain good enough transformation efficiency in *V. natriegens* by electroporation. If it was possible to make chemically competent *V. natriegens* cells, this challenge would most likely be bypassed by eliminating the need to transform the Golden Gate reactions into *E. coli* first.

Perhaps the most important an rewarding reason for continuing the work with *V. natriegens* even though there was multiple time-consuming challenges which is a challenge on its own due to the limited time designated for a master thesis, was the rapid growth time of the organism. Unlike *E. coli*, it was possible to incubate 96-well plates with colonies the day of screening, and to measure GFP expression after only 2-3 hours of growth. The rapid growth time was also very valuable as plasmids at high enough concentrations for use in transformations could be obtained from glycerol stocks after just 5-6 hour of growth. An obvious advantage is the possibility to save a day of laboratory work which normally could be lost if one discovers the day of the planned experiment that a plasmid stock is insufficient. Additionally, colonies would appear the same night as a transformation were performed, again halving the needed time required to obtain a plasmid stock in *V. natriegens*.

So far, both challenges and advantages working with *V. natriegens* have been highlighted. In addition to the rapid growth, it was also possible to isolate plasmids from *V. natriegens* in concentrations high enough for transformation and with sufficient purity for sequencing

directly after isolation. These traits, confirmed by the results in this thesis, are crucial for the assertion that *V. natriegens* is a potential and suitable replacement for *E. coli*. However, there is still multiple frequently utilized methods in molecular biology which must be thoroughly tested before confirming this statement. This includes cloning plasmids based on restriction, targeted gene editing by CRISPRi or CRISPR/Cas9 based methods, homologous recombination and standard cloning methods as Gibson cloning and Golden Gate Assembly. Even though several of these methods was confirmed by the research groups working with *V. natriegens*, there is still work needed to establish these methods for the PhotoSynLab laboratory group (5-8, 11, 64).

As mentioned in the results section 3.3, another discovery was that GFP expression had to be measured the day after transformation, otherwise the probability of declined or non-existing fluorescence could occur during the screening. As the colonies grew in height during storage, a probable explanation for loss of GFP-expression was that as the clones lost contact with the selective media. If the bacteria grew without contact with the antibiotic, the need to maintain the plasmids were eliminated. The pACYC-sfGFP 200N SD plasmids had the low copy number OriT p15A, and storage could either have resulted in a reduction of copy number or loss of the plasmids entirely. Weinstock *et. al* (2016) reported that storage by refrigeration at 4℃ decreased viability of colonies growing on agar plates. In a growth experiment, colonies were stored for 20 days at 4℃ and at room temperature. Upon re-streaking the colonies on fresh agar plates, no growth was observed from the refrigerator colonies, against high viability from the colonies stored at room temperature (6). After testing this in the laboratory, we stored all *V. natriegens* colonies on agar plates in room temperatures and the challenges with viability from colonies was reduced substantially. This trait has been reported for several Vibrio species (6) which has been attributed to loss of catalase activity leading to increased sensitivity to H_2O_2 and decreased viability (65, 66).

A central question posed in this thesis, is if the challenge of obtaining predictable gene expression levels could be overcome by use of artificial promoter and 5′UTR, by creating a vector which could result in predictable and consistent expression of the gene of interest without the need to identify a compatible promoter for non-model organisms. The focus of the discussion will move towards an attempt to answer just that.

4.2 Assembly of pACYC-sfGFP 200N SD with Golden Gate Cloning

The goal of the thesis was to establish and characterize novel artificial promoter and 5′UTR sequences which would recruit the host's own transcription and translation machinery by the GeneEE method (39). The method had previously been proven functional in seven microorganisms, both gram-positive and gram-negative bacteria and the yeast *Saccharomyces cerevisiae.* The promoter and 5′UTR sequences utilized in this thesis consists of 200 random nucleotides followed by a SD-sequence and 7 bp separating the SD-sequence from the start codon, as indicated in **figure 1.3** in the introduction. An early screening performed by PhD candidate and co-supervisor Lisa Tietze with and without including the SD-sequence in the 200N promoter and 5′UTR indicated that the SD-sequence could be required to obtain sufficient amounts of *V. natriegens* clones expressing GFP. Additionally, it was believed that several other organisms could be dependent on a SD-sequence within the 5′UTR for translation (39).

The cloning was performed by the standard method detailed in the material and methods section 2.1.5 which successfully produced a library size of 10,000 clones in *E. coli*. The library was later transformed into *V. natriegens* by the electroporation protocol detailed in section 3.1.3 of the results. Throughout the thesis the pre-screening was defined as the initial screenings of the pACYC-sfGFP 200N SD transformants to identify plasmids with functional promoter and 5′UTR in *V. natriegens.* The actual screening was confirming fluorescence by screening in triplicates, and the time-dependent screening of promoters and 5′UTR resulting in variable GFP expression levels in *E. coli* and *V. natriegens.* Verification of successful transformation in *V. natriegens* was challenging at first. The first indication of functional promoters was fluorescent clones on agar plates. In *E. coli*, colonies with functional promoters were either visibly green, or fluorescent under UV light. In *V. natriegens* the fluorescence intensity was remarkably lower, and colonies were usually neither visibly green nor could be observed as fluorescent under UV light. Therefore, after a transformation with the promoter and 5′UTR libraries that resulted in colonies it was unclear if the clones contained plasmids with functional promoter and 5′UTR as the cat promoter in front if the chloramphenicol resistance gene was kept.

During the pre-screening, there was several factors that had to be accounted for. Firstly, one of the main concerns during the pre-screens of the pACYC-sfGFP 200N SD transformants after the Golden Gate Assembly was that the clones expressing GFP would be a result of

contamination with the WT plasmid pACYC-sfGFP with the native mdh promoter (figure 2.2 in the material and methods section). To minimize the risk of contamination and screening of the wrong clones, the backbone PCR was treated with DpnI as detailed in the material and methods section which worked well enough to continue the pre-screening. Secondly, a challenge with Golden Gate Assembly is that a portion of the transformants always will contain incorrectly assembled plasmids or in some cases, only the backbone fragment. Colony PCR was utilized as a pre-confirmation that the cloning was successful and to continue the pre- screening until sequence results arrived as the final confirmation of correctly assembled and unique promoter and 5′UTR sequences. When confirming expected insert lengths by colony PCR, there was a need to create a PCR protocol for *V. natriegens* as the standard protocol used by the laboratory group did not produce visible bands by agarose gel electrophoresis. The PCR protocol required an additional step of freeze-thawing of the master-mix with sample material added. An additional source of error was the discovery that the first round of screening had to be discarded as the results consisted of 53 of 57 plasmids with the exact same promoter and 5′UTR sequences. This was not discovered until after the sequencing results arrived, and after troubleshooting the only explanation was poorly mixed promoter and 5′UTR libraries when transforming the plasmids into *V. natriegens.*

4.2.1 Creating a negative control plasmid for GFP-expression

A negative control plasmid was created to eliminate the possibility that GFP was a result of a sequence upstream of the sfGFP gene being utilized as a promoter instead of the artificial promoter and 5′UTR. As detailed in the results, the negative control backbone plasmid did not result in any fluorescence even though the sfGFP gene was maintained (as displayed in results section 3.4 by confocal laser scanning microscopy). This confirmed that it was in fact functional promoter and 5′UTR driving GFP expression in the pACYC-sfGFP200N SD plasmid.

4.3 Confocal microscopy of GFP expression in *V. natriegens*

The promoter and 5′UTR resulted in variable GFP expression levels (from low to medium to high), and confocal laser scanning microscopy was performed as detailed in section 3.4 of the results for visualization. Additionally, it was confirmed that the pACYC184 was suitable as a negative control plasmid and that the wildtype *V. natriegens* cells did not result in any measurable autofluorescence in **figure 3.20**-**3.21**. This strengthens the validity of the methodology utilized for the time dependent GFP expression level measurements.

The fluorescence intensities are determined by how strong the fluorescence is. For the populations of cells in **figures 3.17 – 3.19** cells without fluorescence can be observed when studying the overlays of phase contrast and GFP expression, even though the populations have been transformed with plasmid DNA. This is especially visible for the positive control population in **figure 3.19**. A possible explanation for this is due to the low copy number and probable reduction in this, which could potentially result in larger heterogeneity in a population of cells, with only a proportion of cells maintaining a plasmid. As mentioned in the results section 3.3 it was observed that *V. natriegens* lost fluorescence completely after only days of storage.

The methods which are confirmed as established and working (milestones) in the laboratory environment of the PhotoSynLab group is presented in the workflow below. The methods which have been detailed so far in the discussions are indicated by arrowheads pointing upwards, and the methods which will be detailed in the following sections are indicated by arrowheads pointing downward.

