SYNTHETIC BIOLOGY IN *MYCOPLASMA PNEUMONIAE*

Bernhard Paetzold

TESI DOCTORAL UPF / ANY 2013

DIRECTOR DE LA TESI

Dr. Luis Serrano Pubull & Dra. Maria Lluch-Senar

DEPARTAMENT:

EMBL/CRG RESEARCH UNIT IN SYSTEMS BIOLOGY

To my parents and everyone who read this from start to end

“Science is an edged tool, with which men play like children, and cut their own finger”
-Sir Arthur Eddington

Acknowledgements

I would like to express my very great appreciation to my Supervisors Luis Serrano and Maria Lluch-Senar for giving me the opportunity to conduct such an exciting research project, their time, their patience and the constant useful feedback.

I also would like to thank Carlo Carolis and Tony Ferrar who helped at certain points in thesis when there were too many colonies to screen or constructs to clone and Jae Seong Yang for calculating p-values of the secretome data. For good advices, time for discussion and endless patience I would like to thank in alphabetical order: Almer, Besray, Christina, Erik, Kiana, Marc, Marie, Martin, Peter, Raik, Tobias and Veronica.
A big thank you also goes to Silvia and Miriam for an endless supply of petM14 vector and many other favours, as well as Sira, Hannah and Violetta for never losing their patience when I asked in the last minute for the most complicated orders and Eva for sharing with me many constructs out of her treasure box, the freezer.

My grateful thanks are also extended to the members of my Thesis committee Mark Isalan, Jaume Pinyol and Juan Valcarcel Juarez

Special thanks also go to Javier and Andreu for saving the world with me at Lunch and all the others that I forgot for their feedback and advise on countless occasions.
Finally I wish to thank my girlfriend Veronika and my parents for their support and encouragement throughout my study.

Abstract

M. pneumoniae is one of the smallest self-replicating organisms. Many organism wide studies have been performed on *M. pneumoniae* and provide a wealth of information on the bacteria. However, the genetic tools to manipulate and engineer this microorganism are currently insufficient. Therefore one focus of this project is to extend the genetic toolbox for *M. pneumoniae*. In parallel, we developed first a proof of concepts study for the use of *M. pneumoniae* as therapeutical vector. For this purpose the secretome of *M. pneumoniae* has been investigated to define all secretion signals in this bacterium. This knowledge was used to modify *M. pneumoniae* to secrete three proteins with therapeutical applications. Further we could show for two of the proteins that they are active after secretion and therefore that *M. pneumoniae* can be used to engineer therapeutic vectors for treating lungs diseases.

Resumen

M. pneumoniae es uno de los microrganismos auto-replicativos más pequeños descritos hasta el momento. En nuestro grupo se ha empleado como modelo en Biología de Sistemas y se ha caracterizado a diferentes niveles (genoma, transcriptoma, proteoma...etc). Sin embargo, las herramientas genéticas para manipular y emplear este microorganismo como modelo en Bilogía Sintética, son insuficientes. Por lo tanto, uno de los objetivos de este proyecto es ampliar el repertorio de herramientas moleculares de *M. pneumoniae*. En paralelo, hemos desarrollado por primera vez una "prueba de concepto" para el uso de *M. pneumoniae* con finalidades terapéuticas. Con este propósito, primero hemos determinado el secretoma de *M. pneumoniae* que ha permitido identificar todas las señales de secreción de esta bacteria. Posteriormente, se ha aplicado este conocimiento para obtener cepas de *M. pneumoniae* capaces de secretar tres proteínas con aplicaciones terapéuticas. Hemos demostrado que dos de las tres proteínas son activas tras la secreción, abriendo nuevas perspectivas en el uso de *M. pneumoniae* para el tratamiento de enfermedades pulmonares.

Preface

In the last years technological advances in DNA synthesis and DNA assembly have allowed scientists to redefine the possibilities of Biotechnology. While in the past it was cumbersome to obtain, clone and change a gene, nowadays it can be synthesized in a matter of days.

Therefore, biotechnologists can design *in silico* complete metabolic pathways and build logical circuits based on genes. However, while complete bacterial genomes have been synthesized and assembled, this technological revolution has produced very few successful applications.

Part of this problem is that we do not completely understand the cells in which we implement the genetic circuits. This problem can be solved either by trial and error experiments or by modeling a complete bacterium. We follow the vision that using a simpler organism as chassis, all interactions can be understood and taken into consideration during the design. However, using this simpler organism brings many other drawbacks, like the slow growth or the lack of genetic tools. In this thesis I expanded the knowledge about genetic tools in the minimal organism *M. pneumoniae* and provided a first proof of concept for its use in medical applications.

Table of contents

Publications during the Thesis

Bertero MG, Verschueren E, Paetzold B, Serrrano L (2010) Enhanced SnapShot: Macromolecular machines. Cell 143: 652, 652.e1. doi:10.1016/j.cell.2010.10.035.

Maier T, Marcos J, Wodke JAH, Paetzold B, Liebeke M, et al. (2013) Large-scale metabolome analysis and quantitative integration with genomics and proteomics data in Mycoplasma pneumoniae. Mol BioSyst. Available: http://pubs.rsc.org/en/content/articlelanding/2013/mb/c3mb70113a. Accessed 19 April 2013.

Paetzold B, Carolis C, Ferrar T., Serrano L., Lluch M. (2013) In situ overlap and sequence synthesis during DNA assembly. ACS synthetic biology (under review)

List of figures

List of tables

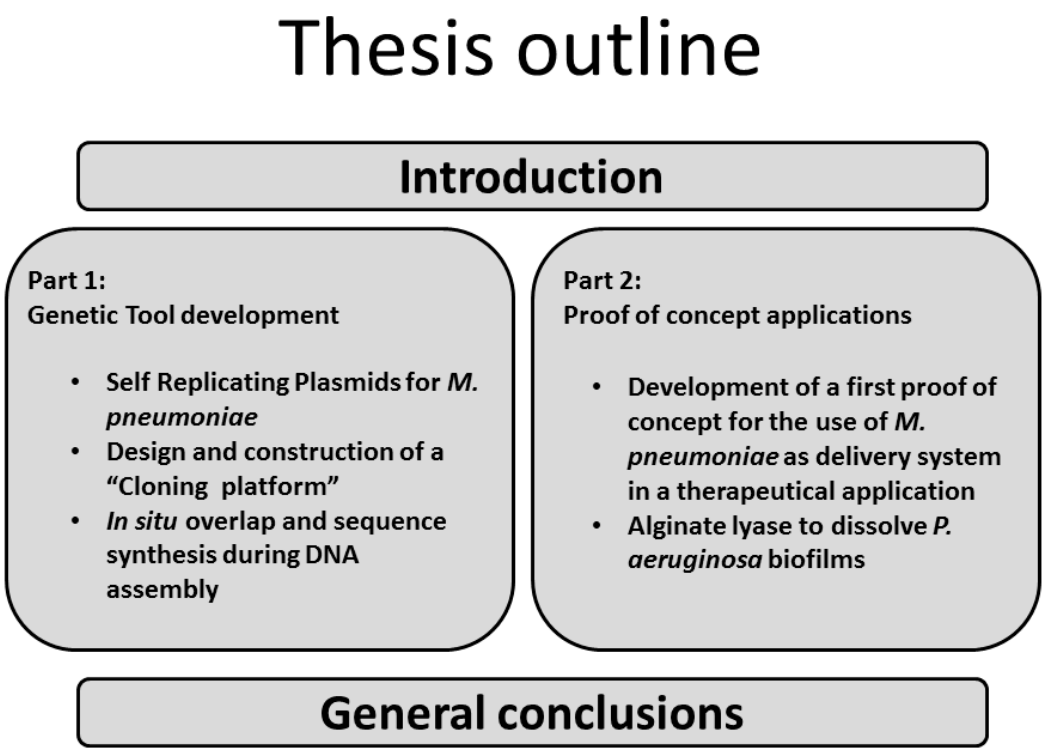


Figure 1 Outline of the Thesis structure

Review of State of the art

Synthetic Biology

Synthetic biology is a new research field that emerged several years ago. Its appearance is based on big advances in key enabling technologies like DNA synthesis and molecular cloning (Gibson et al., 2008, 2009; Hoover and Lubkowski, 2002; Quan et al., 2011; Shao et al., 2009). These new techniques made it possible to clone complete metabolic pathways or signaling cascades from one organism to another. These possibilities allowed scientists to rethink the applications in biotechnology. Biotechnology was up till then, limited to the recombinant expression and purification of a protein of interest, or transferring a certain trait to a new host. In contrast, synthetic biology is about the engineering of complex systems, which either can produce a certain metabolite as a product or act like electronic circuits making decision based on a

variable input (Andrianantoandro et al., 2006). However, even to experts in the field the precise definition of synthetic biology remains elusive (Pei et al., 2012).

One of the first and most famous example of complex engineering in synthetic biology is the production of artemisinic acid in yeast (Paddon et al., 2013; Ro et al., 2006). Artemisinic acid is the precursor for the antimalarial drug artemisin. The precursor artemisinic acid is normally isolated from the plant *Artemisia annua* L. This plant is slow growing and the supply varies greatly over the years as its price. A total chemical synthesis of artemisinin is difficult and expensive. An international consortium supported by the Bill and Melinda gates foundation managed to clone the complete metabolic pathway for the production of artemisinic acid into yeast. This reduced the production cost of the precursor and enabled its flexible production according to short term market demands.

Since the start of the artemisinin project the possibilities of cloning, as well as data analysis have evolved rapidly. However, the main bottlenecks of metabolic engineering are still the same today:

- Incompletely annotated pathways in public databases like KEGG (Kanehisa and Goto, 2000).
- Poorly characterized and unpredictable individual genetic parts.
- Interference of the engineered network with the host metabolic or genetic network.

Whilst the two first drawbacks will be solved by the time, the problems arising from unwanted interference of the engineered network with the host are more difficult to overcome. The later problem can be approached either by combinatorial screening of large libraries or by the use of less complex host organisms (Figure 2).

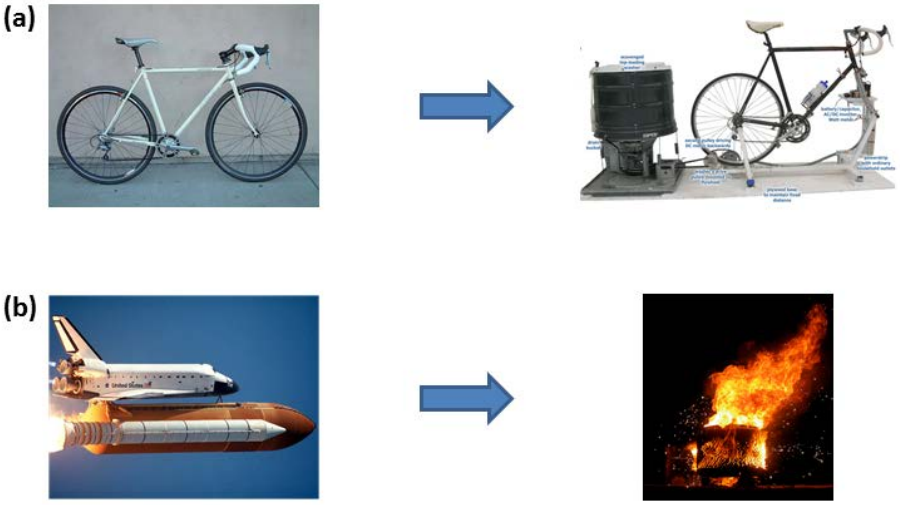


Figure 2 Schematic representation to illustrate the rationale behind choosing a "simple" over a "complex" host chassis. (a) Simple host network (bicycle) can be easily analyzed and a synthetic circuit (washing machine) can be successfully integrated. (b) A Complex host network (Rocket) is very powerful and has many advantages. However, it is difficult to connect a new synthetic network to it due to unwanted interaction of the genuine host network and the engineered circuit.

Smart drugs

Synthetic biology is not limited to constructing metabolic pathways but also includes engineering cells like electronic devices (Figure 3). For this purpose genetic circuits are designed *in sillico* and cloned into cells. Genetic circuits represent a combination of genes and their products (mRNA, proteins etc.) which interact with each other to perform basic logic functions. These genetic circuits work like electronic circuits and process a certain input signal to create an output from it. The input can be an environmental condition or a chemical substance. The genetic circuit would process this input and produce as output another signal, a protein or small molecule. An engineered bacterium or virus with a genetic circuit integrated, which responds specifically to a pathogen, would be a living pill or smart drug. A smart drug only releases or produces an active compound when encountering a pathogen or a specific signal indicating the target of its action (Gupta et al., 2013; Saeidi et al., 2011; Wu et al., 2013).

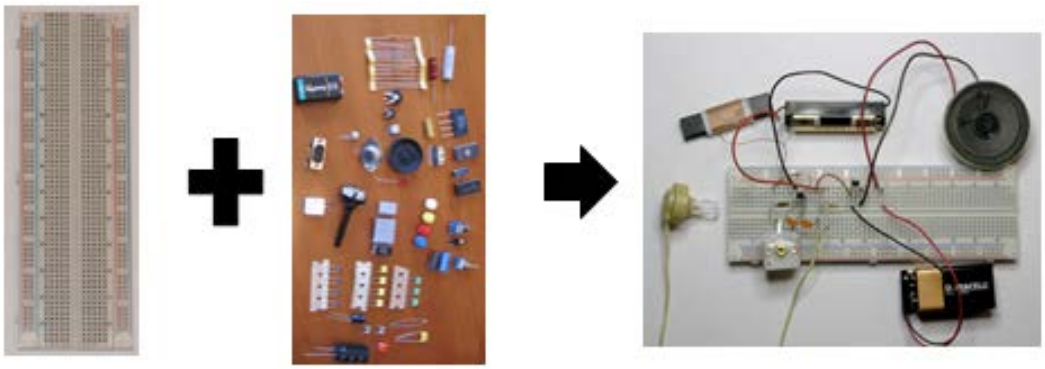


Figure 3 Scheme showing the combinations of different electronic parts on a breadboard to build a simple radio. Synthetic biology aims to achieve the same with a biological chassis, usually an organism, as breadboard and gene, promoters, etc. as parts.

The idea of using living systems to treat human diseases is not novel and many different attempts have been made (Caragata and Walker, 2012; Phillips, 1997; Wittebole et al., 2013). Much effort went into the use of viruses as therapeutic vectors (Thiel, 2004; Waehler et al., 2007). Viruses as therapeutic vectors provide benefits regarding: delivery, cell targeting or proliferation control. However, the amount of DNA that can be packaged in a virus capsid is limited and viral gene delivery always comprises the risk of uncontrolled gene insertion into the host cell.

In contrast, bacteria as therapeutic vectors comprise the intrinsic risk of uncontrolled proliferation or the triggering of an immune response. But they also provide a number of advantages:

- Complete metabolic networks can be integrated, which are able to synthesize complex molecules.
- Complex detector modules can be designed to sense the concentration of biomarkers and respond accordingly (Wu et al., 2013).
- The risk of integration in the host genome is extremely low.
- In most cases antibiotics are sufficient as kill switch to stop uncontrolled bacterial proliferation.

The perfect chassis for the engineering of a living pill would be a bacterium with these properties:

- Easy and cheap to culture.
- Parasite but non-pathogenic.
- No lypolischaride envelope (LPS) nor cell wall, reducing the risk of an immune response.
- Easy genetic manipulation.
- Susceptible to antibiotics for the easy eradication after the end of the treatment.

Equipped with the correct genetic circuit such an organism could be used as a living vector. Without integrating any genetic material in host the cell would sense the environment and would respond to changes in the environment by secreting specific components.

Mammalian cell engineering

A number of laboratories work on engineering mammalian cells for therapeutic applications. From the application point of view the use of mammalian cells as smart drugs has many advantages. They can produce complex and glycosylated proteins. In personalized engineering approaches an immune response could be completely avoided. Cells could be programmed to differentiate according to certain environmental stimulus and therefore provide a bigger variety of functions than for example bacteria or viruses. They could also be engineered to assemble in a 3D context and form tissue networks (Miller et al., 2012; Purnick and Weiss, 2009; You et al., 2004). However, the drawbacks and problems arising from engineering mammalian cells are equally big than the benefits:

- The proliferation of engineered mammalian cells needs to be tightly controlled as they might otherwise become a cancer.
- A simple drug like antibiotics to eradicate uncontrolled proliferating cells is not available.
- The genetic engineering of mammalian cells is intrinsically difficult and slow.

Therefore the most successful applications of engineering mammalian cells today are aiming to better understand pathogenesis or to screen medical compounds (Weber and Fussenegger, 2012).

State of the art in "living pills"

Currently there are two main methodologies for engineering living organisms in synthetic biology applications. One of them is to use *Escherichia coli* as chassis that is easy to engineer but problematic in many later applications. The other approach is to take a bacterium as chassis that is suitable for the desired application but difficult to modify.

In the first approach usually the engineering of a complex genetic circuit stands in the foreground, while a later application might fail because the chassis can ultimately not be adapted for the environmental prerequisites of the application.
The second approach tries to focus on the application and needs arising from it. Usually in these organisms no high throughput techniques are established and the genetic tools are underdeveloped. Therefore, the engineering is kept to the basics and usually following the KISS rule of engineering "Keep it simple, stupid".

A living pill against *Pseudomonas aeruginosa*

First attempts to engineer a living pill with an underlying genetic circuit were made by Gupta et al., (2013) and Saeidi et al., (2011). Both projects engineered an *E. coli* bacterium that can sense and destroy *Pseudomonas aeruginosa*, an opportunistic human pathogen mostly found in the lung. Unfortunately, the choice of chassis limits the applications of these living pills to the gastro intestinal tract, which is not the main niche of *P. aeruginosa*.
Both groups designed an *E. coli* strain that senses the secretion of N-Acyl homoserine lactone (AHL) produced by *P. aeruginosa.* To sense the pathogen they used the protein LasR from *P. aeruginosa* which acts as transcriptional activator in the presence of AHL. As response to the presence of *P. aeruginosa* both systems produce a toxin of the colicin class. To release the toxin into the extracellular space the two groups designed different solutions. The design of Saeidi et al., (2011) used a lysis module that ultimately destroys the cell membrane of the engineered cell and thereby releases the previously produced toxin. In contrast, Gupta et al., (2013) used a more elegant approach of secreting the toxin using the cellular secretion machinery (Figure 4).

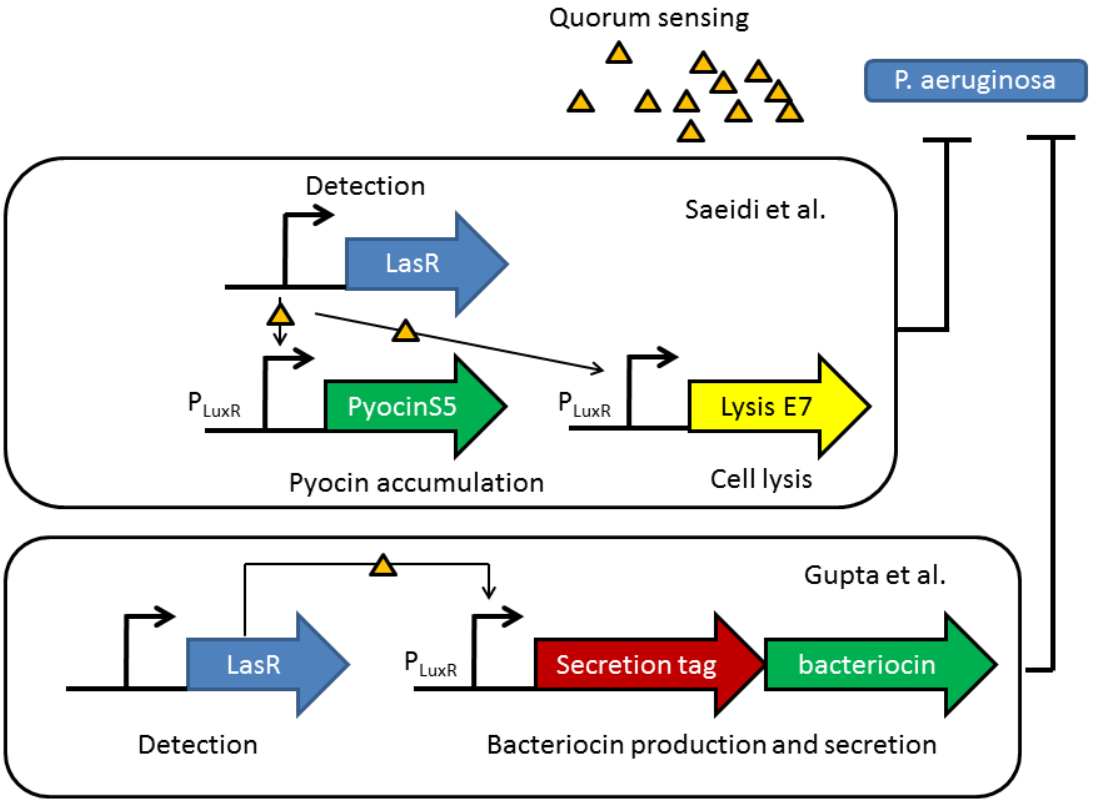


Figure 4 Scheme of the genetic circuits used by Saeidi et al. and Gupta et al. Both used the transcriptional factor LasR that binds to a specific N-Acyl homoserine lactone (AHL) produced by *P. aeruginosa*, which is involved in quorum sensing. LasR in conjunction with AHL activates transcription from the P_{LuxR} promoter. In the system of Saeidi et al. Pyocin S5 is produced and accumulated in the cell. At the same time a protein for cellular lysis is produced which will lyse the *E. coli* after reaching a certain threshold. In the system of Gupta et al. AHL induces the production of a bacteriocin similar to Pyocin S5. It is fused to a secretion tag and exported.

The use of *E. coli* as chassis allowed both groups to design a complex sense and response layout. However, it blocks the application of this living pill in the lung where *P. aeruginosa* infections are a big unmet medical need. The engineered genetic circuit cannot be easily moved to a more suitable chassis bacterium as in the best case the fine tuning of the dose response will be lost or in the worst case the genetic circuit will not function at all. While the genetic code is compatible between many species, regulatory elements like promoters, ribosome binding sites, codon usage etc. are more limited and change their characteristics when moved between species. Therefore fine-tuned circuits are usually not transferable between organisms. Even if the problem of incompatible promoters is solved usually the fine-tuned balance between parts is destroyed as soon as the genes are put in the different context of another organism. Noteworthy, the fine tuning is usually the most time consuming part in current genetic engineering projects. In the case of the genetic circuit designed by (Saeidi et al., 2011) the fine tuning took around 3 years (personal communication). This prohibits the strategy to assemble and

optimize a genetic circuit in *E. coli* and then transfer it to a chassis organism suitable for the desired application.

Recently, the group of Saeidi et al., (2011) published an improved version of their design (Hwang et al., 2013). Basically everything except the detector module was exchanged and the design resembles significantly the one of Gupta et al., (2013). Additionally a chemotaxis module was added allowing the engineered cells to migrate towards the target microbe and another module was added to secrete DNaseI to disperse biofilms. However, the underlying problem that the chassis *E. coli* is only can only live in a location, where *P. aeruginosa* infections play a minor role was not solved. Priority was given to engineering and all cloning was done in Top10 *E. coli* cells.

ActoBiotics

ActoBiotics are genetically engineered *Lactococcus lactis*, a bacteria developed by the Belgium Company Actogenix. Because of a unique combination of chassis and application these bacteria are up to now the only commercially developed living pill. The technology of ActoBiotics is based on engineering of *L. lactis*, a bacterium which normally lives in the gut of humans. The company is testing three engineered bacteria in clinical trials. The bacterial strain AG013 is in a phase 2/3 trial, it secretes human Trefoil Factor against ulcerative oral mucositis. AG014 and AG015 are both in a phase 1 clinical trial and target inflammatory bowel disease. AG014 produces an anti-TNF antibody and Ag015 the anti-inflammatory cytokine IL-27. All the bacterial strains are engineered to continuously secrete the therapeutic protein in the oral or digestion tract (Van Huynegem et al., 2009; Steidler et al., 2009).

These applications in conjunction with the chosen chassis have the great advantage that no immune reaction has to be circumvented, as the bacterium is naturally colonizing the oral and digestion tract. Also complex genetic engineering of a sense and response circuit was unnecessary since the targeted disease required the constant secretion of the therapeutic protein.

The same effect could also be achieved by orally administrating to the patient a purified, therapeutic protein. But the *in situ* production of the therapeutic protein reduces the required amounts of protein and avoids expensive purification steps. This significantly reduces the production cost and therefore the therapeutic application.

The project

Overview

Our main aim is to engineer a *Mycoplamsa pneumoniae* for applications in the health sector. Therefore, we decided to focus on unmet medical needs in the biological niche of our bacterium, the lung.

The chassis: *M. pneumoniae*

M. pneumoniae is one of the smallest self-replicating organisms of the Mollicutes ('soft skin') class. These cell wall lacking bacteria evolved from more conventional progenitors in the Firmicutes taxon by a process of massive genome reduction (Weisburg et al., 1989). Mycoplasmas are obligate parasites that live in relatively unchanging niches requiring little adaptive capability (Moran, 2002). As originally annotated *M. pneumoniae* has the potential to express 688 gene products(Dandekar et al., 2000), although a recent reannotation in our group comibining mass spectroscopy and deep sequencing has increased this number to 727. Its reduced genome size is thought to be linked to its normally parasitic lifestyle, lack of a peptidoglycan wall and a limited capacity for amino-acid and cofactor biosynthesis (Himmelreich et al., 1996). Knowledge of its biochemistry and molecular biology indicates the lack of a TCA cycle, as well as many energy yielding systems, mainly relying on glycolysis to synthesize ATP (Wodke et al., 2013; Yus et al., 2009). Other characteristics are the use of the stop codon UGA as a Tryptophan codon and an obligate dependence upon the availability of cholesterol, a basic component of their plasmatic membrane. The bacterium possesses a very slow growth rate in culture with a division time of approximately 8 hours at its best.

M. pneumoniae and Synthetic Biology

M. pneumoniae has several features that make it interesting to use as a chassis in Synthetic Biology:

- It has a "simple" metabolic and genetic network, which reduces the risk of unwanted interference of the engineered network.
- It has no cell wall and extensive studies on the factors involved in pathogenicity, as well as on essential genes, facilitate the design of a non-pathogenic strain.
- It has a natural low rate of homologous recombination reducing the risk of uncontrolled reversion of the engineered organism to wild type.
- It uses the codon UGA to encode for tryptophan, while most other organisms use UGA as a stop codon. Reducing the risk of uncontrolled spread of the engineered parts to other organisms.

The intrinsic genetic and metabolic network of *M. pneumoniae* is comparably well understood (Güell et al., 2009; Karr et al., 2012; Kühner et al., 2009; Maier et al., 2011, 2013; Wodke et al., 2013; Yus et al., 2009). In theory this will allow one day to design genetic networks that will work ad hoc or at least with less trial and error experiments than usually necessary in *E. coli*, as there is less unwanted interference with the host genetic network. The group of Covert et al. made already a whole cell computational model of *Mycoplasma genitalium* which has smaller genome than *M. pneumoniae* (Karr et al., 2012). Based on this model theoretical assumptions about the impact and behavior of synthetic gene networks can be made (Purcell et al., 2013). While the current existing model is a patchwork of many other smaller models and the predictions and implications for engineering are rather limited, it is reasonable to assume that in the near future the "*in silico*" modeling will help to efficiently design and implement genetic circuits. Efforts towards a complete model of *M. pneumoniae* are currently made.

Despite *M. pneumoniae* being a gram positive pathogen, it lacks a peptidoglycan wall and the associated response of the immune system. It causes atypical pneumoniae and is usually characterized by mild symptoms (Dugdale, 2013). Furthermore the main virulence factors are relative well understood (Ansarin et al., 2011; Hallamaa et al., 2008; Hansen et al., 1981; Johnson et al., 2011; Kannan and Baseman, 2006; Kannan et al., 2011; Krause et al., 1982; Shimizu et al., 2011). It has already been shown that certain mutants are unable to cause a sustained infection and cause only a reduced

immune response (Hansen et al., 1981; Krause et al., 1982; Shimizu et al., 2011). Based on this knowledge it is possible to engineer a non-pathogenic strain that could be used later as a chassis for therapeutic applications in the lung.

However, a critical evaluation of an engineered non-pathogenic strain will be necessary. A study exists connecting the co-culturing of mammalian cells and the Mycoplasma strains *M. genitalium* and *M. hyorhinis*, to malignant changes in infected cells (Namiki et al., 2009). For *M. hyorhinis* a potential gene involved in these changes has been identified (Liu et al., 2007). While no data exists for *M. pneumoniae*, one report links the malignant properties of Mycoplasma to an inflammation reaction (Jiang et al., 2008). Therefore an engineered strain which avoids an immune response should not cause malignant changes. However this is currently speculation and these indications make it necessary to thoroughly evaluate *M. pneumoniae* for any carcinogenic effect before it could be used in real life a therapeutic application.

The role of *M. pneumoniae* in diseases like asthma is disputed and not clear (Ansarin et al., 2011; Waites and Talkington, 2004; Wang et al., 2012b). Some reports also indicate a role of *M. pneumoniae* in encephalitis (Bitnun and Richardson, 2010). However, the evidence for the involvement of *M. pneumoniae* in any central nervous system disease is only based on sporadic case reports and a clear mechanism is missing. It is therefore reasonable to assume that a non-pathogenic strain that has lost its ability to enter cells would not cause any of those rare symptoms upon infection. In the case of unexpected complications it should always be possible to clear a host from *M. pneumoniae* by a standard antibiotic treatment, as well as by engineering a Trojan horse safecty mechanism.

An interesting finding with dual use was recently reported. The *M .pneumoniae* strain M129 has no functional arginine deiminase pathway as the involved enzymes are truncated or missing (Barile et al., 1966). However Rechnitzer et al., (2013) showed that once *M. pneumoniae* M129 is complemented with a functional arginine deiminase pathway it becomes toxic to mammalian A549 cells. This toxicity can be a potential problem in therapeutic applications requiring long term persistence of *M. pneumoniae* in the host. It has to be ensured that the engineered strain can not obtain a functional version of this pathway by horizontal gene transfer. However, if rigorously controlled this toxicity could also serve as a mechanism to attack cancer cells in a potential anticancer seek and destroy applications.

Genetic use restriction technology (GURT) for *M. pneumoniae*

M. pneumoniae shows a very low rate of recombination (Krishnakumar et al., 2010) and horizontal gene transfer (Pascal Sirand-Pugnet, 2007). This feature is potentially based in a mutation in the gene RecU (Sluijter et al., 2010). The low recombination rate is an inconvenience during the engineering process as incorporation of exogenous genetic parts by homologous recombination is not feasible. However, this property could be advantageous once an engineered strain is obtained since reversions to a wild type genetic background by recombination should occur at a very low rate.

M. pneumoniae uses UGA codon for tryptophan instead of stop, complicating the recombinant expression of mycoplasma proteins in other bacteria. This can be seen as a natural Genetic use restriction technology (GURT) or bioconfinement technology (Sang et al., 2013). While this feature hinders an inter species transfer of functional parts it is of course not hindering a transfer in between different Mycoplasma species or other species that use the UGA as a tryptophan codon.

However, while providing many theoretical advantages as a chassis in synthetic biology there are also numerous drawbacks associated with the use of *M. pneumoniae*. The main drawbacks are its slow growth rate and the underdeveloped genetic tools. The lacking genetic tools make the genetic engineering in *M. pneumoniae*, a very work intensive and time consuming process. While changing the growth rate of *M. pneumoniae* is beyond the scope of a single thesis, we tried to engineer genetic tools to ease and speed up the engineering workflow.

Genetic tools

State of the Art Transposon Tn4001

Currently the only useful genetic tool available in *M. pneumoniae* is the mini-transposon mini-Tn4001. Transposons are natural mobile genetic elements, which change their position in a genome. The basic unit of a transposon is the insertion sequences (IS). Bacterial IS elements are between 768 bp to 1426 bp in size and flanked by short inverted repeats on each end of the element. In between the inverted repeats is usually a single open reading frame coding for a transposase. In the natural context two

IS elements, which are close to each other, can form a composite transposon. This composite transposons, carry a DNA sequence unrelated to the transposon in between the IS elements. These composite transposons usually carry drug resistance gene and can be transferred in between different bacteria. Additionally to the direct transfer of resistance genes IS sites can indirectly influence the phenotype of a host. IS elements can indirectly trigger biofilm formation in Staphylococcus epidermidis (Conlon et al., 2004) and indirectly influence antibiotic resistance by gene inactivation (McEvoy et al., 2013) or the formation of hybrid promoters as it was shown for *Staphylococcus aureus* (Maki and Murakami, 1997).

After the discovery of transposons, they were improved by genetic engineering to become useful tools for molecular biology (Hahn et al., 1999; Knudtson and Minion, 1993; de Lorenzo et al., 1990; Pour-El et al., 2002). A big step towards the use of transposons as genetic tools was the design of so called mini-transposons (Huisman et al., 1987; de Lorenzo et al., 1990; Pour-El et al., 2002). Mini-transposons have in contrast to composite transposons only one IS element and carry between the inverted repeats the genetic cargo for delivery (Figure 5). Once in the host either the cargo or the IS element gets integrated in the genome. If the cargo is integrated in the genome the transposase is lost and the cargo stays stably integrated in the genome. Most transposon vectors are limited to a set of species and exhibit different integration characteristics in the individual species. In *M. pneumoniae* only the transposon Tn916 and Tn4001 work (Hedreyda et al., 1993). Of the two available transposons for Mycoplasma species only Tn4001 was accessible by engineering and a set of variants have been cloned (Algire et al., 2009; Knudtson and Minion, 1993; Lluch-Senar et al., 2007; Pich et al., 2006). These variants of the mini-Tn4001 have been widely used genetic tools for either essentiality studies (Dybvig et al., 2010; Glass et al., 2006), functional studies (Lluch-Senar et al., 2007) or genetic manipulation of Mycoplasmas (Halbedel and Stülke, 2007; Schmidl et al., 2007).

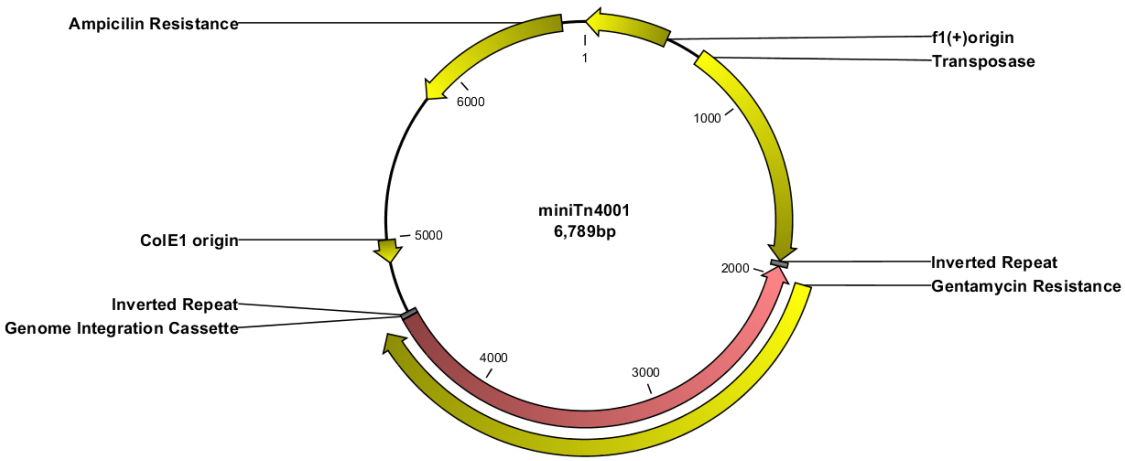


Figure 5 Vector map of a mini-transposon vector mini Tn4001. It is a suicide vector that carries a gentamycin resistance cassette which will be integrated in the host genome after transformation.

While other mini-transposon based vectors are used as reliable tools for engineering and some are even discussed for gene therapy (Izsvák et al., 2009; Nikel and de Lorenzo, 2013). The mini Tn4001 vectors used in *M. pneumoniae* and its basic unit the IS256 contains a number of drawbacks associated with its use as genetic tool for precise engineering:

- The complicated and not yet completely understood integration mechanism of the mini-Tn4001 bears risks for the integrity of the cargo and stability of the insertion (Loessner et al., 2002; Prudhomme et al., 2002a).
- Transposase independent reversal of the genomic integration (Hennig and Ziebuhr, 2008)
- Potential promoter formation in inverted repeat junctions if formed upon circularization (Prudhomme et al., 2002a).
- Random integration in the genome complicates standardized and reproducible characterization of genetic elements.

Most of the studies investigating the side effects of IS256 have been conducted in other species than *M. pneumoniae.* Therefore, until experimentally validated it remains an open question, if these phenomena also in occur when using a mini Tn4001 in *M. pneumoniae*. These drawbacks can be neglected for gene disruption studies or the overexpression of a single gene in *M. pneumoniae*. But they propose a burden for the precise engineering as it is required for synthetic biology.

Apart from delivering genetic parts or circuits to the *M. pneumoniae* genome, these parts need to be transcribed and ultimately translated into proteins. It is possible to use promoters from the host to control transcription. However, these promoters are part of a

larger genetic network and might exhibit surprising dynamics. To circumvent this problem one can introduce a host independent transcription system (Herrero et al., 1993). This idea is inspired by bacterial phages which bring upon infection their own transcriptional toolset with them. Usually a phage polymerase is under control of a host promoter and all subsequent genes of the phage are expressed by the phage polymerase. The T7 phage is the best studied and exploited system in this context. The potential of putting genes for overexpression under the control of a T7 Polymerase promoter was early recognized (Studier and Moffatt, 1986). A problem was that even small amounts of expressed T7 polymerase flooded the *E. coli* cells with mRNA and impaired normal growth. This was solved by putting the T7 polymerase under an inducible promoter and counteracting the leaky expression of T7 Polymerase by its natural inhibitor T7 Lysozyme (Studier, 1991).

To create a system for inducible expression many choices exists in *E. coli*. One class is the transcriptional activators and the other is the transcriptional repressors. The transcriptional activators are unbound to DNA in the inactivated state. Usually upon addition of a small molecule the transcriptional activators bind to the DNA and recruit a host polymerase to drive expression of the gene under control of the transcriptional activators. One example of a transcriptional activator is the arabinose-inducible promoter PBAD in *E. coli* (Guzman et al., 1995). In contrast to the activators the transcriptional repressors are in general bound to the DNA promoter in the absence of their inducer. By binding to the DNA they block access of the Polymerase to the promoter. Many transcriptional repressors work as multimers, providing cooperativity and DNA looping to increase repression efficiency (Narang, 2007). Examples for this class of repressors are: the Tet repressor, the LacI repressor and the cI857 repressor. Transcriptional activators for inducible expression require that the host polymerase interacts with the activator. Therefore, these systems are usually limited to the host in which they were discovered. In contrast, transcriptional repressors can be moved in between species as the bound promoter can be adapted to the host species.

Also inspired by bacteriophages many systems have been developed for gene delivery by recombination. One has to distinguish between site specific recombination and homologous recombination. Site specific recombination depends on the presence of combination sites on a donor and a recipient molecule. The most known system using site specific recombination is the Cre-Lox system. The Cre-Lox system and the very

similar FLP-FRT system have been extensively used in mammalian cell and mouse studies. It consists out of four recombination sites usually two on the donor and two on the recipient. Upon expression of a recombinase the content between the two sites of the recipient is exchanged with the content of the donor or deleted depending on the setup. The recombination systems have been extensively improved since its discovery and multiple insertions at high efficiency are possible (Bode et al., 2000a). In contrast, the RED recombination system is independent of recombination sites and uses homologous recombination to deliver or knock out stretches of DNA (Sharan et al., 2009). It is currently commercially available under the name gene bridges and found wide application in the manipulation of the *E. coli* genome or the modification of bacterial artificial chromosomes which are difficult to modify *in vitro*. While the system provides more flexibility it also is more complex as it requires 3 proteins to work and depends on a complex set of plasmids to control unwanted homologous recombination.

