Development of protocols and workflows for a fast gene synthesis and de novo synthesis of viral genomes

Dissertation

zur Erlangung des Doktorgrades rerum naturalium

an der Universität Hamburg,

MIN Fakultät für Mathematik, Informatik und Naturwissenschaften,

Fachbereich Biologie

Vorgelegt von Julia Weigl

aus Regensburg

Dezember 2018

This doctoral thesis was completed from 1st of November 2015 until 23rd of October 2018 in the Molecular Biology Research Group at Thermo Fisher Scientific GeneArt GmbH Regensburg in cooperation with Heinrich-Pette-Institut Hamburg.

Submitted by Julia Weigl on 2018-10-23 in Hamburg, Biozentrum Klein Flottbeck.
Disputation: 2018-12-05

First evaluation: Prof. Dr. Thomas Dobner, HPI Hamburg

Second evaluation: Prof. Dr. Ralf Wagner, University of Regensburg

TABLE OF CONTENTS

ABBREVIATIONS

0	percent
°C	degree Celsius
bp	Base pairs
CFTR	cystic fibrosis transmembrane conductance regulator
cfu	Colony forming units
cm	Centimetre
CP	Control primer
DBP	DNA binding protein
DNA	deoxyribonucleic acid
DNS	Desoxyribonukleinsäure
dNTP	deoxyribonucleoside triphosphate
e.g.	for example
et al.	and others
EtBr	Ethidium bromide
FD	Fast Digest
fwd	forward
g	gram
GFP	Green fluorescent protein
GPAdV	Guinea Pig Adenovirus
h	hour(s)
H_2O	water
HF	High-Fidelity
kb	Kilo bases
L	liter
LiAc	Lithium acetate
m	milli
M	molar (mol/liter)
Matα	Mating type alpha
min	minute
ms	Milli seconds
MW	molecular weight
nf	Nuclease free water
NC	Negative control
ng	Nano grams
OL	overlap
ori	Origin of replication
PBD	Plackett-Burman-Design
PC	Positive control
PCR	polymerase chain reaction
Pol	Polymerase
rcf	relative centrifugal force
rev	reverse
rpm	revolutions per minute
RT	room temperature
sec	Second
SCR	Sequential chain reaction
SS	Salmon Sperm
Taq	polymerase (*Thermus aquaticus*)
THR	Terminal homology region
V	Volt
WT	wild type
µ	micro

ABSTRACT

Synthetic biology became one of the most admired branches within new age technologies. The past 30 years brought innovations forth with high impact on medicine, ecology and technology. Indispensable in all areas is the production of synthetic DNA in large scales and high quality.

In this thesis, the first project was to optimize the established production pathway of gene synthesis on the basis of polymerase chain reactions (PCR) in time, to abridge the generating of desired genes. A set of various sequences from different origin and complexity ranging from 300 – 1500 bp in size with a moderate GC-content, was tested on the new developed protocols. A significant reduction of the turnaround time by 47% was achieved, keeping the reliability of correct constructs. The utilisation of the Plackett-Burman statistical tool afforded the effective determination of critical factors by an experimental set-up with a fractional factorial design. In this study, protocols were developed to serve the need of fast access to genetic material as for DNA vaccination against tumours.

The goal of the second project was the consecutive building of large DNA fragments of a wild type adenoviral sequence. The guinea pig adeno virus was found to be responsible for infectious outbreaks among laboratory guinea pig populations, leading to the severe disease bronchopneumonia of immunocomprised animals, ending in death of the infected animals. Histopathological investigations of the isolate GER1 occasioned to the generation of a complete *in silico* sequence of the genomic DNA. This sequence was the source material to develop synthesis protocols for challenging large constructs. The synthesis of the complete adenoviral genome in eight blocks was successfully performed and can be used for isolation and cultivation of the virus after transfection into guinea pig cells. Furthermore, an official annotation can now be conducted based on transcriptome analysis. In this thesis, eight blocks were built by using the capability of *Saccharomyces cerevisiae* to homologous recombination. With the Genome Partitioner tool a sufficient higher order assembly strategy was developed, which is applicable on the wild-type sequence. Due to long GC-peaks, allocated secondary structures and the multiple presence of common restriction sites other cloning strategies were found unsuitable. Adenoviruses inherit a specific inverted terminal repeat sequence (ITR) flanking both 5’ and 3’ terminal ends, that is responsible for integration into the host genome and initiation of replication. These ITRs interfere the complete assembly of a whole viral linear genome acting like overlapping sequences and may lead to mis-assembly. A sequence optimization of critical sequence areas might give the possibility to build even larger fragments and the whole genome. To avoid the recombination on the ITR

sites, a vector can be plotted that already includes the ITR sequences. Thus the ITRs can be released by restriction digest connected to the remaining genomic sequence after integration into the target vector. Transfection of guinea pig tracheal cells (GPTEC-T) with the *in vitro* assembled whole construct did not lead to virus formation.

In this study though, the basis was created for deeper determinations on the viral genome and its infection mechanism. This establishment of a guinea pig adenovirus model can then answer further questions on the tumour formation that appears after infection of rodents with human adenoviruses and pathogenicity in the guinea pig host organism.

ZUSAMMENFASSUNG

Synthetische Biologie wurde zu einer der renommiertesten Branchen unter den zeitgenössischen Technologien. Die letzten 30 Jahre brachten Innovationen hervor, die einen starken Einfluss auf Medizin, Ökologie und Technologie haben. Unverzichtbar in allen Bereichen ist dabei die Produktion synthetischer DNS in großem Maßstab und hoher Qualität. In dieser Arbeit befasste sich das erste Projekt mit der zeitlichen Optimierung des klassischen Produktionswegs der Gensynthese, der auf Polymerasekettenreaktion (PKR) basiert, um die Generierung der gewünschten Gene zu beschleunigen. Eine Zusammenstellung verschiedener Sequenzen unterschiedlichen Ursprungs und Komplexität, die zwischen 300 – 1500 bp lang waren und einen moderaten GC-Gehalt hatten, wurde mit den neu entwickelten Protokollen getestet. Eine signifikante Reduktion der Durchlaufzeit von 47% wurde erreicht, bei gleichbleibender Zuverlässigkeit der richtigen Konstrukte. Die Anwendung des Plackett-Burman statistischen Instruments erbrachte die effiziente Identifikation kritischer Faktoren, durch einen teilfaktoriellen Versuchsplan. In dieser Arbeit wurden Protokolle entwickelt, die das Erfordernis nach schnellem Zugang zu genetischem Material bedienen, wie bei der DNA Impfung gegen Tumore.

Ziel des zweiten Projektes war der konsekutive Bau großer DNS Fragmente einer wildtypischen, adenoviralen Sequenz. Der neu entdeckte Meerschweinchen-Adenovirus ist verantwortlich für infektiöse Krankheitsausbrüche unter Meerschweinchenpopulationen in Laboren, die zu einer schweren Erkrankung an Bronchopneumonie in immunschwachen Tieren führten und im Tod betroffener Tiere endeten. Histopathologische Untersuchungen der Isolates GER1 lieferten die *in silico* Sequenz der genomischen DNS. Diese Sequenz war das Ausgangsmaterial, um Syntheseprotokolle zu entwickeln, abgestimmt auf schwierige, große Konstrukte. Die Herstellung des Adenovirusgenoms in acht Blöcken war erfolgreich durchgeführt worden und kann nach Transfektion in Meerschweinchenzellen zu einer Isolierung und Kultivierung des Virus genutzt werden. Des Weiteren kann jetzt nach einer Transkriptomanalyse der Säugerzellen eine offizielle Annotation vorgenommen werden. In dieser Arbeit wurden die acht Blöcke unter Verwendung des Mechanismus der homologen Rekombination von *Saccharomyces cerevisiae* gebaut. Mit dem Genome-Partitioner-Tool wurde eine geeignete hierarchische Assemblierungsstrategie entworfen, die auf diese wildtypische Sequenz anwendbar ist. Durch lange GC-Strecken, verteilte Sekundärstrukturen und das mehrfache Vorhandensein gängiger Restriktionsschnittstellen waren andere Klonierungsstrategien ungeeignet. Adenoviren beinhalten in ihrer DNS Sequenz sogenannte invertierte terminale Repetitionen (ITR), die sowohl das 5‘-, als auch das 3‘-Ende

flankieren, welche für die Integration in das Wirtsgenom und den Replikationsstart verantwortlich sind. Diese ITR behindern einen vollständigen Zusammenbau des viralen Genoms, da sie sich während der Rekombination wie überlappende Sequenzen verhalten und so zu einer Fehlassemblierung führen können. Eine Sequenzoptimierung kritischer Stellen kann einen Zusammenbau größerer Fragmente oder sogar des gesamten Genoms ermöglichen. Um eine Rekombination der ITR zu umgehen, kann ein Vektor entworfen werden, der die Sequenzen der ITR bereits enthält. Nach der Integration können diese, angebaut an die restliche genomische Sequenz, wieder durch Restriktionsverdau mit ausgeschnitten werden. Die Transfektion von Meerschweinchen-Trachea-Zellen mit dem *in vitro* zusammengebauten Volllängekonstrukt hat nicht zur Entstehung von Viren geführt. In dieser Arbeit wurde aber der Grundstein für tiefere Untersuchungen des viralen Genoms und seines Infektionsmechanismus gelegt. Die Schaffung eines Meerschweinchen-Adenovirus-Modells kann dann Antwort auf weitere Fragestellungen liefern, zum Beispiel zur Tumorentstehung nach der Infektion von Nagetieren mit humanen Adenoviren und der Pathogenität im Meerschweinchen Wirtsorganismus.

1 INTRODUCTION

1.1 History of synthetic biology

Today “SynBio” is a set phrase not only in the language use of natural scientist, but also of politicians and any persons who are interested in the development of modern engineering and current achievements of a newly admired market called synthetic biology. But where does it come from and when did it start? Here a short discourse is given on how we came to modern biology. The field of synthetic biology has its roots in the ability to directional engineer cells and DNA. Meanwhile a vast number of small and great companies contribute their work and progress to a worldwide growing market and research area. Starting with automated DNA sequencing in the mid- 1990s, when the first bacterial, archaeal and eukaryotic genomes were fully sequenced (Hutchison 2007), synthetic biology compasses commercial production of synthetic DNA, proteins and engineered organisms from expression optimized mammalian cell lines to humanized mice. The importance to researchers became clear very fast, that the synthetic construction of natural pathways can open a much deeper insight to functional drains in cells and therefor are the key to medical research on human health (Cameron et al. 2014). The modularity and complexity increases steadily since then and powerful tools had been created to modulate the behaviour and output of synthetic genes (Annaluru et al. 2014). Today the common cloning techniques GoldenGate and Gibson assembly (Engler et al. 2008; Gibson et al. 2009; Gibson 2009) are used day by day as a matter of course for researchers and students all over the world. The creation of a bacterium with an entirely synthetic genome (Gibson et al. 2010) in 2010 was a milestone that was built shortly after the publication of the most common cloning methods. As the demand had rose, the prices decreased dramatically, what made many research groups able to afford metabolic engineering. Now, whole metabolic pathways were designed and directed to predict correlations of enzymatic functions within the living biological system of an host cell (Cameron et al. 2014). Recently, the breakthrough of the year 2015 was found in genome editing as announced by Science magazine (Science News Staff 2016). The CRISPR/Cas- (clustered, regularly interspaced short palindromic repeats-CRISPR-associated proteins) system was found and first described by the research group of Emanuelle Charpentier (Jinek et al. 2012), enabling genome-wide transcriptional control. During all the years, small companies had started to pick their way through the market, such as GeneArt GmbH, a small company that was founded in 1999 in Regensburg by

Prof. Dr. Ralf Wagner, Dr. Marcus Graf and Dr. Hans Wolf. Dedicating itself to the synthesis of DNA it is now part of the world leader in serving science, Thermo Fisher Scientific.

1.2 The assignment of SynBio to nowadays

Currently synthetic biology finds its way more and more into the daily newsfeed of non-research society. Reputable newspapers like the German “Frankfurter Allgemeine Zeitung” report and review the chances and risks of this technological field with enlightening articles in a way intelligible to all (Frankfurter Allgemeine Zeitung GmbH 2012) and ask critical questions on ethical issues at the same time. Moreover, the ease of genetic manipulation bothers critics ever since the methods had become popular and protein and DNA technologies directed. Scientists call upon society to improve the formal and informal education (Arno G. Motulsky 1983). Nowadays synthetic biology brings together engineers and biologists to develop environment friendly fuel, find ways to eliminate plastics from our aqua sphere and analyse an individual’s genes to find genetic loading for widespread disease like breast cancer. Industrial scale applications had made incredible steps possible in immunological research and treatment of human disease. Bacteria had been engineered to invade cancer cells, delivering proteins in dependence of environmental signals (Chien et al. 2017). The Golden Rice Project had published its first results after engineering the provitamin A (-carotene) biosynthetic pathway into (carotenoid-free) rice endosperm. By this, millions of affected humans can be saved from a hypovitaminosis resulting in nyctalopia and bone growth disorder in children (Ye 2000). Besides all medical approaches, synthetic biology delivers ingredients for consumable goods like cosmetics, washing detergents or food additives. The Max-Planck-Gesellschaft for synthetic biology reports about the status quo in medicine, chemistry and energy industry. The global SynBio market was valued with 3.02 billion USD in 2016 and is predicted to grow to 8.84 billion USD by 2022 from medicine to agriculture (Synthetic Biology Market by Tool & Technology - Global Forecast 2022 | MarketsandMarkets).

1.3 Fast gene synthesis

1.3.1 Synthetic DNA

Synthetic DNA combines both the creation of nucleic acids and whole artificial genomes. Pathways can be engineered either synthetic or natural by assembling molecules to DNA fragments of large

sizes and even a whole *Mycoplasma genetalium* genome (Gibson et al. 2008). The human genome project, which had the aim of decoding of the complete human genome, is aimed to be completed officially by April 2004. Now it is followed by the human genome project write, which was announced to has the purpose of better understanding the interactions between our genes (Boeke et al. 2016). New methods are developed with great variety in approaches such as the DATEL (DNA Assembly with Thermostable Exonuclease and Ligase), a method promising to be independent from sequence and applicable to manual and automated high-throughput assembly of DNA fragments (Jin et al. 2016). The combination of computational models that can describe *in vivo* happening biological processes give us the opportunity to increase our knowledge on all sorts of applications feasible. But to accomplish these requirements, manufacturing of the raw product has to be improved steadily to make it affordable and competitive.

In this thesis the DNA fragments are constructed in a production system that starts with sequential chain reaction (SCR). Partly overlapping oligonucleotides (L, m) as shown in Figure 1 are assembled to one double strand fragment by a polymerase. The resulting product is the templated used for multiplication in a further standard polymerase chain reaction (SPCR) step that includes primers annealing to the construct to make the polymerase enzyme extend the strand. After this process multiple copies of the sub-fragment are produced.

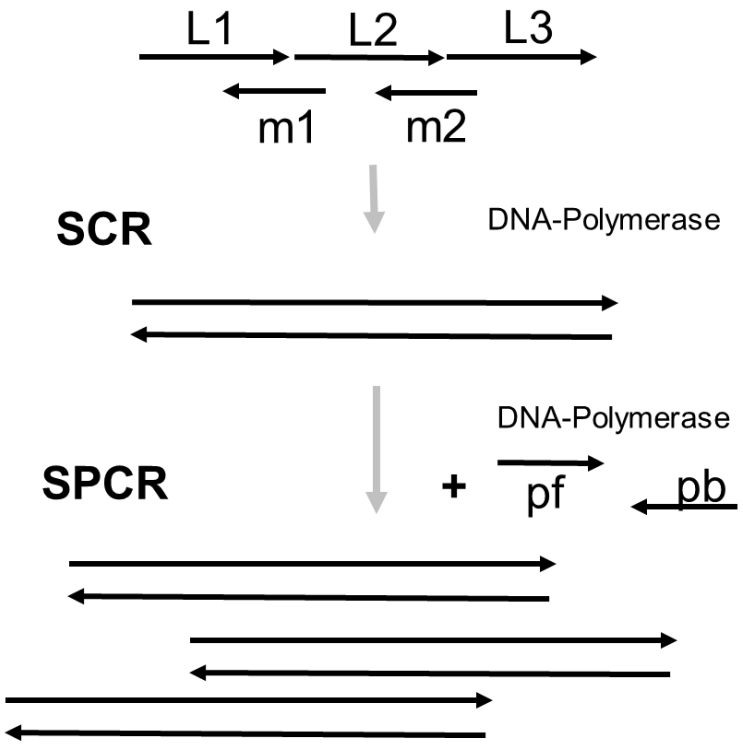


Figure 1: Standard process of gene synthesis from oligonucleotides to multiple fragments; SCR = standard chain reaction; SPCR = standard polymerase chain reaction; L/m = oligonucleotides; pf/pb = primer forward and backward

These fragments then undergo an enzymatic error correction before using them for further progress. In a third reaction, the fusion polymerase chain reaction (FPCR), multiple sub-fragments that match by an overlapping region are assembled to on gene construct.
For the creation of large scale DNA production in longer size methods like Gibson assembly, seamless cloning or Golden Gate (Gibson et al. 2008; Kok et al. 2014; Engler et al. 2008; Gibson et al. 2009) are set routinely in academic and industrial research. Still all of these methods have their limits on sequence base, which can be restriction sites, repetitive areas or toxicity. A further method, yeast assembly, is a remarkable alternative to those mentioned before with its ability to simply use the recombination capabilities of *Saccharomyces cerevisiae* as assigned in this thesis (Gibson et al. 2008; Sherman 2002).

1.3.2 Needs in industry and research respectively applications

Synthetic biology is a rapidly developing field, where synthetic DNA and modified synthetic genes are the essential tool. Hence, the attention of scientific research has shifted from the synthesis of oligonucleotides and genes to their application. Parallel to the speed in research and production the DNA has to be delivered as fast and reliable in highest quality. In March 2017 the reputable newspaper “Zeit” reports on its online feed (Löfken 2017) about hard-drives made of DNA. With their strategy „DNA Fountain“ (Erlich and Zielinski 2017) the group of researchers around Yaniv Erlich announced in Science magazine the ability to store a computer operating system, a movie and other files with a total of 2.14×10^6 bytes in DNA oligonucleotides and to completely be able to decode the data again. These breaking news disclose a prospect of technical applications for DNA in the near future. Back to natural science, medicine already established methods to fight severe cancerous diseases by using artefacts derived from synthetic DNA. The approach is to create vaccines that utilize tumour mutations can elicit the immunological T-cell response by introducing synthetically generated long peptides, DNA or RNA into the body of patients (Melief 2017). Ott et al. had shown that personalized vaccines targeting neoantigens (tumour specific antigens) can lead complete regression of tumours in patients (Ott et al. 2017). It is from a high importance, that the sequences of these neoantigens are produced fast and dependable. The improvements on the underlying gene synthesis workflow in this thesis, had targeted this requirements in turnaroundtime of production and reliability.

1.4 Adenoviruses

1.4.1 Adenovirus structure and function

Adenoviruses are ubiquitous, non-enveloped, double-stranded DNA viruses (Rux and Burnett 2004). The prefix “adeno” means being derived from gland, which comes from the fact that they were first isolated from human tonsils. An essential process in gene expression, called splicing was discovered first in this group of viruses, the early transcription of the viral DNA is established by the hosts RNA-polymerase. These primary transcripts contain introns and appear in a very high number, so that the investigation of eukaryotic splicing is based on this mechanism. Today, more than 100 members of the family Adenoviridea have been described that can infect humans and other mammals, birds, reptiles, amphibians, and even fish (Kaján et al. 2012). Human adenoviruses encompass more than 70 different types that are resumed into six species sub-groups (A – F) by reference to their ability of agglutination of erythrocytes (Madigan et al. 2013). The typical adenoviral icosahedral structure is shown in Figure 2.

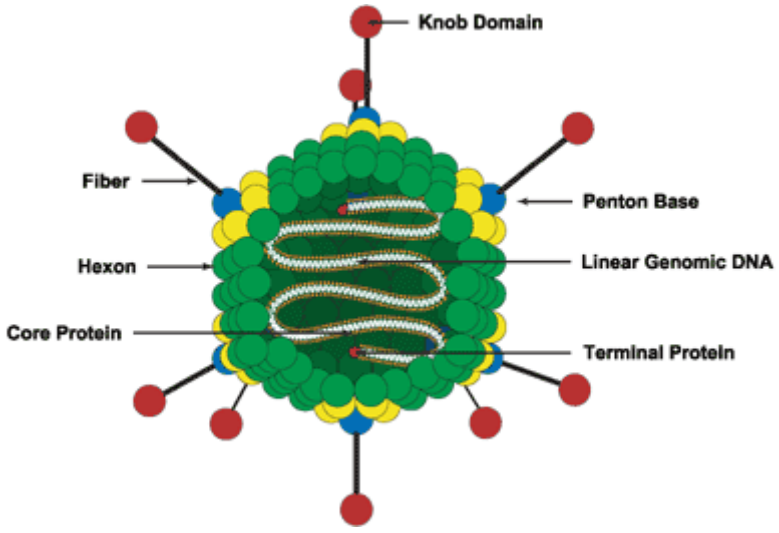


Figure 2: Schematic structure of adenoviruses; www.coral-club-eu.com

Its capsid can be 90 – 110nm in size (Rux and Burnett 2004) consisting of up to 252 capsomeres and 240 hexons. Hexons are the major capsid component and play the essential role in induction of immune response. 12 pentons are located around the capsid on 12 edges, comprised from a penton base to stabilize the capsid and the fiber glycoproteins that consist of tail, shaft and knob. This complex provides stability by burying hydrophobic surfaces. Fibers are the first viral components that interact with epithelial cells in tissue during infection. Polypeptides VI, VIII and IX keep the hexon capsomeres packed. The core of the virion contains four polypeptides, V, VII, X

and the terminal protein (TP). V is building the connection to the capsid and VII is playing a histone-like role. The genome of Adenoviruses is about ~36 kb in size and has at its 5' terminal end a covalently bound protein that may facilitate the circularization of the viral genome during replication (Russell 2009) together with a distinct feature, the inverted terminal repeats (ITR) that can range from 100 – 1800 bp depending in the genus (Doerfler 1996).

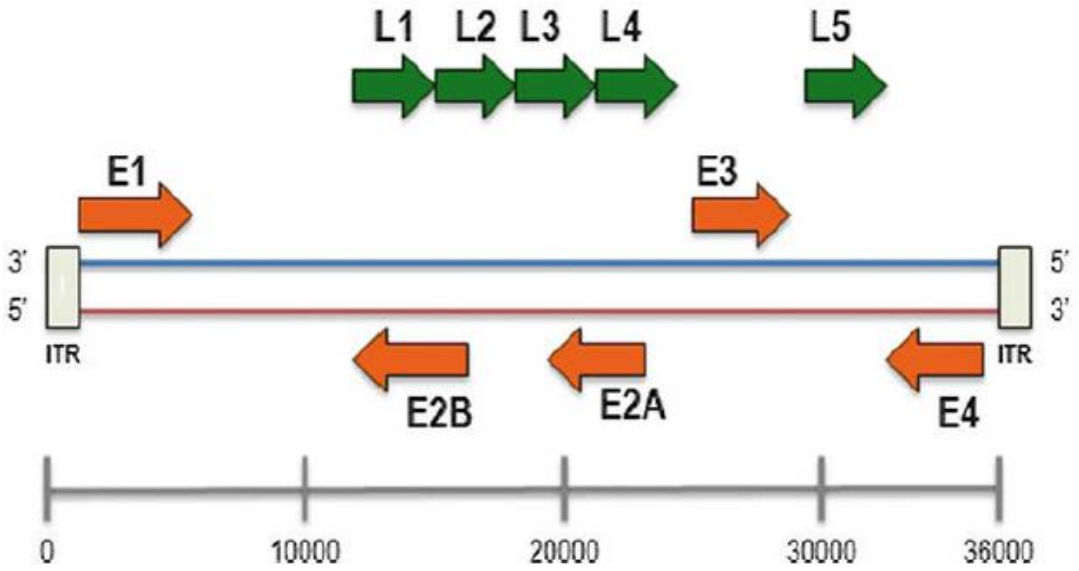


Figure 3: Schematic map of the adenovirus genome: shown here is the Ad5 genome including the early proteins E1–4, the late proteins L1–5, and the terminal repeats (ITR) as they are located along the linear genome; (Milavetz and Balakrishnan 2015).

Figure 3 shows an overview of the genetic map of adenovirus on the example of human Ad5. First expressed during viral replication are the so called early genes E1 (A, B), E2A, E2B, E3 and E4 (orange arrows) that encode regulatory proteins. Both, E1 and E4 lacking viral mutants are not able to replicate (Blackford and Grand 2009). In the later phase when replication has started the late genes L1 – 5 (green arrows) encoding structural proteins are expressed, leading to the production of more viruses and cell death (Wilson 1996). Infection by adenoviruses starts with binding to an extracellular receptor and the transport of the virus into the cell. During transport to the nucleus, the viral genome is unpacked (Milavetz and Balakrishnan 2015). Replication of the genome takes place in the viral core. Early proteins of adenoviruses regulate the DNA replication, especially the terminal protein as it works like a primer inducing the process with a covalently bound cytosine residue. The replications results in a double stranded and a single stranded product. Special among this viral mechanism is the forming of a ring structure of the single strand product through annealing of the ITR regions that leads to the synthesis of a new complementary strand from 5' terminal end (Madigan et al. 2013).

1.4.2 Diseases of adenoviral infection

Adenoviruses are known to cause a range of diseases like acute respiratory disease, pneumonia, hepatitis, hemorrhagic cystitis, colitis, pancreatitis, meningoencephalitis, and disseminated disease (Echavarría 2008). Viral infections among adults are mostly being self-limiting, but can persist asymptomatically for years (Garnett et al. 2002). Early epidemiological studies have shown that around 90% of six-year-olds are seropositive for at least one type of human adenovirus and that the percentage of seropositive individuals is close to 100% in adults (D'Ambrosio E et al. 1982). Most of the patients are children between one and five years old and represent 5 – 7 % of the respiratory tract infections among pediatric patients. Especially children receiving hematopoetic stem cell transplantation (HSCT) are at high risk of adenoviral dissemination in which mortality rates occasionally exceed 50% (Kojaoghlanian et al. 2003). Epidemic keratoconjunctivitis is an ocular surface infection caused by HAdV (human adenovirus) and can only be treated for the symptoms with limited antiviral drugs (Ghebremedhin 2014). The most commonly drug cidofovir is a cytosine analogon which serves as a substrate for the adenoviral polymerase, is incorporated into viral DNA and thereby blocks DNA replication (Lenaerts and Naesens 2006).

There are three ways for adenoviruses to infect the host cells. The first one is the lytic infection of epithelial cells, where new viruses are produced and the cell dies. The second one is a latent infection of lymphoid cells, where less virus is produced and the cell death deposed compared to the lytic infection. The third way is the oncogenic infection. Here, the DNA is integrated into the host genome and replicated without the production of new infectious virions (Ghebremedhin 2014). It is known that adenoviruses can cause undifferentiated sarcomas in rodents, but currently there is no evidence, that adenoviruses are oncogenic in humans (Doerfler 1996). Still this occurrence and the mechanism behind are of high interest to be determined more closely.

1.4.3 Adenoviruses in industry and medicine

The approach of using genes for disease therapy requires a vector that transports the certain gene into the desired target cell, functioning as a vehicle. Adenoviruses are meanwhile commonly used as vector in gene therapy. The first evaluations were done on treatment of cystic fibrosis, an autosomal recessive disorder caused by mutated cystic fibrosis transmembrane conductance regulator (CFTR) protein, in epithelial lung cells. Very accommodating is the virus' property to deliver a high number of recombinant viruses so that the initial propagation is omitted. In the

following, the systematic of viral vaccination is shown on the example of treatment of cystic fibrosis.

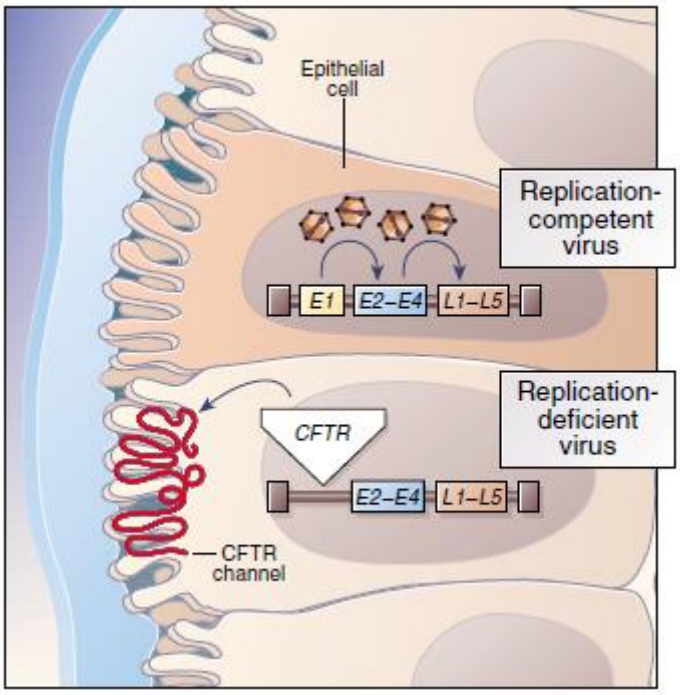


Figure 4: Adenoviral vector on the example of cystic fibrosis: Here a replication competent virus is shown in contrast to the modified one lacking the early gene region E1 carrying the cystic fibrosis transmembrane conductance regulator (CFTR) protein (Wilson 1996).

In the upper part of Figure 4 a regular virus is shown that is able to replicate in the infected lung epithelial cell. Below that, the artificial adenoviral vector is shown, that lacks the early region E1 gene and carries the gene for CFTR channel instead. The deletion of E1 hinders the virus from replicating and thus from multiplying and leading to cell death. Using Adenoviruses as vectors requires of course some changes to its genomic map. Thus, the E1 regions are deleted in the adenoviral vectors of first generation. Some vectors also lack E3 to have more space for foreign genetic material to be placed between the ITRs. E4 gene products are also involved in the same cell cycle regulatory key mechanisms acting independently from E1A/B. Thus, to build adenoviral vectors, these sequences have to be deleted (Täuber and Dobner 2001). The second generation is therefore lacking E1 regional genes together with the E4 genes. A third generation was developed lacking all viral genes and thus having a high capacity for foreign DNA as there are only the terminal regions including the ITRs left (Volpers and Kochanek 2004).

Further examples show the worldwide impact of adenoviral vectors. Duchenne muscular dystrophy is a severe disorder linked to the X chromosome, where mutations occurred present in the dystrophin gene. It was shown in animal models, that the directed gene transfer of utrophin

mediated by adenoviruses mitigates the disorder (Cerletti et al. 2003). The Ebola virus outbreak in Africa between 2014 and 2016 led to the investigation of the usability of adenoviruses as vaccines against Ebola disease. The adenoviral vaccines were successful in guinea pigs and primates by expressing a variant of the antigenic glycoprotein. The first clinical trials followed and showed the safety and immunogenicity of an Ad5-vectored vaccine expressing the glycoprotein Makona in humans (Wu et al. 2016). These examples urge the production of the antigen genetic material in particular.

1.4.4 Guinea Pig Adenovirus

The Guinea Pig Adenovirus (GPAdV) bothers animal facilities worldwide. Once the virus has infected one animal it is easily transmitted through direct contact leading to severe bronchopneumonia and death of immunocomprised and juvenile animals (Butz et al. 1999). The virus is usually found in nuclei of epithelial cells in guinea pig lung tissue (Charles River Research Animal Diagnostic Services). There were two outbreaks documented, where the infected animals were no longer suitable for pulmonary research. Indeed, the virus was detected by histopathological investigations, but it was yet not possible to isolate the whole virus or keep it in cell culture stock. By PCR, the genome of the GPAdV was selectively verified which made it possible to re-construct the sequence (Butz et al. 1999). The viral pathogenesis was investigated by infection of guinea pigs. It revealed, that the infection did not lead to clinical signs in all cases, but ended up in high mortality. The first outbreak was captured by the group of Susanne Naumann 1981 in Hannover, Germany. Electron microscopy showed the virus particles isolated from lung tissue, but once the pneumonia infection occurred it was time-limited so that no isolation of viruses was possible (Naumann et al. 1981). Two years later, the same group announced that the infection was reproducible in new born guinea pigs, but not in adults. Now an incubation period from 5 – 10 days was detected and experiments showed that it was not oncogenic in rodents. The studies indicated that it is about a virus specific for guinea pigs, when the cases appeared spontaneously and the virus did not cross react with other adenoviruses either human or fowl. Until today, there is very limited access to small animal models on GPAdV, as the common mouse model is not permissive for the infection. Thus, the guinea pig model is the only model in which virus induces pneumonia in its natural host. The isolate used in this thesis is called GER1 and is available as an *in silico* sequence provided by Prof. Thomas Dobner and Dr. Helga Hoffmann-Sieber from the Heinrich-Pette-Institut located in Hamburg, Germany.

1.5 Aim of the thesis

1.5.1 Fast gene synthesis to serve speed on markets and research

In this thesis, new PCR protocols on an abridged gene synthesis workflow were developed. The growing market of synthetic biology comes along with new needs on applications and fast access to products. Research is quickened by the rise of new intelligent and automated technologies like next generation sequencing and the computational support of robots in modern industry. The closer researchers can have a look into new fields, the faster they want to investigate their recent discoveries. Of course, the technology of building DNA fragments is not a novel one and works sufficient for the acute purpose, but still improvement needs to be done to serve the upcoming claims. Rather than following economic interests, the fastening on gene synthesis serves new technologies in medical treatment of patients. In 2017 nature published an article that shows how promising the fast development of innovative therapies are contributing to humans contracted from cancerous tumours (Melief 2017). To be able to improve and further implement this DNA vaccination approach, the fast detection and delivery of neoantigens brought along with adenoviral vectors is essential. Here, the assembly of oligonucleotides into DNA fragments shall be fastened accompanied by reliability of a correct quality amplification product of PCR in gene synthesis.

1.5.2 Large fragment DNA synthesis of wild type sequences on the example of Guinea Pig Adenovirus

Synthetic viruses are a new opportunity to understand and prevent infectious diseases and oncogenic tumour development. The Guinea Pig Adenovirus is still not suitable to be kept cultivation and availability of natural host models is very limited. The sequence could be ascertained and compiled to a full genome and annotated with all exact locations of structural genes and other features by transcriptome analysis. Therefor the genome has to be synthesized in as large fragments as possible that can be transfected into a guinea pig derived cell line. In this thesis, protocols shall be developed on the assembly of large DNA fragments that are not sequence optimized and thus keep the original codon structure. The mechanism of homologous recombination from *Saccharomyces cerevisiae* combined with a computational partitioning approach shall deliver double stranded functional DNA constructs of the wildtype Guinea Pig Adenovirus genome.

2 MATERIAL, METHODS AND TOOLS

2.1 Material

2.1.1 Bacteria

Strain	Organism	Company
DH5α	*Escherichia coli*	Thermo Fisher
DH10B	*Escherichia coli*	Thermo Fisher
TransforMax™ EPI300™	*Escherichia coli*	Lucigen EC300110
One Shot™ OmniMAX™ 2 T1R Chemically Competent	*Escherichia coli*	Invitrogen™
Vmax	*Vibrio natriegens*	SGI

Table 1: Bacterial cell lines

2.1.2 Yeast

Strain	Organism	Company
Y187	*Saccharomyces cerevisiae; MATα, ura3-52, his3-200, ade2-101, trp1-901, leu2-3, 112, gal4Δ, met–, gal80Δ, MEL1, URA3::GAL1UAS - GAL1TATA-lacZ*	Takara Bio Inc.

Table 2: *Saccharomyces cerevisiae* cell line

2.1.3 Mammalian cells

Strain	Organism	Vendor
GPTEC-T	*Cavia porcellus*; Guinea Pig Trachea Epithelial Cells – large T Antigen	Prof. Dr. Adriana Kajou, Albuquerque, New Mexico

Table 3: Guinea Pig cell line

2.1.4 Plasmids

Plasmid	Used in	Vendor
pUC	Transformation of *E. coli*	Thermo Fisher
pYES8D_GFP	Transformation of yeast	Self-construction; backbone provided by Chang-Ho Beak (Baek et al. 2015) Thermo Fisher

Table 4: Plasmids

2.1.5 Vectors

Vector	Features	Vendor
pMA	ccdB, ColE1ori, Amp-resistance	Internal Thermo Fisher GeneArt
pEYES	repE, Cen6-ARS4, His3 marker, CAT marker, oriV,	Internal Thermo Fisher GeneArt
pYES8D	pUCori, Amp-resistance, deleted Trp marker, deleted 2µ-ori	Chang-Ho Beak (Baek et al. 2015) Thermo Fisher

Table 5: Vectors

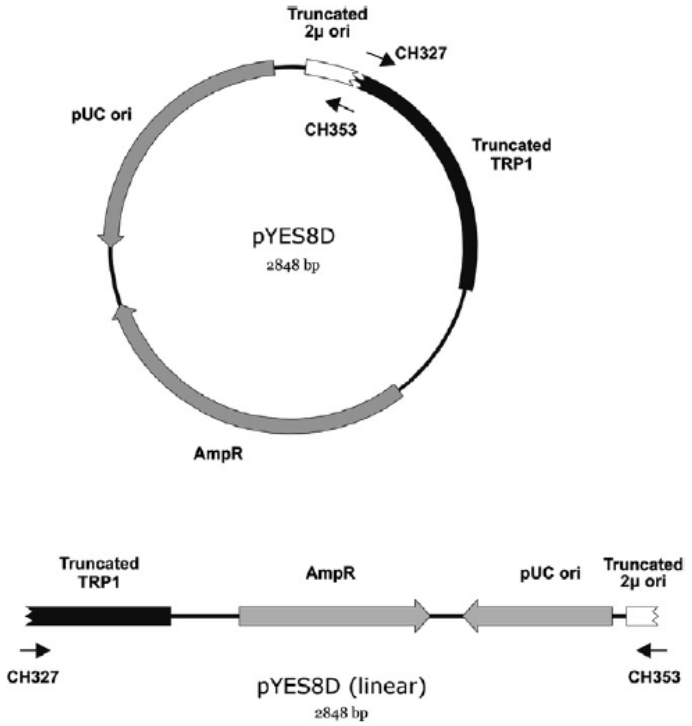


Figure 5: pYES8D map circular (Baek et al. 2015)

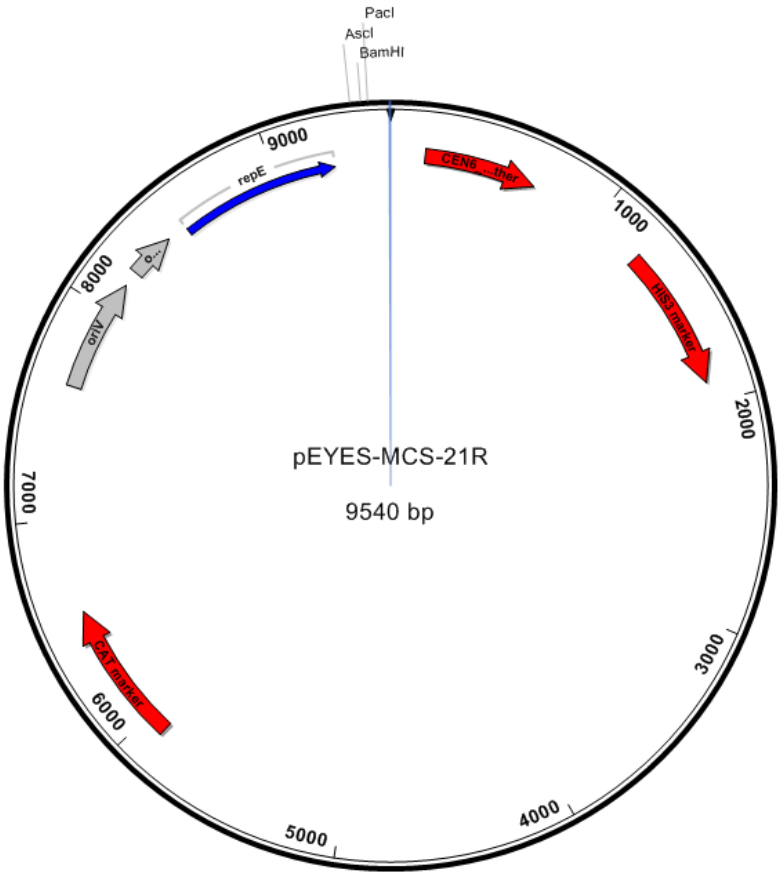


Figure 6: pEYES map circular; Seqbuilder DNASTAR

2.1.6 Media

2.1.6.1 Liquid media

Media	Company
EMEM Eagles Minimum Essential Medium	ATCC®
LB	In-house GeneArt GmbH
SOC	Invitrogen™
TB	In-house GeneArt GmbH
YPD	Gibco™

Table 6: Liquid media

2.1.6.2 Agarose media for yeast plates

Plates for *Saccharomyces cerevisiae* were prepared, autoclaved and then provided with sterile filtered glucose. For every plate 15 ml of liquid, hot agarose medium was filled into 8.5 x 1.5 cm sterile petri dishes. After the medium hardened out the plates were stored top down at 4 °C.

Supplement	Amount	
Yeast nitrogen base	6.7 g	
Agar Agar	20 g	
Yeast Synthetic Drop-out Medium Supplement	1.92 g	Without histidine/ tryptophan
Glucose 40 % sterile filtered	50 ml	
H_2O desalted	to 950 ml	

Table 7: Agarose media

2.1.7 Enzymes and Kits

2.1.7.1 Polymerases

Polymerase	Company	Catalogue number
Phusion High-Fidelity DNA Polymerase (2u/µl)	Thermo Scientific™	F-530L
Platinum™ PCR SuperMix High Fidelity	Invitrogen™	12532024
Platinum™ SuperFi™ DNA Polymerase (2u/µl)	Invitrogen™	12351050

Table 8: Polymerases

2.1.7.2 Restriction enzymes

Enzyme	Recognition site	Company	Catalogue number
FastDigest BamHI	5' G↓G A T C C 3' 3' C C T A G↑G 5'	Thermo Scientific™	FD0054
FastDigest PacI	5' T T A A T↓T A A 3' 3' A A T↑T A A T T 5'	Thermo Scientific™	FD2204
FastDigest SgsI	5' G G↓C G C G C C 3' 3' C C G C G C↑G G 5'	Thermo Scientific™	FD1894
FastDigest XhoI	5' C↓T C G A G 3' 3' G A G C T↑C 5'	Thermo Scientific™	FD0694

Table 9: Restriction enzymes

2.1.7.3 Kits

Kit	Company	Catalogue number
Gene JET Gel Extraction and DNA Clean-up Micro Kit	Thermo Scientific™	K0692
High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems™	4368814
NEBuilder® HiFi DNA Assembly Cloning Kit	New England Biolabs®	E5520S
PureLink™ Quick Plasmid Miniprep Kit	Thermo Scientific™	K210011

Table 10: Kits

2.1.8 Buffers and standards

Buffers	Company	Catalogue number
10x loading buffer	Invitrogen™	10816015
FD buffer	Thermo Scientific™	B64
Phusion High GC buffer	Thermo Scientific™	F519L
Phusion High-Fidelity buffer	Thermo Scientific™	F518L
Phusion High-Fidelity buffer detergent free	Thermo Scientific™	F520L
Platinum™ SuperFi™ buffer	Invitrogen™	12355005

Table 11: Buffers

Standards	Company	Catalogue number
1 kb DNA ladder	New England Biolabs®	N3232L
1 kb Plus E-Gel ladder	Invitrogen™	10488090
1 kb Plus ladder	Thermo Scientific™	SM1334
Qubit™ 1X dsDNA HS Assay Lambda Standard	Invitrogen™	Q33233

Table 12: Standards

2.1.9 Chemicals, equipment and consumable material

Chemicals	Company
Adenine hemi-sulphate	Thermo Scientific™
Agar-Agar	Carl Roth®
Agarose	Invitrogen
Ampicillin	Carl Roth®
Arabinose	Sigma-Aldrich®
CaCl2	Carl Roth®
Chloramphenicol	Carl Roth®
DMSO	Sigma-Aldrich®
dNTP mix 100 mM	Thermo Scientific™
EtBr	PanReac AppliChem
Ethanol	Carl Roth®
Fetal bovine serum	Gibco™
Glucose	Carl Roth®
Guanosine	Carl Roth®
Kanamycin	Carl Roth®
Lipofectamine 2000	Invitrogen™
Lithium acetate	Sigma-Aldrich®
NaAc sodium acetate	Carl Roth®
NaOH sodium hydroxide	Carl Roth®
PEG 3350	Sigma-Aldrich®
Penicillin-Streptomycin	Thermo Scientific™
Salmon Sperm DNA	Invitrogen™
TE-buffer	Sigma-Aldrich®
Tris	Carl Roth®
Trypsin-EDTA	Gibco™
Yeast nitrogen base	Sigma-Aldrich®
Yeast synthetic Drop-out supplements	Gibco™

Table 13: Chemicals

Equipment	Company
AcroPrep™ Advance 96-Well Filter Plates for Ultrafiltration	PALL Laboratory
Biophotometer	Eppendorf
Bunsen burner	Campingaz
Centrifuges 5424, 5430	Eppendorf
E-Gel 10 sample 1% Agarose	Invitrogen™
E-Gel 10 sample 2% Agarose	Invitrogen™
E-Gel 48 sample 2% Agarose	Invitrogen™
E-Gel base	Invitrogen™
Electroporator Micropulser	Biorad
Gel electrophoresis chamber	Biorad
Incubators	Thermo Scientific™
Magnetic stir apparatus	Heidolph
Microwave	Panasonic
Qubit 4	Thermo Scientific™
Shaking incubator HAT Multitron	Infors AG
Thermocycler Mastercycler Pro S	Eppendorf
Thermomix comfort	Eppendorf
Vortex Genie 2	Scientific Industries Inc.