Milestone	**Discussed**
Growth conditions (temperature and incubation times) (*V. natriegens*)	↑
Media composition and antibiotics (*V. natriegens*)	↑
Transformation protocol (*V. natriegens*)	↑
Plasmid with reporter gene and compatibility (*V. natriegens*)	↑
Colony PCR protocol (*V. natriegens*)	↑
DNA sequencing (Eurofins Genomics) (*V. natriegens*)	↑
Confirming functionality of promoter and 5′UTR sequences (*V. natriegens*)	↑
Screening in triplicates to confirm fluorescence (*V. natriegens*)	↓
Time-dependent screening of plasmid identities (*E. coli* and *V. natriegens*)	↓
Sequence analysis (*E. coli* and *V. natriegens*)	↓

4.4 Time-dependent expression in *V. natriegens* and *E. coli*

Experimental setup screening for successful mutants

The goal of the time-dependent GFP-expression analysis was to identify at which growth phases promoter and 5′UTR activity and GFP-expression began. Before the time-dependent expression measurements of *V. natriegens* and *E. coli* could be performed, a pre-screen for successful promoter and 5′UTR sequences resulting in GFP expression were performed in *V. natriegens*. The pre-screening was performed in triplicates of each plasmid inoculation to ensure that the clones in fact did fluorescence and to check for experimental errors during the screen. If not all three replicates were fluorescent, the construct would be eliminated from further screening. All of the data depicted in the result section are from calculated averages of triplicates. Another important factor for measurement in triplicates is the ability to assess replicability. What is indicated by replicability is that if the plasmids were to be transformed into *V. natriegens* and re-screened during similar experimental conditions utilizing the same instruments, the GFP expression levels should not deviate far from the measurements presented in this thesis. Additionally, the instruments could be calibrated for increased standardization (67).

The need to distinguish clearly between reproducibility and replicability and how these terms are commonly misused is perhaps one of the factors of the difficulty of reproducibility synthetic biologists are encountering, especially for non-model organisms. This will be further outlined in the concluding part of the discussion (68, 69). As mentioned in section 4.2.1, a negative control plasmid was created to ensure that fluorescence would only originate from correctly assembled pACYC-sfGFP 200N SD plasmid libraries. For the time-dependent screening, the plasmid pACYC184 was used for as a negative control for fluorescence and the plasmid pACYC-sfGFP as a positive control. Each 96-well plate contained at least negative controls. A blank measurement of BHIN + V2 and chloramphenicol or LB and chloramphenicol depending on the organism was measured to remove the background from the growth media. The negative control was crucial to assess where fluorescence expression began, especially in *V. natriegens* as the measurements resulting from certain plasmids with correctly assembled promoters were close to background. These constructs would potentially have been omitted from the time-dependent screening, which would have led to the loss of some high-expressing constructs in *E. coli*, detailed in the following sections of the discussion. The GFP expression from the positive control plasmid pACYC-sfGFP was significantly higher in *E. coli* compared to *V. natriegens.* For *E. coli*, GFP expression varied

from *coli*, GFP expression levels were up to ~4000, while in *V. natriegens* GFP expression levels were up to ~900. The lower GFP expression levels observed in *V. natriegens* from the artificial promoter and 5′UTR in the plasmid libraries pACYC-sfGFP 200N SD were not unique. This tendency will be further discussed, as plasmid compatibility in the two organisms could be a potential factor.

4.4.1 Promoter activity at different growth phases

In order to identify at which growth phases promoter and 5′UTR activity and GFP-expression began, expression was measured from 30 minutes before expression was expected to occur**.** Transcription of genes are affected by which growth phase a bacterial culture is in, and the proteome changes with shifts between growth phases. While certain genes are turned on during the exponential phase and off towards the stationary phase, other genes are first turned on during the stationary phase. The genes activated during the stationary phase achieved in laboratory conditions are often stress-related due to decreased nutrient availability. By measuring the GFP expression and OD_{600} over a period of 6 hours, it was possible to identify at which growth phase promoter activity began (70, 71, 72). The exponential growth phase for *E. coli* continues until an OD_{600} of between 0.6 and 1.0, marking the beginning of the stationary growth phase. However, a recent publication claims that the exponential growth phase ends already around an OD_{600} of 0.3 (72).

Measurement were conducted every 25-minutes for *E. coli* to roughly hit each doubling-time, but only for each 20-minutes for *V. natriegens* which would be every 1,5 doubling times. The measurements were continued for approximately 5 hours. Measurements could have been performed more often too hit each estimated doubling-time for *V. natriegens* but was chosen against to avoid disrupting the growth as much as possible. When moving the plates from the shaking incubator to the Tekan instrument there was a decrease in temperature from 37℃ to approximately 22℃. Additionally, as the measuring was performed in a non-sterile environment the measurements was not performed more often for *V. natriegens* due to the increased risk of contamination. A variation from different promoter and 5′UTR, with activity beginning at different time-points was expected. After earlier measurements not included in the results section, the majority of the promoter and 5′UTR was expected to produce GFP during the exponential phase around OD_{600} of 0.5, or later near the end of the exponential phase.

The main reason for observing time dependent GFP expression was to identify when the promoter and 5′UTR's were activated. When attempting to regulate gene expression, the desire is to identify or design promoters (with or without including the 5′UTR) that produce proteins at a predictable time point. This is necessary when designing synthetic circuits, and for industrial scale production where avoiding overproduction before cells have reached the stationary phase are essential. Studying expression at different growth phases relevant due to the established fact that for engineered promoters only one specific sigma factor will recognize the promoter. With artificial regulatory sequences as the promoter and 5′UTR utilized in this thesis, this specificity could potentially not apply as several motifs potentially could be recognized as sigma factors. Additionally, since the activity of the promoter and 5′UTR are observed in two different organisms, a potential result could be that the variable functionality observed are somehow related to different identification of sigma factors enabling transcription initiation. Before continuing this discussion, the promoter activity observed in *E. coli* and *V. natriegens* will be detailed below (73).

4.4.2 Promoter activity in *E. coli*

The results of the measurements are displayed in figures 3.9-3.11 for *E. coli*, and figures 3.12-3.14 in *V. natriegens*. The measurements from the negative control (pACYC184 in *E. coli* or *V. natriegens*) are included in each graph. For *E. coli*, all plasmids expressed above the negative control at the first measurement, when the OD_{600} is on average ~0.380 in the early exponential phase. An increase in GFP expression can be first observed from the plasmids EC3. 5D and EC4. 4B after 25 minutes when the OD_{600} is on average ~0.460. This could indicate that these promoter and 5′UTR's attract transcription factors during the exponential phase. After 75 minutes and an OD_{600} average of ~0.550, the majority of the curves indicate an increase of promoter activity. This could indicate that the majority of the promoter and 5′UTR's identified during the screening are activating during the late exponential to stationary phase. After this time-point, a steady increase in GFP expression is evident for most plasmids indicated by steep curves with a logarithmic pattern until around 200 minutes when the OD_{600} have reached and average of ~0.800. After 20 minutes, the curves visibly flatten out indicating that a plateau will be reached. The majority of the promoter and 5′UTR's discussed so far apparently prioritize growth before producing GFP. As protein production is demanding for the cells and consumes nutrients which would otherwise contribute to the growth metabolism, this could be seen as a successful strategy. If the cells contained promoter and 5′UTR's resulting in large amounts of GFP produced before reaching mature cell sizes, a

negative energy balance would occur. This could result in the cells dying instead of thriving and continuously produce GFP as long as nutrients were available (74).

There are some potential outliers from the general trends. In figure 3.9 the plasmids EC3. 4B, EC1. 1G, EC3. 5D and EC1. 6B hits a plateau already around 175 minutes. In figure 3.10 the plasmids EC3. 5H and EC4. 4H also hits a plateau around 175 minutes. The promoter and 5′UTR's in these clones potentially could be poorer producers, which is indicated by lower normalized endpoint GFP expression levels in table 3.3 of the results. In figure 3.11 the highest expressing plasmid (EC3. 6B) had an GFP expression at 671 at the first measurement. The OD_{600} measurement does not indicate a large variation from the density measured in other wells. This could indicate a promoter and 5′UTR which is active from a very early growth phase. After 125 minutes, the curve indicates some decrease in exponentiality, earlier than the majority of the high-expressing promoter and 5′UTR's in figure 3.11 (72, 75).

4.4.3 Promoter activity in *V. natriegens*

In figure 3.12 – 3.14 the curves produced by the time-dependent screening in *V. natriegens* at first sight potentially looks to have more variation in promoter activity. However, this is most likely an artefact resulting from the overall lower GFP expression levels which highlights small differences in expression between measurements. For *V. natriegens*, all plasmids except VN4. 4H (figure 3.12) expressed above the negative control after ~65 minutes of measurements, when the OD_{600} ~0.500.

The GFP expression seem to increase from most plasmids until 100 minutes for most plasmids at an OD_{500} of ~0.600. This indicates that the majority of the constructs could be recruiting transcription factors during the exponential phase (72, 75). The curves in figure 3.12 and 3.13 cannot be described as having a logarithmic shape between those timepoints, but the shape is clearly steeper indicating an increase in fluorescence. This could indicate that the fluorescence continued to increase when the measurements was concluded. For figure 3.14 detailing the high GFP expressing plasmids, the two highest expressing plasmids VN1. 2G and VN1. 4D have a logarithmic shaped curve. This was the shape that was expected from the fluorescence measurement due to the maturation time required to form the GFP chromophore, but probably due to the low increase in expression levels most promoter and 5′UTR's in *V. natriegens* did not obtain this shape (76). Compared to *E. coli*, the GFP expression in *V. natriegens* for the majority of the plasmids reach a plateau after much shorter measuring time. This occurs for most plasmids after approximately 130 minutes of measuring

time, when the OD_{600} is ~0.650, still in the exponential growth phase. This was unexpectedly early. By studying the shape of the curves at the final measuring point the majority of the curves could still be increasing. During measurements, there was an observable decrease in the speed development of the GFP expression after this timepoint. It could possibly be that producing the GFP protein are a heavier load for *V. natriegens* than *E. coli*, or that the temperature fluctuations during the measurements affect the organism more. It would be interesting to re-screen the promoter and 5′UTR's in *V. natriegens* for longer periods of time in a closed system to avoid potential contamination and temperature fluctuations. Potentially, by studying the apparently low producing promoter and 5′UTR's in *V. natriegens*, they could turn out to be activated later during the stationary phase which was not reached during these measurements. There are several potential outliers displaying somewhat different promoter activities. In figure 3.12 the plasmid VN4. 4H reaches the highest GFP expression after 200 minutes, when the OD_{600} is ~0.740. *Similarly*, the plasmid VN4.1H increases marginally later than the majority of the constructs and continues to increase at the end of the measurements when the OD_{600} is 0.760. This could potentially be a promoter and 5′UTR which achieve higher protein production during the late exponential phase or stationary phase, as the expression apparently still was increasing when measurements was concluded. In figure 3.13 the plasmid VN3. 4F declines from 93 until 83 after 160 minutes, the most drastic declined in expression observed for *V. natriegens.* This could indicate a promoter and 5′UTR which are active in the early exponential phase (72, 75).