Chronic Obstructive Pulmonary Disease

Chronic Obstructive Pulmonary Disease (COPD) is an inflammatory disease of the lung affecting millions of people and leading progressively to a respiratory deficiency and finally death (Barnes, 2004). This lung failure is the consequence of a progressive destruction of lung epithelial cells due to a chronic imbalance between the level of neutrophil elastase and its inhibitor alpha-1-antitrypsin (A1AT). Alpha-1 antitrypsin is a plasmatic protein mainly produced by the liver, which in normal condition regulates the activity of neutrophil elastase. Neutrophil elastase is naturally induced after infection or exposition to lung irritants. Nevertheless, continuous exposure to irritants or genetic deficiency in the serpin gene coding for A1AT leads to an inactivation of the protective role of A1AT against inflammation processes occurring in the lung. In the current state of the art intravenous injections of purified A1AT from human plasma are used to help people suffering of COPD but the cost and efficacy of this treatment is matter of debate (Hay and Robin, 1991; Petrache et al., 2009; Sandhaus et al., 2008).

Based on the success of the actobiotics from Actogenix we want to use a synthetic biology approach to address COPD. We propose to use an attenuated form of a natural parasite of the lung, to secrete AA1T at the epithelium surface in response to the level of proteases, a direct consequence of an inflammatory response.

Pseudomonas aeruginosa biofilms

Bacterial biofilms provide a large unmet medical need in the modern world. Biofilms can protect bacteria from antibiotic treatment or the immune system, causing difficult to treat chronic infections. *P. aeruginosa* can appear in a non mucoid (free living) or mucoid (biofilm) state. Once *P. aeruginosa* bacteria infect the lung of a cystic fibrosis patient it forms persistent biofilms. These biofilms in the lung of cystic fibrosis are characterized by an overproduction of alginate (Govan and Deretic, 1996). Thus in principle alginate lyase could be used to dissolve biofilms. In fact, there are multiple encouraging reports in the literature on the use of alginate lyase to dissolve *P. aeruginosa* biofilms in the lung (Alipour et al., 2009; Alkawash et al., 2006; Lamppa et al., 2011), but also controversy on its beneficial effect (Christensen et al., 2001; Lamppa and Griswold, 2013).

We wanted to use the ability of *M. pneumoniae* to colonize the lung and combine it with a mechanism that allows it to dissolve the *P. aeruginosa* biofilm. We are aware that it is doubtful whether a cystic fibrosis patient should ever be infected with bacteria to fight a bacterial infection. However we choose this as a proof of concept as the therapeutic protein in this application was of bacterial origin and therefore easy to secrete. The designs of (Gupta et al., 2013; Saeidi et al., 2011) do not target the mucoid state of P.aeruginosa and were not yet available at the time when we planned the project. The latest publication of Hwang et al., (2013) claims also an antibiofilm effect by incorporating DNaseI in the secretion system. However rhDNaseI is already provided to CF patients as standard treatment and is obviously not enough to solve the problems arising from *P. aeruginosa* biofilms.

Results

Development of self-replicating plasmids for *M. pneumoniae*

Abstract

Mycoplasma pneumoniae is a model organism for systems biology. The wealth of knowledge available on *M. pneumoniae* makes it an ideal chassis for applications in synthetic biology. For sophisticated engineering precise genetic tools are needed that allow to clone and evaluate single genes in a fast and reproducible manner. Unfortunately, the molecular biology tools in *M. pneumoniae* are poorly developed and currently limited to a transposon system.

We followed work by Cordova et al., (2002) to design self-replicating plasmids in *M. pneumoniae*. We first identified the genomic origin of replication region by bioinformatics tools available on the internet. We divided the identified region, into three overlapping fragments of approximately equal size and cloned them into a vector backbone. We further determined the ability of each vector to transfer antibiotic resistance to *M. pneumoniae* cells and to express a fluorescent marker protein.

All plasmids have low transformation efficiency. Two out of the three fragments tested were able to replicate in Mycoplasma and we could isolate extrachromosomal plasmidic DNA from transformed *M. pneumoniae* cultures after multiple passages. The overlapping region between the two fragments could also transfer antibiotic resistance to *M. pneumoniae*. However, the plasmid carrying the shortest fragment showed a significantly reduced fluorescence signal compared to the other plasmids. The longer fragments showed impaired growth compared to a culture transformed with a mini-transposon.

We further identified problems associated with gentamycin as selection marker for low efficiency transformations and created a new reporter system for *M. pneumoniae*. This system allows a faster and more reliable readout of transformations than the one previous state of the art.

Introduction

Mycoplasma pneumoniae is one of the smallest self-replicating organisms and therefore well suited for system biology studies. The genetic tools for *M. pneumoniae* are currently restricted to a mini-transposon based on the transposon Tn4001, that randomly integrates in the genome (Algire et al., 2009; Halbedel and Stülke, 2007; Hedreyda et al., 1993). For many advanced experiments additional genetic tools are necessary. The mini-transposon mini-Tn4001 consists out of one insertion Sequence (IS256) element of the composite transposon tn4001, an *E. coli* origin of replication, a selection marker and a variable part depending on the application. Variants of the mini-Tn4001 were successfully used as genetic tools for either essentiality studies (Dybvig et al., 2010; Glass et al., 2006), functional studies (Lluch-Senar et al., 2007) or genetic manipulation of Mycoplasmas (Halbedel and Stülke, 2007; Schmidl et al., 2007).

While mini-Tn4001 was a useful genetic tool for these studies, it has intrinsic limitations and drawbacks. With only the mini-Tn4001 vector available, it is difficult to make rescue experiments. In these experiments, a gene gets knocked out in the genome and the resulting mutant strain gets rescued by complementing it with the same gene delivered on a plasmid. In the current state of the art this can only be done by transforming the bacteria with a second transposon vector. However, this transposon vector will integrate again randomly in the genome and most likely disrupting another gene, which potentially can influence the final phenotype. The IS256 element, the basic unit of the mini-Tn4001 vector, has been shown to have a natural role in phenotypic variation of Staphylococcus strains (Conlon et al., 2004; McEvoy et al., 2013).

In synthetic biology genetic parts or systems need to be tested in a standardized and highly reproducible way. A potential influence of the disrupted gene or the location of the insertion on the expression of the cargo cannot be excluded when using a mini-transposon. Also the integration mechanism of mini-Tn4001 is not yet fully understood and there are reports about potential side products of mini-Tn4001 (Hennig and Ziebuhr, 2008; Loessner et al., 2002; Prudhomme et al., 2002b). It is worth noting that these observations were reported for different bacteria than *M. pneumoniae* and it is not clear if these observations also apply to *M. pneumoniae*.

One solution to overcome these limitations could be the use of extrachromosomal self-replicating plasmids. While for some species like *E. coli*, a variety of self-replicating plasmids exists, there is none for *M. pneumoniae*. A general strategy on how to develop

a self-replicating plasmid for Mollicutes was described in (Cordova et al., 2002). This strategy is based on identifying the origin of replication in the genome by a GC-skew diagram and locating conserved DnaA binding sites. Then this region is cloned into a vector and tested for self-replication. When applying this strategy to *M. pneumoniae* we encountered the problem that the DnaA binding sites are not conserved in *M. pneumoniae*. We used bioinformatics and wet lab experiments to overcome this problem and construct a set of self-replicating plasmids that could be used for studies in *M. pneumoniae.*

Results and discussion

We wanted to design a set of self-replicating plasmids, allowing us to test in a fast and standardized way individual genetic parts for the engineering of *M. pneumonia.* We followed the strategy previously described by Cordova et al., (2002).The main performance criteria were reproducibility and the time to read out. As benchmark we used a mini-transposon vector based on IS256. A detailed explanation of the bioinformatics analysis, as well as a comparison between the origin of replication from 4 different Mollicutes including *M. pneumoniae* can be found in Cordova et al., (2002). Briefly, the origin of replication of the bacterial chromosome is defined by bioinformatics tools and then it is tested on a vector for the ability to transfer antibiotic resistance and extrachromosomal replication. We used the DoriC database to confirm the origin of replication as it combines several standard analysis methods in one output (Gao and Zhang, 2007). Unfortunately, we could not use the standard DnaA binding boxes defined in *E. coli* and B subtilis as a guideline to identify all parts necessary for the origin of replication as done in previous studies (Cordova et al., 2002; Maglennon et al., 2013). Therefore, we choose stretch of about 6Kbp that contained the putative Ori and the DnaA gene and split it up in three smaller overlapping pieces and cloned these fragments in a vector backbone to test their ability to self-replicate in *M. pneumoniae* (Figure 6).

We also designed a new vector-reporter backbone to reduce the readout times of our experiments. The readout of previous studies was the ability to grow in antibiotic media and subsequent southern blotting (Cordova et al., 2002; Maglennon et al., 2013). While providing a wealth of information, the southern blotting technique as a basic readout is slow and work intensive. Further the gentamycin antibiotic resistance alone proved

unreliable as indicator for a successful transformation, as spontaneous background growth occurs in selective media (Hedreyda et al., 1993). The background is generally low enough to permit the use of gentamycin as selectable marker in conjunction with high efficiency transposon transformations. However, it is a problem for low efficiency transformations and analysis of growth impaired strains. We dismissed the idea of using tetracycline as a selective marker, which gives no background, because it is still used in clinical settings as antibiotic. Instead we added to the gentamycin selection marker a functional fluorescent protein as an additional read out and later a vector with puromycin resistance was obtained. The new vector backbone consisted out of the genomic region comprising, the putative origin of replication, a reporter cassette expressing a dsRed fluorescent protein fused to the *M. pneumoniae* ORF pmp200 described by Zimmerman and Herrmann, (2005), an *E. coli* origin of replication for assembly and propagation in *E. coli* and a selection cassette for either resistance to gentamycin or in later constructs puromycin (Algire et al., 2009). A vector map for one of the gentamycin constructs is shown in Supplementary figure 1 and for the puromycin constructs in Supplementary figure 3. As a positive control (S2) we designed the same vector with an IS156 element instead of the *M. pneumoniae* genomic regions (Supplementary figure 2).

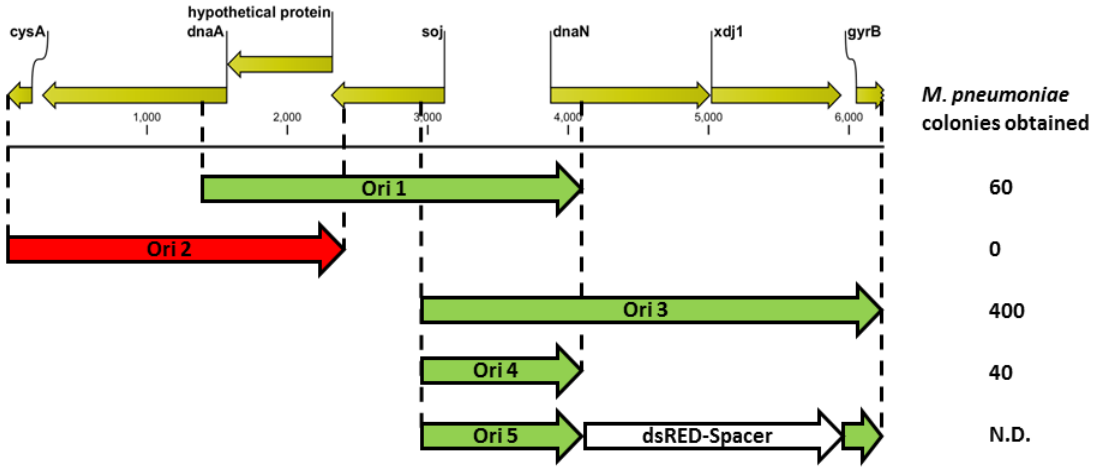


Figure 6 Schematic view of the *M. pneumoniae* chromosomal origin of replication and the cloned fragments. The fragments which transferred antibiotic resistance to the transformed *M. pneumoniae* strain are shown in green. In red is shown the fragment that did not transfer antibiotic resistance. The white arrow indicates a spacer between the two pieces of the *M. pneumoniae* genome. The spacer was a gene for the fluorescent protein dsREd which all constructs carried. At the right hand side is indicated the number of *M. pneumoniae* colonies which were obtained after transformation. All colonies obtained had a circular shape and but lacked the typical Mycoplasma "eye" structure.

The new vector allowed monitoring the cultures not only by visual inspection for growth but also to monitor individual cells by fluorescence microscopy in cover slide chamber dishes. While in the first week of a plasmid transformation usually little growth is visible by eye in the standard 25cm^2 culture dishes. We could observe individual cells under the microscope usually after 3-4 days. In the first days all cultures transformed with plasmids carrying Ori 1, Ori 2 or Ori 3 showed clear defined fluorescent bacteria (Figure 7b). After six days of incubation the cultures containing constructs with Ori 2 lose their morphology and their fluorescence became faint (Data not shown). For these construct we did not observe macroscopic growth above the background from spontaneous mutants,indicating that non replicating plasmids are initially transcribed before the DNA is lost.

In contrast, cultures carrying Ori 1 and Ori 3 keep their shape and fluorescence after day and can be grown to larger cultures. We then compared individual cells from the self-replicating plasmids, to cells that were transformed with the S2 transposon based vector. We observed that the self-replicating plasmid carrying cultures are more heterogeneous in the expression of the red fluorescence protein than the cultures having the vector backbone integrated in the genome by transposition (Figure 7). This could be due to a higher degree of background growth or a stochastic loss of the plasmid during cell division.

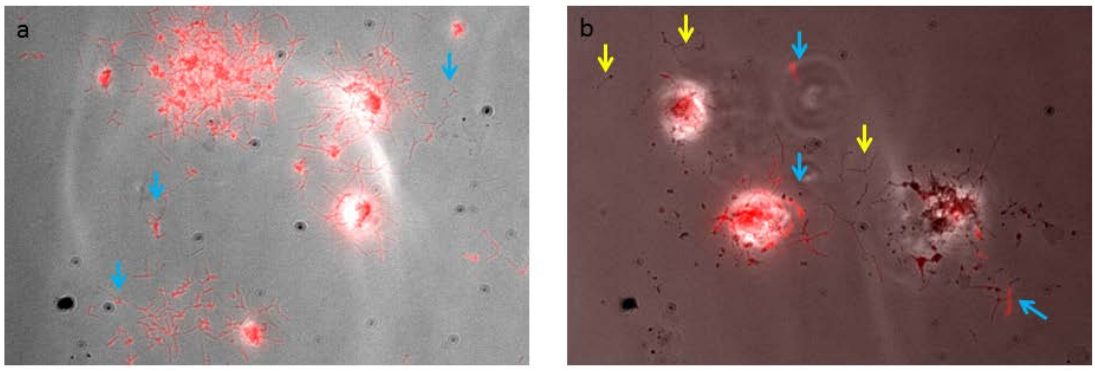


Figure 7 Comparison of one culture transformed by a transposon vector S2 (a) and one culture transformed with a self-replicating plasmid carrying Ori 3 (b) at day 6. While the cells transformed with the transposon vector in (a) are all uniformly fluorescent (highlighted by a blue arrow), many non-fluorescent cells are visible in (b) (highlighted by a yellow arrow). This indicates that the plasmid are not equally distributed in between the cells or that a significant amount of spontaneous resistant background growth is contaminating the culture.

With the new vector system it was possible to reliably and fast answer if a plasmid was self-replicating or not. However, we observed significant background growth in gentamycin selective media if cultures were transformed with water instead of DNA. We assumed that is also affecting our plasmid cultures and a part of our culture was

originating from spontaneous resistant cells. To characterize pure cultures of our plasmid a further reduction of background growth was necessary. Therefore, we moved to the antibiotic puromycin as selective agent which was published as highly efficient selective agent in *M. pneumoniae* which is not used in clinical treatments (Algire et al., 2009). It should be noted that the use of puromycin as selectable marker in *E. coli* is not reliable and a lot of background is usually obtained (personal communication Merryman, C. and own experiments). Therefore it is necessary to complement any puromycin vector with an additional selection marker for *E. coli*.

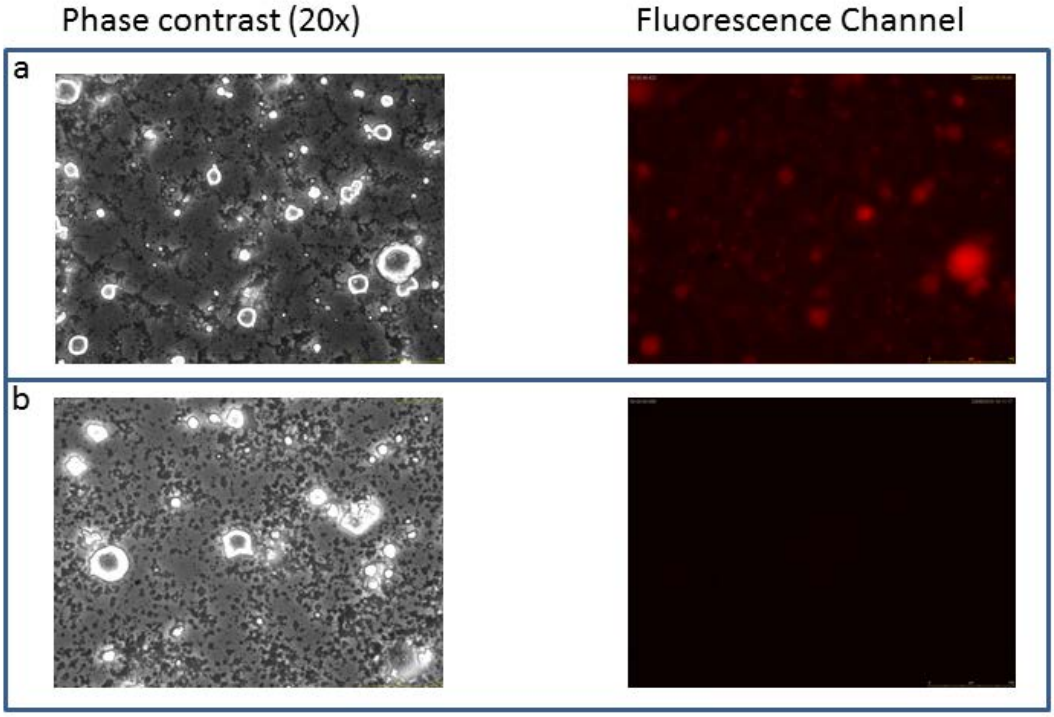


Figure 8 All panels show microscope pictures of 25cm^2 culture dish at 20x magnification. The fluorescence image was taken with a leica N2.1 Filter cube and at 2s of exposure. The images were taken from cultures grown in full modified Hayflick media containing 200µg of gentamycin. The cultures were started from a 1:100 dilution after a standard transformation of *M. pneumoniae*. In (a) is shown a culture transformed with a transposon expressing a dsRed fluorescent protein, while the culture shown in (b) was only transformed with water instead of DNA. The culture of (a) was incubated for 8 days while the culture in (b) was incubated for 4 weeks. While growth is clearly visible in both cultures no fluorescence is visible in the culture resulting from spontaneous background growth.

To distinguish whether the growth we observed was exclusively caused by homologous recombination or if the plasmids were maintained extrachromosomal, we tried to isolate free plasmids from *M. pneumoniae* cultures. The plasmid preparations were done from transformed *M. pneumoniae* cells after several passages. We tested vectors carrying Ori 1, Ori 3, Ori 4, the S2 transposon vector and a growing spontaneous resistant negative control which had been originally transformed with water. None of the eluates from the

Minipreps showed detectable DNA pattern after digestion. Therefore, we used the eluate to transform Invitrogen Top10 chemically competent cells. No *E. coli* colonies were obtained from the eluate of the negative control culture (spontaneous resistant), nor from the Ori 4 culture. On the other hand, we found colonies in the transformation done with the Miniprep from the Ori 1 and Ori 3 cells, confirming the presence of extrachromosomal plasmids. However, this result does not fully exclude the possible integration of some of our vector into genome. To our surprise we also found colonies in transformation of the eluate from the cultures originally transformed with the S2 transposon vector

The extrachromosomal plasmids isolated from *M. pneumoniae* and transformed in *E. coli* were intact and showed the same restriction pattern than the original plasmids used to transform *M. pneumoniae* (Figure 9). In contrast, the extrachromosomal plasmids isolated from *M. pneumoniae* cells transformed with a S2 mini-transposon and later isolated from *E.coli*, showed a different restriction pattern than the original vector. Sanger sequencing revealed that these plasmids had lost part of the IS element. A vector map of the mini-transposon (S2), indicating the deletions is shown in Supplementary figure 2. The shortened transposon vector might be explained by the complex way how

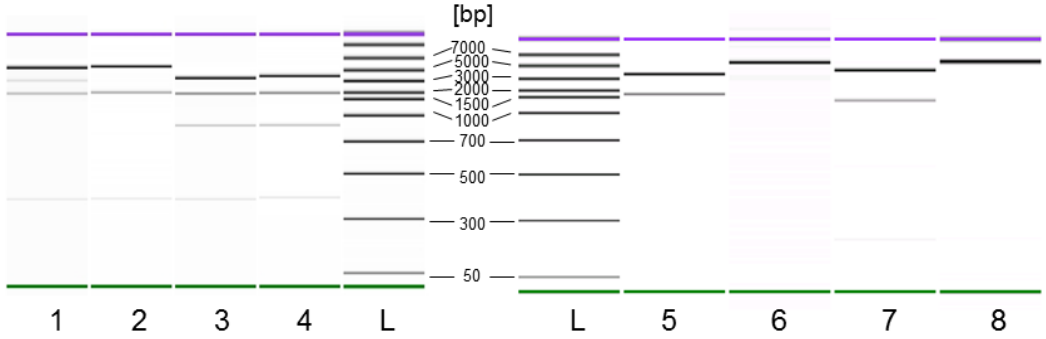


Figure 9 Capillary electrophoresis showing the restriction pattern of plasmids reisolated from *M. pneumoniae* cultures. We obtained *E. coli* colonies after transformation with eluates from minipreps of *M. pneumoniae* cultures transformed originally with Ori 1, Ori 3 or a mini-transposon (S2). These colonies were scaled up and the plasmidic DNA was purified and digested. As control also the original plasmid used for transformation was digested. The left panel shows the plasmid and control for the self-replicating plasmids. All samples were digested by the enzyme PstI and matched the expected pattern. Lanes: [1] Control plasmid Ori 3, [2] plasmid reisolated from *M. pneumoniae* culture carrying Ori 3, [3] control plasmid Ori 1, [4] plasmid reisolated from *M. pneumoniae* culture carrying Ori 1. The results show that the plasmids are extrachromosomal maintained as they could be recovered from *M. pneumoniae* cultures even after multiple passages. The right panel shows the plasmids isolated from a culture originally transformed by the S2 transposon vector. Lanes [5] and [6] show a digest by *Xho*I and lanes [7] and [8] show a digest by *Apa*I. Lane [5] shows intact mini-transposon S2 cut by *Xho*I, [6] Plasmid isolated from a Mycoplasma culture previously transformed with S2 digested by *Xho*I, [7] Intact S2 plasmid cut by *Apa*I, [8] Plasmid isolated from a Mycoplasma culture that was transformed with S2 digested by *Apa*I. The results indicate that the DNA reisolated from a *M. pneumoniae* culture transformed with a mini-transposon vector did not match the S2 transposon originally used to transform *M. pneumoniae*. Two reisolated plasmids, a shorter and a longer version, were completely analysed by Sanger sequencing. The results revealed that the isolated constructs are truncated version of the S2 transposon vector. The truncated regions are indicated in Supplementary figure 2.

the transposon vectors based on Tn4001 integrate in the genome (Prudhomme et al., 2002b). The isolation of circularized IS extrachromosomal DNA was already described earlier for TN4001 in *Staphylococcus epidermidis* and *Staphylococcus aureus* (Loessner et al., 2002). It would be interesting to investigate if this is limited to the TN4001 transposon family or if similar truncations occur with other transposon vectors like PiggyBac or sleeping beauty" that are currently discussed for human gene therapy.

Unusual colony morphologies were observed from *M. pneum*oniae cells transformed with the different plasmids in comparison to WT (eye shape morphology). We then investigated whether this unusual colony morphology reflects changes in the growth rate of the transformed cultures. For this we compared the growth rate of culture cultures that were transformed with different self-replicating plasmids carrying the puromycin resistance and a culture transformed with a transposon. The plasmid carrying cultures were significantly slower in growth (Figure 10a). The culture carrying Ori 3 grew faster and more efficiently than the culture carrying Ori 1, which is in agreement with the higher number of colonies obtained after transformation for Ori 3 (Figure 6). We believe that the slower growth of the cultures transformed with Ori 1 and Ori 3 originates from an uncontrolled separation of plasmids upon cell division. The cellular machinery controlling chromosome segregation ensures that each daughter cell receives one chromosome after cell division. We think that while our plasmids replicate at the same rate as the genome, the signals to be actively segregated on mother and daughter cell are missing on the plasmids. Therefore after each cell division the cells can have each one plasmid or one cell has no plasmid and the other cell has two plasmids.

We were surprised to see that the cells transformed with Ori 4 and Ori 5 grow faster than the other plasmid. When we transformed cells with equal amounts of DNA for the different constructs the cells carrying Ori 4 and Ori 5 took the longest time to recover before the biomass for a first split was reached. When we analysed the cultures under the microscope to assess the expression of red fluorescent protein, we observed bright fluorescence for the cultures carrying plasmids with Ori 1 and Ori 3 as well as in the transposon transformed cultures. However the culture carrying plasmids with Ori 4 had a significantly dimmer fluorescence and Ori 5 showed no fluorescence at all (Figure 10).

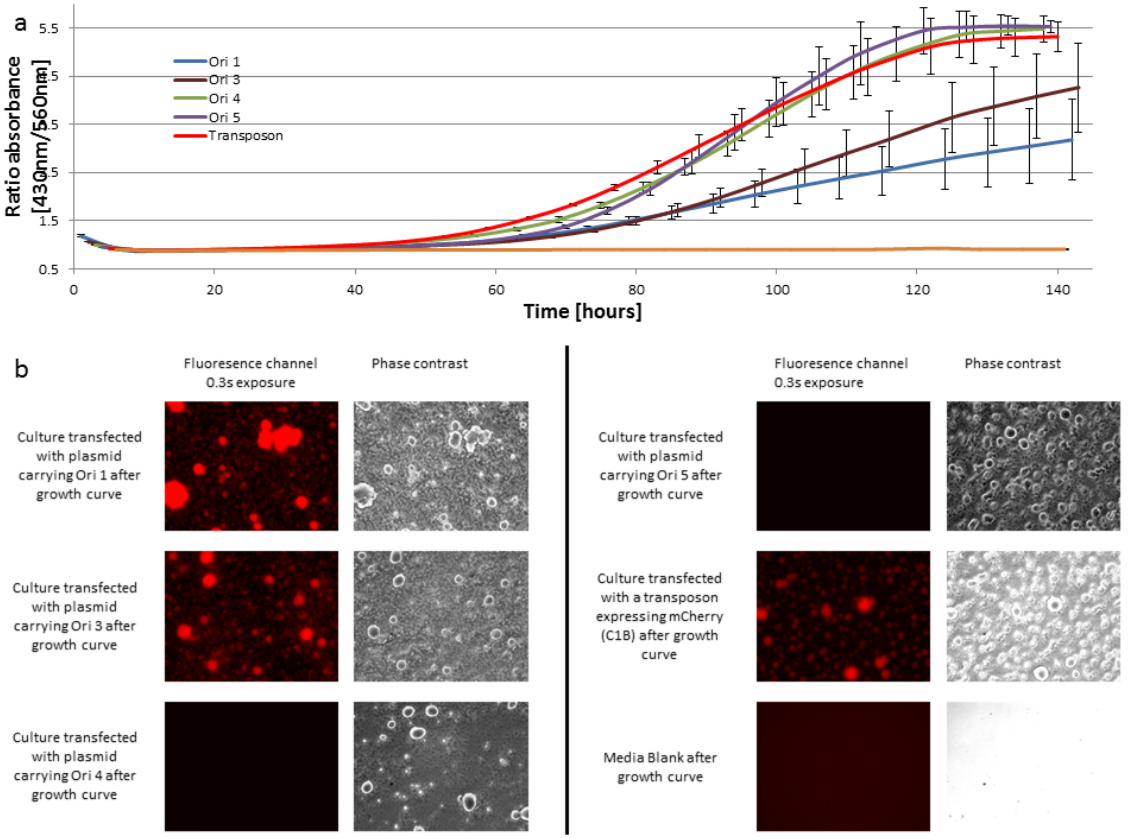


Figure 10 Growth curves of five different Mycoplasma strains (a) one is transformed with a mini-transposon vector and four with different self-replicating plasmids carrying Ori 1 to Ori 5 (Figure 6). The cultures had been normalized before according to the ATP content and previously passaged multiple times. A Media Blank is shown in orange. (b) Microscope pictures at 20x magnification, the fluorescence image was taken with a leica N2.1 Filter cube and at 0.3s of exposure. The images were taken from cultures grown in a 24 well plate during the growth curve of 4 days after inoculation.

The slow recovery of cultures transformed with Ori 4 and Ori 5 in combination with normal growth later on indicates a very low number of initially transformed cells and a potential insertion by homologous recombination of the plasmids in the genome. To test without doubt if the plasmids carrying Ori 4 and Ori 5 are integrated in the genome a southern blot is necessary. We have not performed this experiment yet but plan to do so in the near future.

Conclusion

Our aim in this study was to establish self-replicating plasmids for *M. pneumoniae* to allow the rapid and consistent testing of parts for the engineering of *M. pneumoniae.* While we could establish self-replicating plasmids in *M. pneumoniae*, we also observed altered growth characteristics compared to cultures that had the antibiotic resistance integrated in the genome by transposition. We speculate that the slower growth could originate from an uneven distribution of plasmids during cell division. However, it is

important to note that based on our data we cannot exclude other reasons for the impaired growth. Alternatively the cell division machinery is affected by the doubling of the genomic region involved in cell division. Further investigation to clarify this would be necessary.

The slow growth and the unusual colony morphology rendered the constructs carrying Ori 1 and Ori 3 unsuitable for our desired application, the fast and rapid screening of genetic parts in *M. pneumoniae.* The constructs carrying Ori 4 and Ori 5 showed improved characteristics regarding the growth but the very slow initial recovery, in conjunction with the reduced expression of the fluorescence protein, left them also unsuitable for our needs.

In summary, we were able to construct self-replicating plasmids for *M. pneumoniae* adapting a previously published strategy. However, we failed in our attempt to generate a genetic tool for the fast and consistent screening of genetic constructs in *M. pneumoniae.*

Material and Methods

Molecular biology methods

All primers were ordered from Sigma-Aldrich. As standard procedure all PCRs were setup with a temperature gradient as annealing temperature for optimisation. Either Phusion (Thermo scientific product # F-534) or KOD high fidelity polymerase (EMD Millipore product # 71085) was used according to the manufacturer's recommendation. All primers are listed in Table 1 and the obtained fragments in Table 2.

Plasmids were assembled using the SLIC technique (Li and Elledge, 2007). For the production of single strands T4 DNA polymerase (NEB product # M0203) was used. The constructs containing Ori 2 and a control with a transposon were assembled out of 3 pieces in a one-step reaction. Transformations were done in *E. coli* Top 10 cells as described in the general material and methods. The numbers of *E. coli* colonies obtained were in the lower 2 digit range. Single *E. coli* colonies were picked and plasmids purified by using Miniprep Kit (Qiagen product # 27104). The plasmids were analysed by restriction digestion and then sequenced. Constructs containing Ori 1, Ori 3, Ori 4 and Ori 5 were assembled out of 2 fragment cloning, for this the backbone consisting out of the gentamicin resistance and the RFP was amplified by PCR and then the

respective fragment cloned by SLIC. Then, constructs S3 and S4 were completely sequenced. After no mutation was found in the backbone of S3 and S4, only the insert was sequenced for the derived vectors S5 and S6.

The second generation plasmids with the puromycin resistance were constructed by amplifying the *E. coli* origin of replication, the ampicillin and puromycin resistances from the plasmid pJet 1.2. Pac using the primers Fwd and Rev pJet -Pac-Ori. The plasmid pJet 1.2 Pac was previously ordered from genscript and the synthesized and amplified region corresponds to the $Puro^R$ and spiralin promoter region of Mini-Tn4001PsPuro (GenBank accession no. FJ872396). The chromosomal origins and the fluorescent protein were amplified from the previously obtained gentamycin constructs using the primers Fwd and Rev Ori-Pac. The constructs were assembled in an isothermal assembly mix.

Ori 5 was constructed by cutting the Ori 5 vector with *Xba*I (NEB product # R0145) then the small region was synthesized from the oligos "Primer 1/2/3 for B ori" by using the protocol of Gibson et al., (2010)

The complete plasmid sequences will be made publicly available in genebank, as soon as possible.

Isolation of extrachromosomal plasmids from *M. pneumoniae* culture

A fully grown culture from a $75cm^2$ flask was harvested and washed in PBS. The plasmids were extracted using a standard Miniprep Kit (Qiagen product # 27104), while following all recommendations of the manufacturer for low yield plasmids. The DNA was eluted by using the provided elution buffer but the incubation time was extended to 5 min and performed at 50°C.

Table 1 Primers used to generate the vector reporter backbone and amplify the fragments of the chromosomal origin of replication

Name	Fragment	Sequence
Pr1-Ad1r	Ori 1	ACAGTTGGCCGGCCACTGGAATTGCCTTTAACAATTACTTCACCTTCTTCA
Pr2-Ad3f	Ori 1	CTCTTGGGCCCGTACCTCTCATATCAGGGTTGCTATTGAGGGTTTCACG
Pr3-Ad3f	Ori 2	CTCTTGGGCCCGTACCTCTCAACTGCCTTACAAACGCCCTTGAACAA
Pr4-Ad1r	Ori 2	ACAGTTGGCCGGCCACTGGAAACCAACATCCAACAAGAAGTACCAAACTG
Pr5-Ad3f	Ori 3	CTCTTGGGCCCGTACCTCTCAGGTGTTATTTAATCTTTCTGGAT

Pr6-Ad1r	Ori 3	ACAGTTGGCCGGCCACTGGAATTAAGGGTTAGTTTGACAGTACTCGCA
Pr8-Ad3r	dsRed	TGAGAGGTACGGGCCCAAGAGacgtaatacgactcactatagg
Pr9-Ad2f	dsRed	AGCTGCACGACGTGCATGATTAGCTTTGGACACACACTAGTA
Pr13-Ad1f	Gentamycin resistance	TTCCAGTGGCCGGCCAACTGTCAGTTCAATTTGGGTTTATAGC
Pr14-Ad2r	Gentamycin resistance	ATCATGCACGTCGTGCAGCTGTAATACGGTTATCCACAGAATC
Pr15-Ad3f	IS256	CTCTTGGGCCCGTACCTCTCATTCTGTGGATAACCGTATTACC
Pr16-Ad1r	IS256	ACAGTTGGCCGGCCACTGGAAGAGCCGGAAGTATAAAGTGTAA
Fwd pJet -Pac-Ori	Puromycin resistance	TCAGTTCAATTTGGGTCACTCAAAGGCGGTAATA
Rev pJet -Pac	Puromycin resistance	CTAATCATGCACGTCGATGAGGTGGTTAGCATA
Rev Ori-Pac	Ori 1/2/3/4	TACCGCCTTTGAGTGACCCAAATTGAACTGAC
Fwd Ori-Pac	Ori 1/2/3/4	GCTAACCACCTCATCGACGTGCATGATTAGCTT
Primer 1 for B ori	Ori 5	AAGTTACAAATTGGATCTTTAATCGCTTAATTTTTAATGTGGTATAATT GTTTGGATTCG
Primer 2 for B ori	Ori 5	TGCACGTTGGAAAGTTAAAAGCCATAAGCAAAATTTATGGCGAATCCA AACAATTATACC
Primer 3 for B ori	Ori 5	TTTTAACTTTCCAACGTGCACATTAAATGGAAGACAATTAAAAGATCTC CTACAATATTCT

Table 2 Amplified PCR fragments and respective template used.

Name	Primer 1	Primer 2	template	Size
Ori 1	Pr1-Ad1r	Pr2-Ad3f	Genomic Dna	2694bp
Ori 2	Pr3-Ad3f	Pr4-Ad1r	Genomic Dna	2395bp
Ori 3	Pr5-Ad3f	Pr6-Ad1r	Genomic Dna	3304bp
dsRed	Pr9-Ad2f	Pr10-Ad3r	pMt-clpB-luc-pmp200-RFP	1070bp
Gentamycin resitance	Pr13-Ad1f	Pr14-Ad2r	mini-TNP PmT85	2548bp
IS256	Pr15-Ad3f	Pr16-Ad1r	mini-TNP PmT85	1717bp
Puro backbone	Fwd pJet -Pac-Ori	Rev pJet -Pac	pJet1.2-PAC	3420bp
Origin	Rev Ori-Pac	Fwd Ori-Pac	S1/S2/S3/S4/S5	2287bp-4443bp

Growth curve

Before inoculation all cultures were normalized according to their ATP content (Saglio et al., 1979) using the *ATP* Bioluminescence *Assay* Kit HS II (Roche product # 11699709001). The cultures were grown in Nunc 24 well plate (Nunc product # 142475) at 37°C and the absorbance at 430nm and 560nm was measured in regular intervals in a Tecan infinite M200 Pro plate reader. For each construct quadruple wells were set up. To avoid local effects from cell clusters, multiple measurements per well were performed each with at a slight offset.

Selective media gentamycin

Full modified Hayflick media was prepared as described in the general material and methods. For selection with gentamycin (life technologies product # 15750-60) a concentration of 200µg/ml was normally used. Higher concentrations for the regular refreshing of gentamycin did not improve the selection efficiency.

A synthetic “cloning platform” to supply genetic tools in *Mycoplasma pneumoniae*

Abstract

Mycoplasma pneumoniae is a model organism for system biology well characterized by numberous “-omics” studies. However, the genetic tools in *M .pneumoniae* to validate the derived hypothesis from this wealth of information are underdeveloped. We designed a set of genes that should help to overcome the problem of lacking genetic tools in *M. pneumoniae*. The set contained a T7 Polymerase for independent expression, four repressor proteins (Tet / Lac / CI857 / T7 Lysozyme) and a set of recombination sites for the targeted integration of further genes. The design was done with the software Gene designer 1.0 and optimized according to the state of the art. After we introduced the first version in *M. pneumoniae* we found that the majority of the proteins were not expressed. Then we restructured the promoter sequences of the all non-expressed genes and transformed it into *M. pneumoniae.* We saw a slight improvement for some proteins but no major breakthrough was obtained. In the course of the experiments we made the observation that fusion of a gene of interest can mediate expression. Therefore, we decided to split up the cloning platform in individual genes and systematically screened the influence of N-terminal fusions on their protein expression level. We obtained for each gene at least one version that was reasonable expressed. Then we used this expressed gene to reassemble the platform. Unfortunately, also this version provided no functionality.

Introduction

Synthetic biology is a new discipline in biotechnology that aims the engineering of whole organisms. While the technologies which changed biotechnology to synthetic biology evolved rapidly. It is still hampered by a limited understanding of the biological systems it wants to alter or complement with a new genetic network. One possible solution to this problem is the use of so called minimal bacteria as chassis, which have a small genomes and comparably simple genetic networks. The hypothesis is that these

minimal bacteria can be completely understood one day and therefore a precise and directed engineering will be possible.