Table 14: Equipment

2.2 Methods

2.2.1 Culture and treatment of cell lines

2.2.1.1 Yeast

2.2.1.1.1 Inoculation and growth of *Saccharomyces cerevisiae*

Yeast cultures were grown from single colonies picked from fresh selective-medium plates, at 30 °C on YPD media. For pre-cultures single colonies 5 ml of YPD medium were inoculated and grown for 16 hours. High accretion was achieved by vigorously shaking 150 rpm in a 50 ml Falcon tube. From this culture 3 ml were inoculated to 30 ml YPD and further incubated shaking until cells reached a maximum optical density OD_{600} = 2.5 in YPD.

2.2.1.1.2 Preservation of *Saccharomyces cerevisiae*

Yeast strains have been stored for short periods of time at 4 °C, on YPDA medium in Petri dishes. Passages of the stock were prepared in 4 weeks intervals. Yeast strains can be stored in 40% (v/v) glycerol at –80 °C. To preserve yeast-strains indefinitely, a fresh YPD-liquid culture of cells grown in appropriate media is mixed with sterile glycerol. The caps are tightened and the vials shaken before freezing. Transferring a small portion of the frozen sample to an YPD plate can revive yeast for colony forming (Sherman 2002).

2.2.1.1.3 Transformation with PEG/LiAc

The following protocol for transformation of yeast is based on the classic method described by (Gietz and Schiestl 2007). A single colony of yeast Y187 was inoculated to 5 ml of liquid YPD medium and incubated at 30° shaking at 150 rpm for 16 h. To determine the OD_{600} 100 µl of the pre-culture were added to 900 µl YPD, mixed well and measured on a spectrophotometer. For blank 1 ml of YPD was used. Afterwards 2.5 $x10^8$ cells were added to 50 ml of pre-warmed YPD medium and again incubated for 4 – 5 h until the OD_{600} reached at least 2. The cultures was harvested in a 50 ml conical tube by centrifuging at 3,000 g for 5 minutes. The cell pellet was first washed with 10 ml sterile H_2O and then with 10 ml 0.1 M LiAc and again pelleted. To proceed with the transformation the cells were re-suspended with 0.1 M LiAc to a concentration of 10^8 cells/100 µl. For each transformation 100 µl of cell suspension were transferred to 301 µl of transformation mix (see Table 15) and vortexed.

Transformation mix	Volume in µl
PEG 50% (w/v) in sterile H_2O	240
LiAc 1 M	36
SS DNA (10mg/ml)	25
Total volume	301

Table 15: Transformation mix for *saccharomyces cerevisiae* with LiAc/PEG method

Afterwards the DNA/Plasmid mix was added. For an assembly reaction 300 ng of each linear insert and 500 ng of linearized target vector were combined. In every set up, one positive control containing 500 ng plasmid DNA, one negative control containing 500 ng of linearized target vector and one negative control containing no DNA were included. The cell suspension was incubated at 30 °C for 30 min and then placed at 42 °C for heat-shock after adding 36 µl of DMSO. After 15 min the cells were placed on a cool rack for 2 min and then pelleted at 4,000 rpm for 45 sec. The pellet was diluted in 100 µl of sterile H_2O for plating on YPD amino acid selection plates for auxotrophic selection. The plates were incubated at 30 °C for 3 - 4 days.

2.2.1.1.4 Plasmid DNA Preparation from *Saccharomyces cerevisiae*

For preparation of plasmid DNA from *Saccharomyces cerevisiae* a single colony was picked from a plate with a sterile pipet tip and diluted in 50 µl of water. The suspension was directly plated on appropriate YPDA plate for auxotrophic selection. After 48 h a confluent cell layer had grown and was harvested from the plates by adding 2 ml of water on the surface, suspension with a cell spreader and transfer to a 2 ml reaction tube. The cells were pelleted by centrifugation at 12,000 rcf for 1 min.

The plasmid preparation was performed with the Purelink Plasmid MiniPrep Kit with using a modified centrifugation protocol. After dissolving the yeast pellet in 250 µl R3 buffer containing RNase. 50 µl of lyticase solution containing 25 u of enzyme were added, mixed well and incubated for 45 min at 37 °C. In addition, 100 µl volume of glass beads were added after incubation to the reaction and vortexed for 1 min. The suspension was then cooked with 250 µl of L7 buffer for 10 min at 95 °C and then quickly chilled on a cool metal rack before 350 µl of N4 buffer were added. The precipitation was performed for 30 min on ice. To proceed with the plasmid solution, the mixture was centrifuged for 10 min at 12,000 rcf and the supernatant was transferred on a spin column while being careful to not bring any of the precipitate along. Another 1 min centrifugation step was taken before performing two washing steps with 700 µl W9 buffer containing ethanol. The column was then placed on a sterile 1.5 ml collection tube and eluted in 75 µl H_2O.

2.2.1.2 Bacteria

2.2.1.2.1 Heat shock transformation of competent bacteria

For transformation competent *E. coli* DH10B cells were used. The cells were thawed on ice and incubated 25 min on ice after mixing 2 µl (150 ng) of plasmid DNA to 100 µl bacteria. Then a heat shock was performed for 45 sec at 42 °C and the cells were immediately transferred back on ice for 2 min. 900 µl of pre-warmed SOC medium was added to the cells before they were incubated for 1 h at 37 °C with shaking at 350 rpm to let them grow. The suspension was centrifuged for 30 sec at 12,000 rcf, room temperature and then 900 µl of the supernatant were removed. The bacteria pellet was re-suspended in the remaining 100 µl and plated on LB_{Cam} plates for incubation over night at 37 °C.

2.2.1.2.2 Electroporation of competent bacteria

For the transformation of electrocompetent DH10B *E. coli* cells 100 ng of DNA were diluted with H_2O to a final Volume of 5 µl. Each reaction was performed with 20 µl cells (conforming on portion per reaction). The cells were thawed on ice and the DNA added gently to avoid bubble formation. The cell/DNA mixture was transferred into a sterile and pre-chilled 0.1 cm cuvettes. The cuvettes were placed in the appropriate device of the micro pulser and the program was set to “Ec1”. The suspension was then pulsed 1x with 1.8 kV and the ms tracked. The samples were in an appropriate range from 4.5 – 5 ms. The cuvettes were rinsed with 500 µl SOC medium immediately after pulsing to take up the cells. The suspension was transferred into a 1.5 ml reaction tube and placed on a thermomixer for incubation at 37 °C for 1 h. Following, the cells were directly plated onto a LB plate with appropriate antibiotics for selection.

2.2.1.2.3 DNA Mini-Preparation

For preparation of plasmid DNA from *Escherichia coli* a single colony was picked from a plate with a sterile pipet tip and inoculated to 5 ml of LB medium containing appropriate antibiotics and supplements such as 0.02% arabinose. After 16 h shaking with 350 rpm at 37 °C the cells were pelleted by centrifugation at 12,000 rcf for 1 min. The samples containing plasmid construct B2 were grown on 30 °C.

The MiniPrep was performed with the Purelink Plasmid MiniPrep Kit with using the centrifugation protocol. After dissolving the yeast pellet in 250 µl R3 buffer containing RNase the suspension was mixed with 250 µl of L7 buffer and incubated for 10 min at RT. For precipitation 350 µl of N4 buffer was added and the tube gently inverted to have the suspension mixed. To proceed with the plasmid solution, the mixture was centrifuged for 10 min at 12,000 rcf and the supernatant was transferred on a spin column while being careful to not bring any of the precipitate along. Another 1 min centrifugation step was taken before performing two washing steps with 700 µl W9 buffer containing ethanol. The column was then placed on a sterile 1.5 ml collection tube and eluted in 75 µl TE buffer.

2.2.1.2.4 DNA Midi-Preparation

Plasmids isolated by Midi preparation was performed from 15 ml inoculated TB medium, containing appropriate antibiotic chloramphenicol (25 µg/ml) and supplemented with 0.02% arabinose. The procedures were performed by the local department at Thermo Fisher Scientific GeneArt GmbH in Regensburg.

2.2.1.3 Mammalian

2.2.1.3.1 Growth and preservation of GPTEC-T cells

The Guinea Pig derived cells were grown in EMEM medium containing 10% FCS and 1% Penicillin-Streptomycin. Every other day the cells were washed with PBS and dosed with fresh medium.

2.2.1.3.2 Transfection using Lipofectamin2000

Fresh GPTEC-T cells were harvested from a full grown 14 cm culture plate. For this, all medium was removed and the attached cells washed with 5 ml PBS. By adding 5 ml Trypsin solution the cells detached from the surface after 5 min incubation at 37 °C. The cells were taken up with another 10 ml of medium and transferred to a 15 ml Falcon tube to be collected by centrifugation at 4500 rpm for 5 min at room temperature. The supernatant was discarded and the pellet carefully re-suspended in 5 ml EMEM medium/10% FCS and the cells counted in a Neubauer chamber.

For every transfection sample 1 ml containing 2.5 x 10^5 cells was seeded in a 6-well culture plate. To prepare the transfection mix two 1.5 ml reaction tubes were prepared with 150 µl EMEM medium without supplements and either plasmid DNA or the appropriate amount of Lipofetamin^{2000} (twice the amount of DNA in µl). A mock control with only Lipofectamin^{2000} was carried along and treated the same as the other samples. The tubes were snapped to have the liquid inside mixed properly and then quickly spun down before the tubes DNA containing were combined with the complementary Lipofectamin^{2000} ones. After 20 min of rest at room temperature, the mixtures was slowly dropped onto the seeded cells. After 5 hrs the medium was changed to get rid of the remaining Lipofectamin^{2000}. The empty wells in the culture plate were filled with non-transfected cells to control the normal cell viability. The cells were harvested for processing after 48 hrs.

2.2.1.3.3 RNA isolation from GPTEC-T

The transfected GPTEC-T cells were harvested by removing the medium and washing them with 1 ml of PBS before the detaching of cells was induced with 0.5 ml Trypsin. With trypsin on top the plate was incubated at 37 °C for 5 min. By pipetting the liquid up and down the cells were rinsed from the wells bottom, transferred to a 1.5 ml reaction tube and pelleted. The cell pellet was washed with 0.5 ml of PBS and pelleted again. To every pellet derived from one well, 1 ml Trizol was pipetted on top and the cells were re-suspended with the pipet before leaving them for 5 min at room temperature. 200 µl of Phenol-Chloroform-Isoamylalcohol were added with care and the closed reaction tubes shaken to promote phase separation. After incubation at room temperature for 3 min the samples were centrifuged at 12,000 g for 15 min at 4 °C. The aqueous phase was clearly visible and transferred carefully (about 400 µl) into an RNase free 1.5 ml reaction tube containing 600 µl of isopropyl and incubated again for 10 min at room temperature. The centrifugation step was repeated as before and the supernatant was discarded. The remaining RNA pellet was washed with 1 ml of – 20 °C cold ethanol and the centrifuged for 15 min at 4 °C and 7500 g. Again the supernatant was discarded completely until no alcohol was left. The remaining pellet was taken up in 20 µl of RNase free H_2O and stored at -80 °C for further processing.

2.2.1.3.4 Reverse transcription of RNA to cDNA

To transcribe the isolated RNA (see 2.2.1.3.3) into single stranded RNA the High Capacity cDNA Reverse Transcription Kit (see Table 10: Kits) was used. The reaction was performed without RNase inhibitors and set up as seen below.

Master mix:	Volume per reaction	Component
[µl]	2	10x RT Buffer
	0.8	25x dNTP Mix (100 mM)
	2	10x RT random Primers
	1	MultiScribe™ Reverse Transcriptase
	3.2	RNase free H_2O
∑	10	

Table 16: Reaction mix reverse transcription

The master mix was prepared on ice after kit components had been thawed on ice as well. To always 10 µl of RNA sample another 10 µl of master mix was added and mixed well by pipetting up and down. The samples were placed in a thermocycler and to be run with following protocol.

Reverse transcription protocol:	
Temperature	Time
25 °C	10 min
37 °C	120 min
85 °C	5 min
4 °C	∞

Table 17: Cycler protocol for reverse transcription

The samples were kept on – 80 °C until further processing.

2.2.2 DNA processing

2.2.2.1 Polymerase chain reaction

2.2.2.1.1 Sequential chain reaction

For the production of DNA fragments in purpose of reducing time the oligonucleotides were assembled and elongated with a shortened sequential chain reaction (SCR) protocol. Therefor a master mix was prepared as follows per reaction in Table 18:

Master mix SCR:	Volume per reaction	Component
[µl]	10	5x Phusion buffer detergent free
	1	Phusion HF DNA polymerase
	1	dNTPs (10 mM each)
	15	H_2O
∑	27	

Table 18: SCR/SPCR Master mix with Phusion polymerase.

The master mix was kept cool until further use. The oligonucleotide mix consisting of overlapping single strand DNA molecules was prepared for the reaction ending up at a concentration of 0.15 µM. The reaction was set up as shown in Table 19:

SCR reaction:	Volume per reaction	Component
[µl]	10	Oligonucleotides 0.15µM
	27	SCR/SPCR Master mix
	13	H_2O
∑	50	

Table 19: SCR reaction composition

All steps were carried out on cold metal racks and kept cool until starting of the thermocycler. The cycling protocol was composed as shown in Table 20:

SCR protocol:		
98 °C	4 min	
98 °C	10 sec	27 x
60 °C*	30 sec	
72 °C	30 sec	
72 °C	4 min	
4 °C	∞	
*touchdown -0,8 °C/cycle; arriving at 38.4 °C after 27 cycles		

Table 20: SCR thermocycler protocol

The reaction was kept on ice until being processed in SPCR (see 2.2.2.1.2).

2.2.2.1.2 Sequential polymerase chain reaction

To multiply the DNA strands resulting from SCR reaction (see 2.2.2.1.1) a sequential polymerase chain reaction (SPCR) was performed. Using the same master mix composition as for SCR (see Table 18: SCR/SPCR Master mix with Phusion) the reaction was set up as shown in Table 21:

SPCR reaction:	Volume per reaction	Component
[µl]	7	SCR reaction
	27	SCR/SPCR Master mix
	5	Primer forward 10µM
	5	Primer reverse 10µM
	11	H_2O
∑	55	

Table 21: SPCR reaction composition

Here specific primer pairs are added to the reaction that are necessary to start the exponential reaction resulting in multiple copies of the template fragment. All steps were carried out on cold metal racks and kept cool until starting the thermocycler. The cycling protocol was composed as shown in Table 22:

SPCR protocol:		
98 °C	4 min	
98 °C	10 sec	20 x
58 °C	15 sec	
72 °C	30 sec	
72 °C	4 min	
4 °C	∞	

Table 22: SPCR thermocycler protocol

The reaction was kept on ice until being processed in SPCR (see 2.2.2.1.3.2).

2.2.2.1.3 Fusion of two fragments to one construct

2.2.2.1.3.1 Error correction of fragments

The fragments resulting from SCR/SPCR were denatured and error corrected before being assembled to a construct like described in 2.2.2.1.3.2.

Denaturation:	Volume per reaction	Component
[µl]	3	SPCR reaction fragment A
	3	SPCR reaction fragment A
	3.33	10x Ampligase buffer
	10.67	H_2O
∑	20	

Table 23: Denaturation fragments from SPCR

In a thermocycler the reaction was started with the following protocol shown in Table 24:

Denaturation protocol:	
98 °C	2 min
4 °C	5 min
37 °C	5 min
4 °C	∞

Table 24: Denaturation protocol for fragments A and B before enzymatic error correction

12 µl of this reaction were used for an enzymatic error correction using *Taq* ligase and endonuclease. The correction was performed in a thermocycler at 37 °C for 60 min and was then kept on ice immediately after the protocol had finished until further progression.

Error correction set up:	Volume per reaction	Component
[µl]	12	SPCR reaction fragment A/B
	2	*Taq* ligase
	2	T7NI endonuclease
	0.5	10x Ampligase buffer
	3.5	H_2O
∑	20	

Table 25: Error correction reaction composition

2.2.2.1.3.2 Fusion polymerase chain reaction

To result in bigger constructs 2 fragments (A and B) originating from SCR/SPCR reactions were fused together and integrated into a vector by a fusion PCR reaction. A 2 µl aliquot of the error corrected fragments was combined into a reaction mix as shown in Table 26:

FPCR reaction:	Volume per reaction	Component
[µl]	1	Vector pMA-T (50 ng/µl)
	2	Error corrected fragments
	10	5x Phusion HF buffer
	0.4	Phusion HF polymerase
	1	dNTPs (10 mM)
	35.6	H_2O
∑	50	

Table 26: FPCR reaction composition

The reaction was directly put into a thermocycler and started with running the following cycler protocol:

FPCR protocol:		
98 °C	2 min	
98 °C	10 sec	27 x
70* °C	25 sec	
72 °C	45 sec	
72 °C	3 min	
4 °C	∞	
*touchdown -0,9 °C/cycle; arriving at 45.7 °C after 27 cycles		

Table 27: FPCR cycler protocol

Afterwards 2 µl of the reaction were transformed into 100 µl OmniMAX™ or DH5α *E. coli* cells (as described in chapter 2.2.1.2.1.) and plated on LB_{Amp} medium.

2.2.2.1.4 Colony polymerase chain reaction

2.2.2.1.4.1 *Saccharomyces cerevisiae* colony PCR

For verification of correct assembly a PCR reaction from yeast colonies was performed. To open the cells a single colony was picked from a grown plate with a sterile pipet tip, diluted in 0.02 M NaOH and solubilized as in the following Table 28.

Subblocks	Blocks	
18 µl	9 µl	0.02 M NaOH

Table 28: NaOH solubilization for yeast colony PCR

3 µl of solubilised cells were added to every PCR reaction. For control of correct assembling 6 pairs of primers for subblocks and 3 pairs for the block assembly (see 3.2.3.5 and 3.2.4.4) were used to amplify the overlapping regions between each connected fragment. The dilution was cooked on 99 °C for 10 min and the cooled down to 4 °C. PCR was performed under the following conditions using the given protocol below in Table 29.

Master mix:	Volume per reaction	Component
[µl]	3	NaOH colony solubilisation
	5	Betaine (5 M)
	12	Platinum™ PCR SuperMix High Fidelity
	5	10 µM Primer
	5	10 µM Primer
∑	25	

Table 29: Master mix for cPCR *Saccharomyces cerevisiae* for GPAdV constructs

The reaction was set up on a cool rack until the reaction tubes were placed in a thermo cycler running the following protocol:

cPCR protocol yeast:		
95 °C	4 min	
95 °C	30 sec	30 x
60 °C	1: 30 sec	
72 °C	1 min	
72 °C	4 min	
4 °C	∞	

Table 30: Colony PCR cycler protocol from *Saccharomyces cerevisiae* transformed with GPAdV constructs

3 µl of each PCR reaction were loaded on an E-Gel 2% Agarose gel and verified by gel documentation.

2.2.2.1.4.2 Escherichia coli cPCR

2.2.2.1.4.2.1 From fast gene synthesis products

To verify that the constructs derived from the new protocols for SCR/SPCR/FPCR (see 2.2.2.1.1 to 2.2.2.1.3.2) were assembled correctly, colony PCR from the grown bacterial colonies was performed. Therefor one colony was picked up with a sterile pipet tip and diluted in 20 µl Platinum™ PCR SuperMix High Fidelity. Each 0.5 µl of M13 primers (10 µM) forward and reverse was added before the reaction was started in a thermocycler with the protocol in Table 31.

cPCR protocol E. coli for fast gene synthesis:		
95 °C	4 min	
95 °C	30 sec	30 x
55 °C	30 sec	
65 °C	4 min	
65 °C	6 min	
4 °C	∞	

Table 31: Colony PCR cycler protocol from E. coli transformed with fast gene synthesis constructs

For verification 1 µl of the reaction was loaded on a 1% agarose gel to check the correct size of the into vector pMA inserted construct.

2.2.2.1.4.2.2 From *E. coli* for GPAdV constructs

After the assembly in yeast, the colonies were screened for correct clones and the plasmid of the positive ones isolated. This DNA preparation was used to be transformed into *E. coli* TransforMax™ EPI300™, DH10B or Vmax *Vibrio natriegens* cells. From these transformations, a second cPCR screening was done. With a sterile pipet tip one single colony was picked and diluted in 10 µl H_2O in the reaction mix that was composed as shown in Table 32:

Master mix:	Volume per reaction	Component
[µl]	1	Colony dilution
	10	5x SuperFi Buffer
	5	10 µM Primer
	5	10 µM Primer
	1	dNTPs (10 mM each)
	0.5	SuperFi Polymerase
	25.5	H_2O
∑	50	

Table 32: Colony PCR cycler protocol from *E. coli* GPAdV constructs

PCR protocol for *E. coli* colonies on GPAdV constructs:		
98 °C	4 min	
98 °C	30 sec	30 x
70 °C*	30 sec	
72 °C	1 min	
72 °C	4 min	
4 °C	∞	
*touchdown -0,8 °C/cycle; arriving at 46 °C after 30 cycles		

Table 33: Protocol for *E. coli* colony PCR on GPAdV constructs

2.2.2.1.5 Amplification PCR of subblocks

These subblocks that were not suitable to digestion (5, 14, 15, 23 45, 46, 47). Thus, they were amplified by PCR to have a linear DNA fragment to assemble into blocks. The subblocks were ordered as plasmid construct from internal (GeneArt Regensburg), then retransformed into *E. coli* see 2.2.1.2.1 and isolated as MIDI preparations (2.2.1.2.4). The template was diluted 1 : 100 and 1 µl (2 – 7 ng) was used for amplification. Please see Table 52 for exact concentrations. The master mix per reaction is given in Table 34.

Master mix:	Volume per reaction	Component
[µl]	1	MIDI 1 : 100 dilution
	10	5x Phusion GC buffer
	1.25	10 µM Primer
	1.25	10 µM Primer
	1	dNTPs (10 mM each)
	0.4	Phusion polymerase
	35.1	H_2O
∑	50	

Table 34: PCR Mix for amplification of subblocks

The reaction was prepared on a cool rack and performed with the cycler protocol shown in Table 35.

PCR protocol for subblock amplification:		
98 °C	4 min	
98 °C	30 sec	30 x
70 °C*	30 sec	
72 °C	1 min	
72 °C	4 min	
4 °C	∞	
*touchdown -0,8 °C/cycle; arriving at 46 °C after 30 cycles		

Table 35: PCR cycler protocol for amplification of subblocks

2.2.2.1.6 Amplification PCR of block 7

After the correct assembly of block 7 in *Saccharomyces cerevisiae* the plasmid was isolated from yeast cells as described in chapter 2.2.1.1.4. The preparation was used as template for the amplification of block 7 with a concentration of 43.87 ng/µl. 1 µl (4.3 ng) was taken from a 1 : 10 dilution and used in the PCR reaction. The reaction mix was composed as shown in Table 36.

Master mix:	Volume per reaction	Component
[µl]	1	MINI yeast 1 : 10 dilution
	10	5x SuperFi buffer
	5	10 µM Primer
	5	10 µM Primer
	1	dNTPs (10 mM each)
	0.5	SuperFi polymerase
	27.5	H_2O
∑	50	

Table 36: PCR reaction mix for the amplification of block 7 from yeast plasmid preparation

The reaction was prepared on a cool rack and performed with the cycler protocol shown in Table 37.

PCR protocol for block 7 amplification:		
98 °C	4 min	
98 °C	30 sec	30 x
65 °C	30 sec	
72 °C	3 min	
72 °C	4 min	
4 °C	∞	

Table 37: PCR cycler protocol for amplification of bock 7

2.2.2.1.7 Linearization PCR of pYES8D

The target vector for assembly of blocks into segments does not contain restriction enzyme sites for linearization, but is opened by PCR (Baek et al. 2015). Therefore a certain primer set is used that binds exactly on the abridged tryptophan gene and 2-ori (CH327_Trp.fwd and CH353_2µ.rev). As template, 0.56 ng were added to the reaction mix, which was composed as shown in Table 38.

Master mix:	Volume per reaction	Component
[µl]	5	MINI pYES8D 0.56 ng
	10	5x SuperFi buffer
	1	10 µM Primer
	1	10 µM Primer
	1	dNTPs (10 mM each)
	0.5	SuperFi polymerase
	31.5	H_2O
∑	50	

Table 38: Master Mix linearization PCR pYES8D

The cycler protocol was started after preparing the reaction on ice as shown in Table 39.

PCR protocol for linearization pYES8D:		
98 °C	2 min	
98 °C	30 sec	30 x
55 °C	30 sec	
72 °C	3 min	
72 °C	10 min	
4 °C	∞	

Table 39: PCR protocol for linearization of pYES8D

2.2.2.2 Restriction digest

For digestion of circular plasmid DNA the Fast Digest restriction enzymes from Thermo Fisher Scientific were used. 1 µl of every digest was used to verify the size of the expected construct on an E-Gel. The samples were prepared on ice until incubation for 1 h at 37 °C. The following Table 40 show the individual recipes for subblocks, blocks and vectors.

a)

pEYES-MCS-21R	
DNA Preparation pEYES-MCS-21R plasmid	1 µg
Fast Digest SgsI (2.5 U/reaction)	1 µl
Fast Digest PacI (2.5 U/reaction)	1 µl
Fast Digest BamHI (2.5 U/reaction)	1 µl
FD buffer	2 µl
H_2O	Fill up to final volume
∑	20 µl

b)

Subblocks	
DNA Preparation pEYES-MCS-21R plasmid	1 µg
Fast Digest XhoI (2.5 U/reaction)	1 µl
FD buffer	2 µl
H_2O	Fill up to final volume
∑	20 µl

c)

Blocks	
DNA Preparation pEYES-MCS-21R plasmid	1 µg
Fast Digest PacI (2.5 U/reaction)	1 µl
FD buffer	2 µl
H_2O	Fill up to final volume
∑	20 µl

Table 40: Recipes for enzymatic digestions
a)For the linearization of target vector plasmid pEYES-MCS-21R the fast digest enzymes SgsI, BamHI and PacI were used and incubated in the described recipe; after incubation 1 µl of Calf Intestinal Alkaline Phosphatase was added and incubated for another 30 min at 37 °C; all enzymes were heat inactivated by incubation at 65 °C for 10 min. b) The subblocks were cut out of the production vector using fast digest enzyme XhoI; the reaction was stopped by heat inactivation at 80 °C for 10 min. c) The blocks were cut out of the production vector using fast digest enzyme PacI; the reaction was stopped by heat inactivation at 80 °C for 10 min.

2.2.2.3 Agarose gel electrophoresis

2.2.2.3.1 Agarose gel electrophoresis using Ethidium bromide (EtBr)

For analysis, the DNA fragments were mixed with 10x blue buffer loading dye (1x final) and loaded on a 1% agarose gel. The gel was prepared with 0.2 µg/mL EtBr and 0.5x TAE buffer. The electrophoresis ran at 220 V for 40 min. The bands were analysed and documented via UV light with the Science Imaging system from INTAS.

2.2.2.3.2 Agarose gel electrophoresis using E-Gel® EX system

For analysis of the DNA fragments, the DNA samples were mixed with H_2O (15 µl final) and loaded on either a 1% or 2% precast agarose gel. The E-Gel® EX gels were delivered pre-stained with SYBR Gold II stain. The electrophoresis ran for 9 min for the 10 well gels and 20 min for the 48 well gels on the E-Gel® iBase™ Power System. The bands were analysed and documented via UV light with the Science Imaging system from INTAS.

2.2.2.4 DNA clean-up

2.2.2.4.1 Gel extraction

To separate assembled or amplified DNA products from either vector, template or side products a 1% Agarose gel was composed as shown in Table 41:

Guanosine agarose gel for gel extraction	
TAE	60 ml
Agarose	0.6 g
Guanosine 0.1 M	60 µl
EtBr 0.7 mg/ml	2 drops

Table 41: Guanosine agarose gel 1% for gel extraction

2.2.2.4.2 Vacuum ultrafiltration

PCR products derived from FPCR (see 2.2.2.1.3.2) were cleaned before being transformed into *E. coli* cells. Each reaction was pipetted into one filter well and the plate placed on the vacuum manifold. The vacuum was turned on allowing all liquid to drain through the filter. Afterwards each wells bottom was washed with 50 µl Tris (10 mM) to take up the purified DNA. The suspension was either directly used for further applications or stored a -20 °C.

2.2.2.5 Sequencing

2.2.2.5.1 Sanger Sequencing

Sequencing was performed at GeneArt (Biopark, Regensburg) using 200 ng plasmid or linear DNA and 10 µM of sequencing primer. Oligonucleotides were provided by GeneArt GmbH internally. The results were analyzed with the Seqman Pro software from DNAStar.

2.2.2.5.2 RNA sequencing by NGS

The isolated RNA sample from the GPTEC-T cells after transfection was performed at the sequencing facility at Heinrich-Pette-Institut, Hamburg.

2.2.2.6 Concentration measurement of DNA

2.2.2.6.1 Using NanoDrop™ 8000 Spectrophotometer

To check concentrations via NanoDrop 2 µl of reaction or elution sample were used. Before measuring the sample, the channels were cleaned and 2 µl of desalted water were loaded for initialization. The blanking was performed with a 2 µl of either appropriate buffer or a blank control reaction (as used in PCR or digest reactions).

2.2.2.6.2 Using Qubit 4 Fluorometer

Concentration measurement using the Qubit 4 Fluorometer was performed with the Qubit™ dsDNA HS Assay Kit. Each measurement was set up as shown in Table 42 and incubated 2 min at room temperature before measurement.

Qubit assay	
Working solution (ws)	n (number of samples plus 2 standards) x 200 µl
For each 200 µl ws	1 µl high sensitivity reagent + 199 µl dsDNA high sensitivity buffer
First Standard	10 µl standard #1 + 190 µl ws
Second standard	10 µl standard #1 + 190 µl ws
Sample	2 µl + 198 µl ws

Table 42: Qubit assay set up and composition of working reagent, standards and samples

2.3 Statistical methods

2.3.1 Plackett-Burman Design of Experiments

As described above in chapter 1.3.1 and 1.5.1 one goal in this thesis was to optimize the known standard PCR protocols in a direction leading to speed up a whole production process. Instead of trying any imaginable combination of all factors, a strategy had to be used to come to a significant conclusion with as less experimental set ups as possible. The Plackett-Burman statistical technique is as famous as suitable to optimization purposes like in this thesis, where the effect of varying different components is to be studied separately (R. L. Plackett and J. P. Burman 1946). It finds implementation in many scientific exercises of experimental research or production process optimization. The Plackett-Burman-Design (PBD) identifies critical parameters as here, the times of different PCR steps, without involving the interaction affects (Ekpenyong et al. 2017). Method to investigate the dependence of some measured quantity on a number of independent variables (factors). The goal is to determine the main effect. The objective in this thesis was to find the shortest PCR times for denaturation, annealing and elongation steps and the lowest number of cycles by keeping the same quality of product. Within these factors the time is either as long as the standard protocol or shortened to a certain defined time, meaning a 2-factor interdependency. With using a excel tool, that was generated as a PBD tool, scenarios are set up that bring the answer to the question which factor has the highest influence. The fractional factorial design bears a certain number of scenarios, meaning it passes some combination opposed to a full factorial design. This number is always that multiple of 4, which is next greater than

the number of factors (Beres and Hawkins 2001), but best no potency of 2. Here we have 4 factors, leading to 12 scenarios to be tested as shown in the example below, instead of 16 which a full factorial design would deliver.

Number of scenarios:	**12**		
Number of factors:	**4**		
		+	-
Factor A:	Number of cycles	30	20
Factor B:	Denaturation time	30 sec	10 sec
Factor C:	Annealing time	30 sec	15 sec
Factor D:	Elongation time	1 min	30 sec

Figure 7: Plackett-Burman Design: Example of number and subject of factors and the 2-factor-interdependency

The PBD is applicable on any question where the factors can be defined as either original (+) or changed (-), with what a 2-factor-interdependency is given. Here it is the number of cycles in the original protocol or the reduced one and same with the original times for PCR steps and the shortened ones. After this is defined, the scenarios are designed. For that a programmed Excel template (designed by Tobias Reusch, Technical assistant, Thermo Fisher Scientific GeneArt GmbH) was filled and used to set up the experiments (see results part 3.1.1).

	Faktor A	Faktor B	Faktor C	Faktor D
1	+	+	-	-
2	-	+	+	-
3	+	-	+	+
4	+	+	-	+
5	+	+	+	-
6	-	+	+	+
7	+	-	+	+
8	-	+	-	+
9	-	-	+	-
10	-	-	-	+
11	+	-	-	-
12	-	-	-	-

Table 43: Plackett-Burman Design: Scenario set up with Excel template designed by Tobias Reusch (see description above); the changed factors are coded with (-) and marked light blue.

In Table 43, the set-up of the 12 scenarios derived from the 4 investigated factors and resulting experiments on PCR is shown. To have a formulated transcript, the table was translated into the certain times and numbers that are to be investigated (see Table 44). With this model the protocols were programmed in the thermocyclers.

To determine the highest impact on result reliability, the number of correct bands on an agarose gel was counted and translated into percentage. The 2-factor table consisting of – and + was complemented with a row named result, where the percentage was filled in after every scenario. This whole design can only be applied if it is known, that the reciprocity is negligible small compared to the main virtue. Otherwise both effects are blended with each other and not traceable anymore. This PBD is called a fractional factorial design with resolution III, meaning the 2-factor-interdependency, what makes it mostly admired for determining main factors (Marko Kapitza 2011).

2.3.2 Genome Partitioner and homologous recombination

To get a strategy of how to build a complete genome the Genome Partitioner software was applied. As the Guinea Pig Adenovirus is newly found and its genome not annotated, which means that the open reading frames and genes coding for structural protein are not located among the sequence. Therefore it was important to get the complete wild-type sequence synthesized without optimizations to the original sequence. With using the Genome Partitioner web based interface, a bottom up approach for building genome scale constructs from the 37.070 kb genome was designed. The tool was developed at Eidgenößische Technische Hochschule (ETH) in Zurich as a free access online software tool (Matthias Christen, Luca Del Medico, Heinz Christen, Beat Christen 2017). Its goal is to partition *in silico* DNA sequences to generate a retrosynthetic pathway for higher-order assemblies. It delivers the ready to order sequences and primers to have the accomplishment of the assemblies checked by PCR (Christen et al. 2017). The programming language is Python using the Biopython package, which allows to give out GenBank files as well as a graphic output file and a documentation file to track all algorithmic procedures. In this thesis the approach was completely built on homologous recombination *in vivo* carried out in *Saccharomyces cerevisiae*. The overlap based assembly method enables the combining of 5 DNA parts in a one step process by joining short overlapping sequences after transformation into Y187 yeast cells (see Table 2: *Saccharomyces cerevisiae* cell line) making use of its DNA repair machinery. For that approach, disfavoured DNA sequence features like secondary structures, high GC-stretches and repetitions within the overlapping regions are eliminated by the algorithm

without changing the sequence. Therefor this methods represents not only splitting DNA into parts, but using a logical computational design algorithm. The website allows to upload the sequence as a GenBank file as shown in Figure 8.

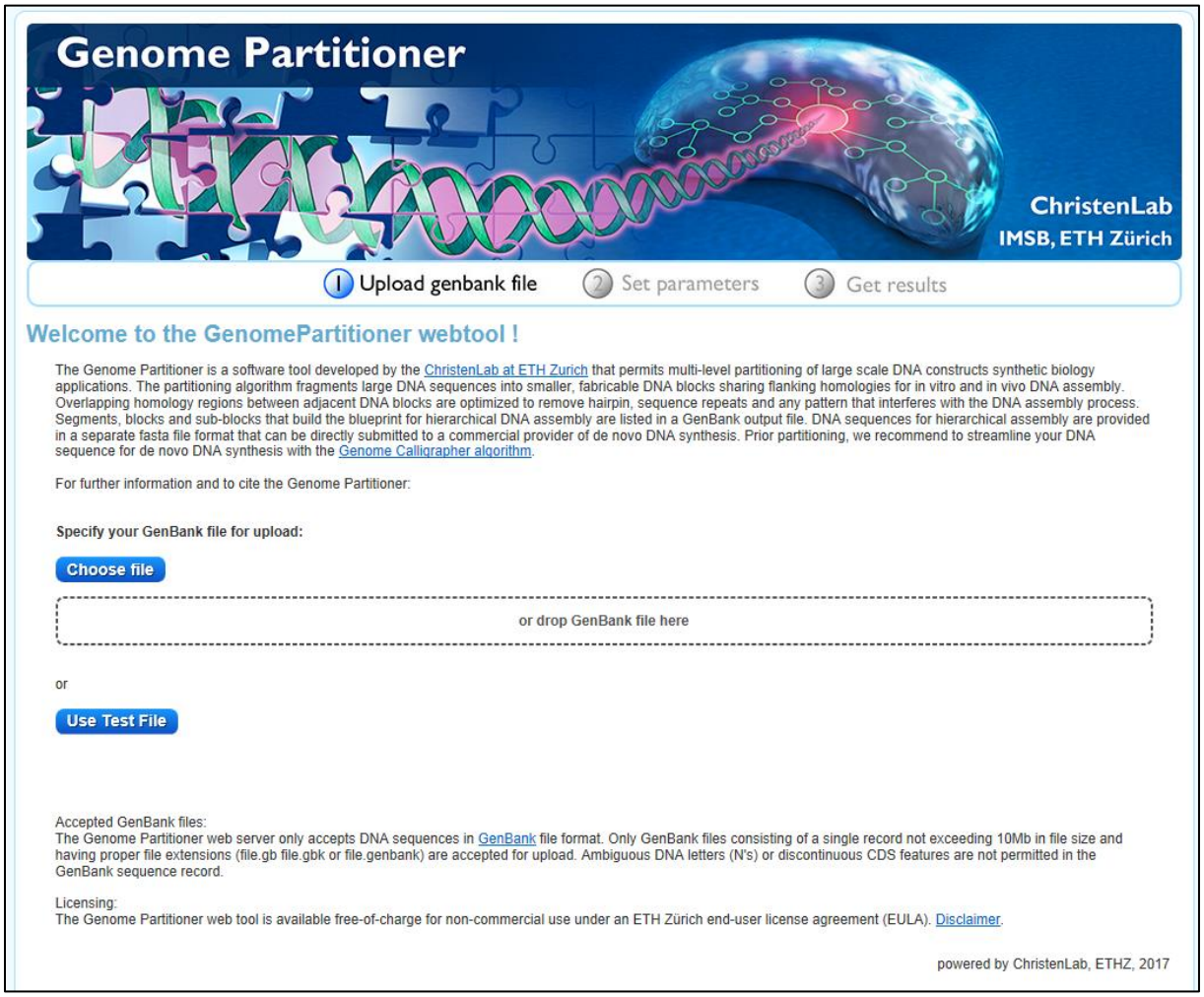


Figure 8: Genome Partitioner web interface: Entering the sequence in GenBank file format; Figure is derived from https://christenlab.ethz.ch/GenomePartitioner

The next page guides through planning the assembly steps from subblocks (smallest parts) to blocks (middle step) to segments (final halves).

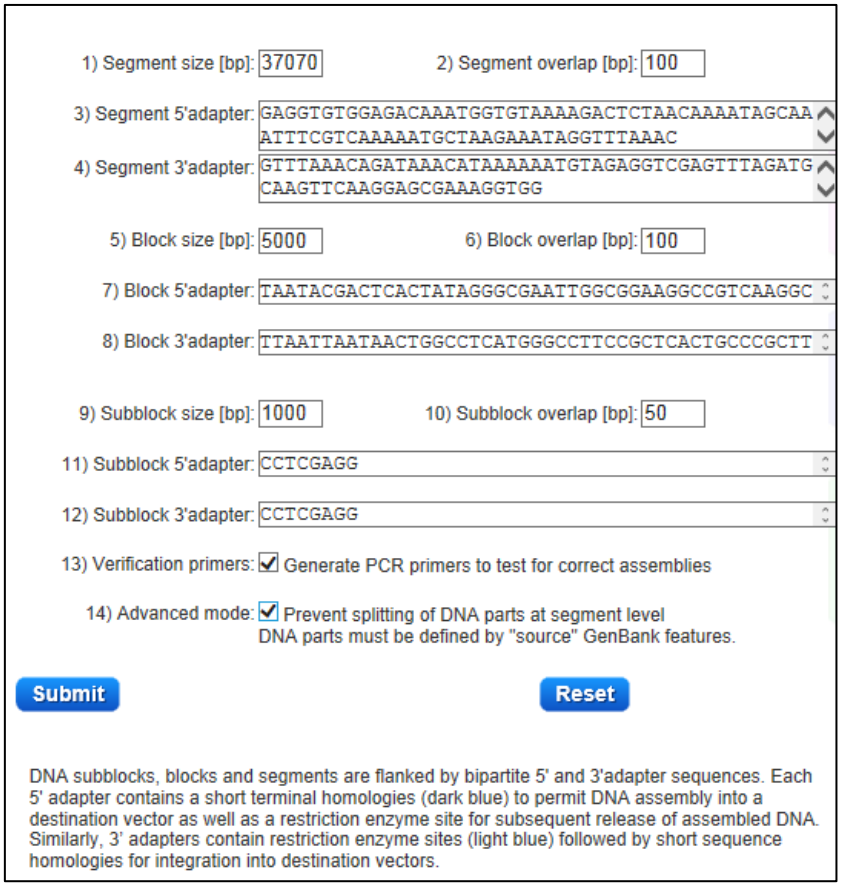


Figure 9: Genome Partitioner interface: Specifying the partitioning parameters including restriction sites and overlaps to vectors (= 3'- and 5'-adapters); Figure is derived from https://christenlab.ethz.ch/GenomePartitioner.

The specific parameters are set, such as size at each assembly level, size of overlaps and the so called 3'- and 5'-adapters) (see Figure 13: Homologous recombination, mechanism of double strand break; (Barlow and Rothstein 2010).). In this work, the size of subblocks was set to 1000 with overlaps of 50 bp to each other. At this level the restriction enzyme XhoI was used, which defines the restriction site in the delivery vector for ordering. The block size was set to 5000 bp having a 100 bp overlap to the flanking blocks in addition to the restriction site for PacI. The same was done for the segments resulting in two 18 kb parts with a PmeI site and 100 bp overlap to each other. Here a prevention of annotated regions was redundant as the sequence is a newly found wild type one as described before. The adapters are already planned and integrated at the lowest level for a higher order assembly as shown in Figure 10.

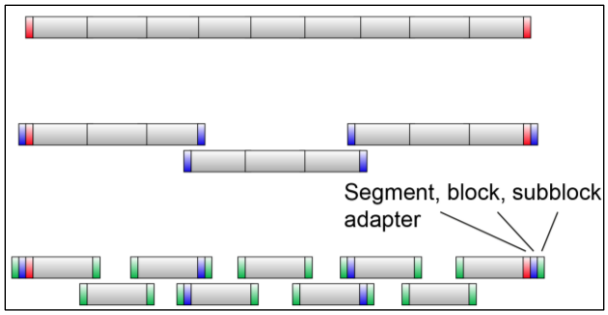


Figure 10: Genome Partitioner interface: Higher orde assembly steps showing the arrangement of the 3'- and 5'-adapter flanking the subblocks, blocks and segments.