When assessing at which growth phase the inoculations are in, it is important to mention that the OD_{600} measurements were not calibrated for *V. natriegens*. As the light scattering is affected by a multitude of factors including cell size and shape, optical density measurements should be considered with caution. To achieve accurate density measurements and determinations of growth phases, colony forming units from multiple dilution series, especially as *V. natriegens* is a smaller bacterium than *E. coli* and differences in cell size particularly affects the cell concentrations. This ultimately means that the conclusions drawn in the discussion on which growth phases *V. natriegens* are in during the measurements of GFP expression are incorrect. Until a growth experiment is performed and a standard curve are created to calibrate the optical density measurements for *V. natriegens* it will continue to be a highly probable source of error. However, no such calibrations have been standardized yet for this organism (77).

As several publications utilizing OD_{600} when characterizing *V. natriegens* growth and metabolism without mention performing calibrations and utilizes the same cell dry weight as for *E. coli* calibration was decided against due to time restrictions (7, 78). Another possible source of error is that the measurements of time dependent GFP expression were performed differently for *E. coli*. Due to longer doubling times, the 96-well plates were inoculated the night before measurement and had to be diluted the morning after as the cells had reached the stationary phase. As the dilution were large enough, the OD600 inclined that the cells were again in the exponential phase, and measurements could be initiated after an additional 2 hours of growth. This was supported by the fact that the GFP fluorescence began after the expected cell density was reached. If the cells in fact were still in lag-phase, a different pattern of expression was expected from the screening. Due to a discarded first round of time-dependent expression, the promotor activity patterns were more predictable.

Although the standard deviations seen in the results from the time-dependent screening are very large for some samples, I made the decision not to discard the measurements. This was mainly due to the methodology utilized for the experiments. Large variations were expected as the micro-environment in the wells were affected by a number of conditions. One of them which can explain some of the deviation within the triplicates was that during the inoculation, colonies was picked manually with toothpicks. As the colonies were in the form of frozen glycerol stock, the number of cells resulting from each pick were expected to vary. Additionally, even though the sample volume was mixed well when added to the glycerol, several factors could have led to variations within each sample. The original individual liquid samples that the plasmids were grown up from was expected to have different densities, as they were grown up on separate days which most likely caused some variation in growth time due to different inoculation times overnight. Again, when picking colonies to inoculate these samples the number of cells originally present would be variable. As each sample volume and the volume of glycerol was identical when preparing each glycerol stock, the densities would remain variable. As detailed earlier in the discussion a probable explanation for loss of GFP expression during storage was due to the low copy number of the pACYC-sfGFP 200N SD plasmid. Low copy numbers have been reported to result in occurrences of subpopulations which are entirely without plasmids or cause larger heterogeneity in cell-to-cell copy numbers producing variability in reporter gene expression (65). This could indicate that subpopulations of different compositions were growing in the wells, contributing to large standard deviations.

Another possible source of variation between the triplicates could be poor homogenization in the stock. Even though one attempts to avoid this by keeping the glycerol stocks frozen, the upper layer would inevitably thaw to a degree when inoculating the well, and the stocks were used for inoculations several times. This would again affect when each well reached an OD_{600} high enough to begin expressing GFP, which would lead to a following error until the conclusion of the measurements. As some of the variability of promoter strength and activity could be a result of some colonies reaching a higher OD_{600} earlier than others, the values used to compare endpoint GFP expression levels were normalized with OD_{600} measurements when comparing GFP expression in *E. coli* and *V. natriegens* in the following section.

4.5 Comparison of GFP expression levels in *V. natriegens* and *E. coli*

The measurements detailed in this section were the last of the measurements from the time dependent screening, when the GFP expression and promoter activity was highest. By sequencing the resulting promoters, identifications of motifs and consensus sequences could possibly help to understand the importance of these in promoter 5′UTR and sequences. This was done by measuring the promoter and 5′UTR activity as a function of GFP expression. If multiple promoter and 5′UTR sequence were high expressing in either only *E. coli* or *V. natriegens*, it would be interesting to find consistent proof for which motifs that could be important for expression in which organism. This could potentially lead to new insights on the transcription and translational machinery of previously uncharacterized organisms where the genome is not completely described. One of the central questions in today's synthetic biology work, is the possibility of a shift in focus from the need to characterize an organism before establishing use of genetic tools in a non-model organism, or if synthetic or most likely artificial promoters could bypass this need.

In table 3.2 and figure 3.11 of the results section 3.3.2, normalized endpoint GFP expression levels in the range from low to high in *V. natriegens* are compared with measurements of the identical plasmid in *E. coli*. The interesting finding by this comparison is that plasmids that led to some of the lowest GFP expression in *V. natriegens* (VN1. 12F, VN3. 7G and VN4. 4H) resulted in low to medium expression levels in *E. coli*. When comparing the plasmids which resulted in medium GFP expression levels in *V. natriegens* (VN1. 4D, VN2. 4F and VN3. 6B) this resulted in some of the highest measurements in *E. coli*. Although two of the highest GFP expression levels in *V. natriegens* (VN4. 1 C and VN4. 1H) also resulted in high measurements in *E. coli*, this was not true for the highest measured plasmid in *V. natriegens*, VN1. 2G (183) which resulted in a low-medium measurement in *E. coli* (365). The trends so far discussed fit well with expected functionality, as earlier attempts of comparing promoter functionality in *E. coli* and *V. natriegens* resulted in high variability between the two organisms. Additionally, several promoters consistently resulted in lower response in *V. natriegens* (5). However, in this case the promoters were optimized for function in the model organisms *E. coli* (40, 37). The resulting GFP expression levels so far discussed for this thesis does however point towards lesser functionality in *E. coli* compared to *V. natriegens* and vice versa. This points towards an inconsistency in both organisms. As an attempt to deduce the cause for these inconsistencies, the normalized endpoint GFP expression levels in the results

section 3.3.2 will be further discussed in section 4.6, together with the sequence analysis results from promoter and 5′UTR.

4.6 Sequence analysis of promoter and 5′UTR sequences

In this section, the resulting promoter and 5′UTR sequences will be characterized by analysing the DNA sequences and comparing fluorescence levels of the green fluorescence protein from *E. coli* and *V. natrigens*. The variable functionality of the promoter and 5′UTR in the two organisms was presented in section 4.5 of the discussion, with consistently lower GFP expression levels in *V. natriegens*. Continuing the discussion in section 4.4, by studying **table 3.3** and **figures 3.12 – 3.14** of the normalized endpoint GFP expression levels in the results section 3.3.2 it is evident that the promoter and 5′UTR sequences overall resulted in consistently lower expression levels in *V. natriegens*. While the highest GFP expression measured in *E. coli* was 2219 from EC3. 6B, it was 183 in VN1. 2G. For comparison, only four plasmids in *E. coli* resulted in lower GFP expression than 183 (EC3. 7G, EC3. 4B, EC3. 4F and EC1. 6B).

To explain what potentially cause the variabilities observed during the time dependent screening, the insert promoter and 5′UTR sequences were analysed in the online software BPROM and Improbizer. The results from the BROM analysis are displayed in table 3.4, indicating the position of possible -10 and -35 elements, the position of the TSS, total promoter length and the SD-sequences. The results from the Improbizer analysis are displayed in tables 3.5-3.9. The results in tables 3.5-3.8 indicates possible conserved nucleotide positions of the localized motifs, while tables 3.9 display possible identified consensus sequences. The functionality of the searches conducted by the tools is explained in the results, for BPROM in section 3.5 and Improbizer in section 3.6.

The output from the sequence analysis indicated the location of several consensus sequences in all promoter and 5′UTR, as well as complete SD-sequences and start codons (with few exceptions). Overall, the motifs detected by Improbizer it was indicated that the -10 position was G-rich, while the -35 position was A/T rich which fits well with the actual consensus sequences for *E. coli* which are TTGACA (-35)/ TATAAT (-10) (19-22). *V. natriegens* share consensus with *E. coli* at these positions which should lead to favourable conditions for binding of the RNAP for both organisms.