M. pneumoniae is one of those minimal bacteria. It has a genome of only 800 kbp and was extensively studied in the last years (Güell et al., 2009; Kühner et al., 2009; Lluch-Senar et al., 2013; Maier et al., 2011, 2013; Yus et al., 2009, 2012). Currently, the genetic tools of *M. pneumoniae* are limited to a mini-transposon vector. This vector randomly integrates the gene of interest in the genome (Halbedel and Stülke, 2007). While this technique can be used for the disruption of genes or to integrate a small set of genes for proof of concept studies, it is unsuitable for the precise engineering of a bacterium. Homologous recombination, a technique used in most bacteria for editing the genome is not effectively working in *M. pneumoniae* (Krishnakumar et al., 2010; Sluijter et al., 2010). One unusual genetic tool available for Mycoplasmas is the genome transplantation technique developed at the Craig Venter Institute. With this protocol it is possible to transfer a complete Mycoplasma genome into yeast, maintain it there, manipulate it and then transplant it back into a Mycoplasma cell (Benders et al., 2010; Hutchison et al., 2008; Karas et al., 2013; Lartigue et al., 2007, 2009). Once the genome is maintained in the yeast it is possible to manipulate the Mycoplasma genome with all the genetic tools available in yeast. However, the protocol is very work intensive and the transfer back from yeast is inefficient. This limits the approach only to individual complex engineering tasks and makes it unsuitable for the everyday engineering experiments.

Furthermore, no genetic tools are available for controlled gene expression or silencing of a natural expressed gene. We decided to use the possibilities of gene synthesis to develop additional genetic tools for *M. pneumoniae*.

We ordered a ~11kbp stretch of synthetic DNA from the company DNA 2.0. The synthesized DNA consisted out 7 different genes. The genes are clustered in 3 modules: a transcription/gene expression module, a gene regulation/repressor module, and a recombination module (Figure 11).

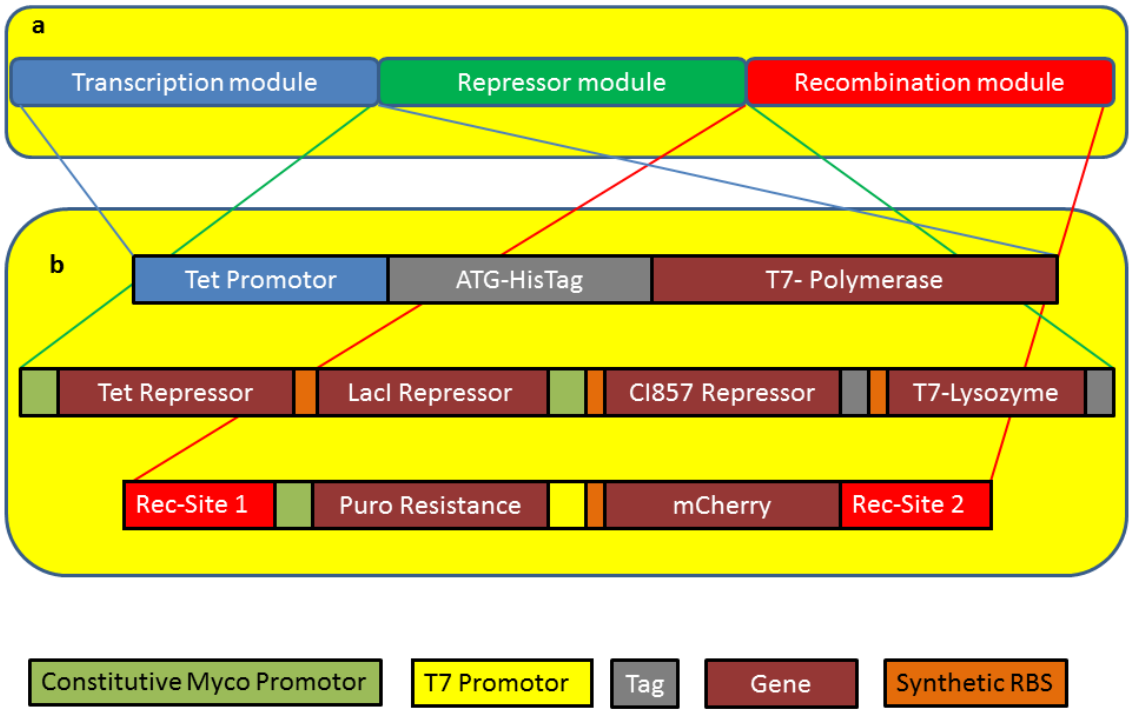


Figure 11 Schematic representation of the initial design of the "Cloning Platform". (a) Overview of the complete 11Kbp fragment is given. It consists out of transcription module (T7 Polymerase under the control of an inducible promoter). A repression module that contains 4 well characterized repressors and a recombination module. (b) Detailed view of the individual modules. The transcription module has Tet promoter that was designed in the lab followed by a T7 polymerase fused to a 6His tag at the N-terminus. The repressor module comprises the Tet repressor and LacI repressor under the control of the constitutive promoter of *mpn376* gene of *M. pneumoniae* genome. A shortened version of the *mpn376* promoter was then inserted after the LacI repressor to transcribe the CI857 repressor and the T7 lysozyme gene. All 4 repressor had synthetic RBS sites previous to ATG codon that were designed using the Salis RBS calculator (Salis et al., 2009). The recombination module contained a gene coding for resistance against puromycin (Algire et al., 2009) under a constitutive promoter and a mCherry fluorescent protein under the control of T7 promoter with LacI binding sites. The recombination module is flanked by Cre/Lox and Flp recombination sites.

Results and Discussion

Design of the platform

The scheme in Figure 12 shows in detail the functional genetic elements of the platform, with the individual parts not drawn to scale. The platform consists out of transcription module comprising a T7 Polymerase (Studier and Moffatt, 1986) under control of Tet inducible mycoplasma promoter. The T7 Polymerase is preceded and followed by a Mycoplasma terminator hairpin. This hairpin was used three times in the platform and it separates each module, being slightly modified each time to reduce unwanted recombination effects. All ribosomal binding sites (RBSs) in the construct are optimized for expression by using the RBS calculator (Salis et al., 2009). All coding sequences are

terminated by a triple stop codons comprising out of 2 stop codons in frame with the original transcript and a third that allows a +1 frame shift (Adachi and Cavalcanti, 2009). The repressor module consists out of a Tet, LacI and CI857 repressors and T7 lysozyme (Hochschild and Lewis, 2009; Lutz and Bujard, 1997; Studier, 1991). The repressor module is under the control of the *mpn376* promoter and after the first 2 repressor genes a short version of the *mpn376* promoter is repeated to enhance transcription of the 2nd half of the module.

To be able to investigate the expression levels of the different promoters we confirmed either that an antibody was available or fused the target protein to a tag recognized by a specific antibody. The Polymerase was fused to an N-terminus His tag, while the cI857 repressor was fused at the C-terminus to a PFT-c-myc tag and the T7 lysozyme was fused at the C-terminus to a FLAG-tag. The location of the tag was chosen either according to literature examples or based on available protein structures. The Tet and Lac repressors were left un-tagged as commercial antibodies are available for both proteins.

The basic version of the platform contains no recombinase. It has both Lox P sites and FRT sites framing a puromycin resistance gene under the spiralin promoter (Algire et al., 2009) and a mCherry fluorescent protein, under a Lac controlled T7 promoter allowing induced expression. The specific Lox sites used should allow unlimited integration events mediated by the Cre recombinase as they get destroyed upon integration (Kameyama et al., 2009). The FRT sites will be limited to 3-4 integrations mediated by the Flp recombinase (Bode et al., 2000b).

Each module and feature is separated by a unique restriction site in order to ease later modifications and improvements. A wiring scheme of the basic version of the platform is provided in Figure 13. The addition of anhydrotetracycline to the media, will induce the expression of T7-polymerase. Leaky expression of T7 Polymerase is inhibited by T7 lysozyme. The T7-Polymerase transcribes the mCherry gene. However, this transcription is blocked in the presence of the LacI repressor. Inhibition by the LacI repressor is released by the addition of IPTG to the media. We did neither introduce an IS256 element nor a *M. pneumoniae* origin of replication in the ordered DNA. We split the delivery up in two parts which allowed us to reduce the cost of the delivery by 60% compared to the standard market price. The DNA insert was designed in silico using the software gene designer 1.0 (Villalobos et al., 2006).

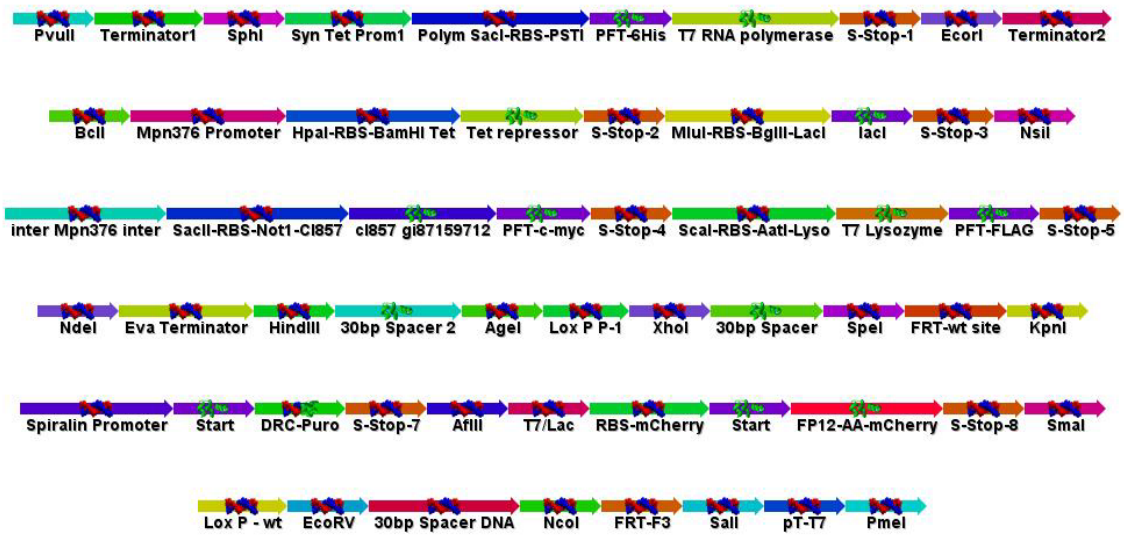


Figure 12 Detailed schematic representation of the designed platform. All genetic elements used are listed at their respective place in the platform. The size of the individual parts is not to scale

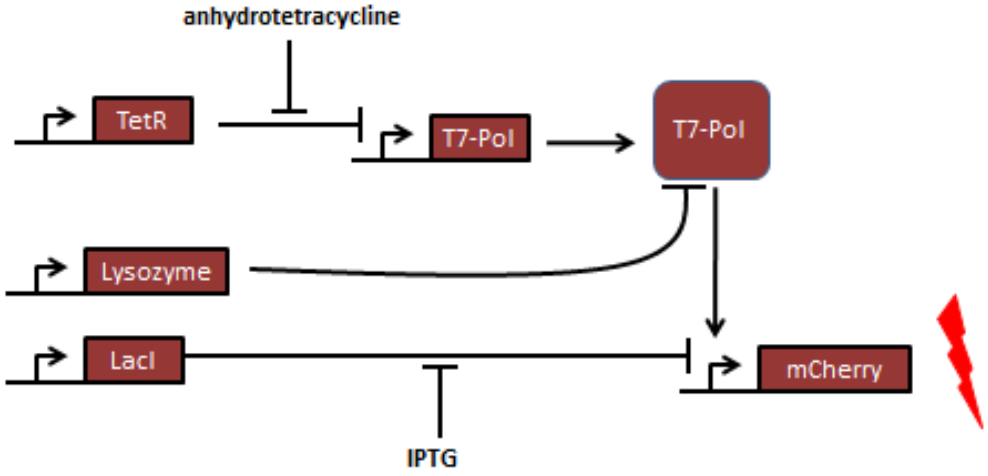


Figure 13 Theoretical wiring scheme of the cloning platform. The Tet repressor controls the transcription of the T7 polymerase. The repression is abolished by the addition of anhydrotetracycline. Leaky expression of T7 Polymerase is inhibited by a constitutively expressed T7 lysozyme. Only upon induction T7 polymerase expression the levels of T7 Polymerase exceed the T7 lysozyme. The expression of the reporter protein mCherry is under the control of a T7 promoter flanked by LacI binding sites. In the basal state the expression of mCherry is depending on the presence of T7 polymerase and is additionally repressed by LacI. In the presence of IPTG LacI releases it binding site and mCherry is expressed if a T7 polymerase is present.

Characterisation and troubleshooting of the cloning platform in *M. pneumoniae*

After we received the synthetic DNA we incorporated an IS256 element derived from the mini-tn4001 plasmid in the vector and transformed it into *M. pneumoniae*. We confirmed by PCR and Sanger sequencing of amplified PCR products that the complete platform was integrated in the genome. Then we tested if the individual modules were functional and if the respective proteins were expressed by Western blot (Figure 14). The transcription module (T7 polymerase) and most of the repressor module were not detectable in Western blots of *M. pneumoniae* lysates. The CI857 repressor and the mCherry reporter were the only proteins detected. The presence of mCherry was surprising as it is under control of a LacI controlled T7 promoter and should not be expressed without the presence of T7 Polymerase. We assumed that the expression resulted from a read through of the antibiotic resistance gene.

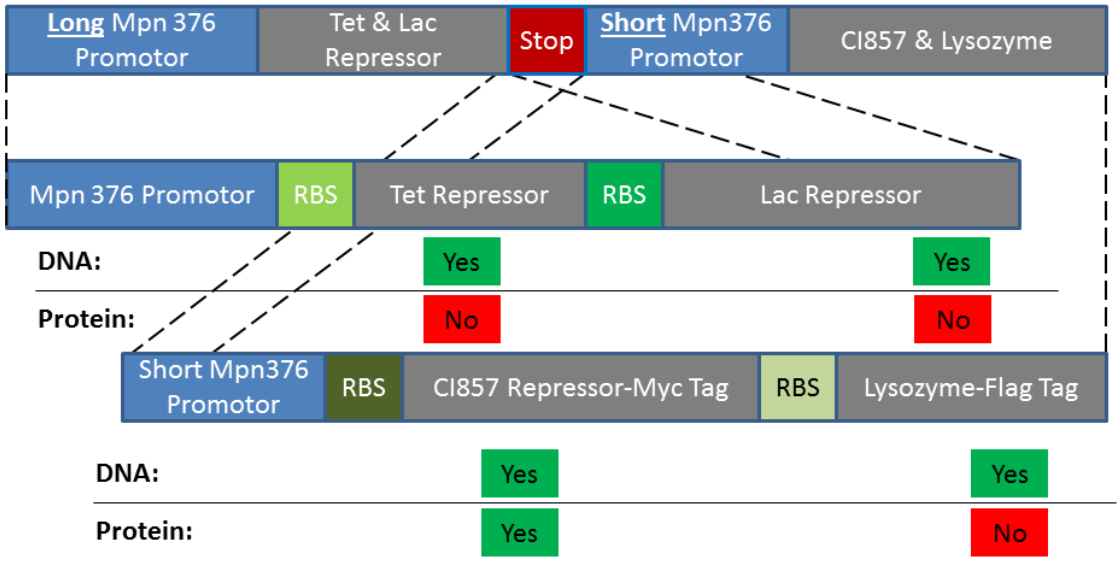


Figure 14 Schematic view on the expression detection of the core platform genes and protein after transformation into *M. pneumoniae*. For all genes the DNA was detected by PCR and subsequently sequenced. Only the CI857 repressor was detected clearly in Western blot. Both the Tet and the Lac repressor could not be detected on protein level. The Western blot with a FLAG-tag antibody showed a band that was specific for the transformed culture but at the wrong molecular size.

Based on the negative results from the first version we redesigned the Platform. The designed Tet promoter controlling the T7 polymerase expression was exchanged to another tetracycline inducible promoter described previously in the literature (Breton et al., 2010). The Tet and LacI repressor were put under control of EfTu (Elongation factor Tuf) promoter from *M. pneumoniae*. The EfTu promoter is one of the strongest promoters available in *M. pneumoniae* (Güell et al., 2009). Finally the RBSs, which were designed with the RBS calculator (Salis et al.,

2009), in front of not expressed proteins were removed. We assembled the restructured version of the platform out of 5 pieces using isothermal assembly cloning (Gibson et al., 2009).

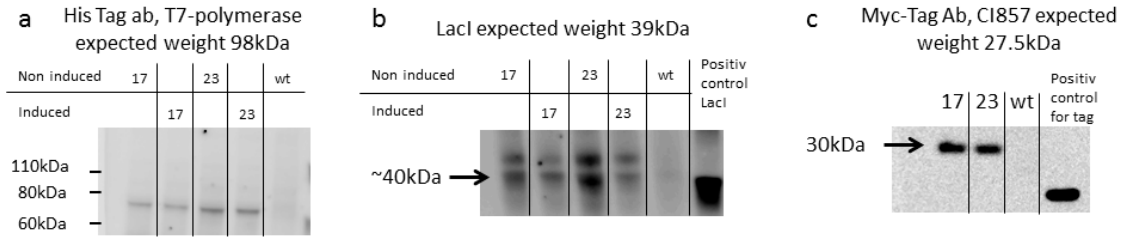


Figure 15 Western blots of the three platform proteins which could be detected in the restructured second version of the platform. Two transformed *M. pneumoniae* strains "17" and "23" were compared in each panel, (a) and (b) show additional for each strain a culture that was induced with anhydrotetracycline and a non-induced culture. (a) Western blot using an anti His-tag antibody to detect the T7 polymerase, which is fused at the N-terminus to a His-tag. No band could be detected at the expected weight of 98 kDa. However a band at ~70kDa which is specific to the transformed cultures is visible. This could be a nicked version of the T7 polymerase as it was observed in other organisms (Ikeda and Richardson, 1987). The positive control had to be removed before developing the blot as it was too intense compared to the signal from the Polymerase. (b) The same cultures on a Western blot developed with a primary antibody to the LacI protein. A band at the expected size specific to the transformed strains is visible. The positive control of the LacI protein is shown on the right. An additional band specific to the transformed cultures is visible above the expected size of the LacI protein. The band could represent a readthrough product. (c) Western blot of the two cultures treated with an antibody detecting Myc-Tag. The CI857 protein was fused at the C-terminus to the Myc tag. A specific band at the expected molecular weight is visible in both cultures. A positive control carrying the Myc tag is shown on the right. The positive control is a different protein than the CI857 fusion and therefore has a different molecular weight.

The redesign of the Platform yield some improvements (Table 1). From this version of the platform we could detect three proteins in a Western blot (Figure 15), while the functionality of the platform could still not be confirmed. While the band for the T7 Polymerase was not at the expected size, it is reported in the literature that T7 Polymerase can occur as a nicked but still active version (Ikeda and Richardson, 1987). It is worth noting the Western blots for T7 Polymerase and LacI were developed with a highly sensitive ECL reagent, so that the visualized bands represent very little amounts of protein. In the Western blot with an antibody against the LacI repressor (panel b of Figure 15) two distinct bands with a higher molecular weight than LacI are visible. These bands are specific for the strains carrying the platform and do not appear in the wt control. They could be potential read through products, despite the 3 stop codons at the end of the *lacI* gene.

Table 3 Summary of the analytical results from the 2nd restructured version of the platform. The complete DNA was integrated and could be detected in *M. pneumoniae*. Three proteins LacI, CI857 and mCherry could be detected by Western blot. The Tet repressor was not expressed at detectable levels and the T7-lysozyme protein was detected but at a wrong molecular weight. Functionality could not be detected for any protein except mCherry.

Gene	DNA	WB	FUNCTION
T7 Polymerase	OK	Potentially nicked version detected	No induction signal
Tet Repressor	OK	NO	(see above)
LacI	OK	OK	(see above)
CI857	OK	OK	N.A.
T7-Lysozyme	OK	No	N.A.
mCherry	OK	OK	OK

In order to troubleshoot the cloning platform further we split up the individual modules of the platform and test them individually. The Tet Repressor (TetR) was never expressed in any of the previous versions of the platform despite being under the control of a very strong promoter. Confusingly while the Tet repressor was never expressed at detectable levels, the LacI repressor was expressed in the second version of the cloning platform, while both are in the same operon (Figure 16).

Figure 16 Schematic representation of the repressor operon in the second version of the cloning platform

Based on various hints from previous experiments and from the literature, we hypothesized that the first base pairs of the gene could have a crucial role in gene expression. Secondary structure of the 5' mRNA is already described in the literature to influence expression levels (Bentele et al., 2013; Kudla et al., 2009; Scharff et al., 2011) while the extent is subject to debate (Supek et al., 2010). In theory the amino acid sequences of the platform were back translated choosing codons to minimize secondary structure at 5' of any mRNA sequence. However, we had noticed that primers annealing at the 5' end of the Tet repressor mRNA had a strong secondary structure. One possible explanation is that that all functional parts in the platform were separated by unique restriction sites causing severe constraints on the codon choice in the back translation process.

Based on this hypothesis we made a pilot experiment in which we fused the Tet-repressor coding sequence to different parts of the *M. pneumoniae* S200pmp peptide

and its promoter (Zimmerman and Herrmann, 2005)(Figure 17a). When the ATG start codon of the Tet repressor coding sequence follows directly the promoter sequence of the S200 peptide no expression was visible. However, when Tet repressor was fused to the first 15 bases of the S200 coding sequence we obtained a low level expression which reached a maximum when the first 45bp of the S200pmp coding sequence were fused to the Tet repressor (Figure 17b). We concluded that the starting region might contain or exclude certain signals like secondary structures, codon choice, motifs or else that are necessary to allow for translation or transcription.

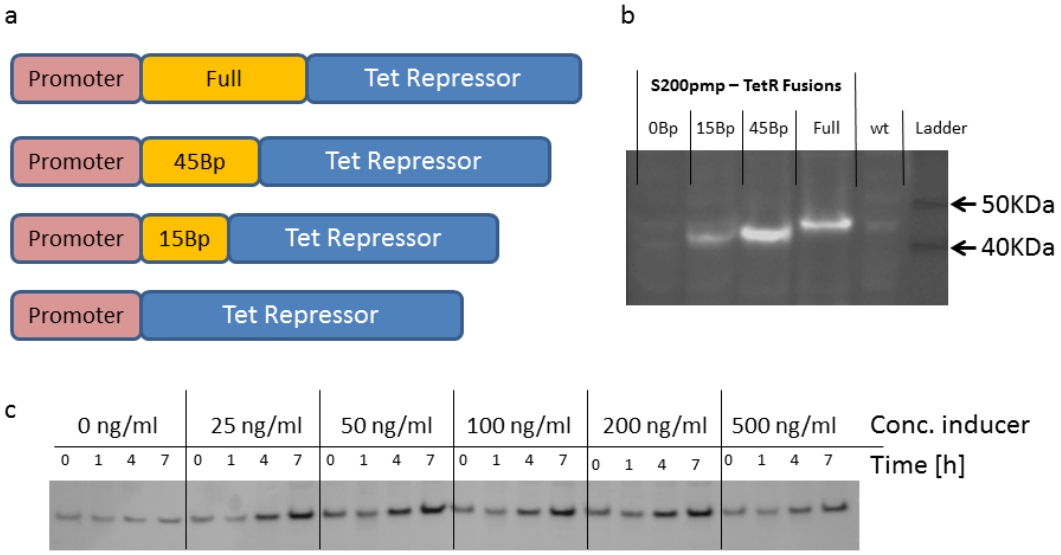


Figure 17 (a) Schematic representation of the different fusion products between the S200pmp gene and the Tet repressor coding sequence. The construct without fusion had directly the upstream promoter region of S200pmp before the ATG codon of the Tet repressor coding sequence, which would include any 5' UTR sequence. (b) Western blot of strains transformed with the different constructs shown in (a) On the right side of the plot a sample of *M. pneumoniae* M129 wild type (wt) is shown as control. The expected size of the Tet repressor is 37kDa and 40kDa for the longest fusion. While the specific bands run slightly above the expected size they show the expected pattern of increasing size with increasing fusion length. (c) Test for the functionality of the TetR fusion. It is a Western blot against the protein Venus under the control of a synthetic promoter with Tet-R binding sites. The Venus constructs were transformed in a strain expressing 45bp fusion construct of (a). The culture was split up in 6 aliquots and each aliquot was grown in the presence of a different concentration of the inducer anhydrotetracycline. For each concentration of inducer aliquots were taken at 4 time points (0h, 1h, 4h and 7h). A high level of basal expression is visible in the sample without inducer (0ng/ml) however it is stable over the time course of the expression. In the presence of inducer an increased expression of Venus is observed. However, the system seems to be already fully induced in the presence of 25ng/ml of anhydrotetracycline. This indicates a low concentration of TetR repressor in the cells.

Functionality of Tet repressor fused to pmp 200

With the Tet repressor expressed at a low level we started an experiment to check the functionality of the repressor. We used a construct containing a Venus protein under the control of a synthetic promoter with Tet repressor binding sites. The same promoter was used in the first version of the platform to control the expression of T7 polymerase. We could see a slight induction of expression upon addition of anhydrotetracylin (Figure

17c). However, it should be noted that the basal expression was high and the observed induction was weak.

Fusion screening

Based on this result we started a project in which we took all proteins parts of the platform and systematically investigate the influence of different promoters and fusion to *M. pneumoniae* proteins (Figure 18). However, the size of the fusion (3-90 bp) made it difficult to introduce these sequences in a primer for PCR amplification. We therefore tested whether a simultaneous de novo synthesis of a sequence and assembly of the gene of interest, is possible with the one-step isothermal assembly (Gibson et al., 2009, 2010).

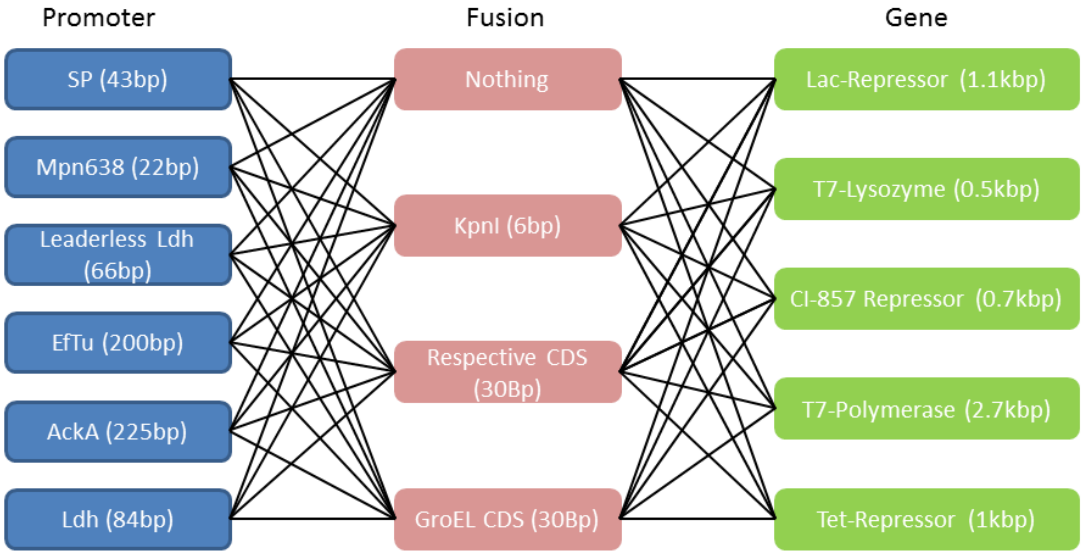


Figure 18 Graphical scheme for the constructs cloned in the small scale library. We used 6 promoters, 4 fusions and 5 genes. We choose only to clone a subset of 47 constructs from the possible 120 combinations. All promoters are from *M. pneumoniae*. SP stands for a synthetic promoter. Leaderless Ldh is truncated version of the normal Ldh promoter. The respective CDS stands for the first 30bp of the gene corresponding to the cloned promoter. The SP and leaderless Ldh promoter do not have their own CDS and were only cloned with the other fusions. All sequences are listed in detail at the end of this document.

We tested this approach by cloning 47 constructs combining 6 different promoters, 4 different N-terminal fusions and 5 genes (Figure 18). Of the 120 theoretically possible constructs we choose 47 for cloning. The library of 47 constructs was assembled in a miniTn4001-Puro-1 backbone (GenBank accession number: KC816623). The constructs consisted of a PCR insert a dsDNA between 500-2700 bp and a *de novo* sequence between 30-255 bp (Figure 19). The techniques are described in detail in the part "*In situ* overlap and sequence synthesis during DNA assembly" of this thesis.

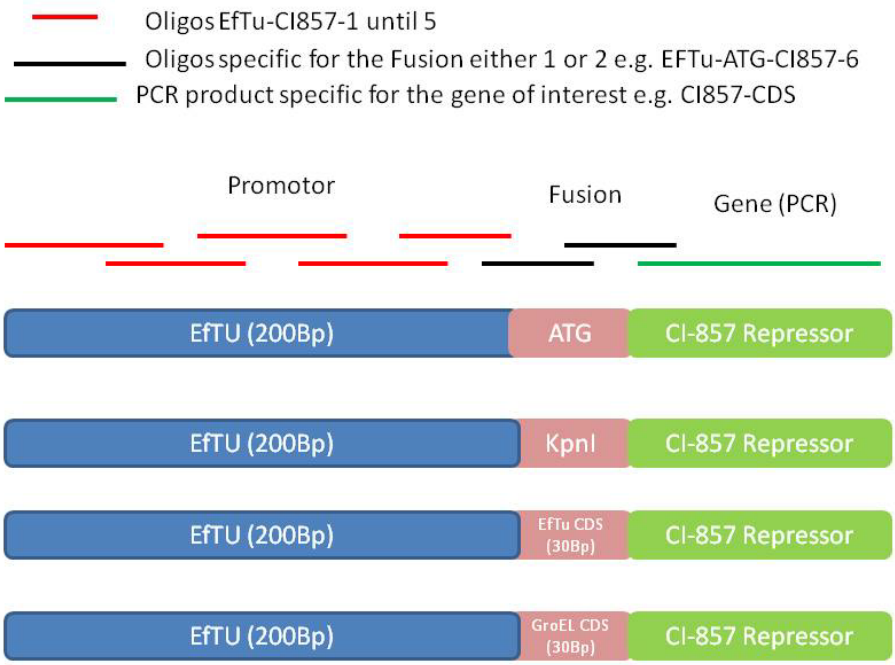


Figure 19 Schematic representation of the cloning strategy to build up the various fusion constructs. The promoters (Blue) and fusion parts (Pink) were built up *de novo* from commercial oligos. The green part represents the CDS of the gene of interest (green) which is amplified by PCR. All ingredients are mixed in a normal isothermal assembly master mix as described by Gibson et al. and then transformed into *E. coli*. The sizes are not to scale.

The constructs were assembled in *E. coli* into a mini-transposon vector. After sequence verification and scale up the constructs were transformed into in *M. pneumoniae*. The *M. pneumoniae* cultures were split twice and then a pellet was harvested and analyzed for protein expression levels by Western blotting. The results showed that the coding sequence of T7 lysozyme (Figure 20), T7 Polymerase (Figure 21b), Tet repressor and the LacI repressor (Figure 21c) depend on a fusion to a *M. pneumoniae* gene for efficient expression of the protein. In contrast, the CI857 repressor showed levels independent of the fusions made, with the expression levels only determined by the promoter (Figure 21a). Also the degree by which the construct depended on the fusion differed between individual coding sequences. The T7 lysozyme coding sequence showed an on/off type behavior for expression depending on the fusion to the coding sequence (Figure 20). In contrast the LacI and T7 polymerase coding sequence showed bands specific to certain transformed strains but at the wrong molecular weight. The construct Ldh–KpnI-T7 Pol showed no band at the expected size for T7 Polymerase of 95kDa, while two specific bands representing possible nicked or degraded products are visible. The same promoter without a fusion or a fusion to the first amino acids of the Ldh coding sequences showed a distinct band at the expected size in addition to the lower bands (Figure 21b). The LacI repressor coding sequence showed a similar

behavior. Only the construct, SP-GroEL-Lac showed a distinct band at the same height as the expected weight and the positive control (Figure 21c). Interestingly the samples 1-3 showed a band that is specific to the transformed strains. This band runs higher than the positive control, while the actual constructs are smaller as they lack an N-terminal fusion.

We calculated the accessibility score for ATG start codon of each construct (Scharff et al., 2011) to investigate whether the observed effects are based on the mRNA structure around the start codon. For the here tested constructs we could observe no correlation between the accessibility of the ATG codon and the amount of protein expressed.

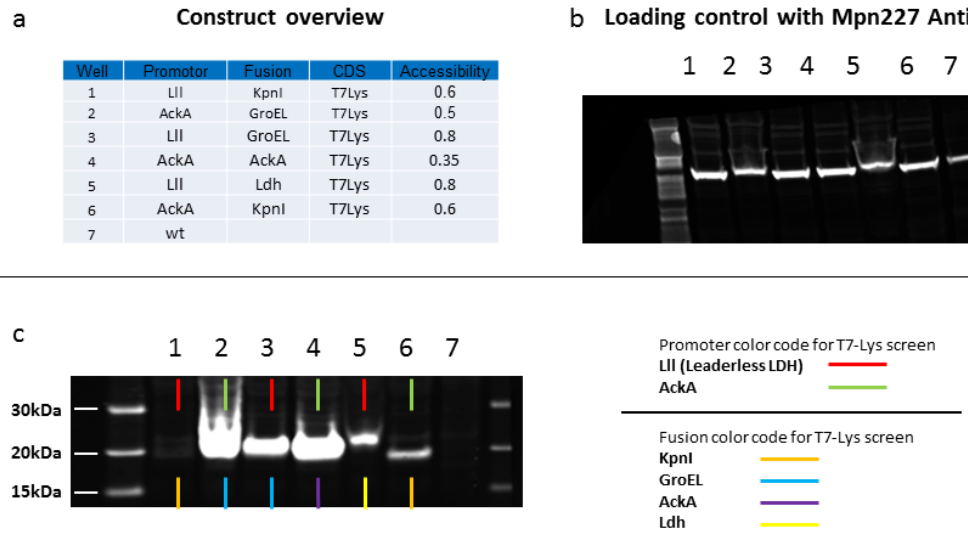


Figure 20 (a) Overview of all constructs in this figure. The first column identifies the lane for the construct on the Western blots. The column promoter indicates the promoter used in the respective construct. The fusion column indicates which fusion was used, KpnI stands for the introduction GGTACC in between the promoter and the ATG codon. GroEL indicates that coding sequence of T7-Lys was fused to the first 30bp of the *M. pneumoniae* GroEL coding sequence. The same scheme applies to the AckA and Ldh gene. The column accessibility gives the probability that the ATG start codon is unbound and is a measure for secondary structure around the ATG start codon (Scharff et al., 2011). (b) Loading control of all samples with an antibody against Mpn 227 (Fus), it is recorded at an excitation wavelength of 800nm on the same blot. Additionally, all samples were normalized according to the protein concentration determined by a BCA test before being applied to the gel. (c) Image of the 700nm excitation of the blot. A primary antibody against the FLAG-tag is used. All constructs under the control of the Lll promoter are marked with a red stripe in the upper row, all constructs having the AckA promoter are marked green. The lower row indicates the different fusion to the T7 Lysozyme coding sequence. The result shows that the fusion has a significant influence on the expression level of the protein, which cannot be explained by secondary structure effects. The expected weight of the T7 lysozyme is ~18kDa.

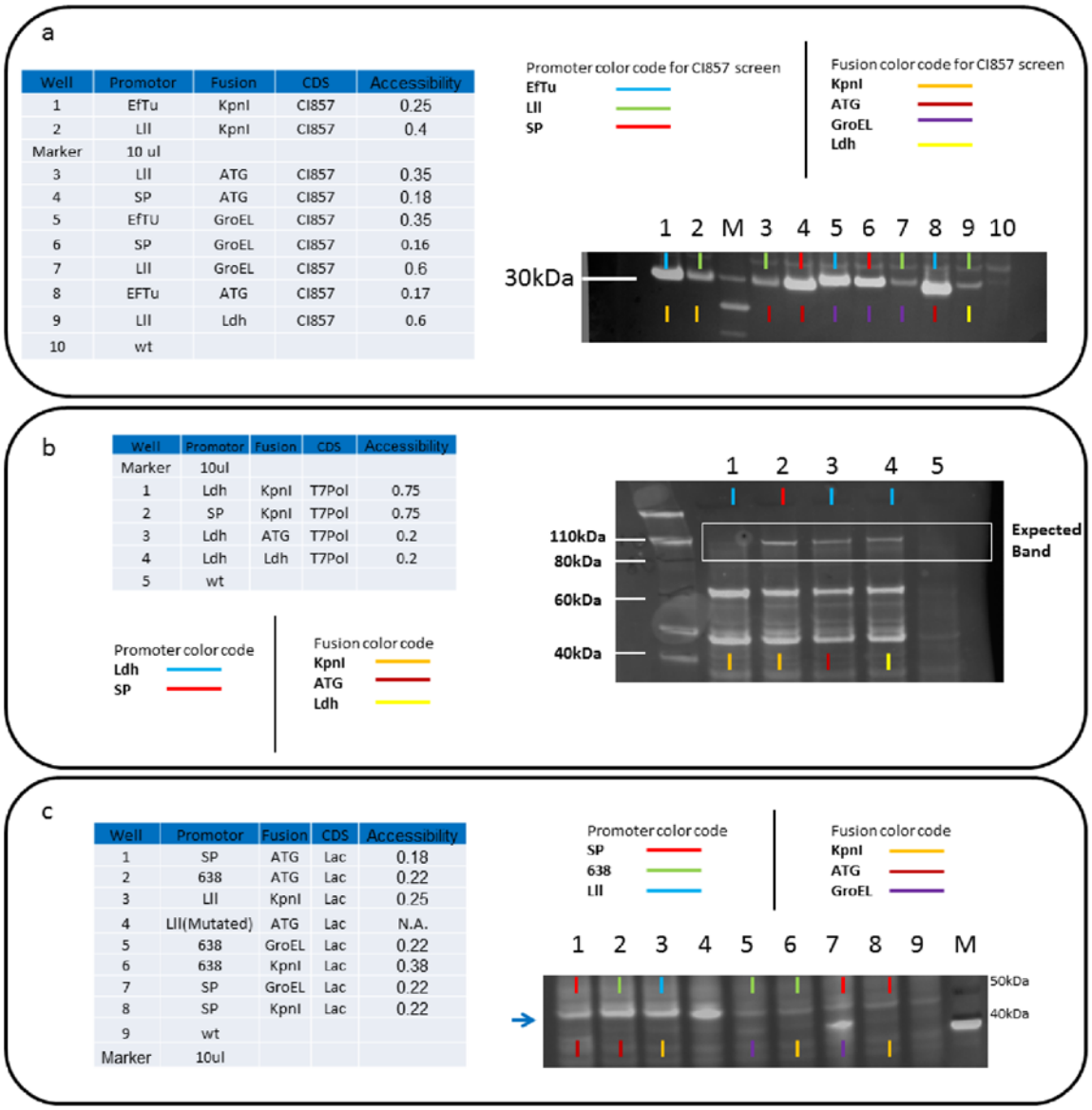


Figure 21 Results for the screening of the CI857 repressor, the T7-Polymerase and the LacI repressor. All samples were normalized according to their total protein content by a BCA assay. The table in each panel lists the screened promoter, fusion, coding sequences and the accessibility as measure of secondary structure around the ATG codon, the setup is as in Figure 20, additionally "ATG" in the column fusion stands for a direct connection of the coding sequence to the promoter without any fusion. (a) CI857 screening. The blot shows the *chemiluminescent* detection of the Myc-tag. The color code for marking constructs with similar promoter or fusion is above the blot. The expected weight of the CI857 is 27kDa. A specific band at the expected size is visible for most constructs. A comparison of the different fusions and promoters shows that the biggest influence on protein expression comes from the promoter. This is in contrast to the results of the T7 lysozyme coding sequence. Secondary structure seems to have no detectable influence. (b) T7-Polymerase screening. The setup of the table is the same as panel a, please note the color code is different. All samples from transformed strains showed specific bands that did not appear in the wt control. Only sample 2-4 showed a specific band at the correct molecular weight of T7 polymerase (white box). The lower bands could represent a nicked version of the T7 polymerase as it was observed earlier. An influence of secondary structure cannot be observed. (c) LacI repressor screening. A specific band with the correct size is only detected in sample 7. Sample 1-4 also show a specific band but at a molecular weight slightly higher than expected, while the actual constructs are smaller than sample 7. This effect could be based on the amino acid sequence of the GroEL fusion in construct 7.