The higher order design is shown in Figure 11, where each assembly integrates matching DNA parts into a vector. From there they can be released by restriction digest and propagated to the next assembly step.

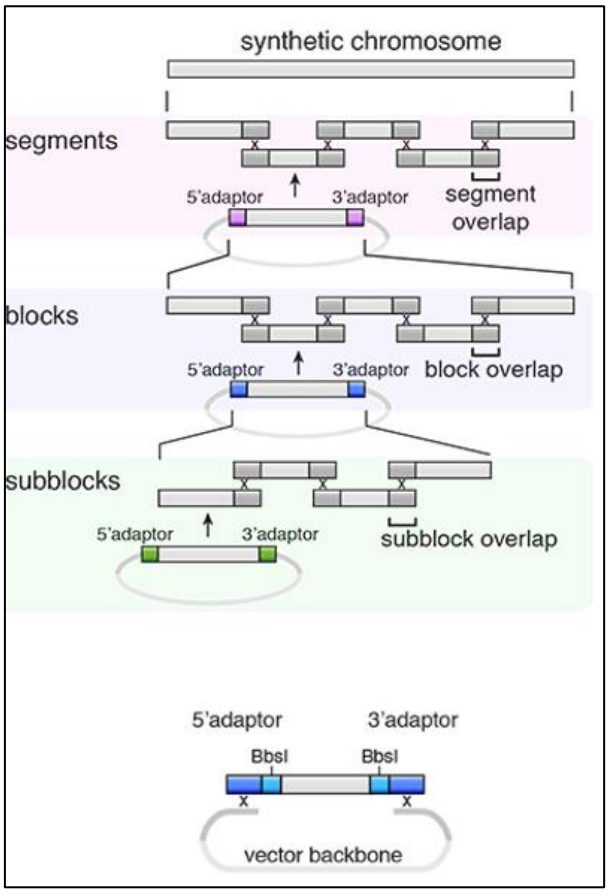


Figure 11: Genome Partitioner interface; retrosynthetic higher-order assembly; from subblocks to synthetic chromosome; Figure is derived from https://christenlab.ethz.ch/GenomePartitioner.

After the data was filled into the tool interface, a concrete plan including unique primer pairs, amplifying the overlaps between the parts is given. Using the primers for colony screening PCR reaction ensures the connection of DNA parts to each other and also in the correct order. The algorithm of the tool prevents the terminal homology regions (THR) from being similar. An

addition logfile delivers all the data that ran through the algorithm including the exact length of every subblock, block and segment and also location and size of removed hairpins, repetitions or other disturbing features within the THR (see 7.2).

```
-----------------------------------------------------------------
segments created 2, min,mean,max: 18732, 18732, 18732
original length 37070, partitioned length 37464
-----------------------------------------------------------------
blocks created: 8, min,mean,max,std: 4833, 4874, 4915, 18.1738
original length 37070, partitioned length 38992
-----------------------------------------------------------------
subblocks created: 48, min,mean,max,std: 851, 870, 881, 0.0000
original length 37070, partitioned length 41760
-----------------------------------------------------------------
overlap analysis summary
hairpins            : 0
direct repeats      : 0
multiple subseq > 8 : 0
bad segments        : 0 of 2
bad blocks          : 0 of 8
bad subblocks       : 0 of 48
-----------------------------------------------------------------
```

Figure 12: Genome Partitioner interface; showing a part of the logfile with the designed parts and information on the output after applying the algorithm to the sequence; Figure is derived from https://christenlab.ethz.ch/GenomePartitioner.

With this information the subblocks were ordered in-house at Thermo Fisher Scientific GeneArt GmbH.

2.4 Homologous recombination system in yeast

The DNA fragments described as subblocks, blocks and segments shall be assembled using the system of *Saccharomyces cerevisiae* on the principle of homologous end joining. With this pathway, yeast repairs its own DNA lesions error-free and thus is a reliable method of connecting several fragments at overlapping sequence sites. Here an overview is given on homologous recombination in yeast with the mechanism of double strand breaks.

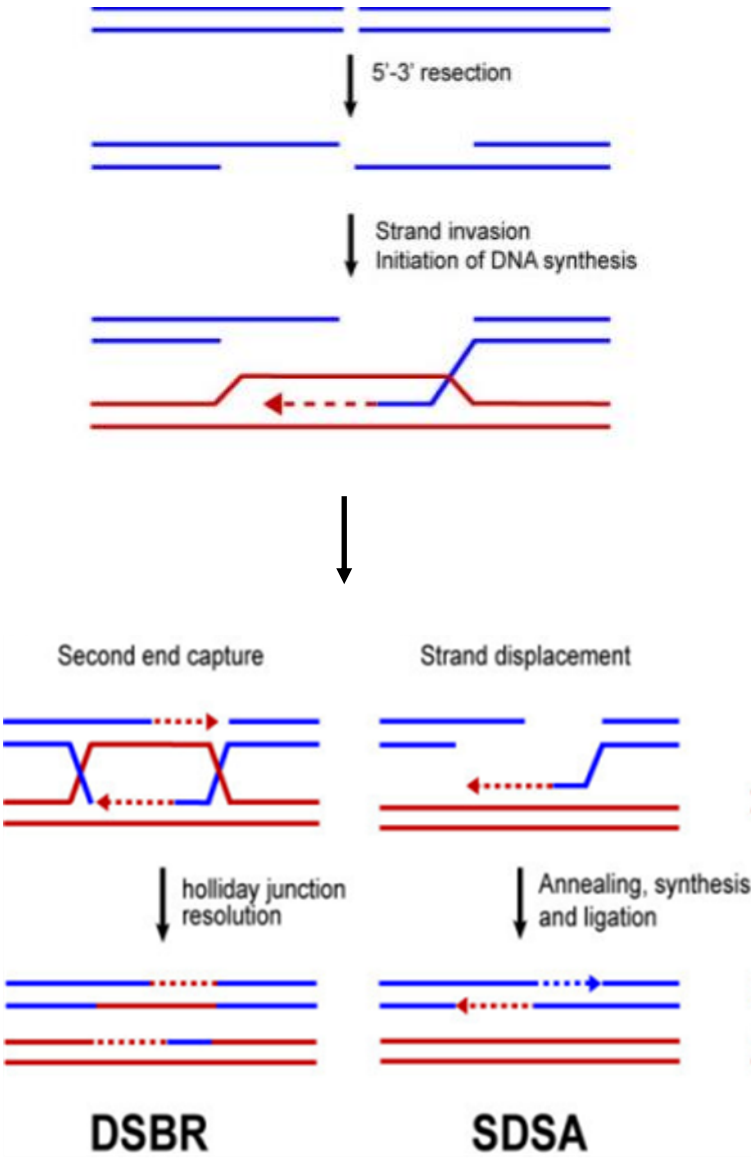


Figure 13: Homologous recombination, mechanism of double strand break; (Barlow and Rothstein 2010).

The re-joining of DNA ends requires homology in sequence to the fragments that are to be assembled. In DBSR both ends of the break invades the acceptor (vector) and leads to formation of a joint molecule resulting from replication of both donor 3' ends. Synthesis dependent strand pairing (SDSA) delivers a product that comes from a subsequent displacement of DNA synthesis of one strand. The resulting single strand DNA is complementary to the other double strand break and is combining with it. DNA replication promotes that all gaps are filled (Barlow and Rothstein 2010).

3 RESULTS

3.1 An abridged workflow for gene synthesis

3.1.1 Plackett-Burman-Design tool

3.1.1.1 Optimization of SCR, SPCR and FPCR

Gene synthesis as here performed is mainly comprised from three polymerase chain reactions (PCR), the sequential chain reaction (SCR) for oligonucleotide assembly, the standard polymerase chain reaction (SPCR) to multiply the product and the fusional polymerase chain reaction (FPCR), where products of SPCR are combined. The first subject to optimize in generating fast gene synthesis protocols was the one of SCR. In this reaction oligonucleotides with an overlapping region are assembled to small sub-fragments of the desired DNA construct. According to the explanation of the Plackett-Burman-Design (PBD) tool in Excel, the scenarios are shown in Table 44, formatted with the numbers of cycles, times in each reaction step. During denaturation the double strand DNA template is separated by heating up to 98 °C resulting in 2 single strands (Mullis et al. 1986). This step is needed in the beginning of every PCR cycle and its duration depends on the length and GC content of the target DNA. The higher GC content and the longer the template is, the longer is the time needed to separation. During the annealing step the primers are binding to the denatured strands contributing to their design also in GC content and length. The efficiency is given by an adapted annealing temperature to the primer pairs. In elongation phase of each cycle, the DNA polymerase extends the single strands and completes the synthesis of new strands of DNA. Depending on the length of the template a certain duration for this step is adjusted. The number of cycles, where these steps are performed are repeated, leading to the multiplication of newly synthesized DNA strands.

Every scenario was performed with the same reaction master mix and oligonucleotide set. The agarose gel was then screened for bands showing the correct size and a distinct signal. The durations were noted in Table 44 for 12 scenarios.

	Number of cycles	Denaturation time	Annealing time	Elongations time	Correct bands (%)
1	30	30 sec	15 sec	30 sec	87.5
2	20	30 sec	30 sec	30 sec	87.5
3	30	10 sec	30 sec	1 min	100.0
4	30	30 sec	15 sec	1 min	87.5
5	30	30 sec	30 sec	30 sec	100.0
6	20	30 sec	30 sec	1 min	93.7
7	30	10 sec	30 sec	1 min	100.0
8	20	30 sec	15 sec	1 min	87.5
9	20	10 sec	30 sec	30 sec	87.5
10	20	10 sec	15 sec	1 min	87.5
11	30	10 sec	15 sec	30 sec	87.5
12	20	10 sec	15 sec	30 sec	93.7

Table 44: Plackett-Burman Design: Scenario set up formulated changes and results row showing the correct bands in %; the changed factors are marked light blue.

In an additional column, the success rate of 96 sample sequences was checked on a 1% agarose gel to determine wrong or missing bands. To express the influence of changes on the parameters in numbers. The sum of the numeral number of original factors (+, white) and also of the new ones was added up (-, light blue) within each row. For the row “number of cycles” these were 562.5 and 537.5, having a difference of 25. The difference divided by the number of scenarios delivers the score shown in red and green in Table 45 for each factor. A high score represents a high impact on the result.

	Number of cycles	Denaturation time	Annealing time	Elongation time
sum +	562.5	543.75	568.75	556.25
sum -	537.5	556.25	531.25	543.75
total sum	1100	1100	1100	1100
difference	25	-12.5	37.5	12.50
Impact on SCR-product	**2.0833**	**-1.0417**	**3.1250**	**1.0417**

Table 45: Plackett-Burman Design: Results translated into a score of each factor/row; Red: high impact on result; Green: low impact on result

Shown in a graph (see Figure 14) it is visible, that the factors having a high impact are the number of cycles and the annealing time (red bars), whereas denaturation has no and the elongation time only little impact on the result as seen on an agarose gel (green bars).

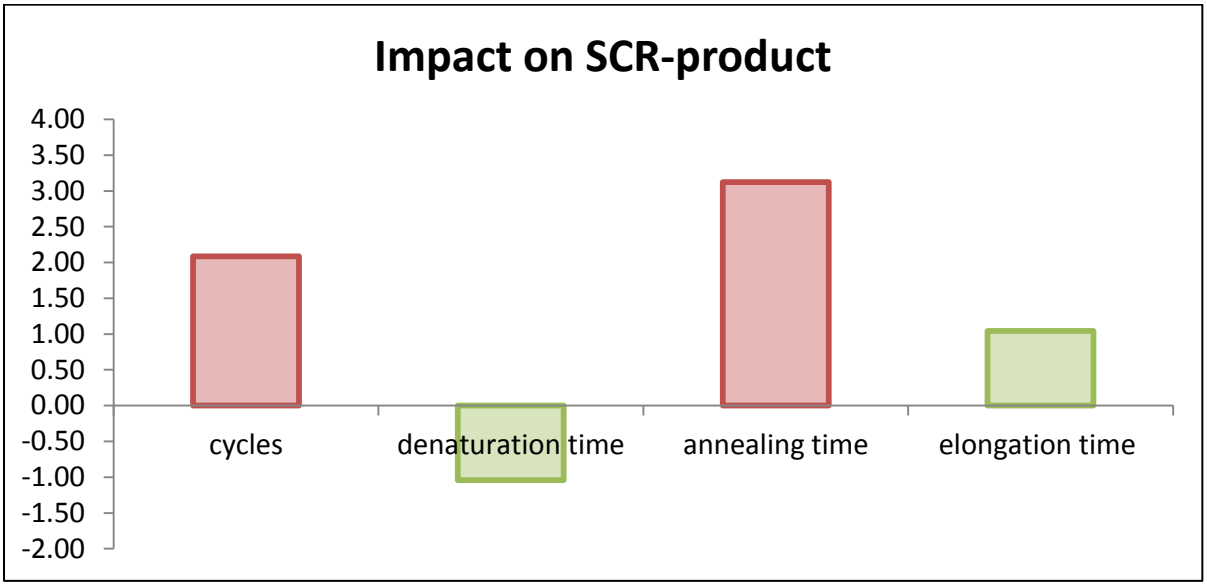


Figure 14: Plackett-Burman Design: Graph on the statistical outcome of the PBD on SCR; Red: The factors having the highest and second highest impact on the SCR product as seen on an agarose gel; Green: Impact of denaturation time and elongation time with no and only little impact on SCR.

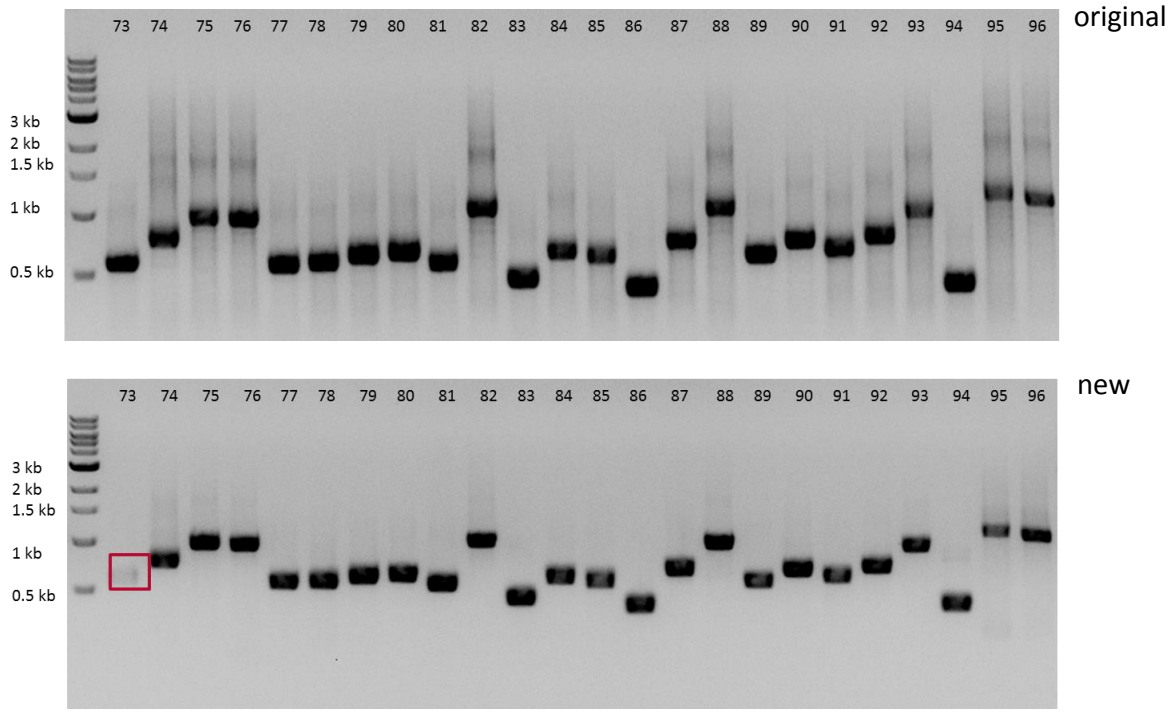


Figure 15: Plackett-Burman test set for SPCR: Here an excerpt of the 96 sequence analysis data is shown (sample 73 – 96); marked in a red box is one sequence that failed after shortening the protocol for the thermo cycler; 7 µl of 1kb öladder from NEB were used as areference.

Conferring to the statistical evaluation shown in Figure 14 and Table 45 experiments were done for SCR, SPCR and FPCR. The output protocols are described in 2.2.2.1.1 and 2.2.2.1.3.2 and are summarized in Table 46.

SCR protocol:		
98 °C	4 min	
98 °C	10 sec	27 x
60 °C*	30 sec	
72 °C	30 sec	
72 °C	4 min	
4 °C	∞	
*touchdown		

SPCR protocol:		
98 °C	4 min	
98 °C	10 sec	20 x
58 °C	15 sec	
72 °C	30 sec	
72 °C	4 min	
4 °C	∞	

FPCR protocol:		
98 °C	2 min	
98 °C	10 sec	27 x
70* °C	25 sec	
72 °C	45 sec	
72 °C	3 min	
4 °C	∞	
*touchdown		

Table 46: New protocols for SCR, SPCR and FCR in overview.

A total reduction in time of 47% was reached after application of PBD, with the most significant time efficiency in FPCR optimization respectively 56%. The detailed numbers and results are listed in Table 47.

PCR	Time before (min)	Time optimized (min)	Time reduction (min)	Time reduction (%)	Success rate after optimization
SCR	83	55	28	**34%**	100%
SPCR	78	42	36	**46%**	100%
FPCR	120	53	67	**56%**	97 %
Total process	282	150	132	**47%**	

Table 47: Plackett-Burman process optimization overview: Here the time reduction is shown for SCR, SPCR and FPCR as well as the success rate conferring to a sequence set of 96 sequences.

3.2 Results on Guinea Pig Adeno Virus construction

3.2.1 Analysis of the GPAdV sequence

3.2.1.1 DNA statistics and characterization

The Guinea Pig Adenovirus Sequence was provided by Prof. Dr. Thomas Dobner and contributes to the isolate Ger1. Before starting to develop a concept for synthesizing DNA fragments, the sequence had to be analyzed. Table 48 shows an overview of length and parameters. The first thing to do was to determine the composition of bases. It revealed that the *in silico* data sequence has a slightly increased average GC content.

Sequence type	DNA
Length	37,070 bp
Organism	Virus
Name	Guinea Pig Adenovirus
Description	GER1
Molecular weight (double-stranded)	22.91 MDa

Table 48: Sequence data sheet overview; from clc file in CLC sequence viewer 8.0 by Quiagen

Base(s)	Presence in %
A	18.85 %
G	30.43 %
T	19.12 %
C	31.60 %
A+T	37.97 %
G+C	62.03 %

Table 49: DNA statistics by DNAStar SeqBuilder showing the overall base content in GPAdV sequence

Looking a bit deeper, there were several repetitions found within the sequence between 12 -17 bp, that can lead to mis-annealing of oligonucleotides, when synthesizing the desired area. Furthermore, the sequence has 36 GC-rich stretches of 40 – 50 bp and at least 20 AT-rich stretches. The example below shows in Figure 16, where a stretch of 50 bp reach GC peak of more than 90% (shaded grey), which leads to a high melting temperature of oligonucleotides in gene synthesis and is therefore difficult to produce with standard protocols.

a)

Startpos1	Startpos2	Länge1	Länge2
323	2951	17	17
369	4322	13	13
458	2804	13	13
2908	4756	13	13
3075	4316	13	13
441	1864	12	12
1504	5265	12	12
1895	4051	12	12
3423	4757	12	12
3498	4190	12	12

b)

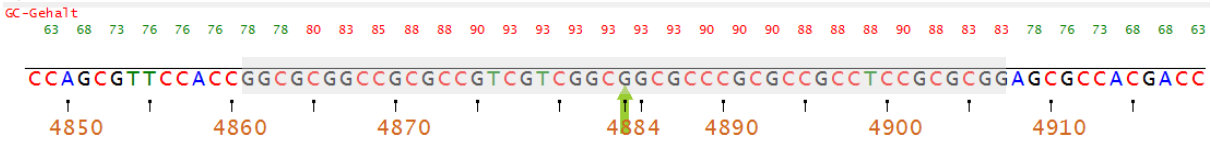


Figure 16: Sequence analysis of Guinea Pig Adenovirus; a) Examples for repetitions; b) Example for GC stretch.

The inverted terminal repeats (ITR) were found to be 167 bp long and identical in sequence. The ITR on both ends contains a 26 bp repetition (13 bp direct repeat). Both are shown in Figure 16 with the 3'- and 5' ITR marked in grey and the repeats in dark blue.

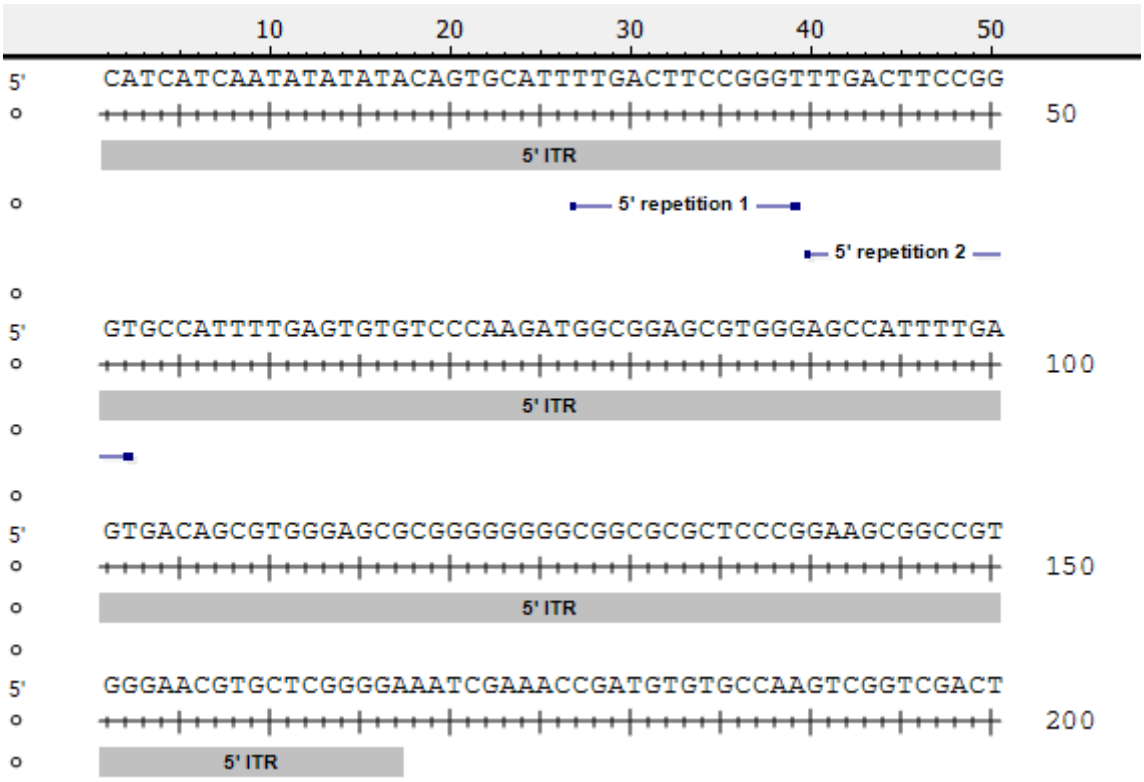


Figure 17: 5' ITR region; Inverted terminal repeat on 3' end raching from base 1 to base 167 with length of 167 bp (grey bar); direct repeats of 13 bp each are marked with a dark blue bar.

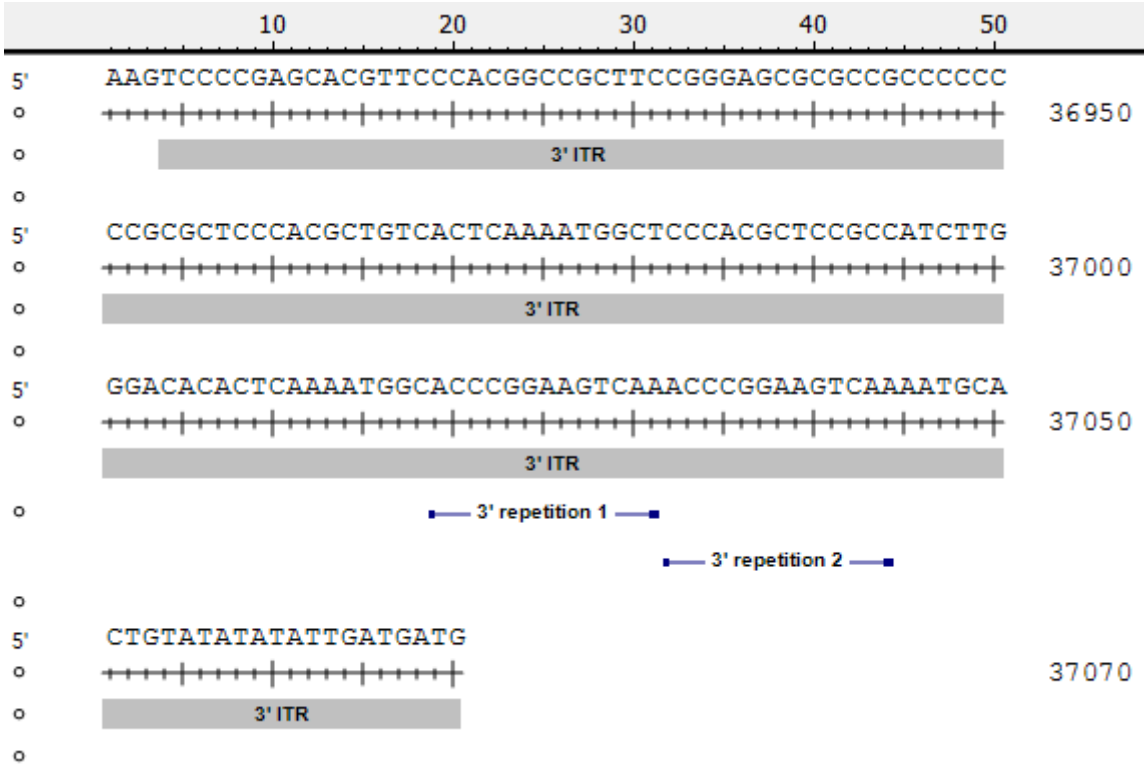


Figure 18: 3' ITR region; Inverted terminal repeat on 3' end raching from base 36904 to base 37070 with length of 167 bp (grey bar); direct repeats of 13 bp each are marked with a dark blue bar.

3.2.1.2 Analysis of possible cloning strategies

To get an idea of which cloning strategies could be suitable the sequence was screened for cutting sites of restriction enzymes. Therefor the sequence was analysed with the DNAStar Seqbuilder software to detect restriction enzyme sites. It revealed AbsI, PacI and PmeI (MssI) as the only restriction enzymes not cutting in the sequence. See Table 50 for information about their contributing recognition sites and cutting pattern.

Enzyme	Recognition site and cutting pattern
AbsI	5' C C↓T C G A G G 3' 3' G G A G C T↑C C 5'
XhoI	5' C↓T C G A G 3' 3' G A G C T↑C 5'
PacI	5' T T A A T↓T A A 3' 3' A A T↑T A A T T 5'

PmeI (MssI)	5' G T T T↓A A A C 3' 3' C A A A↑T T T G 5'

Table 50: Restriction enzymes used for GPAdV assembly based on concept of Genome Partitioner tool; Here the enzymes AbsI, XhoI, PacI and PmeI are shown with their recognition site and cutting pattern

The enzyme AbsI was chosen for flanking the subblocks. As visible in the table above, AbsI includes the cutting site of XhoI. The enzyme AbsI is very uncommon and of poor specifity, thus XhoI was used instead although it was found to cut 4 times within the sequence at position 1563, 4239, 11622 and 18401. These sites appeared within 4 subblocks as described in 2.2.2.1.5. Therefore, XhoI, PacI and PmeI were used for the higher order assembly of the viral genome using the software Genome Partitioner and thus for the three assembly steps (see 2.3.2).

3.2.2 Partitioning

The following table gives an overview of the parameters including the subblock, block and segment adapters. The adapters include the overlap sequence to the target vector (green), in case of pYES8D also the complementation sequence to tryptophan or the 2µ-origin (orange) and the restriction site (blue) to excise the assembled construct for the next higher assembly step. All sequences are given in 5' – 3' orientation. In addition, a primer set for control of correct assembled construcst is generated.

feature_type	Source
segmentsize	37070
segment_window	0.1
overlap_segment	100
blocksize	5000
overlap_blocks	100
subblocksize	1000
overlap_subblocks	50
hairpin_size	8
max_stretch_length	8
max_iteration	20

block_adapter_left	TAATACGACTCACTATAGGGCGAATTGGCGGAAGGCCG TCAAGGCCTAGGTTAATTAA
block_adapter_right	TTAATTAATAACTGGCCTCATGGGCCTTCCGCTCACTGCC CGCTTTCCAGTCGGGAAA
subblock_adapter_left	CCTCGAGG
subblock_adapter_right	CCTCGAGG
segment_adapter_left	GAGGTGTGGAGACAAATGGTGTAAAAGACTCTAACAAA ATAGCAAATTTCGTCAAAAATGCTAAGAAATAGGTTTAA AC
segment_adapter_right	GTTTAAACAGATAAACATAAAAAATGTAGAGGTCGAGT TTAGATGCAAGTTCAAGGAGCGAAAGGTGG
primer_generation	Yes

Table 51: Parameters of Genome Partitioner; overview of setting of partitioning the GPAdV sequence including adapter sequences and algorithmic presettings (hairpin size, repeats, iterations); a primer set is generated; size of subblocks, blocks and segments is set as well as their contributing overlap size.

Figure 19 shows the principle of construction. All 48 subblocks (numbered from 0 to 47) are shown in light and dark green, where each 6 are the matching ones to result in a block (turquoise and blue). In summary 8 blocks are built to end up with 4 blocks segment (light ans dark pink). The purpose was to end up with one segment of 37070 bp (see Table 51). But as both segments inherit one ITR each acting like homologous regions and thus hinder the partitioning algorithm from building one construct (see Figure 10) the output were 2 segments represting each one half of the genome.

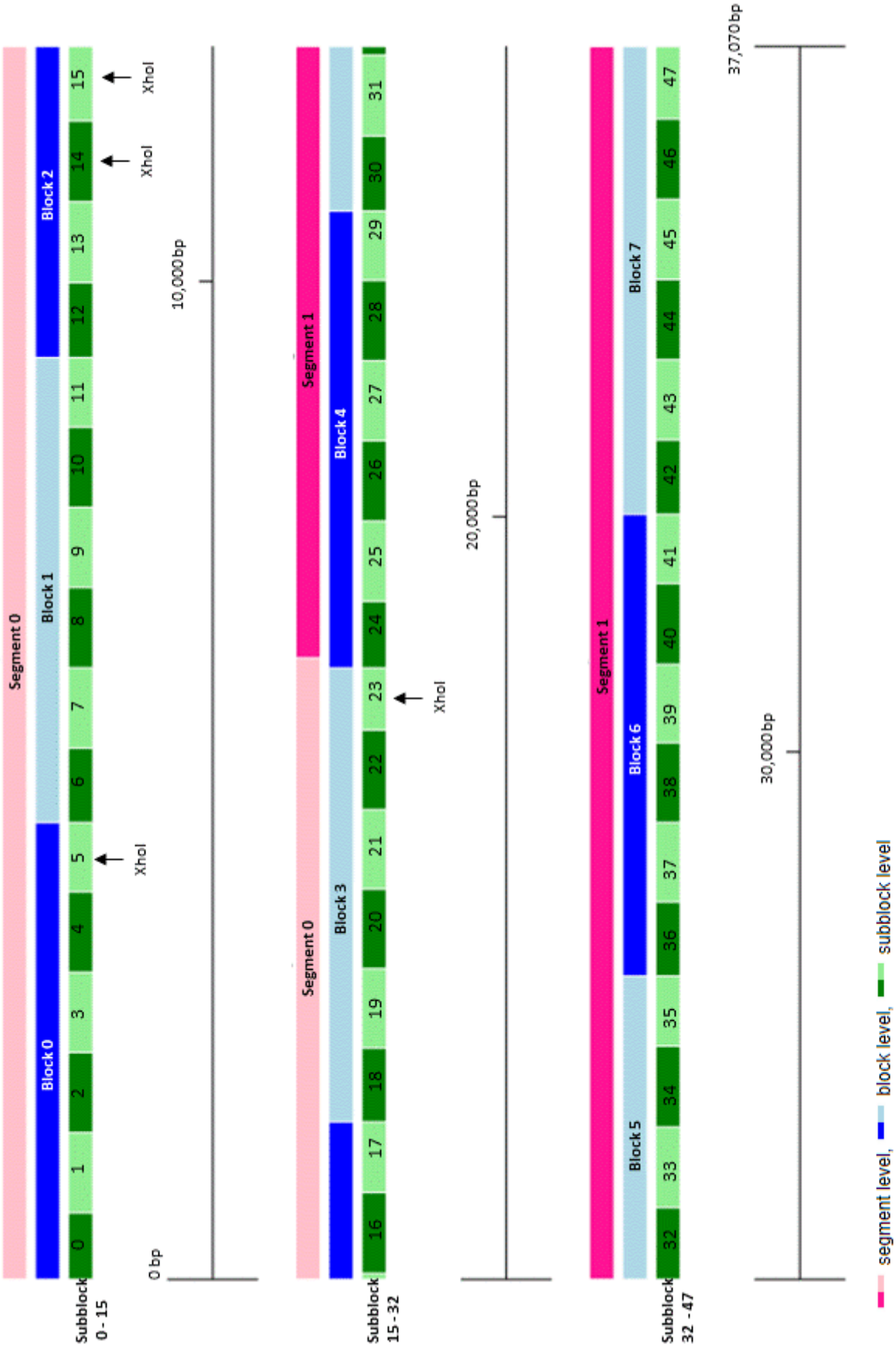


Figure 19: Set up of the higher order assembly srategy by Genome Partitioner; subblocks containg a XhoI restriction site are labelled.

3.2.3 Assembly of subblocks into blocks

3.2.3.1 Digestion of subblocks

All 48 subblocks were ordered in-house by standard gene synthesis and arrived as cloned constructs in a production vector (pMK) with a kanamycin resistance marker. For each digest 1 µg of plasmid DNA was used and digested with XhoI (see Table 9 and chapter 2.2.2.2). From each sample 1 µl was run on a 2% E-Gel as described in 2.2.2.3.2 using 5 µl of 1 kb Plus ladder. The results are shown in Figure 20.

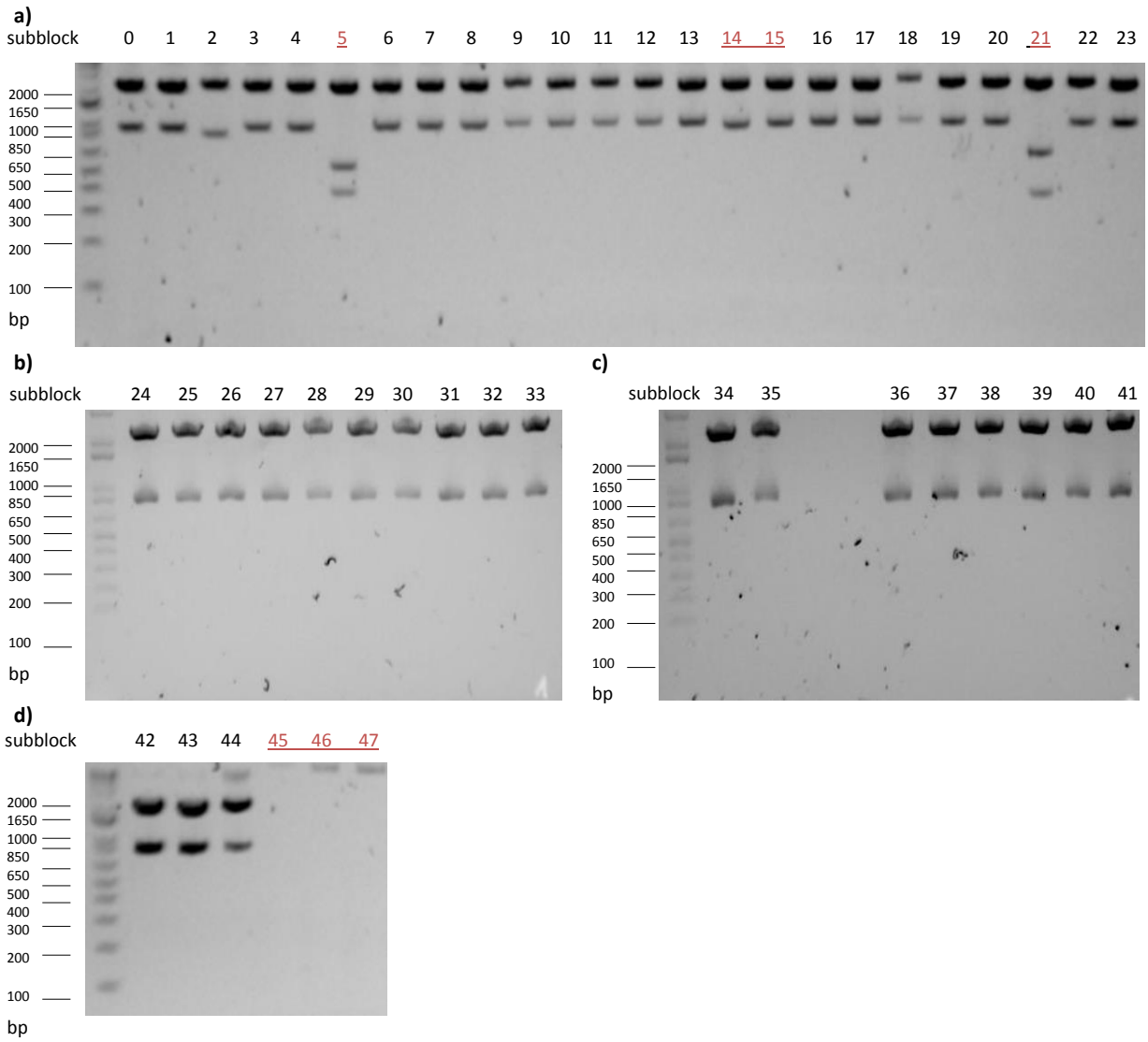


Figure 20: Assembly of subblocks: Digestion of subblocks 0 – 47 with XhoI to release from production vetor; the lower bands represent the insert (subblock) of a size between 851 bpand 877 bp; the upper bands show the vector backbone with a size of 2281 bp; a) Digest of subblock 0 – 23: marked in red are the subblocks containing a internal XhoI restriction sites; b) Digest of subblock 24 – 33; c) Digest of subblock 34 – 41; d) Digest of subblock 42 – 47; marked in red are the subblocks where the digest failed (45 – 47). Here some bands right below the wells of gel are visible that might represent genomic DNA.

The digest of 41 subblocks showed the expected result with one band between 851 bp and 877 bp (see Appendix 7.2). For subblocks 5 and 21 the digest shows 3 bands, one resulting from the vector backbone and two resulting from the insert being cut at a third internal XhoI binding site. The same is true for the samples containing subblock 14 and 15 although not visible, as the internal third XhoI binding site lies close to the 3' end of the insert (Figure 20 a). The digest of subblocks 45 – 47 did not show any digestion result. Weak bands appearing very close to the application well give a hint on non-digested vector constructs. Hence, subblocks 5, 14, 15, 21 and 45 – 47 were amplified by PCR (see 2.2.2.1.5).

3.2.3.2 Amplification of subblocks by PCR

As described above subblocks 5, 14, 15 and 23 with an internal XhoI restriction site were amplified by PCR as well as subblocks 45 – 47. Therefor PCR was performed from a 1:100 dilution of the plasmid preps as template.

Subblock	Conc. ng/µl	1 : 100 dilution ng/µl	Size
5	282	2.82	854
14	411	4.11	847
15	412	4.12	854
23	432	4.32	847
45	759	7.59	854
46	807	8.07	854
47	204	2.04	854

Table 52: Overview of subblocks to be amplified by PCR; Shown are the concentration of the plasmid prep and its 1 : 100 dilution in H_2O and the length of the desired product.

The PCR was set up and run as described in Table 35. 1 µl of each reaction was loaded on a 1% E-Gel (see 2.2.2.3.2) to verify the correct size. The result after gel documentation is shown in Figure 21.

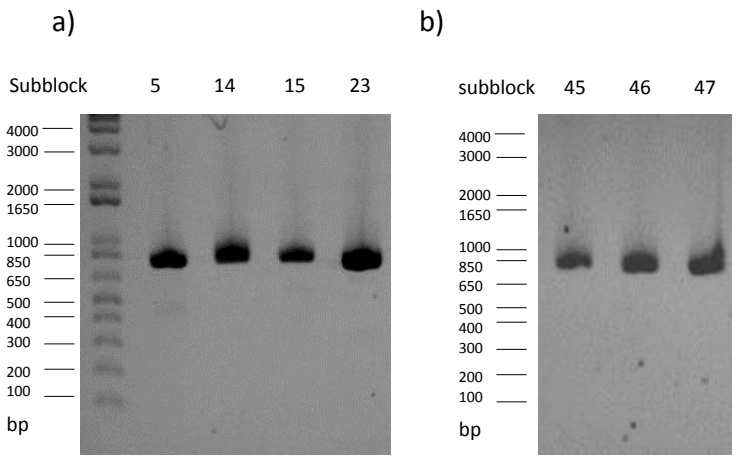


Figure 21: Amplification of subblocks: 5 µl of 1kb Plus ladder were used; a) Subblocks 5 (854 bp), 14 (847 bp), 15 (854 bp) and 23 (847 bp) have an internal XhoI restriction site and were thus amplified by PCR from the ordered plasmid; b) Subblocks 45, 46 and 47 (854 bp each) were also amplified by PCR due to a failed restriction digest; shown here all samples were loaded on a 1% E-Gel and documented.

All reactions showed distinct bands at the expected level contributing to the ladder and predicted size. The PCR was performed with the matching subblock primer pairs (see 7.1) that bind specifically at the 5' and 3' ends of each defined subblock. The reactions were then used for the assembly of blocks in the higher step.

3.2.3.3 Linearization of pEYES-MCS-21R

The target vector for the block assembly is pEYES-MCS-21R that is described in chapter 2.1.5. The vector was grown in bacteria and the plasmid isolated. To open the vector a digest was performed with AscI and PacI as restriction sites flanking the overlapping region for homologous recombination. The figure below shows an excerpt of the pEYES-MCS-21R plasmid at its recombination sites. Marked in orange are the recognition sites for the restriction enzymes with the cutting site. Next to those are the overlap regions (marked pink) of 50 bp each located for the integration of the flanking subblocks building a block by recombination as described in Figure 11. As visible in Figure 22 the overall size of the vector is 9540 bp, the multiple cloning site with the restriction sites is t 66 bp long.