The results from the GFP expression endpoint comparisons of low to medium to high GFP expression levels in *V. natriegens* with GFP expression levels from the identical promoter and 5′UTR sequences in *E. coli* indicated high variability between the two organisms. Exploring

the results from the sequence analysis in tables 3.4-3.9, leads to several interesting observations highlighting why these comparisons are essential. In the last section of the results, the strength of the matches to consensus sequences from BPROM were evaluated with the scores from Improbizer, as well as any overlapping results from the two tools. To conclude the comparisons of the low to medium to high expression in *V. natriegens* displayed in **figure 3.13** the resulting expression levels the promoter and 5′UTR produced when transformed back into *E. coli* were evaluated together with the sequence analysis results. These promoter and 5′UTR sequences were interesting as one of the overall goals of the thesis was to analyse how the promoter and 5′UTR libraries behaved differently in the two organisms. The experience during the laboratory experiments fit well with the expected variations detailed by reports from several of the publications mentioned earlier in the discussion on the variable functionality of genetic tools in *E. coli* vs. *V. natriegens* or non-model organisms overall (5, 73). Table 4.1 below presents the findings detailed in the final results section. The table is simplified to present the overall conclusions drawn after studying the results of the sequence analysis by BPROM and Improbizer. The score levels described for Improbizer are determined by the overall strength indicated by the colour codes in table 3.9 in the results. Per example, if several motifs are identified of medium and high strength, the overall score is indicated as high.

Table 4.1: Simplified table revisiting the results detailed in results section 3. 7, comparing results from the sequence analysis by BPROM and Improbizer. The comparison was also done to discover any overlapping results from the two tools, which could help strengthen the validity of the identified motifs. The plasmids detailed are the low-medium-high GFP expressing promoter and 5′UTR sequences in *V. natriegens* and the resulting GFP expression levels when transformed back into *E. coli*. */***The second lowest and the lowest match identified by BPROM for the -35 position. **The highest recorded GFP expression level measured for *E. coli*.

Plasmid identity in *V. natriegens and E. coli*	**Score level by BRPOM -10/ -35**	**Overall score level of motifs by Improbizer indicated by colour codes**	**Actual measured GFP expression level (low-medium-high)**
VN1 12F	High/ low	Strong	**Low**
EC1 12F			Medium
VN3 7G	High/ Low	Strong	**Low**
EC3 7G			Low
VN4 4H	Medium / medium	Strong	**Low**
EC4 4H			Low
VN1 4D	High/ medium	Strong	**Medium**
EC1 4D			High
VN2 4F	High/ high	Strong	**Medium**
EC2 4F			High
VN3 6B	Medium/ low*	Medium	**Medium**
EC3 6B			High**
VN1 4C	Medium/ low	Strong	**High**
EC1 4C			High
VN1 2G	High/ low***	Strong	**High**
EC1 2G			Low
VN4 1H	Medium/ low	Medium	**High**
EC4 1H			High

Only a few partial overlaps were detected, indicating that the tools evaluated the sequences differently. When studying table 4.1, it is evident that there were large discrepancies between the predicted sequence strength by the two tools and the actual resulting GFP expression levels. The plasmid identities chosen were interesting as they resulted in inconsistent expression levels in the two organisms. The promoter and 5′UTR sequences **1. 12F**, **3. 7G** and **4. 4H** were overall scored as having strong consensus sequences by both BROM and Improbizer but did not produce high fluorescence levels. This was unexpected at the least for *E. coli* as a model organism which the tools were optimized for. The promoter and 5′UTR in the plasmid **3. 6B** which resulted in the highest measured fluorescence level in *E. coli* also had the second lowest score for the -35 position, and an overall medium score for consensus sequences, and the promoter and 5′UTR **1. 2G** which resulted in the highest fluorescence level in *V. natriegens*, but low in *E. coli* had the lowest score for the -35 position.

Additionally, promoter and 5′UTR **4. 1H** which resulted in very high fluorescence levels in both organisms was only scored as medium/ low by both tools. The expectation was that the tools would hopefully locate indications of strong motifs for the high producers and vice versa. The output from BPROM and Improbizer revealed that there are inconsistencies which are not explainable by the sequence analysis performed in this thesis. Improbizer consequently located a strong match flanking the SD-sequence, which was expected and to a degree confirmed that the tool actually was able to detect probable consensus sequences. It is possible that the additional number of motifs detected by Improbizer could be related to the -40 and -60 positions upstream from the TSS as motifs were often were located more upstream than by BPROM, where A/T rich sequences have been identified in strong promoters (24). This also holds true for the -40 position in *V. natriegens* (26).

The spacing between the -10 and -35 positions have been reported to be an optimum of 17 bp in *E. coli* (19), and 16 bp in *V. natriegens* (26). BPROM results for these promoter and 5′UTR sequences showed distances from 11-18 bp. There was no clear indication of an effect on the fluorescence levels. This may be an indication of the tool's limitations and could possibly be influenced by the length of the motifs BPROM provides, i.e. if the motifs were of different lengths and more correctly positioned the actual distance could be around the 17 bp optimum spacing. In theory, there was only allowed a difference of 3-4 bp for continued recognition by RNAP (20, 23).

Table 3.3.1 displays which promoter and 5′UTR sequences containing the 7-bp optimum (20, 35) and the normalized endpoint GFP expressions in both organisms. The 7-bp optimum distance was present in all promoter and 5′UTR except for 1. 12F which resulted in low fluorescence levels in both organisms, 2. 4F which resulted in medium fluorescence levels in *V. natriegens* and high fluorescence levels in *E. coli* and 4. 1 C which resulted in high fluorescence levels in both organisms. Overall, the optimum distance does not seem to correlate with the expression levels. However, the distance between the RBS and the ATG gene start varied only with 3 bp, from 4 bp to 7 bp separating the two motifs. Overall, the fluorescence levels are variable even though the 7 bp optimum is present in the promoter and 5′UTR, which could potentially indicate that this is less crucial for gene expression (20, 35).

When studying each promoter and 5′UTR (table 3.3-3.4) and comparing with the measured GFP expression levels, it is difficult to determine if BPROM is a good predicter of motifs in *E. coli* or *V. natriegens* overall. BPROM did in some cases locate motifs in the beginning of the sequence, which probably indicated that these are far from the actual -10 and -35 positions

related to the core promoter upstream from the +1 TSS (29). Additionally, there was a tendency of low scores for the -35 position, even for the highest GFP producing promoter and 5′UTR sequence in *E. coli* and for another high producing promoter and 5′UTR in *V. natriegens.* This is another factor potentially illustrating the limitations of predictor software tools. The goal of comparing results from the sequence analysis for promoter and 5′UTR leading to variable expression levels in *E. coli* and *V. natriegens* was the possibility of detecting consistent differences in the scores provided by BPROM and Improbizer. If BRPROM would per example have provided consistent high scores for motifs leading to high expression in *E. coli* but not so for *V. natriegens*, it would have been possible to conclude that as the tool is optimized for *E. coli* inconsistencies were due to poor optimization for the non-model organism *V. natriegens*. Unfortunately, the results were not consistent enough to determine if the tools utilized in this thesis were best fit for either organisms, which probably are due to the relatively low amount of data available. If the screening would be repeated with considerably more promoter and 5′UTR's, it could be possible to determine how well in-silico sequence analysis tools as BRPOM and Improbizer work for the two organisms for prediction of what constructs that will result in high productivity. Improbizer overall seems more arbitrary considering positions of the motifs. The tool has been utilized in multiple publications focused on promoter analysis, and limitations regarding specificity due to learning principles utilized by the tool have been discovered as exemplified by Keilwagen *et. al* (2011) (79).

To improve the experiments performed for the sequence analysis, additional tools than utilized here should be included. A larger dataset would probably be necessary for the ability to detect patterns in the results and to draw conclusions on the differences in gene expression by *E. coli* and *V. natriegens.* It would be interesting to perform time-dependent expression measurements for longer periods of time with calibrations for the non-model organism *V. natriegens* to identify the exact link between growth stages and GFP expression levels. This could possibly lead to a very different results and subsequently comparisons of the two organisms.

Concluding remarks and future outlook

Throughout the discussion the focus was the measured GFP expression levels resulting from an artificial promoter and 5′UTR plasmid library in the model organism *E. coli* and the non-model organism *V. natriegens*. An attempted explanation of the variable functionality of the promoter and 5′UTR's were made by studying output from sequence analysis by BPROM and Improbizer. Although this is an informative and well-used methodology for studying gene expression levels, a shift in focus by researchers especially when discussing non-model organisms are currently underway (24).

A conclusion which could be drawn from the results discussed in this thesis is that the next step to improve the predictability of gene expression in *V. natriegens* could be establishing cloning methods in the organism and optimize the tools utilized for determining growth phases. This would make it possible to avoid having to utilize plasmids sourced from *E. coli* when characterizing the functionality of plasmid libraries in the non-model organism and to perform accurate measurements of which growth phases the resulting promoter and 5′UTR's activate during and to identify the responsible sigma factors (6, 73). The goal of achieving predictable and consistent gene expression may not be reached unless the scope is expanded beyond the promoter and 5′UTR sequences (80). Although the functionality of promoters in synthetic biology have been increased when moving towards use of synthetic sequences with artificial elements (29, 37, 39), consistent gene expression are still limited by factors as the availability of sigma factors and context dependency of the promoter sequences (40, 35, 73).

Improved predictability of regulating gene expression has been achieved by including the 5′UTR with the promoter, acknowledging a possible physical context dependency of the genetic elements involved in translation. A recent publication by Balzer *et. al (2020)* demonstrated that by implementing a dual 5′UTR system where two unique 5′UTR were separated by a spacer, several constructs resulting in increased expression of reporter genes and protein production were identified (81). A more holistic focus beyond engineered core promoters, towards including the 5′UTR and assessing context dependency of genetic elements as a factor causing variable and non-reproducible results could be worthwhile and time saving, compared to the more reductive focus on unravelling the functionality of non-model organisms gene transcription systems which could potentially require decades of research (33, 73, 80, 81).