We then reassembled four new versions of the platform based on the results from the fusion experiments. The constructs were designed to successively build up a basic set of

functionalities from the cloning platform (Figure 22). The reporter module was the same as in the first version of the platform. As transcriptional module we choose a T7 polymerase under the control of the Ldh promoter and fused to the first amino acids of the *M. pneumoniae* Ldh protein. As repressor we used the fusion of Groel and LacI under the control of the Synthetic promoter, as it was the only one expressed with correct molecular weight. Further we incorporated a T7 lysozyme, to counter leakey expression of the T7 Polymerase, under the control of the AckA promoter and fused to the start of the *M. pneumoniae* AckA gene. The AckA and Ldh promoters are induced and repressed by different carbon sources for growth in the media (Halbedel et al., 2007). In the designed setup a switch of growth medium should also change the equilibrium between the T7 polymerase and T7 lysozyme, from an excess of T7 lysozyme to an excess of T7 polymerase.

Contrary to our expectations Construct 1 showed already bright red fluorescence. We still investigated whether Construct 2 showed enhanced fluorescence or an inducible expression could be achieved with construct 3 or 4 upon exposure to IPTG or substitution of glucose to glycerol in the growth Media. Unfortunately all constructs showed uniform expression levels of mCherry as judged by fluorescence microscopy. We later identified a short stretch in the designed RBS site for the mCherry protein that can serve as potential promoter in *M. pneumoniae.* Further experiments to clarify this are necessary.

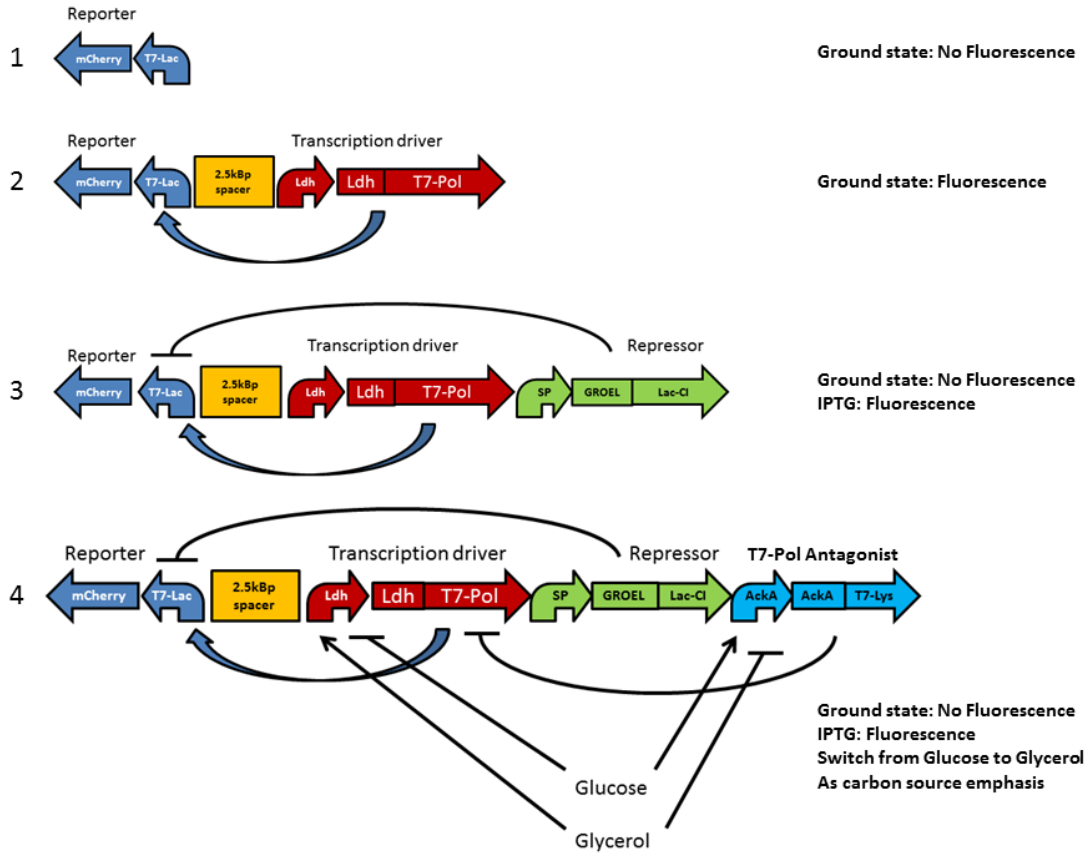


Figure 22 New version of the cloning platform reassembled out of parts that were found in the fusion screen. On the right hand side of the figure the functionality is described. Construct 1 is only the mCherry reporter under the control of T7-lac promoter.

Conclusions

The results from this study indicate that we have a gap in our knowledge about translational or transcriptional control in *M. pneumoniae*. The results from the fusion experiments show that the coding sequence after the ATG plays an important role in the determination of protein expression levels. This effect could not be explained by the analysis of the 5' mRNA secondary structure. Similar results, although only with a single constructs were already obtained earlier (Loechel et al., 1991). In order to understand this effect a number of experiments future experiments necessary. First the mRNA concentration for each construct needs to be determined to investigate if the effect observed in the Western blot is originating at the transcriptional or translational level. Then a systematic shuffling off the first codons from a set of coding sequences has to be done to define which codon combinations promote expression. An experimental procedure feasible to do this is currently developed with Jae-Seong Yang. Based on these results a hypothesis can be made what exactly is the nature of this effect.

Ultimately design rules can be deduced from this knowledge to reliably predict protein expression in *M. pneumoniae*.

Material and Methods

Molecular biology methods

All DNA manipulation techniques are covered in the general section and the Material and methods section of the chapter "*In situ* overlap and sequence synthesis during DNA assembly". All assemblies were done by the isothermal assembly protocol. For difficult assemblies overhangs were extended to 30 bp and for the first integration of a IS256 element into delivered version of the platform 50bp overhangs were used.

The complete sequence of the version 1.0 and 2.0 of the platform is given in genebank format in the sequences part of this section. Also the complete sequence of construct 4 (Figure 22) is provided there.

Induction by anhydrotetracycline

A non-adherent strain was transformed with a mini-transposon vector transferring resistance to gentamycin and carrying a venus fluorescent protein under control of designed tet promoter. The strain was allowed to recover and split multiple times. Then, it was transformed a second time with a mini-transposon vector carrying a 45bp fusion of pmp200 to the tet repressor coding sequence, under the control of the pmp200 promoter. After one passage, the culture was split up in 6 aliquots and each culture was grown for 1 day. Then each culture was induced by the addition of anhydrotetracycline (Fluka product # 37919) or a blank was added. The cultures were grown in the presence of the inducer and at 4 time points (0h, 1h, 4h and 7h) samples were taken. The aliquots were centrifuged to obtain cell pellets, washed with PBS and lysed with a Bioruptor in a buffer of 8M Urea (Sigma product # U5378), 100mM $NH_4(CO_3)_2$ (Sigma product # 09830**)**. Subsequently, the total protein concentration was determined by a BCA assay (Pierce product # 23225) and equal amounts of total protein was loaded on a Novex 4-12% bis tris SDS gel (life technologies product # WG1403BOX) and subsequent western blotting with a iBlot setup (life technologies product # IB3010-01)

Table 4 Primers used to generate PCR fragments for assembly and modification of the platform. Sequencing primers are not indicated neither the primer used to generate the fusion library, these are described elsewhere.

Name	Sequence
Fwd Primer Tnp 50bp	AACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGGTTTAAACCTTCTGTGGATAACCGTATTACC
Rev Primer Tnp 50bp	CTATAGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCCCTTCATTTTCGAGCCGGAAGTATAAAGTGTAA
T7 Pol Fwd	CACCATCATCACCACCATAA
T7Pol Rev Hom efTu	TTTTGTAATTGCTGACATTGGGACTTGAACCCATAAC
efTU Rev Hom T7	AAGGTTATGGGTTCAAGTCCCAATGTCAGCAATTACA
efTu Fwd Hom Tet	TTTGACTTGTCGAGCCTGGACATGTGTTTGAATTACGTCT
Lac Rev Hom SpirPuro	CTAACTTTTAATTCTGGCGGATGCATCTACTTATCGCTATTGA
SpirPuro Fwd Hom Lac	ATAGCGATAAGTAGATGCATCCGCCAGAATTAAAAGTTAG
Tnp Fwd Hom spirPuro	TTAGAGGTAAGTAGCTTAAGCGAGTCAGTGAGCGAGGAA
SpirPuro Rev Hom Tnp	CTTCCTCGCTCACTGACTCGCTTAAGCTACTTACCTCT
Tnp rev Hom pYES1L	GCTCACTGACTTTAATTAACTGCGGCGAGGGAGCCGGAAGTATAAAGTGTAA
Tetfwd Hom efTU	AGACGTAATTCAAACACATGTCCAGGCTCGACAAGTCAAA
Fwd Tet Prom	GTCCCTGTTTGCATTAT
Rev Tet Prom	TGATCGGCTAAAGTGTT
Full pmp200 peptide	GACTTGTCGAGCCTGGACATTTTTTGGCAGCATTTGC
45bp pmp 200 peptide	GACTTGTCGAGCCTGGACATGTTAGCTGTTTCGGTTT
15bp pmp200 peptide	GACTTGTCGAGCCTGGACATTTTCTTGTCTTCCATCTTA
0bp pmp200 peptide	GACTTGTCGAGCCTGGACATCTTAATATGAAATAA
Rev Tet overhang PstI	ATCTTGCCCCCCGGGCTGCAAGCTACCCGCCGTATT
Fwd Tet	CCAGGCTCGACAAGT
Fwd pmp200 Hom Tnp puro	AATTATCGATACCGTCGACCAGTTTTTGGCAGCATTT
S2 tnp Fwd pmp200 Tet	TATCGATACCGTCGACCTCGACTTTTTGGCAGCATTT
Rev pmp200	GATCCTACCAGTTCAA
S2 tnp Rev pmp 200 Tet	TGATATCGAATTCCTGCAGCGTGTGTCCAAAGCTAA
Rev pmp200 Hom Tnp puro	TTGTGTAAAAGGGCCCCCCCGATCCTACCAGTTCAA
Fwd pmp200	TTTTTGGCAGCATTT
C1B mCherry R	ACACTTTATACTTCCGGCTCGAACTTCCCTTCATT

C1B TNP R	AGGAAAATGAAGGGAAGTTCGAGCCGGAAGTATAAA
C1B mCherry F	CTTATTCGACTCCCTATAGAAATTGTGAGCGCTCACAATTGTGTATGACTAGGAAAC
C1B BB F	AATTGTGAGCGCTCACAATTTCTATAGGGAGTCGAAT
C1B ccdb R	CCCTTATTCGACTCTATAGGCGGTATCGATGATAT
C1B BB R	CGGGGATATCATCGATACCGCCTATAGAGTCGAATAA
C1B Puro F	GGGGAATATAACCCGGGAAGTACTCGAGGCAGAAA
C1B ccdb R-1	ACTTCTTTCTGCCTCGAGTACTTCCCGGGTTATAT
C1B TNP F	TTAGAGGTAAGTAGCTTAAGTTCTGTGGATAACCGT
C1B Puro R	TAATACGGTTATCCACAGAACTTAAGCTACTTACCTCT
C2B T7Pol R	TGCACTGGCCAGGGGGATCACTGCAGGAATTCGATAA
C2B T7Pol F	CACGCCCGGGCGACGGATGGCGACGGTATCGATAA
C3B LacI F	AGGTTATCGAATTCCTGCAGATGAATTGATCGCCATA
C3B T7Pol R	ACCTATGGCGATCAATTCATCTGCAGGAATTCGATAA
C3B LacCI R	TGCACTGGCCAGGGGGATCAATGCATCTACTTATCGCTA
C4B T7 Lys R	TGCACTGGCCAGGGGGATCAAAGCTTCTACTTATGTCTA
C4B T7 Lys F	ATAGCGATAAGTAGATGCATCGTTAATAATGATGAT
C4B LacI R	TTCAATCATCATTATTAACGATGCATCTACTTATCGCTA

Table 5 Amplified PCR fragments for later assembly and respective template used.

Name	Primer fwd	Primer Rev	Size	Template
Restructuring of platform from version 1.0 to version2.0				
IS256-v1.0	Fwd Primer Tnp 50bp	Rev Primer Tnp 50bp	1810	pmT85
T7 Pol-efTu	T7 Pol Fwd	T7Pol Rev Hom efTu	2759	Platform v1.0
efTu	efTu Fwd Hom Tet	efTU Rev Hom T7	248	gDNA
Tet-Lac	Tetfwd Hom efTU	Lac Rev Hom SpirPuro	2204	Platform v1.0
Puro	SpirPuro Fwd Hom Lac	SpirPuro Rev Hom Tnp	987	Platform v1.0
Tnp	Tnp Fwd Hom spirPuro	Tnp rev Hom pYES1L	1641	pmT85
pYes1L	pre linearized Life technologies		10000	
TetR for Fusion to pmp200				
pmp200	Fwd pmp200 Hom Tnp puro	Rev pmp200 Hom Tnp puro	380	S2
TetR for Fusion to pmp	Fwd Tet-pmp	Rev Tet-pmp	1026	Platform v1.0

Full pmp200	Fwd pmp200 Hom Tnp puro	Full pmp200 peptide	290	S2
45bp pmp 200	Fwd pmp200 Hom Tnp puro	45bp pmp 200 peptide	248	S2
15bp pmp 200	Fwd pmp200 Hom Tnp puro	15bp pmp200 peptide	218	S2
0bp pmp 200	Fwd pmp200 Hom Tnp puro	0bp pmp200 peptide	203	S2
		Fusion reassembly		
Puro module	C1B Puro F	C1B Puro R	1045	Platform v1.0
ccdb	C1B ccdb R	C1B ccdb R-1	336	Platform v1.0
mCherry	C1B mCherry R	C1B mCherry F	1078	Platform v1.0
IS256	C1B TNP R	C1B TNP F	1717	Platform v1.0
Ldh-Ldh-T7Pol	C2B T7Pol R	C2B T7Pol F	2881	Platform v1.0
SP-GroEL-Lac	C3B LacI F	C3B LacCI R	1238	Platform v1.0
AckA-AckA-T7Lys	C4B T7 Lys R	C4B T7 Lys F	811	Platform v1.0
Backbone pJ251	C1B BB F	C1B BB R	2542	Platform v1.0

Western blots

Western blot were performed using Novex NuPage 4-12% Bis-tris gels (life technologies product # WG1403BOX) in either MOPS (life technologies product # NP0001) or MES Buffer (life technologies product # NP0002) according to the manufacturers instruction. The protein transfer to a nitrocellulose membrane was performed with an iBlot system (life technologies product # IB1001EU) according to the manufacturer's recommendation and with standard settings. For chemiluminescent detection the membrane was blocked with 5% BSA solution (Sigma product # A9418) and washed in Tris-Buffered Saline containing 0.05% Tween 20 (Sigma product # P7949). The primary antibodies were incubated in Wash buffer with 0.05% BSA for 1h and after multiple washes incubated with the secondary antibody for 1h. For detection the ECL substrate plus (Pierce product # 32132) was used and in case no signal was detected the blot was incubated with the SuperSignal West Femto Substrate (Pierce

product # 34094). A Fujifilm LAS 3000 imaging system was used to record the images. Alternatively, the LI-COR Odyssey infrared imaging system was used when an internal standard (Mpn227) was used as loading control. Blots imaged with the LI-COR system were blocked with the commercial Odyssey blocking buffer (LI-COR product # 927-40000) to improve the signal to noise ratio. All antibodies, respective supplier and used dilutions are given in Table 5

Table 6 All antibodies used in the study, the production host, source and used dilution.

Target	Produced in	Provider	Dilution
Mpn227	Rabbit	Gift Prof Herrman	1:3000
Tet repressor	Rabbit	Novus biologicals	1:1000
T7 Polymerase	Mouse	Millipore	1:10000
Flag Tag	Mouse	Sigma	1:1000
Myc Tag	Rabbit	Sigma	1:5000
LacI	Mouse	abcam	1:1000
DsRed	Rabbit	Clontech	1:1000
GFP (Venus)	Mouse	Roche	1:1000
Anti mouse IgG-HRP	Sheep	Jackson Immunoresearch	1:5000
Anti Rabbit-IgG-HRP	goat	Sigma	1:5000
Anti Rabbit-IgG-800CW	Goat	LI-COR	1:10000
Anti mouse-IgG-680RD	Goat	LI-COR	1:10000

In situ overlap and sequence synthesis during DNA assembly

Bernhard Paetzold[1,2]*, Carlo Carolis[2,3], Tony Ferrar[1,2], Luis Serrano[1,2,4], Maria Lluch-Senar[1,2]*

[1]EMBL-CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain.

[2]Universitat Pompeu Fabra (UPF), Dr. Aiguader 88, 08003 Barcelona, Spain.

[3]Biomolecular Screening and Protein Technologies Unit, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain.

[4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluís Companys 23, 08010 Barcelona, Spain.

* To whom correspondence should be addressed. Tel: +34933160259; Fax: 34933160099; Email: bernhard.paetzold@crg.eu, Correspondence may also be addressed to maria.lluch@crg.eu

ABSTRACT

Modern cloning methods are independent from restriction enzyme recognition sites. However, nearly all current cloning methods still require the introduction of overlaps by PCR, which can introduce undesired mutations. Here, we investigated whether overlaps needed for DNA assembly can be synthesized *in situ* and tested if "*de novo*" synthesis of sequences can be simultaneously combined with the assembly of larger double stranded DNA fragments. We showed in a set of 44 cloning experiments that overlaps of 20 base pairs needed for DNA assembly can be synthesized *in situ* from single stranded oligonucleotides. Short sequences of 30 to 255 bp can be synthesized from single stranded oligonucleotides concurrently with the DNA assembly and both techniques can be combined. The assembly of similar constructs by state of the art techniques would have required multiple rounds of cloning or tedious sample preparations while our approach is a one-step reaction.

INTRODUCTION

Since the first restriction enzyme based cloning was made in the early 80s,(Cohen et al., 1973) the techniques of molecular cloning have made significant advancements. Restriction enzymes were the method of choice during decades in molecular cloning. However, restriction cloning requires the presence or absence of enzyme recognition sites, thereby complicating the design of DNA constructs and limiting high throughput assembly. These problems were solved with the invention of Sequence- and Ligation-Independent Cloning (SLIC).(Li and Elledge, 2007) The SLIC method was the first sequence-independent cloning method reported. SLIC allows the easy generation of scarless constructs and the assembly of DNA fragments independent of restriction sites. It is based on the introduction of overlaps between vector and insert by PCR. These overlaps are processed to single stranded overhangs by the 3' to 5' exonuclease activity of T4 Polymerase and are annealed with their homolog partner strand. The annealed construct is transformed into bacteria, which subsequently ligate the construct. One drawback of this technique is that the exonuclease activity is difficult to control and it tends to produce longer single stranded regions than required. As result, single stranded gaps, which are inefficiently repaired by the host, are introduced in the final construct, thus impairing cloning efficiency.

Gibson et al. improved the SLIC method further and created a very popular cloning method.(Gibson et al., 2009) The main difference in the new method is the use of an enzyme mix, containing an exonuclease to produce single stranded ends, a high fidelity polymerase to fill up single stranded gaps and a heat stable ligase to ligate the final constructs. This method called one-step isothermal assembly can be used to clone double stranded DNA (dsDNA) fragments seamlessly, even when they are hundreds of kb long,(Gibson et al., 2009) or to synthesize short sequences (200 bp - 450 bp) from single stranded oligonucleotides (Gibson et al., 2010). We hypothesized that it must be possible to combine both applications of the one-step isothermal assembly and thereby simplify existing cloning workflows even further.

Currently, all sequence-independent cloning methods available for *E. coli* need homologous overlaps for the assembly of the fragments, which are usually introduced by primers in a PCR reaction. This can be challenging in cases where PCR amplification and subsequent sequencing is problematic. In addition, it requires a new PCR reaction for each overlap needed during subcloning increasing the number of steps involved in sample preparation. Herein, we show that it is possible to synthesize these overlaps *in situ* from oligonucleotides during the cloning assembly. The *in situ* synthesis of overlaps suppresses the need for redundant PCRs each time a DNA fragment is subcloned into a new vector. We further show that it is simultaneously possible to assemble *de novo* sequences from oligonucleotides and clone long DNA fragments. This approach allows the effortless and flexible introduction of modified DNA stretches from 30 bp to 255 bp. These modifications can be used to enhance protein expression or to characterize a protein of interest by the combination of customized ribosomal binding sites (RBS), affinity tags, antibody tags, secretion signals, localization signals and promoters.

RESULTS AND DISCUSSION

Simultaneous assembling of one dsDNA fragment and the *de novo* synthesis of an additional sequence

Our initial task was to assemble a set of constructs in which a gene of interest was fused to a target sequence at the beginning of the gene. However, the size of the fusion (3-90 bp) made it difficult to introduce these sequences in a primer for PCR amplification. We

therefore tested whether a simultaneous *de novo* synthesis of a sequence and assembly of the gene of interest, is possible with the one-step isothermal assembly.(Gibson et al., 2009, 2010) In the state of the art, these methods can be used either to assemble multiple overlapping dsDNA molecules or, alternatively, to synthesize *de novo* DNA molecules from overlapping oligonucleotides. The combination of both in a one-step reaction has not yet been described.

We tested this approach by cloning 47 constructs combining 6 different promoters, 4 different N-terminal fusions and 5 genes (Figure 18). Of the 120 theoretically possible constructs we choose 47 for cloning. The library of 47 constructs was assembled in a miniTn4001-Puro-1 backbone (GenBank accession number: KC816623). The constructs consisted of a PCR insert (dsDNA between 500-2700 bp) and a *de novo* sequence (between 30-255 bp). These *de novo* sequences were built up from combinations of 2 to 8 oligonucleotides each around 60bp with an overlap of 20 bp. For the complete experiment 77 oligonucleotides were used (Supplementary table 2). The *de novo* sequence contained the promoter and a short sequence that should be fused to the PCR insert. In the first round of cloning, we obtained 29 out of 47 constructs with the correct sequence. After two more rounds of transformation and screening we obtained 42 out of 47 constructs (Supplementary table 1). We reused the initial assembly reaction for the second and third round of transformation and colony screening. Two out of four constructs with 7 oligonucleotides were not obtained with the correct sequence, even though both constructs with 8 oligonucleotides were obtained with the correct sequence. We observed background in the transformation, partly due to self-ligation of the vector (32% of all colonies) but also due to the generation of an unknown product during the assembly process (29% of all colonies). Nonetheless, this strategy enabled a convenient one-pot assembly of constructs that would be difficult to make in any other way. We evaluated colony PCR hits and error rates in detail for the constructs obtained in the third round of the transformation (Table 2). To increase transformation efficiencies for these constructs, the assembly reaction mixtures were purified by MinElute columns before transformation. Twelve colonies for each construct were screened with a hit rate of 63% in the colony PCR. For each construct we selected 4 colonies that showed a hit in the colony PCR and sequenced the construct. On average we obtained a correct sequence for 50% of the clones. However the individual rates for each construct varied from 25% to 100% of correct sequences. As general trend longer synthesized stretches contained more errors. This is not surprising as the synthesis of oligonucleotides is error

prone and the errors accumulate with an increasing number of oligonucleotides used for a construct. The Errors observed were: 14 insertions, 5 mutations and 2 deletions; all located in the part corresponding to the synthetized oligonucleotides. Only one truncation and one mutation were detected in the parts introduced as dsDNA PCR products, in the screening of the complete library.

Table 7 Detailed statistics from 9 constructs of the "Assembling one dsDNA insert and the de novo synthesis of a sequence between 30–255 bp". All 9 constructs were obtained in the last round of transformation. The identifiers of the constructs are given in the first column and refer to Supplementary table 8. Percentage of positive clones from 12 colonies when screened by PCR is given in column two. The result from 4 positive clones sent to sequencing is given in the third column. The indication failed corresponds to no sequencing result. The "DNA synthesis length" column specifies the length of sequence synthesized by oligonucleotides in the corresponding construct. The "Errors total" column indicates the number of all errors found per construct, including multiple errors in one construct. The next column reports the error rate per synthesized base pair, the dsDNA parts were not counted towards this error rate. The last column reports what type of errors were found. The bottom row reports the average for column 2, 4 and 7 and the sum for all other columns.

Construct	Hit rate colony PCR	Sequencing result Mutated / Failed / Correct	correct clones	DNA synthesis length	Errors total	Error per bp	Insertions / Deletions / Mutations
Construct 13	58.33%	3/0/1	25.00%	87	5	0.057	4/1/0
Construct 15	41.67%	2/0/2	50.00%	114	2	0.017	2/0/0
Construct 22	83.33%	1/0/3	75.00%	96	2	0.020	2/0/0
Construct 19	66.67%	1/1/2	50.00%	69	1	0.014	1/0/0
Construct 30	66.67%	2/0/2	50.00%	96	3	0.031	1/0/2
Construct 31	83.33%	3/0/1	25.00%	96	5	0.052	3/0/2
Construct 36	41.67%	0/0/4	100.00%	46	0	0	0/0/0
Construct 46	75.00%	1/0/3	75.00%	73	1	0.013	0/0/1
Construct 41	58.33%	2/1/1	25.00%	52	9	0.173	1/8/0
Total	63%	15/2/19	52%	729	21	0.03	14/2/5

In situ generation of overlaps from oligonucleotides

After observing that sequence synthesis from oligonucleotides and DNA assembly can be done simultaneously in one-step isothermal assembly master mix, we investigated whether the overlaps needed for assembly could be added simultaneously *in situ*.

To synthesize *in situ* the overlaps necessary for the enzymatic assembly, we added 4 different oligonucleotides to the enzymatic assembly mix. Each oligonucleotide overlaps 20 bp with the insert (*DnaA* gene, PCR amplified using *E. coli* genome as template) and 20 bp with the linearized vector pETM14(Dümmler et al., 2005) (Figure

23 a) (GenBank accession number: KC816624). After transforming Top 10 *E. coli* competent cells with 1 µl of the assembly mix, 22 colonies were obtained. After purification and sequencing of plasmid DNA from 12 randomly picked colonies, 4 contained the correct DnaA insert.

In the negative control (transformed cells with the digested vector, plus insert without oligonucleotides) we obtained 3 colonies. We purified and sequenced the plasmidic DNA of two of those 3 colonies. In one case we found a truncated version of the backbone vector and in the other both sequencing reactions failed. Confirming that, the oligonucleotides are the key component of the *in situ* enzymatic assembly.

We obtained unusual low yields of plasmidic DNA when purifying the DnaA constructs form bacterial cultures suggesting that the DnaA protein, which is involved in the chromosome replication(Messer et al., 2001), could have a toxic effect on the host when cloned in a high copy number vector. This could be one factor leading to the high degree of background in this specific experiment.

As a general trend, we observed that 4 oligonucleotides produced more colonies than 2 (Figure 23c). The optimal concentration of oligonucleotides in the final master mix was around 45 nM (Figure 23b). Very high concentrations of oligonucleotides (>150 nM) inhibited the DNA assembly step, as well as, reduced the efficiency of the transformation. Overall, we demonstrated that it is possible to *in situ* synthesize overlaps in a one-step isothermal assembly reaction, and we found robust parameters for this assembly. This facilitates the high-throughput cloning of the same fragment into many non-standard vectors by only exchanging the stitching oligonucleotides.

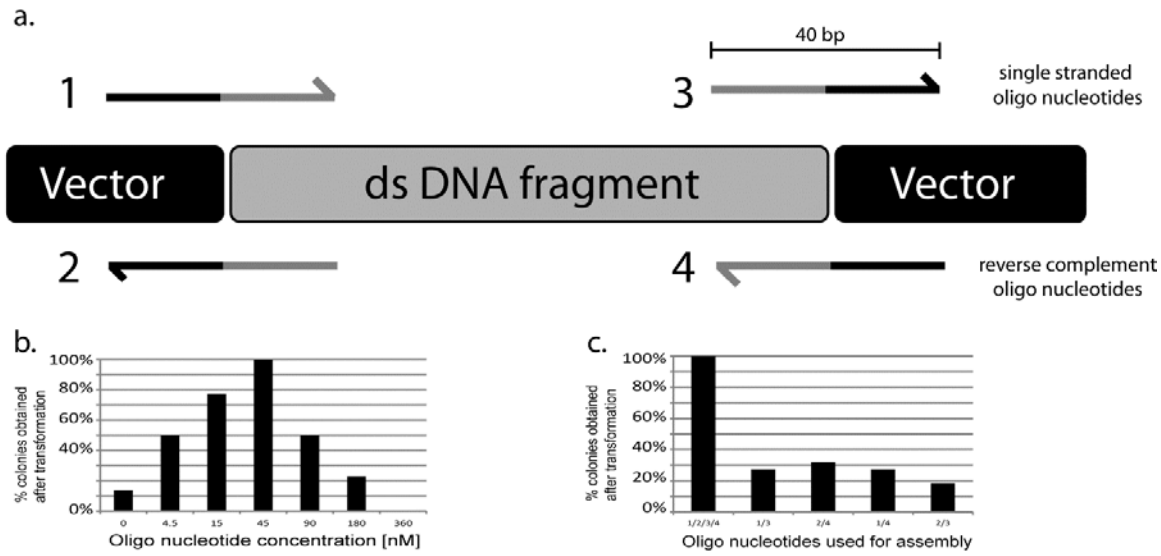


Figure 23 a.) Schematic representation of the oligonucleotides used for the *in situ* generation of overlaps. The oligonucleotides are shown in respect to the assembled ds DNA fragments. The bar indicates the size of the oligonucleotides (they are not in scale with the overlapping regions). b.) Effect of oligonucleotide concentration on the number of colonies obtained after transformation c.) Number of colonies obtained depending on different combinations of oligonucleotides. The oligonucleotides names corresponds to those in scheme a.).

Assembly of two inserts by *in situ* generation of overlaps and *de novo* assembly of a promoter and RBS (156 bp)

After demonstrating that the *in situ* synthesis of overlaps from oligonucleotides is possible and that it can be combined with the assembly of ds DNA fragments, we tested whether it is possible to combine the two methods and clone multiple fragments (Figure 24). For this, we chose to clone the importin-α (1.6 kb) and importin-β (2.6 kb) genes in the pCDF-Duet vector, while building up a 156 bp spacer in between them, comprising a T7 promoter and a modified RBS (GenBank accession number: KC816625). To achieve this, three different oligonucleotide concentrations were tested in the final master mix: 550 nM, 55 nM and 5.5 nM. As negative controls, we used a one-step isothermal assembly master mix without Taq ligase and one complete master mix missing the oligonucleotides (Supplementary table 4). We screened 12 colonies of the assembly mix with 55 nM oligonucleotides for the presence of both inserts and found 8 colonies with a PCR signal for importin-α and 4 with importin-β. Three of the colonies were positive for both inserts. The plasmidic DNAs of double positive clones were purified and no errors were detected after sequencing.

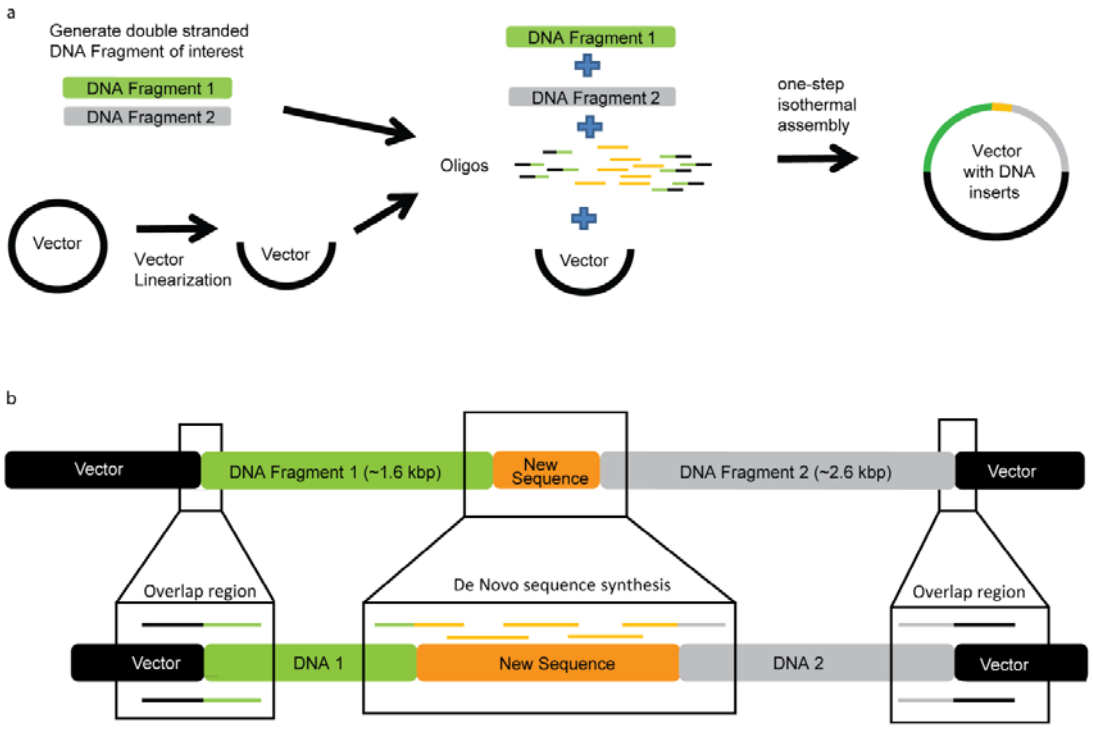


Figure 24 Schematic workflow and oligonucleotide design a.) The general workflow. Ds DNA fragments and a linearized vector are obtained. Subsequently, the fragments, the vector and the oligonucleotides are added to the one-step isothermal assembly master mix b.) Distribution of the oligonucleotides on the construct for the assembly of importin-α (1.6 kb) and importin-β (2.6 kb) genes into the pCDF-Duet vector. A 156 bp spacer was synthesized in between them, comprising a T7 promoter and a modified RBS. The overlap between the individual oligonucleotides and the vector and DNA fragments is 20 bp.

We obtained 26 colonies in the assembly of the negative control without oligonucleotides and 10 in the negative control without ligase. Sequencing results showed that the background is result of misprimed product of the PCR linearization of the backbone. This product produced two homologous ends that were able to combine during the reaction. Such by-products can be easily eliminated by PCR screening and do not affect the overall convenience of this method.

We have shown that the synthesis of sequences from single stranded oligonucleotides can be combined with the assembly of DNA fragments into a vector. We also demonstrated that the overlaps needed for DNA assembly can be synthesized *in situ* during the assembly. By combining these two approaches we could assemble, in a one-step reaction, 2 genes flanking a *de novo* built up sequence.

This technique is also advantageous for individual constructs when *de novo* sequences have to be constructed which are too large to be introduced in a primer. For example, with our technique a secretion signal of 30 amino acids or more can be easily

synthesized in a one-step cloning reaction. By only exchanging individual oligonucleotides, mutations can be introduced in the amino acid sequence to investigate their effect on the phenotype.

The *in situ* synthesis of overlaps has multiple applications. First, it can be used to subclone ds DNA fragments with the versatility of one-step isothermal assembly cloning, while omitting the need for any PCR reaction. Another application is in the generation of libraries for expression screening in which two genes need to be cloned in many vectors, while varying the order of the genes, the promoters and optimized RBS sites. The order of the genes can be easily changed in the *in situ* assembly of the overlaps by only exchanging the overlapping oligonucleotides. In the same way, a set of RBS sites can be screened or new promoters can be tested. Compared to a standard cloning, the advantages of this technique become greater with increasing sample numbers, as oligonucleotides can be reused between constructs.

We evaluated the error rates and location of errors for a small subset of the cloned library. The error rate per base pair for the synthesized stretches is about a magnitude larger than the error rate reported for the underlying technique by Gibson et al.(Gibson et al., 2010) While we cannot directly explain this, it is worth noting the oligonucleotides used came from a different vendor then in the original study (Sigma in our case and IDT in the case of Gibson et al.(Gibson et al., 2010)).

In conclusion, we have generated a library of 42 fusion genes in single-step reactions by using a new approach based on the one-step isothermal assembly cloning technique. This avoids a large number of PCR amplifications and multiple intermediate cloning steps. We show that sequences between 30-255 bp can be synthesized during the assembly of a vector with a PCR fragment of the size range between 500-2700 bp. The design and assembly of the constructs was very easy compared to other state of the art methods and is compatible with automation. Therefore we anticipate that this simple approach, when combined with rigorous screening methods, can be easily implemented in many applications and could be widely used in both molecular and Synthetic Biology.

MATERIAL AND METHODS

DNA Fragment Preparation

The vectors were linearized by inverted PCR or with a restriction enzyme (EcoRV-HF, New England laboratories). PCR fragments were amplified by using either Phusion (Finnzymes) or Kod (Merck Millipore) High fidelity polymerases. A list of all primers used to generate PCR fragments and the amplified sequences can be found in the supplementary information. Vector and inserts were purified using the Qiagen MinElute Kit or Qiagen Gel purification Kit following the instructions of the manufacturer.

All oligonucleotide sequences, the combinations and vectors used in this study are summarized in the Supplementary table 8Supplementary table 1Supplementary table 2Supplementary table 3. The oligonucleotides were ordered from Sigma as desalted (Desalt) or reverse phase purified (RP).

Oligonucleotide design for de novo synthesis

All oligonucleotides were designed to be around 60 bp with a 20 bp overlap to the next oligonucleotide, a dsDNA fragment or the Vector. Properties like melting temperature or secondary structure were ignored for the design process as described earlier(Gibson et al., 2010). If the length of the synthesized stretch made it necessary individual oligonucleotides were designed shorter or longer than 60bp and the overlapping region was extended. A list of all oligonucleotides can be found in Supplementary table 2

Assemblies

The assemblies were done using a modified version of the protocols previously described or the manual for the “Gibson Assembly Master Mix” available at neb.com.(Gibson et al., 2009, 2010) Briefly, oligonucleotides were mixed in equimolar ratios from a 100 µM stock. Then, the mixture was heated to 95°C for 5 min and slowly cooled to 4°C with a ramp of 0.1°C /s. The oligonucleotide mixture was diluted to a concentration of 900 nM and 1µl was added to a master mix for Isothermal assembly resulting in a final concentration of 45 nM of each oligonucleotide.