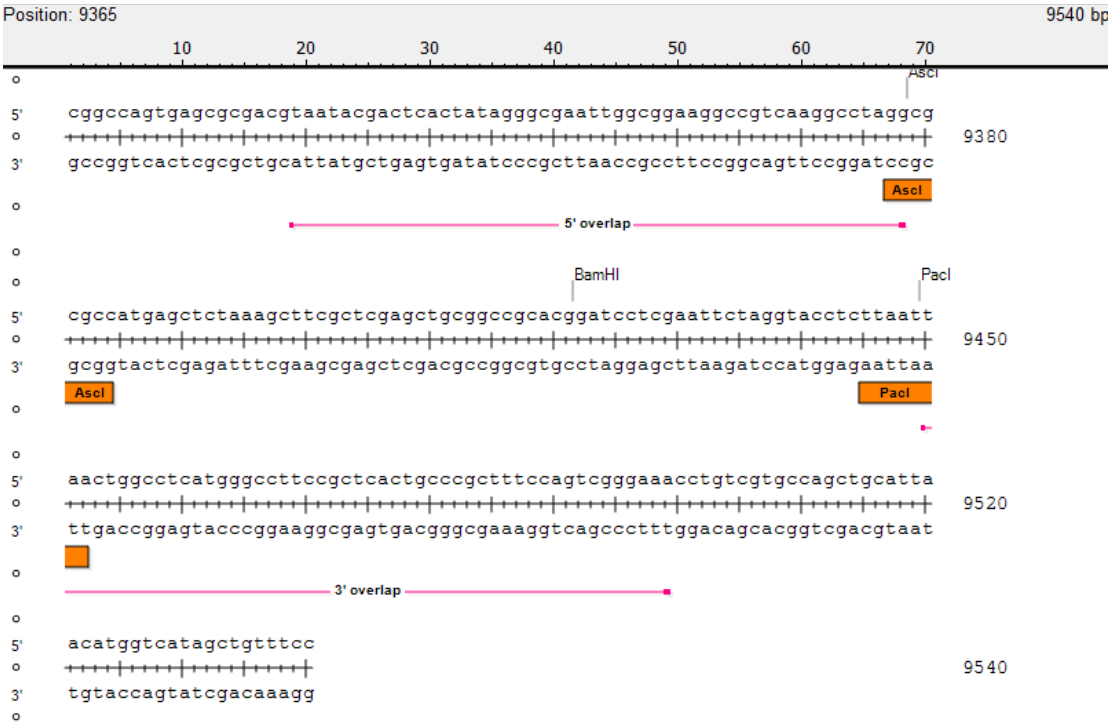


Figure 22: Target vector pEYES-MCS-21R multiple cloning site: The vector is linearized by digestion with the restriction enzymes AscI and PacI (marked orange); for efficient limearization a third enzyme BamHI was used to prevent the insert from recombinig; the homologouse regions for insertion of the subblock DNA fragments are marked pink with a size of 50 bp at each 5' and 3' end.

The digest was performed together with a third restriction enzyme BamHI. By that, the fragment between the overlapping regions is cut once more, which helps to prevent re-ligation. 1 µg of plasmid was digested for every assembly as described in 2.2.2.2. After 1 h at 37 °C, 1µl of CIAP was added to the reaction, which catalyses the hydrolysis of 5'-phosphate termini and thus reduces self-ligation of linear vectors (Alberts and Börsch-Supan 2001). A following heat inactivation stopped the enzymes from being active. 1 µl of the reaction were loaded on a 1% agarose gel. The run was documented as in the following figure in doublets and a control with undigested plasmid vector was run along.

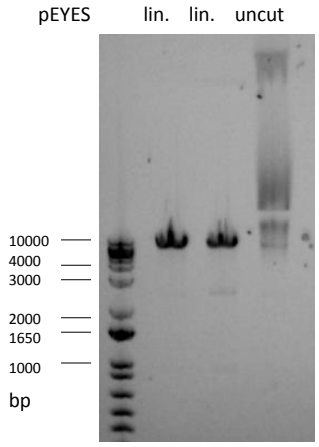


Figure 23: Linearization of pEYES-MCS-21R: 5 µl of 1 kb Plus ladder were used; The digest was perfomed with the restriction enymes AscI, PacI and BamHI; the multiple cloning site of 66 bp was realeased and the target vector treated with CIAP to prevent self ligation; a aliquot of 1 µl of the reaction (50 ng) was loaded on a 1% agarose gel that ran 60 min at 180 V; the first two bands show the aliquot of digest reaction, the third one a control of 50 ng untreated plasmid.

The performance of the digest is nicely visible by comparing the bands appearing from the reaction sample to the control band of uncut vector. The samples labelled as “lin.” (linear vector) in Figure 23 show distinct bands at the expected level of almost 10000 bp (9540 bp) in contrast to the control sample of untreated and thus plasmid target vector. Here the expected run process is shown.

To make sure there is no further material transferred into the assembly reaction, the digest was cleaned up by gel extraction as described in 2.2.2.4.1. For that the size was checked again on a 1% agarose gel containing 1% guanosine dilution (100 mM) using a wider well. Three reactions were be pooled to a volume of 60 µl, loaded at once on the gel and ran 80 min at 180 V. Shown in Figure 24 is the result of a 250 ng aliquot for documentation, before the band was cut from the gel for further process.

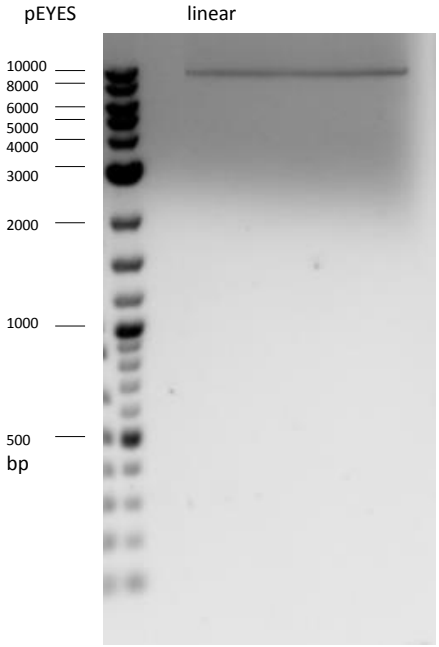


Figure 24: Gel extraction of pEYES: 5 µl of 1 kb Plus ladder were used; The digest reaction containing the linearized plasmid was loaded in a 1% gel in a 70 µl volume well; here 250 ng cut vector was loaded in a volume of 60 µl and run for 80 min at 180V.

3.2.3.4 Assembly of 8x 6 subblocks to blocks

Yeast transformation with PEG/LiAc is based on the common method described by Gietz and Schiestl 2007 and the method described and used by Christen et al. 2017. Former studies showed that a proportion of 300 ng insert (here subblock) and 500 ng target vector (here pEYES) is sufficient for the transformation of 10^8 competent yeast cells. Here the strain Y187 from TAKARA was used (see Table 2: *Saccharomyces cerevisiae* cell line). The cells were prepared freshly for every transformation experiment by picking a single colony from a grown full media YPDA plate (see 2.1.6.2, 2.2.1.1.2 and 2.2.1.1.1). The reagents PEG3350 and LiAc are decisive in this method

for acting on the membrane to make the transformation efficient and to help DNA passing through the cell wall (Kawai et al. 2010). Every step of the transformation procedure was kept sterile by using sterilized utensils and by working under flame of a Bunsen burner. The procedure includes a heat shock of 15 min at 42 °C. To all experiments controls were carried along. One negative control was made to distinguish that the linearized vector is free from background and left over plasmid vector. A second negative control shows that the untransformed cells cannot grow on histidine lacking media plates when treated with the same transformation mix, but without DNA (see Table 15). A positive control with transforming the circular plasmid pEYES brought inference about the efficiency and performance of the transformation itself. The table below shows the yield in colonies after 3 - 4 days of incubation and selection of positive transformants on the auxotrophic marker histidine.

Block	Subblocks 300ng each	Target vector 500 ng	Colony forming units Block	NC lin. vector	PC	NC cells	Transformation rate
0	0 - 5	pEYES	49	17	100	0	21 cfu/µg
1	6 - 11	pEYES	44	5	117	0	19 cfu/µg
2	12 - 17	pEYES	6	0	147	0	3 cfu/µg
3	18 - 23	pEYES	57	5	117	0	25 cfu/µg
4	24 - 29	pEYES	336	2	312	0	146 cfu/µg
5	30 - 35	pEYES	224	2	312	0	97 cfu/µg
6	36 - 41	pEYES	120	2	312	0	52 cfu/µg
7	42 - 47	pEYES	3	0	147	0	1 cfu/µg

Table 53: Assembly of blocks: Number of colony forming units in selective media (-His) plates after transforming 300 ng of each linearized subblock and 500 ng of linear target vector (2300 ng DNA/transfection); two negative controls were carried along: one with linear target vector (500 ng) and one with untransformed cell, both treated with the same transfection mix; a positive control with the plasmid vector (500 ng) was carried along as well; the table shows which subblocks were combined to their respective blocks.

A graph below shows the number of colonies for every block comparing it to the number of colonies derived from the positive and negative controls.

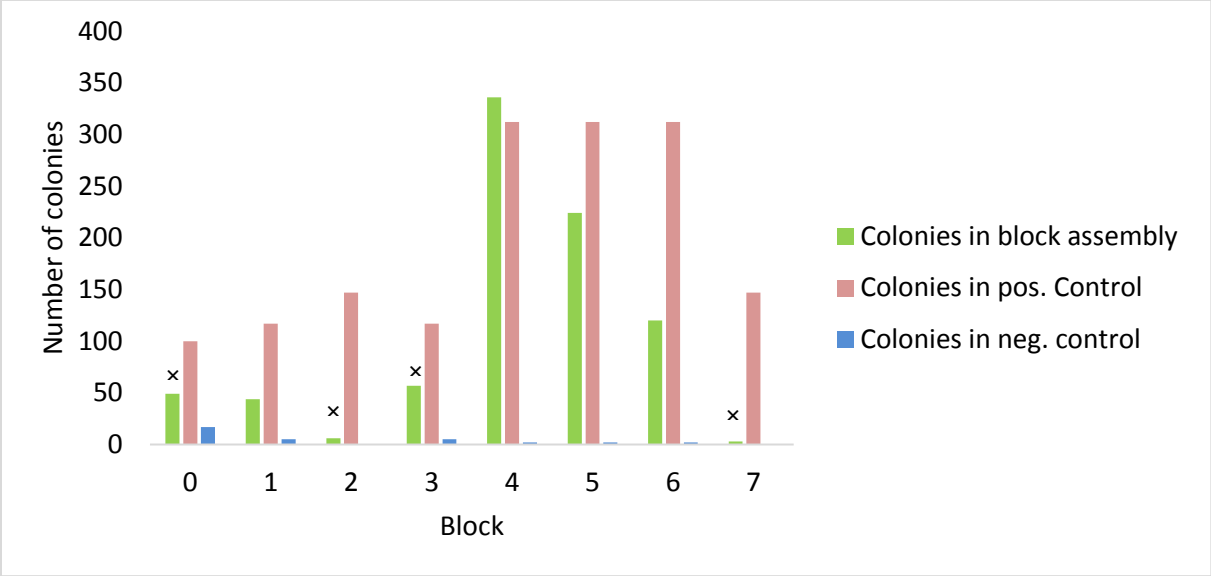


Figure 25: Number of colonies for block assemblies: here the numbers of colonies for each block assembly is shown in green; numbers of colonies for the appropriate controls are shown in light red for the positive control (PC, plasmid pEYES) and blue for the negative control (NC, linear pEYES)

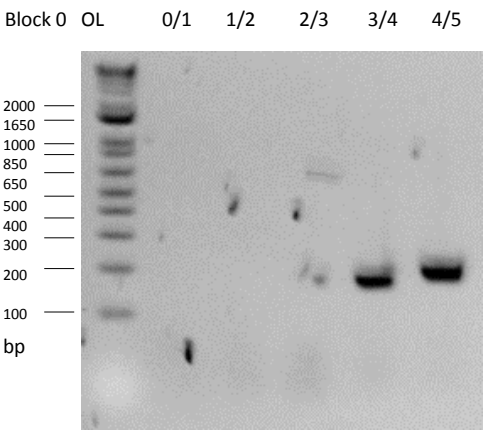
The overview above clearly shows an uneven transformation rate and number of colonies within the assemblies of blocks (green bar). Here has to be noted that the subblocks 5, 14, 15 and 23 were linear DNA products resulting from PCR instead of digestion, as well as subblocks 45 – 47 (Table 53). These DNA fragments were within the assemblies of block 0, 2, 3 and 7 (marked with ×). Block 0 and 3 were assembled with 1 PCR product, block 2 with 2 PCR products and block 7 with 3 PCR products. The positive control (PC) showed also various numbers of colonies, but lies with an average of 196 colonies in the expected range (Gietz and Schiestl 2007). With the negative control of purged linearized vector, between 2 and 17 colonies appeared after transformation for samples 0, 1, 2, 3, 4, 5 and 6, whereas no colonies grew for samples 2 and 7.

3.2.3.5 Yeast colony PCR from block constructs

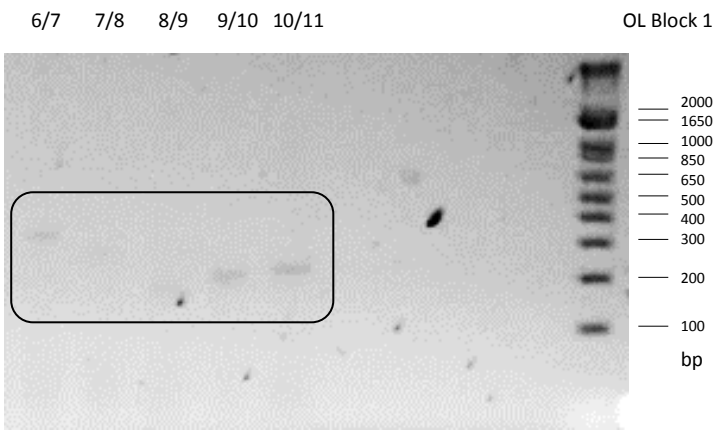
To screen the grown colonies for correct assemblies of blocks, colony PCR of the transformed *Saccharomyces cerevisiae* cells was performed as described in chapter 2.2.2.1.4.1. Picking the cells was followed by boiling the colonies in 0.02 M NaOH to break open the cell wall, which is essential for a successful colony PCR performance (Blount et al. 2016; Walhout and Vidal 2001). Each colony was re-suspended in 18 µl of 0.02 M NaOH and 3 µl were used as template for each reaction. Per colony, 5 control primer (CP) pairs (see 7.1) were used for amplification of the junction of the connected subblocks. The protocol for the cPCR of yeast contains 1 M betaine in the recipe for the reaction mix. Betaine is supposed to destabilize dsDNA through the reaction and makes PCR on

GC rich templates more effective (Weissensteiner and Lanchbury 1996; Frackman et al. 1998). In the following, the results documented on 2% E-Gels are shown. For each run 3 µl of sample were loaded.

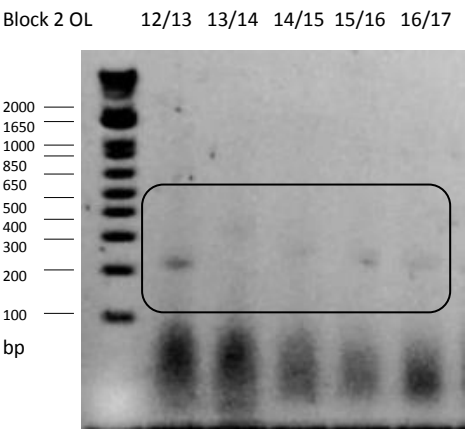
a)



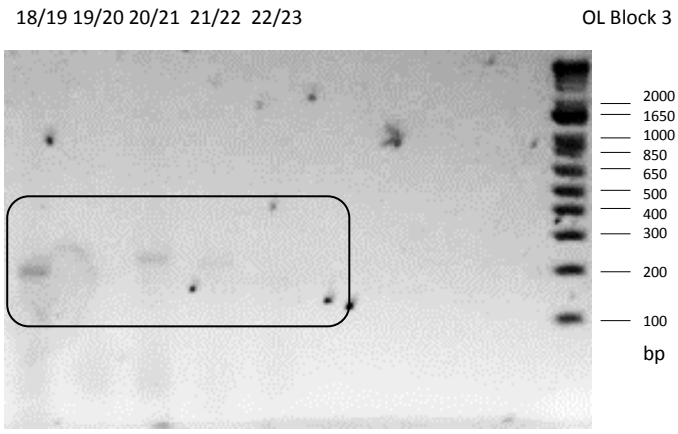
b)



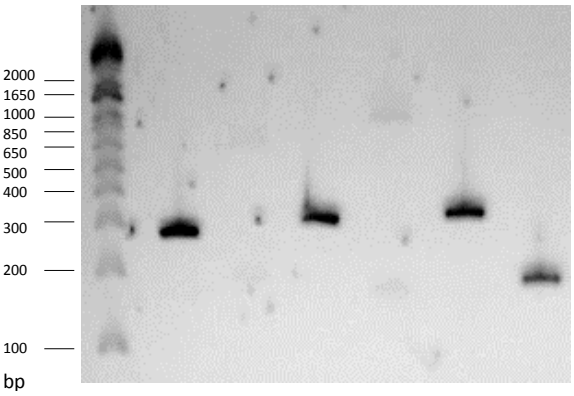
c)



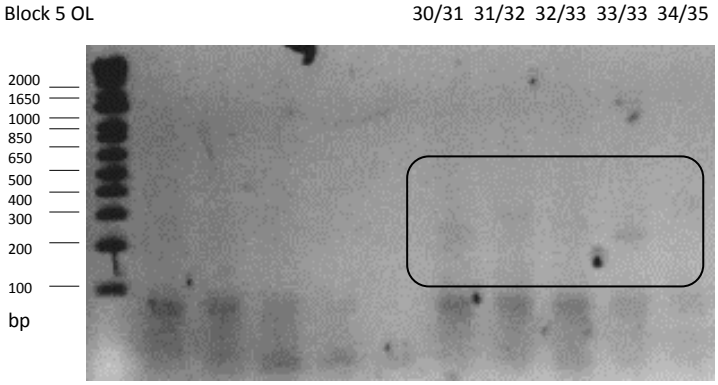
d)



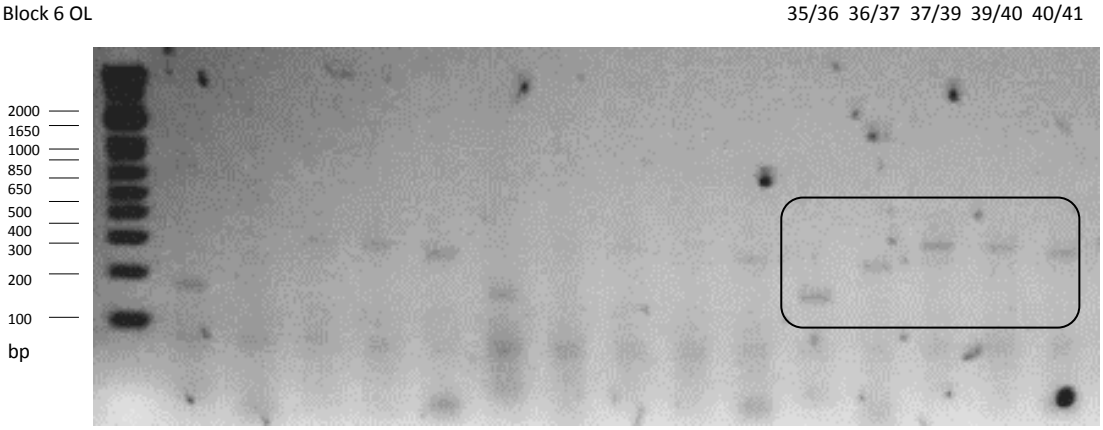
e)



f)



g)



h)

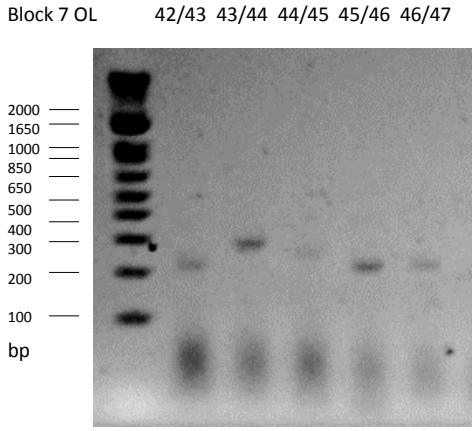


Figure 26: Yeast colony PCRs from the assembled blocks: 3 µl of each PCR reaction was loaded on a 2% E-Gel; amplification products are the overlapping regions (OL) between the assembled subblocks; 5 µl of 1 kb Plus ladder were used;

- **a) Block 0: The gel documentation shows the connected region between subblock 3/4 and 4/5 in the expected size of 148 bp and 176 bp; the first 3 connections are missing;**
- **b) Block 1: Amplification of the regions 6/7, 9/10 and 10/11 are visible in the expected size of 317 bp, 188 bp and 203 bp;**
- **c) Block 2: All connecting regions are amplified and slightly visible bands appeared in the expected sizes of 174 bp, 247 bp, 207 bp, 187 bp, 174 bp;**
- **d) Block 3: Amplification of the regions 18/19, 20/21 and 21/22 are slightly visible in the expected size of 206 bp, 228 bp and 208 bp;**
- **e) Block 4: Amplification of the regions 25/26, 27/28 and 28/29 are visible in the expected size of 264 bp, 267 bp and 193 bp;**
- **f) Block 5: All connecting regions are faint bands in the expected sizes of 190 bp, 240 bp, 210 bp, 171 bp, 279 bp;**
- **g) Block 6: All amplification of the regions are slightly visible in the expected sizes of 128 bp, 211 bp, 258 bp, 254 bp and 230 bp;**

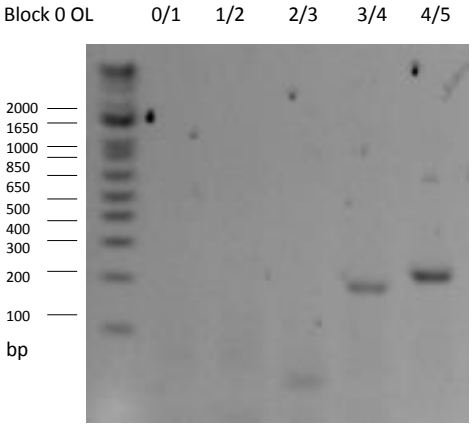
h) Block 7: All connecting regions appeared as bands in the expected sizes of 186 bp, 253 bp, 226 bp, 198 bp, 207 bp;

The gel pictures show the amplified product of the 5 recombined overlapping homologous regions (OL) between 6 subblocks. The bands in general appeared quite weak and often (for example block 5) only as slight shadows. Suspending one yeast colony in NaOH lowers the concentration of template due to dilution. Additionally, most of the DNA in is genomic DNA from yeast. Block 0 shows no results on the first three primer pairs. Still there were 3 of the subblocks (3, 4 and 5) assembled, so that this colony was proceeded to isolate the plasmid and transform it to *E. coli* to check if a functional plasmid is there and can be determined closer by cPCR. The same was done for block 4, where 2 expected bands appeared as double bands in wrong size. This can also happen, when the primer pairs bind to a genomic sequence instead of the plasmid. This occurrence might indicate either only few replicates of the plasmid or a mixture of re-ligated, incomplete assembled and correct assembled plasmids were produced within one yeast colony. The samples (block 7, 6 or 2) where all expected amplification led to bands appearing correctly, still gave faint bands. All positive and positive appearing colonies were spread and plated on fresh media plates to be incubated 48 h. All colonies grew again as complete cell layer. The layer was then harvested and the plasmid DNA isolated as described in 2.2.1.1.4.

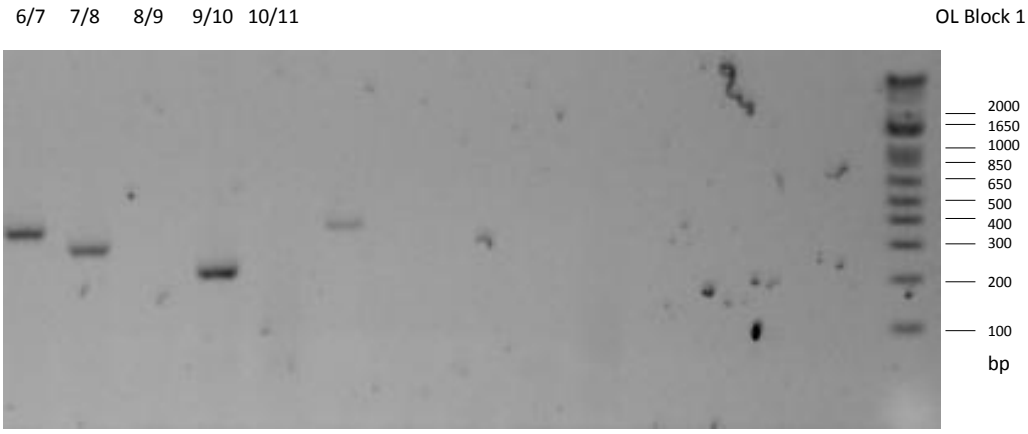
3.2.3.6 Transformation of isolated plasmids from *Saccharomyces cerevisiae* into *E. coli*

The plasmids derived from the transformed yeast cells was transformed into *E. coli* cells to increase the yield and analyze the bacterial transformants by cPCR (see 2.2.2.1.4.2.2). 1 – 5 µl yeast plasmid preparation containing 100 ng DNA was adjusted to 5 µl with H_2O and electroporated into 20 µl of fresh electro-competent DH10B *E.coli* cells as described in 2.2.1.2.2. The cells were regenerated in SOC medium and plated on LB plates with chloramphenicol for antibiotic selection. For control of the procedure 1 µl pUC plasmid was transformed with every experiment. From the grown plates single colonies were picked for cPCR and a safety plate was generated from each colony picked. The plasmids of block 2 and 7 were not transformable by electroporation as no colonies had grown for these samples. For screening the same primer pairs were used as before for cPCR on yeast colonies (3.2.3.5). The result is shown by gel documentation of 2% E-Gels in Figure 27.

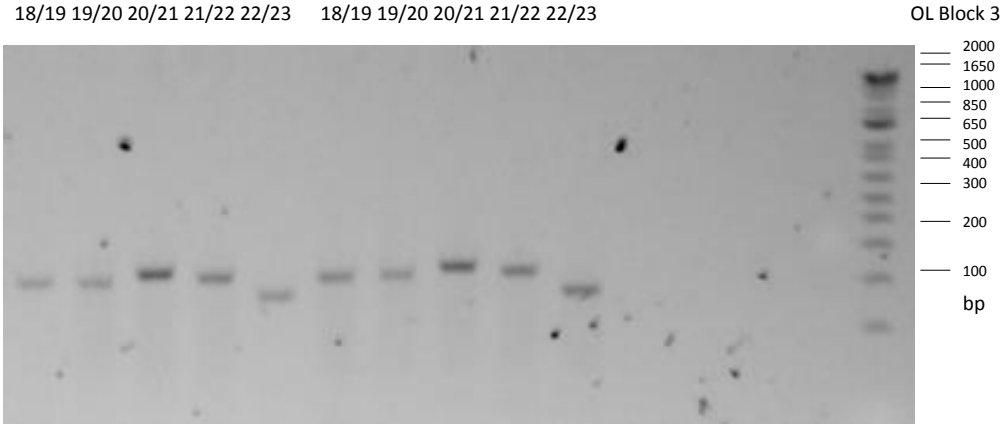
a)


Block 0 OL
0/1
1/2
2/3
3/4
4/5
2000
1650
1000
850
650
500
400
300
200
100
bp

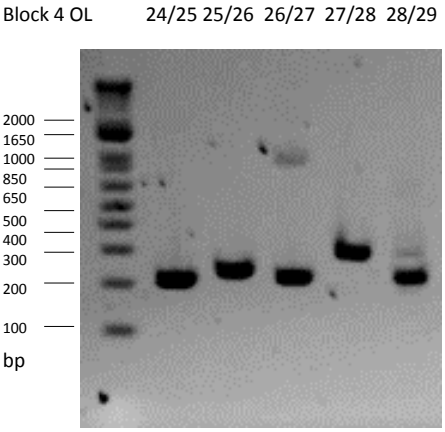
b)


6/7
7/8
8/9
9/10
10/11
OL Block 1
2000
1650
1000
850
650
500
400
300
200
100
bp

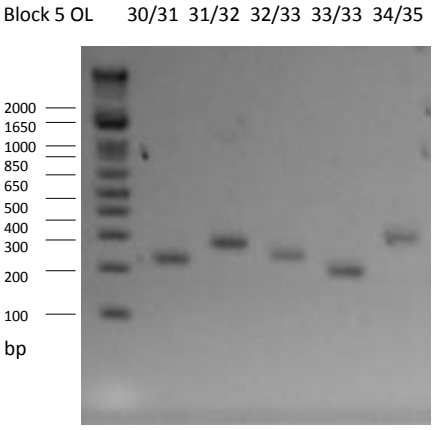
c)


18/19 19/20 20/21 21/22 22/23
18/19 19/20 20/21 21/22 22/23
OL Block 3
2000
1650
1000
850
650
500
400
300
200
100
bp

d)



e)



f)

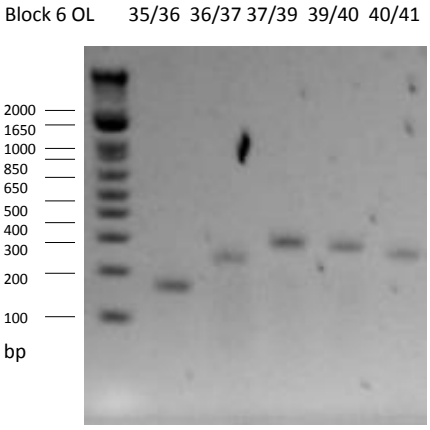


Figure 27: *E. coli* colony PCRs from the assembled blocks: 1 µl of each PCR reaction was loaded on a 2% E-Gel; amplification products are the overlapping regions (OL) between the assembled subblocks; 5 µl of 1 kb Plus ladder were used;

- **a) Block 0: The gel documentation shows the connected region between subblock 3/4 and 4/5 in the expected size of 148 bp and 176 bp; the first 3 connections are missing;**
- **b) Block 1: Here the result from 2 colonies is shown, whereas the second colony only gave the first band as a clear result; amplification of the regions 6/7, 7/8 and 9/10 are visible in the expected size of 317 bp, 259 bp and 161 bp;**
- **c) Block 3: Here the result from 2 colonies is shown; all connecting regions are amplified ad resulted in clear bands appearing in the expected sizes of 206 bp, 202 bp, 226 bp, 208 bp, 155 bp;**
- **d) Block 4: Here 3 µl of the reaction were loaded; all connecting regions are amplified ad resulted in clear bands appearing in the expected sizes of 174 bp, 247 bp, 207 bp, 187 bp, 174 bp;**
- **e) Block 5: All connecting regions are slight, shadowy bands in the expected sizes of 181 bp, 204 bp, 190 bp, 267 bp, 193 bp;**
- **f) Block 6: All amplification of the regions are clearly visible in the expected sizes of 190 bp, 240 bp, 210 bp, 171 bp and 279 bp;**

The procedures for transformation of block 2 and amplification of block 7 are described in chapters 3.2.3.7 and 3.2.3.8. As the blocks 0 and 1 still did not show the overlapping region between all assembled subblocks, the colonies were screened with 5 different primer pairs amplifying the complete sequence with expected bands at 4x 1000 bp and 1x 854 bp over the subblocks including their assembly region.

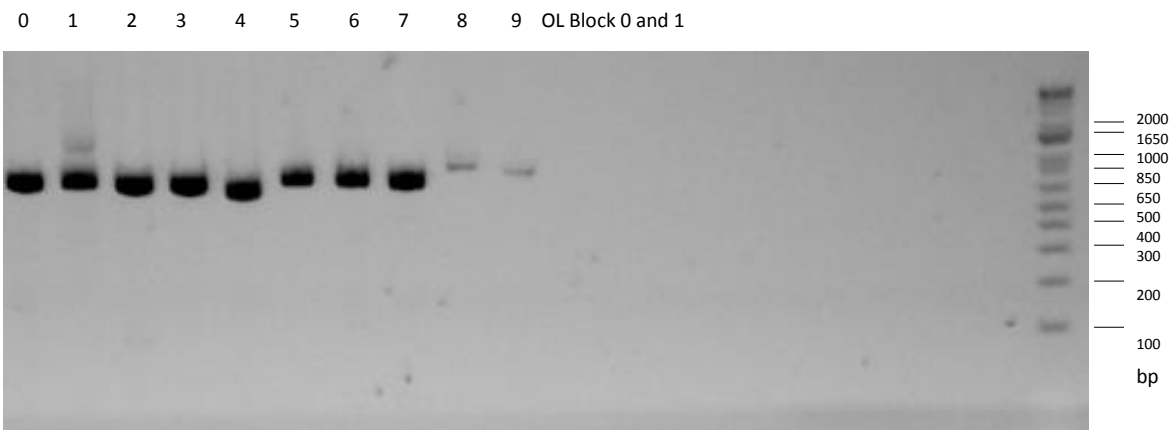


Figure 28: Amplification of the connected subblocks within block 0 and block 1: Here the amplification product is shown of the sequencing primer pairs for block 0 and block 1; all bands are clearly visible resulting in the expected sizes of 1 kb for 0, 1, 2, 3 and 5, 6, 7, 8; for columns 4 and 9, 854 bp appeared as expected; 5 µl of 1 kb Plus ladder were used as a reference.

The alternative set of primers demonstrated, that the assemblies for blocks 0 and 1 had worked as well. A reason for the former observation that not all primers worked efficiently can be that the sequence in general contains many secondary structures hindering the primers to bind well and thus affect the PCR efficiency (Snyder et al. 2008).

3.2.3.7 Amplification of block 7

As the transformation of pEYES with block 7 into *E. coli* yielded no colonies, the insert (block 7) was amplified by PCR from the whole plasmid construct (block 7 in pEYES). Therefore an amplification primer pair was designed binding exactly at each 3' and 5' end of the expected construct (see 7.1). For the reaction a 1 : 10 dilution of the plasmid isolation from a single yeast colony was used as template. The melting temperature of forward and reverse primer were 70 °C and 62 °C, respectively three annealing protocols were compared in the cycler protocol, one with starting at 70 °C with a touch down of -0.08 °C/cycle, one with a constant temperature at 65 °C and one with a constant temperature at 60 °C. The elongation time was set to 3 minutes as the template is about 4.5 kb in size. As block 0 and 1 were incomplete on colony PCR results from yeast and B2 was also transformable via electroporation, these blocks were carried along to also try PCR amplification. Figure 29 shows the PCR results of the protocol with a constant annealing temperature at 65 °C, which worked best. 1 µl of each reaction was loaded on a 1% E-Gel for verification and documentation. The reactions were set up in doublets.

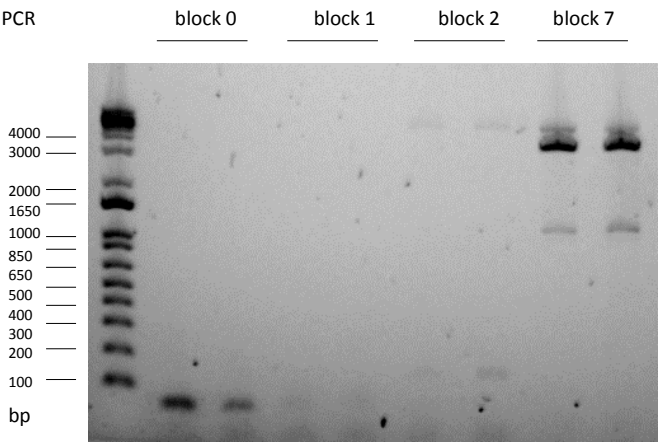


Figure 29: Amplification of blocks: 5 µl of 1 kb Plus ladder were used as a reference; Block 7 had to be amplifies by PCR as ther was no transformation into *E. coli* possible; here the PCR result is shon on a 1% E-Gel in doublets; block 0, 1 and 2 were also carried along; the last two rows show the bands appearing from sample block 7 with 3 bands from which one is in the correct size of 4.5 kb, one at 3 kb and one at 100 bp appeared; for block 2 there a 2 weak bands slightly visible in the corect size of 4.5 kb; the samples from block 0 and 1 showed no result.

The last two lanes loaded with the samples of block 7 were loaded clearly showed three distinct bands. The highest on is at the expected size of 4.5 kb. The second one is most prominent at a level around 3000 bp, both of the third and lowest band appears at around 1000 bp, which could be incomplete assembled subblocks in pEYES. This result indicates, that within one colony a variety of plasmids could be assembled. The lowest band might give a hint on a plasmid pEYES variant where only one subblock was integrated, whereas for the band at 3000 bp it could be 4 subblocks integrated into the target vector. Only the band on 4.5 kb is supposed to be correct having all 6 subblocks assembled as shown in 3.2.3.5. Thus, the upper band was cut and cleaned up by gel extraction as described in chapter 2.2.2.4.1 and used for further process.

3.2.3.8 Chemical transformation of block 2

Block 2 never showed correct colonies after electroporation into electro-competent DH10B *E. coli* cells. This might be a hint that the DNA assembled in yeast is toxic for bacterial cells. Thus, a different method for transformation was carried out on block 2 (Saida et al. 2006). 2 µl of the plasmid isolated from a single *Saccharomyces cerevisiae* colony were transformed into chemically competent D10B *E. coli* cells (see 2.2.1.2.1). After transformation the cells were plated on LB plates containing chloramphenicol (25 µg/ml) for antibiotic selection and grown 48 h at 30 °C to allow slow growth and efficient plasmid replication. After two days 25 colonies appeared on the plate. To check the presence of the right construct, a single colony was picked and cPCR was performed as described in 2.2.2.1.4.2.2 with the same control primer pairs as used before for screening the

yeast colonies. The figures below shows the result of the PCR reactions and verification on a 2% E-Gel.

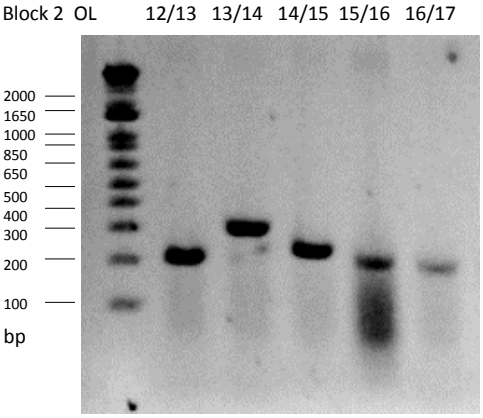


Figure 30: *E. coli* colony PCRs from the assembled blocks: 1 µl of each PCR reaction was loaded on a 2% E-Gel; amplification products are the overlapping regions (OL) between the assembled subblocks; all connecting regions are amplified and distinct visible bands appeared in the expected sizes of 174 bp, 247 bp, 207 bp, 187 bp, 174 bp; 5 µl of 1 kb Plus ladder were used as a reference.

The results pointed out, that the method for transforming and growing different plasmids is critical to a successful outcome. Figure 30 clearly shows distinct bands for every connected region between subblocks 12 – 17 building block 2, so that it can be assumed to be a complete integrated construct in target vector pEYES.

3.2.4 Assembly of blocks into segments

3.2.4.1 Linearization of pYES8D

To get started with the assembly of the derived blocks from the former assembly step, the target vector pYES8D (Baek et al. 2015) had to be linearized. As described in chapter 2.1.5, this vector enhances the positive selection of transformed yeast cells on amino acid selective media by selection on a completion of tryptophan marker gene and the 2µ-ori for induction of replication.

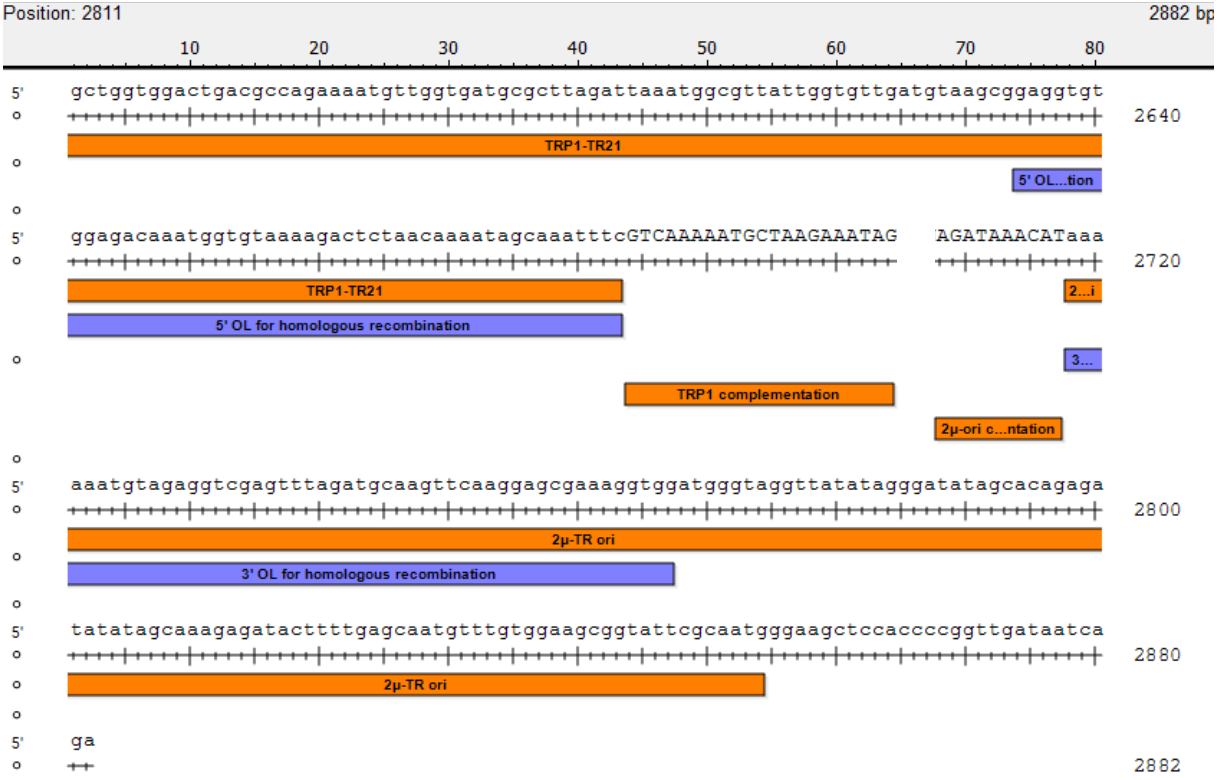


Figure 31: Plasmid pYES8D (Baek et al. 2015) complementation area: The completing sequences of the tryptophan gene and the 2µ-origin of replication gene are shon in this figure (orange); the overlapping sequence (OL) for homologous recombination is marked purple; the figure was compiled by SeqBuilder.

Figure 31 shows that this target vector does not contain a common multiple cloning site with a variety of recognition sites for restriction enzymes, and thus has to be linearized by PCR. The vector comprises a deleted version of the tryptophan gene and the 2µ-ori (orange). The tryptophan gene is lacking the 21 bp sequence 5’ GTCAAAAATGCTAAGAAATAG 3’ and the ori the 10 bp sequence 5’ AGATAAACAT 3’. These are brought by the inserts (blocks) to the vector as they flank the blocks 5’- and 3’ ends. After correct assembly, the vector will be functional of replication and can be selected on tryptophan lacking media plates. The protocol for PCR set up is described in chapter 2.2.2.1.7. 1 µl of the reaction and a 1 : 10 dilution were loaded on a 1% E-Gel and documented. The result is shown Figure 32.

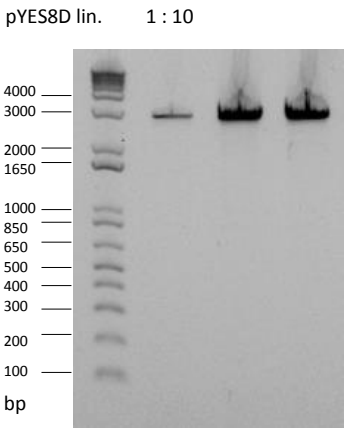


Figure 32: Linearization pYES8D by PCR: 5 µl of 1 kb Plus ladder were used as a reference; On the 1% E-Gel 1 µl of the PCR reaction was loaded together with a 1 : 10 dilution; all samples show distinct bands in the expected size of 2848 bp.

The samples appeared to deliver distinct bands around 3000 bp, so that a correct amplification of the desired 2848 bp (Baek et al. 2015) long target vector is assumed. To gain enough product, 8 reactions were pooled and dsDNA was measured on the Qubit4. A clean up of the product was not necessary, as the template is not a functional vector, which was also verified by transformation experiments.

3.2.4.2 Digestion of blocks 0 – 6

As block 7 was amplified by PCR, it was blocks 0 – 6 that were to cut from pEYES as linear DNA fragments to be assembled in the next levels step to 2 segments. All blocks are flanked by a PacI restriction site on both 5' and 3' ends. The clones found positive were grown in liquid TB_{CAM} medium to gain enough plasmid from MIDI preparation (2.2.1.2.4). Per reaction 1 µg was digested with 1 µl Fast Digest PacI enzyme and heat inactivated (see 2.2.2.2). 1 µl of the reaction was loaded on a 1% agarose gel and run at 220 V for 40 min.