Conclusion

The establishment and characterization of artificial constitutive promoter and 5′UTR libraries in *E. coli* and *V. natriegens* were performed by analysis of the promoter and 5′UTR DNA sequences and comparison of fluorescence levels of the green fluorescent protein.

In the thesis, growth conditions, a compatible plasmid with selection markers and the sfGFP reporter gene, protocols for preparing competent cells, transformation by electroporation and colony PCR for the non-model organisms *V. natriegens* are detailed. The functionality of the 36 artificial constitutive promoter and 5′UTR's described and characterized were variable in both organisms and led to an overall higher protein production levels in *E. coli.* This was expected by the information available from the research community working to establish molecular biology methods for *V. natriegens.*
The DNA sequence analysis of the promoter and 5′UTR sequences resulted in surprisingly inconsistent predictions of expected protein production levels of both organisms. This was not expected for the model organism *E. coli*, which the tools were optimized for. The predictions were based on how close the motifs identified by the tools BPROM and Improbizer were scored on containing probable -10 and -35 consensus sequences. The expected dependency on the correct spacing of the -10 and -35 positions were not confirmed, but due to the inconsistencies of the two tools functionality, definite conclusions should not be drawn from the analysis performed in this thesis.
PhD candidate Lisa Tietze discovered in early work with *V. natriegens* that the organism probably depends on the artificial promoter and 5′UTR sequence containing a defined Shine-Dalgarno sequence (GGAG) to achieve GFP expression in a sufficient amount of clones. The time dependent screening of GFP expression showed that the majority of the 36 promoter and 5′UTR's was probably activated during the late exponential to stationary growth phase in *E. coli.* Further experiments to determine the actual growth phases of *V. natriegens* are necessary to provide a conclusion on the promoter and 5′UTR activity in this organism. However, *V. natriegens* is ultimately a promising potential model organism for biotechnology in small and large scale if the efforts to establish use of the organism is continued.
Perhaps the most valuable conclusion to be drawn by the results presented here, supported by the current debate emerging in the synthetic biology research community, is the importance of including the probable context dependency of genetic elements. This is probably particularly

valuable working with non-model organism to achieve predictable and consistent regulation of protein production with good reproducibility throughout the synthetic biology and biotechnology research community.

List of references

1. Austin, B., Zachary, A., and Colwell, R.R. Recognition of Beneckea natriegens (Payne *et al.*) Baumann *et al.* as a Member of the Genus Vibrio, as Previously Proposed by Webb and Payne. International journal of systematic and evolutionary microbiology. 1978; 28: 315-317. https://doi.org/10.1099/00207713-28-2-315

2. PAYNE WJ, EAGON RG, WILLIAMS AK. Some observations on the physiology of Pseudomonas natriegens nov. spec. Antonie Van Leeuwenhoek. 1961;27:121-128. doi:10.1007/BF02538432

3. PAYNE WJ. Studies on bacterial utilization of uronic acids. III. Induction of oxidative enzymes in a marine isolate. J Bacteriol. 1958;76(3):301-307.

4. EAGON RG. Pseudomonas natriegens, a marine bacterium with a generation time of less than 10 minutes. J Bacteriol. 1962;83(4):736-737. doi:10.1128/JB.83.4.736-737.1962

5. Tanya Tschirhart *et al.*Synthetic biology tools for the fast-growing marine bacterium *Vibrio natriegens.* ACS Synthetic Biology 2019 8 (9), 2069-2079.
DOI: 10.1021/acssynbio.9b00176

6. Weinstock, M., Hesek, E., Wilson, C. *et al.* Vibrio natriegens as a fast-growing host for molecular biology. Nat Methods 13, 849–851 (2016). https://doi.org/10.1038/nmeth.3970

7. Hoffart E, Grenz S, Lange J, Nitschel R,Müller F, Schwentner A, Feith A, Lenfers-LückerM, Takors R, Blombach B. 2017. High substrate uptake rates empower Vibrio natriegens as production host for industrial biotechnology. Applied Environmental Microbiology 83:e01614-17. https://doi.org/10.1128/AEM.01614-17.

8. Lee, H.H., Ostrov, N., Wong, B.G. *et al.* Functional genomics of the rapidly replicating bacterium Vibrio natriegens by CRISPRi. Nat Microbiol 4, 1105–1113 (2019). https://doi.org/10.1038/s41564-019-0423-8

9. Pfeifer E, Michniewski S, Gätgens C, Münch E, Müller F, Polen T, Millard A, Blombach B, Frunzke J. 2019. Generation of a prophage-free variant of the fast-growing bacterium Vibrio natriegens. Appl Environ Microbiol 85:e00853-19. https://doi.org/10.1128/AEM.00853-19

10. Rosano GL and Ceccarelli EA (2014) Recombinant protein expression in Escherichia coli: advances and challenges. Front. Microbiol. 5:172. doi: 10.3389/fmicb.2014.00172

11. Schleicher L, Muras V, Claussen B, Pfannstiel J, Blombach B, Dibrov P, Fritz G and Steuber J (2018) Vibrio natriegens as Host for Expression of Multisubunit Membrane Protein Complexes. Front. Microbiol. 9:2537. doi: 10.3389/fmicb.2018.02537

12. Pelosse, M., Crocker, H., Gorda, B. *et al.* MultiBac: from protein complex structures to synthetic viral nanosystems. BMC Biol 15, 99 (2017). https://doi.org/10.1186/s12915-017-0447-6

13. Dalia TN, Hayes CA, Stolyar S, Marx CJ, McKinlay JB, Dalia AB. Multiplex Genome Editing by Natural Transformation (MuGENT) for Synthetic Biology in Vibrio natriegens. ACS Synth Biol. 2017;6(9):1650-1655. doi:10.1021/acssynbio.7b00116

14. Karin L. Meibom, Melanie Blokesch, Nadia A. Dolganov, Cheng-Yen Wu, Gary K. Schoolnik. Chitin Induces Natural Competence in Vibrio cholerae. Science (2005);310(5755): 1824-1827. DOI: 10.1126/science.1120096

15. Gulig PA, Tucker MS, Thiaville PC, Joseph JL, Brown RN. USER friendly cloning coupled with chitin-based natural transformation enables rapid mutagenesis of Vibrio vulnificus. Applied Environmental Microbiology. 2009;75(15):4936-4949. doi:10.1128/AEM.02564-08

16. Dalia AB, Lazinski DW, Camilli A. Identification of a membrane-bound transcriptional regulator that links chitin and natural competence in Vibrio cholerae. mBio. 2014;5(1):e01028-13.doi:10.1128/mBio.01028-13

17. Centeno-Leija S, Huerta-Beristain G, Giles-Gómez M, Bolivar F, Gosset G, Martinez A. Improving poly-3-hydroxybutyrate production in Escherichia coli by combining the increase in the NADPH pool and acetyl-CoA availability. Antonie Van Leeuwenhoek. 2014;105(4):687-696.

18. Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. Cells As Experimental Models. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9917/

19. Engstrom MD, Pfleger BF. Transcription control engineering and applications in synthetic biology. Synth Syst Biotechnol. 2017;2(3):176-191. doi:10.1016/j.synbio.2017.09.003

20. Molecular Biology, David P. Clark and Nanette J. Pazdernik. Academic/AP cell press.

21. Nelson DL. *Lehninger principles of biochemistry.* 6th ed. New York: W.H. Freeman; 2012.

22. T Murlidharan Nair, B.D. Kulkarni. On the consensus structure within the *E. coli* promoters. Biophysical Chemistry 1994; 48(3):383-393. Doi:https://doi.org/10.1016/0301-4622(93)E0056-B.

23. Paget MS. Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution. Biomolecules. 2015;5(3):1245-1265. doi:10.3390/biom5031245

24. Bervoets I, Charlier D. Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology. FEMS Microbiol Rev. 2019;43(3):304-339. doi:10.1093/femsre/fuz001

25. Mendoza-Vargas A, Olvera L, Olvera M, *et al.* Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. PLoS One. 2009;4(10):e7526. doi:10.1371/journal.pone.0007526

26. Aiyar SE, Gaal T, Gourse RL. rRNA promoter activity in the fast-growing bacterium Vibrio natriegens. J Bacteriol. 2002;184(5):1349-1358. doi:10.1128/jb.184.5.1349-1358.2002

27. Duval, M., Simonetti, A., Caldelari, I., Marzi, S.Multiple ways to regulate translation initiation in bacteria: Mechanisms, regulatory circuits, dynamics. Biochimie. 114: 18-29. (2015). https://doi.org/10.1016/j.biochi.2015.03.007

28. Bervoets I, Charlier D. Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology. FEMS Microbiol Rev. 2019;43(3):304-339. doi:10.1093/femsre/fuz001

29. Gilman, James, and John Love. "Synthetic promoter design for new microbial chassis." Biochemical Society transactions vol. 44,3 (2016): 731-7. doi:10.1042/BST20160042

30. De Nijs, Y., De Maeseneire, S.L. and Soetaert, W.K. (2020), 5′ untranslated regions: the next regulatory sequence in yeast synthetic biology. Biol Rev, 95: 517-529. doi:10.1111/brv.12575s

31. Mutalik VK, Guimaraes JC, Cambray G, *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Methods. 2013;10(4):354-360. doi:10.1038/nmeth.2404

32. Berg, Laila *et al.* "The expression of recombinant genes in Escherichia coli can be strongly stimulated at the transcript production level by mutating the DNA-region corresponding to the 5'-untranslated part of mRNA." Microbial biotechnology vol. 2,3 (2009): 379-89. doi:10.1111/j.1751-7915.2009.00107.x