The PCR fragments were added either in equimolar amounts or as a twofold excess of the vector (50-100 ng). Unless otherwise stated stitching oligonucleotides were also added to a concentration of 45 nM, in a final volume of 20 µl as previously

described.(Gibson et al., 2009, 2010) In contrast to the original protocol the isothermal assembling was done at 45°C for 2 h instead of 50°C and 1 h. This modification is based on previous observations reported that an incubation at 45°C for 1h increases the efficiency of the assembly compared to 50°C and that an incubation at for 2h at 50°C also yields more clones that 1h at 50°C(Gibson et al., 2010). We confirmed the assumption that 2h at 45°C are superior to 1h or 2h at 50°C by a pilot experiment (data not shown). The combinations of oligonucleotides and PCR fragments for each construct are listed in Supplementary table 1.

Bacterial Transformation

Top10 (life technologies) or DH5α *E. coli* cells made competent with Z-Competent Kit (Zymo research) were used for the transformations following instructions of the manufacturer. The first transformation round for assembling one double stranded insert and the de novo synthesis of a sequence between 30–255 bp was done with DH5 cells in combination with the Z-competent Kit. The manufacturer gives a transformation efficiency for this cells 10^8 - 10^9 transformants/µg of plasmid DNA. The 2nd round was performed with Top 10 cells which have a transformation efficiency of 10^9 transformants/µg of plasmid DNA. In the third round also Top 10 cells were used but the assembly reaction mixture was purified by a min Elute column (Qiagen) before transformation and 5ul of the elution was used for transformation. The removal of PEG8000 and various proteins increases the yields of colonies significantly.

The transformations for all other described experiments were done by adding 1 µl of the assembly master mix to Top 10 cells and following the instruction of the manufacturer. Selection on LB plates was performed with the antibiotic required by the vector.

Screening for full-length inserts assembled from oligonucleotides and non-overlapping dsDNA fragments

To screen for the successful assembly of our constructs we performed colony PCR. In the case of the fusion library, we picked 8 colonies from each transformation and screened using the Primers “Fwd Screen Puro” and “Rev Screen Puro”. For each positive construct, 4 colonies were picked and prepared using the Millipore Montage Plasmid Miniprep HTS 96 Kit. If fewer than 4 colonies were positive in colony PCR,

new unscreened colonies were selected and the plasmid purified. All purified plasmids were sequenced by Sanger sequencing (GATC, Germany).

The assembly in the pCDF_Duet Vector was screened by PCR with the primers used to generate the inserts. Colonies giving a positive signal for the inserts were purified using a Qiagen Miniprep Kit and sent for sequencing. All primer sequences used for screening are listed in Supplementary table 1.

In silico construct design

The constructs were designed using the gene designer software and the CLC Main workbench.(Villalobos et al., 2006)

Supporting Information

Supplementary material is available free of charge via the internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*Ph: + 34933160259, Email: bernhard.paetzold@crg.eu, Correspondence may also be addressed to maria.lluch@crg.eu

Author contributions:

B.P. wrote the manuscript and designed the experiments. B.P, C.C. and T.F. contributed experimental data. M.L.S. and L.S. provided supervision and feedback; the work was performed in the lab of L.S.

Notes:

The authors declare no competing financial interest.

ACKNOWLEDGEMENT

The authors would like to thank Mark Isalan for his critical reading of the manuscript and Sira Martinez for her technical support.

This work was supported by the European Research council (ERC) advanced grant GA number 232913, the Fundacion Marcelino Botin, the Spanish Ministerio de Economía y Competitividad BIO2007-61762, the ISCIII, Subdirección General de evaluación y fomento de la investigación PI10/01702 to the ICREA researcher LS. Funding for open access charge: European Research council (ERC) advanced grant GA number 232913

ABBREVIATIONS

SLIC, Sequence- and Ligation-Independent Cloning; RBS, ribosomal binding site; PCR; Polymerase Chain Reaction; *E. coli*, Escherichia coli; dsDNA, dsDNA. bp: base pair;

Engineering *Mycoplasma pneumoniae* as therapeutic vector for lung diseases

Contributions:

Part of the work presented in this chapter was carried out by the Sanofi exploratory Unit in Montpellier and Toulouse. The quantification and characterisation of Alpha-1-antitrypsin has been completely done by Sanofi Montpellier. Also the characterisation of the secretion in the lung of mice and conditional delivery were performed in Mice. I would like to acknowledge in Montpellier especially Jean Bernard Ferrini but also Regine Floutard, Julie Izac, Nathalie del-Negro and Omar Jbilo. The *P. aeruginosa* biofilm assay was developed and all tests were performed at the Sanofi Sanofi Exploratory Unit in Toulouse by Richard Legoux, Anne-Claude Noirot-Jagu and Mourad Kaghad. Both projects were sponsored and managed by Edgardo Ferran and Luis Serrano

Abstract

Huge advances have been made in the last decade of biomedical research, yet many unmet medical needs still exist and cannot be addressed by classical therapy approaches. Synthetic biology holds the potential to design new living pills that could close a part of this gap. However, real proofs of concepts for biomedical applications of synthetic biology are rare. Here, we show that it is possible to engineer the bacterium *Mycoplasma pneumoniae*, a model organism for systems biology to produce and secrete 3 different therapeutic proteins. The three chosen proteins have a varying level of complexity in regards to the production and secretion in a bacterium. We showed for two out of the three proteins that they are active after secretion. We further verified that one of them is also secreted, when *M. pneumoniae* is in the lung of mice. In a next step, we designed a conditional delivery system that releases a protein cargo only in the presence of proteases associated with inflammation. In summary, we could establish a proof of concept for the engineering of *M. pneumoniae* as living pill.

Introduction

Synthetic biology is a young discipline that aims the complex engineering of living systems. It is fuelled by the recent advances in DNA synthesis and DNA assembly techniques (Gibson et al., 2008, 2009). Possible applications range from the large scale production of biofuels to the synthesis of fine chemicals by engineering of metabolic pathways. Another application of these new possibilities is the engineering of genetic circuits like electronic ones (Menolascina et al., 2012; Wang et al., 2012a). These circuits can make simple decisions based on the environment and could serve as the base for so called smart drugs. One example of such a genetic circuit would be the conditional production and secretion of a toxin once a pathogenic cell is detected (Gupta et al., 2013; Saeidi et al., 2011). In the state of the art, the design of such systems is done by combinatorial testing, as most genetic circuits behave unpredictably once assembled (Kosuri et al., 2013). A number of groups work on standardizing parts for reliable and predictable protein expression (Mutalik et al., 2013a, 2013b). However, the predictiveness of these systems is still low and the success rate for the *in silico* prediction of *in vivo* behaviour of a genetic circuit is close to random. The main drawback for the successful prediction of the behaviour of a genetic circuit, is the complexity of the host. The genetic and metabolic networks of the host, bacterial or mammalian cells are too big to be fully simulated in a computer and can therefore not be completely accounted in the design process. One possible solution to this problem is to use “minimal” bacteria with a smaller genetic and metabolic network. These bacteria can be better understood and finally modelled in a computer(Karr et al., 2012; Purcell et al., 2013), so that individual parts and circuits can be tested *in silico* before actually engineered in reality.

M. pneumoniae is one of the smallest self-replicating organisms. It has been well characterized by genome, transcriptome, proteome and metabolome studies and it is considered as a model organism for systems biology (Güell et al., 2009; Kühner et al., 2009; Maier et al., 2011, 2013, 2013; Yus et al., 2009). Further *M. pneumoniae* lacks a cell-wall and it is well studied in terms of virulence and pathogenicity (Hansen et al., 1981; Kannan and Baseman, 2006; Krause et al., 1982). Strains with reduced pathogenicity were identified by Layh-Schmitt et al., (1995) and Shimizu et al., (2011). These traits will make *M. pneumoniae* in the long term a suitable chassis for complex synthetic biology applications, especially in the field of smart drugs.

While *M. pneumoniae* was studied intensively at level of different “omics”, no comprehensive study has been done on the secretome of the bacteria. However, we need a detailed quantitative understanding of the secretion signals used by the bacteria to engineer *M. pneumoniae* for delivering therapeutic proteins.

To identify the secretion signals used by a bacteria, the secretome which represents all secreted proteins can be determined *in silico* and experimentally (Bendtsen et al., 2005; Hathout, 2007; Manoil and Beckwith, 1985; Meissner et al., 2013; Petersen et al., 2011; Tjalsma et al., 2000; Tolonen et al., 2011). The *in silico* analysis allows to distinguish the transport mechanism and identify putative signal peptides, while the experimentally analysis by quantitative mass spec analysis allows to identify unkown secretion systems as well as quantify how efficiently a secretion signal is working.

Three main pathways for secretion exist in bacteria: the Sec pathway, the twin-arginine translocation (Tat) pathway and unconventional secretion (Dalbey and Kuhn, 2012). In *M. pneumoniae* no evidence for the existence of the Tat pathway or its associated machinery was found. However, bioinformatic evidences for the sec export pathway (Dandekar et al., 2000) and experiments suggesting unconventional secretion have been reported (Balasubramanian et al., 2008; Dallo et al., 2002; Hegermann et al., 2008).

To gain insight into the secretion system of *M. pneumoniae* and to ultimately use it as chassis to deliver therapeutic proteins, we studied its secretome by combining *in silico* and experimental approaches. Based on these results we designed a battery of vectors that allow the expression and secretion of three different proteins with different biochemical properties and potential therapeutic applications in lung diseases. The tested therapeutic proteins were a poly M alginate lyase A1-III, Alpha-1-antitrypsin (A1AT) and p53.

The alginate lyase A1-III was chosen because it is a monomeric bacterial protein, which is secreted in its natural context. It is easy to track in secreted medium (Kitamikado et al., 1992) and has a potential as a therapeutic protein against chronic *P. aeruginosa* lung infections (Alipour et al., 2009; Eftekhar and Speert, 1988; Hatch and Schiller, 1998).

The second protein is A1AT it is a human protein, that is glycosylated in the natural context but functional recombinant version have been produced in bacteria. It is usually secreted by mammalian cells. The structural analysis revealed that the protein fold is under strain and major rearrangements occur upon cleavage by its target neutrophil elastase. A1AT plays a key role in Chronic Obstructive Pulmonary Disease (COPD) which is an inflammatory disease of the lung affecting millions of people and leading

progressively to a respiratory deficiency and finally death. The complex and strained structure of A1AT made it uncertain if the protein can be produced by *M. pneumoniae* or secreted through the sec pathway requiring the protein to be unfolded when passing the membrane. Normal blood levels of A1AT are very high and in the mg/ml range. However, it was recently shown that a targeted delivery of A1AT to the surface of neutrophils might have enhanced therapeutic effects (Gehrig et al., 2012). This makes A1AT a very interesting target for a proof of concept application of targeted protein delivery.

Among several human therapeutic gene-products that could be engineered in *M. pneumoniae* for intracellular delivery, we have chosen p53 because its functional role is well documented (Zilfou and Lowe, 2009),,an extensive experimental toolbox has been developed and it is a tertameric protein. It is further tightly regulated on the post translational level, which should reduce the unwanted interference with the host p53 dynamics.

Overall, the three proteins represent different challenges on the protein production and secretion machinery. We have an enzyme (algynate lyase) , an strained fold (A1AT) and a tetrameric protein, thus covering many of complex challenges for both production and secretion of a functional version of the protein.

Results and discussion

Our aim was to produce and secrete therapeutic proteins for lung diseases in *M. pneumoniae*. First we analysed the secretome of *M. pneumoniae*. We then used this knowledge to build a series of vectors to produce and secrete recombinant therapeutic proteins in *M. pneumoniae* (Figure 25). Normal *M. pneumoniae* media is rich in proteins which interfere with MS analysis. Therefore we used the minimal media of *M. pneumoniae* as basic growth media for our experiment (Yus et al., 2009) and replaced bovine serum albumin (BSA) as lipid carrier with (2-hydroxy)propyl-β-cyclodextrin (Hyprop).

After growing *M. pneumoniae* cells during 24h and 72h the supernatant media was harvested and ratios of intracellular to extracellular protein concentrations were obtained by dimethyl labelling and mass spectroscopy analyisis. A similar strategy was already earlier described by Tolonen et al., (2011). On average we obtained ratios for

281 out of 688 proteins from *M. pneumoniae* per experiment. All obtained data is summarized in the Supplementary table 9.

The proteins with highest extracellular to intracellular ratios were further analysed by the algorithm Signal P 3.0. Based on the predicted secretion signals we designed 11 vectors comprising a promoter and a secretion signal fused to the therapeutic proteins. As negative control we fused the 50 N-terminal residues of a cytosolic protein (Mpn332) to the therapeutic proteins (Table 3). As promoter we choose either the promoter corresponding to the coding sequence of the secretion signal or if it was part of an operon with a distal promoter we put it under the control of the EfTu promoter. A detailed list of all constructs made, indicating the signal peptide and promoters can be found in Supplementary table 6 for the alginate lyase, Supplementary table 7 for the A1AT constructs and Supplementary table 8 for p53. The tables also indicate whether the secretion signals are predicted to belong membrane anchored lipoproteins (SpII) or sec mediated secreted proteins (SpI).

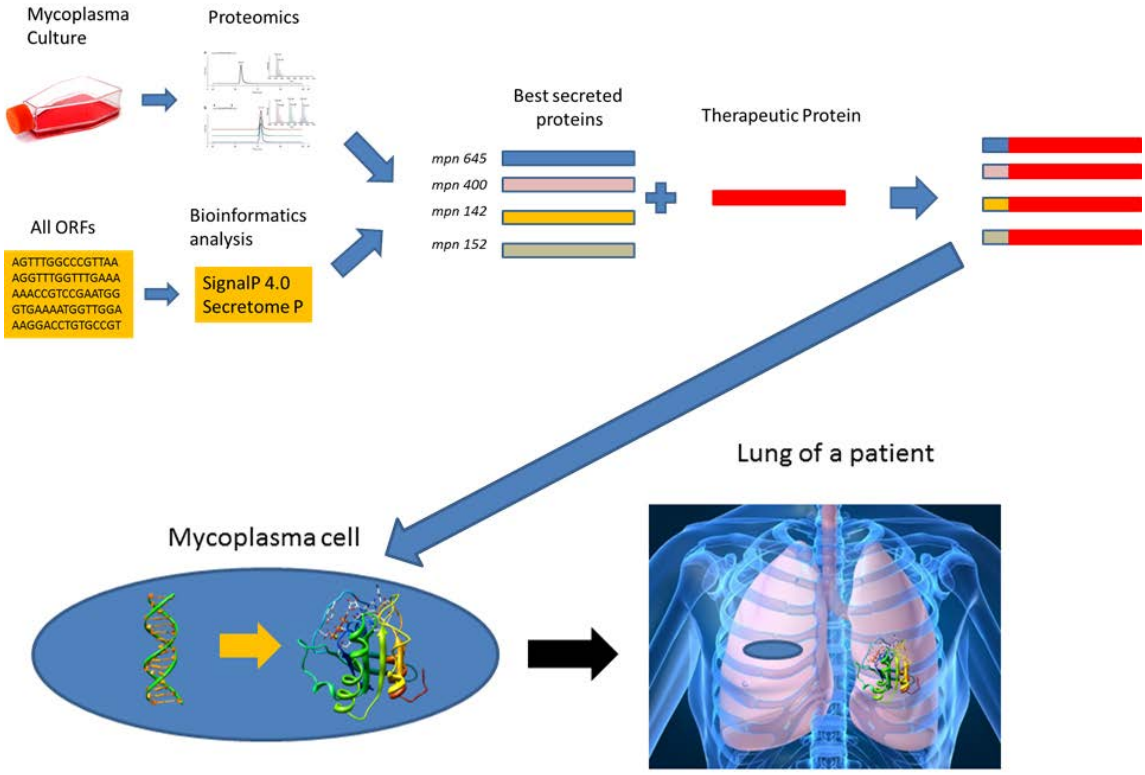


Figure 25 Schematic overview of the strategy to obtain a first proof of concept application. The steps undertaken in this thesis are marked with blue arrows. With a black arrow is marked the last step for clinical tests which are beyond the scope of this study.

We cloned the three therapeutic proteins fused to the secretion peptide and the respective promoter in a miniTn4001-Puro-1 vector (GenBank accession number: KC816623). The constructs were assembled and scaled up in *E. coli* before

transformation in *M. pneumoniae*. We determined the levels of the secreted protein either by an activity assay for the alginate lyase constructs or by ELISA measurements for A1AT and p53. In all cases, the constructs under the control of the EfTu promoter showed the highest level of secreted protein, while no signal was detected for the negative control.

Table 8 Overview of the secretion signals used to construct secretion vector. The "Accession" column gives the *M. pneumoniae* gene identifier. The following columns indicate the positions of the signal peptide cleavage site predicted by the SignalP 3.0 sever. NN give the position for the Neuronal Network score and HMM is the site obtained by the hidden Markov model. We defined the signal peptide according to the highest predicted value and included in the cloning 5 more Amino acids upstream.

Accession	NN	HMM	signal sequence
Mpn142	25-26	25-26	MKSKLKLKRYLLFLPLLPLGTLSLANTY
Mpn142	25-26	25-26	MKSKLKLKRYLLFLPLLPLGTLSLANTY
Mpn152	21-22	27-28	MKFKYGAIVFSGLLGVSAILAACGARGKFN
Mpn200	21-22	22-23	MKFKYGAIVFSGLLGVSAILAACGT
Mpn213	28-29	27-28	MKLSAIISLSVAGTVGTTAVVVPTTITLVNK
Mpn332	Negativ control		MPAVKKPQILVVRNQVIFPYNGFELDVGRERSKKLIKALKNLKTKRLVLV
Mpn400	35-36	35-36	MKLNFKIKDKKTLKRLKKGGFWALGLFGAAINAFSAVL
Mpn489	28-29	28-29	MGYKLKRWPLVAFTFTGIGLGVVLAACSALN
Mpn506	21-22	22-23	MKFKYGAIFFSGFLGLSAILAACGT
Mpn588	24-25	24-25	MRLQFKLLGFLTLLGTSTILSACAATQ
Mpn592	24-25	30-31	MGFKLKGFGFLTLFASQAFLTACSATLTVANTN
Mpn645	25-26	25-26	MKLKLKFLLISLLGSSLLLSACSSAATQ

Quantitative analysis of the secreted p53 by *M. pneumoniae*

After transforming *M. pneumoniae* cells with each of the engineered p53 constructs, p53 secretion was verified by using a p53 pan sandwich ELISA (Roche). To assess the dynamics of p53 accumulation in the supernatant, the absolute concentrations of p53 in the media was measured at various time points after inoculation. We characterized protein secretion both in modified Hayflick (Full Media) and McCoy media which is commonly used for mammalian cell culture. The results for the secretion in Hayflick media are summarized in Supplementary figure 4 and the results for secretion McCoy media are summarized in Figure 26. The constructs Mpn-142 and Mpn-645 under control of the Eftu promoter showed the highest concentration of p53 in the supernatant media. Interestingly the p53 concentration in the supernatant for most constructs peaked at 48h when the cells enter into stationary phase and then slowly dropped again. This could either indicate either a regulation of secretion or gene expression which is linked to the switch from exponential to stationary phases of growth. Also, a balance between secretion and extracellular breakdown by proteases could explain this result.

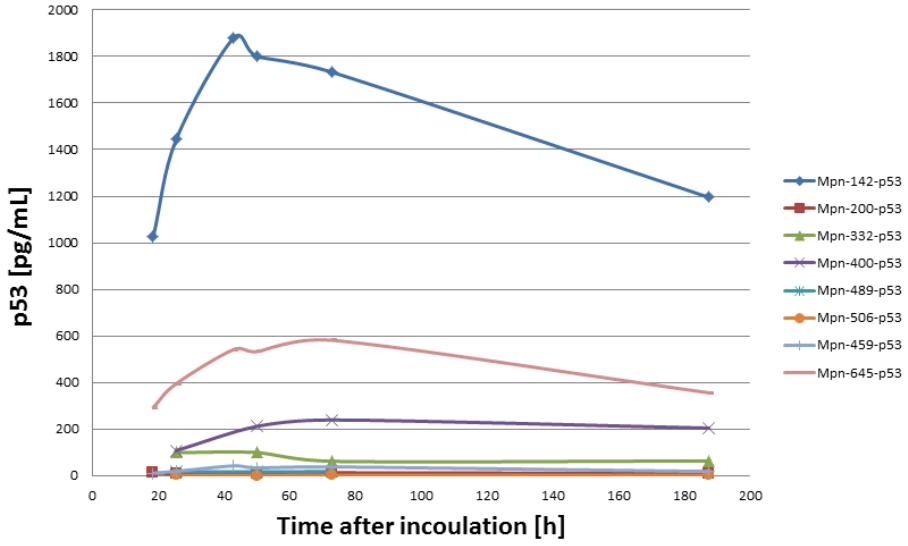


Figure 26 Results for the secretion of p53 by modified *M. pneumoniae* into McCoy Media. The amounts were determined at different timepoints using a Roche p53 pan ELISA sandwich assay.

Quantitative and functional analysis of the secreted alginate lyase A1-III by *M. pneumoniae*

We further analysed the secretion and activity of alginate lyase in *M. pneumoniae*. First the gene was cloned in our secretion constructs and then transformed into *M. pneumoniae*. To evaluate the secretion efficiency we used an assay for alginate lyase

activity. As with the p53 constructs the two constructs with the alginate lyase under control of the EfTu promoter and with the secretion signals of Mpn-142 and Mpn-645 performed the best. Both showed degradation of the alginate in the media already at day 1 (Figure 27). While the majority of the constructs showed no sign of alginate degradation even after 5 days of incubation. The constructs with the secretion signals of Mpn-200 and Mpn-400 which are both under their respective promoter showed degradation of alginate at day 2 and the construct with the secretion signal of Mpn-489 and its respective promoter showed degradation only at day 5 after inoculation. The negative control which is fused to the 50 N-terminal amino acids of the cytosolic protein Mpn-332 showed no sign of degradation. No differences in growth rate of different *M. pneumoniae* strains were observed in comparison with wt (data not shown) indicating that the promoter activity is a determining factor for the production and secretion of alginate lyase.

In order to test if this system could be implemented in a less immunogenic strain, we tested whether the secretion constructs also work in a non-adherent strain (Supplied by Professor Herman). This strain lacks part of the adhesion and gliding machinery, main factors in virulence and pathogenicity (Layh-Schmitt et al., 1995). The results obtained in the non-adherent strain matched up with the results obtained in wt M129 strain (Supplementary figure 5) implying that the same secretion vectors can be used in non-pathogenic strain.

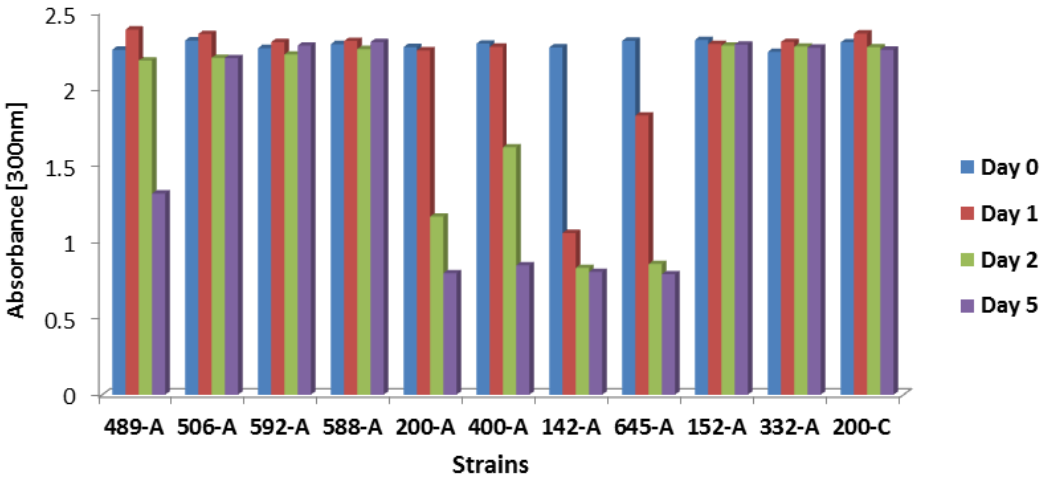


Figure 27 Results from alginate lyase activity assay in full medium supernatant from cultures secreting alginate lyase (M129 strain). All cultures were normalized to the same inoculation count at day 0 and over the next 5 days samples of the media were taken. The assay is based on the degradation of polymeric alginate to shorter sugar fragments.

In order to improve the levels of protein production, we designed a new Mpn-142 secretion signal (Mpn-142(Opt)) with a minimized secondary structure in the 5' of the mRNA. The construct was still under the control of the EfTu promoter but the codons for the secretion signal were changed from the wt sequence following the recommendations of the company DNA 2.0. We had observed earlier that *M. pneumoniae* grows better in Hayflick (full) media than in our modified minimal media used to characterize the secretome. To quantify the maximal amounts of alginate lyase produced by the different strains, we thus established a quantitative assay compatible with full media based on the previously used qualitative assay (Kitamikado et al., 1990). The absolute alginate lyase activity from the media of 2 strains was determined at different time points (Supplementary figure 6 andSupplementary figure 7). The highest activity levels was observed for the Mpn-142(Opt) strain (Supplementary figure 6). In the media of the *M. pneumoniae* strain transformed with Mpn-142(Opt) we measured an activity corresponding to ~0.1 Units of alginate lyase. This corresponds to ~0.01mg/ml of the alginate lyase (Sigma A1603) which we used as standard in our quantitative assay.

We finally tested supernatants of the strains Mpn-142(Opt) and Mpn-142 in a *P. aeruginosa* biofilm assay. We saw no anti biofilm activity while a positive control of media containing the same amount of alginate lyase activity, of a commercial alginate lyase (Sigma A1603) showed antibiofilm activity. We extensively troubleshot this problem and the results of this investigation will be reported in a separate section of this thesis.

It has been reported that DNAse enhances the breakdown of *P. aeruginosa* biofilms (Alipour et al., 2009). Based on this study we investigated whether *M. pneumoniae* M129 media supernatant contains Dnase activity that could already supply this function. For this propose, we measured lambda phage DNA degradation in the media from a M129 WT culture. After 1h of incubation at 37°C most of the DNA was digested indicating a strong Dnase activity in the supernatant of *M. pneumoniae* M129 wt supernatants (Supplementary figure 8). Only minor signs of degradation were observed in the control reaction corresponding to the media that had not been in contact with cells.

Quantitative analysis of the secreted A1AT by *M. pneumoniae*

In order to test the activity of AA1T secreted by *M. pneumoniae*, first we measured the protein expression level in the supernatant for each strain using an ELISA kit from USCN. As shown in Figure 28a, secretion of AAT is better using full medium. The best secretion levels are obtained, as in the previous cases, from the Mpn-142 and Mpn-645 strains with a yield respectively of 250 ng/ml and 130 ng/ml using the full medium and 152 ng/ml and 20 ng/ml using the minimal medium after 73h of culture. As with p53 we observed that the protein concentration in the supernatant peaked after 2-3 days and then decreased (Figure 28a). We tested whether the degradation of the secreted protein by an extracellular protease is the mechanism causing this peak. Therefore, we spiked supernatant of a wt *M. pneumoniae* culture grown either in minimal or full Hayflick media with 300nM A1AT and incubated it for 96h. We measured the inhibitory effect on neutrophil elastase at different time points (Supplementary figure 10). All sample showed a constant maximal inhibition indicating no significant degradation in the supernatant over time. We performed a similar test with BSA in supernatant and could also observe no proteolytic effect on a SDS Gel (data not shown).

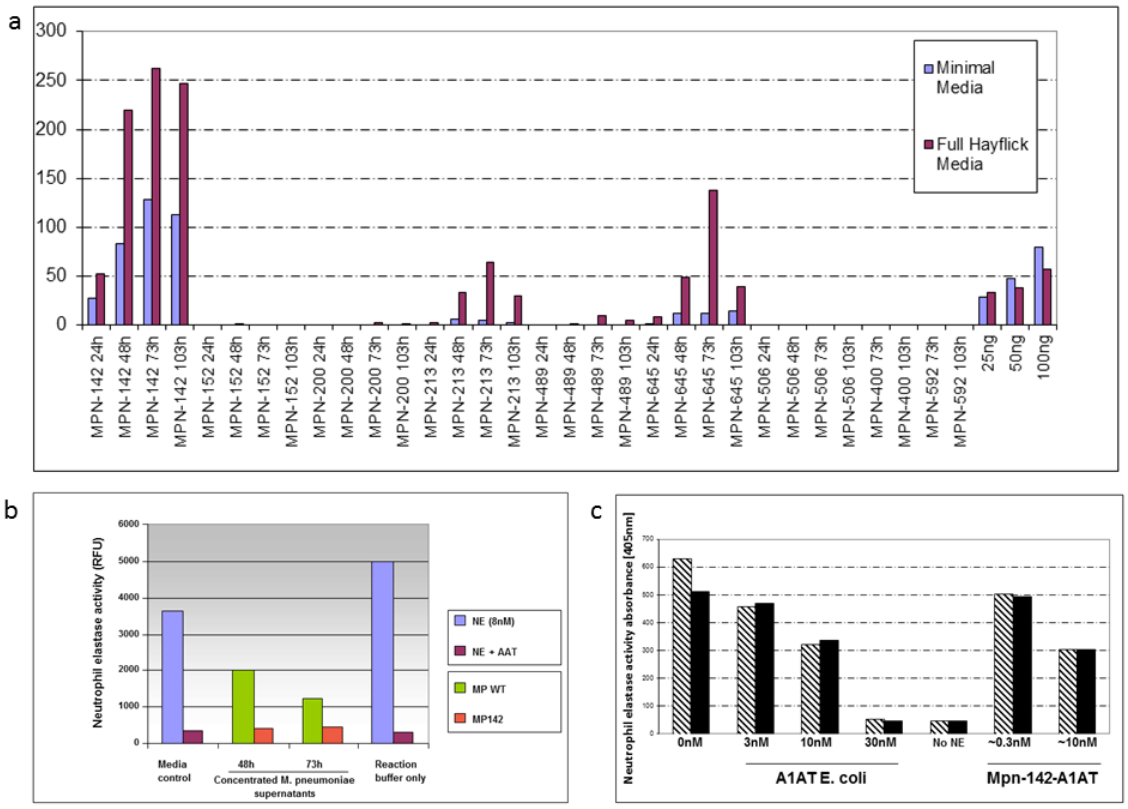


Figure 28 (a) shows the quantitative analysis using ELISA of secreted AAT by the different strains of *M. pneumoniae* cultivated in the minimal or full medium. (b) shows a neutrophil elastase activity assays using concentrated minimal media supernatant culture from WT and MP142 strains. (c) shows a neutrophil elastase activity assays using purified AAT from MP142 strain and purified AAT from *E. Coli* (Acrys).

Then we analyzed if the secreted A1AT was functional. Unfortunately the full Hayflick media which yields the highest proteins concentrations also contains high level of Horse A1AT. This Horse A1AT interferes with functional neutrophil elastase assays while not showing up in the ELISA measurements (Supplementary figure 9). The amounts of A1AT secreted in the minimal media is below the minimum concentration necessary for the functional assay 10 to 30 nM (500ng/ml – 1500ng/ml) (Figure 28a). To reach the concentration of 10 to 30 nM of AAT in the supernatant of Mpn-142 culture, we have concentrated the samples from minimal media using ultrafiltration with an ultra-30K device (Amicon). As shown in Figure 28b, concentrated supernatant showed a complete inhibition of NE activity compared to WT concentrated supernatant. However, we observed a decrease of NE activity in the minimal media alone in comparison with our control reaction in buffer only suggesting unspecific inhibition of NE activity following the volume reduction.

To further verify the activity of AAT secreted by *M. pneumoniae*, a purification process was performed. We used a Strep Tag II which was introduced at C-terminus of the protein in conjunction with a Strep Tactin column to enrich the protein. The binding of the protein to the column was weak most likely because of the low concentration of A1AT in the media compared to the Kd of the Strep Tag II. Neither the less we could significantly enrich and purify the A1AT from the full media. We then tested the purified protein in a Neutrophil elastase activity test and could show that the A1AT produced by *M. pneumoniae* is as active as A1AT produced by *E. coli* (Figure 28c).

In vivo analysis of A1AT secretion by *M. pneumoniae*

We then investigated if our engineered *M. pneumoniae* can secrete A1AT in the lung of mice or if the secretion is limited to the *in vitro* setting of a culture dish. For this C57BI6 mouse were intranasal infected with the engineered *M. pneumoniae* strain Mpn-142-A1AT. Bronchoalveolar lavage (BAL) was conducted at day 1, 3 and day 6. The BAL fluid was centrifuged and the tested by Western lot for the presence of human A1AT. The A1AT produced by *M. pneumoniae* is recognised by an antibody specific for human A1AT but runs with a lower molecular weight as it lacks the glycosylation produced in humans.

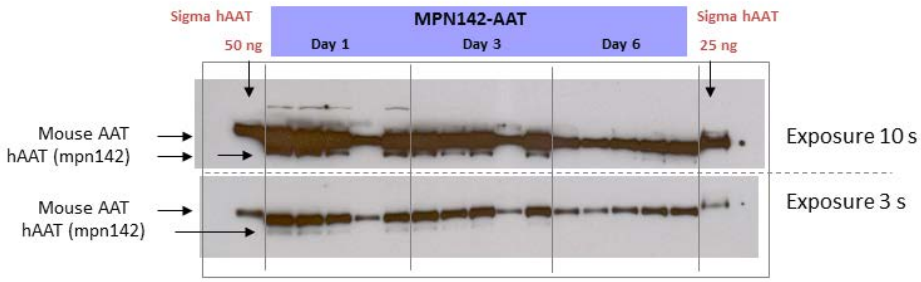


Figure 29 Western Blot on supernatants from BAL of mouse infected with *M. pneumoniae* strain Mpn-142-A1AT. The same image is represented with two different exposure times. A control with human A1AT (Sigma) was loaded at left and the right of the blot. The band corresponding to mouse A1AT is indicated as well as the band corresponding to the human A1AT produced by the engineered *M. pneumoniae* strain. As primary antibody the A1AT antibody of Mybiosource was used.

A quantification of the A1AT in the BAl was not possible as the ELISA Kit cannot detect A1AT after it reacted with Neutrophil elastase. However, in this experiment we showed that our engineered *M. pneumoniae* strains can produce and secrete a therapeutic protein in the lung of a host.

Conditional delivery construct design

It was recently reported that the activity of Neutrophil elastase close to the surface of neutrophils is more important in chronic lung diseases like COPD or Cystic Fibrosis than overall Neutrophil elastase levels (Gehrig et al., 2012). We hypothesized that a triggered release of A1AT close to activated neutrophils would have the same beneficial effect as a very high dose administered globally to the lung of a patient. Therefore, we wanted to improve our *M. pneumoniae* for a conditional delivery of A1AT in the vicinity of activated neutrophils.

Currently, no reliable system for inducible expression is available in *M. pneumoniae*. Therefore, we want to exploit the fact that certain Matrix Metalloproteases (MMP) are co-expressed with Neutrophil elastase by neutrophils during inflammation. We decided to constantly express an A1AT fused to a membrane anchored protein. For the triggered release we included a MMP cleavage site in between the membrane anchor and the A1AT. This system should release A1AT from the *M. pneumoniae* surface if the MMP corresponding to the cleavage site is in the environment. From the MMPs available we choose MMP9 because of its expression pattern and comparable high substrate specificity.

We evaluated different Mycoplasma membrane proteins for their potential as carrier of the A1AT constructs. We choose 3 membrane proteins, P65, P1 and P30 as anchors according based on the criteria:

- Previous reports indicate that a fusion to this protein is stable (P65) (Kenri et al., 2004)
- They are reported to be stable expressed in the non-pathogenic strain M6 (P1) (Krause and Balish, 2001)
- Data is available on the stability of truncated versions (P30).(Chang et al., 2011)

Further, we decided to not only clone the A1AT protein but also included a set of constructs having EYFP fused to the same membrane anchors. The EYFP constructs will allow us to characterize easier the release of the cargo upon exposure to MMP9. We used a modified Gibson cloning approach to construct 12 Vectors. The 12 vectors represent 6 fusions of the carrier with A1AT and 6 fusion of the carrier with EYFP. Each of the 6 fusions corresponds to 3 fusion with MMP9 cleavage site and without (Figure 30a).

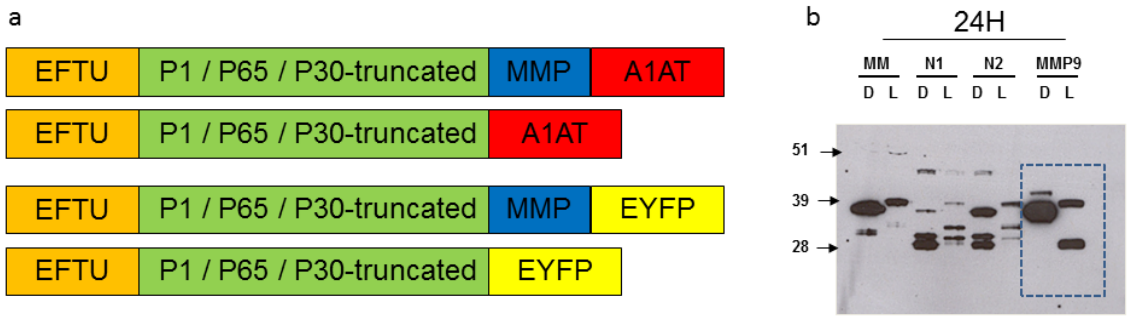


Figure 30 (a) Schematic representation showing the 4 groups of constructs cloned. The orange part is the EfTu promoter that showed high expression levels in previous studies. The green part represents the M. pneumoniae membrane proteins that will be used as carriers. P1 and P65 will be cloned as full length proteins and the cargo will be fused to the C-Terminus. P30 will be cloned as a truncated version with cargo fused to the C-Terminus of the truncated protein. In Blue is the MMP9 cleavage site represented. In Red and yellow is the Cargo with red being the A1AT protein and yellow being EYFP, respectively. (b) Western blot of a M. pneumoniae strains carrying P30-EYFP (D) and P30-MMP9-EYFP (L). The cultures were exposed for 24h to either minimal media (MM), activated neutrophil supernatant from donor 1 (N1) or donor 2 (N2) or recombinant and activated MMP9 (MMP9). The size of the fusion construct is 42kDa (D) and 44kDa (L). The released EYFP has an expected weight of 28kDa

We initially characterized the constructs carrying EYFP using Western blot (Supplementary figure 11) and fluorescence microscopy (data not shown) and observed varying expression levels. Then we tested if EYFP is cleaved from the membrane and released into the supernatant by exposure to either pre-activated recombinant MMP9 or supernatant from Neutrophils which have been stimulated by LPS (Figure 30b). The cultures in the minimal media alone showed little sign of degradation after 24h. The supernatants from both neutrophils showed an unspecific degradation pattern, interestingly releasing by coincidence the free EYFP. The samples of the P30-MMP9-EYFP strains carrying the MMP9 cleavage site showed no or only a very weak band for

the released EYFP. However, the recombinant MMP9 produced a clear signal for free EYFP in the strain P30-MMP9-EYFP while the fusion without the cleavage site produced no signal. We assume that in supernatants of the neutrophils a number of other proteases are active that degrade all surface exposed proteins on the *M. pneumoniae* membrane. This effect seems to be more pronounced than the specific cleavage of the fusion construct by MMP9. The higher weight of the degradation products compared to the free EYFP suggests that the protease in the supernatants cleave the P30 protein which is used as a membrane anchor. If the other constructs which use P1 and P65 as a membrane anchor have the same problem has still to be determined and experiments for this are ongoing.