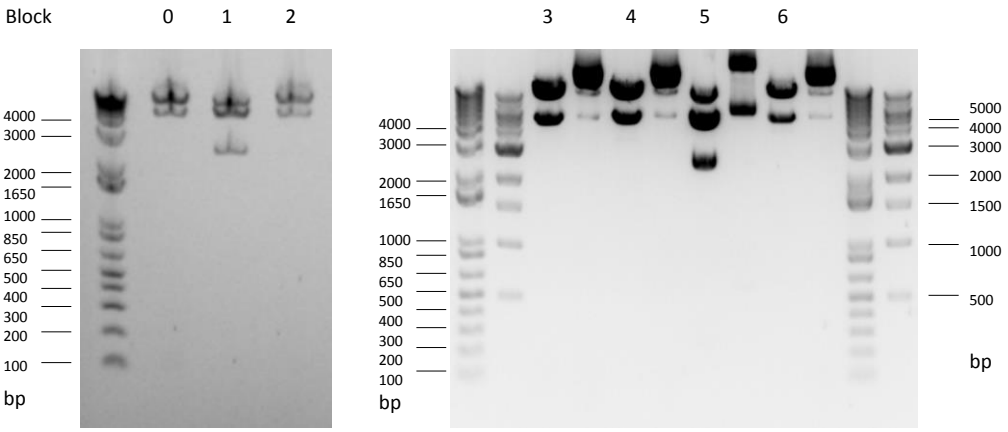


Figure 33: Digestion of assembled blocks in pEYES: 3 µl of every reaction was loaded on a 1% agarose gel and ran 40 min at 220V; blocks 0 – 6 were digested with PacI enzyme; next to block 3 – 6 a sample of the undigested construct was carried along as a digestion control. All digested samples delivered the expected bands appearing at 4.5 kb and 9.5 kb of vector backbone; block 5 and block 1 had a third unspecific band appearing at 2.5 kb. 5 µl of 1 kb Plus ladder and 7 µl of NEB 1kb ladder were used as a reference.

Figure 33 shows the documentation of the digests of blocks in pEYES. All samples clearly show the desired band at 4.5 kb which is the predicted size of the block insert indication a complete assembled construct. Atop of these bands, the vector backbone with around 9.5 kb is visible as well. Block 1 and 5 show an unspecific third band appearing at 2.5 kb. This can either be contamination with a different construct or as presumed before, plasmids where only a part of the subblocks were assembled and integrated into pEYES during yeast assembly. Then a mixture of fully assembled and partly assembled constructs appears within one yeast colony, which might lead to a carryover to bacterial cells and there promotes a mixed culture as well. To get rid of both, backbone and undesired bands, the digest reactions were completely loaded on a 1% agarose gel to be cleaned up by gel extraction (see 2.2.2.4.1) and measured on Qubit .

3.2.4.3 Assembly of 2x 4 blocks into segments

After the restriction digest to release the assembled blocks from the target vector backbone, the next assembly step was set up. Block 0 – 3 and block 4 – 7 shall each be assembled to segments. The procedure was the same as before (see 3.2.3.4) but now the linear target vector pYES8D was used to prevent former vector being carried over. Therefore it has a different selection marker on tryptophan (histidine in pEYES) and an improved positive selection system (see Figure 31) by completion of the marker site and the 2µ-ori site.

Segment	Blocks 300 ng each	Target vector 500 ng	Colony forming units Block	NC lin. vector	PC GFP-pYES8D	NC cells	Transformation rate
0	0 - 3	pYES8D	12	0	50	0	7 cfu/µg
1	4 - 7	pYES8D	8	0	49	0	5 cfu/µg

Table 54: Assembly of segments: Number of colony forming units in selective media (-Trp) plates after transforming 300 ng of each linearized subblock and 500 ng of linear target vector (1700 ng DNA/transfection); two negative controls were carried along: one with linear target vector (500 ng) and one with untransformed cell, both treated with the same transfection mix and both show no cell growth as supposed; a positive control with the plasmid vector (500 ng) was carried along as well; the table shows which subblocks were combined to their respective blocks.

To check the success of transformation, a positive control construct was carried with the experiment. GFP-pYES8D is the simple sequence of green fluorescent protein (GFP) integrated in pYES8D including the completing sequences for the tryptophan auxotrophic marker gene and the 2µ-ori. Again two negative controls were set up, one with transforming the linearized vector and one without any DNA being transformed. Both negative controls showed the expected result with no cells growing on selective media plates. The positive controls showed 49 and 50 colonies appearing after 3 days. The segment assemblies showed colonies as well, but very few compared to the positive control. A safety plate was generated by spreading some cell material on it of the colonies that were picked for colony PCR verification.

3.2.4.4 Yeast colony PCR from segment constructs

As described in 2.2.2.1.4.1, a single yeast colony was picked with a sterile tip in diluted in 0.02 M NaOH and boiled on 95 °C for 10 min to get access to the DNA material inside the cells. Colony PCR was performed as before with 3 µl of the dilution for every reaction and primer pair. The primers amplified also here the connected region the adjacent blocks that were transformed. Thus 3 primer pairs for every segment colony were used. Verification and documentation was done by loading a 2% E-Gel with 3 µl of the PCR reaction. To verify, that the cPCR reaction itself had worked, a control primer pair was tested on the colonies, that amplifies the mating type region in the yeast genome, delivering a band at 369 bp (Illuxley et al. 1990).

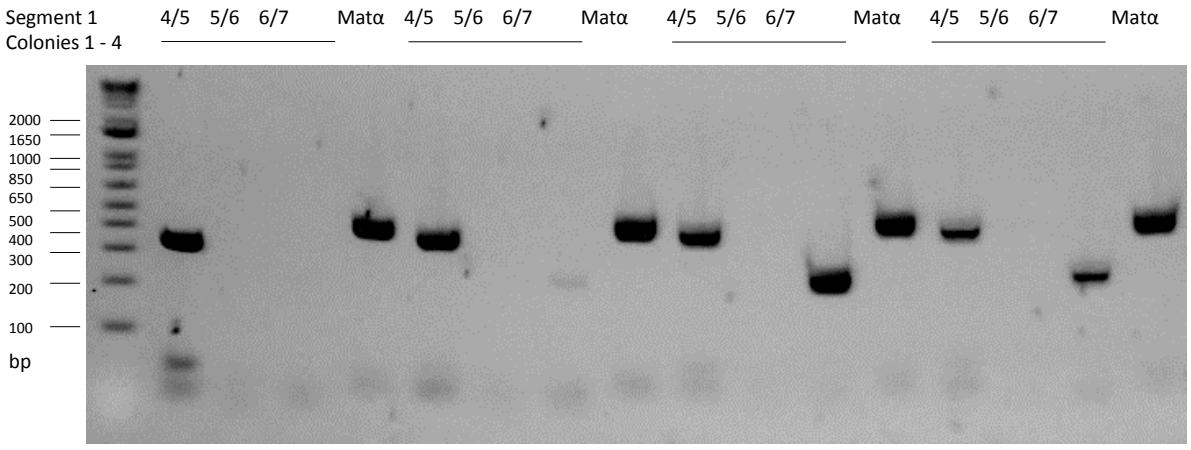


Figure 34: Yeast colony PCRs from the assembled segment 1: 3 µl of each PCR reaction was loaded on a 2% E-Gel; amplification products are the overlapping regions (OL) between the assembled blocks. Here, 4 colonies are shown that were screened for overlaps; OL 4/5 resulted in a distinct band in every colony screened at the expected size of 309 bp; in colonies 3 and 4 the OL of block 6 to 7 is also resulted in a distinct band at 161 bp; the control primer showed in every colony the expected size at 369 bp, standing for mating type alpha. 5 µl of 1 kb Plus ladder were used as a reference.

Unfortunately, the screened colonies of segment 0 showed no results on the gel (not shown), except of the mating type cPCR control primers. Segment 1 resulted in distinct bands for the first overlap between blocks 4 and 5 at a level of 309 bp as expected. For two colonies (3 and 4) prominent bands appeared at the size of 161 bp, which are the amplification product of the overlap from block 6 to 7. Here also the mating type control primers resulted in the expected amplification product from mating type alpha.

3.2.5 Transfection of GPTEC-T cell line with block constructs of Guinea Pig Adenovirus

The constructed blocks were transfected into guinea pig tracheal cells GPTEC-T (see Table 3: Guinea Pig cell line) to investigate the RNA on viral transcripts. Therefor the linear blocks were used and assembled *in vitro* with the NEbuilder Kit as described in chapter 2.2.1.3.2 using Lipofectamin2000. The assembled blocks were used for transfection of (GPTEC-T) to investigate the RNA on viral transcripts. Based on the predicted annotation shown in Figure 35 (annotation by Dr. Helga Hoffman-Siebers, HPI, Hamburg), the blocks were referred to the appropriate genes located along the linear Guinea Pig Adenovirus genome. It revealed that the pre-terminal protein (PTP), which binds to the ends of the linear DNA to induce replication, is mostly located on block 2, but a part aligns to block 1. Due to this finding, block 1 and 2 were *in vitro* assembled and directly used to transfect the GPTEC-T cells expecting the transcription of PTP and the polymerase (compare Figure 35). A second assembly was performed using the NEbuilder kit, which was the joining of all 8 blocks. With this sample another aliquot of cells was transfected to investigate the

possible formation of virus particles in cell culture. The other blocks were transfected individually, to detect the transcription of the early structural proteins E1A, E1B small unit, E1B, E3 large unit and E4orf6. Also Hexon and Penton, that have conserved sequences, were predicted to be transcribed followed by the DNA binding protein (DBP), the protease and the late genes L1 and L4. Upon the experiment a confirmed annotation was planned after performing a complete transcriptome analysis. The results of isolated RNA is shown in Table 55. This isolated RNA was reverse transcribed into copy DNA (cDNA) and then amplified with specific primers matching the genes for E1A, polymerase, Penton, Hexon, DBP, L1 and E4orf6. Unfortunately, the bands appearing in the samples from cDNA also appeared in the RNA samples, indicating plasmid contamination of the RNA samples.

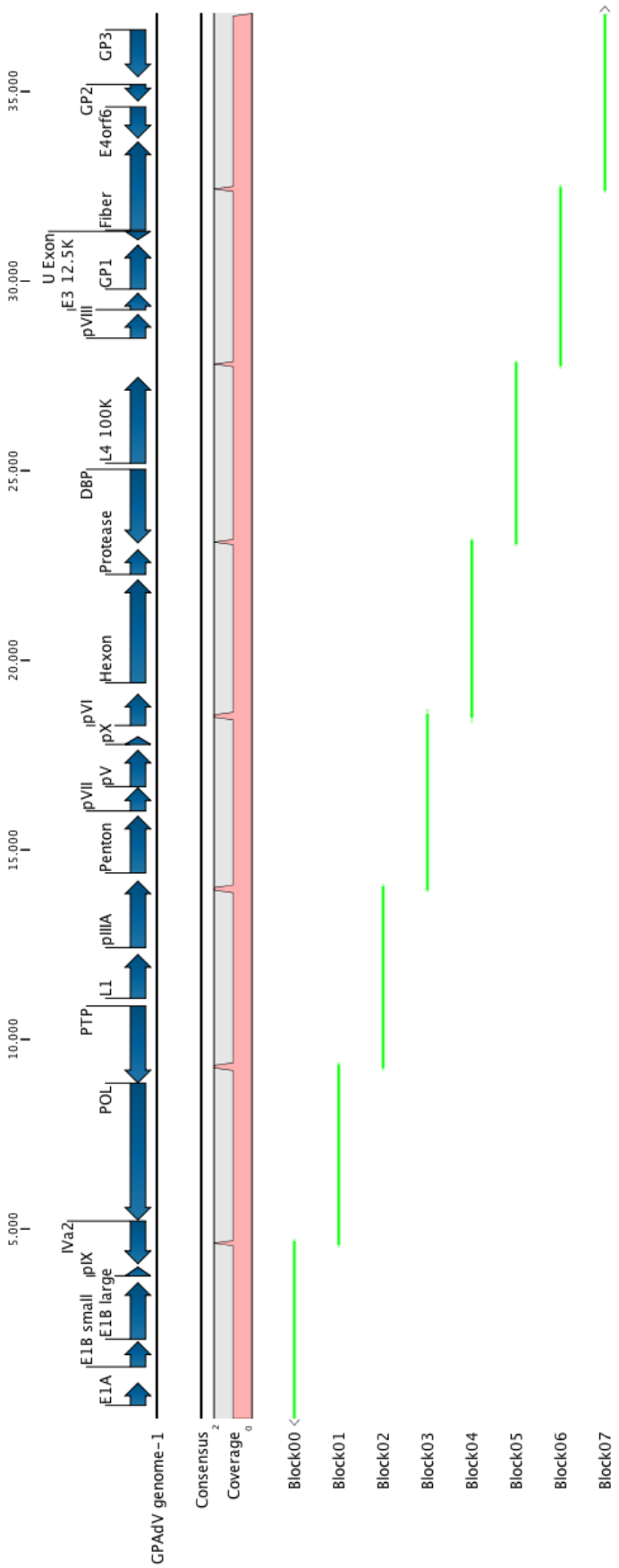


Figure 35: Overview of the predicted annotation of the GER1 Guinea Pig Adenovirus isolate (by Dr. Helga Hoffmann-Sieber, HPI, Hamburg): Main domains starting from 5' end: E1A, E1B (small and large unit), polymerase, pre-terminal protein (PTP), L1, Penton, Hexon, protease, DNA binding protein (DBP), L4, E3, Fiber, E4orf6 as well as the glycoproteins 1 – 3 and the polypeptide; blue arrows mark the genes; consensus sequence and alignment coverage are shown in grey and pink bars; blocks are marked by green bars.

3.2.5.1 RNA analysis

The transfected cells were harvested after 48 h to isolate and measure the RNA levels (see method description in chapter 2.2.1.3.3. Each sample was the measured on Qubit 3 and analyzed on a Bioanalyzer (performed by NGS facility department of HPI, Hamburg). All samples were measured as a 1 : 10 and a 1 : 100 dilution.

Sample	Concentration (ng/µl)	
	1 : 10 dilution	1 : 100 dilution
Block 0	104	13.7
Block 1+2	91	11.9
Block 3	98	13.1
Block 4	108	13.7
Block 5	114	14.2
Block 6	91	10.9
All blocks	60	7.4
mock	85	10.3

Table 55: RNA concentrations: The transfected GPTEC-T cells were harvested and the entire RNA isolated; shown here are the concentrations measured on a Qubit 4.0.

3.2.5.2 Transcription of RNA into cDNA and analysis by PCR

For the transcription of RNA into cDNA 10 µl were used of all RNA samples (see concentrations in Table 55). After the procedure was processed as described in chapter 2.2.1.3.4 using the High Capacity cDNA Reverse Transcription Kit from AppliedBiosystems, the samples were stored at -80 °C until a PCR was done with defined primers for the E1A, Polymerase (Pol), Penton, Hexon, L1, E4orf6 and DNA binding protein (DBP) regions. The PCR reaction was performed by Dr. Helga Hoffmann-Sieber at HPI in Hamburg.

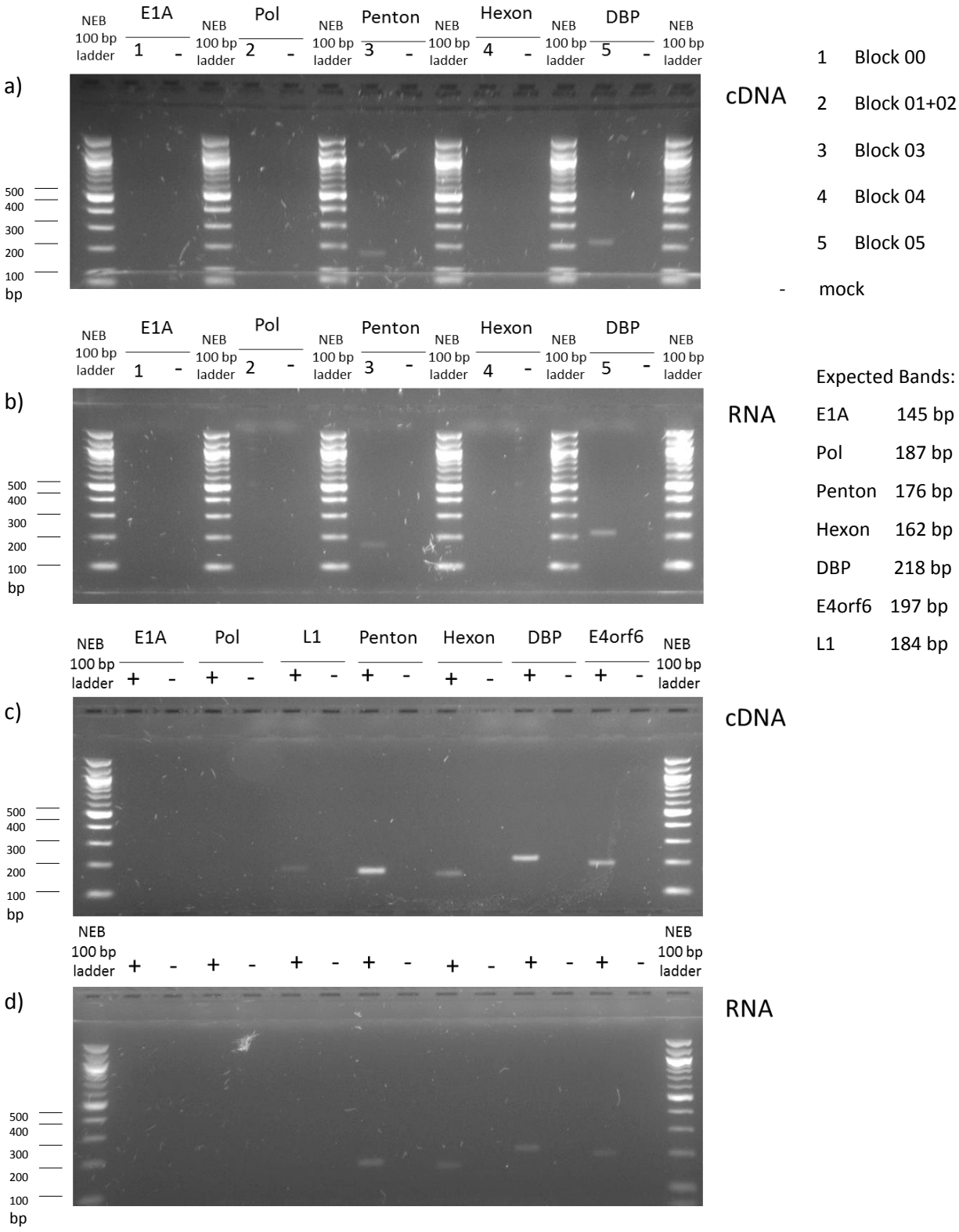


Figure 36: PCR from cDNA samples: as a reference the 100 bp lader from NEB was used; a) Here the results of PCR using specific primers to amplify the E1A, Pol, Penton, Hexon and DBP region is shown; for each sample the reaction and an empty control reaction were loaded on an agarose gel; c) Here the sample containing the complete block assembly (0 – 7 in one piece) shows clear bands L1, Penton, Hexon, DBP and E4ofr6 primer pairs; the picture a) and c) show the results on cDNA, whereas picture b) and d) show the results on the RNA samples.

Picture a) shows bands on the expected size of 176 bp and 218 bp for the regions of Penton and the DNA binding protein. In Figure 36 c) visible bands could be detected representing L1, Penton, Hexon, DBP and E4orf6 regions. The same bands appearing from the cDNA samples (pictures a and c) also appeared from the RNA samples (pictures b and d), which indicates DNA contamination in the RNA samples. The reaction has to be repeated for a clear determination of amplification product.

4 DISCUSSION

4.1 Relevance of an abridged gene synthesis

Synthetic biology is a steadily increasing field including biotechnology, pharma industry, academic research, medical technology and the industry producing all tools that are needed reaching from mass spectrometry devices, chemicals, computational devices as well as synthetic DNA. The development of new methods in all of the areas mentioned proceeds more and more rapidly and thus forces the production of genes and artificial DNA to become faster. All synthesis approaches that are common are based on the enzymatic assembly of small oligonucleotides by enzymatic reactions including ligases and mainly polymerases. In a complete manufacturing process ordering, oligonucleotide synthesis and assembly, cloning, quality control and also shipment are involved to fulfil the demand of quick access for industry and research. The underlying protocols of SCR, SPCR and FPCR in this thesis for gene synthesis build the foundation for the implementation. Thus, reduction of the turnaround time is a crucial way to keep up with progressive advance and need. Comparable technologies like sequencing have been further improved leading to high throughput sequencing methods that were inconceivable in former times.

4.2 Application of the Plackett-Burman-Design tool on gene synthesis

The Plackett-Burman-Design (PBD) tool allowed the concurrent consideration of numerous parameters. These are determined with a certain amount of scenarios necessary, instead of a full factor combination approach. In this thesis, 12 scenarios were investigated on a final test set of 96 various fragments. To apply the PBD tool to any approach, the model's parameters have to be defined prior to filling the excel tool as shown in 3.1.1.1. (Beres and Hawkins 2001). As for this study time reduction of PCR protocols to speed up a gene synthesis workflow was the goal, the parameters were defined as denaturation time, annealing time, elongation time and the number of cycles in PCR. The ascertained duration achieved in the resulting protocols (see Table 46: New protocols for SCR, SPCR and FCR in overview.) showed to be eligible for application on this set of samples. The result of the PBD (see Figure 14) gives a score which indicates a high impact of the determined factor when the score is high and poor impact when the score number is low. It had clearly shown the sensitivity of the reaction to changes in the annealing time expressed by a score of 3.1250. In contrast to that result, the denaturation time and elongation time with a score at -

1.0417 and 1.0417 have less impact on the reliability of yielding products. Additionally the number of cycles performed during the reaction resulted in a score of 2.0833. Poor yield of the SCR product leads to a less efficient reaction in SPCR. The repeating of denaturation, annealing and elongation in cycles states the amount of amplification product and is therefore only reduced a little to 27 cycles instead of 30. The amount of SCR product is critical for the following SPCR reaction, where an exponential amplification takes place. In contrast to SCR, 20 cycles in SPCR still delivered enough DNA fragments suitable for proceeding with FPCR, where the sub-fragments are fused together. In this third FPCR the most significant time reduction was accomplished by 56%. In SCR a time reduction of 34% was achievable clearly showing that this is the most sensitive part of the process to optimization. Resumed, the turnaround time in the underlying gene synthesis and assembly process was decreased by 47% and according to that can be established in almost half of the time.

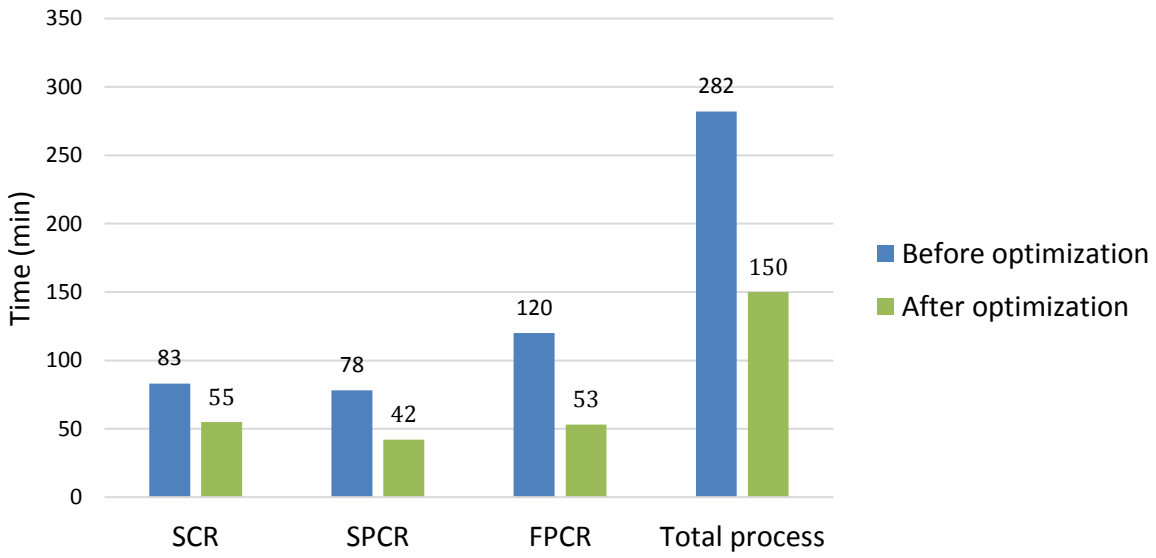


Figure 37: Turnaround time in comparison: overview on SCR, SPCR, FPCR and the total process; the blue bars represent the initial times and the green bars the times after PBD optimization; time is given in minutes.

4.3 Outlook on the application of the abridged gene synthesis

The optimization of gene synthesis PCR protocols was successfully performed by using the universal applicable statistical method of a fractional factorial design by Plackett and Burman. In this study, to reach a total reduction time of 47% the sequences producible were limited to 35 – 58% in GC content and 300 – 1500 bp in size. This represents a main part of the ordered sequences and is thus worth for further optimization. Other factors involved in the process, like temperatures in the different protocols, the involved enzymes and devices can be object of detailed

examination. Gene synthesis in production companies does not only happen in the assembly itself, but much more in ordering, nucleotide synthesis and supply chain management. All components involved can be investigated for improvement with the Placket-Burman-Design tool. For now, the protocols were implicated in parts in the daily production process and applied to feasibility of certain projects.

The advantageous improvements resulting from this studies are not only for research use, but also for medical application. In emerging disease like influenza it is critical to gain very quick access to antigen genetic material for the development of efficacious vaccines (Nogales et al. 2014). In former studies on the Ebola outbreak in 2014, adenoviral vaccines were effectively introduced expressing the Ebola virus glycoprotein, which represents an appropriate application of this protocols.

Today, there are companies developing therapies based on personalized DNA vaccination of patients who put confidence in being treated in time (Shevelev and Pyshnyi 2018; Melief 2017). With the achieved protocols as a result of this thesis for fast gene synthesis, so called neo-antigens (see 1.4.3) were recently successfully produced and cloned into an adenoviral vector, exactly for this purpose. These very individual and specific sequences require a reliable producibility, correctness and delivery.

4.4 Investigation of the Guinea Pig Adenovirus sequence and partitioning

The *in silico* adenoviral sequence was analysed to determine the structure of bases and the significant inverted terminal repeats (ITR) characteristic for adenoviruses. The sequence of this newly found virus is 37,070 bp long with a total GC content of 62.03%. The Genome Partitioner creates the fragments in a higher order assembly approach, meaning that the very first constructs (subblocks) that are assembled, already comprise the sequence for next assembly level (blocks) and the one after (segments). The computationally added sequences inherit all cutting sites and the transition sequences to either the related fragment or target vector or both, named subblock adapters (see an example of subblock 0 in Figure 38.

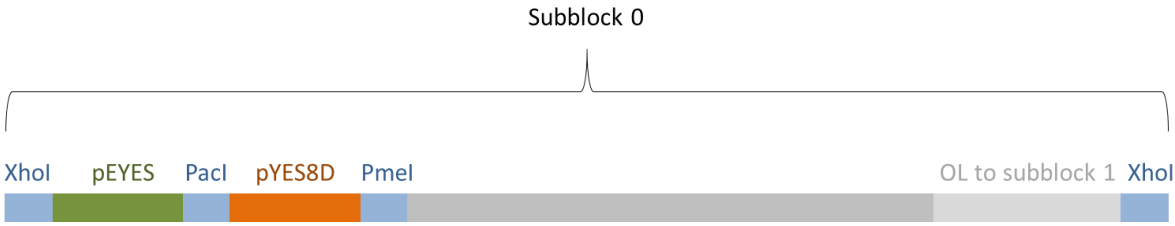


Figure 38: Sequence structure of subblock 0 including the subblock adapter and overlapping region (OL) to subblock 1; the adapter inherits all restriction sites XhoI, PacI and PmeI (blue) and connection regions to the target vectors pEYES (green) and pYES8D (orange) that are needed for each assembly level contributing to the higher order assembly.

Here subblocks 0 – 5 are assembled to block 0, so subblock adapter 5 contains both the connection to pEYES and the PacI restriction site as well as the overlapping part to subblock 6. This system is consistent for all created fragments and thus convenient for the purpose of reaching large constructs by eliminates the risk of mistakes that can easily occur designing the adapters manually. The underlying algorithm selects terminal homology regions (THR) that represent the overlaps, which do not include secondary structures, but are unique among all THRs and have a moderate GC content. This is achieved by computational shifting a window frame along the sequence, ensuring that the subblocks will not differ in size for more than 10%. If any disturbing motif is detected in an adapter region, the region is shifted so that the motif is split in two parts or lies absent of this area (Christen et al. 2017). The principle is based on homologous overlaps between the fragments (subblocks, blocks and segments) due to a higher order assembly. The segments each represent the halved genome and the tool creates an overlap of 100 bp.

Another problem was caused by the presence of several repetitions in the genomic sequence each more than 12 bp long. Appearance of these in the THRs would lead to a mis-assembly, meaning that the subblocks or blocks are assembled in a wrong order. Here, the tool guarantees no repetitions in the THRs that are longer than 7 bp.

The ITR was found to flank the 5'- and 3' ends with 167 bp in reverse order. This special feature plays a crucial role for the genome partitioning by the algorithm underlying the web interface. This special feature of an inverted terminal repeat must technically be seen as an overlap at the aspect of homologous recombination. As shown in Figure 9 the segment size was set to the full length of 37.070 kb, but due to the algorithm the result was two segments each 18.732 kb. With the assembly method it might happen, that the two segments assemble on the ITR regions resulting in an "inside out" version of the viral genome with the ITR in the middle.

GC-rich areas in DNA sequences are not only critical for amplification, but are also a crucial aspect in gene synthesis as the system is also based on chain reactions including polymerases and oligonucleotide primers (Mamedov et al. 2008). One possibility to get rid of GC-stretches is to search for the coding sequences that have to be kept and leverage genome optimization upon the

remaining areas (Benjamini and Speed 2012). The redundancy of the genetic code enables to change codons without altering the amino acid sequence. Secondary structures and repetitions appearing throughout the sequence make gene synthesis defiant and a synthesis of this linear genome a challenge. An annotation therefore would be necessary to perform optimization upon the GPAdV genome. Here, the viral genome sequence is recent and neither officially annotated nor have deeper investigations been performed on its behaviour *in vivo*. Therefore, it is important to really synthesize the wild type artificial DNA. As this thesis' goal is to build large constructs of the adenoviral genome, any optimization to the original sequence is improper.

One disadvantage of the strategy used in this study is, besides the restrictions due to the ITR, that the system is not modifiable once the first assembly step took place. In case, the target vector shall be exchanged or the restriction enzyme, it is hard to adapt all the fragments by PCR or redesign and order new subblocks. The approach therefor is very rigid, but reliable concerning functionality. The primers that are designed by the tool all have the same melting temperature and hence are suitable for one particular PCR protocols applicable on all cPCRs.

4.5 Assembly of blocks in *Saccharomyces cerevisiae*

To assemble the double stranded linear DNA of each six subblocks into respectively one block, the target shuttle vector pEYES was used. The growing colonies are then selectable for the presence of plasmid by a marker encoding the gene for auxotrophic histidine metabolization. The plasmids containing each a subblock were digested to excise the desired insert. A problem appeared, when some subblocks were found to inherit an internal restriction cutting site.

For those subblocks 5, 14, 15, 23 and 45- 47 amplification of the sequences was performed (see 3.2.3.2), as a digest was not applicable due to. As described in Figure 25, the colony formation appearing after the assembly reactions varied considerably in number. Although the same amount of insert was added to the transformation mix, the colonies for example for block 4 were 100x higher than in the assembly of block 7 or 2. This result points toward the idea, that the composition of inserts to assemble plays a crucial role for homologous recombination in yeast. In contrast to block 4, where all subblocks were prepared by digestion, the blocks 2 and 7 included 2 respectively 3 inserts prepared by PCR. In blocks 0 and 3, only one insert was derived from amplification. A resulting dissonance of inserts might lead to a less effective recombination event. The mixture of differently prepared donor DNA seems to play a significant role for assembling efficiency.

Although photometric measurement was performed to detect only double stranded DNA, there can be an undefined amount of abbreviated DNA strands remaining after PCR. Impure PCR

reactions can be purified for example by gel extraction. A potential improvement can also be the removal of template from the PCR reaction by digestion with DpnI, what leads to a cleavage of only methylated DNA and thus selects for the amplification product (Mierzejewska et al. 2014).
Furthermore the number of cfu among the positive control, transforming 500 ng of plasmid target vector, indicates variations in transformation efficiency. This can be due to fluctuation of competence of the cells or sensitivities to slight variations of the transformation mix.

4.5.1 Screening of colonies from *Saccharomyces cerevisiae* and *Escherichia coli*

The colonies resulting after subblock transformation were screened for a complete and correct assembly of the inserts. Figure 26: Yeast colony PCRs from the assembled blocks: 3 µl of each PCR reaction was loaded on a 2% E-Gel; amplification products are the overlapping regions (OL) between the assembled subblocks; 5 µl of 1 kb Plus ladder were used; clearly shows the inefficient cPCR reaction resulting in mostly very faint bands. Poor template material is the most likely reason for this monitoring. To get access to the DNA material of yeast cells, the cell wall has to be busted. Here a boiling of the picked colonies was performed in 0.02 M sodium hydroxide. Although the accessibility was given by this method, the dilution of genetic material that comes along leads to a lower yield of template. In terms of the high content of genomic DNA in yeast in contrast to *E. coli* preparations, the amount of accessible plasmid is low in relation to the genomic DNA. Enzymatic treatment with lyticase prior to cPCR could enhance the digestion of the cell wall and avoid a counterproductive dilution of template.
The reaction set up containing 1M betaine final gave the best results among the tested protocols and recipes in former studies. Betaine acts as an isostabilising agent, enhancing the stability of the DNA duplex with oligonucleotide primer during PCR (Frackman et al. 1998). The extended annealing time supports this effect, resulting in detectable amplification product. Still, gel documentation had proven to be hard due to the very faint bands appearing that just give a hint. It was clearly shown here, that multiplication of the plasmid in *E. coli* is necessary to yield workable amounts.
In contrast to yeast, the screening PCR performed from *E. coli* colonies revealed clear bands on gel documentation. The primer pairs designed by the Genome Partitioner tool all have the same melting temperature, so that one PCR protocol can be applied to all assembly determinations. This allows a quick and straightforward workflow of colony screening from bacteria. In case of block 4, only in *E. coli* cPCR all primers had bound to template. This can be due to an unintended primer binding in the yeast genome, leading to a false negative result in PCR.

Further investigations had to be done for the transformation of block 2 and 7, which did not lead to colony growth in *E. coli*. For block 7, different target cells were tested like stable cell lines of *E. coli*, strain DH5α and a different species *Vibrio natriegens*, all without success. To gain the whole construct for the next order assembly step, the complete construct was amplified by PCR as described in 3.2.3.7. Block 2 instead was successfully transformed chemically with adapted growth conditions of 48 h at 30 °C. That this method worked for block 2 indicates that with the plasmid toxic genes might have been brought into the bacteria, which potentially results in significant defects in growth or cell death after gene expression (Saida et al. 2006). Toxicity is not only occurring after the translation process within the cell, but is also frequently due to toxicity of the DNA sequence itself. Codon optimization prior to transformation can evade this problem (Kimelman et al. 2012).

4.5.2 Preparation of blocks and next step assembly to segments

As defined in the Genome Partitioner tool, the blocks were created with a flanking PacI restriction site on both ends each. Block 7 was amplified as discussed before in chapter 4.5.1. Growing conditions were optimized by using TB liquid medium for cell growth and 0.02% arabinose, to switch the l-arabinose-inducible oriV from low to high copy (Wild and Szybalski 2004). Former experiments of this study showed, that the yield of plasmid DNA was increased by using TB instead of LB due to its higher levels of yeast extract (Wood et al. 2017). The digest with PacI showed other bands at around 2500 bp and 2000 bp (block 1 and 5 see Figure 33), which indicates that there is the chance of yeast recombining different plasmids. Some of them might inherit only 2 – 3 fragments, resulting in a smaller digested product on gel documentation.

For the next assembly step, the target vector was changed to pYES8D prevent any carryover of former plasmids leading to false positive results in colony forming. The analysis of the colonies appearing for the assembly of segment 0 revealed, that the completions were assembled into the pYES8D target vector, but the inserts respectively blocks 0 - 3. This result might be hinting on a selective recombination in yeast cells, as only the short missing sequences that are necessary for a functional plasmid were singled out.

In segment 1, the colony PCR presented 2 of 3 expected bands in gel documentation. This can be a hint on an incomplete assembly of only 3 of 4 blocks. Another explanation might be, that the blocks were assembled I a wrong order due to homologies in the overlapping sequences of different blocks. Then a different connecting region would appear that the primers are not suitable

for. Here no conclusive resolution was yet performed to demonstrate a successful segment assembly.

In summary, 8 correct blocks were assembled representing the wild type adenoviral genome. A graphic overview is given to the status quo of this study.

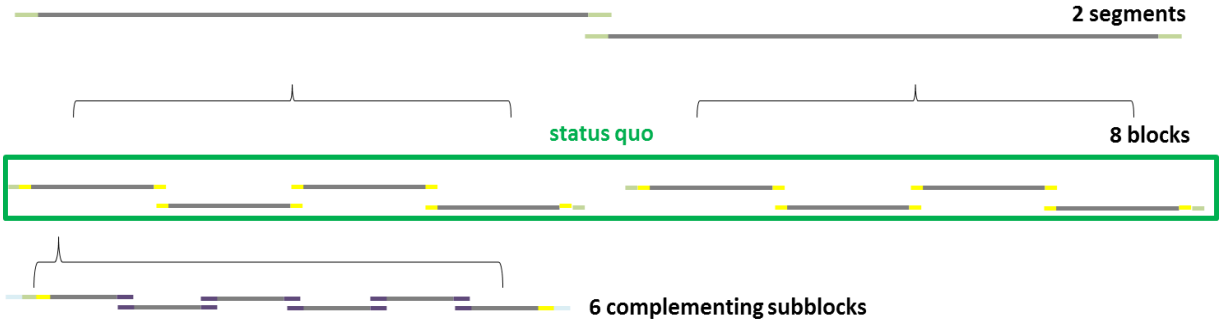


Figure 39: Overview of the large fragments built of the Guinea Pig Adenovirus genome; 6 exemplary subblocks are shown that are matching to build a block; in the green frame the status quo is shown of finalized 8 blocks; in the upper part, the 2 segments of the next level assembly are given.

4.5.3 Comparison of pEYES and pYES8D as target shuttle vectors

In this study, a vector with a newly created function improving the positive selection of transformed cells was introduced. The shuttle vector pYES8D is described in the chapters 2.1.5 and 3.2.4.3. After correct integration of the completing sequences for the tryptophan marker gene and the 2µ-origin brought by the inserts to assemble, the plasmid is functional of replication in yeast and selectable on the auxotrophic marker tryptophan. The vector was tested by inserting the simple sequence of green fluorescent protein (GFP) flanked by the completions (Baek et al. 2015). For each block assembly experiment this construct was carried along as a positive control for transfection, which was affirmed by colonies appearing after 3 – 4 days (see Table 54). The other vector pEYES used for subblock assembly is described in Table 53.

Both were transformed as plasmids for positive controls in the assembly experiments with transforming 500 ng of each. Comparing the results in the number of colony forming units, pEYES resulted in around 5x more colonies than pYES8D. As both vectors differ considerably in size (pEYES 9540 bp and pYES8D 2848 bp), more colonies were expected for pYES8D. Another aspect is the difference in their replication origins. The pEYES inherits CEN6/ARS4, which is known to result in 4 – 8 plasmids per cell. In contrast, pYES8D uses the 2µ, which is endorsed for yielding high copy numbers of 28 – 58 per cell (Karim et al. 2013).

In contrast to pEYES, the transformation of the linear vector pYES8D never lead to colony forming. The risk of incomplete digestion or re-ligation of pEYES instead, resulted in colonies appearing on

the negative control plates. Thus, the usage of the double selective pYES8D vector improves the analysis of colonies by prevention of false positive cell growth. Another aspect is, that the transformation of smaller vectors leads a more effective recombination event due to the ratio of target vector size and number and size of inserts to assemble (Gietz and Schiestl 2007).

4.6 Transfection of GPTEC-T cells and RNA analysis

It is known that the transfection of recombinant structural genes can lead to virus formation in cell culture as shown on the example of an assembly of human papillomavirus type 16, resulting in virion-like particles (Zhou et al. 1991). In this study, no viral particles could be detected after transfection. Thus, further investigations have to be performed to identify possible obstacles hindering the formation of a replication enabled genomic full-length DNA. The analysis of the transcribed DNA hints at a contamination with DNA, as the same bands appeared for the PCR of the RNA samples. Thus, the experiment has to be repeated with using DNAse on the RNA samples to get rid of the DNA (Huang et al. 1996).

4.7 Outlook

Synthesizing large fragments of genomic sequences is a desired goal for many approaches in both academic and industrial research. Here, the underlying subject was a newly found wildtype guinea pig adenovirus genome. The complete synthesis of 8 blocks was successfully performed using the computational tool Genome Partitioner and the *in vivo* tool of homologous recombination in yeast. Future studies on assembling the whole construct require some improvements of the concept. One is the dilemma of the inverted terminal repeats (ITRs) flanking the viral genome, that compete with the created overlapping sequences for recombination. To prevent the fragments from assembling in a wrong orientation, the ITR could be synthesized separately including a short part of the sequence 5’ and 3’ outside the ITR region as specific overlap (OL) and pre-integrated into the target vector pYES8D. The idea of a redesigned target vector is shown in graphic sketch in Figure 40.

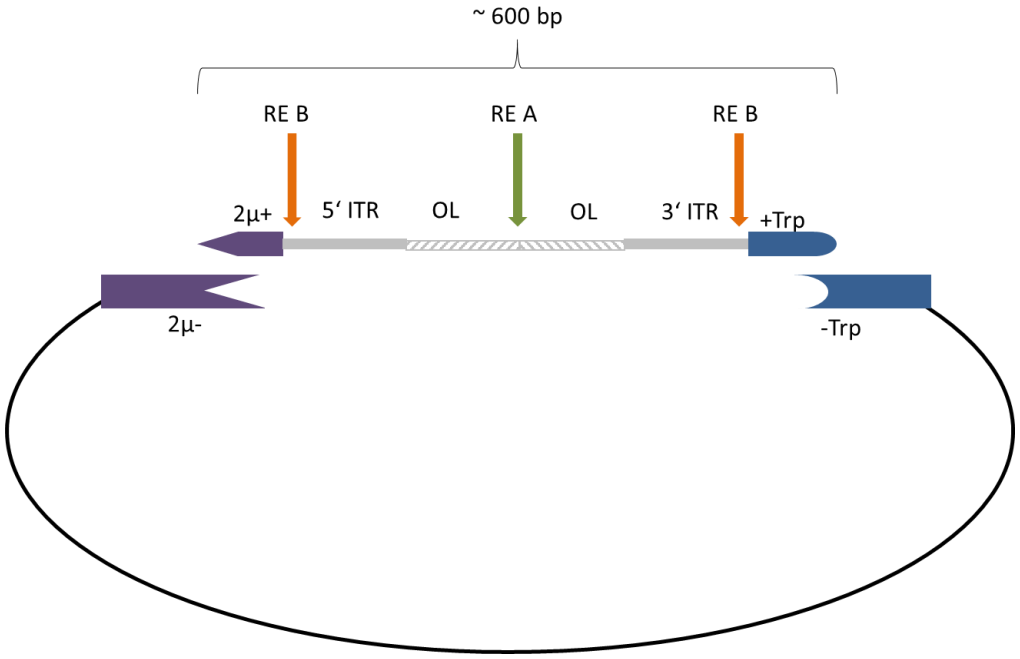


Figure 40: Design of an assembly strategy to prevent the dilemma with ITRs competing with overlapping sequences (OL) in fragments design.

By this approach, the ITRs can be integrated in pYES8D by recombination and completion of the tryptophan marker gene and the 2µ-origin. The insert to synthesize must inherit of course the 5’- and 3’-terminal ITRs, some overlapping sequence homologous to the remaining viral genome that can be assembled separately. One restriction enzyme (RE A) then opens the target vector (green arrow) for homologous recombination on the last level and the final construct van then be released by digestion with the restriction enzymes B (RE B, orange arrow).