33. Mutalik, V., Guimaraes, J., Cambray, G. *et al.* Quantitative estimation of activity and quality for collections of functional genetic elements. Nat Methods 10, 347–353 (2013). https://doi.org/10.1038/nmeth.2403

34. Borujeni, A. E., Salis, H. M. Translation Initiation is Controlled by RNA Folding Kinetics via a Ribosome Drafting Mechanism. Journal of the American Chemical Society 138(22), 7016-7023 (2016). DOI: 10.1021/jacs.6b01453

35. Yurovsky A, Amin MR, Gardin J, Chen Y, Skiena S, Futcher B. Prokaryotic coding regions have little if any specific depletion of Shine-Dalgarno motifs. PLoS One. 2018;13(8):e0202768. doi:10.1371/journal.pone.0202768

36. Justin R.J. Vigar, Hans-Joachim Wieden. Engineering bacterial translation initiation — Do we have all the tools we need? Biochimica et Biophysica Acta (BBA) - General Subjects. 2017;1861(11):3060-3069. https://doi.org/10.1016/j.bbagen.2017.03.008

37. Cavicchioli, R., Ripple, W.J., Timmis, K.N. *et al.* Scientists' warning to humanity: microorganisms and climate change. Nat Rev Microbiol 17, 569–586 (2019). doi: https://doi.org/10.1038/s41579-019-0222-5

38. Salis HM, Mirsky EA, Voigt CA. Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol. 2009;27(10):946-950. doi:10.1038/nbt.1568

39. Rahmi Lale, Lisa Tietze, Jenny Nesje, Ingerid Onsager, Kerstin Engelhardt, Che Fai Alex Wong, Madina Akan, Niklas Hummel, Jörn Kalinowski, Christian Rückert, Martin Frank Hohmann-Marriott. A universal method for gene expression engineering. bioRxiv 644989; doi: https://doi.org/10.1101/644989

40. Jin L, Nawab S, Xia M, Ma X, Huo YX. Context-dependency of synthetic minimal promoters in driving gene expression: a case study. Microb Biotechnol. 2019;12(6):1476-1486. doi:10.1111/1751-7915.13489

41. Jin, L., Jin, W., Ma, Z. *et al.* Promoter engineering strategies for the overproduction of valuable metabolites in microbes. Appl Microbiol Biotechnol 103, 8725–8736 (2019). https://doi.org/10.1007/s00253-019-10172-y

42. Leavitt, J. M., Alper, H. S. Advances and current limitations in transcript-level control of gene expression. Current Opinion in Biotechnology. 34,98-104 (2015). https://doi.org/10.1016/j.copbio.2014.12.015

43. Failmezger J, Scholz S, Blombach B, Siemann-Herzberg M. Cell-Free Protein Synthesis From Fast-Growing Vibrio natriegens. Front Microbiol. 2018;9:1146. doi:10.3389/fmicb.2018.01146

44. IGEM. Registry of standard biological parts - Anderson promoter catalogue; 2017. Accessed: 2020-04-08. Available from: http://parts.igem.org/Promoters/Catalog/Anderson

45. Valla, S., and Lale, R. DNA cloning and assembly methods. 1st ed. Humana Press; 2014. :119:133.

46. Lykke, A. (2019). *Construction and characterisation of artificial promoters in Bacillus subtilis.* Unpublished master thesis. Norwegian University of Science and Technology, Trondheim, Norway.

47. Synthetic biology: BioBricks; 2007. Accessed: 2020-04-08. Available from: https://openwetware.org/wiki/Synthetic_Biology:BioBricks

48. Green, R., Rogers R. J. Chemical Transformation of E-coli. NIH. 2013: 529: 329–336. doi:10.1016/B978-0-12-418687-3.00028-8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4037286/pdf/nihms577557.pdf

49. QIAprep spin Miniprep Kit: Qiagen: 2020. Accessed: 2020-04-20. Available from: https://www.qiagen.com/dk/products/discovery-and-translational-research/dna-rna-purification/dna-purification/plasmid-dna/qiaprep-spin-miniprep-kit/#orderinginformation

50. Pédelacq, J., Cabantous, S., Tran, T. et al. Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol 24, 79–88 (2006). https://doi.org/10.1038/nbt1172

51. New England Biolabs inc.: PCR Using Q5® High-Fidelity DNA Polymerase (M0491): 2020. Accessed: 2020-04-20. Available from: https://international.neb.com/protocols/2013/12/13/pcr-using-q5-high-fidelity-dna-polymerase-m0491

52. New England Biolabs inc.: DpnI: 2020. Accessed: 2020-04-20. Available from: https://international.neb.com/products/r0176-dpni#Product%20Information

53. New England Biolabs inc.: T4 Polynucleotide Kinase: 2020. Accessed: 2020-04-20. Available from: https://international.neb.com/products/m0201-t4-polynucleotide-kinase#Product%20Information

54. Qiagen: QIAquick PCR Purification Kit: 2020. Accessed: 2020-04-20. Available from: https://www.qiagen.com/us/shop/pcr/qiaquick-pcr-purification-kit/

55. BioNordika: NanoDrop - Microvolume Spectrophotometer: 2020. Accessed: 2020-04-20. Available from: https://bionordika.no/produkter/analytisk-kjemi/spektroskopi/microvolume-spectrophotometer

56. Vladimir Potapov, Jennifer L. Ong, Rebecca B. Kucera, Bradley W. Langhorst, Katharina Bilotti, John M. Pryor, Eric J. Cantor, Barry Canton, Thomas F. Knight, Thomas C. Evans, and Gregory J. S. Lohman. ACS Synthetic Biology 2018 7 (11), 2665-2674 Doi: 10.1021/acssynbio.8b00333

57. New England Biolabs inc.: Tm Calculator: 2020. Accessed: 2020-04-20. Available from: https://tmcalculator.neb.com/#!/main

58. New England Biolabs inc.: PCR Protocol for Taq DNA Polymerase with Standard Taq Buffer (M0273): 2020. Accessed: 2020-04-20. Available from: https://international.neb.com/protocols/0001/01/01/taq-dna-polymerase-with-standard-taq-buffer-m0273

59. Wang Q et.al. Quick and efficient method for genetic transformation of biopolymer-producing bacteria. Journal of Chemical Technology & Biotechnology. 2009;6(85). Doi: 10.1002/jctb.2284

60. Lee H, Ostrov N. Vibrio natriegens - DNA transformation by electroporation. Protocols.io. Accessed 2020-04-20. Available from: https://www.protocols.io/view/vibrio-natriegens-dna-transformation-by-electropor-habb2an

61. Solovyev V, Salamov A (2011). Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In Metagenomics and its Applications. Agriculture, Biomedicine and Environmental Studies (Ed. R.W. Li). Nova Science Publishers. 61-78.

62. UC Santa Cruz. Improbizer. Accessed 2020-04-20. Available from: https://users.soe.ucsc.edu/~kent/improbizer/improbizer.html

63. Guss, A. M., Olson, D. G., Caiazza, N. C., & Lynd, L. R. 2012. Dcm methylation is detrimental to plasmid transformation in Clostridium thermocellum. Biotechnology for biofuels, 5(1), 30. Doi: 10.1186/1754-6834-5-30

64. Eichmann J., Oberpaul M., Weidner T., Gerlach D., Czermak P. Selection of High Producers From Combinatorial Libraries for the Production of Recombinant Proteins in Escherichia coli and Vibrio natriegens. Frontiers in Bioengineering and Biotechnology. 2014;254(7): 2296-4185. Doi: 10.3389/fbioe.2019.00254

Casadevall, A., & Fang, F. C. Reproducible science. Infection and immunity, 2010:78(12), 4972–4975. Doi: 10.1128/IAI.00908-10

65. Jahn M, Vorpahl C, Hübschmann T, Harms H, Müller S. Copy number variability of expression plasmids determined by cell sorting and Droplet Digital PCR. Microb Cell Fact. 2016;15(1):211 doi:10.1186/s12934-016-0610-8

66. In-Soo Kong, Tonya C. Bates, Anja Hülsmann, Hosni Hassan, Ben E. Smith, James D. Oliver, Role of catalase and oxyR in the viable but nonculturable state of Vibrio vulnificus, FEMS Microbiology Ecology, 2014;50(3). Doi: 10.1016/j.femsec.2004.06.004

67. Geoff Baldwin, Traci Haddock-Angelli, Jacob Beal, Ari Dwijayanti, Marko Storch, Natalie Farny, Cheryl Telmer, Alejandro Vignoni, Richard Tennant, Paul Rutten 2019. Calibration Protocol - Plate Reader Fluorescence Calibration. Accessed 2020-04-20. Available from: https://www.protocols.io/view/calibration-protocol-plate-reader-fluorescence-cal-6zrhf56

68. Casadevall A, Fang FC. Reproducible science. Infect Immun. 2010;78(12):4972-4975. doi:10.1128/IAI.00908-10

69. Vaux DL, Fidler F, Cumming G. Replicates and repeats--what is the difference and is it significant? A brief discussion of statistics and experimental design. EMBO Rep. 2012;13(4):291-296. Published 2012 Apr 2. doi:10.1038/embor.2012.3670.