Conclusion

We could show that it is possible to engineer *M. pneumoniae* to produce and secrete therapeutic proteins. We could produce and detect in the supernatant 3 different proteins, two of which are of human origin (p53/A1AT) and of which only two are normally secreted (alginate lyase/A1AT). We could prove that our secretion vector worked reliable with the same construct yielding the highest yields for all three proteins. Further, we could show that two out of the three proteins are correctly folded and active after secretion. Especially for A1AT it was unclear if the secretion of an active form would be feasible, as the tertiary structure of A1AT is complex and it has to be secreted in unfolded state through the Sec pathway. However, we could show that engineered *M. pneumoniae* strains still secrete their therapeutic cargo in the lung of infected mice. Unfortunately, quantification of A1AT in the BAL was not possible and further experiments to assess the effectiveness and retention of *M. pneumoniae* will be necessary. For future experiments we will have to engineer first a nonpathogenic strain in order to reduce the immune response of the host to the engineered bacteria. We also engineered a conditional delivery system to release A1AT only in the proximity of activated neutrophils. This system is based on the cleavage of a membrane bound A1AT by MMP9. While we could show that recombinant MMP9 is specifically releasing our cargo. We also observed that our membrane bound protein is degraded in supernatants from activated neutrophils. We assume that the anchor protein P30 is degraded by proteases secreted from activated neutrophils. Future experiments will show if the other two membrane anchor P1 and P65 will also be degraded or be stable in the supernatant

of activated neutrophils. Taking all our data together we could prove the suitability of *M. pneumoniae* as chassis for medical applications in human lung diseases..

Material and Methods

Experimental determination of the secretome

M. pneumoniae is usually cultured in modified Hayflick media, a rich media containing many proteins from added horse serum. The highly abundant proteins from the rich media cover the signal from low abundant secreted proteins leaving them unsuitable for mass spectrometric (MS) analysis. Therefore, we used the minimal media of *M. pneumoniae* as growth media for our experiment (Yus et al., 2009). This minimal media still contains bovine serum albumin (BSA) as lipid carrier in high amounts. We replaced BSA with 5mM (2-hydroxy)propyl-β-cyclodextrin (Hyprop) (Sigma H107 CAS Number 128446-35-5) (Greenberg-Ofrath et al., 1993) and could so obtain a protein free Media compatible with downstream MS analysis.

To produce our supernatants we grew wt M129 in normal media for 3 days. Before splitting the cultures were washed twice with PBS while attached and twice after scraping. We then split it 1:10 in a 150cm^2 flask containing 40ml of Hyprop media. We started at each repeat of the experiment two flask one for each time point. The cells were allowed to attach for 24h and were the attached cells were washed twice again with PBS to remove all trace amounts from the horse serum. We then let the cells grow another 72h before removing the supernatant and harvesting the cells of the first flask. The attached cells were resuspended in exactly the same amount of Hyprop media as the removed supernatant. In the other flask only the supernatant was removed and the attached cells were washed twice with PBS before 35ml of fresh media was added. After 24h this flask was harvested as the first one. The cell suspension was always processed identical and in parallel to the supernatants to avoid any bias from experimental procedure on the outcome. The samples were precipitated with 60% acetone (Sigma product # 179124) and 10% trichloracetic acid (TCA) (Sigma product # T9159) as final concentration. The mixture was spun for 1h at 35000g (4°C). The supernatant was discarded and the pellets resuspended in 1.5 ml of TCA/acetone and spun 2h at 16000g (4°C). The supernatant was removed and the pellet was dried completely in a speed vac before being redisolved in a buffer of 8M Urea and 100mM

$NaHCO_3$ using a bioruptor system. The total protein amounts in the samples were determined using BCA assay (Pierce product # 23225). The UPF-CRG Proteomics facility then normalised, digested and labelled the samples. We used dimethyl labelling to label the different samples. Three Labels were used heavy, medium and light. Equal amounts of the samples mixed and analysed by nano LC/MS/MS to obtain ratios of intracellular to extracellular protein concentrations as previously described (Boersema et al., 2009; Tolonen et al., 2011). We calculated the p-Value of a bimodal distribution by standard methods. We a conservative p-value of 0.001 as threshold to define a protein as secreted.

Molecular biology methods

All constructs were cloned in mini-Tn4001-Puro vector (GenBank accession number: KC816623). First, we generated a set of vectors containing the secretion signals and promoters. All fragments comprising different secretion signals and promoters were amplified by PCR from *M. pneumoniae* genomic DNA. For the secretion signals carrying their own promoter we introduced the PstI and EcoRI sites for restriction cloning during the PCR and cloned the fragment by fast ligation (see general material and methods). In the cases when the secretion signal had no own promoter and the EfTu promoter was used we generated the constructs by the isothermal assembly method. Overhangs of 20bp needed for the assembly were introduced by PCR. Detailed sequences of all secretion signals in combination with their promoters are listed in the sequences section and all primers used are listed in Table 6.

The resulting vectors derived from mini-Tn4001-Puro vector were linearized using the restriction enzymes EcoRI (NEB R0101) and PstI-Hf (NEB R3140) and subsequently dephosphorylated using Antartic phosphatse (NEB M0289). The different therapeutical proteins were synthetized and cloned into the EcoRI and PstI-Hf digested vectors. Also the nucleotide sequences of all therapeutic proteins used are listed in Table 6.

The constructs for the conditional release of proteins in the presence of matrix metallo-proteases were assembled using isothermal assembly method. All constructs were cloned in mini-Tn4001-Puro vector (GenBank accession number: KC816623) which was linearized by EcoRV-HF (NEB product # R3195). All fragments were prepared as blunt end and combined using the *in situ* overhang synthesis method. If needed the MMP9 cleavage was synthesized from oligos during the assembly. All sequences are

listed at the end of this document and all primers in Table 7. For the conditional release we introduced the amino acid sequence SGSGPLGLSGSGPLGLGSGS between the membrane anchor and the cargo. It contains two MMP-9 PLGL cleavage motifs and consists mostly out the small amino acid glycine and serine as spacers to provide flexibility.

Table 9 Primers used to generate PCR fragments and assembly of the secretion vectors as template genomic *M. pneumoniae* DNA was used. The primers for Mpn142 and Mpn645 were used directly in an assembly mix. The assembled sequences are listed at the end of this document

Name	Sequence
Mpn152 Rev	GTGTGCCTGCAGGCTgCCATCAACTTGGTTAAATTTGCCCCTTGCC
Mpn 152 fwd	AAGCTTGATATCGAATTCGCTTTTAAAAATACTTTTACTTCAGTAACTC AAAC
Mpn200 Sig F	AAGCTTGATATCGAATTCGACAGTAGTTTAAACTGATTCTTTACCTC
Mpn200 Sig R	GTGTGCCTGCAGGCTgCCTTTACCGCGTGTACCACAAG
Mpn400 Sig F	AAGCTTGATATCGAATTCGCGTAAATTTTCTCCTTTAGGGATTACT
Mpn400 Sig R	GTGTGCCTGCAGGCTgCCATTAACGATTAGAACTGCGGAAAAAGCA
Mpn489 Sig F	AAGCTTGATATCGAATTCTTCACCTTCACCTATTTTATTAGC
Mpn489 Sig R	GTGTGCCTGCAGGCTgCCATTGGAGGTATTGAGTGC
Mpn506 Sig F	AAGCTTGATATCGAATTCGATTAAATTTTCATCTTAAAAGCTTTTATTT TTACC
Mpn506 Sig R	GTGTGCCTGCAGGCTgCCTTTACCCTTTGTACCACAGGCAGC
Mpn588 Sig F	AAGCTTGATATCGAATTCTTCAATTAATCATTGATGGTTTAAGTGTCT C
Mpn588 Sig R	GTGTGCCTGCAGGCTgCCAAAGTTTGGCTGGGTTGCAG
Mpn592 Sig F	AAGCTTGATATCGAATTCAACAGACCTTTAGAAGAAGTGCGA
Mpn592 Sig R	GTGTGCCTGCAGGCTgCCATTTTTGTGATTAGTGTTAGCTACTGTTAG CGT
Mpn142 only signal_No1	TAGAGACGTAATTCAAACACATGAAATCGAAGCTAAAGTTAAAACGT TATTTACTGTTTT
Mpn142 only signal_No2	TGTTGGCTAGTGACAACGTCCCTAGCGGTAAAAGTGGTAAAAACAGT AAATAACGTTTTA
Mpn142 only	GACGTTGTCACTAGCCAACACCTACCTCCTCCAAGGcAGCCTGCAGCC

signal_No3	CGGGGGGCAAGA
Mpn645 only signal_No2	GTAGCTGCTGAAGAACAAGCGCTTAACAACAAACTAGAACCTAAAAG AGAAATTAATAGAA
Mpn645 only signal_No3	CGCTTGTTCTTCAGCAGCTACTCAAGTAATTTCTGGcAGCCTGCAGCC CGGGGGGCAAGA
Mpn645 only signal_No1	TAGAGACGTAATTCAAACACATGAAACTGAAACTTAAATTTCTATTAA TTTCTCTTTTAG
eFTu Fwd	AAGCTTGATATCGAATTCGAAGACCTTTTGTGCTAACGCCAG
eFTu Rev	GTGTTTGAATTACGTCTCTAATTTTACATAAGTTTG
Mpn459 F	GACGTAATTCAAACACATGGCTTTCATGCCATGTTTTTCATATAGC
Mpn459 Rev	GTGTGCCTGCAGGCTgCCAGTAACATAAACATCTCGTGCTTGGGC
Mpn213 Rev	GTGTGCCTGCAGGCTgCCTTGGTGGGTCTTATTTACAAGCGTTATAGT TGTAGG
Mpn213 F	GACGTAATTCAAACACATGAAGCTTAGTGCTATTATCTCCCTATCAGT CG
Mpn332	ATGCCAGCTGTAAAAA
Mpn332	GACTAACACCAAACGTTT

Table 10 All primers used to generate the constructs for the conditional delivery constructs

Name	Sequence
A1AT FWD	CGCGGCTCAAAAGACTGA
A1AT REV	CCGGGCTGCAGGAATTCGATttaTTTCTGGGTGGGATTCAC
eFTU FWD	CGGTATCGATAAGCTTGAT
eFTU REV-P1	GTTTTTTTGGTTTGGTGCATGTGTTTGAATTACGTCTCTAATTTTAC
eFTU REV-P30	CTTCGAGGTGGTAACTTCATGTGTTTGAATTACGTCTCTAATTTTAC
eFTU REV-P65	CCTGGTTTATTTATATCCATGTGTTTGAATTACGTCTCTAATTTTAC
eFTU REV-P90	AACTTTAGCTTCGATTTCATGTGTTTGAATTACGTCTCTAATTTTAC
eYFP FWD-MMP X3	TTGGGCTGGGCAGTGGTAGCATGGTTAGTAAAGGTGAAGAATTGTT
eYFP FWD-P1	CAACCACCTAAAAAACCCGCTATGGTTAGTAAAGGTGAAGAATTG
eYFP FWD-P30	CAAGAACAGTTAGCGGAACAGATGGTTAGTAAAGGTGAAGAATTG

eYFP FWD-P65	GTGGTGATGAATTTTACGAAATGGTTAGTAAAGGTGAAGAATTG
eYFP FWD-P90	GTCCAACCACCTAAAAAGGCTATGGTTAGTAAAGGTGAAGAATTG
eYFP REV	CCGGGCTGCAGGAATTCGATttaGCGGGACCCGGTGCTTC
eYFP-MMP X2	TACTAACCATGCTACCACTGCCCAGCCCAAGAGGGCCTGAGCCACTTAAG CCTAATGGGC
MMP-A1AT X2	GCGTCCCCTTGGGGGTCTTCGCTACCACTGCCCAGCCCAAGAGGGCCTG AGCCACTTAAG
MMP-A1AT X3	GGGCAGTGGTAGCGAAGACCCCCAAGGGGACGCGGCTCAAAAGACTGA CACTTCCCACCA
P1 FWD	ATGCACCAAACCAAAAAAACTG
P1 REV	AGCGGGTTTTTTAGGTGGTT
P1-A1AT X1	AACCACCTAAAAAACCCGCTGAAGACCCCCAAGGGGACGCGGCTCAAAA GACTGACA
P1-A1AT X2	TGTCAGTCTTTTGAGCCGCGTCCCCTTGGGGGTCTTCAGCGGGTTTTTTA GGTGGTT
P1-MMP X1	AACCACCTAAAAAACCCGCTTCGGGTAGTGGCCCATTAGGCTTAAGTGG CTCAGGCCCTC
P30 FWD	ATGAAGTTACCACCTCGAAGAAAG
P30 REV	CTGTTCCGCTAACTGTTCTTG
P30-A1AT X1	AAGAACAGTTAGCGGAACAGGAAGACCCCCAAGGGGACGCGGCTCAAA AGACTGACA
P30-A1AT X2	TGTCAGTCTTTTGAGCCGCGTCCCCTTGGGGGTCTTCCTGTTCCGCTAACT GTTCTT
P30-MMP X1	AAGAACAGTTAGCGGAACAGTCGGGTAGTGGCCCATTAGGCTTAAGTG GCTCAGGCCCTC
P65 FWD	ATGGATATAAATAAACCAGGTTGAAATC
P65 REV	TTCGTAAAATTCATCACCACG
P65-A1AT X1	GTGGTGATGAATTTTACGAAGAAGACCCCCAAGGGGACGCGGCTCAAA AGACTGACA
P65-A1AT X2	TGTCAGTCTTTTGAGCCGCGTCCCCTTGGGGGTCTTCTTCGTAAAATTCAT CACCAC

P65-MMP X1	GTGGTGATGAATTTTACGAATCGGGTAGTGGCCCATTAGGCTTAAGTGG CTCAGGCCCTC
P90 FWD	ATGAAATCGAAGCTAAAGTTAAAAC
P90 REV	AGCCTTTTTAGGTGGTTGG
P90-A1AT X1	TCCAACCACCTAAAAAGGCTGAAGACCCCCAAGGGGACGCGGCTCAAAA GACTGACA
P90-A1AT X2	TGTCAGTCTTTTGAGCCGCGTCCCCTTGGGGGTCTTCAGCCTTTTTAGGT GGTTGGA
P90-MMP X1	TCCAACCACCTAAAAAGGCTTCGGGTAGTGGCCCATTAGGCTTAAGTGG CTCAGGCCCTC

Neutrophil elastase activity assay.

We measured the activity of A1AT indirectly through the inhibition of neutrophil elastase. The activity of neutrophil elastase was measured either by a colorimetric or fluorescence based assay. We used the commercially available substrate N-Methoxysuccinyl-Ala-Ala-Pro-Val p-nitroanilide (Sigma product # M4765) or the cleavage of a quenched fluorescence substrate N-Methoxysuccinyl -Ala-Ala-Pro-Val-AMC (Merck millipore product # 324740). The reaction buffer is 0.1M Tris ph 7.5. In the colorimetric assay the release of free 4-nitroaniline was followed at 400nm and in the fluorescence assay an excitation of 370nm was used and the emission was recorded at 445nm.

Enrichment of A1AT for neutrophil elastase activity measurements

To measure concentrated minimal media supernatant, we first improved the culture condition to reduce the carryover of horse A1AT originating in the pre culture to a minimum. We grew a pre-culture in full Hayflick medium, the media was aspirated and cells were washed twice with PBS. Then, the cells were scraped and split 1:10 in minimal media. The minimal media was aspirated and replaced with fresh media after 24h. The culture was grown for 48 and then the supernatant was harvested. To measure A1AT activity on concentrated media 200 µl of supernatant were reduced to a volume of 20 µl using an Amicon ultrafiltration device with a 30 kDa cut-off and then the A1AT activity was tested.

To enrich A1AT by affinity chromatography we used the Strep-TagII fused to A1AT. We harvested supernatant from 300cm^2 culture dish with cells grown in full Hayflick media. The media was run over a 1ml StrepTrap HP column (GE Healthcare product # 28-9075-46). The column was then washed with 6 column volumes of wash buffer (100mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA) and subsequently eluted with wash buffer supplemented with 2.5mM desthiobiotin. The concentration of A1AT was determined by ELISA.

In vivo secretion of A1AT

Eight-week-old female and male C57BL/6 mice were purchased from Charles River. The Committee of Animal Studies at Sanofi approved the protocol for animal experimentation. This protocol and all laboratory procedures complied with French legislation, which implements the European Union directives. Animals were received at least 1 week before the experiment to allow acclimatization. The animals were kept under a natural daylight cycle and given food and water ad libitum.

Mice were anesthetized via inhalation of isofluorane (5 %) and inoculated intranasally using 50 µl *M. pneumoniae* MPN142-A1AT at 10^8 CFU or 50 µL saline solutions. After 1, 3 and 6 days, mice were sacrificed to examine levels of A1AT in bronchoalveolar lavage.

Conditional delivery of A1AT by MMP-9 protease

Activated neutrophil supernatant.

Neutrophils prepared from three healthy donors. Neutrophils (5x105) were cultured in *M. pneumoniae* minimal media and stimulated with LPS for 24 hours at 37°C. The supernatant was removed and then tested for MMP-9 activity using the Sensolyte MMP-9 assay kit (ANASPEC product # 71134). The supernatants with the highest MMP-9 activity were used for release experiments.

Purified MMP-9 was obtained from EMD Millipore (product # 444231) and activated prior to use with p-aminophenylmercuric acid (*APMA*)-activated. The activated MMP-9 was added to minimal media and cells expressing the conditional delivery constructs were exposed for 24h to the media before being analysed by western blotting using an anti GFP antibody produced from Roche (Table 5).

Dnase activity in supernatant of *M. pneumoniae* cells

M. pneumoniae M129 cells were grown in full Hayflick Media for 48h. An aliquot of 200ul was taken and Lambda DNA (NEB product N3011) was added, as negative control served fresh sample full Hayflick media. Both samples were incubated for 2h at 37°C and subsequently inactivated for 20min at 80°C. The samples were then analysed on a 1% agarose gel and stained with Gel RED (Biotium product # 41003).

P53 and A1AT quantification by ELISA

We used commercially available Kits for the quantification of human A1AT (USCN Product No.: SEB697Hu) and a p53 pan ELISA (Roche Product No.: 11828789001). For the assay we followed the manufacturer's instructions.

Alginate lyase assay

To measure alginate lyase activity in full media we used the assay developed by Kitamikado (et al., 1990). Briefly, 0.1% of alginate substrate is added to the media and with the cells. At various time points 0.2ml of media supernatant is put in a test tube and 2.0ml of an acidic albumin solution (3.26g sodium acetate, 4.56 ml of glacial acetic acid, 1.0g of bovine albumin fraction V are filled up to 1l with water and ph adjusted to 3.75 with HCl). In the presence of polymeric alginate a white precipitate is formed. A small aliquot of the mixture is then transferred to a plate and the absorbance is measured.

We tested different wavelengths for the signal to noise ratio and found 300nm to be the most sensitive, while everything up to 660nm gave good reliable readings.

Alginate lyase to dissolve *Pseudomonas aeruginosa* biofilms

Contributions:

Part of the work presented in this chapter was carried out by the Sanofi exploratory Unit Toulouse. The *P. aeruginosa* biofilm assay was developed and all tests were performed at the Sanofi Sanofi Exploratory Unit in Toulouse by Richard Legoux, Anne-Claude Noirot-Jagu and Mourad Kaghad. The project was sponsored and managed by Edgardo Ferran and Luis Serrano

Abstract

Pseudomonas aeruginosa infections are a severe threat to people suffering from cystic fibrosis. The bacteria form strong biofilms in the lung of these patients, which render the bacteria resistant to antibiotic treatment. The main component of these biofilms is alginate, a sugar polymer. Alginate lyases degrade this polymeric sugar and have long been discussed as biotherapeutic agent to dissolve *P. aeruginosa* biofilms. However, there are contradictory reports in the literature on the efficacy of alginate lyase against *P. aeruginosa* biofilms. We systematically investigated the cause of this controversy. We found out that most positive reports used a crude extract as alginate lyase source and the negative reports used a purified recombinant protein. We fractionated and tested one of the commercially available crude extracts towards anti biofilm activity. By using a number of chromatography methods and nano Lc MS/MS we identified a number of candidates as the putative active compound causing biofilm disruption.

Introduction

Cystic fibrosis (CF) is an incurable genetic disease that has a large range of symptoms affecting many different parts of the body (Bye et al., 1994; Kaneshiro and Zieve, 2012). One of the symptoms is the build-up of viscous mucus in the lung of CF patients. This viscous mucus results in decreased mucociliary clearance and clogging of the airways. The administration of recombinant DNase thins the mucus to ease lung clearance (Quan et al., 2001). However, bacterial infections are a common side effect of the decreased

mucociliary clearance and have to be treated with high dosed antibiotics. While many bacterial infections can be managed like this, around 80% of young adults suffering from CF are chronically infected with the opportunistic pathogen *P. aeruginosa.* This bacteria causes chronic infections that are controllable but persistent and recurrent (Pedersen, 1992). *P. aeruginosa* produces a strong biofilm in the lung of CF patients. It protects the bacteria from antibiotics and the host immune system (Leid et al., 2005). The consequence is chronic inflammation resulting in slow but continuous decrease of lung function, which is the primary cause of death for CF patients.

The biofilm formed by *P. aeruginosa* in the lungs of CF patients is a complex and dynamic mixture. It consists out of 5 major components, three exopolysachrides (alginate, Pel and Psl) , extracellular DNA and proteins (Mann and Wozniak, 2012). While Pel and Psl play an important role in the biofilm formation and stability (Colvin et al., 2012; Ma et al., 2007; Zhao et al., 2013), the most studied exopolysacharide of the three is alginate. Alginate is a polymer of the sugars β-d-mannuronate (M) and its C-5 epimer, α-l-guluronate (G). The polymer composition can be homopolymeric stretches of M or G or heteropolymeric MG stretches. Alginate is produced by many bacteria and algae. It has wide applications as industrial product.

While there are at least 4 components of the Pseudomonas biofilm, an overproduction of alginate was associated with the chronic lung infections of CF patients (Hentzer et al., 2001). Therefore, many groups focused on the alginate production pathway or the degradation of alginate by alginate lyase, as possible drug targets (Alipour et al., 2009; Alkawash et al., 2006; Eftekhar and Speert, 1988; Hatch and Schiller, 1998; Lamppa et al., 2011; Mai et al., 1993). The effect of alginate lyase on biofilm degradation is discussed controversially in the literature. While multiple studies reported a decreased biofilm inhibitory concentration (BIC) of antibiotics against *P. aeruginosa* biofilm cultures (Alipour et al., 2009; Alkawash et al., 2006; Eftekhar and Speert, 1988; Hatch and Schiller, 1998), at least two reports observed either no antibiofilm activity for alginate lyase (Christensen et al., 2001) or an effect that is independent of the enzymatic activity from alginate lyase (Lamppa and Griswold, 2013). The reason for these contradictory observations is currently unknown.

Two publications investigating alginate composition of *P. aeruginosa* strains concluded that the major component is the M sugar mannuronate. L-guluronate appeared to be a minor component and was observed only in in heteropolymeric MG stretches, Poly G

stretches were not detected by both studies (Annison and Couperwhite, 1987; Schürks et al., 2002). The NMR analysis made in these studies suggested that the contribution of M in the polysaccharide is close to 70%. The 13-C NMR signal propagation in sugar triplets showed that MMM sugar triplets contribute to ~15%, MMG to~30%, MGM to ~12%, GGM to ~20% and GMG to ~17%. However, GGG triplets showed no detectable contribution to the signal (Schürks et al., 2002). This composition is different from what is expected by a random polymerisation, as it is observed for example in algae. Schürks et al., (2002) speculated that this is based on: "a biological mechanism leading to a desired chain structure and that the preference of GM-pairs may be linked to possible binding sites of bivalent cations." A structural role of the G sugars in the *P. aeruginosa* biofilm is plausible as in general alginate biofilms with Poly G stretches are more solid because of a specific interaction with Ca^{2+} ions (Franklin et al., 1994).

Noteworthy, both studies which investigated the alginate biofilm composition used cultured clinical isolates to generate a suitable amount of sample for the analysis method (13-C and 1-H NMR). The culturing period could have influenced the alginate composition but a study on biofilm obtained from the lung is not available.

In *P. aeruginosa* two alginate lyases have been identified so far AlgL and PA1167. AlgL is a Poly M alginate lyase (E.C. 4.2.2.3). AlgL has a specific activity of 1211 U/mg against poly M blocks, 15 U/mg against poly G blocks and 453 U/mg against poly MG blocks. In contrast, the alginate lyase PA1167 has a significantly reduced overall activity and changed specificity. In details the specific activity of PA1167 is 8 U/mg against poly M blocks, 0.94 U/mg against poly G blocks and 51 U/mg against poly MG blocks (Yamasaki et al., 2004).

Like other groups we hypothesized that if the major component of the *P. aeruginosa* biofilm is poly M alginate, we can degrade it by a poly M alginate lyase. The M residues in the *P. aeruginosa* alginate can be acetylated to various degrees on the O-2 and/or O-3 position. We decided to use the alginate lyase A1-III from Sphingomonas sp., which has a similar activity towards poly M alginate blocks as the *P. aeruginosa* alginate lyase AlgL and cleaves highly acetylated alginate (Murata et al., 1993; Yamasaki et al., 2004). The alginate lyase A1-III has been proposed by other groups as well for dispersion of *P. aeruginosa* biofilms (Lamppa et al., 2011). We planned to deliver the alginate *in situ* by engineered bacteria similar to previous studies that used

Pyocin 5 as therapeutic protein (Gupta et al., 2013; Saeidi et al., 2011). Therefore, we engineered *M. pneumoniae* to secrete the alginate lyase A1-III from *Sphingomonas sp*. We could detect active alginate lyase in the supernatant of *M. pneumoniae* while we failed to produce a signal in a pseudomonas biofilm assay. A crude alginate lyase extract prepared from *Flavobacterium multivorum* (formerly known *as Sphingobacterium multivorum*) purchased from Sigma (A1603) showed a clear and reproducible signal in our test. In this study we investigated the reason behind this problem.

Results and Discussion

We developed a visual assay to validate activity of the alginate lyase produced by modified *M. pneumoniae* against *P. aeruginosa* biofilms. The assay measures the formation of a clear halo on agar media loaded by bacteria around a paper disk containing the alginate lyase sample (Figure 31 panel a). We obtained such halos on a biofilm of the *P. aeruginosa* CHA strain and using a commercial crude alginate lyase extract from Sigma (A1603) as test compound. The diameter of the halo was proportional to the Alginate Lyase concentration (Figure 31a). The shiny aspect of the *Pseudomonas* CHA biofilm is lost inside the halo (Figure 31b). We could not observe a halo on *Staphylococcus aureus* biofilms, which do not contain alginate (Figure 31c) and other biofilms (data not shown). We concluded from this the effect of the crude alginate lyase extract was specific to *P. aeruginosa* and not a general toxic effect.

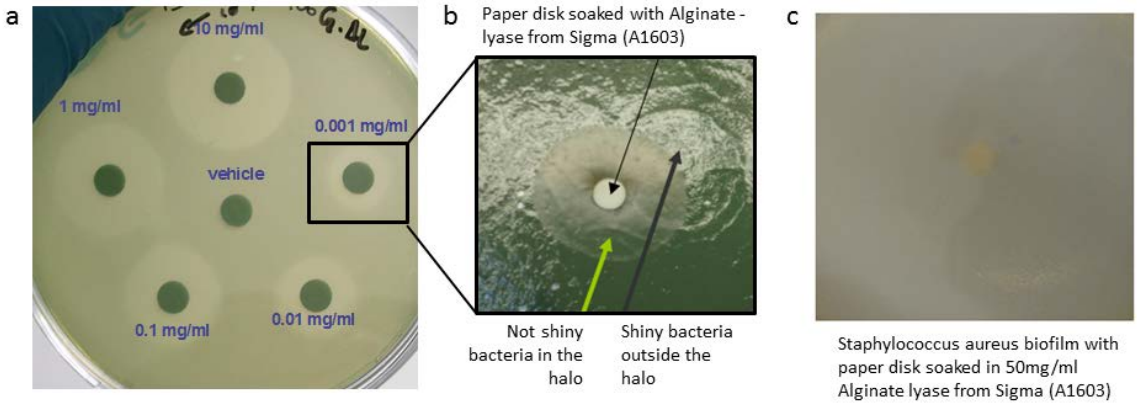


Figure 31 "Halo Test" developed to detect action of Sigma alginate lyase (A1603) on the biofilm created by a mucoid *P. aeruginosa* strain (CHA strain). (a) Influence of Sigma alginate lyase protein concentration on the biofilm Halo size. (b) Closeup view on one of the Halos. The bacteria outside the halo are shiny while the bacteria close to the paper disk appears dull. (c) Sigma alginate lyase on a *Staphylococcus aureus* biofilm. No Halo is observed excluding a unspecific toxic effect of the crude alginate lyase extract.

Using the Halo assay we compared a set of alginate lyases with different substrate specificities in regards to their antibiofilm activity. We tested the A1-III poly M alginate lyase from *Sphingomonas sp*. which was used by (Lamppa and Griswold, 2013) and the A1-II' alginate lyase from the same organism which exhibits a broad substrate specificity (Miyake et al., 2004). We normalized all samples according to their alginate lyase activity against brown seaweed alginate (Sigma W201502). We observed no effect for the A1-III nor the A1-II' alginate lyase on the *P. aeruginosa* biofilm (Figure 32). However, the crude extract of sigma which we used as positive control, produced clearly visible halos in the biofilm for each experiment.

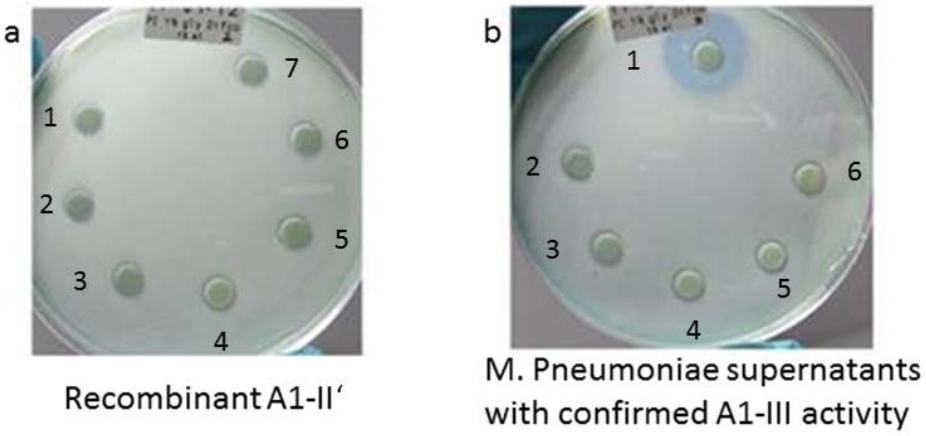


Figure 32 Halo test with different enzymes. (a) Halo test with recombinant produced and purified alginate lyase A1-II' which has broad substrate specificity. The samples are a dilution series from upper left to right, starting with a blank (1) and then samples of A1-II' in a concentration range from 0.37mg/ml (2) to 37ng/ml (7) in 10 fold dilution steps. (b) Halo test with supernatants from A1-III secreting *M. pneumoniae* with confirmed Poly M alginate lyase activity. (1) Positive control with 0.01 mg/ml Sigma alginate lyase (A1603), (2) media blank, (3) the supernatant of a WT strain, (4) supernatant of a A1-II' secreting strain with an alginate lyase activity equivalent to a concentration of 0.01mg/ml from the Sigma alginate lyase, (5) supernatant sample of an A1-III secreting *M. pneumoniae* strain with an alginate lyase activity equivalent to 0.001mg/ml (6) and 0.01mg/ml (7) of Sigma alginate lyase. While the alginate lyase from Sigma produces a clear halo in the P. aeruginosa biofilm, the two other alginate lyases fail to produce a halo despite similar activity levels.

We reviewed the literature and found that all reports describing an effect of alginate lyase against *P. aeruginosa* biofilms used a crude preparation of alginate lyase. A summary of the reviewed studies is given in Table 4. A crude extract from *Bacillus circulans* enriched for alginate lyase activity was used by (Alkawash et al., 2006; Eftekhar and Speert, 1988; Mai et al., 1993). Alipour et al., (2009) used the crude extract sold by Sigma from *Flavobacterium multivorum* which gave a positive signal in our test. In contrast, two studies reporting no effect of alginate lyase activity on *P. aeruginosa* biofilms used recombinantly produced and highly purified alginate lyases for their experiments. An exemption is the study of Hatch and Schiller, (1998) who used a recombinantly produced protein and showed increased antibiotic penetration through

an alginate biofilm after treatment with this alginate lyase. It is worth noting however that this experiment did not involve a live biofilm and that the two other studies also reported efficient breakdown of a *P. aeruginosa* alginate polymer by their enzymes.

Table 11 Table summarizing the outcome of different studies and the respective alginate lyase used

Study	Result	Alginate lyase used
Alkawash et al.	Positiv	Crude extract from *bacillus circulans* (ATCC 15518)
Alipour et al.	Positiv	claim in the title that used AlgL but used in the end Sigma
Eftekhar et al	Positiv	Crude extract from *bacillus circulans* (ATCC 15518) most likely similar to Alkawash
Mai et al.	Positiv	Crude extract *bacillus circulans*
Christensen et al	Negativ	Recombinant *A. vinelandii* alginate lyase gene, AlgL
Lamppa et al	Negativ	recombinant A1-III
Hatch et al	Positiv	Recombinant AlgL, but no real biofilm assay for evaluation only digestion tests of purified alginate polymer

To identify the active compound in the crude extract from Sigma we purified the crude extract further by anion exchange chromatography and successive gel filtration. As a first purification step we performed an anion exchange chromatography and tested the resulting fraction for halo activity. Based on the halo activity we made two pools joining separate fractions and run these two pools on a S75 gel filtration column. The individual fractions were then again tested for halo activity and we could identify two major peaks of antibiofilm activity (Figure 33a/b). Then, we determined the alginate lyase activity for each fraction (Figure 33 c) and plotted the surface area of the formed halos (Figure 33 b). The profile of the halo activity for both pools resembles the profile of alginase activity. We analyzed the fractions by SDS Gel electrophoresis and could

identify two bands which followed in their intensity the observed activity patterns. A protein at a molecular weight of ~ 40kDa seems to be responsible for the peak in fraction B5 and a protein at a molecular weight of ~30kDa seems to be responsible for the peak at fraction B2.

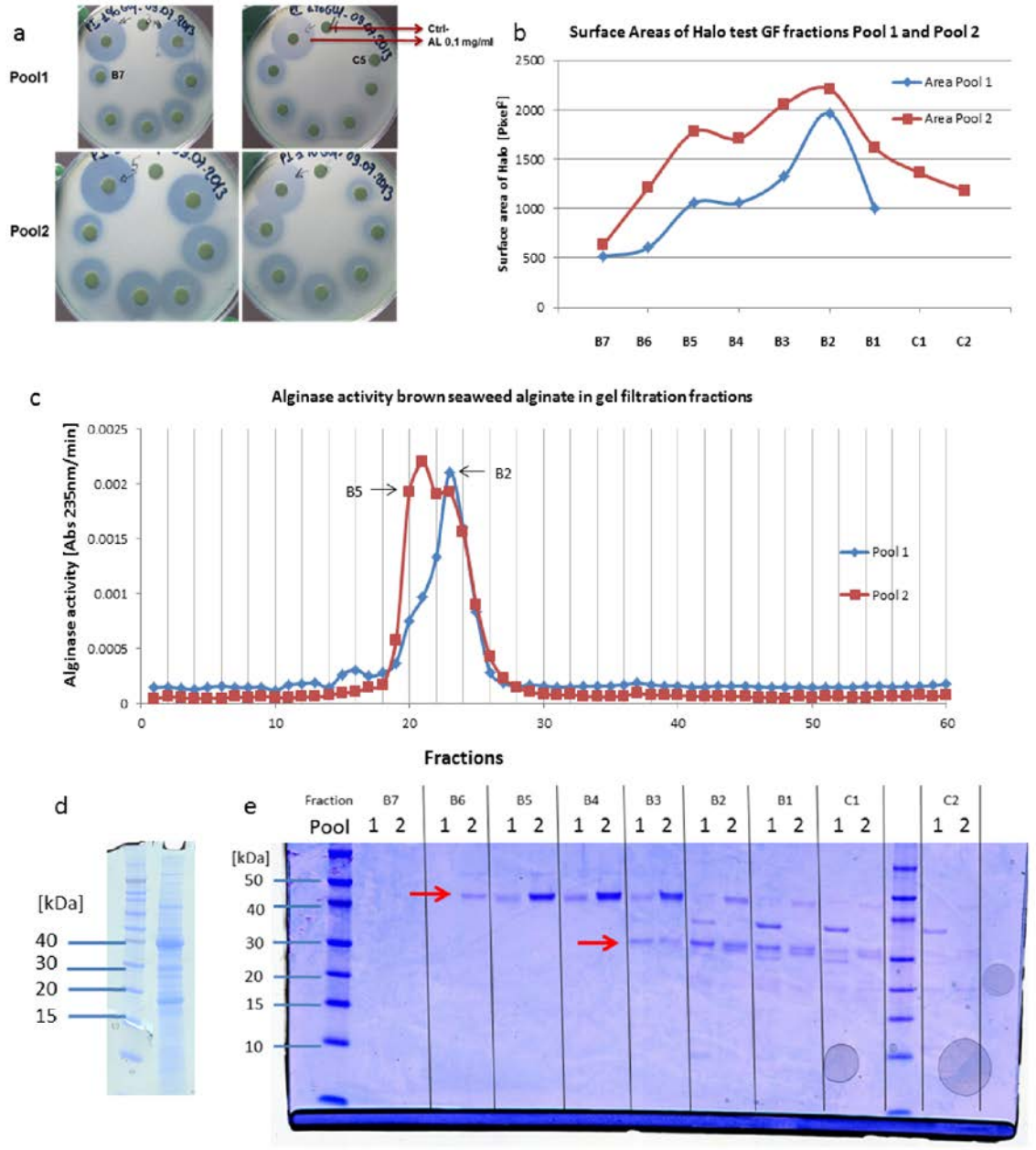


Figure 33 Figure showing the Halo forming activity in relation to different proteins in the crude alginate lyase extract from Sigma (A1603). (d) SDS gel shown from the crude extract before purification. We purified the crude extract by anion exchange chromatography and pooled fraction with a high halo activity. Subsequently these pools were further purified by gel filtration on a Superdex 75 column. The fractions of the gel filtration runs were analyzed for antibiofilm activity by the halo assay (a) and (b) and alginate lyase activity (c). A SDS gel comparing directly the fraction of the two runs is shown in (e).

We analyzed the identified bands by peptide mass finger printing but as the organism from which the alginate lyase crude extract is produced is not yet sequenced we could not unambiguously attribute the peptides to a single band. We used the two best scored peptides for each band and conducted a blast search. For the upper band the two blasted peptides did not point to similar proteins. The spectra on the left was attributed to a thioredoxin from *Citrobacter rodentium* while the sequence obtained from the spectra on the right was found in the hypothetical protein Fjoh_4072 from *Flavobacterium johnsoniae* UW101 (Figure 34a). The analyses of the highest scoring peptides from the lower band were both attributed to an alginate lyase. The peptide sequence for the spectra on the left was found in an alginate lyase from *Lacinutrix sp.* 5H-3-7-4. The peptide sequence obtained from the spectra on the right was found in an alginate lyase *Mesoflavibacter zeaxanthinifaciens* (Figure 34b). We were surprised by the identification of the lower band as an alginate lyase as we expected a protein unrelated to the alginate lyase activity. All found peptides are very short and only provide limited evidence for the identity of the protein. Especially for the upper band we have no clear hypothesis what the molecular function could be that causes the halo activity. To identify unambiguously both proteins and to clone them for recombinant production, we ordered the bacterial strain *Sphingobacterium multivorum* which is the old name for *F. multivorum* from the ATCC. We will sequence the genome completely and use the data to identify the complete sequence. In parallel, we will extract the relevant DNA sequences by degenerate primers based on the identified peptides followed by RACE PCR and Sanger sequencing as described by Clergeot et al., (2012) for the protein CDiT1.