The results in this study show, that a full synthesis of a wild-type viral genome as the guinea pig adenovirus one is possible in theory without any optimization or changes to the sequence. Still, changes to the sequence that do not affect the translated product, can help to evade the challenges discussed in particular. Yeast appeared to be a convenient tool in this case, as it shows no sensitivity to potentially toxic sequences. It is known, that the optimization of open reading frames is used to attenuate viruses and leads to a loss of virulence (Nogales et al. 2014; Coleman et al. 2008). Here, virulence is targeted and thus, open reading frames have to be excluded from optimization. Once the assembly is finished, a further approach on transfecting the GPTEC-T cells with a full length construct is promising to result in replication competent viral particles. The prosperous transfection of GPTEC-T mammalian cells with the created blocks will deliver the chance of a complete transcriptome analysis and an official annotation of the guinea pig adenovirus.

5 ACKNOWLEDGEMENT

Das ist's ja, was den Menschen zieret,
und dazu ward ihm der Verstand,
daß er im innern Herzen spüret,
was er erschafft mit seiner Hand.

(Johann Christoph Friedrich von Schiller, Das Lied von der Glocke, 1799)

Ich möchte mich herzlich bei Herrn Prof. Dr. Thomas Dobner und Dr. Axel Trefzer für die Möglichkeit zur Durchführung dieser Arbeit in Form einer Kooperation von Thermo Fisher Scientific mit dem Heinrich-Pette-Institut bedanken, sowie bei Herrn Dr. Michael Liss, der mich in vergangenen drei Jahren im Team für Forschung und Entwicklung bei GeneArt betreut und stets gefördert hat. Die Einbettung sowohl in ein wissenschaftliches, als auch industrielles Arbeitsumfeld haben mich fortwährend bereichert und mir eine große fachliche und persönliche Weiterentwicklung beschert.

Für die Erstellung des Zweitgutachtens danke ich herzlich Herrn Prof. Dr. Ralf Wagner vom Lehrstuhl für medizinische Mikrobiologie der Universität Regensburg.

Ebenso herzlich möchte ich Frau Dr. Helga Hofmann-Sieber danken, die mich in den Aufenthalten in Hamburg stets betreut hat. An dieser Stelle ein Dankschön von Herzen an alle Kollegen der Abteilung 1 für die wunderbare Zeit in Hamburg.

Besonderer Dank gilt meinen Kollegen bei GeneArt, ohne deren herzliche, mentale, freundschaftliche und fachliche Unterstützung diese Arbeit für mich nicht möglich gewesen wäre und die meine Promotionszeit zu etwas ganz Besonderem gemacht haben. Besonders erwähnen möchte ich dabei Frau Dr. Claudia Chiocchini, Herrn Dr. Nikolai Netuschil, Frau Tatjana Smarschevski, Frau Dr. Ruth Kapzan, Stefanie Jost und viele weitere, die mir vielmehr liebevolle Wegbegleiter als nur Kollegen waren.

Der größte Dank gebührt meinen Freunden und meiner Familie, die mich liebend, immer unterstützt haben und unerschütterlich an mich geglaubt haben. Insbesondere verdanke ich alles was ich bin meinen Eltern, die mich mit größter Liebe durch alle Höhen und Tiefen begleitet haben. Deshalb widme ich diese Arbeit meinen Eltern Jeanette und Karlheinz, meiner Schwester Carolin, Florian und auch meinem kleinen Neffen Vito, bei dem ich kaum erwarten kann ihm alles zu erzählen.

6 PUBLICATION BIBLIOGRAPHY

Alberts, Bruce; Börsch-Supan, Martina (2001): Lehrbuch der molekularen Zellbiologie. Der "kleine" Alberts. 2., korrigierte Aufl. Weinheim: Wiley-VCH.

Annaluru, Narayana; Muller, Héloïse; Mitchell, Leslie A.; Ramalingam, Sivaprakash; Stracquadanio, Giovanni; Richardson, Sarah M. et al. (2014): Total synthesis of a functional designer eukaryotic chromosome. In *Science (New York, N.Y.)* 344 (6179), pp. 55–58. DOI: 10.1126/science.1249252.

Arno G. Motulsky (1983): Impact of genetic manipulation on society and medicine. In *Science* Vol. 219, pp. 135–140.

Baek, Chang-Ho; Chesnut, Jonathan; Katzen, Federico (2015): Positive selection improves the efficiency of DNA assembly. In *Analytical Biochemistry* 476, pp. 1–4. DOI: 10.1016/j.ab.2015.01.021.

Barlow, Jacqueline H.; Rothstein, Rodney (2010): Timing is everything: cell cycle control of Rad52. In *Cell division* 5, p. 7. DOI: 10.1186/1747-1028-5-7.

Benjamini, Yuval; Speed, Terence P. (2012): Summarizing and correcting the GC content bias in high-throughput sequencing. In *Nucleic acids research* 40 (10), e72. DOI: 10.1093/nar/gks001.

Beres, Diane L.; Hawkins, Douglas M. (2001): Plackett–Burman technique for sensitivity analysis of many-parametered models. In *Ecological Modelling* 141 (1-3), pp. 171–183. DOI: 10.1016/S0304-3800(01)00271-X.

Blackford, Andrew N.; Grand, Roger J. A. (2009): Adenovirus E1B 55-kilodalton protein: multiple roles in viral infection and cell transformation. In *Journal of Virology* 83 (9), pp. 4000–4012. DOI: 10.1128/JVI.02417-08.

Blount, Benjamin A.; Driessen, Maureen R. M.; Ellis, Tom (2016): GC Preps: Fast and Easy Extraction of Stable Yeast Genomic DNA. In *Scientific reports* 6, p. 26863. DOI: 10.1038/srep26863.

Boeke, Jef D.; Church, George; Hessel, Andrew; Kelley, Nancy J.; Arkin, Adam; Cai, Yizhi et al. (2016): GENOME ENGINEERING. The Genome Project-Write. In *Science (New York, N.Y.)* 353 (6295), pp. 126–127. DOI: 10.1126/science.aaf6850.

Butz, Nicole; Ossent, P.; Felix R. Homberger (1999): Pathogenesis of Guinea Pig Adenovirus Infection. In *Lab animal science* (49(6)).

Cameron, D. Ewen; Bashor, Caleb J.; Collins, James J. (2014): A brief history of synthetic biology. In *Nat Rev Micro* 12 (5), pp. 381–390. DOI: 10.1038/nrmicro3239.

Cerletti, M.; Negri, T.; Cozzi, F.; Colpo, R.; Andreetta, F.; Croci, D. et al. (2003): Dystrophic phenotype of canine X-linked muscular dystrophy is mitigated by adenovirus-mediated utrophin gene transfer. In *Gene therapy* 10 (9), pp. 750–757. DOI: 10.1038/sj.gt.3301941.

Charles River Research Animal Diagnostic Services: Guinea Pig Adenovirus | Charles River Research Animal Diagnostic Services. Available online at https://www.criver.com/sites/default/files/resources/GuineaPigAdenovirusTechnicalSheet.pd, checked on 10/10/2018.

Chien, Tiffany; Doshi, Anjali; Danino, Tal (2017): Advances in bacterial cancer therapies using synthetic biology. In *Current opinion in systems biology* 5, pp. 1–8. DOI: 10.1016/j.coisb.2017.05.009.

Christen, Matthias; Del Medico, Luca; Christen, Heinz; Christen, Beat (2017): Genome Partitioner: A web tool for multi-level partitioning of large-scale DNA constructs for synthetic biology applications. In *PLoS ONE* 12 (5), e0177234. DOI: 10.1371/journal.pone.0177234.

Coleman, J. Robert; Papamichail, Dimitris; Skiena, Steven; Futcher, Bruce; Wimmer, Eckard; Mueller, Steffen (2008): Virus attenuation by genome-scale changes in codon pair bias. In *Science (New York, N.Y.)* 320 (5884), pp. 1784–1787. DOI: 10.1126/science.1155761.

D'Ambrosio E; Del Grosso N; Chicca A; Midulla M. (1982): Neutralizing antibodies against 33 human adenoviruses in normal children in Rome. In *The Journal of Hygiene* (89(1), 155-161. Available online at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2134173/pdf/jhyg00025-0157.pdf, checked on 10/21/2018.

Doerfler, W. (Ed.) (1996): Medical Microbiology. Adenoviruses. Chapter 67. With assistance of editor Baron S. Available online at https://www.ncbi.nlm.nih.gov/books/NBK8503/.

Echavarría, Marcela (2008): Adenoviruses in immunocompromised hosts. In *Clinical microbiology reviews* 21 (4), pp. 704–715. DOI: 10.1128/CMR.00052-07.

Ekpenyong, Maurice George; Antai, Sylvester Peter; Asitok, Atim David; Ekpo, Bassey Offiong (2017): Plackett-Burman Design and Response Surface Optimization of Medium Trace Nutrients for Glycolipopeptide Biosurfactant Production. In *IBJ* 21 (4), pp. 249–260. DOI: 10.18869/acadpub.ibj.21.4.249.

Engler, Carola; Kandzia, Romy; Marillonnet, Sylvestre (2008): A one pot, one step, precision cloning method with high throughput capability. In *PloS one* 3 (11), e3647. DOI: 10.1371/journal.pone.0003647.

Erlich, Yaniv; Zielinski, Dina (2017): DNA Fountain enables a robust and efficient storage architecture. In *Science (New York, N.Y.)* 355 (6328), pp. 950–954. DOI: 10.1126/science.aaj2038.

Frackman, Susan; Kobs, Gary; Simpson, Dan; Storts, Doug (1998): Betaine and DMSO: Enhancing Agents for PCR. In *Corporation, Promega* (Number 65), p. 27.

Frankfurter Allgemeine Zeitung GmbH (2012): Reflexionen zur Synthetischen Biologie: Dann lasst uns einen neuen Baum des Lebens pflanzen. With assistance of Manuela Lenzen. Edited by Frankfurter Allgemeine Zeitung GmbH. Available online at http://www.faz.net/aktuell/feuilleton/buecher/rezensionen/sachbuch/reflexionen-zur-synthetischen-biologie-dann-lasst-uns-einen-neuen-baum-des-lebens-pflanzen-11692664.html, checked on 10/8/2018.

Garnett, C. T.; Erdman, D.; Xu, W.; Gooding, L. R. (2002): Prevalence and Quantitation of Species C Adenovirus DNA in Human Mucosal Lymphocytes. In *Journal of Virology* 76 (21), pp. 10608–10616. DOI: 10.1128/JVI.76.21.10608-10616.2002.

Ghebremedhin, Beniam (2014): Human adenovirus: Viral pathogen with increasing importance. In *European Journal of Microbiology and Immunology* 4 (1), pp. 26–33. DOI: 10.1556/EuJMI.4.2014.1.2.

Gibson, D. G.; Benders, G. A.; Axelrod, K. C.; Zaveri, J.; Algire, M. A.; Moodie, M. et al. (2008): One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. In *Proceedings of the National Academy of Sciences* 105 (51), pp. 20404–20409. DOI: 10.1073/pnas.0811011106.

Gibson, Daniel G. (2009): Synthesis of DNA fragments in yeast by one-step assembly of overlapping oligonucleotides. In *Nucleic acids research* 37 (20), pp. 6984–6990. DOI: 10.1093/nar/gkp687.

Gibson, Daniel G.; Glass, John I.; Lartigue, Carole; Noskov, Vladimir N.; Chuang, Ray-Yuan; Algire, Mikkel A. et al. (2010): Creation of a bacterial cell controlled by a chemically synthesized genome. In *Science (New York, N.Y.)* 329 (5987), pp. 52–56. DOI: 10.1126/science.1190719.

Gibson, Daniel G.; Young, Lei; Chuang, Ray-Yuan; Venter, J. Craig; Hutchison, Clyde A.; Smith, Hamilton O. (2009): Enzymatic assembly of DNA molecules up to several hundred kilobases. In *Nat Methods* 6 (5), pp. 343–345. DOI: 10.1038/nmeth.1318.

Gietz, R. Daniel; Schiestl, Robert H. (2007): High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. In *Nature protocols* 2 (1), pp. 31–34. DOI: 10.1038/nprot.2007.13.

Huang, Z.; Fasco, M. J.; Kaminsky, L. S. (1996): Optimization of Dnase I removal of contaminating DNA from RNA for use in quantitative RNA-PCR. In *BioTechniques* 20 (6), 1012-4, 1016, 1018-20. DOI: 10.2144/96206st02.

Hutchison, Clyde A. (2007): DNA sequencing: bench to bedside and beyond. In *Nucleic acids research* 35 (18), pp. 6227–6237. DOI: 10.1093/nar/gkm688.

Illuxley, Clare; Green, Eric D.; Dunbam, Ian (1990): Rapid assessment of S. cerevisiae mating type by PCR. In *Trends in Genetics* 6, p. 236. DOI: 10.1016/0168-9525(90)90190-H.

Jin, Peng; Ding, Wenwen; Du, Guocheng; Chen, Jian; Kang, Zhen (2016): DATEL: A Scarless and Sequence-Independent DNA Assembly Method Using Thermostable Exonucleases and Ligase. In *ACS Synth. Biol.* 5 (9), pp. 1028–1032. DOI: 10.1021/acssynbio.6b00078.

Jinek, Martin; Chylinski, Krzysztof; Fonfara, Ines; Hauer, Michael; Doudna, Jennifer A.; Charpentier, Emmanuelle (2012): A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. In *Science (New York, N.Y.)* 337 (6096), pp. 816–821. DOI: 10.1126/science.1225829.

Kaján, Gyozo L.; Davison, Andrew J.; Palya, Vilmos; Harrach, Balázs; Benko, Mária (2012): Genome sequence of a waterfowl aviadenovirus, goose adenovirus 4. In *The Journal of general virology* 93 (Pt 11), pp. 2457–2465. DOI: 10.1099/vir.0.042028-0.

Karim, Ashty S.; Curran, Kathleen A.; Alper, Hal S. (2013): Characterization of plasmid burden and copy number in Saccharomyces cerevisiae for optimization of metabolic engineering applications. In *FEMS yeast research* 13 (1), pp. 107–116. DOI: 10.1111/1567-1364.12016.

Kawai, Shigeyuki; Hashimoto, Wataru; Murata, Kousaku (2010): Transformation of Saccharomyces cerevisiae and other fungi: methods and possible underlying mechanism. In *Bioengineered bugs* 1 (6), pp. 395–403. DOI: 10.4161/bbug.1.6.13257.

Kimelman, Aya; Levy, Asaf; Sberro, Hila; Kidron, Shahar; Leavitt, Azita; Amitai, Gil et al. (2012): A vast collection of microbial genes that are toxic to bacteria. In *Genome Research* 22 (4), pp. 802–809. DOI: 10.1101/gr.133850.111.

Kojaoghlanian, Tsoline; Flomenberg, Phyllis; Horwitz, Marshall S. (2003): The impact of adenovirus infection on the immunocompromised host. In *Reviews in medical virology* 13 (3), pp. 155–171. DOI: 10.1002/rmv.386.

Kok, Stefan de; Stanton, Leslie H.; Slaby, Todd; Durot, Maxime; Holmes, Victor F.; Patel, Kedar G. et al. (2014): Rapid and reliable DNA assembly via ligase cycling reaction. In *ACS synthetic biology* 3 (2), pp. 97–106. DOI: 10.1021/sb4001992.

Lenaerts, L.; Naesens, L. (2006): Antiviral therapy for adenovirus infections. In *Antiviral research* 71 (2-3), pp. 172–180. DOI: 10.1016/j.antiviral.2006.04.007.

Löfken, Jan Oliver (2017): Datenspeicher: Festplatten aus DNA speichern mehr als jeder Chip. Edited by www.zeit.de. Available online at https://www.zeit.de/wissen/2017-03/dna-datenspeicher-erbgut, updated on 3/3/2017, checked on 10/9/2018.

Madigan, Michael; Martinko, John; Stahl, David; Clark, David (Eds.) (2013): Brock Mikrobiologie. With assistance of Michael Madigan, John Martinko, David Stahl, David Clark. 13th ed.: Pearson.

Mamedov, T. G.; Pienaar, E.; Whitney, S. E.; TerMaat, J. R.; Carvill, G.; Goliath, R. et al. (2008): A fundamental study of the PCR amplification of GC-rich DNA templates. In *Computational Biology and Chemistry* 32 (6), pp. 452–457. DOI: 10.1016/j.compbiolchem.2008.07.021.

Marko Kapitza (2011): Design of Experiments. Seminar "Robust Design". In : SpringerReference, Universität zu Köln. Berlin/Heidelberg: Springer-Verlag.

Matthias Christen, Luca Del Medico, Heinz Christen, Beat Christen (2017): https://christenlab.ethz.ch/GenomePartitioner. A web tool for multi-level partitioning of large-scale DNA constructs for synthetic biology applications. https://christenlab.ethz.ch/GenomePartitioner. Edited by christenlab.ethz.ch. Eidgenößische Technische Hochschule (ETH) Zürich. Available online at https://christenlab.ethz.ch/GenomePartitioner.

Melief, Cornelis J. M. (2017): Cancer: Precision T-cell therapy targets tumours. In *Nature* 547 (7662), pp. 165–167. DOI: 10.1038/nature23093.

Mierzejewska, Karolina; Siwek, Wojciech; Czapinska, Honorata; Kaus-Drobek, Magdalena; Radlinska, Monika; Skowronek, Krzysztof et al. (2014): Structural basis of the methylation specificity of R.DpnI. In *Nucleic acids research* 42 (13), pp. 8745–8754. DOI: 10.1093/nar/gku546.

Milavetz, Barry I.; Balakrishnan, Lata (2015): Viral epigenetics. In *Methods in molecular biology (Clifton, N.J.)* 1238, pp. 569–596. DOI: 10.1007/978-1-4939-1804-1_30.

Mullis, K.; Faloona, F.; Scharf, S.; Saiki, R.; Horn, G.; Erlich, H. (1986): Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. In *Cold Spring Harbor Symposia on Quantitative Biology* 51 (0), pp. 263–273. DOI: 10.1101/SQB.1986.051.01.032.

Naumann, Susanne; Kunstyr, I.; Langer, Ina; Maess, J.; Hörning, R. (1981): Lethal pneumonia in guineapigs associated with a virus. In *Laboratory Animals* 15; 235-242.

Nogales, Aitor; Baker, Steven F.; Ortiz-Riaño, Emilio; Dewhurst, Stephen; Topham, David J.; Martínez-Sobrido, Luis (2014): Influenza A virus attenuation by codon deoptimization of the NS gene for vaccine development. In *Journal of Virology* 88 (18), pp. 10525–10540. DOI: 10.1128/JVI.01565-14.

Ott, Patrick A.; Hu, Zhuting; Keskin, Derin B.; Shukla, Sachet A.; Sun, Jing; Bozym, David J. et al. (2017): An immunogenic personal neoantigen vaccine for patients with melanoma. In *Nature* 547, 217 EP -. DOI: 10.1038/nature22991.

R. L. Plackett; J. P. Burman (1946): DESIGN OF OPTIMUM MULTIFACTORIAL EXPERIMENTS. Edited by Biometrika. Oxford Academic (Issue 4). Available online at https://academic.oup.com/biomet/article-abstract/33/4/305/225377, checked on 9/18/2018.

Russell, W. C. (2009): Adenoviruses: update on structure and function. In *The Journal of general virology* 90 (Pt 1), pp. 1–20. DOI: 10.1099/vir.0.003087-0.

Rux, John J.; Burnett, Roger M. (2004): Adenovirus structure. In *Human gene therapy* 15 (12), pp. 1167–1176. DOI: 10.1089/hum.2004.15.1167.

Saida, F.; Uzan, M.; Odaert, B.; Bontems, F. (2006): Expression of Highly Toxic Genes in E. coli: Special Strategies and Genetic Tools. In *Current Protein and Peptide Science* 7 (1), pp. 47–56. DOI: 10.2174/138920306775474095.

Science News Staff (2016): And Science's 2015 Breakthrough of the Year is… With assistance of Robert Coontz. Edited by Science. Available online at https://www.sciencemag.org/news/2015/12/and-science-s-2015-breakthrough-year, updated on 2/24/2016, checked on 10/8/2018.

Sherman, Fred (2002): Getting started with yeast. In : Guide to Yeast Genetics and Molecular and Cell Biology - Part B, vol. 350: Elsevier (Methods in Enzymology), pp. 3–41.

Shevelev, G. Y.; Pyshnyi, D. V. (2018): Modern approaches to artificial gene synthesis: aspects of oligonucleotide synthesis, enzymatic assembly, sequence verification and error correction. In *Vestn. VOGiS* 22 (5), pp. 498–506. DOI: 10.18699/VJ18.387.

Snyder, Thomas M.; Tse, Brian N.; Liu, David R. (2008): Effects of template sequence and secondary structure on DNA-templated reactivity. In *Journal of the American Chemical Society* 130 (4), pp. 1392–1401. DOI: 10.1021/ja076780u.

Synthetic Biology Market by Tool & Technology - Global Forecast 2022 | MarketsandMarkets. Available online at https://www.marketsandmarkets.com/Market-Reports/synthetic-biology-market-889.html?gclid=EAIaIQobChMIp_vv5Ib33QIVGOh3Ch0o9gidEAAYAiAAEgJKzPD_BwE, checked on 10/8/2018.

Täuber, Birgitt; Dobner, Thomas (2001): Adenovirus early E4 genes in viral oncogenesis. In *Oncogene* 20 (54), pp. 7847–7854. DOI: 10.1038/sj.onc.1204914.

Volpers, Christoph; Kochanek, Stefan (2004): Adenoviral vectors for gene transfer and therapy. In *The journal of gene medicine* 6 Suppl 1, S164-71. DOI: 10.1002/jgm.496.

Walhout, A. J.; Vidal, M. (2001): High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. In *Methods (San Diego, Calif.)* 24 (3), pp. 297–306. DOI: 10.1006/meth.2001.1190.

Weissensteiner, T.; Lanchbury, J. S. (1996): Strategy for controlling preferential amplification and avoiding false negatives in PCR typing. In *BioTechniques* 21 (6), pp. 1102–1108. DOI: 10.2144/96216rr03.

Wild, Jadwiga; Szybalski, Waclaw (2004): Copy-control pBAC/oriV vectors for genomic cloning. In *Methods in molecular biology (Clifton, N.J.)* 267, pp. 145–154. DOI: 10.1385/1-59259-774-2:145.

Wilson, J. M. (1996): Adenoviruses as gene-delivery vehicles. In *The New England journal of medicine* 334 (18), pp. 1185–1187. DOI: 10.1056/NEJM199605023341809.

Wood, Whitney N.; Smith, Kyle D.; Ream, Jennifer A.; Lewis, L. Kevin (2017): Enhancing yields of low and single copy number plasmid DNAs from Escherichia coli cells. In *Journal of Microbiological Methods* 133, pp. 46–51. DOI: 10.1016/j.mimet.2016.12.016.

Wu, Shipo; Kroeker, Andrea; Wong, Gary; He, Shihua; Hou, Lihua; Audet, Jonathan et al. (2016): An Adenovirus Vaccine Expressing Ebola Virus Variant Makona Glycoprotein Is Efficacious in Guinea Pigs and Nonhuman Primates. In *The Journal of infectious diseases* 214 (suppl 3), S326-S332. DOI: 10.1093/infdis/jiw250.

Ye, X. (2000): Engineering the Provitamin A (-Carotene) Biosynthetic Pathway into (Carotenoid-Free) Rice Endosperm. In *Science* 287 (5451), pp. 303–305. DOI: 10.1126/science.287.5451.303.

Zhou, Jian; Sun, Xiao Yi; Stenzel, Deborah J.; Frazer, Ian H. (1991): Expression of vaccinia recombinant HPV 16 L1 and L2 ORF proteins in epithelial cells is sufficient for assembly of HPV virion-like particles. In *Virology* 185 (1), pp. 251–257. DOI: 10.1016/0042-6822(91)90772-4.

7 APPENDIX

7.1 Oligonucleotides

Primers used for fast gene synthesis project:

Primer	Sequence	T_M in °C
M13 fwd	TTGTAAAACGACGGCCAG	68
M13 rev	CATGGTCATAGCTGTTTCC	68
In-house primer for specific oligonucleotides		

Table 56: Primers used for fast gene synthesis project

Primers for amplification of subblocks:

Primer	Sequence	T_M in °C
SB5.fwd	CGGGCGCTTGATTCTGCCCG	68
SB5.rev	TTTCCCGACTGGAAAGCGGG	64
SB14.fwd	CACGTGAATAGTGTAGGAAC	58
SB14.rev	CCCACAAATGCATCAACCCG	62
SB15.fwd	CGGCATCGTCTCTCGAGACG	66
SB15.rev	CGTAACGCCGAATCTGCCGC	66
SB23.fwd	GCGGAGCAGTTCGTTACCG	62
SB23.rev	TTTCCCGACTGGAAAGCGGG	64
SB45.fwd	CACGCCTCTTTGGAACAGC	60
SB45.rev	GCACCGGCGGGCCATGGATC	70
SB46.fwd	CGCGTCGCTGGCTCATGCTG	68
SB46.rev	CATGCGTCATCGCGTGCGTG	66
SB47.fwd	AGCGATGGGCGCCCCAGTTG	68
SB47.rev	TTTCCCGACTGGAAAGCGGG	64

Table 57: Primers for amplification of subblocks

Control primers for block assembly:

Primer	Sequence	T_M in °C
CP_SB0.fwd	GGTCGGAGCCCGTTTCTGTT	55.88
CP_SB1.fwd	TTTATTGTGGGCAGCGCCGC	55.88
CP_SB2.fwd	GTCAGCGTGGTGTGCATGGT	55.88

CP_SB3.fwd	CGAGTGGGCATCCGGTCTTT	55.88
CP_SB4.fwd	GTCAATATGGCAACCGGCGC	55.88
CP_SB6.fwd	GCATCTCTTCTTCGTCGGCG	55.88
CP_SB7.fwd	GCCAGCACGAAAGACGCGAT	55.88
CP_SB8.fwd	CGATCACGAATATGGCCGGC	55.88
CP_SB9.fwd	CGACTTGCTGTAGCGAGGGT	55.88
CP_SB10.fwd	GCCGGCTGCACGTTCTTGAT	55.88
CP_SB12.fwd	GCGGTCGCAACGTGAAAGCT	55.88
CP_SB13.fwd	TCCGTAAAGCGGCCAGAAGG	55.88
CP_SB14.fwd	CATCGCGTTGGATCCGGAGT	55.88
CP_SB15.fwd	GCGTCACCCGTACACTGTGT	55.88
CP_SB16.fwd	CTGCTCTTACTTGTGGCGCC	55.88
CP_SB18.fwd	ACGGTAGTGGAACTGTCGCC	55.88
CP_SB19.fwd	TGATCCGCCGCGTCCTCTAA	55.88
CP_SB20.fwd	TGGCCATCCTTATCTCGCCC	55.88
CP_SB21.fwd	CATCCCGTTGTTACGCACGC	55.88
CP_SB22.fwd	ACCGGCGTCACTTACCGAGT	55.88
CP_SB24.fwd	GCGGCAGTAGAAAGCGGTAC	55.88
CP_SB25.fwd	CACCCACGTGAACAACCAGC	55.88
CP_SB26.fwd	GGTTAACAACCAGGGGCCGT	55.88
CP_SB27.fwd	CCTGGACGGCACTTTCTACC	55.88
CP_SB28.fwd	CCCTAGGCGGTTTTTCCCGT	55.88
CP_SB30.fwd	TGTCGTGACGCAACTGCACC	55.88
CP_SB31.fwd	CGGCCTGCGGAAAGGTAACA	55.88
CP_SB32.fwd	AACACCTGACGCGACAAGCG	55.88
CP_SB33.fwd	TCGAAGGTGTCCGGCGTCAA	55.88
CP_SB34.fwd	TACGAGGACCAATCGGAGCC	55.88
CP_SB36.fwd	CTTCCAGAGGTACTGCTCCC	55.88
CP_SB37.fwd	AACAGATACGCGAACGGCGC	55.88
CP_SB38.fwd	CCCCCAAGTTTAATCGGCCC	55.88
CP_SB39.fwd	ACCCTCCGAAGAAGGCCTGT	55.88
CP_SB40.fwd	GGAGTGTCCACCATCTCGCA	55.88
CP_SB42.fwd	GCGGGACTCACCCTGAAACT	55.88

CP_SB43.fwd	GCAAATAATCGCGGGACCGG	55.88
CP_SB44.fwd	ACGCTGTGCGTCTCGCCAAA	55.88
CP_SB45.fwd	ACAGGTTGGGAGACCGACAC	55.88
CP_SB46.fwd	CCGTGTCCGTAGAAGACCCT	55.88
CP_SB0.rev	TAACCCCCATCGTCGTCGGT	55.88
CP_SB1.rev	CCCCGCGGTGGAAAAATCCA	55.88
CP_SB2.rev	CTGCTGATGCAGACCCACCA	55.88
CP_SB3.rev	GTGGGGCTCATCACCGTGTT	55.88
CP_SB4.rev	GGATACAAAAGCCCGACCCC	55.88
CP_SB6.rev	CAATGGAACGATCCGCCACC	55.88
CP_SB7.rev	CGAGATACCCCCGACGATGT	55.88
CP_SB8.rev	TGGGAGAATACCCCCATCCC	55.88
CP_SB9.rev	AAGACCCTGGCCCACAAGCT	55.88
CP_SB10.rev	ACTGCTCAGACCGTTCTCCG	55.88
CP_SB12.rev	CCCGAAATAGCGTTCCCGCT	55.88
CP_SB13.rev	CGTCGTCGTCCCGTTAACCA	55.88
CP_SB14.rev	GTTGCCCACGTAAGCCTCCA	55.88
CP_SB15.rev	TTGGAATCCCCGCTGGCCTT	55.88
CP_SB16.rev	CGCGCCGAGTCAGCAAAAAG	55.88
CP_SB18.rev	GCTGGCCTCGTAGGGACTAA	55.88
CP_SB19.rev	TGTCCCGGAACGTGATAGGC	55.88
CP_SB20.rev	CGCCGATAATCTCCGTCGTC	55.88
CP_SB21.rev	GTTTGCACGCCTATGCCCTG	55.88
CP_SB22.rev	TCCCGGAATGGCTCCTATGG	55.88
CP_SB24.rev	AAGAGGCAAGCACACCGGTG	55.88
CP_SB25.rev	GGGCATGACGTCGCCTTTCA	55.88
CP_SB26.rev	TGTCGATGTAGCCGTCCACG	55.88
CP_SB27.rev	GGAGGGATCCGGTATCTGTC	55.88
CP_SB28.rev	AAGGTGTGCGAGACGGGTTC	55.88
CP_SB30.rev	GTACGACGAGGGAGATCACG	55.88
CP_SB31.rev	ACCTTCAAGAGGGTCCAGCG	55.88
CP_SB32.rev	CGTGGGACGCGGTGATTTAG	55.88
CP_SB33.rev	CGCGCAGATTTTCGGGTTCC	55.88

CP_SB34.rev	GACCCTTTTTTTGCGCGCCG	55.88
CP_SB36.rev	TGCGCGGGTGGAGTAATCCT	55.88
CP_SB37.rev	CCGACAGCGGCGGTTTTACA	55.88
CP_SB38.rev	GCAGCGCGTTCATGAGGTAC	55.88
CP_SB39.rev	CGCCGCCAAAGTGGGGTTAA	55.88
CP_SB40.rev	CAGAGAGATGGCGTTGACGC	55.88
CP_SB42.rev	AGAACACGGTCATGCCCCGT	55.88
CP_SB43.rev	GCGAGCCTGCGTGAGTAACT	55.88
CP_SB44.rev	TCTAGGGAGTGCGCTTTGCG	55.88
CP_SB45.rev	AGACGCCTGCGGTTGCATCT	55.88
CP_SB46.rev	TGTGTGGTCAGCGGCTTGGT	55.88

Table 58: Control primers for block assembly

Sequencing primer for block 0 and 1:

Primer	Sequence	T_M in °C
B0_seq1.fwd	TAATACGACTCACTATAGGGCG	54,55
B0_seq1.rev	TGTAACACAGGCTGCACAACACC	47,83
B0_seq2.fwd	TGTTGTGCAGCCTGTGTTACAT	54,55
B0_seq2.rev	CGCGGTCACCGTCGACGAC	26,32
B0_seq3.fwd	CCGTCGTCGACGGTGACC	27,78
B0_seq3.rev	TCGGTGGAGCACTTGCACAGGAT	43,48
B0_seq4.fwd	CCTGTGCAAGTGCTCCACCGA	38,10
B0_seq4.rev	TGCCCGGAACGCCGTAGC	27,78
B0_seq5.fwd	AGCGCTACGGCGTTCCG	29,41
B0_seq5.rev	TTTCCCGACTGGAAAGCGG	42,11
B1_seq1.fwd	TAATACGACTCACTATAGGGCG	54,55
B1_seq1.rev	ACGCCCACGGCGCCCCGC	11,11
B1_seq2.fwd	ATTGCGGGGCGCCGTGGGC	21,05
B1_seq2.rev	CGCAGCATCGCCAAGCTG	33,33
B1_seq3.fwd	CAGCAGCTTGGCGATGCTGC	35,00
B1_seq3.rev	CACCAGTCTCAAAAAGGCCG	45,00
B1_seq4.fwd	GCGGCCTTTTTGAGACTGGTGT	45,45
B1_seq4.rev	TCGTCGACTGCGACAACCAC	40,00
B1_seq5.fwd	CGTGGTTGTCGCAGTCGACGAC	36,36

B1_seq5.rev	TTTCCCGACTGGAAAGCGG	42,11

Table 59: Sequencing primer for block 0 and 1

Amplification primer block 7:

Primer	Sequence	T_M in °C
B7.fwd	CCCCGCTCGCGGTCGACGAA	70
B7.rev	CCACCTTTCGCTCCTTGAAC	62

Table 60: Amplification primer block 7

Primers for linearization of pYES8D by PCR:

Primer	Sequence	T_M in °C
CH353_2µ.fwd	GAAATTTGCTATTTTGTTAGAGTCTTTTACACCATTTGTC	67
CH327_Trp.rev	AAAAAATGTAGAGGTCGAGTTTAGATGCAAGTTCAAG	69

Control primer assembly segments:

Primer	Sequence	T_M in °C
CP_Bl0.fwd	GGTACGGGTAGCGCTTTGGT	55.88
CP_Bl1.fwd	TTGGCCGCGACCTGTCTCAA	55.88
CP_Bl2.fwd	GAGTGGACCCGGCCAATGTA	55.88
CP_Bl4.fwd	TTCTGCGTTAGACGGCGACC	55.88
CP_Bl5.fwd	GACTCCCACCTGTCGGAAGT	55.88
CP_Bl6.fwd	AAGCGACGGACACCTCACGT	55.88
CP_Bl0.rev	ACCCGGAAAACGTCTTCGCC	55.88
CP_Bl1.rev	GGGACCGTCTGTTTTCCCGT	55.88
CP_Bl2.rev	TCCTCGTTGTCGCCCGTCAT	55.88
CP_Bl4.rev	ATGGAGCGAGGTGATGGGTC	55.88
CP_Bl5.rev	CGCGTTGAATGGAAGCTCCG	55.88
CP_Bl6.rev	AAAAGCGGATCGGCGACCTC	55.88

Table 61: Control primer assembly segments

7.2 Logfile Genome Partitioner tool

analysing file: webserver input by user feature type: source
.....checking gaps or overlaps
feature type source not foundsegment list (size, start, end):
18585 [0, 18585]

18585 [18485, 37070]
segment list (size, start, end)
0 18585 0 18585
1 18585 18485 37070
---
segment 0 block_overlap 2
GCTACCCGTACGTCCTCGCCCCCCGGAGCAACCGCCTTCGACCCGTTTTCCCTGTTGGGCGCTCGCCCCAAGA
TGACCAACGTAGAAATGGGTATGCGTC
direct repeat CTCGCCCC and CTCGCCCC from 14 to 68 size 8 gap 39
shift 41 new length 4915 4765 segment 0 block_overlap 2 [9237, 13995] to [9237, 14036] and block_overlap 3 [13895, 18585] to [13936, 18585]
---
block partitioning of segment:0[0:18585]=18585|18732
block: 0[0:4679]=4679|4874
block: 1[4579:9337]=4758|4874
block: 2[9237:14036]=4799|4915
block: 3[13936:18585]=4649|4833
---
segment 1 block_overlap 1
ATCTACGAAGAGGAGGAACGGGATTCTCACCACCACCCAGAACGCGACCCCCCCTACGACTACGAAGAGAGG
GAACCCGGCCCTCTGGTCATCGACGAGG
direct repeat CTACGAAG and CTACGAAG from 2 to 66 size 8 gap 49
shift 34 new length 4908 4840 segment 1 block_overlap 1 [4579, 9337] to [4579, 9371] and block_overlap 2 [9237, 13995] to [9271, 13995]
---
block partitioning of segment:1[18485:37070]=18585|18732
block: 4[18485:23164]=4679|4874
block: 5[23064:27856]=4792|4908
block: 6[27756:32480]=4724|4840
block: 7[32380:37070]=4690|4874
---
blocks created: 8, min,mean,max,std: 4833, 4874, 4915, 18.17376445
original length 37070, partitioned length 38992
block overlap optimisation
hairpins: removed 0, remaining 0
direct repeats: removed 2, remaining 0
remaining subsequences > 8: 0
==========subblock
section
block list (size, start, end)
0 4679 0 4679
1 4758 4579 9337
2 4799 9237 14036
3 4649 13936 18585
4 4679 18485 23164
5 4792 23064 27856
6 4724 27756 32480
7 4690 32380 37070

```
---------------------------------------------------------------
subblock partitioning of block:0[0:4679]=4679|4874
subblock: 0[0:717]=717|870
subblock: 1[667:1521]=854|870
subblock: 2[1471:2325]=854|870
subblock: 3[2275:3129]=854|870
subblock: 4[3079:3933]=854|870
subblock: 5[3883:4679]=796|870
---------------------------------------------------------------
subblock partitioning of block:1[4579:9337]=4758|4874
subblock: 6[4579:5375]=796|870
subblock: 7[5325:6179]=854|870
subblock: 8[6129:6983]=854|870
subblock: 9[6933:7787]=854|870
subblock: 10[7737:8591]=854|870
subblock: 11[8541:9337]=796|870
---------------------------------------------------------------
subblock partitioning of block:2[9237:14036]=4799|4915
subblock: 12[9237:10039]=802|876
subblock: 13[9989:10849]=860|876
subblock: 14[10799:11659]=860|876
subblock: 15[11609:12469]=860|876
subblock: 16[12419:13279]=860|876
subblock: 17[13229:14036]=807|881
---------------------------------------------------------------
subblock partitioning of block:3[13936:18585]=4649|4833
subblock: 18[13936:14725]=789|863
subblock: 19[14675:15522]=847|863
subblock: 20[15472:16319]=847|863
subblock: 21[16269:17116]=847|863
subblock: 22[17066:17913]=847|863
subblock: 23[17863:18585]=722|864
---------------------------------------------------------------
subblock partitioning of block:4[18485:23164]=4679|4874
subblock: 24[18485:19202]=717|870
subblock: 25[19152:20006]=854|870
subblock: 26[19956:20810]=854|870
subblock: 27[20760:21614]=854|870
subblock: 28[21564:22418]=854|870
subblock: 29[22368:23164]=796|870
---------------------------------------------------------------
subblock partitioning of block:5[23064:27856]=4792|4908
subblock: 30[23064:23865]=801|875
subblock: 31[23815:24674]=859|875
subblock: 32[24624:25483]=859|875
subblock: 33[25433:26292]=859|875
subblock: 34[26242:27101]=859|875
```

```
subblock: 35[27051:27856]=805|879
---------------------------------------------------------------
block 6 subblock_overlap 0 to 5 number of identical sub-sequences: size= [8, 9] amount= [2, 1]
largest substring GGACCTCGC length 9 in subblock_overlap 2 from 28 to 37
CGGACCGCCGGGATTAAAAGGCGAAGACGGACCTCGCGGCTTCGTAGGAA
   and substring GGACCTCGC length 9 in subblock_overlap 3 from 33 to 42
GCGCGGCGGCATCAACAATCCCGCCCCGGGCAAGGACCTCGCCAGCAACC
---------------------------------------------------------------
option shift seq b1 b2  lb1  lb2 max counts
    1   -13   6  3  4 851 877   7 1
    2   -18   6  2  3 846 882   7 1
    3    37   6  3  4 901 827   7 1
    4    32   6  2  3 896 832   9 1
shift -13 new length 851 877 block 6 subblock_overlap 3 [2336, 3184] to [2336, 3171] and
 subblock_overlap 4 [3134, 3982] to [3121, 3982]
---------------------------------------------------------------
subblock partitioning of block:6[27756:32480]=4724|4840
subblock: 36[27756:28546]=790|864
subblock: 37[28496:29344]=848|864
subblock: 38[29294:30142]=848|864
subblock: 39[30092:30927]=835|851
subblock: 40[30877:31738]=861|877
subblock: 41[31688:32480]=792|866
---------------------------------------------------------------
subblock partitioning of block:7[32380:37070]=4690|4874
subblock: 42[32380:33176]=796|870
subblock: 43[33126:33980]=854|870
subblock: 44[33930:34784]=854|870
subblock: 45[34734:35588]=854|870
subblock: 46[35538:36392]=854|870
subblock: 47[36342:37070]=728|870
-----------------------------------------------------
subblocks created: 48, min,mean,max,std: 851, 870, 881, 0.0
original length 37070, partitioned length 41760
subblock overlap optimisation
hairpins: removed 0, remaining 0
direct repeats: removed 0, remaining 0
size   number of removed subsequences in overlaps
  9       1
remaining subsequences > 8: 0
block-list with subblock-overlap results
feature  nr   hp   dr  str  sub  elm
-------------------------------------------
  block   0    0    0    0
  block   1    0    0    0
  block   2    0    0    0
  block   3    0    0    0
  block   4    0    0    0
```

```
  block   5    0    0    0
  block   6    0    0    0
  block   7    0    0    0
-------------------------------------------
  block tot    0    0    0
-------------------------------------------
segment-list with block-overlap results
feature  nr   hp   dr  str  sub  elm
-------------------------------------------
segment   0    0    0    0
segment   1    0    0    0
-------------------------------------------
segment tot    0    0    0
-------------------------------------------
     total    0    0    0
-------------------------------------------
bad segment numbers:
[]
bad block numbers:
[]
bad subblock numbers:
[]
sequence file: ./upload297B635c7/1vudnio0hsfm65u78vhlpsm7g2/output.gb block_type: subblock
number of primers tested: 12000 mean temperatur: 56.5568416667 search_range 150 primer_size 20
statistic of discarded primer pairs for subblock overlap
-------------------------------------------------------------
 150263 Tm_mismatches
      0 homologs
      0 self homologs
    776 hairpins
    608 direct repeats
     57 last 8 inside
      0 Tm_diff increase first try
      0 Tm_diff increase second try
      0 Tm_diff increase third try
      0 search range increase first try
      0 search range increase second try
      0 search range increase third try
      0 no pair found
-------------------------------------------------------------
sequence file: ./upload297B635c7/1vudnio0hsfm65u78vhlpsm7g2/output.gb block_type: block
number of primers tested: 1800 mean temperatur: 56.5268888889 search_range 150 primer_size 20
statistic of discarded primer pairs for block overlap
-------------------------------------------------------------
 202391 Tm_mismatches
      0 homologs
      0 self homologs
```

783 hairpins
778 direct repeats
57 last 8 inside
0 Tm_diff increase first try
0 Tm_diff increase second try
0 Tm_diff increase third try
1 search range increase first try
0 search range increase second try
0 search range increase third try
0 no pair found