70. Shimada T, Makinoshima H, Ogawa Y, Miki T, Maeda M, Ishihama A. Classification and strength measurement of stationary-phase promoters by use of a newly developed promoter cloning vector. J Bacteriol. 2004;186(21):7112-7122. doi:10.1128/JB.186.21.7112-7122.2004

71. Kobayashi A, Hirakawa H, Hirata T, Nishino K, Yamaguchi A. Growth phase-dependent expression of drug exporters in Escherichia coli and its contribution to drug tolerance. J Bacteriol. 2006;188(16):5693-5703. doi:10.1128/JB.00217-06

72. Shimada T, Tanaka K, Ishihama, A. The whole set of the constitutive promoters recognized by four minor sigma subunits of Escherichia coli RNA polymerase. PLoS ONE 2017:12(6): e179181. doi: 10.1371/journal.pone.0179181

73. Han L, Chen Q, Lin Q, et al. Realization of Robust and Precise Regulation of Gene Expression by Multiple Sigma Recognizable Artificial Promoters. Front Bioeng Biotechnol. 2020;92(8). doi:10.3389/fbioe.2020.00092

74. Kafri M, Metzl-Raz E, Jona G, Barkai N. The Cost of Protein Production. Cell Rep. 2016;14(1):22-31. doi:10.1016/j.celrep.2015.12.015

75. Sezonov G, Joseleau-Petit D and D'Ari R. Escherichia coli Physiology in Luria-Bertani Broth. Journal of Bacteriology. 2007;189(23): 8746-8749. doi: 10.1128/JB.01368-07

76. Hebisch E, Knebel J, Landsberg J, Frey E, Leisner M. High variation of fluorescence protein maturation times in closely related Escherichia coli strains. PLoS One. 2013;8(10):e75991. Published 2013 Oct 14. doi:10.1371/journal.pone.007599177.

78. Long CP, Gonzalez JE, Cipolla RM, Antoniewicz MR. Metabolism of the fast-growing bacterium Vibrio natriegens elucidated by 13C metabolic flux analysis. Metab Eng. 2017;44:191-197. doi:10.1016/j.ymben.2017.10.00879.

80. Cardinale S, Arkin AP. Contextualizing context for synthetic biology--identifying causes of failure of synthetic biological systems. Biotechnol J. 2012;7(7):856-866. doi:10.1002/biot.201200085

81. Le SB, Onsager I, Lorentzen Ja, Lale R. Dual UTR-A novel 5′ untranslated region design for synthetic biology applications, Synthetic Biology, ysaa006. doi: 10.1093/synbio/ysaa006

Appendix A

Primer sequences

Table A1: Primer sequences utilized in this thesis and descriptions of the use.

Primer name	**Sequence (5′ to 3′ direction)**	**Description**
ALY_529_TH1_sfGFP_fw	gtacgcgtacggtctcgn ATGCGTAAAGGCGAAGAGCTG	Creating the pACYC-sfGFP plasmid backbone
ALY_530_TH1_rv_200	gtacgcgtacggtctctggca GGCGTAATCATGGTCATAGCTGTTTC	
BB_Prefix_Fwd	GAATTCGCGGCCGCTTCTAGAG	Amplification of the GeneEE inserts
BB_Suffix_rev	CTGCAGCGGCCGCTACTAGA	
ALY532 seq forward	CCGATTCATTAATGCAGCTG	Confirming elimination of the native mdh promoter, colony PCR with Taq polymerase and sequencing with Q5 of the 200N SD promoter and 5′UTR region
ALY533 seq reverse	AACTTCAGCGTCAGTTTACC	

Appendix B

DNA ladder

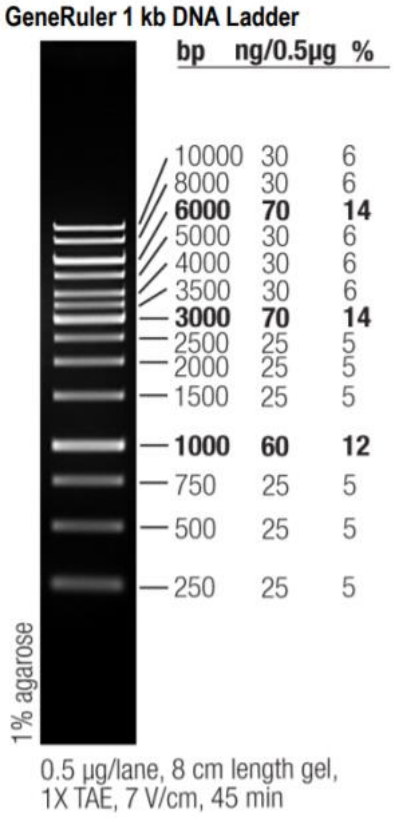


Figure B1: Thermo Scientific™ GeneRuler™ 1 kb DNA ladder utilized for determining size of DNA fragments by agarose gel electrophoresis. Image downloaded from https://assets.thermofisher.com/TFS-Assets/LSG/manuals/MAN0013004_GeneRuler_1kb_DNALadder_250ug_UG.pdf 28.07.2020.

Appendix C

GeneEE insert template

5′ - GAATTCGCGGCCGCTTCTAGAGGGTCTCaTGCC-200N-**GGAG**-7N-atgc

GAGACCTACTAGTAGCGGCCGCTGCAG - 3′

Figure C1: GeneEE insert template for the 200N SD insert. BsaI sites are in red color and underlined. The Shine-Dalgarno (SD) sequence are in bold and underlined. The sequence contains a 7N insert after the 200 random nucleotides stretch. The position of the 200N sequence is shaded in turquoise.

Appendix D

Pre-screen of triplicates by fluorescence and absorbance measurements in *V. natriegens*

Figure D1: Plate 1 of pre-screen of *V. natriegens* clones transformed with the promoter and 5′UTR libraries. Results from fluorescence and absorbance measurement with a Tecan infinite M200 Pro plate reader. The positive and negative control including a blank is indicated on the plate position A1-3.

<>	Blank	Pos	Neg	4	5	6	7	8	9	10	11	12
A	69	903	74	82	102	77	87	99	83	99	88	93
B	101	94	82	81	91	77	85	108	82	99	83	88
C	92	90	80	89	87	75	78	104	81	121	82	89
D	90	98	82	82	92	81	78	106	100	122	81	82
E	93	97	98	90	102	80	79	81	108	114	151	78
F	79	91	99	92	100	78	97	83	112	81	143	81
G	78	107	92	98	110	101	97	83	101	81	144	86
H	79	101	91	96	84	104	89	86	104	83	106	86

<>	Blank	Pos	Neg	4	5	6	7	8	9	10	11	12
A	0,0517	0,7121	0,6257	0,7365	0,7296	0,7187	0,7526	0,741	0,7655	0,7193	0,7512	0,7825
B	0,7169	0,7338	0,7456	0,7292	0,7426	0,7415	0,7717	0,743	0,7548	0,7284	0,7202	0,7143
C	0,712	0,6859	0,6696	0,7032	0,6675	0,7255	0,7338	0,6926	0,7328	0,7322	0,7306	0,7227
D	0,6741	0,7502	0,7382	0,6233	0,7395	0,7336	0,7483	0,724	0,6443	0,7256	0,7406	0,7686
E	0,6961	0,755	0,7377	0,7156	0,6799	0,7005	0,7388	0,7476	0,7	0,6822	0,7922	0,7253
F	0,7065	0,7393	0,7542	0,7334	0,6527	0,6628	0,7456	0,6969	0,7247	0,7334	0,7517	0,7604
G	0,657	0,7348	0,649	0,742	0,7303	0,7263	0,7426	0,7162	0,7256	0,7016	0,7602	0,7318
H	0,6751	0,712	0,6577	0,7334	0,7355	0,7535	0,6691	0,7458	0,7609	0,746	0,7898	0,7326

Figure D2: Plate 2 of pre-screen of *V. natriegens* clones transformed with the promoter and 5′UTR libraries. Results from fluorescence and absorbance measurement with a Tecan infinite M200 Pro plate reader. A blank value is indicated in A1, and a positive control in H2. The first row are blank values due to number of clones.

<>	1	8	9	10	11	12
A	70	83	73	82	129	86
B	71	75	74	80	125	83
C	68	75	74	80	115	80
D	70	72	85	74	117	71
E	68	95	88	80	78	80
F	68	108	87	75	77	85
G	70	113	82	82	78	141
H	70	1164	85	84	79	131

<>	Blank	8	9	10	11	12
A	0,0525	0,6785	0,5985	0,681	0,7057	0,6647
B	0,0528	0,7003	0,7023	0,7208	0,7205	0,667
C	0,0529	0,7005	0,6862	0,6948	0,6869	0,6328
D	0,0526	0,651	0,6725	0,5954	0,6695	0,5036
E	0,052	0,6196	0,7187	0,7069	0,7124	0,6375
F	0,0522	0,6977	0,6873	0,5753	0,7142	0,6723
G	0,0519	0,7049	0,6964	0,7013	0,7131	0,6983
H	0,0522	0,7348	0,7039	0,7208	0,6542	0,687

Appendix E

Sequence results confirming that the promoter and 5′UTR are unique from the native mdh promoter.