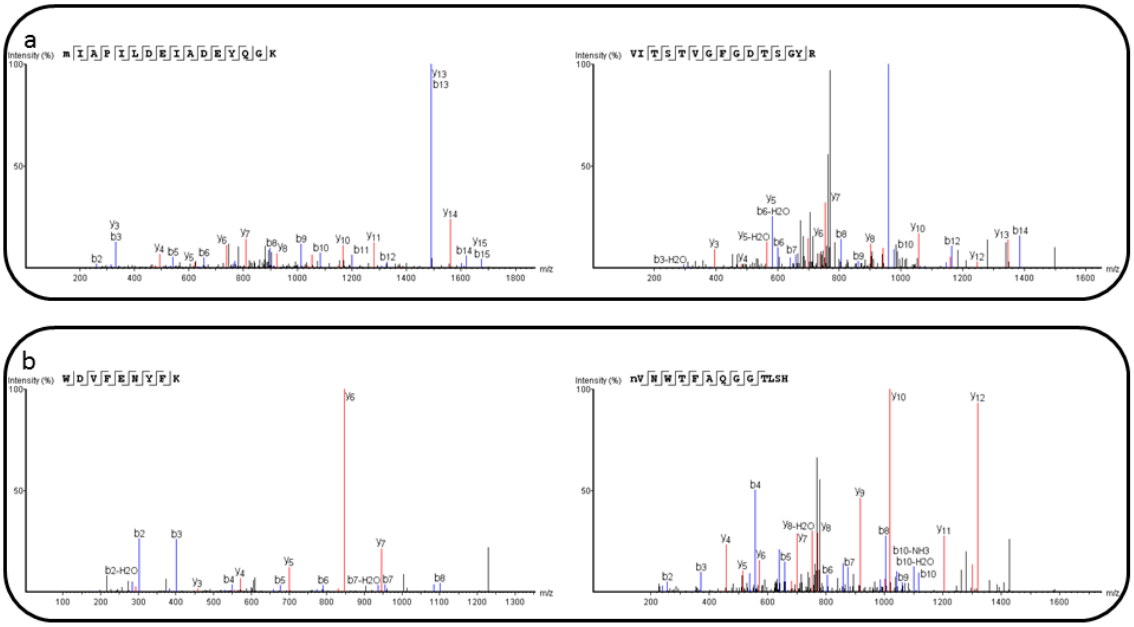


Figure 34 Mass spectrometry analysis of the upper band (a) and lower band (b) identified as active compounds causing the halo activity. Both panels show MS/MS spectra of the two highest scored peptide ions of tryptic digest nano LC/MS run. The y ions are marked in red and the corresponding b ions are in blue. The obtained sequences are given in the upper right corner with the fragmentation sites indicated.

In the meantime we revaluated alginate lyase activity as the driving force in the halo formation. In a first step, we investigated the substrate specificity of the Sigma alginate lyase crude extract. The product is classified by Sigma as poly M alginate lyase (E.C. 4.2.2.3). This is the same enzyme class as the alginate lyase A1-III, which was tested by Lamppa and Griswold, (2013) and our group as not active. Surprisingly, the category E.C. 4.2.2.3 of the Brenda database does not list an entry for *F. multivorum* nor *S. multivorum* but only in the category for poly G alginate lyases, E.C. 4.2.2.11. We tested the substrate specificity of the crude extract sold by Sigma and confirmed that it is a poly G alginate lyase of the enzyme class 4.2.2.11 (Figure 35). We therefore cloned the Alginase A1-II from *Sphingomonas species* which is a pure Poly G alginase in contrast to A1-II' which is indiscriminate to the substrates(Miyake et al., 2004). We are currently expressing and purifying the A1-II alginate lyase to test it for activity in the Halo test.

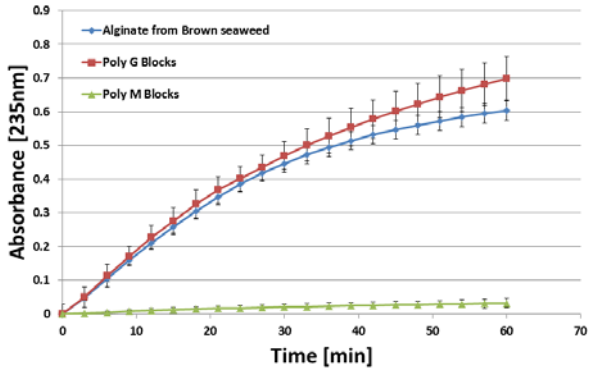


Figure 35 Substrate specificity of the crude alginate lyase extract from Sigma (A1603) towards natural alginate from brown seaweed (Sigma W201502), purified poly G blocks (Elicityl ALG610) or poly M blocks (Elicityl ALG601). Alginase activity is measured as increase of absorbance at 235nm upon degradation of polymeric alginate and formation of unsaturated bonds.

Conclusion

We confirmed with our halo assay, the previously reported observation, that a recombinant produced poly M alginate lyase from *Sphingomonas sp*. has no effect on *P. aeruginosa* biofilms. Neither has the alginate lyase A1-II' an effect on *P. aeruginosa* biofilms, which has an equal activity against poly M, poly MG and poly G substrates. However we could show that a crude extract from *Flavobacterium multivorum* has a halo forming activity on *P. aeruginosa* biofilms. We identified two proteins being responsible for this Halo forming activity. In gel filtration experiment these two proteins peaked at the same time as alginate lyase activity. We obtained the sequences of two peptides for each protein by nano lc MS/MS analysis of tryptic digests from bands isolated of a SDS gel. The protein with the lower band is likely to be an alginate lyase while we could not yet confirm the identity of the upper band. The experiments to identify, clone and purify the two proteins are on-going. We could show that the crude extract sold by Sigma, is in contradiction to the provided specifications a poly G alginate lyase (E.C. 4.2.2.11) and not a poly M alginate lyase (E.C. 4.2.2.3). This led us to the hypothesis, that an alginate lyase specific to poly G or poly MG is necessary to affect *P. aeruginosa* biofilms. This effect could be based in the structural role the G sugar residues in the alginate biofilm. Experiments to purify and test a poly G alginate lyase in the biofilm assay are on-going. In paralell we will continue to unambigously identify and clone the proteins causing the halo formation from the genome of *Flavobacterium multivorum.*

Material and Methods

Alginate lyase activity assay

We used two types of assay. To compare purified proteins in between each other and to determine specificities we measured the increase of absorbance at 235nm. We used the substrates at a final concentration of 0.1% in 0.1M Sodium phosphate buffer (ph7.0). All measured were performed in Tecan M200 infinite pro plate reader at 37°C.

As substrates we used either sodium alginate from brown seaweed alginate (Sigma W201502), Mannuronate oligosaccharides DP20-35 (Elicityl ALG601) or Guluronate oligosaccharides DP25-45 (Elicityl ALG610)

Halo Test

For the halo test the *P. aeruginosa* cells were grown on agar plates with PI media (Bactopeptone 20g/l, $MgCl_2$ 1.4g/l, K_2SO4 10g/l, glycerol 1%, pH 7.2). First the Petri dish was seeded with a bacterial culture and paper disks (MDB-oxoid CT0998-B) were placed on the agar soaked with the solution to be tested for antibiofilm activity for example 20 µl of an alginate lyase solution at different concentrations (10, 1, 0.1, 0.01 and 0.001 mg/ml in the example shown here) and incubate for 16 hours. Formation of halos was then evaluated by visual inspection or measurement of the halo diameter.

Gel filtration and anion exchange chromatographies

For anion exchange chromatography we used a HiTrap Q HP column (GE Healthcare product # 17-1153-01) in conjunction with an ÄKTA-FPLC system. The sample alginate lyase (Sigma # A1603) was dissolved 20mM Tris at ph8 as buffer A and applied to the preequilibrated column. As buffer B we used the same buffer with 1M NaCl added. After 20ml we increased the increased the salt concentration to 500mM over 50ml and collected fractions of 1ml.

For gel filtration chromatography a Superdex 75 column was used (GE Healthcare product # 17-5174-01) with 20mM Tris ph 8 and 150mM NaCl as running buffer. The fraction size was 0.5ml.

To achieve the presented resolution of proteins in the crude extract, we first performed an anion exchange chromatography, pooled the fractions 37-42, 48-51 and 55-60,

concentrated them and performed subsequent gel filtration. After the chromatography runs all fraction were supplemented with 10% glycerol.

Proteomics analysis

All experimental proteomics procedures were performed by the UPF/CRG proteomics Unit. We performed an initial tryptic digestion and peptide finger print mass spec analysis on the crude alginate lyase extract (Sigma A1603). Our aim was to screen for known toxins or other proteins with biofilm dissolving properties. We did not yield any reasonable lead from this. Therefore we cut from a SDS gel the bands protein bands we identified as peaking with the halo activity. The bands were digested by trypsin and analysed by nano-LC MS/MS. The resulting peaks were analysed as DeNovo assembly using the software PEAKS (Bioinformatics solutions inc.). The raw data of each run will be made available as soon as possible.

General Material and Methods

Bacterial strains and growth conditions

Mycoplasma pneumoniae

M. pneumoniae cells were grown in modified Hayflick medium per liter it contains:

1. 16 g PPLO Difco broth (BD product # 255420)
2. 24 g HEPES (Sigma product # H3375)
3. 25mg Phenol red (Sigma product # P3532)
4. 200 ml heat inactivated horse serum (life technologies product # 16050122)
5. 1% Glucose (Sigma product # G7528)
6. 333mg Penicillin (sigma P7794)
7. Adjust pH to 7.7
8. For selective media at 5µg/ml of puromycin or 200µg/ml of gentamycin depending on the experiment

Components 1-4 were combined and filled up to 800ml, ph adjusted and autoclaved. Components 5-7 were added only before use.

For agar plates Components 1-4 were mixed and 1% Agar (BD Bacto Agar product # 214030) added. After autoclaving the other components including antibiotics were added and the plates poured.

The composition of the minimal media is described elsewhere (Yus et al., 2009).

All constructs were transformed into *M.pneumoniae* by electroporation using the method of Hedreyda et al., (1993) with slight modification. A 150cm^2 was started from *M. pneumoniae* M129 stock and grown for 4 days and the resulting cells were used for up to 6 transformations. Before transformation the cells were extensively washed with ice cold transformation buffer (8mM Hepes/272mM Surcose/ ph7.4). For each transformation roughly 10µg DNA was used, the cells were incubated for 15min before the electroshock on ice with the DNA and after. A 1mm cuvette (Biorad product # 165-2083) was used. The shock was delivered by a Gene pulse machine with the settings 1250Volts, 25µFD, 100Ω. The transformed cells were allowed to recover for 2h in 1ml of antibiotic free medium. Then serial dilution of the cells was made and cultures were started in 25cm^2 flasks with 10ml modified Hayflick media containing antibiotics when necessary.

E. coli

As standard conditions *E. coli* cells were cultured at 37ºC in LB or 2xTY with the respective antibiotics added. For standard transformations we used commercial competent cells Top10 (life technologies product # C4040-06) or DH5α cells in conjunction with the *Mix & Go E. coli* Transformation Kit (ZYMO RESEARCH product # T3001). For transformations of constructs >10kbp electroMAX Stbl4 (life technologies product # 11635018) and grown according to the manufacturer's instructions.

PCR

For the generation of fragments subsequently used for cloning we performed PCR with the high fidelity polymerases Phusion (Thermo scientific product # F-534) or KOD high fidelity polymerase (EMD Millipore product # 71085). As standard procedure an annealing gradient between 50°C-70°C was performed for each primer pair in 20µl reaction mix, according to the manufacturer's instructions. Results were visualised by standard agarose gel electrophoresis using Gel Red (Biotium product # 41003).as stain. For large scale screening >96samples a Labchip GX system was used in conjunction with 5k or 12k DNA chips depending on the requirements of the experiment.

For colony PCR Taq polymerase was used. Each colony was picked in 50µl of water and resuspended. Of this suspension 1µl was used as template in 11µl PCR reaction mix.

Molecular assemblies

Unless otherwise stated all constructs were assembled using the isothermal assembly method (Gibson et al., 2009).

Normal restriction ligation cloning was performed using the Rapid DNA Ligation Kit (Roche product # 11635379001) following the manufacturers recommendations.

Genomic DNA purifications

M. pneumoniae genomic DNA was prepared using the llustra bacteria genomicPrep Mini Spin Kit (product # 28-9042-58) following the manufacturer's recommendations.

For diagnostic PCR on *M. pneumoniae* cultures, supernatant from culture was treated with StrataClean resin (Agilent Technologies product # 400714) according to the

instructions of the manufacturer for the detection of *M. pneumoniae* in mammalian cell cultures. The resin supernatant was used as template for standard PCR.

Conclusions

The aim of this thesis was to engineer *M. pneumoniae* for therapeutical applications. First, I wanted establish a set of genetic tools for *M. pneumoniae*, which comply with the high demands of synthetic biology on precision and reproducibility. Despite not being able to obtain the optimal genetic tools, I significantly increased the knowledge about genetic manipulation in *M. pneumoniae*. I implemented a previously described way to generate plasmids in Mollicutes, by using the chromosomal origin of replication of *M. pneumoniae* (Cordova et al., 2002). However, the resulting plasmids did not perform up to my expectations. They either fell short on promoting growth in selective media or the expression of a heterologous reporter protein.

Therefore, I focused my work on the design and integration of a cloning platform in the genome of *M. pneumoniae*. For this purpose, a synthesized stretch of DNA comprising a transcription module, a gene regulation module and a recombination module was ordered. Despite many parameters were considered in the design, the first version of the platform was not functional. I troubleshot the platform multiple times and established techniques in the lab which allow modifying large pieces of DNA as a routine operation. Further, an existing cloning technique was improved to generate overhangs *in situ* and synthesize new sequences during the assembly of larger molecules. However, I could ultimately not obtain a functional version of the platform and more work will be required to achieve it.

We realized during the troubleshooting of the platform, that despite the wealth of knowledge we already have about *M. pneumoniae*, our current understanding of expression control in this bacterium is incomplete. While I could provide a workaround to express heterologous proteins, more research is necessary to understand the underlying mechanism. Otherwise, it will slow down the future attempts to engineer *M. pneumoniae*. One option to investigate the effect, could be the generation of a combinatorial library including all possible codons for the first 10 amino acids of given heterologous protein. A technique that could generate these constructs is currently developed in collaboration with Jae-Seong Yang. It is based on the CPEC cloning method in combination with degenerate primers. Alternatively, the constructs could be ordered from the company Gene9 which offers extremely competitive prices for large

libraries. However, the screening of this large amount of constructs will be a bottleneck, since high throughout techniques are not available for *M. pneumoniae*. The use of *E. coli* to study this effect would simplify the investigation because of the available methods (Kosuri et al., 2013). A report of a similar observation exists from a group investigating *M. pneumoniae* promoters in *E. coli* for protein expression (Loechel et al., 1991) and could be used as a starting point.

While I came short in the initial goals regarding the construction of genetic tools, I provided a clear proof of concept that *M. pneumoniae* can be used for therapeutical applications. Because I was unable to solve the problem of genetic tools on time we followed the KISS approach to develop a first proof of concept. In collaboration with our partners at Sanofi we showed that it is possible to engineer *M. pneumoniae* to secrete a set of functional proteins with therapeutic applications. I further designed a first conditional delivery system releasing a therapeutic protein cargo upon the presence of a protease, indicating inflammation. It was shown by our collaborators that our system actually secretes a therapeutic protein in the lung of a mouse model and not only *in vitro*. However, the *M. pneumoniae* strain currently used as chassis causes inflammation in the lung of the host. Next steps should be focused on engineering the chassis to become non-pathogenic. This would enable the engineered bacteria to stay long term in the lung without causing inflammation. A critical evaluation of this non-pathogenic strain towards rarely observed symptom ms of *M. pneumoniae* infections in humans will be necessary.

I also attempted to engineer a version of *M. pneumoniae* that dissolves *P. aeruginosa* biofilms. While the secreted protein was active and degraded polymeric alginate, we observed no biofilm dissolving effect. When troubleshooting this problem I discovered a number of misleading reports have been published mostly because of a wrongly categorized alginate lyase sold by the company Sigma. Another factor was the use of crude extracts instead of recombinant proteins in some studies. By logic combination and systematic investigation we are currently investigating what is the active compound dissolving *P. aeruginosa* biofilms. The outcome of this investigation will be very useful for the cystic fibrosis research community. Ultimately, this could lead to development of new protein based drug like pulmozyme but which would help to eradicate chronic *P. aeruginosa* infections in combination with antibiotics.

The engineering of complex metabolic pathways or genetic circuits that include more than a few parts is currently still a cumbersome process. Most successful examples have either been achieved through significant amounts of workhours or combinatorial screening approaches. A good example is the nitrogen fixation pathway that the Voigt group tried to move from one bacterium to another. In a first step the group refactored the operons in the original organism and then cloned the whole pathway in a new organism. While they could show that each individual part was able to complement a knock out of the same part in the original bacterium, they were unable to assemble out of the new parts a pathway that would be as efficient as the wild type or complemented version (Temme et al., 2012). Only a combinatorial screening allowing multiple promoters in front of genes and random distribution of the genes lead to a pathway that provided the same output than the original (Voigt et al. SB 6.0 conference Video available online). This clearly illustrates that we do not yet understand the underlying biology enough to rationally design complex pathways in biology. Minimal organisms like *M. pneumoniae* have the potential to solve this bottleneck and enable the rational design of pathways and circuits. However, the results presented in this thesis illustrate that we are not there yet. We still have to invest more efforts in understanding and modeling *M. pneumoniae* to ultimately enable the rational *in silico* design of networks, which will work as predicted once implemented in *M. pneumoniae.* However we could show that it is possible to engineer the chassis *M. pneumoniae* and it is only a question of time and hard work until we can yield the full potential of *M. pneumoniae* in synthetic biology.

Bibliography

Adachi, M., and Cavalcanti, A.R.O. (2009). Tandem stop codons in ciliates that reassign stop codons. J. Mol. Evol. *68*, 424–431.

Algire, M.A., Lartigue, C., Thomas, D.W., Assad-Garcia, N., Glass, J.I., and Merryman, C. (2009). New selectable marker for manipulating the simple genomes of Mycoplasma species. Antimicrob. Agents Chemother. *53*, 4429–4432.

Alipour, M., Suntres, Z.E., and Omri, A. (2009). Importance of DNase and alginate lyase for enhancing free and liposome encapsulated aminoglycoside activity against Pseudomonas aeruginosa. J. Antimicrob. Chemother. *64*, 317 –325.

Alkawash, M.A., Soothill, J.S., and Schiller, N.L. (2006). Alginate lyase enhances antibiotic killing of mucoid Pseudomonas aeruginosa in biofilms. APMIS *114*, 131–138.

Andrianantoandro, E., Basu, S., Karig, D.K., and Weiss, R. (2006). Synthetic biology: new engineering rules for an emerging discipline. Mol. Syst. Biol. *2*, 2006.0028.

Annison, G., and Couperwhite, I. (1987). Composition of alginate synthesized during the growth cycle of Pseudomonas aeruginosa. Aust. J. Biol. Sci. *40*, 435–441.

Ansarin, K., Abedi, S., Ghotaslou, R., Soroush, M.H., Ghabili, K., and Chapman, K.R. (2011). Infection with Mycoplasma pneumoniae is not related to asthma control, asthma severity, and location of airway obstruction. Int. J. Gen. Med. *4*, 1–4.

Balasubramanian, S., Kannan, T.R., and Baseman, J.B. (2008). The surface-exposed carboxyl region of Mycoplasma pneumoniae elongation factor Tu interacts with fibronectin. Infect. Immun. *76*, 3116–3123.

Barile, M.F., Schimke, R.T., and Riggs, D.B. (1966). Presence of the arginine dihydrolase pathway in Mycoplasma. J. Bacteriol. *91*, 189–192.

Barnes, P.J. (2004). Mediators of chronic obstructive pulmonary disease. Pharmacol. Rev. *56*, 515–548.

Benders, G.A., Noskov, V.N., Denisova, E.A., Lartigue, C., Gibson, D.G., Assad-Garcia, N., Chuang, R.-Y., Carrera, W., Moodie, M., Algire, M.A., et al. (2010). Cloning whole bacterial genomes in yeast. Nucleic Acids Res.

Bendtsen, J., Kiemer, L., Fausboll, A., and Brunak, S. (2005). Non-classical protein secretion in bacteria. BMC Microbiol. *5*, 58.

Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., and Blüthgen, N. (2013). Efficient translation initiation dictates codon usage at gene start. Mol. Syst. Biol. *9*.

Bitnun, A., and Richardson, S.E. (2010). Mycoplasma pneumoniae: Innocent Bystander or a True Cause of Central Nervous System Disease? Curr. Infect. Dis. Rep. *12*, 282–290.

Bode, J., Schlake, T., Iber, M., Schübeler, D., Seibler, J., Snezhkov, E., and Nikolaev, L. (2000a). The transgeneticist's toolbox: novel methods for the targeted modification of eukaryotic genomes. Biol. Chem. *381*, 801–813.

Bode, J., Schlake, T., Iber, M., Schübeler, D., Seibler, J., Snezhkov, E., and Nikolaev, L. (2000b). The transgeneticist's toolbox: novel methods for the targeted modification of eukaryotic genomes. Biol. Chem. *381*, 801–813.

Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S., and Heck, A.J.R. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. Nat Protoc. *4*, 484–494.

Breton, M., Sagne, E., Duret, S., Beven, L., Citti, C., and Renaudin, J. (2010). First report of a tetracycline-inducible gene expression system for mollicutes. Microbiology *156*, 198–205.

Bye, M.R., Ewig, J.M., and Quittell, L.M. (1994). Cystic fibrosis. Lung *172*, 251–270.

Caragata, E.P., and Walker, T. (2012). Using bacteria to treat diseases. Expert Opin. Biol. Ther. *12*, 701–712.

Chang, H.-Y., Jordan, J.L., and Krause, D.C. (2011). Domain Analysis of Protein P30 in Mycoplasma pneumoniae Cytadherence and Gliding Motility. J. Bacteriol.

Christensen, B.E., Ertesvag, H., Beyenal, H., and Lewandowski, Z. (2001). Resistance of biofilms containing alginate-producing bacteria to disintegration by an alginate degrading enzyme (Algl). Biofouling *17*, 203–210.

Clergeot, P.-H., Schuler, H., Mørtz, E., Brus, M., Vintila, S., and Ekengren, S. (2012). The corky root rot pathogen Pyrenochaeta lycopersici secretes a proteinaceous inducer of cell death affecting host plants differentially. Phytopathology *102*, 878–891.

Cohen, S.N., Chang, A.C.Y., Boyer, H.W., and Helling, R.B. (1973). Construction of Biologically Functional Bacterial Plasmids In Vitro. Proc. Natl. Acad. Sci. U. S. A. *70*, 3240–3244.

Colvin, K.M., Irie, Y., Tart, C.S., Urbano, R., Whitney, J.C., Ryder, C., Howell, P.L., Wozniak, D.J., and Parsek, M.R. (2012). The Pel and Psl polysaccharides provide Pseudomonas aeruginosa structural redundancy within the biofilm matrix. Environ. Microbiol. *14*, 1913–1928.

Conlon, K.M., Humphreys, H., and O'Gara, J.P. (2004). Inactivations of rsbU and sarA by IS256 Represent Novel Mechanisms of Biofilm Phenotypic Variation in Staphylococcus epidermidis. J. Bacteriol. *186*, 6208–6219.

Cordova, C.M.M., Lartigue, C., Sirand-Pugnet, P., Renaudin, J., Cunha, R.A.F., and Blanchard, A. (2002). Identification of the Origin of Replication of the Mycoplasma pulmonis Chromosome and Its Use in oriC Replicative Plasmids. J. Bacteriol. *184*, 5426–5435.

Dalbey, R.E., and Kuhn, A. (2012). Protein Traffic in Gram-negative bacteria – how exported and secreted proteins find their way. FEMS Microbiol. Rev. *36*, 1023–1045.

Dallo, S.F., Kannan, T.R., Blaylock, M.W., and Baseman, J.B. (2002). Elongation factor Tu and E1 beta subunit of pyruvate dehydrogenase complex act as fibronectin binding proteins in Mycoplasma pneumoniae. Mol. Microbiol. *46*, 1041–1051.

Dandekar, T., Huynen, M., Regula, J.T., Ueberle, B., Zimmermann, C.U., Andrade, M.A., Doerks, T., Sánchez-Pulido, L., Snel, B., Suyama, M., et al. (2000). Re-annotating the Mycoplasma pneumoniae genome sequence: adding value, function and reading frames. Nucleic Acids Res. *28*, 3278–3288.

Dugdale, D.C. (2013). A.D.A.M. Medical Encyclopedia [Internet]. Atlanta (GA): A.D.A.M., Inc.; ©2005. Atypical pneumonia: MedlinePlus Medical Encyclopedia.

Dümmler, A., Lawrence, A.-M., and Marco, A. de (2005). Simplified screening for the detection of soluble fusion constructs expressed in E. coli using a modular set of vectors. Microb. Cell Factories *4*, 34.

Dybvig, K., Lao, P., Jordan, D.S., and Simmons, W.L. (2010). Fewer essential genes in mycoplasmas than previous studies suggest. FEMS Microbiol. Lett.

Eftekhar, F., and Speert, D.P. (1988). Alginase treatment of mucoid Pseudomonas aeruginosa enhances phagocytosis by human monocyte-derived macrophages. Infect Immun *56*, 2788–2793.

Franklin, M.J., Chitnis, C.E., Gacesa, P., Sonesson, A., White, D.C., and Ohman, D.E. (1994). Pseudomonas aeruginosa AlgG is a polymer level alginate C5-mannuronan epimerase. J. Bacteriol. *176*, 1821–1830.

Gao, F., and Zhang, C.-T. (2007). DoriC: a database of oriC regions in bacterial genomes. Bioinformatics *23*, 1866–1867.

Gehrig, S., Mall, M.A., and Schultz, C. (2012). Spatially Resolved Monitoring of Neutrophil Elastase Activity with Ratiometric Fluorescent Reporters. Angew. Chem. Int. Ed. *51*, 6258–6261.

Gibson, D.G., Benders, G.A., Andrews-Pfannkoch, C., Denisova, E.A., Baden-Tillson, H., Zaveri, J., Stockwell, T.B., Brownley, A., Thomas, D.W., Algire, M.A., et al. (2008). Complete Chemical Synthesis, Assembly, and Cloning of a Mycoplasma genitalium Genome. Science *319*, 1215–1220.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat Meth *6*, 343–345.

Gibson, D.G., Smith, H.O., Hutchison, C.A., Venter, J.C., and Merryman, C. (2010). Chemical synthesis of the mouse mitochondrial genome. Nat Meth *7*, 901–903.

Glass, J.I., Assad-Garcia, N., Alperovich, N., Yooseph, S., Lewis, M.R., Maruf, M., Hutchison, C.A., Smith, H.O., and Venter, J.C. (2006). Essential genes of a minimal bacterium. Proc. Natl. Acad. Sci. U. S. A. *103*, 425–430.

Govan, J.R., and Deretic, V. (1996). Microbial pathogenesis in cystic fibrosis: mucoid Pseudomonas aeruginosa and Burkholderia cepacia. Microbiol. Rev. *60*, 539–574.

Greenberg-Ofrath, N., Terespolosky, Y., Kahane, I., and Bar, R. (1993). Cyclodextrins as carriers of cholesterol and fatty acids in cultivation of mycoplasmas. Appl. Environ. Microbiol. *59*, 547–551.

Güell, M., van Noort, V., Yus, E., Chen, W.-H., Leigh-Bell, J., Michalodimitrakis, K., Yamada, T., Arumugam, M., Doerks, T., Kühner, S., et al. (2009). Transcriptome complexity in a genome-reduced bacterium. Science *326*, 1268–1271.

Gupta, S., Bram, E.E., and Weiss, R. (2013). Genetically Programmable Pathogen Sense and Destroy. ACS Synth. Biol.

Guzman, L., Belin, D., Carson, M., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. J Bacteriol *177*, 4121–4130.

Hahn, T., Mothershed, E.A., Waldo, R.H., and Krause, D.C. (1999). Construction and Analysis of a Modified Tn4001Conferring Chloramphenicol Resistance inMycoplasma pneumoniae. Plasmid *41*, 120–124.

Halbedel, S., and Stülke, J. (2007). Tools for the genetic analysis of Mycoplasma. Int. J. Med. Microbiol. IJMM *297*, 37–44.

Halbedel, S., Eilers, H., Jonas, B., Busse, J., Hecker, M., Engelmann, S., and Stülke, J. (2007). Transcription in Mycoplasma pneumoniae: Analysis of the Promoters of the ackA and ldh Genes. J. Mol. Biol. *371*, 596–607.

Hallamaa, K., Tang, S.-L., Ficorilli, N., and Browning, G. (2008). Differential expression of lipoprotein genes in Mycoplasma pneumoniae after contact with human lung epithelial cells, and under oxidative and acidic stress. BMC Microbiol. *8*, 124.

Hansen, E.J., Wilson, R.M., Clyde, W.A., Jr, and Baseman, J.B. (1981). Characterization of hemadsorption-negative mutants of Mycoplasma pneumoniae. Infect. Immun. *32*, 127–136.

Hatch, R.A., and Schiller, N.L. (1998). Alginate Lyase Promotes Diffusion of Aminoglycosides through the Extracellular Polysaccharide of Mucoid Pseudomonas aeruginosa. Antimicrob. Agents Chemother. *42*, 974–977.

Hathout, Y. (2007). Approaches to the study of the cell secretome. Expert Rev. Proteomics *4*, 239–248.

Hay, J.W., and Robin, E.D. (1991). Cost-effectiveness of alpha-1 antitrypsin replacement therapy in treatment of congenital chronic obstructive pulmonary disease. Am. J. Public Health *81*, 427–433.

Hedreyda, C.T., Lee, K.K., and Krause, D.C. (1993). Transformation of Mycoplasma pneumoniae with Tn4001 by Electroporation. Plasmid *30*, 170–175.

Hegermann, J., Halbedel, S., Dumke, R., Regula, J., Gabdoulline, R.R., Mayer, F., Stülke, J., and Herrmann, R. (2008). The acidic, glutamine-rich Mpn474 protein of Mycoplasma pneumoniae is surface exposed and covers the complete cell. Microbiol. Read. Engl. *154*, 1185–1192.

Hennig, S., and Ziebuhr, W. (2008). A Transposase-Independent Mechanism Gives Rise to Precise Excision of IS256 from Insertion Sites in Staphylococcus epidermidis. J. Bacteriol. *190*, 1488–1490.

Hentzer, M., Teitzel, G.M., Balzer, G.J., Heydorn, A., Molin, S., Givskov, M., and Parsek, M.R. (2001). Alginate Overproduction Affects Pseudomonas aeruginosa Biofilm Structure and Function. J. Bacteriol. *183*, 5395–5401.

Herrero, M., de Lorenzo, V., Ensley, B., and Timmis, K.N. (1993). A T7 RNA polymerase-based system for the construction of Pseudomonas strains with phenotypes dependent on TOL-meta pathway effectors. Gene *134*, 103–106.

Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B.C., and Herrmann, R. (1996). Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae. Nucleic Acids Res. *24*, 4420–4449.

Hochschild, A., and Lewis, M. (2009). The bacteriophage λ CI protein finds an asymmetric solution. Curr. Opin. Struct. Biol. *19*, 79–86.

Hoover, D.M., and Lubkowski, J. (2002). DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. Nucleic Acids Res. *30*, e43–e43.

Huisman, O., Raymond, W., Froehlich, K.-U., Errada, P., Kleckner, N., Botstein, D., and Hoyt, M.A. (1987). A Tn 10-lacZ-kanR-URA3 Gene Fusion Transposon for Insertion Mutagenesis and Fusion Analysis of Yeast and Bacterial Genes. Genetics *116*, 191–199.

Hutchison, C.A., Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., and Smith, H.O. (2008). One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome. Proc. Natl. Acad. Sci. U. S. A. *105*, 20404–20409.

Van Huynegem, K., Loos, M., and Steidler, L. (2009). Immunomodulation by genetically engineered lactic acid bacteria. Front. Biosci. Landmark Ed. *14*, 4825–4835.

Hwang, I.Y., Tan, M.H., Koh, E., Ho, C.L., Poh, C.L., and Chang, M.W. (2013). Reprogramming Microbes to Be Pathogen-Seeking Killers. ACS Synth. Biol.

Ikeda, R.A., and Richardson, C.C. (1987). Enzymatic properties of a proteolytically nicked RNA polymerase of bacteriophage T7. J. Biol. Chem. *262*, 3790–3799.

Izsvák, Z., Chuah, M.K.L., Vandendriessche, T., and Ivics, Z. (2009). Efficient stable gene transfer into human cells by the Sleeping Beauty transposon vectors. Methods San Diego Calif *49*, 287–297.

Jiang, S., Zhang, S., Langenfeld, J., Lo, S.-C., and Rogers, M.B. (2008). Mycoplasma infection transforms normal lung cells and induces bone morphogenetic protein 2 expression by post-transcriptional mechanisms. J. Cell. Biochem. *104*, 580–594.

Johnson, C., Kannan, T.R., and Baseman, J.B. (2011). Cellular vacuoles induced by Mycoplasma pneumoniae CARDS toxin originate from Rab9-associated compartments. PloS One *6*, e22877.

Kameyama, Y., Kawabe, Y., Ito, A., and Kamihira, M. (2009). An accumulative site-specific gene integration system using Cre recombinase-mediated cassette exchange. Biotechnol. Bioeng.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

Kaneshiro, N.K., and Zieve (2012). Cystic fibrosis.

Kannan, T.R., and Baseman, J.B. (2006). ADP-ribosylating and vacuolating cytotoxin of Mycoplasma pneumoniae represents unique virulence determinant among bacterial pathogens. Proc. Natl. Acad. Sci. U. S. A. *103*, 6724–6729.

Kannan, T.R., Coalson, J.J., Cagle, M., Musatovova, O., Hardy, R.D., and Baseman, J.B. (2011). Synthesis and Distribution of CARDS Toxin During Mycoplasma pneumoniae Infection in a Murine Model. J. Infect. Dis. *204*, 1596–1604.

Karas, B.J., Jablanovic, J., Sun, L., Ma, L., Goldgof, G.M., Stam, J., Ramon, A., Manary, M.J., Winzeler, E.A., Venter, J.C., et al. (2013). Direct transfer of whole genomes from bacteria to yeast. Nat. Methods.

Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Assad-Garcia, N., Glass, J.I., and Covert, M.W. (2012). A Whole-Cell Computational Model Predicts Phenotype from Genotype. Cell *150*, 389–401.

Kenri, T., Seto, S., Horino, A., Sasaki, Y., Sasaki, T., and Miyata, M. (2004). Use of Fluorescent-Protein Tagging To Determine the Subcellular Localization of Mycoplasma pneumoniae Proteins Encoded by the Cytadherence Regulatory Locus. J Bacteriol *186*, 6944–6955.

Kitamikado, M., Yamaguchi, K., Tseng, C.H., and Okabe, B. (1990). Method designed to detect alginate-degrading bacteria. Appl. Environ. Microbiol. *56*, 2939–2940.

Kitamikado, M., Tseng, C.H., Yamaguchi, K., and Nakamura, T. (1992). Two types of bacterial alginate lyases. Appl. Environ. Microbiol. *58*, 2474–2478.

Knudtson, K.L., and Minion, F.C. (1993). Construction of Tn4001lac derivatives to be used as promoter probe vectors in mycoplasmas. Gene *137*, 217–222.

Kosuri, S., Goodman, D.B., Cambray, G., Mutalik, V.K., Gao, Y., Arkin, A.P., Endy, D., and Church, G.M. (2013). Composability of regulatory sequences controlling transcription and translation in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A.

Krause, D.C., and Balish, M.F. (2001). Structure, function, and assembly of the terminal organelle of Mycoplasma pneumoniae. FEMS Microbiol. Lett. *198*, 1–7.

Krause, D.C., Leith, D.K., Wilson, R.M., and Baseman, J.B. (1982). Identification of Mycoplasma pneumoniae proteins associated with hemadsorption and virulence. Infect. Immun. *35*, 809–817.

Krishnakumar, R., Assad-Garcia, N., Benders, G.A., Phan, Q., Montague, M.G., and Glass, J.I. (2010). Targeted Chromosomal Knockouts in Mycoplasma pneumoniae. Appl Env. Microbiol AEM.00024–10.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-Sequence Determinants of Gene Expression in Escherichia coli. Science *324*, 255–258.

Kühner, S., van Noort, V., Betts, M.J., Leo-Macias, A., Batisse, C., Rode, M., Yamada, T., Maier, T., Bader, S., Beltran-Alvarez, P., et al. (2009). Proteome organization in a genome-reduced bacterium. Science *326*, 1235–1240.

Lamppa, J.W., and Griswold, K.E. (2013). Alginate Lyase Exhibits Catalysis-Independent Biofilm Dispersion and Antibiotic Synergy. Antimicrob. Agents Chemother. *57*, 137–145.

Lamppa, J.W., Ackerman, M.E., Lai, J.I., Scanlon, T.C., and Griswold, K.E. (2011). Genetically Engineered Alginate Lyase-PEG Conjugates Exhibit Enhanced Catalytic Function and Reduced Immunoreactivity. PLoS ONE *6*, e17042.

Lartigue, C., Glass, J.I., Alperovich, N., Pieper, R., Parmar, P.P., Hutchison Iii, C.A., Smith, H.O., and Venter, J.C. (2007). Genome Transplantation in Bacteria: Changing One Species to Another. Science.

Lartigue, C., Vashee, S., Algire, M.A., Chuang, R.-Y., Benders, G.A., Ma, L., Noskov, V.N., Denisova, E.A., Gibson, D.G., Assad-Garcia, N., et al. (2009). Creating Bacterial Strains from Genomes That Have Been Cloned and Engineered in Yeast. Science *325*, 1693–1696.

Layh-Schmitt, G., Hilbert, H., and Pirkl, E. (1995). A spontaneous hemadsorption-negative mutant of Mycoplasma pneumoniae exhibits a truncated adhesin-related 30-kilodalton protein and lacks the cytadherence-accessory protein HMW1. J. Bacteriol. *177*, 843–846.

Leid, J.G., Willson, C.J., Shirtliff, M.E., Hassett, D.J., Parsek, M.R., and Jeffers, A.K. (2005). The Exopolysaccharide Alginate Protects Pseudomonas aeruginosa Biofilm Bacteria from IFN-γ-Mediated Macrophage Killing. J. Immunol. *175*, 7512 –7518.

Li, M., and Elledge, S. (2007). Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat. Methods *4*, 256, 251.

Liu, W., Ren, T., Jiang, B., Gong, M., and Shou, C. (2007). Mycoplasmal membrane protein p37 promotes malignant changes in mammalian cells. Can. J. Microbiol. *53*, 270–276.

Lluch-Senar, M., Vallmitjana, M., Querol, E., and Pinol, J. (2007). A new promoterless reporter vector reveals antisense transcription in Mycoplasma genitalium. Microbiology *153*, 2743–2752.

Lluch-Senar, M., Luong, K., Lloréns-Rico, V., Delgado, J., Fang, G., Spittle, K., Clark, T.A., Schadt, E., Turner, S.W., Korlach, J., et al. (2013). Comprehensive Methylome Characterization of Mycoplasma genitalium and Mycoplasma pneumoniae at Single-Base Resolution. PLoS Genet *9*, e1003191.