7.3 Guinea Pig Adenovirus sequence

1 CATCATCAAT ATATATACAG TGCATTTTGA CTTCCGGGTT TGACTTCCGG GTGCCATTTT
61 GAGTGTGTCC CAAGATGGCG GAGCGTGGGA GCCATTTTGA GTGACAGCGT GGGAGCGCGG
121 GGGGGGCGGC GCGCTCCCGG AAGCGGCCGT GGGAACGTGC TCGGGGAAAT CGAAACCGAT
181 GTGTGCCAAG TCGGTCGACT GCGCGTGAAA TCGATTAGCC CTGATATTTT GCGCTTTTTC
241 CGGGATGGGC GGCGCTTTTG CCGCGGATTT GCGCCCTGGT GGCGGTTTCG TTTGGCGCGG
301 CGGCGATTTG GGTCGGAGGG TCGTGAGAAA TGGCTCGTAT CGTGAGTACC GGCGTGTGGT
361 GGCCCCCCGT GGACGTCCAC TCCTCTTTTT GTGTGTGCGA CCCGAACGTG TGTCCCACGG
421 CCGCGGCCGC TCGCGCCCGT CGGGGTTCGG AGCGACAGGC GGCCGCGGCG GCCGCGGCGG
481 CGGAGAAGGG GAAACGAAAT TTGTCTGCGG CGCACCGCGG CGAGGGGTCG GAGCCCGTTT
541 CTGTTTTGCG CTGGGAAGCC CCCTCCGGTG AGGTCAGTCT GGGAGACGGA CTGGAGACCG
601 CGGGCACGCA CGTGTGCGGT TGGGACTACG AGACCCCGGG CGCTTGTTCC TGTTTTGAGG
661 GAAGAGAGAT GACGGAGAAG GACCTGTTGT GTTTTGAGAA TCTGTACGGC ATGGAAGACG
721 GCGGGGGTTT GTCGACCGAC GACGATGGGG GTTATCTGCG CGTGGAGGAA GAGGATTTTG
781 AGTTGGATTA TCCCGAGAGC CCCGGCGCGG ACTGTAAGAG TTGCGAGTGG CACCGGCGGC
841 GGGGCATGCA GGTGTTGTGC AGCCTGTGTT ACATGCGCCT GAGCGAGTGG ATAGGGGGTG
901 AGTATGTGCG TGGCGAGCGC GTGTACCGTG GGAAATGGGG TAGTAGGAAG TAGGAAGCGG
961 CCAGGTGGAA GTTGGAAGCT TCCACCGTGA CGGATGGAAC CGGCGGAAAC GGGCGGTGGG
1021 TTGCGAAAAT TGCCCCCACG GTGCGAACGG GCTGTAGATA ACGATGTAAG GTAAAAACCG
1081 ACGCGCTTTT TTTTTTTGGC TTGCACAGGC GAAGAGGCGG AGACGTGTGA CCGGGTTGCG
1141 TGTGAGGGCT TTGAGACGGA AGGTATTATG ACCGAGATGG AGGAGCGGCC GATGGATCTG
1201 CGAACCAGGA AGTTTTAAAA TAAATAGAAG TAAAACCGCA AGTGTTTGTG TTGTCGTTAA
1261 TGAGGCGTGT AGGGGCGGGG GTGGCGGTTG CCGGGGTTGT TTTATTGTGG GCAGCGCCGC
1321 CTCCTAGAGT TAAATTGCCG TTTGGCGCAC CATGGAGCTG GCAGAGTACC TGGAGAACTT
1381 TGCCGTGCTG CGTCGCGTGT TGTATCAGAG CAGCGACCGC TACGGGGACT GGTGGCGTTG
1441 GTTGACGGGG TACAAGCTAG CCCGTCTGGT GCACGAGGTG CGCGCCGAGC ACCGGTACAG
1501 CTTCGAGGCC CTGTGTGGGG ACGAGGTGCT GCAGCGGCTG CTGCGCGATC AGGGTCGCCG
1561 GCTCGAGCCT TTCTTTCATT CCGTGTTGGA TTTTTCCACC GCGGGGCGCT CGGTGACCAG
1621 TTTGGCTTTT GCCGTGCACG TTTTGGGTGC TCTAACTCGT GACGGTCGCG GTCAGCCCCT
1681 GTCCGACGAT TTTTGGTTGG ATACCGTGTG CGTGGCCGTG CACCGGGCGA TATGTCTGAC
1741 CCTGCAGTGC CACCGGCGTC GGAGCCGAGG CAGCAGGGGG GCGCGTCGGC GGGAGGCATC
1801 CTCGGGCAGT TTAGATTTGG CGGACGCTTT GCCCCAGCCG CCCCCGTCGC CGCCGTCGTC
1861 GACGGTGACC GCGCAAGTGG AGAGGATGCA GAGGCGGGAC AGCAGTCGGG GGATGGACGC
1921 CGAGGGGGCG GCGGAGGCGG CGGAGGCGAC GGCGGCGGCG ACGCAGAGGA GAGAGGAGCA
1981 GGTGCGTCGC GCTCTGCGGT CGGATCTGTC GACCACGCCG CGGCGGGGGG CGTTTTAGTA
2041 GCGGCGGGAG GTAGGGGTTG TCGACCTCTG CCGGCTCCCG TGCCGGTGCG TCCGCAGCCC
2101 GTGGCGGCCG TGCCGGTGTT GGTGGCCGCG GAGCCGGGCG CCGGCGCCGG CGGAGGGGTT
2161 GTGGGGGGAG GCGTGCAGGT GGGGGTGGCG GCGGGGGGTT ACGACGCCGC CGTCAGCGTG
2221 GTGTGCATGG TGCCCATGGG TCTGGGCGAC GTGATGATCG TGCAGGACGG CGGGGAGGAG
2281 CAGGTGAGGG AGCAGAATTA CATCGCCGCC ATGGCTCACG GTATGATGGT GGGTCTGCAT
2341 CAGCAGGCGC AGGGGGCGGG AGAGGAACCC AGTTTGGAGG GGGTGTTGCG GTCTTTGCAG
2401 CAGCAGATGG AAGCCAGGGC CGGCGGGCAG CAGCAGCGGC GGGGCGGCGG GTCGGACGAC
2461 CAGCAGGCGG CCACCGCTCG GGGGGCCTCG GTGTCCTTCG AGAGCCTGCT GTTGCGCCAG
2521 GAGCGTCAGG GGTCCGGGGC CGTGTTGTTA GAGAGTTACA GCTTCGAAAA CATCGCGGCT
2581 TACGTGATGG GGCCGCAGGA CGACTGGGAG CAGTGCATCG CGCGGCACGC CAAGCTGTGG
2641 CTGGATCCGC GCACCGTGTA CCGCGTGACC GCGCCCGTGC ACATCAACAG CGTTTGCTAC
2701 GTGGTGGGGA ACGGAGCGCG GGTGATCGTG GAGTGCGAGC ACAACACGGT GTTCAACGTG
2761 GCGGGGCGAC GTCGGCCCGT TTCCATCTAC AGCATGTGGG CGGTGACCTT TCAGAACGTG
2821 GTGTTCGAGT GTGCGGCGCG TCGCCGCGTC ATCCTGTGCA AGTGCTCCAC CGACGTGAAC
2881 TTTCACGGCT GCAACTTTAT TAATTTTCCG GGCGTGTGTC TGGACGTGCG GGAGGGGGGT
2941 CGCATTAGGG GTTGTTATTT TTTGGGTTGC CGAGTGGGCA TCCGGTCTTT CAGCGCGCCG
3001 CGCGTCACTG TGCGCAGCTG CACCTTTGAA AAGTGTCTGG TGGGTGTCAT GGCCAAGTTC

3061 AAAATCAAGC TGGTACACAA CGCCGCGGTA GACACGTACT GTTTCTGTCT GCTGCGCGGC
3121 GCGGCCGTGG TAAAAAATAA CACGGTGATG AGCCCCACGC GTATCACCGA CCAGAATAAC
3181 ATCAGCGTGG TCACCTGCGC CTCGGGTCAG GTGGTGCCCC TGACGGCCGT ACACGTGGTG
3241 GGCAATCCGC GCACCCGATG GCCGCTGTTC GTGCAGAACA CCTTCAGCTG CTGCCGCGTG
3301 TATCTGGGAA ACCGACGCGG CAGCGTGAGT TTTACCGGCT GCGCCATGCA TTTCTGCGCC
3361 CTCATCATGG AGCGCGAAGT GATGCCGAAG GTGTCGCTGG CCGGCGTGTT TGACCAGAGC
3421 CTGTCGGCCG TTCGCGTGGT TTCTCGCGAG ACCGGAGGCA CCGTGGCGCG ACAGTGCGAG
3481 TGTGGTCACG TGCACGTGCT ACAGATACCT CTGGTGTGCG GAGTGACGGA GGAGCTGCGG
3541 GTGAACCATC GCGTGATATC GTGCGATACC ATCGACCACT CTTCCTCGGG CGAGGACTGA
3601 CCGATCGGCC GGGTACCCGG CGCGTGTGGT GAGGCGGTGG GGGTGGGTGG GAAAGGAGGA
3661 GAGATGGCGG GCAGAGGGGA TGAACGGGTG TACGGGAGTT ACGTTAACCG TCATGTGTAT
3721 GTTTTTTTTT TTTTGGTACA GTCAATATGG CAACCGGCGC GCACACCCTG TGGACGGTGA
3781 ACCCACCCCC CGGACCTTAC GGGACCGTGT ACTCTTACTG CACCCTGGCG CGATTACCCG
3841 GATGGGCCGC GGAGCGCTAC GGCGTTCCGG GCAGCACCAT TGACGGGCGC TTGATTCTGC
3901 CCGGAGAGAA TGCCAATTCT CCCTTCGTGG GAGTGAGGGG TCGGGCTTTT GTATCCCCCG
3961 AGGCCACCGG AGGATTGACC GCGGTTGTGG GTGGGGCCAA TGGGAACTGA TCAATAAAAA
4021 GCAATATTGG AGTCAAAAGA TTTATTATGA GTGTGCCTTT TTATTATTAT TATTGAGGGG
4081 GTCGGGAGGC GACGGGCAGG GGTGGAGGGG GAGGTGGAGG AGGGGGAAGA TCCCCGTGTG
4141 GCGACCGCCG CTGTCGGTTG TCGACCGCGC GACGGGCGCG CTGTTGCCGG TATTGGGCGT
4201 TCCATCTATC GCGGGCGGAC AGCTTACCGT GGATGTACTC GAGGCAGCGG TAGAGGACGC
4261 TGTAAGAGTG CAAACACAGG GGTCGCAGGC CGTCGGCGGG GTGCAGGTAG CACCACTGCA
4321 GGTCGGCGTG AGGTGGCTGA GCGTTGTACA CGATCCAATC GTAGCAGTGC TGACGGGCGT
4381 AGTGTCCGCA CACGTCTTTC AACAGCACGG AGATGGGTAC GGGTAGCGCT TTGGTGTAGG
4441 TGTTGATGAA TCGGTACAGC TGGTTGGGTT GCATCTGGGG GCTCATAAAG TGAAACTTGG
4501 CCTGGATTTT CAGGGTGGAG ATGTTGCCGG CCGCGTCCTT GCGCGGGTGC ATGTTGTGTA
4561 GCACCACGAG CACGGCGTAG CCGGTGCAGC GGGGGTAGCG GTCGTGCAGC TTGGAGGGGA
4621 AAGCGTGAAA GAATTTAGCG ATGCCGCGGT ACCCGCCGAG TTCCTCCATG CACTCGTCCA
4681 TGACGATGGC GATGGGGCCG CGGGCGGCGG CGCGGGCGAA GACGTTTTCC GGGTGCGCCA
4741 CATCGTAGTT TTTTTCCTGC AGCAGATCCG CGTAGGTGAG CGGTTGAAAG GCGGGCAGCA
4801 GCGTGCCCGA CAGCGGCACG ATGGTGTCTT CCGGACCGCG TTTGAAGTTG GCTTCGCATA
4861 TCTGGGCTCT CCAGGCCATG ATTTCCTGGG GAGGAATCAT GTCTAGGTGG GGTGCCACGA
4921 AGAAGATGGT TTCGGGAGAC GGTGAAATGA GATGCGAGGA CATGAGATTG CGAATCAGTT
4981 GCGATTTTCC GCAGCCCGTG GGTCCGTAGA TGACGCCGAT GAACGGTTGA ACCCCGTGAT
5041 TTACGGTAGC GGCGCGACCC TCCGCGTCCA GCAGGGCCGT TCTGTCGCGA TCTATGTCGC
5101 GCAGGGCGTC TTCGGTGGAC AGCAGCGAGC GCAGCAGCTG CGGGCCGGCG AGGGACAGCA
5161 TCTCTTCTTC GTCGGCGAAG CGGCGTAGGC GGTGCAGTCC GGGGGCGTGC GGGGTGTCCT
5221 TTAGCATGGC GTGTAGGTGT GCGAGGCGTT TCCAGAGCGA GTGCAAGCGT TCCAGGGCTC
5281 CGTCATCCAG CAGACGTGCG TGTTGCGCGG GTTGGGGCGG GAGTTGCTGT AGGGTACCAG
5341 TCGGCCGTCC GCCAGGAGCA CCATTGTGGC GTCCTTCCAG GGTCGTAGGG TGCGCGTGAG
5401 GGTGGTCTGG GCGACCGTGA AGGCCTGGGC CGACCCCTGC GCGTTGGTCA GCGTGCGTCG
5461 CAGGCTGAGG CGTTTGGTGG CCAGGTGTCG CTGCGGTGGC GGATCGTTCC ATTGCGGGGC
5521 GCCGTGGGCG TCGGCGCGCG GCGGCCCGTA CTCGGCGAGA CAACGCCTCA GCGTCTCGTA
5581 GGTGAGAGAG TGCACGTTGT GTCCCTTGGC GCGCAGTTTG TCCGGTCCCC CCGCCGGCCG
5641 TCCGCAGGAG CCGCAGGTAA TGTCCCGCAG CGCGTACAGC TTGGGCGCCA GGTAGATGGA
5701 AGAGGCGGCG TAGGCGTCCG CGCCGCAGTG CGCGCAGCGC GTCTCGCACT CTACGGCCCA
5761 GGTGAGGCGG GGGTCGGCGG GGTCGAAGAC GAGCGCGGCG CCCGCGGACT TCAGTCGGTG
5821 GGCGCCCGAC GTCTGCATGA GGCGGTGACC GCGCTCGGTG ACGAACAGGC TGTCGGTGTC
5881 GCCGTAAACT GACGTCACCG GTCGACGTTC CAGGGGCGTG CCGCGATCGC CGGCGTAGAG
5941 CAACTGACAC CAATCTGACA TGAAGGCCCG TGTCCACGCC AGCACGAAAG ACGCGATCTG
6001 GGAGGCGTAC CGGTCGTTGT CCACCAGGGG GTCGCAGTCG GTCAGGGTGA GGATGCAGGT
6061 GTCTACGTCA TCAACATCTA AAAATGTGAT TGGCTTGTAT CTGTAGGGGG TGGAGTTTGT
6121 GGGGAGGGTA TTTAAATCGT GACGTGAGTC GTTGACGTCA TTTGCGATCG TATCTGCGTC
6181 CGTGGACGTG AGCGTTTCGC GCGCGCGTGG CAACCGTCTC GGCGAGGGTG AGTGACGGGG

6241 GTCCGACGGG GGCGATACAT CGTCGGGGGT ATCTCGGCGG CGGCCGTCGG TTTCGGGGTC
6301 CTCGGGCAAC GTGGGCGAGT GTCGATGCAG AAAGGTGGTG AAGGCGGGGA TGACTTCCGC
6361 GCAGAAGTCG TCCGTTTCTA AGAAGGCGGC GCTTTCCACG CTGAGTGCCC CTTTGGCCAG
6421 GTCGGCGGCC AGACGGTCGT CCAGCTGGTC GGCGAAGACG TTGCGTTTGT TGTCCAGCCG
6481 CGTGGCGAAA GAACCGTACA GGGAGTTGGA CAGCAGCTTG GCGATGCTGC GCAGGGTGGC
6541 GTTGCGGTCG GCGTCCGCAC GCTCTTTGGC GGCGATGTTC ACGCTGACGT ACTCCCGGGC
6601 CAGACAGTTC CAGGCGGGAA ACACGGTGGT GCGTTCGTCG GGTACGACGC GCACGCGCCA
6661 GCCGCGGTTG CAGAGGGTGA CCGCGTCCAG GCTGGTGCAC ACCTCTCCGC GCAGACGTTC
6721 GTTGGTCCAG CACAGTCGAC CGCCCCGTCG CGAGCAGAAG GGCGGCAGCA CGTCCAACAG
6781 ACGTTCGTCT GGAGGGTCGG CGTCGATCAC GAATATGGCC GGCAGGAGGC GACTGTCAAA
6841 GTAGGACACC TCGCGACCTC TGGCGAACAG CGCGTTATAC TGTCGTAGCA CGTGCGCGGC
6901 GTCGCGCGGG CTGAGGGGCC GCCCGCAGGG CAGCGGGTGC GTGAGAGCGC TGGCATACAT
6961 GCCGCATATG TCGTACACGT ACACGGGATG GGGGTATTCT CCCAGAAACG TGGGGTAGCA
7021 TCGTCCGCCA CGGATGCTCT GTCTGACGAA CTCGTACATC TCGCGCGAGG GGGCCAGCAG
7081 CGTATCGCCC AGGTCCGCGG CGGCGTCGTG CCTGCCGCAG GCGCGGAAGA GCAGCTGTTT
7141 GAACAGGGCG TGAGAGTTGG AGCTGATGGT GGGCCGTTGG AAGACGTTGA AGGCGCAGCG
7201 TGGCAGACCG GCGGCCTCGC GCACGAAGCG CTCGTAACTG CGCAGTAACG TGTGCACCAG
7261 CGCGGCGGTG ACCGCCACGT CCTGTTCGCA GTACCGCAGG GTGTGCGCCA CCAGGTCGTA
7321 GGTTTTTTCG CCGGTGCGCC GCCGGCGTCG GCGCCAGGCG TCGCGGTTGT GTTCATACTC
7381 GGCGTGGTCC TGCCAGTACC GTCGCGCCGG AAATCCGTCC TCCTCCGCCT CGTAGGCGCC
7441 GGTCATGTAA AAGTCGTTGA CGGCCTCGTA GGGGCAGTGT CCCTTGTGCA CCGCGAGGTC
7501 GTAGGCGCCC GCGGCCTTTT TGAGACTGGT GTGCGTGAGG GCCAGGGTAT CCCGCACCAT
7561 GATCTTTACG TATTGATGTT TGAGGTCGTG GTCGCGTACG TCGCCCTCGG CCCACAGCCG
7621 GCAGTCGCGC AGCGACTTGC TGTAGCGAGG GTTCGGCAGG CAGAGAGTGA TGTCGTTGAA
7681 CAGCAGCCTG CCGCGGCGGG GCAGGAAGTT GCGGCGCACC GAGAATGCGG CCGGGTATCG
7741 CACGCGGTGA GCCAGAGTCT GCGCCGCCAG CACGATCTCG TCGAACCCGC TGATGTTGTG
7801 ACCCACTACG TATAGTTCCA CGAAGTGAGG TGGTCCCGCC AGCTTGTGGG CCAGGGTCTT
7861 TAGCGTCTCG TGGGACAGAC AGTCGACGTG TTCCAGACCG CGTTCTCGGG CGACGCGTTC
7921 CAGCCAGGGA TTGGCGTTGA CCACGTCGTC CCACACGTGC CGCGCCGCAT CGTGCGCCAG
7981 GCCGTCCCGG AAAGCTTTGA ACATGTATCC CACGTGCCCT TTCTGCGGAT GCATGGCAAA
8041 GTACGTGCGC GGATGAGCGT CCGAGGCCTG CCATCCCGCC CGGGCGGCCG CGCGCTCAGC
8101 CAGCGTCACC AGGGCGTCGT CGCCCGAGAG GTGGAATACG AGCATGAAGG GCACCAGCTG
8161 TTTGCCGTGG CGTCCGTGCC AGGTGTACGT TTCCACGTCG TAAGTGACGA ACAGGCGACG
8221 CACGGACGGC GGCGAGCCGA TGGGGAAAAA CTGAATCTTT TTCCACCAGT GACGCGTGTT
8281 ACCGTGCACG CGGTGAAAAA AGTACTGTCG TCGTCGGGCG TCGCAGGTGT GATCGCGCAG
8341 ATAGGTGGAC CCGCAGTGCG CGCAGCGGGT GGCCGGCTGC ACGTTCTTGA TCAGGGCGAG
8401 GGCCCGAGGG TGGTTCTGAT CGACCATCAA TTTTAAGGCA AAGGGAAGTG TAAAATAATC
8461 GGTGGCGGTG TGGGCTTCCA GGGGGTCTCG ATTCTCCGCG TCCTCGTCGT CGTGGTTGTC
8521 GCAGTCGACG ACCTGCCCGT CGCCGACCTC CGTCATACGG CAGCGACCCG GACCGCAGCG
8581 GGTCACCAGC GTGGCGGAGA ACGGTCTGAG CAGTCGGAGG AGGTGGCGCG CCGACGCCAC
8641 CTCCCTCCAG CCCGGGCGTT GCGCGTCTAG ATTGACGCGG TGACATAGGG CCTGCTTGTG
8701 CAGCGCTTGT GTCCACTCGC GGTGATAGGT GACGCGCAAT CGCTGTCCGG TGTCGATGTC
8761 AGTGAGGTCG CACTGCAGGA TGGCTCGCGA GGCCCTGAGG GTGCCGCCCC GCCGTGCGGC
8821 CGCGTCGTCG GGCGAGGCTA ACGGCGGGCT TGCCCTCCGC AACGGTGACG GAACGCCAGG
8881 GCCGGGGCCC GGAGCAGCGG CGGCGCGGAG GTCACTTCGG CGTAGGGCAC GCGCTGCTGC
8941 GCCGCCGGCA GCTCCCGACC CTGGCGGCGT AGATGCGTGG CGTAAGCTAT GACGTCCCGG
9001 TTGATGTCTC GGATGTTTCT GTTGTCCGAG AAGGCCACCG GTCCCGACAC TCGGAATCTG
9061 AAAGAGAGTT CCACAGAATC AATGTCGCGG TCGTTGGCCG CGACCTGTCT CAAGATCTGG
9121 GCCACGTCGC CGCTCTTTTC GTGAAATTGG ATGTCTTGCA TGAAGCGGTC TATCTCGTCC
9181 TCGGAGGGGG GTTCTTCGCC GGCGCGTCGC ACCGTGGCGG CCAAATCAGG GCCCACGCGT
9241 TGCATGAGAC GGAGGAACGC GTCGTGTCCG ACCTCGTTCC ACACGCGTGT GTAGACCACC
9301 TCTCCCCGCT CGTCCCGCGC TCGCATGATG GCCTGAGCCA GGTTGACTTC CACGTAACGG
9361 GAAAACAGAC GGTCCCTGAC CAGGGTGAAG TGCAGGTAGT TCAGCGTGGT GGCCATGTGC

9421 TCGGCGATGA AAAAGTAGAT GACCCAACGG CGCAGGGTGT GCTCAGTGAC GTTATCGGTC
9481 CTCTGCAGTT CTTCCATGGC GGAGTAAAAG TGGACCATGA AATCGAACAT GCCCTCCTGC
9541 CTGGTTTGCG GCGACAGCAC TTCGCGCAGC TGGGCGATGG CCTCCGAGAC GGCGGCGCGG
9601 ACCTCGCTAG CAAAATCACG CGGATCGGGA GAGGCGGCGA CCTCGGGGGA GGGGGGAGGG
9661 GAGGGTGAGG CGGGAGGCGA CGCCCGCGTG CCGGGCATGG GACGCGTGCT CGGCCCTTCC
9721 AGGGGGTCGG CGGGCGTGCT GGGAGGGACG CGACGTCTGC GGGTGCGGAG AGGCAGACTT
9781 TCGATGAAGC GCGCCACGGT TTCGCCGCGA CGCCGCCTCA TGGTTTCGGT GACGGCGCGG
9841 CCTCTTTCCC GCGGTCGCAA CGTGAAAGCT CCGCCGCGCA TGACGTAGGG TAGGGCGGGC
9901 ATGTCGGGCA ATGAGAGCGC GCTGATTGCC TTTTTGATGG TTTCCGTGAC AGGTTCGCGT
9961 CCGCGTCTCA GGAGGTCGTC CGTGTCCAGG GAGCAGAAGT ATCCCACGAA GGTTCGCAGC
10021 CAGCGATCGT CGCAAGGTAG GCACAGCGGG AACGCTATTT CGGGTCTATC GACGTCGGGG
10081 GGCGTAAAGT TTTCGCTTAG TAAAAAATGA AAGTAAGCGG TACGCAGCCG CCGGGTGGCT
10141 TCCAAAACGA CCAGATCGTC GTCGGGCGCG CTGCGGTGCC TCCATCGCTG CGACAGTCCC
10201 CACGCCGCCT CGCCGCAGTC GCTGAGGGAT TTGTACTGCT CGTGGAACCA GCGTTCCACC
10261 GGCGCGGCCG CGCCGTCGTC GGCGGCGCCC GCGCCGCCTC CGCGCGGAGC GCCACGACCC
10321 CCAAAGGTAG AGATTCCAAA ACCGCGTTGT GGCTGTATTA GGGCCAAATC GGCCACCACG
10381 CGATCGGCCA GCACGGCCTG CTGCACTTCC GCTAGAGTGC GTTGAAAATC GTCTAGGTCT
10441 ACGAAGCGGT GGTAGGTACC GGTGTTGATG GTGTACGTGC ACGTGGCGGT AACCGACCAG
10501 TCCACGTGCT GCGTTCCGGG GTGCACCGTA TCGTAAAACC TCAGACGACT GTAGGCGCGG
10561 CTGTCGAAAA CGTAATCGTT ACACGTCCGC ACCACGTATT GATAGCCCAC CAGTAGGTGT
10621 GGAGGCGGAA ATCCGTAAAG CGGCCAGAAG GCGGTGTACG GGCTGCGCGG GCTGAGATCG
10681 TGTAACATCA GGCGGGCGTA GCGGTACACG TAGCGCGCCA TCCAGGACAG ACCGGCGGTG
10741 GTCACGGAGG CCGTTACCCA CTGGTTGGCT CTACCCCAAA TATTACGTAG CGGTCGAAAC
10801 ACGTGAATAG TGTAGGAACT CTGACCGGTG AGTCTGGCGA CGTCCTGAAC GTTCTGATGA
10861 GCATAAAAAA AAAAAGAAAC GACACACAAA CGCACAAACA AAGACAGAGA AAGATACGAT
10921 TGGGTTGGTT AACGGGACGA CGACGATAAC CCGCGACGGC GGCGGTGACG ACGGCAGCAA
10981 CGGCGGCGGC GTGGCGTGAA ACGAGGTGTG CATGTGGTTG TGCTTTATTT TTTTTTTTAC
11041 AGCGGTTGGG TCGCTTGGTT GGCAGGAGCC ATGCAGCATC CCGCCGTGAA TTACGGCGCC
11101 CGCGTTCGAG ACTACGCGTC GTCGTCGACG ACGCGACGTC GCGGTGGTGC GGGTTCCGTC
11161 GCCGATGCGA CAAGCAGCGC TTCTCCCGCA TACGCGGGCG CGACGGCCGA CATCGCCGCC
11221 GCCGCCGCGG TCGATGCCGC CGTTGCGACC TCGGACGTTG AAGGGGAGGG TTTGGCGCGA
11281 CTGTATTGCG ACCCCGAAAC CCATCCCAGA TGCGTGATGA AGCGCGACGG TTGTGTGGCC
11341 GCCGTGCCTC AAAAAAACGT GTTGCGCGAC GCGTGTGACC CGCCCGGCGG CAGCGAGGCG
11401 GAGGATCTGC GGCACGCGCG CTATGCGGCC GGTCGCGACA TCGCGTTGGA TCCGGAGTCG
11461 GTGATTCGCG ACGACGACTT TGTGCCTTGC GACCGTTCCG GCATCAGCCG TGCTCGGGCG
11521 CAGCTGCAGG CCGCGGACCT TAAGACGGCC TTCGAGCAGA CCGTGCGACA GGAAAGCAAT
11581 TTTCAGCGTT CCTTCAACGC TACCGTGCGC GGCATCGTCT CTCGAGACGA CGTACTAGTC
11641 GGGTTGATGC ATTTGTGGGA CTTTGTGGAG GCTTACGTGG GCAACCCTAC CAGTAAGGCG
11701 CTCAACGCGC AGTTGTTCCT CATCGCTCAG CATTCGCGAG ACGAGGGACT CTTTAGGGAG
11761 GCCTTGCTTA ACATCTGCGA TCGCGAATCT AAGTGGCTGC TGGACCTGAT CAACGTCTTG
11821 CAGACCATCG TGGTACAGGA GAGACATCTA AGCGCCGCGG AGAAGGTGGC GGCCGTTAAC
11881 TATTCGGTGA TCACGCTGGC CAAACACTAC GCTCGCCGCA TTTATCAGAG CCCATATGTA
11941 CCTCTGGATA AAGAGCTTAA GATCGCTACT TTTTACATGC GAATGGTTTT GAAGATATTG
12001 ACCCTGGCGG ACGATTTAGG AGTGTATCGT AACGAACGCA TGCAGCGGGT GGTGGGTCTG
12061 TCCCGTCGGC GCGAACTTAG CGACGCTCAA CTGTTGTGTG AGCTACGACG GTGCTTCTCG
12121 TGCACCGAAC CCGCCCCGGC CGGTGTGGTG GGGACGGCTT CCGTACAGGC CATGAATCCT
12181 GCTTCTCGGG CGGTGTTCGA TCTGAGGGAG GACGACGATG ACGAGGATGA TATGGAGGAC
12241 GAGTCGGCGG CCTGAACGGA AGCCGTGTGA GCGCGTGGGC GTCACCCGTA CACTGTGTTT
12301 TGTTTTTTTG TTTCCCGTGT GTTCCGTTTT TTCTAGTATA TTCTATCCGC CTGGGTTCCC
12361 AACCCTAGTC GCCTACCCTC CACCCCCCAA AAAACACATA AAAAGATATG GCGGCACACG
12421 TCGCGGCGGG CGCCGCCGCT CGCCCGCGGG CGGCAGATTC GGCGTTACGT GCGGTGCTGC
12481 AGGCCAAGGC CAGCGGGGAT TCCAACTGGG ACGAGGTGAT GCGGAGGGTT TTGAGCCTGA
12541 CGACCGGCAA CGCACACTTT GGAAATCTGC CGCGCGCCAC GCGATTTGAC ACCATCCTGG

12601 AAGCGGTGGT ACCGTCGCGT ACCGATCCCA CCCACGAGAA ACTGGTGGCG GTGGTAAATC
12661 TGCTGATCAG CGGCGGGGCG GTGAGACCCG ACGAGGCGGC CCAGATATAC ACGGCCCTGT
12721 TGGATCGCGT GTCCAAGTAC AACAGTCAGA ACGTGCAGAG CAACTTGGAT CGCCTGGTGA
12781 CGGACGTGCG CGAGGCGGTG GCGGTGAGGG AGAGACAGAT GGCCGCCGGC AACGTATCCT
12841 CACTGGTGGC GCTGAACGCC TTTCTGGGAA CCCTGCCGGC TACGGCCGAA AGAGGCGGTC
12901 AAGCGGATTA CGTGTCCTTC GTGTCCGCCC TGCGCATCAT GGTTAGCGAA GCGCCCCAAT
12961 CCATGGTATA CCGTTCCGGC CCCGACACCT ACTTTCAGAC AAGTCGTCAC GGCACGCAAA
13021 CGGTTAACCT GACGCGCGCC TTTGCTAATT TGGCTCCTTT ATGGGGGGTG GCCGCCCCCG
13081 CGGACGGTGC CGCCACTGTG GGTAATCTGT TAACCCCCAA TACCCGTTTA CTGCTCTTAC
13141 TTGTGGCGCC GTTTACGGAC GCCAACAGTC TGTCGCGGGA CAGTTACCTG GGTCATCTCG
13201 TAAACCTGTA CCGCGAGGCA ATCGGGGTGA CGCATTTGGA GGAACGTACG TTCCGAGAAG
13261 TGACGGAGGT CAGTCGCGCC CTGGGACAGG AAGACCAGGG CCAGACCTTG CAGCAGACGC
13321 TTAACTTTTT GCTGACTCGG CGCGCCGCCC CACGTCGCGG TAACGTGCGA TTGAGCGAGC
13381 AGGAGACGCG TCTGCTGCGT TTCGTGCAAG GTGCGGTAAG AGACCTAATG TTCAACGAGG
13441 GTTACGCGCC GTCGGACGCC ATCGACCGCG CCTCCCGTCT GTTGGAAGAC GGCACGTACG
13501 AACCGCACGC CTGGTTTTTG GATTACGTGT TCGATTATTT TCATCGGGCG GCCGCCATGT
13561 CTCCCGACTA CTTCGTGCAA GCGGTGCTGA GCCCGCAGTG GGTGCCGCCG CCCGGTTTTT
13621 TCGATCGCCA ATTTGTGTTT CCCGAAGACG ACCCGCCGGC TTCCAGCAGC GGCAGTCTGG
13681 TGTGGGACGA CGGGGATTCT TACAAGACCC CGTCTTCGAC GTCCTCCTCC GCCTCCGTTT
13741 CTCCGGCGTC CTCTCTGCTG TCCCTTCGGG GAGTGGACCC GGCCAATGTA TCCTTACCAC
13801 CCTCCAGCGT CATGTCCCTG TCCGGGTCGC CTTCCGCTTC GGCCGTTTCT CTTCCCGTGG
13861 CGTCGTCTTC CGCCGCCAGT CGTCGAAGCA GCGGCGCTAC CCGTACGTCC TCGCCCCCCG
13921 GAGCAACCGC CTTCGACCCG TTTTCCCTGT TGGGCGCTCG CCCCAAGATG ACCAACGTAG
13981 AAATGGGTAT GCGTCGCGCG GCGCTAATCA ATAATTCCGT TGACGATTTG GCCGCGCGTC
14041 TGAGTCGCGT GACCGCGTCC GCCTCGCCCG TGCGACGGCG GCGATCCCGT TCGCCCCTGT
14101 TCGAACGCAC CGGTTCGCCG GACGACCCCA TGACGGGCGA CAACGAGGAT GACGCTTTTG
14161 GTTATTTGCG TCCCCGCGGC AGATTTATGT AACGGACGCG CCAGGTTTTG AAAAGGAGGT
14221 AGCGCTTACT TACCATGATA ACGACCAATC TGTGTGTTGA GTCTCCGTTC GTCTGTCTGT
14281 TTACGAGGCC GCGTTCATAG ATAGCTGACG TGCTGTATGT GAAAAACATA AAATAAAACA
14341 TTTGTAATAA AAAATCACTT TGTGCGCGGT GCGTGGCTAC GTGTGTTTTT TTTTTGTTA
14401 AACGCGTGTA TCGTGGTCGC GTTTCAGCCA GCCGCACGGT TCGCTACCCG TCGTCCGTCG
14461 TCATCATGGC CGCCGCTTTG TATTCCGAAC CCCCACCCGC TTACGAAACG GTAGTGGAAC
14521 TGTCGCCGTC TCTGGGCGCT CCTTACGTGC CCCCCCGTTA CATGGCCGCG CAGGCCGGTC
14581 GCAACTCCAT TACCTACGGC GCGGGTGGAA ACGCTCGACA GAGAGACACC ACCCGCCTGT
14641 TTCTGATTGA TAACAAATCG GCGGACATCG AATCGCTGAA CTATCAGAAC GATCGCAGCA
14701 ACTACGTGAC CACGGTCATT CAAAATGACG ATTTTAGTCC CTACGAGGCC AGCACACAAA
14761 CCATCAATTT TGACGACCGT TCCAACTGGG GTGCTGAGTT CCGCAGCGCC CTGAACACCA
14821 ACCTGCCCAA CGTGACGGAC TTCATGCGCA GCAACAGCTT TCGCGCCAAG CTGATGGTGG
14881 CCCGCGAGGG CGATGCCAAA CAGCCGCGCT ACGAGTGGGT CACCCTCACG TTACCGGAGG
14941 GAAACTTTAG CAGCGACAAG GTGATCGATC TGATGAACAA CGCCGTGGCT GAGCACTACG
15001 CCGCCGTGGC GTCGCAGCAC GGCGTCGCGG AAGACGACAT CGGGGTTAAG TTTGACACGC
15061 GCAACTTTAG ACTGGGATTC GACCCCCGTA TCGGTCTCAA TACCGACGGG CAATACACTC
15121 ACGAGGCTTT TCACGCCGAC ATCGTGTTGC TGCCCGGATG CGCCGTCGAT TTTACGCGGA
15181 GCCGACTGAA CAACATGTTG GGCATCCGCA AGAGGCTGCC GTTTCGTCGG GGCTTCGTCA
15241 TCTCGTACGA AACTTTGGTG GGCGGCAATA TTCCCGCCCT GGCGGATCCC GCCGTGCCGA
15301 AGGGTGATCC GCCGCGTCCT CTAACGACCG ACGCTAAGGG TCGAAGTTAT CACGTAGGAC
15361 AGGACCCGTC CGTACCGGCG ACCTTCACGG CCTACCGCAG CTGGTATCTG GCGTATAACT
15421 ACGGACCCCG CGACGGTCCC GCGCGCACAG AAACCGTGCT GACCACGGAA GACATCACGG
15481 GCGGCATCGA GCAGGTGTAC TGGAGCCTGC CCGACGTGGT CAAACCGCCT ATCACGTTCC
15541 GGGACAGCCA GCTGTTTGAT CAGCTGCCCG TGGTGGCGGC CGAACTGATG CCGGTTCGCG
15601 CTCGCACGTT CTACAACACG CAAGCCGTGT ACGCGCAGCT GATCCAGGAC GCGTCTAACC
15661 GCACGCACGC GTTCAATCCG TTTCCCGAAC ATCAAATTTT AATGCAGCCC CCCGCTAGCA
15721 CCATCGCCGC CATCAGCGAA AACGTGCCGT CGGTCACCAA TCACGGCGTG CTGCCCATTA

15781 AAAATCACGT ACCCGGCGCG CAACGAGTGA CCATCACCGA CGCCAGGCGC CGGACTCTGC
15841 CCTACGTGTA CAAGACCTTG GGTGTTCTGA GTCCCAAAGT CCTGTCCAGC GCTACGCTGT
15901 GAAGTGGTGT TACGGTTTTT TGCAAAGTTC GTGGGTGTAA AACGTGTCTC GATTTTTTCT
15961 TCCAAGTTAT GGACATTTTT TTTTTGATTA TTAATTCGCG CCCGTCGTCT TGTTTTTTTT
16021 TTGTTCTGTT CTTCGTAATA AAAAAAAACA CAAAGAGTCG ACCGTCGCCG TCGTTATAAA
16081 CGCCGCCACG TCCGTGCCCG TCGTCTTACC CAAACACATG GCCATCCTTA TCTCGCCCGA
16141 TAACAACACG GGGTGGGGAT TGGGCATGTC GCGCATGTAC GGCGGAGCTC GCTTTCGCTC
16201 CGAGGGTCAC CCGGTGCGAG TGCGCCAACA CTTTCGCGCG CCTTGGGGCT CCCTGAGAGC
16261 TTCTATCGAG CGCGTCGTGC GGCGTTCTTC AGCTCCCGCC GTGGTGACCG CCGACGTTCC
16321 CGCGGTGGCC GTCACCGACG CCGAACTGGG TCCCCCACCC GCCAGACGAC GGAGATTATC
16381 GGCGTCTTCC GCGCCCGCAG CCGTAGCCTT GCGCACGCCC GCGGTAATCG TACGACCCCG
16441 ACGACGCCGT CGTTTTACGG AACCGCTAAG AGAACCTTAC GTCTTGAGAA GGCGCGCTCC
16501 CAGAGCCGCG GTGGTGCTGA ACGTGGCCCA CGGACGGGCT CGTACACGTC ACAGACGACA
16561 ACGTCGTCGT CAGGTTCGCG CCAGGCTGGT AAACGCCTCC AGGGTTAGGA GACGACGCGC
16621 CCCCGCCGCC GTAGTGACCC TGCCCGCCTG ATGGACATCA TTTTTTTTTT GTTTTTTGTT
16681 TTGTTTGCTT TTATTTTTTT TTTTTTGAAC GTCGCGCTTC AGTCAGTTAT GATAGGAGCC
16741 GCGGCCATCC GCGCGGCGCG CGACAGGCGC GTCAAACGTC GATCGGTGCC CTATTCGGCC
16801 TCGGTTCGCC ATCGTTCCGC TCCTGTCCGT CCAGTGACGG CCGTGCCGTT GGACATGTCA
16861 AACCCCACCC CCGGTCTCAC GCCCGTTACG CCGCAACATC CCGTTGTTAC GCACGCCGTC
16921 GCCGCAACTA CTCCAGCCGC CGCCGCCGTA CCCACCGTGC AGGTGCTGGC CCCGGTCGCC
16981 GCGACCGCCG CTCACGCCGC CGCGACGTGT CACTCGCCCG TCGTTAAGGT GGACGCGTCC
17041 ACGTGTACAG AAGAGGCCAT GGATTATCAA AACCAGGGAC ACGCGGCCAC CGCCGTTACC
17101 GTGCGACCCG TGAAGCGGGT GACGCAGGGC ATAGGCGTGC AAACCGTGGA CGTGCACATG
17161 ACGCCGGTGG AAGAACTTTA CCGACCGCCG CCCTACGCTT ACCGTTCCGC CGCTCCTCGC
17221 TACTATCGAA CGTCCAGAAG GCGCGCGCCC GCCGCGCCGC GAGTCGTGCC CGTGGTATAC
17281 CACCCGTCTA TGGCCGGTCC CATATTTCAC CCACCGCCGC CGCCGCCTCC GCCACCGCCG
17341 TTGCAACAGT TTCGTCCCCG CTCGCGCGCG TCCTCGGGCG CGCGCCGCAG AAACACCTCC
17401 AGGGCGCCCG CCGTCACGTC TTCTCGTCGC CGCACCTCCA TCTCCGCGAC GACGGGCAAG
17461 CGCAAACGTT CCGGAGCTCG CGTAGTAGTG ACGCCGGTGC CTTTGGCGCC CGCCAGCACT
17521 AACGCGCGTC CCGCCGCGGC TAGGAGAACC GCTACCGCGC TACCCGTTGC TGTGGAAGTT
17581 AAATCCCGTC GACCCACCGT ACCGGTGGTC ACCTACCACC CGTCTATCAG AAGCCGTTAA
17641 AAATTTGGCT GTTTTTTGTC GTAGTTTTCT TGTATATGTT TCGTGTAGGA GTTTACGCGC
17701 GCACGTAACC ACGCGGCGAC GCCGTTTCGT TTTTTTTTTT GTCTTTCGAC GCACAGTCCA
17761 GTCATGACCG GCGTCACTTA CCGAGTTCGA ATTCCCGTAA GAACTCGCCG TCGTAGGCGA
17821 GTGTACTATC CCACCGCCGG TCTCCGACTA TCGTCCGCCC GCGGCGGAGC AGTTCGTTAC
17881 CGAAGACACC GTCGCATGAG AGGCGGTTTT TTGGCCGCTT TGGCGCCCAT CATCGCGGCC
17941 GCCATAGGAG CCATTCCGGG AATAGCCTCC GTGGCCATGG CCGCCAACCG CCGGTGAAAC
18001 CGCCTACCCG CTCGTTTCGT CCCGTCTCGC CGAAAGCCAG AGCATTTGCG CCCGTTCCGC
18061 TGAATGATTG TGCCGCCGAC TAAAAAAAAA CGTTACACGC ACATACAAAC CGTCTAGCGT
18121 TTTTTTTGAA ACGCGCGTAA TGACGATGCC TGTGTTTTTC TTTTTTTTAT CTTCGACCAT
18181 TCACTCCACC CCGTTTACGA CGCGCGATAC AATGTAAAAT TTTTGTTATG TACGTCCGCC
18241 GTATCGTTTC TGCGTGACGG CAGATGAATT TTGCTTCTTT GGCTCCACGC GTGGGAGAAA
18301 GAGCCATTTT ATCCCACAGT TATTATCTGG GGGACACGGG ACTGCACGGA GGTGCCTTGT
18361 CCTGGGGCTC ACTGTGGACC GGCCTAAAAT CTTTCGGTTC TCGAGTATCC AATTGGGGCT
18421 CGCAGCTGTG GAATAGCCGC GGCGTGCACC TTCTGAGACA GCGTTTGGAC GAGACCGGCG
18481 TGGGAGACAA AATAATCGAC GGCGTCACCG CGGGTATTCA CGGAGCTCTG GATATAGCGC
18541 GACAGCACAT GGACCAGGCG GTAGCCGACA GACTGACGCC TCCCGCGGGC GTACCTGCCG
18601 CCGCCGTGCC CGCGGCCGCG GCCGCGTCCG AAACCTTACC CTCCGCGCCG GCGGATGTCA
18661 CGACCGCCGC GGCCGCCAAC CTGGCTTCTC AGCGCGTCCC GCCCGTGGTG TCTTCGCTGC
18721 CGGCCCAAAC CGTGGAAGTG GCCGGTCCCG CGCCCGTGAA GGAACCTCCT CCGCCTTACG
18781 TTGAAAAAGA AGAGAGCGCG GACGATGGCC GCTGGCGAGA GCGCGTGACC ACCGTGGTGG
18841 AGCCCCCTCC CTCCTACGCC TCGTTGTATG GCGCCAACTC TAAGACCCCG CTGAACGTGT
18901 CCACCGCGCC CGTGGTGCAG CCGGCGGTCG CGCCCCTGCC GCCCTCCGAA CCCGCGGCCG