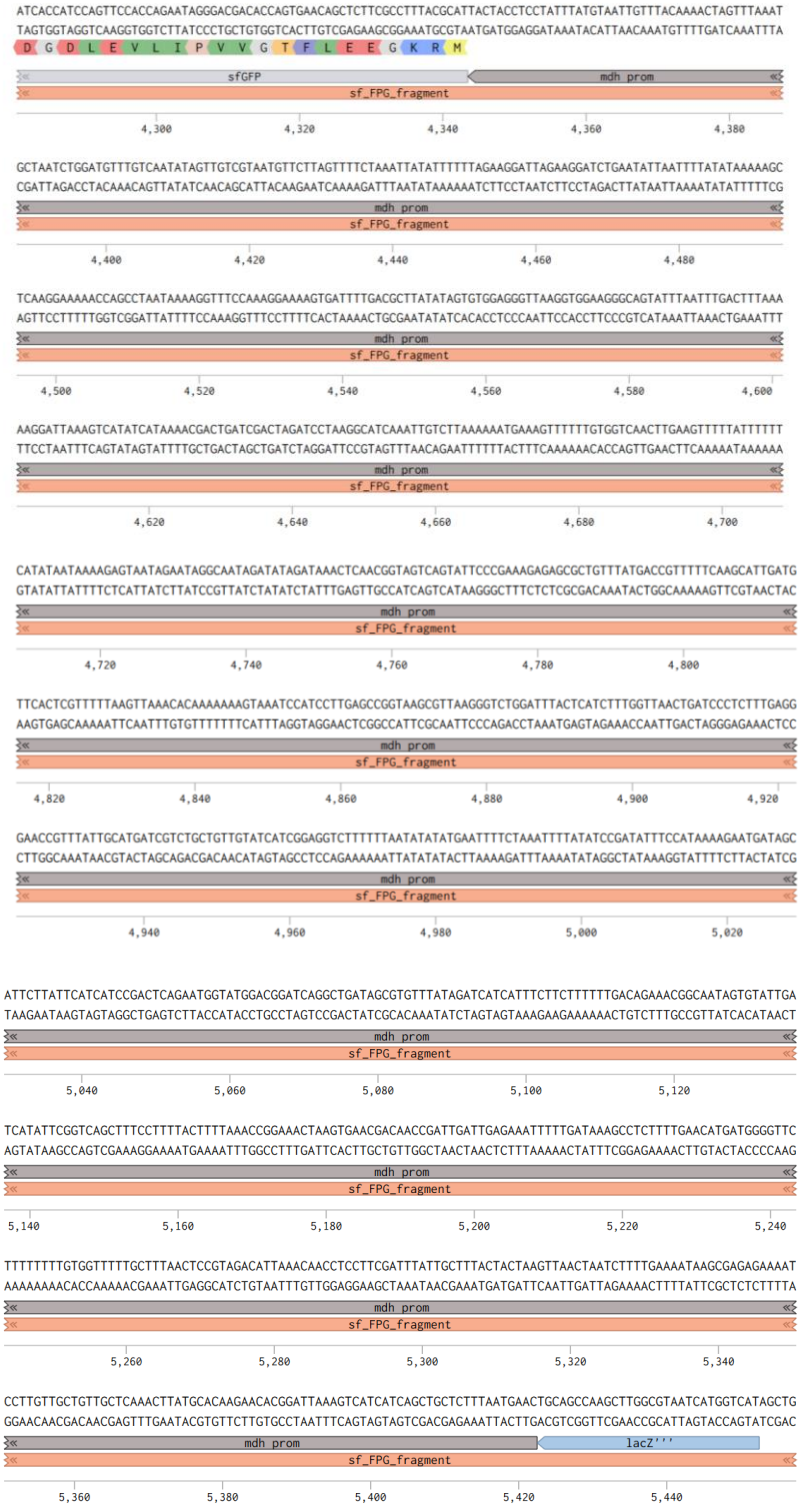


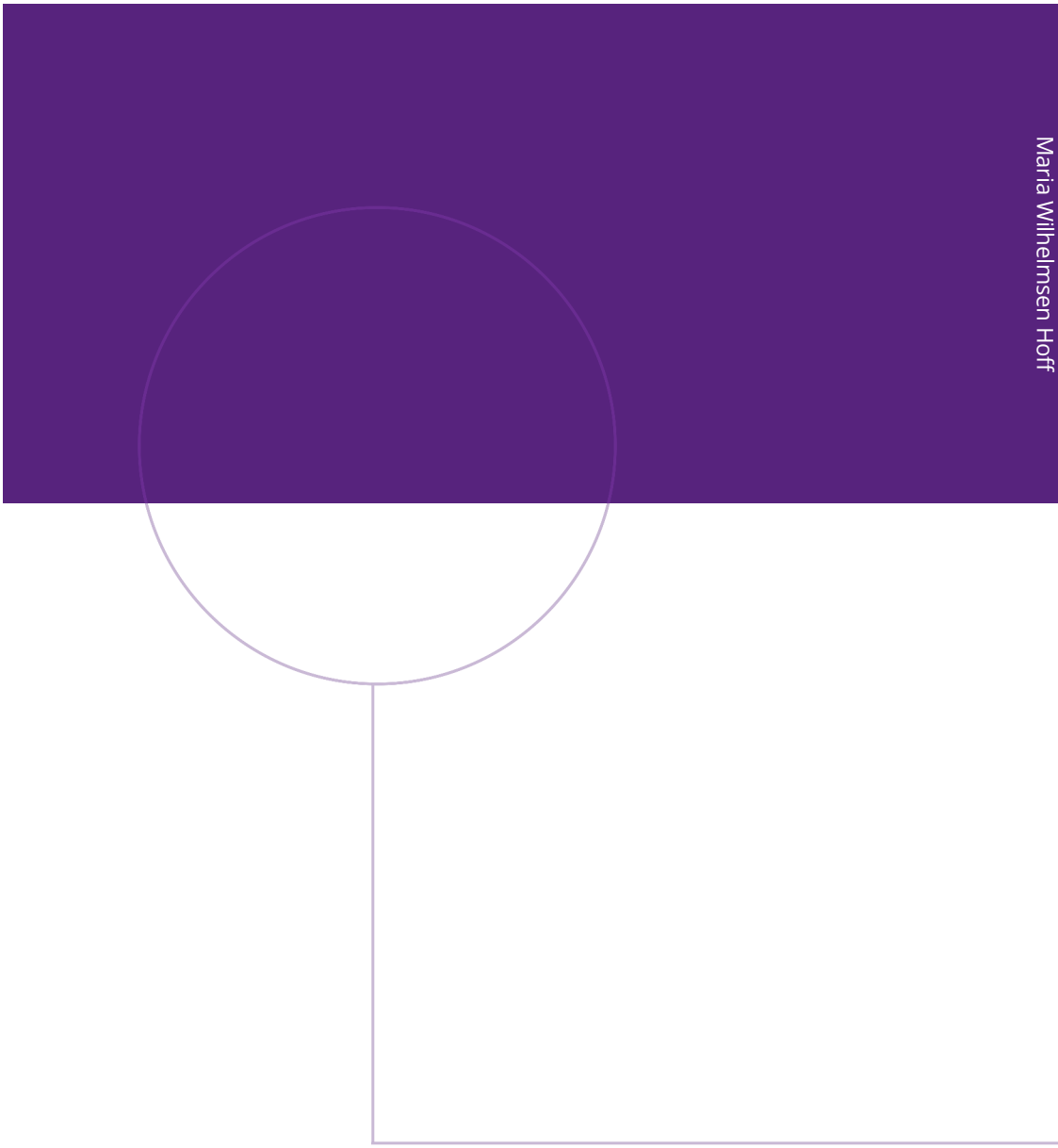
Figure E1: The sequence of the native mdh promoter in the plasmid pACYC-sfGFP. A sequence alignment was performed *in silico* with the online software Benchling with the 36 promoter and 5′UTR, confirming that the sequences were in fact unique from the mdh promoter.

Appendix F

Sequences from identical promoter and 5′UTR sequences from plasmids isolated from *E. coli*. Confirmation that the correct plasmids were transformed back from *V. natriegens*.

Table F1: Sequences from Q5 PCR of the promoter and 5′UTR sequences 4. 2G/ 6, 4. 4B/28 and 3. 4B/ 30 isolated from *V. natriegens* and *E. coli*. The plasmids were transformed back into *E. coli* when sequencing confirmed that the promoter and 5′UTR sequences were functional and unique.

Identity ***V. natriegens***	**Identity** ***E. coli***	**Sequence**
4. 2G	**6**	TGTTGGTATCCATGAGTTTACTGATGAAGCTTAGTACTT CCGGGTATTTCATGTTCACACAGAGTGTAGAGTTACTAT GGCAGTTCTGTTATTTTGACGGTGTGAGGGCGTTTGTTA TCGTGCTAGTAAAATCGTCTAATGTTGTTGTGGAAGCTT GACATAGTCATGTTAATCATGAAGGTAGTGGCTGGGAG TAGGATG
4. 4B	**28**	GCGCATGTGGGCTTTTTGGGACTATTTTTATATGTGGGG TTCTCTTTTGACAAGCGGGTTTGTTGGTGTAATACAGGT AGGAAACACGAGGTGAATAGCTGGTCCGGAACAACAT TGATTCCAGTTTGGTCGATGCATAGGATCGAAGAACAC AACTGCTTCTGTATTTCGTTACTTTAGGGAAATTAGGGT GGGAGAAAATTG
3. 4B	**30**	CTGGTAAAGTTAATGGTATTGAGCTCCGGGTCGGTCTGA AACTGTAACGTATTTAGGTATGTTGTTCGTGTGTGGTCTT TTCGAAAATACTATACGACAATATACCTGGGGAGGATCG TGCGATGTTACACCTGGCGGAGGGTGATACTGGCGTAGG GATTGAGATCATAAGGAAGATTGAGGCGGTGAAGAGGG AGGGGGATG



Maria Wilhelmsen Hoff

NTNU
Norwegian University of
Science and Technology