18961 CCGACGTGGT GGTTCCGGCT ACGACGGCGG CGGCGCCCGT CGTCGTAACG CGGCCCGGCG
19021 GCAGTAGAAA GCGGTACAGA GGAACCCGTT GGCAAACCAC GCTGAACGAC ATCGTGGGTC
19081 TGGGATTTCA CATCGATAAG CGCCGTCGTT GTTACTGACC GCCACGCGCC GCACTGTAAA
19141 TATGTGTTAT GCTGAACGTA ATAAACTGTC ACAGACGTGA AAAAACTTTT GCAAACTTCG
19201 GACCGCCCAT AATTTTTTCA CCGGTGTGCT TGCCTCTTTT TTTTTTTTAT TTATCCTTGT
19261 GTTCCTGCAC TGTACTGGCC TTGCCGTTCA TCGCCGCCGC CCGTGTATTA AAAAACAAAA
19321 AAAAAAAAAC ACAAAGACTT TTTAAACTCG CAGTCCCCAC TCCGCGTTCA TCGCTCGCGC
19381 CGACACCATG GCGGTACCGT CCATGATGCC ACAGTGGTCG TACATGCACA TATCCGGCCA
19441 GGAGGCGGTA GACTACCTGT CTCCCGGCCT GGTGCAATTC GCCCGCGCCA CCGACAGTTA
19501 TTTCCATCTG GGCAATAAGT TTCGTAACCC CACCGTGGCA CCCACTCAAG AGGTCACGAC
19561 CGACCGCTCT CAGAGACTCC AACTGCGGTT TGTGCCTGTG GATCGGGAGG ACACTCAGTA
19621 TGCGTATAAG ACCCGTTTTC AACTGGCCGT GGGAGACAAT CGCGTCTTGG ACATGGGCAG
19681 CACGTACTTC GACATTCGCG GCACCATCGA CCGCGGGCCC TCCTTCAAAC CCTACAGCGG
19741 CACGGCCTAC AACAGTCTGG CGCCCAAGGG CGCGTCCAAC AACACCATGT ACACCCACGT
19801 GAACAACCAG CAGCAGCAGG TCGGCGTGGT GGCACAGGCG GCCTTTTTGA CGGAAAACAT
19861 CGACCCCCAG AACGGTATAC AGGTCCGCGT GGACGCCAAC GGTCAGGCCG TACGCGCTCA
19921 GGCGCGTTTC GAACCCGAAC CCAACGTCGG TAACGAGACT TGGGTGTATC ACGACACGGT
19981 GCAGCGCGAC GTGGGTCCCG TGGCCGGACG TGTGTTGAAA GGCGACGTCA TGCCCATGCC
20041 CTGCTACGGC TCCTACGCCC GCCCCACCGG TGCCGACGGC GGTCAGTCGG TGGACAACCA
20101 GATAGATCTC ACCCTGCTGC GTAGCGGCAA CGCCGCGGGC GCGCCCGAGA TCGCCCTGTA
20161 CGCCGAAAAC GTGGACCTGG AGACTCCCGA CACGCATCTG GTGTCTCGCG TGGATCCCGG
20221 CGAGGCTCGC CTGGCCCCGG CGCTGGGACA GATAGCGCAG CCCAACCGAC CCAATTACGT
20281 AGCCTTCCGC GACAACTTCA TCGGACTGAT GTACTACAAC AGCAGCGGAA ACCTGGGCGT
20341 GCTGGCCGGT CAGTCGTCGC AGCTCAACGC GGTGGTGGAG CTGCAGGACC GCAACACCGA
20401 GCTGTCCTAC CAGCTGCTGT TAGACAACCT GGTGGATCGC ACGCGGTATT TTGCCATGTG
20461 GAATCAGGCC GTGGACAGCT ACGACCCCCA CGTTCGCGTC ATTGAGAATC ACGGCGTGGA
20521 AGACGAAATG CCGAATTACT GTTTTCCCCT TTCGGGCATG GTGTTGGAAG AGGCCACGCA
20581 GGTGGACGCG CAAAACGCCC GGGTTAACAA CCAGGGGCCG TCCTACGCCG GCGTGGGCAA
20641 CGTTCAGGCC ATGGAGATAA ATCTGACTCA GAACCTGTGG CGAGGGTTCC TGTACTCCAA
20701 CGTGGCCCTG TACCTGCCCG ACAACCTCAA GTTCACTCCG CGTAACATCA TTTTGCCCGA
20761 AAACCGCAAC ACCTACGCTT ACATCAACGG GCGACTGCCA CCCAGCGGCA TCGTGGACGG
20821 CTACATCGAC ATCGGCGCCC GCTGGTCTCC CGACGTCATG GACTCCGTCA ACCCGTTCAA
20881 CCACCATCGC AACACGGGTC TGCGCTATCG GTCTCAGCTG CTGGGCAACG GCCGCTACGC
20941 CGTGTTCCAC ATACAGGTGC CCCAAAAGTT TTTCGCCATT CGCAATCTGC TGCTGCTGCC
21001 GGGCACATAC ACCTACGAGT GGTCCTTCCG CAAAGACGTG AACATGGTGC TGCAAAGCAC
21061 CCTGGGAAAC GATCTGCGGG CGGACGGCGC CTCCATCAAC ATCGATAACG TGACCCTGTA
21121 CGCCAGCTTT TTTCCACTGG CACACAGCAC GGCCGCCACC CTGGAGCTCA TGCTACGTAA
21181 CGAGACCAAC GATCAGAGCT TCAACGATTA CCTCTCGGCC GCCAACATGC TGTACCCCAT
21241 CCCCGCCGGG GCCACCACGG TGCCCATCTC CATTCCGTGT CGTAACTGGG CGGGCTTCCG
21301 CGGCTGGAGC TTCGCCCGTC TAAAACAACG CGAAACGCCC TCGCTGGGAT CCCCTTTCGA
21361 CCCCTACTTT ACGTACTCAG GTACCATACC CTACCTGGAC GGCACTTTCT ACCTCAACCA
21421 CACGTTCCGT CGAGTCGCCA TACAGTTCGA CTCGTCGGTC AGCTGGCCGG GCAACGACCG
21481 CCTGCTGTCG CCCAACGAGT TCGAAATCAA GCGGTACGTG GACGGCGAGG GTTACAACAT
21541 CGGCCCGAGC AACATGACCA AGGACTGGTT CCTGGTTCAG ATGCTGGCGC ACTACAACAT
21601 CGGCTATCAG GGCTACCATC TGCCCGAAAA CTTCAAAGAC AGGCTGTATT CGTTTCTCAG
21661 AAACTTTCAG CCCCTGTGTA GACAGATACC GGATCCCTCC CATCCCAATT ATCGCAACGT
21721 TCCCTTTACC CGTCAGCACA ACTCCTCGGG TTTCACATCC CACAACCTGG CCGTGGGCGT
21781 GCCGGAAGGT CACCCCTATC CGGCCAACTG GCCGTACCCG CTAATCGGCG CTCGGGCCGT
21841 GAGAACTTTG ACCCAAAAAA AGTTTTTGGT GGACCGTACG CTGTGGCGCA TTCCCTTTTC
21901 CAGCAACTTT ATGAGTATGG GAGCGCTGAC CGATTATGGA CAAAACCTGT TATACGCCAA
21961 TTGCGGTCAC GCTTTAGACA TGACTTTTGA AGTAGACCCC CTGGACGAGG CTACCGTATT
22021 GTACGTGCTG TTTGAGGTGT TCGACGTCGT TCGCATCCAT CAACCGCATC GTGGGGTGAT
22081 CGAAGCCGTG TACCTGCGTA CTCCCTTCGC GGCCGGAAAC GCCACCACTT AGTTCCACTG

22141 GCACGATTGT GTTTTTTTTT TTGTTTTTTC GCTTTTCGCG CGTCACTTTC CACTCCGCCT
22201 CCTCCCCCCC TTCCCTAGGC GGTTTTTCCC GTCAACACCC GCGCCATGGG CTCGTGCGAA
22261 AACGAGATTA GAGCCATCGC GCGCGACCTG GGAGCCGGGC AAGTGTTCTT GGGCACGTTC
22321 GACAGCACCT TTCCTGGATT CATACGTCGG GAAAAACCGG CCTGCGCCGT GGTGAACACG
22381 GCGCGAAGAC ACACCGGCGG CGCGCACTGG GTGGCTGTGG CGTGGGAACC CGTCTCGCAC
22441 ACCTTCTACA TGTTCGATCC CTACGGGTTC GACGATCAGA CGCTCAAAAA GGTGTTTGAC
22501 TTTAGCTATC AGACTCTAAT GAAACGCAGC GCGCTCACCT CCACGGATGA CCGCTGCATC
22561 AACGTGGAAA AGTCCACCGA ATGCGTGCAG GGACCTCACA GCGCCGCTTG CGGTCTGTAC
22621 TGCTGTCTGT TCGTGTACTC GTTCGCTCGC TGGCCCCAGA GGGCCTTACG CGACAGTCCC
22681 GTCATGCAAA AGTTACAGGG AGTGCCTACT CGTCTGCTGC AGTGCCCCGC CTTTGCGCCC
22741 ACCTTTCACC GCAACCAAGA ATACCTGTAC GCCACGCTTC GCAGACTGTC TCCCTACTTT
22801 AGAAACCGCG AGGCGGCTAT CCGCGAACGC ACACGCGTGG ACAGGCTGCT GGTAGATTCT
22861 GCGTTAGACG GCGACCGTTC GAAAATCCCC GAAAAACACC CCGGCCCGCC ACGCCCTCGT
22921 TCGTAACCGC TCGACTATTC CTCGGTGTGT TTGTGATACG TGACAAATAA AAAAAAACGT
22981 CAACAGCGTG ATACATCTTT TTTACGTTTA TTGATACGAC CACCGCTTTA TTTTAAAAAA
23041 AACGTAATGC GAGAGGGAGG GAAAAAACAA ACTCTTCGTC AGAACAGGGC GTCGTCGTCG
23101 TCCTTTTGAC CCACGGGCAG CACGGTGTTC ACGTACTGGT ACCGCGGATG CCAACGGTAC
23161 TCGGGAACGT AGATGGGCGC CGGGCGACCC ATCACCTCGC TCCATAACCT CTTGGCCAGT
23221 TGGACCGCGG CGATGAGGTC GGGTGCGGAA ATCTTAAAGT CGCAGCTCTT CTGCGGGGCG
23281 GACCTGTTCT GACGATAGGT GGGATTGCAA CACTGAAAAA CCAACAGCGT GGGAAAGGTT
23341 ACGCTGGCTC GCACCTTAGG GTCGCTGACG GTACTGGGAT CCAGGTGTCC GGCGGCGGTG
23401 AGCGCAAAGG GCGTCAGCTT GCACACCTGT CGTCCCAACA CGGGACCCGA CCCCGTCTCC
23461 CAGTTGCAGT GACATTTCAG ACAGGTCAGG AGATGGGTGC CGGCGTTCGC CATTCGCGGG
23521 TAACAGGCCT TCTGAAAGGC CATGGCCTGC GCGAAACCCT GACGGGCTTT GTGCAGCTCG
23581 GTATAAAACA TGCCGCACGA CTGACTAGAA AAACTGCCCG GAGCCTGCGC GCTGTCGTGA
23641 AAACAGCAGG CCGCGTTGTC GTGACGCAAC TGCACCACGT TACGCTGCCA GCGGTTGGTG
23701 ATGATCTTGG CTCTGGTGGG GTTTTCCTTC AAAGCGCGTT GCCCGTTTTC GGAGCTGACG
23761 TCCATCTCCA CGACGTGCTC TTTAGAGATC ATGGCCAGAC CGTGCAGGCA GCGCGGACCC
23821 TCTAGACCGG CGCTGTCGTG CTCCCACGAC ACGCAACCGG TAACGCCGTG ATCTCCCTCG
23881 TCGTACTGTT CGGCCTCGCG CAGGCAAAAG TCCAACACAA AACGAGCCAT GCACGTCACC
23941 AGCGTTCCGT GCGTAGAGAA GGTTAGAGGC ACGTGAACGC GGCGATCGGT GAGGAAGTTC
24001 TGGCAGCACT TCTTAAACAC CTCCAGGGTG GCCCGATCGG CCGTCATCGT CGCTTCGCCC
24061 TCGTCTACCT TCAGCTGCTC GCACAGCGTG AAGGCCACCT CTCGGGCGCG TTGCCAGCGT
24121 TGCTGTAGCG AAGGGGGGTG CGCCGCCGTC ATGGACTGTT CCCCGTCCTC CGTCTCCTCC
24181 GTCCCTTCCT CCGTGCGTCG CGCGTCGCGG CGCGCAGCCA CGCGCGCGGG AACGTCGTCT
24241 TGGTCCGCAG CCGGCGCCTT GTAGGGCCGT TTCGGCGGCG GGGGAGACGT CGACCCCGCG
24301 TCGTCGGAAT ACCGAAGAGG TCGATCGTCG CGCATCAAAA CGGGGGAAGA ACTCAGACGG
24361 TGGACGCCGC TGCCGTAGCG CGACTGTCCG CGGACAGAGT CCGCGGTGTT GGACGCCTCG
24421 CGATCCATGG CGGGAGGGCT GTCTTTCTGC CTGTCGGTCA CCACGCCGGC CTGCGGAAAG
24481 GTAACAACGT CCCGTGAGTC CCGTTCGCCG TCCGCGGTCC GGGTCGTCGT CCACTTACGT
24541 AGCAACACCG GTGGGTTAGT GAACCCCCAC ACGCCCGTCT CACGCACCGT GGCCGTAGGG
24601 GTGGTGGCGG CGCCAGGGGT GACGGAACTG CCGGGGGTAG TGAGCGACGG GGGGTTCGAC
24661 GACGGGCGGC GCCGGTCGTC CGGGGGGAAG GAGGGAAGCG AGGTGGCCGC GACAGATGCC
24721 GCACCGCGCT GGACCCTCTT GAAGGTGTAG CGCGGGGGAG CGGGCGGGGG AGCGGCGTCT
24781 ATCGGGACGG CCGGGGCGGG AGACGAGGAC GTCTCCGTCA TGCGCAATAG GGTAGCGTCG
24841 TCCACCTCGT CTATGTATCG CGCCGCATCC GTCAACAGCG CCAGCCCCGA CGACGGCGGC
24901 GAAACACCGT CGGGCGCGGT GCTTAAAGGG GGCTGACGGC GGAGATCCCC GACGGCGACG
24961 GTACCAGTCA CACCGCCGGT CGACGCGCCG GCGTTCGCCG TCGTCGGGGA AACGCCTCGA
25021 CCCTCCGTCA CGCCGCGGCG TTCCGCGCTG CTCATTTGTT CGCAGATCTA AAACAATCCG
25081 GCGGACCGGC CACCGCGTAA ATATTTCAGT ACAATAAAAA AAATACAAAA AAAAAATAAA
25141 AAGCACCACC AAAGACTAAC TACTCGTCGC GAATCGCGCC ATGGCCACCG ACGCTTCCAC
25201 CGCTCACGGA GAAAAGGCGG CCAGTCCCGC GCCGGGCGGC GACGGGAGCG ACGGCAATTC
25261 GGAGGTGCTG TTCAAACACC TGACGCGACA AGCGCGTATC CTCAAGGACG CGCTCCGCGA

25321 GGAGCCCGCG CACCTGAAGA CGCTCAACCT ACGAGACCTC AGCCTGGCCT ACGAGCACAC
25381 CCTGTTCTCT CCTCCCGTGC CGCCTAGAAA GACGCGCGAG GGCACCTGCG CCCCGGAACC
25441 CCGTCTTAAC TTTTATCCGT GTTTTATGGT TCCCGAGAGC CTGGCCACGT ATCACATGTT
25501 TTTTCTAAAT CACCGCGTCC CACGTTCCTG CCGCTTAAAC CGCCCGACGG CCGAAACGCA
25561 GCTGCGCTGC GTCACGCGCT TACCCGAAAT CACGTCTTTG GAAGAAACGC CCTCCGTGTT
25621 CGAAGGTCTG GGCGACGAGG TCACGCCGAT CGAAGATCCC TCCCTGAGGG AAGACAATAA
25681 AGAATCGGCG TTAGTGGAGC TGGAGGGAGA CTCCCCGCGG GCCGTGACGG TGAAACGCGA
25741 TATATGCGTT ACGCACTTCG CCTATCCCGC GGTTAACCTG CCCCCTAAAG TGATGGAAAC
25801 CTTGGTAGAG TTCTTTATCA GGTCCAAAAC CCAGCCCCTG ACCTCCGAAG GACCCGCCGC
25861 TGCGGAACTG GAGGAACGCG AACCCGACGA CGCCGAACAG CGTCGCAGAG TGCGAGAGGC
25921 CCTGAGTGCG TCGTCGGAAC AGGCGGCGCC AGAGGAGAGC GTGGACGCCG TCATGAAGAG
25981 ACTCACGGCC GCCTGCCTGT GCACGGCGAC CCTGCGCTGC ATGCAGCGGT TCTTTACCGC
26041 GCCGGCTCAA ATACGCAAGC TGGGTGAGAC GCTGCACTAC ACTTTCCACC ACGGTTACGT
26101 GCGCCAGGCG TCGAAGGTGT CCGGCGTCAA CCTCAGCAAC CTGGTATCCT ATCTGGGCGT
26161 GTTGCACGAA AATCGCGTGG GCCAATGCGT GCTGCACAAT CGTCTACCCC CCGACGCGCG
26221 AGTGGAGTAC GCTCGCGACA CCGTCTACCT GTTTTTGCTG TACACGTGGC AGAACGCCAT
26281 GGGCGTGTGG CAGCAGTGTC TGGAACCCGA AAATCTGCGC GGTCTGCGCG ACTCCCTGTC
26341 CCAACACCTG GACCCCTTGG TGCGCGCCGG CGACGAGCGC GAAACGGCCG ATCTGCTGTC
26401 GCGACTGGCC TTTCCCGAAC ACCTGGTGGA ACGCCTGCAG GACGGTCTGC CCGACTTCGC
26461 GTCTCAGAGC CTGCTTCACA ACTACCGTTC CTTCGTGCTG GAACGCTCCG GCATCCTGCC
26521 GGCCATGTGT ACGGCCCTAC CGCTGGACTT CGTCCCCCTG CGCTACGAAG AGGCTCCTCC
26581 GCAGCTGTGG CCTTACGTGT ACATCGCACA GCTGGCCAAC TACCTGACGC TGCACACCGA
26641 CGTGGCGTAC GACCGCACGG GCGACGGGGT GATGGAATGT TACTGCCGCT GCAACCTGTG
26701 CACCCCCCAC CGATCCCTGG CTACCAACCC CGACCTGCTC AGCGAGTCGC TACTGGTGGG
26761 CAGCTTCGAA CTACGCGGAC CGGCGTCGAC GGAGGGCGAC GAACGCGCGA CAAACAGGGG
26821 CCTCACACTC TCGCGCGGAG CCTGGACCAG CGCCTACCTG CGCAAGTTCG AACCGGAAGA
26881 CTACTATCAC GACCGGATAC GGTTTTACGA GGACCAATCG GAGCCCACCT GTCGCCCGCT
26941 AACCGCCTGC GTCATCAATA AACCGGAAAT ACTTGCCACC CTACAACACA TCCGGCAGGC
27001 ACGCGAGCGC TTCCTGTTGG AGAAGGGACA CGGCGTCTAC CTGGACCCCG ACACCGGCGA
27061 CGTCCTCAAC CCGTCGCGAC ATGCGAGCCT CGGCGACCGC CAGCGAACCG GTACCGGCGA
27121 GACCCATCAC CCCTCCGCCA CCCTCGCCCC CGTCTTCCGA CATCCCGCTG GCCCCTCCGC
27181 CGTCGCCCGC CGCGGCGACG GCGCCGGCGC GCAAAAAAAG GGTCGCGGCC GCAAACCGCG
27241 CGCCTCTACG CGAAGAAACC CCAACGCGGA AGGCTTCGGC GGGCAAGCGG ACCAAGACGT
27301 CGGTGACGTG TCGCAAAAAC ACGGGCGGTC GGGACGCCTC GCGCACACCG GCGAAGAAAA
27361 CGTCACGGGG CGCCAAGAGG ACCGCGGCGG CCACGCCGCC GACGACAACG ACCGCGTTAC
27421 CGTCATCGTC GAACCAGATC GGGGCTCGGG ACGAGACGAA CGGTCGTCCA GCTAAAACCG
27481 CCAAACGCCG GCTGGTAACG CGACCGCCCT CGGGCGGCGA CGGGGACGGC GGATCCCGTT
27541 GGGATAAGAA AACCAAACGG GCCGAAGCGC TTCACCGTCG TCTGGTCGCT CCGCCGGCCC
27601 CCGTACCCGA ACTCCCCACC GCGGACCAGC ACCGCGACTC CCACCTGTCG GAAGTAAGCA
27661 GCGAGGAAGA AACCCCGTCG GAATTCGAAG ACGACAACGC GGAAACGCTG TCTCTGGAAA
27721 CCATCTACGA AGAGGAGGAA CGGGATTCTC ACCACCACCC AGAACGCGAC CCCCCCTACG
27781 ACTACGAAGA GAGGGAACCC GGCCCTCTGG TCATCGACGA GGCTCCCGTC GCCGCCGCCG
27841 CCCACCGTCG AAACGGTAAG CTTGACGCCG CCGCCGCCGC CGCCGCCGCC AGAACGGGGT
27901 ACCGCTCCTG GAGGAGGCTC GGAGCTTCCA TTCAACGCGC GTACCTGGCC GGAAACGGAG
27961 ACCCGGGATT CGCCCAACGG TTGCTCATGC ACCAGTTCGG CGTGACCGTG CCCCGTCGCG
28021 TAGTGCGGTA CTACGGAGCA AAAACCGAAC ACCGGGACGA TGCCGCGTCG CCGGCGTGAG
28081 ACCGGTCGCG CGTCGCGTCT CTCGCTCTCG TTCCCTCTTT GCTTTTGCCG CTACCGCTTC
28141 GTGACGCCGT TCTGTTACTT GCTTCATTCA CGGCAGATTC GGAAACCGTC GCCCGGGCGT
28201 ACCCCCGCAG CGACGACGCG TCCATCCGTC ACCTCCGGGA CTCCATCTTT CCCACCCTGT
28261 ACGCCATATT TCAACAAAGC CGAGGACAGC CCACGGAACT GAAAGTGAAG AACAGGTCCC
28321 TGCGCTCGCT TACCAAGAGC TGTCTCTATC ACAAGAACAG CTCGCAGCTG CGTCGCACCC
28381 AGCAGGACGC CGAACGCCTC TTCCAGAGGT ACTGCTCCCG TCTACTGGGC AACGCCGAAC
28441 AACACAGCGA GGACTTCTCC GCGACTACTC CGCCATGAGC AAGGAGATTC CCACTCCCTA

28501 CGTGTGGCTT TATGAGCCTC AAAGGGGAAG GGCCAGAGGC GCCTCCCAGG ATTACTCCAC
28561 CCGCGCATAT TGGATGAGCG CCGGGCCCGC GATGATTTCA CGCCTGAACG CCGTCAGGTT
28621 GGCGCGCAAC GGCCTGCTGC TGCAGCAAGC CGCGATTACG GAAACGCCCC GGGCGGTGCT
28681 TAACCCGCCG GAGTGGCCAC CCGCCCTGCT TCGCCAGCAG CCACAGCGTG TGAACATTCC
28741 CACCCAGTTT TCTCACGCTA CTTTTCCCAC GGCCATCGTG GGAAGCGGAA TGTCCCCGCC
28801 GCCGCCCAAA CGGCGCAGGC GCGGTCGCGA CGCAGACGCC CGGCGTTCGC GTCAGCGCCG
28861 CGTCTCAAAC GCTATTTATA CGACCGTTGC CGGCGCCAGA CCGCAGAGCT CCGAGGCTGC
28921 GGCAGCCGCG CTCGCGGCCG TAGATCGCTC GGCCGAACTC CTCAGGCGCG CCACCTTTCC
28981 CTCCGCGCCG CGATCCGGAG GACTCGGCGC TCAACGGTTC GTGGAGGAAT TCGTTCCCGA
29041 GGTATACTTA AATCCTTTCG CCGGGGAGCC CTCTAACTTT CCCCTCGAAT TTATTCCACA
29101 GTACGACGTG CAGACCGGTA CAGTGCACGA TTACTAACAG ATACGCGAAC GGCGCCGCTT
29161 CGGATCGCGC TCCATCAACG GATTGACGTC TTTTTTTTTT TTTTAAAAAT TTTTTTTTTT
29221 TTTGATCACA TGAACCGGTC GGACGCCTTC GGTCCCGTGG CGGACAAATG CGCGTGGCGC
29281 TCCCACTGCT TTCTTATTGC CGACCACCTA GCCCGCTGTG ACGCGGCCTG TTTTGCCCGC
29341 TCCGGCCTGC ATAAGCTAGA TTATTTTGTA AAACCGCCGC TGTCGGACGA TCTGGGCGAC
29401 GTGCCCGACA GCTACCAACC CGGCCACGGC TTCACGTTGA GCTGGAGCCC TCTGTACGAA
29461 TCCAAACGTG TGTTTCCCCC CTTTTGTAAT CCCGAAAAAT TTGACGACGA CCTCCCCTGG
29521 GCCTTCGTCA AAATCGCGCG TCCTTCGGGA CACGTGGAGG TCGTCTGCCA GTGCGACAAA
29581 CCCACCCCGC ACCCGTTTCT CTTAGATCGC CTCTGTGAGG CGTACCGCCG CGCCTTCACC
29641 CCGCACGATC CCTTTGGCAT CCGGGCCGGC GACCTGGGAC CCGCCACCCC CGTGTGAGTG
29701 ACTGCCACGT GCGGCGCCAC CACCATTTTT TTTTTTTATC TACCGTTCCG CAGGTCAGTC
29761 CGCCCGACCA TGCGTGTGCC GGCACCCCCC GGAGGCGGTC CGGCCGTCGC GGCCCTGGCG
29821 CTGTTTATCG CTCTCCTGTT ATCCTCCCTG GCTCCCGTGG TCGCGGGTAG ACCCGAAGTG
29881 GAAACCGACC ACCGTCACCG CGACCCGTCC GGTCTCGACG GCGACGGCGA CTACTTTCAA
29941 GTGTCCGACT CTTTACAAGG ACCCCCAAGT TTAATCGGCC CACCCGGTCC TCCGGGAATT
30001 CCCGGCCCCC CCGGGCCACC CGGTCCGCGA GGTACGCCCG GCCTGCAGGG TTTTAGAGGC
30061 TTTCAGGGTC CGCCGGGTCC CGTGGGACCC CCCGGACCGC CGGGATTAAA AGGCGAAGAC
30121 GGACCTCGCG GCTTCGTAGG AAACACCGGA GACGAAGGAC CTCCGGGGCC CCCCGGACCG
30181 CCGGGACCGG CCGGTCCACC CGGACCCCCC GGCAGCTGCG AATGCAGCTC CAAACACCAG
30241 TACCTCATGA ACGCGCTGCA GCCCGCCGAC GGCGAGATCA TGCGCTGGGT GCCTCAGCAC
30301 CGACCCATCG TGCAGGCCCT GAACGGTAAC CCCGTCTCGC TGACGCCCTG CGCGCACATG
30361 TTGCACGACT ACCGCAACGG GTCCGATCCC AATCACCTAA AGATAGAGTT CCAACAGCAT
30421 CACCACGGAC GCCGCCCCAC GCCACACGAC CCGGTGTACA GACTGTTTCA GGTGGCCTAC
30481 ACCGAGGACG GTTTGCTGAC TCACGAGGTG GCGTCGCCGT CCCACACCGC GTGGCATCTG
30541 CAGATGATCC CCAACCCGCA CAACCCTGGC AAACCGTCCG TGTTGCTGTA CGCCTCCTCC
30601 GTTAACGCCA AGGACGAGGG CAGCTACACC TGCATCGTCT TTTGGACGGA AGGCGATCAG
30661 GTCGTGACCA ACGCCACCAC GCTGCTCATG GTGTACACGC CTCCTCCCAC CGTGCCGGTG
30721 GAAGATCCGC CGCTGGCCCA GCGACCCTCC GAAGAAGGCC TGTCGGGGTT GGCCGTGTTC
30781 GGCCTGGTCG TCCTAGTCGT GGGTCTCCTC ATTCTGGCGT TCACCGCGTT TCGTCTGTGG
30841 CAGCGACGCA AACGCCTGAC CTTCGTCAGA CGCCGGCTGA CCGGCGGCCC GCGCGGCGGC
30901 ATCAACAATC CCGCCCCGGG CAAGGACCTC GCCAGCAACC CCCTCACCCC CTCTTCCTCC
30961 TCTTAAACGC AGTAATGAAT AAAAAAATCA ATCATCTCAT GTTAATAAAA AAAATTTTTA
31021 ACCCCACTTT GGCGGCGTAA AACGTGTGTC TCTTGTAACT ATCGTCTCTC TCTCTCTTAC
31081 CTATGCAGTT CGCGGCGCAG CGCCGCCAGG GTGCTCGCGC CGCTCGCCTC GATCCAACAC
31141 CCTTCCTCCC AACTGTCGTA CGAGATTCGT AATCGGGCGG CCGCCTTGCG AAACCGCCTA
31201 AAGGACACCC TCTCAGCTAA TCGCTTTCCG TTGTACAGGA CGTAACAGTC TTCGCCGTTC
31261 GGCGCCATCG CCGCAGTCTC GCGGACTGCC TCTCCGGTCG TCTCCCCGCC CGACGCCTTC
31321 GCCGCCATTA TGTCAAATAA CCCGCCCCCT CCCGTAAAAC GCCGCCGCCG GCGCGCGTTA
31381 CCCGACGAGC CGGAGGACGT TGTTGCCACG CCTCCAAACG TCACCGGGCC CCTCAGCTAC
31441 GACCGCCTGG TGTCCATGAT CATGCAACGC CTGTCGGCGC TGGGCGGCGG CGCGGCCTCG
31501 GCGGCCGCGG CCGCCTCGCT GGGCCCCGCC AGCGCCGGAG TGTCCACCAT CTCGCAGAGC
31561 TATCCTCCCC AGGTGGTCAA CTCTCGCATG AGCTCTTCCC TACAATCCCA AGAACCCGAA
31621 GACGAAACAG CCGTCGGCCT CACGGAAAAC GAGAATCGCG CGGTGCGCGT CCAGCCGGCG

31681 CTGCACTACG AGCCGCCGGC GTACGACCCG AACGCCGCCA AACAGGACGA CCGCGTCGTC
31741 GTCACCGCCT CCACCCTGGC TCTCGCCGAA CGCGCCCGAG ACCGCCGCGT CAACGCCATC
31801 TCTCTGGCGT ACCCCCTGAG TTACGGACTG CAGGTGCGCC CCGACGTTCG CTCCGCCGCG
31861 GAAGCCAATC CCGGCGCGCC GGTAAACCTG CAACTGGTGG CTCCGTTGTA CTCGGCGGGC
31921 GGACGCATCT CCCTGCGCTA CCAAGATCCG TTTACCCTGC GAGACGACGA CGTGTCGGTC
31981 GCCGGAGTGG ACACCCTGGC TCTCAAGGTC GCGGCACCGC TCGCCGTAGC GTCCGCCGAC
32041 GCCACGGAAC GAGGGGCGCT CGTCCTCCGA TATCAGTTTC CCCTAGAAAC CCGAGACAAC
32101 CTGCTCTCTT TACGTCTGGG CTCGCGTACA CCCCTTCTGG TGGATCCGTC CACCCACGAC
32161 TTGGTGTTTC ACGTAGGGGC CGGCCTTGCC GTAAGCGACA CCAACGCGCT CACCGTGAAC
32221 CCCAAGGCGC CCCTGACGGT CGGCGCGGGC GCCGACTCCG ACCCCCTGCA GCTGCAGTCT
32281 TCCCCGCCTG TCGTGGTGGG AAGCGACGGA CACCTCACGT TGTCTCTGGG CGCCGCCTCC
32341 AAGCTGGACG ATCAGGGCAG GCTCACGGCC CGCGTGTCAC CCCCGCTCGC GGTCGACGAA
32401 ACCGCCCCCG ACAACTCCCT ACAGTTGCAG CTGGGCGCGT CGGGCGCTCT GACGTTAGAC
32461 GCCGCGAACG CGTTGTCCGT CGAGGTCGCC GATCCGCTTT TTGTCCAAAA CCCCGGTAAC
32521 GCCCTGGCTC TGCGCACGCA AGACCCGCTG CACGTAGACC AGGGACGCCT GACCCTGCGC
32581 CAGGGAGACG GTCTGCAGGT TCAGGGAAAC ACACTGACGC TCCGTCTGTC CAACCCCCTG
32641 CGCATCTCCA ACGGAGCGCT GAGTCTGCAC ATAGACTCGG GACGCGCCCT GTCCGTCAGC
32701 AACGACGGTC GCGTCGGGGT CAACGCCGAC GGCGTGCTGG CGTTCGGAGA CGGACGCCTA
32761 CGCCTGCGCC ACCTGCATCC CCTGGTAAAC TCGCAGGGCG CCTTGTCCCT CAGCACGGCC
32821 CCTCACCTAC AGATAGCCAA CGGCTCTCTG GCGCTGCGCA TGGCGGCGCC CGTGGGAATC
32881 AACTCGCAAA ACGAGGTTAC CCTTAACCTG GCATCCCCCT TCGCCGTGAG CAGCGGCGGG
32941 GCCCTATCCA TCACGGCGGG TCACGGCCTC GCCGTAGAAG GGGCGGGACT CACCCTGAAA
33001 CTGCAGGCGC AGGGACCGTT ACAACTCAAC GATCAGGGAC TGAGCTTAAA CGCGGCTGAC
33061 CCGTTCTCGG TCAACGATCA GAAACAGCTG GTCCTGCAGA CCGACGACTC GTACCTGTAC
33121 GTCAACCCCG CATCGCACCG TCTCACCATG CGACCCATAG GGTTTCAGCA CAGCTCGCAC
33181 GTCAAAATAC GGGGCATGAC CGTGTTCTTT AGGCTCCGCC GTCTGGGCAT CATGCCGGTA
33241 CACATGGTGC TGCGCGCCGG CTCCGACCAG GCGCACCCGG TGACCAAGGA CAGCGCGGGC
33301 GTCGGAGACA CCTTCGAGCT GCTCATCTCC ATGGGCAAGG GCGGTACCGA CATCCGCAGC
33361 AACACCTTTA CCCTGACCGA GATGACCACC TCGGGCGCGG TGGACCCAAA TCGTCTGCGA
33421 TACTGCCTGC CCCACAGCGG CGTGTACCCC TACGCCGCCG ACCGTCTGCA CACCGACGTG
33481 AAGGGTACTA GTTTCTTGCT CAGCGACGGC AGCATCATCC CCTGGTACGT CAAATACAAC
33541 CAAGAGTCGC CCGACGTCAC CAGCCTGAAG ATTGTATTCG ACACCTCCAG TCTGGCCTAC
33601 GACCAACGAC TAGACTTCGA GCCCATCGAG ACCACCTACG AAGGCAACGA GTGGAGCGCC
33661 GTGGAACCCC CACCCATACG TTAAACACGC CACAACGCGC AATAAAGGTG CGCAGCGTCG
33721 TCGGCTTGTA TTTTTTTTA TTGCTCAGCC ATTCCTCACA GCAAATAATC GCGGGACCGG
33781 GCGGGAGGGC ACACGTAATC GTCCATAGCA AAACACTGTC CGTAGCGCGC CAGACGATGC
33841 AGCATCTTCA ACCTGGCCGA CAGGTACACG CACCTCTTAA ACCGTTTCTT TAACCTTCCG
33901 TCGCGGGTGC GGGCGTGGGG ACCGCGCCGC AGCAGGGCCC GCGCCTTCTT GATCAACAGA
33961 CGAGTACGTC GACAGCACAC GCGAGCAGAC ATCTCATCGC AGCGCGCGCA GGACAGACAC
34021 TGCAACACCA CGAAGTTACT CACGCAGGCT CGCGTGGTGT AACTGAATCC GAAACTCATG
34081 CCTCGCGTGC AGCGGCGAAT GTCGCGCGTC TCGGGAAAGT ACAGGTACAC CAGATGCTGA
34141 TCGCGCAGCC ACACGCTGCC GATGTAGTAC ACCGCGGCGC AGCTCTGCGC GTTGGCCCTG
34201 AAACGGTAAA ACGCGTACAC GCGATCGTAG GCCGTGCCCA GCAACATCAC GCGAAAAAAG
34261 GAACACACCA GACGCGCCTG GCTCTGGCAT CGCAACGTGC AAGGCGAGCG CAGGGACGCG
34321 TTAGCGTTGG CGCTGTCGCA GTGACAGTGC ACCGTCCAGC GCTCAAACCC GTGCACCAGG
34381 TGCGAGCGGC GACACTCCAC CAGACCGGGC ATGGGACAGG TGTATACAAA GGTCTCGAAC
34441 ACGCGTCGTT CGTAATCCGT CAACAGGGCC GGCCAGGGAA CCAACAGGCA ACCGTGAAGG
34501 GCGAAGGAGG CCGCCGTCGG CACGCCGCGA ATCTCCGCTA CGATGTGCTC CCGATTGGCT
34561 CCGGGCGGCG CCAGGCCGAA CACGCTGTGC GTCTCGCCAA ACACGTCCAT CGTCTCGCCC
34621 TGCAAAACCC GCAACACGTA GACACGTCCA CTCGCATACA ACCTCGCTCA CGTGCACGCC
34681 ATCTGTCTCA CGCTTTACAA TTACAAAAAA AACAAAATCC TACCTTACCA CCCTCACGCC
34741 TCTTTGGAAC AGCAAGCGGC GACTCGGTCC GTCTTGTCTC GCTCGGTGTA ACTGCCGCAG
34801 ATCAGACAGC GCCCACCCCC CGTCCACCGC AAAGCGCACT CCCTAGACCA CCCAATACAA

34861 TCGGCAGTGG GCAGCGACCC GCACCAAAAG CAACGCGTGA CCCTCCCACC CAGCAGAGCT
34921 TTAGCCTCGT TCTGCAAGCG CTCGAAGCTG TTCCTCACAA ATAACACACG CCGCAGGCAC
34981 TCGGAACTGT TACGGGCCTT GCCGTGCTTG TGTTTGCGCA TGAAACAGTA GCACCAGTGA
35041 AAGCACTTCA GCACCATCCA AGCTTGATAA CGTAACTTGG ACGGCGCCCC GCAGTAAAAA
35101 TCGTCCCACT GGTCACAGCA CGCCTCCCAC AGGTTGCAAT TCCAGCAGGC GTAGTAATGC
35161 ACGTCTCCGT CCGCCTTCAG TGGGGGGATC TTCACGAACT GCATCACACA GCCCCTCCTC
35221 CTAAGCACAA ACAGATACCC GCACTAAAAA AATTTTTCCC CACCCACACC CACACGCCAT
35281 ATAACACCAC CAACACCAGT TTGGAATCTC AAATATACAA CACTTTATTT GCACAGCAAA
35341 AAAAAAAAAA CACACGACAC ATTTTTTTTT CTCAAAACAG GTTGGGAGAC CGACACACGT
35401 CCACAGCCAC CCCTGGACCC CGTCGCCTCC AATAGTGCTC GACGGTATCC CGCAACTCGG
35461 TCAGCAGATC TTGAGGATCT TGCACGTCGT AACGCATGCC CCCCACCGTC AGGTCGCACA
35521 CCACACGCCA ACGACTCTCG CGTCGCTGGC TCATGCTGAC GAACCCCAGA TCCATGGCCC
35581 GCCGGTGCTC CGGCAGATGC AACCGCAGGC GTCTCGCCGA GCCCCTCTCG CCCAGGGCCG
35641 CGTCCAGCAA GCTGCGGTAA CGGCGACGGG CCCGAGAACT CAAACCGGGA CCCTCCATGC
35701 GCATGTGATG CTCGAACACC AACACGTTGC GCACCTGGTC CAACATGCCC CGCACCAGCT
35761 TACCAAGGAT GCTCTCCTTG ACCGCGGTAC CGTCCAGAGT GATCTGGAAC ACCAGAGACT
35821 GAAACAGGGG AACCATCTGA TCCACCCAAT CCCGCTCCGG CTCGGGCTCC GGATGGGCCT
35881 CGGACGCCGG TTCGGGCTCG GGAGTACGCC CTTCCTGCCC TTCCATTTCC CCTCCCGCCT
35941 CGGGGCGAAA CACCACCTCC TCATCCATCC CTTCGTCCCC CCGCGTCACG GCGCCGCCGC
36001 CGCCGCCGCC GCCGCCGCGG TCCTCGTTCT CATCCTCCTC CACCTCCTCG TCGTCATCCG
36061 TGGTGGTGTC GTCGATGTTC GTCGCGTCGC TGTCCTCGCT CTCGGACCTC GCCGCCGCCG
36121 CCGTGGCCGC CACCGACGCC GGACGCCCCA CTCGGTCGTC GGGTTCCAAA AAGTCCGTGT
36181 CCGTAGAAGA CCCTCCATCC TCACCGTCAT CCTCCCACCT CAACCTTCTA GCCACGGGAT
36241 CTCTCGGACG AGGTAAGGAA TCGCCCTCCG CCCTTTCCAC CCGACACACC ACACTCTCGT
36301 GCACCAGCAT CATGCACAGG GGGGAGCCCA CCCTGATGTG ACAGCGATGG GCGCCCCAGT
36361 TGCTGACGGT TACACGCACG CGATGACGCA TGCCTCCCGG AACCAAGCCG CTGACCACAC
36421 AGCTGTTTAA CCCGCTCACG CTGCCGGGCA AGGCCCCGAG GGTACCAAAA TAACCCTCCG
36481 ACAGCAAACC GTACACCCCC GTGTCCACCA CCAGGGTGTC CCCGGCATAC ACCACCACGC
36541 TCTCGCTGCT GTAGAACACC ACGGAGGCCA CCCCCAAACC GAAAGGTTCC GGACTGCGAG
36601 CGCCCGTTCC CAGCTGCTCC ACGCTAAAAC TGCACAGGTC CATGACTGCG CTCTTTTTTT
36661 TTTTTCTCAC CTAAAAAAAA AACACACACT CAGCGAGGTG CCCACTCCAA ACGGCTCCGC
36721 GCCGCCCCCC GCCCGGCTCC CCACCGGCAC TACACCTGAA CCCCGTCGCC AACGCGCCGC
36781 CCGGGTCGAA CTAAAACCGA AACCGCGCCG GCGGTGCTCT TATACTGCTC ACCCGGAACT
36841 CACGCCTCAA TAAATCCCAA AAACGAAAGC GACCGGCAGT TCCGCTTTCC CACGGTAAAA
36901 AAGTCCCCGA GCACGTTCCC ACGGCCGCTT CCGGGAGCGC GCCGCCCCCC CCGCGCTCCC
36961 ACGCTGTCAC TCAAAATGGC TCCCACGCTC CGCCATCTTG GGACACACTC AAAATGGCAC
37021 CCGGAAGTCA AACCCGGAAG TCAAAATGCA CTGTATATAT ATTGATGATG

7.4 List of figures

7.5 List of tables

8 DECLARATION

I hereby declare, on oath, that I have written the present dissertation by my own and have not used other than acknowledged resources and aids.

Regensburg, 23st October 2018

_____

Julia Weigl