Loechel, S., Inamine, J.M., and Hu, P.C. (1991). A novel translation initiation region from Mycoplasma genitalium that functions in Escherichia coli. Nucleic Acids Res. *19*, 6905–6911.

Loessner, I., Dietrich, K., Dittrich, D., Hacker, J., and Ziebuhr, W. (2002). Transposase-Dependent Formation of Circular IS256 Derivatives in Staphylococcus epidermidis and Staphylococcus aureus. J. Bacteriol. *184*, 4709–4714.

De Lorenzo, V., Herrero, M., Jakubzik, U., and Timmis, K.N. (1990). Mini-Tn5 transposon derivatives for insertion mutagenesis, promoter probing, and chromosomal insertion of cloned DNA in gram-negative eubacteria. J. Bacteriol. *172*, 6568–6572.

Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res. *25*, 1203–1210.

Ma, L., Lu, H., Sprinkle, A., Parsek, M.R., and Wozniak, D.J. (2007). Pseudomonas aeruginosa Psl Is a Galactose- and Mannose-Rich Exopolysaccharide. J. Bacteriol. *189*, 8353–8356.

Maglennon, G.A., Cook, B.S., Matthews, D., Deeney, A.S., Bossé, J.T., Langford, P.R., Maskell, D.J., Tucker, A.W., Wren, B.W., and Rycroft, A.N. (2013). Development of a self-replicating plasmid system for Mycoplasma hyopneumoniae. Vet. Res. *44*, 63.

Mai, G.T., McCormack, J.G., Seow, W.K., Pier, G.B., Jackson, L.A., and Thong, Y.H. (1993). Inhibition of adherence of mucoid Pseudomonas aeruginosa by alginase, specific monoclonal antibodies, and antibiotics. Infect Immun *61*, 4338–4343.

Maier, T., Schmidt, A., Güell, M., Kühner, S., Gavin, A.-C., Aebersold, R., and Serrano, L. (2011). Quantification of mRNA and protein and integration with protein turnover in a bacterium. Mol. Syst. Biol. *7*, 511.

Maier, T., Marcos, J., Wodke, J.A.H., Paetzold, B., Liebeke, M., Gutiérrez-Gallego, R., and Serrano, L. (2013). Large-scale metabolome analysis and quantitative integration with genomics and proteomics data in Mycoplasma pneumoniae. Mol. BioSyst.

Maki, H., and Murakami, K. (1997). Formation of potent hybrid promoters of the mutant llm gene by IS256 transposition in methicillin-resistant Staphylococcus aureus. J. Bacteriol. *179*, 6944–6948.

Mann, E.E., and Wozniak, D.J. (2012). Pseudomonas biofilm matrix composition and niche biology. FEMS Microbiol. Rev. *36*, 893–916.

Manoil, C., and Beckwith, J. (1985). TnphoA: a transposon probe for protein export signals. Proc. Natl. Acad. Sci. U. S. A. *82*, 8129 –8133.

McEvoy, C.R.E., Tsuji, B., Gao, W., Seemann, T., Porter, J.L., Doig, K., Ngo, D., Howden, B.P., and Stinear, T.P. (2013). Decreased Vancomycin Susceptibility in Staphylococcus aureus caused by IS256 Tempering of WalKR Expression. Antimicrob. Agents Chemother.

Meissner, F., Scheltema, R.A., Mollenkopf, H.-J., and Mann, M. (2013). Direct Proteomic Quantification of the Secretome of Activated Immune Cells. Science *340*, 475–478.

Menolascina, F., Siciliano, V., and di Bernardo, D. (2012). Engineering and control of biological systems: A new way to tackle complex diseases. FEBS Lett. *586*, 2122–2128.

Messer, W., Blaesing, F., Jakimowicz, D., Krause, M., Majka, J., Nardmann, J., Schaper, S., Seitz, H., Speck, C., Weigel, C., et al. (2001). Bacterial replication initiator DnaA. Rules for DnaA binding and rolesof DnaA in origin unwinding and helicase loading. Biochimie *83*, 5–12.

Miller, M., Hafner, M., Sontag, E., Davidsohn, N., Subramanian, S., Purnick, P.E.M., Lauffenburger, D., and Weiss, R. (2012). Modular Design of Artificial Tissue Homeostasis: Robust Control through Synthetic Cellular Heterogeneity. PLoS Comput Biol *8*, e1002579.

Miyake, O., Ochiai, A., Hashimoto, W., and Murata, K. (2004). Origin and Diversity of Alginate Lyases of Families PL-5 and -7 in Sphingomonas sp. Strain A1. J. Bacteriol. *186*, 2891–2896.

Moran, N.A. (2002). Microbial minimalism: genome reduction in bacterial pathogens. Cell *108*, 583–586.

Murata, K., Inose, T., Hisano, T., Abe, S., Yonemoto, Y., Yamashita, T., Takagi, M., Sakaguchi, K., Kimura, A., and Imanaka, T. (1993). Bacterial alginate lyase: Enzymology, genetics and application. J. Ferment. Bioeng. *76*, 427–437.

Mutalik, V.K., Guimaraes, J.C., Cambray, G., Lam, C., Christoffersen, M.J., Mai, Q.-A., Tran, A.B., Paull, M., Keasling, J.D., Arkin, A.P., et al. (2013a). Precise and reliable gene expression via standard transcription and translation initiation elements. Nat. Methods *10*, 354–360.

Mutalik, V.K., Guimaraes, J.C., Cambray, G., Mai, Q.-A., Christoffersen, M.J., Martin, L., Yu, A., Lam, C., Rodriguez, C., Bennett, G., et al. (2013b). Quantitative estimation of activity and quality for collections of functional genetic elements. Nat. Methods *10*, 347–353.

Namiki, K., Goodison, S., Porvasnik, S., Allan, R.W., Iczkowski, K.A., Urbanek, C., Reyes, L., Sakamoto, N., and Rosser, C.J. (2009). Persistent exposure to Mycoplasma induces malignant transformation of human prostate cells. PLoS ONE *4*, e6872.

Narang, A. (2007). Effect of DNA looping on the induction kinetics of the lac operon. J. Theor. Biol. *247*, 695–712.

Nikel, P.I., and de Lorenzo, V. (2013). Implantation of unmarked regulatory and metabolic modules in Gram-negative bacteria with specialised mini-transposon delivery vectors. J. Biotechnol. *163*, 143–154.

Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. Nature *496*, 528–532.

Pascal Sirand-Pugnet, C.L. (2007). Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome. PLoS Genet. *3*, e75.

Pedersen, S.S. (1992). Lung infection with alginate-producing, mucoid Pseudomonas aeruginosa in cystic fibrosis. APMIS. Suppl. *28*, 1–79.

Pei, L., Gaisser, S., and Schmidt, M. (2012). Synthetic biology in the view of European public funding organisations. Public Underst. Sci. Bristol Engl. *21*, 149–162.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods *8*, 785–786.

Petrache, I., Hajjar, J., and Campos, M. (2009). Safety and efficacy of alpha-1-antitrypsin augmentation therapy in the treatment of patients with alpha-1-antitrypsin deficiency. Biol. Targets Ther. *3*, 193–204.

Phillips, M.I. (1997). Antisense inhibition and adeno-associated viral vector delivery for reducing hypertension. Hypertension *29*, 177–187.

Pich, O.Q., Burgos, R., Planell, R., Querol, E., and Pinol, J. (2006). Comparative analysis of antibiotic resistance gene markers in Mycoplasma genitalium: application to studies of the minimal gene complement. Microbiology *152*, 519–527.

Pour-El, I., Adams, C., and Minion, F.C. (2002). Construction of Mini-Tn4001tet and Its Use in Mycoplasma gallisepticum. Plasmid *47*, 129–137.

Prudhomme, M., Turlan, C., Claverys, J.-P., and Chandler, M. (2002a). Diversity of Tn4001 Transposition Products: the Flanking IS256 Elements Can Form Tandem Dimers and IS Circles. J Bacteriol *184*, 433–443.

Prudhomme, M., Turlan, C., Claverys, J.-P., and Chandler, M. (2002b). Diversity of Tn4001 Transposition Products: the Flanking IS256 Elements Can Form Tandem Dimers and IS Circles. J Bacteriol *184*, 433–443.

Purcell, O., Jain, B., Karr, J.R., Covert, M.W., and Lu, T.K. (2013). Towards a whole-cell modeling approach for synthetic biology. Chaos Interdiscip. J. Nonlinear Sci. *23*, 025112–025112–8.

Purnick, P.E.M., and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. Nat. Rev. Mol. Cell Biol. *10*, 410–422.

Quan, J., Saaem, I., Tang, N., Ma, S., Negre, N., Gong, H., White, K.P., and Tian, J. (2011). Parallel on-chip gene synthesis and application to optimization of protein expression. Nat. Biotechnol.

Quan, J.M., Tiddens, H.A., Sy, J.P., McKenzie, S.G., Montgomery, M.D., Robinson, P.J., Wohl, M.E., Konstan, M.W., and Pulmozyme Early Intervention Trial Study Group (2001). A two-year randomized, placebo-controlled trial of dornase alfa in young patients with cystic fibrosis with mild lung function abnormalities. J. Pediatr. *139*, 813–820.

Rechnitzer, H., Rottem, S., and Herrmann, R. (2013). Reconstitution of an active arginine deiminase pathway in Mycoplasma pneumoniae M129. Infect. Immun.

Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature *440*, 940–943.

Saeidi, N., Wong, C.K., Lo, T.-M., Nguyen, H.X., Ling, H., Leong, S.S.J., Poh, C.L., and Chang, M.W. (2011). Engineering microbes to sense and eradicate Pseudomonas aeruginosa, a human pathogen. Mol. Syst. Biol. *7*, 521.

Saglio, P.H.M., Daniels, M.J., and Pradet, A. (1979). ATP and Energy Charge as Criteria of Growth and Metabolic Activity of Mollicutes: Application to Spiroplasma citri. J. Gen. Microbiol. *110*, 13–20.

Salis, H.M., Mirsky, E.A., and Voigt, C.A. (2009). Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotech *27*, 946–950.

Sandhaus, R.A., Turino, G., Stocks, J., Strange, C., Trapnell, B.C., Silverman, E.K., Everett, S.E., Stoller, J.K., and Medical and Scientific Advisory Committee of the Alpha-1 Foundation (2008). alpha1-Antitrypsin augmentation therapy for PI*MZ heterozygotes: a cautionary note. Chest *134*, 831–834.

Sang, Y., Millwood, R.J., and Neal Stewart Jr, C. (2013). Gene use restriction technologies for transgenic plant bioconfinement. Plant Biotechnol. J. *11*, 649–658.

Scharff, L.B., Childs, L., Walther, D., and Bock, R. (2011). Local Absence of Secondary Structure Permits Translation of mRNAs that Lack Ribosome-Binding Sites. PLoS Genet *7*, e1002155.

Schmidl, S.R., Hames, C., and Stülke, J. (2007). Expression of Mycoplasma proteins carrying an affinity tag in M. pneumoniae allows rapid purification and circumvents problems related to the aberrant genetic code. Appl. Environ. Microbiol. *73*, 7799–7801.

Schürks, N., Wingender, J., Flemming, H.-C., and Mayer, C. (2002). Monomer composition and sequence of alginates from Pseudomonas aeruginosa. Int. J. Biol. Macromol. *30*, 105–111.

Shao, Z., Zhao, H., and Zhao, H. (2009). DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. Nucl Acids Res *37*, e16.

Sharan, S.K., Thomason, L.C., Kuznetsov, S.G., and Court, D.L. (2009). Recombineering: a homologous recombination-based method of genetic engineering. Nat Protoc. *4*, 206–223.

Shimizu, T., Kida, Y., and Kuwano, K. (2011). Cytoadherence-dependent induction of inflammatory responses by Mycoplasma pneumoniae. Immunology no–no.

Sluijter, M., Kaptein, E., Spuesens, E.B.M., Hoogenboezem, T., Hartwig, N.G., Van Rossum, A.M.C., and Vink, C. (2010). The Mycoplasma genitalium MG352-encoded protein is a Holliday junction resolvase that has a non-functional orthologue in Mycoplasma pneumoniae. Mol. Microbiol.

Steidler, L., Rottiers, P., and Coulie, B. (2009). Actobiotics as a novel method for cytokine delivery. Ann. N. Y. Acad. Sci. *1182*, 135–145.

Studier, F.W. (1991). Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system. J. Mol. Biol. *219*, 37–44.

Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. *189*, 113–130.

Supek, F., Smuc, T., and muc (2010). On Relevance of Codon Usage to Expression of Synthetic and Natural Genes in Escherichia coli. Genetics *185*, 1129–1134.

Temme, K., Zhao, D., and Voigt, C.A. (2012). Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. Proc. Natl. Acad. Sci. U. S. A. *109*, 7085–7090.

Thiel, K. (2004). Old dogma, new tricks—21st Century phage therapy. Nat. Biotechnol. *22*, 31–36.

Tjalsma, H., Bolhuis, A., Jongbloed, J.D.H., Bron, S., and van Dijl, J.M. (2000). Signal Peptide-Dependent Protein Transport in Bacillus subtilis: a Genome-Based Survey of the Secretome. Microbiol. Mol. Biol. Rev. *64*, 515–547.

Tolonen, A.C., Haas, W., Chilaka, A.C., Aach, J., Gygi, S.P., and Church, G.M. (2011). Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. Mol Syst Biol *7*.

Villalobos, A., Ness, J., Gustafsson, C., Minshull, J., and Govindarajan, S. (2006). Gene Designer: a synthetic biology tool for constructing artificial DNA segments. BMC Bioinformatics *7*, 285.

Voigt, C. Systematic Engineering Meets Biological Complexity - Part I | Session | BioBricks Foundation SB6.0: The Sixth International Meeting on Synthetic Biology.

Waehler, R., Russell, S.J., and Curiel, D.T. (2007). Engineering targeted viral vectors for gene therapy. Nat. Rev. Genet. *8*, 573–587.

Waites, K.B., and Talkington, D.F. (2004). Mycoplasma pneumoniae and Its Role as a Human Pathogen. Clin. Microbiol. Rev. *17*, 697–728.

Wang, B., Wang, J., Zhang, W., and Meldrum, D.R. (2012a). Application of synthetic biology in cyanobacteria and algae. Front. Microbiol. *3*.

Wang, Y., Zhu, Z., Church, T.D., Lugogo, N.L., Que, L.G., Francisco, D., Ingram, J.L., Huggins, M., Beaver, D.M., Wright, J.R., et al. (2012b). SHP-1 as a critical regulator of

Mycoplasma pneumoniae-induced inflammation in human asthmatic airway epithelial cells. J. Immunol. Baltim. Md 1950 *188*, 3371–3381.

Weber, W., and Fussenegger, M. (2012). Emerging biomedical applications of synthetic biology. Nat. Rev. Genet. *13*, 21–35.

Weisburg, W.G., Tully, J.G., Rose, D.L., Petzel, J.P., Oyaizu, H., Yang, D., Mandelco, L., Sechrest, J., Lawrence, T.G., and Van Etten, J. (1989). A phylogenetic analysis of the mycoplasmas: basis for their classification. J. Bacteriol. *171*, 6455–6467.

Wittebole, X., De Roock, S., and Opal, S.M. (2013). A historical overview of bacteriophage therapy as an alternative to antibiotics for the treatment of bacterial pathogens. Virulence *4*.

Wodke, J.A.H., Puchałka, J., Lluch-Senar, M., Marcos, J., Yus, E., Godinho, M., Gutiérrez-Gallego, R., dos Santos, V.A.P.M., Serrano, L., Klipp, E., et al. (2013). Dissecting the energy metabolism in Mycoplasma pneumoniae through genome-scale metabolic modeling. Mol. Syst. Biol. *9*.

Wu, H.-C., Tsao, C.-Y., Quan, D.N., Cheng, Y., Servinsky, M.D., Carter, K.K., Jee, K.J., Terrell, J.L., Zargar, A., Rubloff, G.W., et al. (2013). Autonomous bacterial localization and gene expression based on nearby cell receptor density. Mol. Syst. Biol. *9*.

Yamasaki, M., Moriwaki, S., Miyake, O., Hashimoto, W., Murata, K., and Mikami, B. (2004). Structure and Function of a Hypothetical Pseudomonas aeruginosa Protein PA1167 Classified into Family PL-7. J. Biol. Chem. *279*, 31863 –31872.

You, L., Cox, R.S., Weiss, R., and Arnold, F.H. (2004). Programmed population control by cell-cell communication and regulated killing. Nature *428*, 868–871.

Yus, E., Maier, T., Michalodimitrakis, K., van Noort, V., Yamada, T., Chen, W.-H., Wodke, J.A.H., Güell, M., Martínez, S., Bourgeois, R., et al. (2009). Impact of genome reduction on bacterial metabolism and its regulation. Science *326*, 1263–1268.

Yus, E., Güell, M., Vivancos, A.P., Chen, W.-H., Lluch-Senar, M., Delgado, J., Gavin, A.-C., Bork, P., and Serrano, L. (2012). Transcription start site associated RNAs in bacteria. Mol. Syst. Biol. *8*, 585.

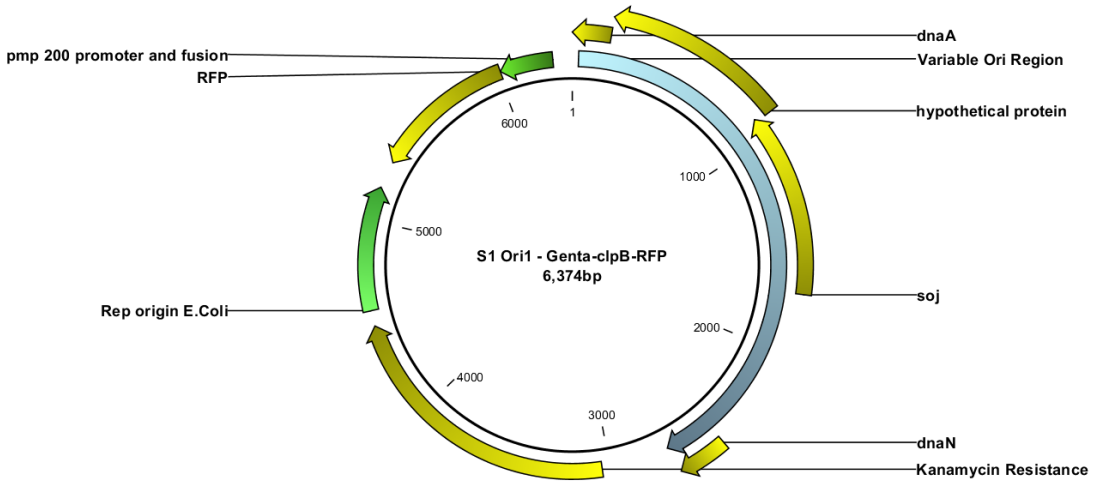
Zhao, K., Tseng, B.S., Beckerman, B., Jin, F., Gibiansky, M.L., Harrison, J.J., Luijten, E., Parsek, M.R., and Wong, G.C.L. (2013). Psl trails guide exploration and microcolony formation in Pseudomonas aeruginosa biofilms. Nature *497*, 388–391.

Zilfou, J.T., and Lowe, S.W. (2009). Tumor Suppressive Functions of p53. Cold Spring Harb. Perspect. Biol. *1*.

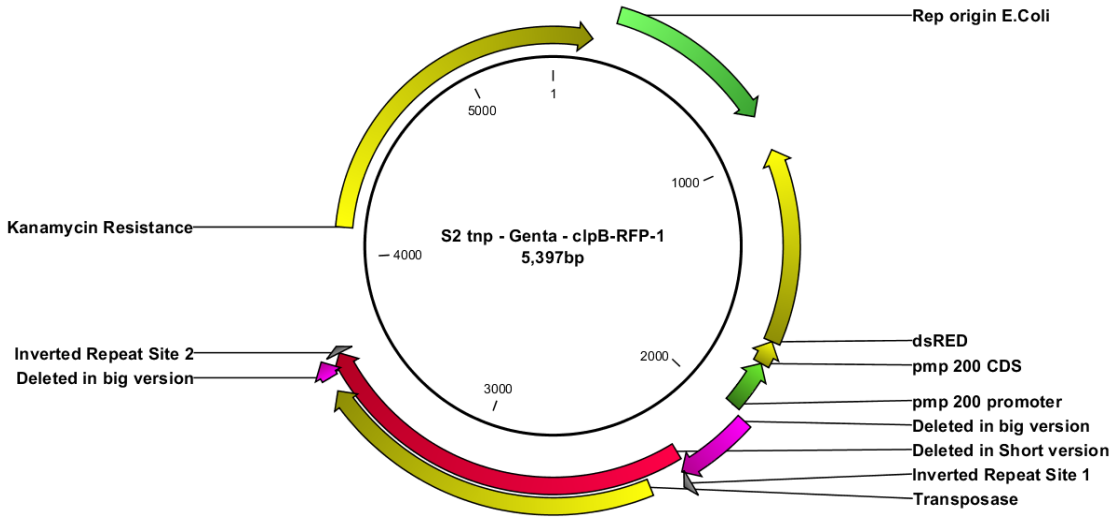
Zimmerman, C.-U., and Herrmann, R. (2005). Synthesis of a small, cysteine-rich, 29 amino acids long peptide in Mycoplasma pneumoniae. FEMS Microbiol. Lett. *253*, 315–321.

Supplementary material

Supplementary material self-replicating plasmids

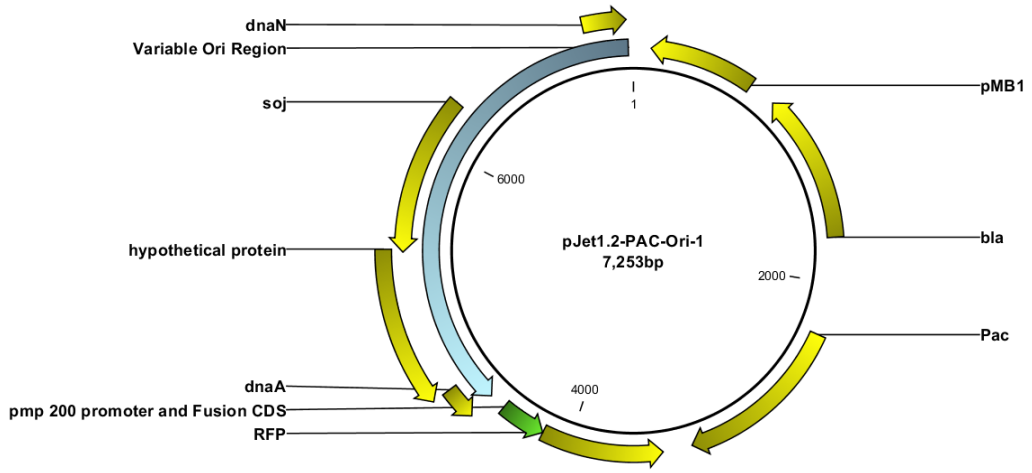


Supplementary figure 1 Vector map of the constructs used to test the fragments of the chromosomal origin of replication for their suitability to construct self-replicating plasmids. In Blue is the region shown that varied between the constructs used to test the different regions



Supplementary figure 2 Mini-transposon Vector S2 that was constructed to benchmark the newly designed plasmids. The backbone is identical to the one shown in Supplementary figure 1 but instead of a chromosomal fragment it has an

IS256 element integrated. In red are marked the parts that were found to be deleted in the two constructs recovered from a *M. pneumoniae* culture. The restriction enzyme ApaI cuts 2050 bp and at 3634 bp. XhoI cuts at 1700 bp and at 3626 bp.



Supplementary figure 3 Second generation of plasmids used to investigate the self-replicating plasmids. Compared to the version shown in Supplementary figure 1 only the fluorescence reporter system and the *M. pneumoniae* origins of replication are the same. The origin of replication for *E. coli* has changed (pMB1) and the ampicilin resistance gene (bla) for selection in *E.coli* has been introduced. The antibiotic resistance gene for selection in *M. pneumoniae* has changed from kanamycin to puromycin (Pac)

Supplementary material *In situ* overlap and sequence synthesis during DNA assembly

Supplementary table 1 Combination of oligonucleotides and dsDNA fragments used for generating a library of 47 constructs. The table shows the size of the PCR fragment cloned (sequences are provided at the end of this document), the number of oligonucleotides used, the length of the DNA synthesis and in which round of transformation one or more clones with the correct sequence was found, as well as, the names of the oligonucleotides used. The constructs that could not be obtained after 3 rounds of transformation are marked with N/A. All constructs were assembled into the miniTn4001-Puro Vector.

Construct name	PCR-CDS	Size double stranded DNA fragment [bp]	Number of Oligos used	Length of synthesized DNA [bp]	Obtained in round	Oligo nucleotiodes used
Construct 1	Lac-CDS	1135	3	52bp	N/A	A1, A2, A3
Construct 2	Lac-CDS	1135	2	25bp	1	A1, A4
Construct 3	Lac-CDS	1135	3	52bp	1	A1, A5, A6
Construct 4	Lac-CDS	1135	2	31bp	1	A1, A7
Construct 5	T7Lys-CDS	526	7	228bp	N/A	A10, A11, A12, B1, B2, B3, B4
Construct 6	T7Lys-CDS	526	7	234bp	N/A	A10, A11, A12, B1, B2, B3, B7
Construct 7	T7Lys-CDS	526	8	255bp	2	A10, A11, A12, B1, B2, B3, A8, A9
Construct 8	T7Lys-CDS	526	8	255bp	1	A10, A11, A12, B1, B2, B3, B5, B6
Construct 9	CI857-CDS	781	6	203bp	2	B8, B9, B10, B11, B12, C1
Construct 10	CI857-CDS	781	6	209bp	1	B8, B9, B10, B11, B12, C6
Construct 11	CI857-CDS	781	7	230bp	N/A	B8, B9, B10, B11, B12, C2, C3
Construct 12	CI857-CDS	781	7	230bp	1	B8, B9, B10, B11, B12, C4, C5
Construct 13	T7Pol-CDS	2737	3	87bp	3	C7, C8, C9
Construct 14	T7Pol-CDS	2737	3	93bp	2	C7, C8, C12
Construct 15	T7Pol-CDS	2737	4	114bp	3	C7, C8, D1, D2
Construct 16	T7Pol-CDS	2737	4	114bp	N/A	C7, C8, C10, C11
Construct 17	CI857-CDS	781	3	69bp	1	D3, D4, D5
Construct 18	Lac-CDS	1135	3	69bp	1	D3, D4, D6
Construct 19	T7Lys-CDS	526	3	69bp	3	D3, D4, D7
Construct 20	Tet-CDS	1052	3	69bp	1	D3, D4, D8
Construct 21	CI857-CDS	781	3	96bp	1	D3, D4, D9
Construct 22	Lac-CDS	1135	3	96bp	3	D3, D4, D10
Construct 23	T7Lys-CDS	526	3	96bp	2	D3, D4, D11

Construct 24	Tet-CDS	1052	3	96bp	1	D3, D4, D12
Construct 25	CI857-CDS	781	3	75bp	1	D3, D4, E1
Construct 26	Lac-CDS	1135	3	75bp	1	D3, D4, E2
Construct 27	T7Lys-CDS	526	3	75bp	1	D3, D4, E3
Construct 28	Tet-CDS	1052	3	75bp	1	D3, D4, E4
Construct 29	CI857-CDS	781	3	96bp	2	D3, D4, E5
Construct 30	Lac-CDS	1135	3	96bp	3	D3, D4, E6
Construct 31	T7Lys-CDS	526	3	96bp	3	D3, D4, E7
Construct 32	Tet-CDS	1052	3	96bp	N/A	D3, D4, E8
Construct 33	CI857-CDS	781	2	46bp	1	E9, E10
Construct 34	Lac-CDS	1135	2	46bp	1	E9, E11
Construct 35	T7Lys-CDS	526	2	46bp	1	E9, E12
Construct 36	T7Pol-CDS	2737	2	46bp	3	E9, F1
Construct 37	Tet-CDS	1052	2	46bp	1	E9, F2
Construct 38	CI857-CDS	781	2	52bp	1	E9, G1
Construct 39	Lac-CDS	1135	2	52bp	1	E9, G2
Construct 40	T7Lys-CDS	526	2	52bp	1	E9, G3
Construct 41	T7Pol-CDS	2737	2	52bp	3	E9, G4
Construct 42	Tet-CDS	1052	2	52bp	1	E9, G5
Construct 43	CI857-CDS	781	3	73bp	1	E9, F3, F4
Construct 44	Lac-CDS	1135	3	73bp	1	E9, F5, F6
Construct 45	T7Lys-CDS	526	3	73bp	1	E9, F7, F8
Construct 46	T7Pol-CDS	2737	3	73bp	3	E9, F9, F10
Construct 47	Tet-CDS	1052	3	73bp	1	E9, F11, F12

Supplementary table 2 The table shows all oligonucleotides used in this study. The oligonucleotides were ordered from Sigma either as purified by reverse phase (RP) or desalted (DS)

Oligonucleotides used Assembling one double stranded insert and the de novo synthesis		
Oligonucleotides used for de novo assembly		
Name	Purity	Sequence
A1	RP	CTCGAGGTCGACGGTATCGATAAGCTTGATTAGTTTTTAGAATTTAAAGGAA
A4	RP	ATCATATAAGGTAACAGGCTTAACGTTCACCATTTCCTTTAAATTCTAAAAACTAATCAA
A5	RP	ATTCTTGCCAAATACTAATTCCTTTGCCATTTCCTTTAAATTCTAAAAACTAATCAAGCT
A6	RP	ATGGCAAAGGAATTAGTATTTGGCAAGAATGTGAACGTTAAGCCTGTTACCTTATATGAT
A7	RP	ATAAGGTAACAGGCTTAACGTTCACCATGGTACCTTCCTTTAAATTCTAAAAACTAATCA
A8	RP	TATTTTATTATTGTTACCTATTAGATAAAAATGAACGACAACAAAATTTTAGTAGTCAAT
A9	RP	ACTTTCGCGTTGCTTGAACTGCACCCGAGCATTGACTACTAAAATTTTGTTGTCGTTCAT
A10	RP	CTCGAGGTCGACGGTATCGATAAGCTTGATCGTTAATAATGATGATTGAAGCTAGTACAA
A11	RP	ATTGGTGTAAGTAATGATAAGACCTTAATCATTGTCTTTTTTGTACTAGCTTCAATCATC
A12	RP	TTATCATTACTTACACCAATTACCCCCATGAGGGAAGAGATAAAGTAGAGAATAATACCT
B1	RP	CTAAAACTTCATATATTGTGGCTGCTGTTTCGACAGCTGTAGGTATTATTCTCTACTTTA
B2	RP	CACAATATATGAAGTTTTAGCACTCATTGACAGTTATTAGAGAAAAAATGAGGACTTATT
B3	RP	TTTTATCTAATAGGTAACAATAATAAAATATTAGTAATAAGTCCTCATTTTTTCTCTAAT
B5	RP	TATTTTATTATTGTTACCTATTAGATAAAAATGGCAAAGGAATTAGTATTTGGCAAGAAT
B6	RP	ACTTTCGCGTTGCTTGAACTGCACCCGAGCATTCTTGCCAAATACTAATTCCTTTGCCAT

B7	RP	TATTATTGTTACCTATTAGATAAAAGGTACCATGGCTCGGGTGCAGTTCAAGCAACGCGA
B8	RP	CTCGAGGTCGACGGTATCGATAAGCTTGATTCAGCAATTACAAAAACAAAACAAATAAAA
B9	RP	CGTTAGCACAAAAGGTCTTCTTGGGGGTAATTCCCTTATTTTTTATTTGTTTTGTTTTTG
B10	RP	GAAGACCTTTTGTGCTAACGCCAGTTTGGCAAATCAAGTTCTGATTTTGCAATTATTTTG
B11	RP	AGAGGCTTATAATTCTTGGAGTAGTGTAATTCATATGGAGCAAAATAATTGCAAAATCAG
B12	RP	TCCAAGAATTATAAGCCTCTCTACAGCTTTATCTCAAACTTATGTAAAATTAGAGACGTA
C1	RP	CTGGGTTAATGGCTTCTTCTTAGTGCTCATGTGTTTGAATTACGTCTCTAATTTTACATA
C4	RP	ATTCTTGCCAAATACTAATTCCTTTGCCATGTGTTTGAATTACGTCTCTAATTTTACATA
C5	RP	ATGGCAAAGGAATTAGTATTTGGCAAGAATAGCACTAAGAAGAAGCCATTAACCCAGGAA
C6	RP	TAATGGCTTCTTCTTAGTGCTCATGGTACCGTGTTTGAATTACGTCTCTAATTTTACATA
C7	RP	CTCGAGGTCGACGGTATCGATAAGCTTGATTTTAGATCAAATCTCAATTGGGAACAGTAA
C8	RP	AGCTTGCCTTCATTTTAATTTGGTTTAAGCCACCAATCATTTACTGTTCCCAATTGAGAT
C9	RP	AATTAAAATGAAGGCAAGCTTGAATAAATTAGTTATGAACACCATTAACATAGCGAAGAA
C12	RP	AATTAAAATGAAGGCAAGCTTGAATAAATTAGTTGGTACCATGAACACCATTAACATAGCGAA
D1	RP	AACCAAATTAAAATGAAGGCAAGCTTGAATAAATTAGTTATGAAGAGTCTTAAAGTAGCA
D2	RP	TCGTTCTTCGCTATGTTAATGGTGTTACCAATGAGTGCTACTTTAAGACTCTTCATAACT
D3	RP	CTCGAGGTCGACGGTATCGATAAGCTTGATTTTAGATCAAATCTCAATTGGGAACAGTAA
D4	RP	TGCCTTCATTTTAATTTGGTTTAAGCCACCAATCATTTACTGTTCCCAATTGAGATTTGA
D5	RP	GGCTTAAACCAAATTAAAATGAAGGCAATGAGCACTAAGAAGAAGCCATTAACCCAGGAA
D6	RP	TTAAACCAAATTAAAATGAAGGCAATGGTGAACGTTAAGCCTGTTACCTTATATGATGTG
D7	RP	TTAAACCAAATTAAAATGAAGGCAATGGCTCGGGTGCAGTTCAAGCAACGCGAAAGTACC
D8	RP	GGCTTAAACCAAATTAAAATGAAGGCAATGTCCAGGCTCGACAAGTCAAAGGTGATTAAC
D9	RP	ACCAAATTAAAATGAAGGCAATGGCAAAGGAATTAGTATTTGGCAAGAATAGCACTAAGAAGAA GCCATT
D10	RP	ACCAAATTAAAATGAAGGCAATGGCAAAGGAATTAGTATTTGGCAAGAATGTGAACGTTAAGCCT GTTAC
D11	RP	ACCAAATTAAAATGAAGGCAATGGCAAAGGAATTAGTATTTGGCAAGAATGCTCGGGTGCAGTT CAAGCA
D12	RP	ACCAAATTAAAATGAAGGCAATGGCAAAGGAATTAGTATTTGGCAAGAATTCCAGGCTCGACAA GTCAAA
E1	RP	GCTTAAACCAAATTAAAATGAAGGCAGGTACCATGAGCACTAAGAAGAAGCCATTAACCC
E2	RP	GCTTAAACCAAATTAAAATGAAGGCAGGTACCATGGTGAACGTTAAGCCTGTTACCTTAT
E3	RP	GCTTAAACCAAATTAAAATGAAGGCAGGTACCATGGCTCGGGTGCAGTTCAAGCAACGCG
E4	RP	GCTTAAACCAAATTAAAATGAAGGCAGGTACCATGTCCAGGCTCGACAAGTCAAAGGTGA
E5	RP	ACCAAATTAAAATGAAGGCAATGAAGAGTCTTAAAGTAGCACTCATTGGTAGCACTAAGAAGAA GCCATT
E6	RP	ACCAAATTAAAATGAAGGCAATGAAGAGTCTTAAAGTAGCACTCATTGGTGTGAACGTTAAGCCT GTTAC
E7	RP	ACCAAATTAAAATGAAGGCAATGAAGAGTCTTAAAGTAGCACTCATTGGTGCTCGGGTGCAGTTC AAGCA
E9	RP	CTCGAGGTCGACGGTATCGATAAGCTTGATGAATTGATCGCCATAGGTTAAAGTAGTATA
E10	RP	GGTTAATGGCTTCTTCTTAGTGCTCATAATTAGTTTATATTATACTACTTTAACCTATGG
E11	RP	TAAGGTAACAGGCTTAACGTTCACCATAATTAGTTTATATTATACTACTTTAACCTATGG
E12	RP	GCGTTGCTTGAACTGCACCCGAGCCATAATTAGTTTATATTATACTACTTTAACCTATGG
F1	RP	GTTCTTCGCTATGTTAATGGTGTTCATAATTAGTTTATATTATACTACTTTAACCTATGG
F2	RP	CACCTTTGACTTGTCGAGCCTGGACATAATTAGTTTATATTATACTACTTTAACCTATGG

F3	RP	GCCAAATACTAATTCCTTTGCCATAATTAGTTTATATTATACTACTTTAACCTATGGCGA
F4	RP	ATTATGGCAAAGGAATTAGTATTTGGCAAGAATAGCACTAAGAAGAAGCCATTAACCCAG
F5	RP	GCCAAATACTAATTCCTTTGCCATAATTAGTTTATATTATACTACTTTAACCTATGGCGA
F6	RP	ATTATGGCAAAGGAATTAGTATTTGGCAAGAATGTGAACGTTAAGCCTGTTACCTTATAT
F7	RP	GCCAAATACTAATTCCTTTGCCATAATTAGTTTATATTATACTACTTTAACCTATGGCGA
F8	RP	ATTATGGCAAAGGAATTAGTATTTGGCAAGAATGCTCGGGTGCAGTTCAAGCAACGCGAA
F9	RP	GCCAAATACTAATTCCTTTGCCATAATTAGTTTATATTATACTACTTTAACCTATGGCGA
F10	RP	ATTATGGCAAAGGAATTAGTATTTGGCAAGAATAACACCATTAACATAGCGAAGAACGAT
F11	RP	GCCAAATACTAATTCCTTTGCCATAATTAGTTTATATTATACTACTTTAACCTATGGCGA
F12	RP	ATTATGGCAAAGGAATTAGTATTTGGCAAGAATTCCAGGCTCGACAAGTCAAAGGTGATT
G1	RP	AATGGCTTCTTCTTAGTGCTCATGGTACCAATTAGTTTATATTATACTACTTTAACCTATGG
G2	RP	GTAACAGGCTTAACGTTCACCATGGTACCAATTAGTTTATATTATACTACTTTAACCTATGG
G3	RP	TGCTTGAACTGCACCCGAGCCATGGTACCAATTAGTTTATATTATACTACTTTAACCTATGG
G4	RP	TTCGCTATGTTAATGGTGTTCATGGTACCAATTAGTTTATATTATACTACTTTAACCTATGG
G5	RP	TTTGACTTGTCGAGCCTGGACATGGTACCAATTAGTTTATATTATACTACTTTAACCTATGG

Oligonucleotides used for PCR amplification of double stranded DNA fragments		
Fwd Lac CDS	RP	GTGAACGTTAAGCCTGTT
Fwd Tet CDS	RP	TCCAGGCTCGACAAGTCAA
Fwd CI857 CDS	RP	AGCACTAAGAAGAAGCCATT
Fwd T7 Pol	RP	AACACCATTAACATAGCGAAGA
Fwd T7 Lys	RP	GCTCGGGTGCAGTTCAA
Rev Tet CDS	RP	ATCTTGCCCCCCGGGCTGCAGGAATTCGATTACGCGTCTAGTTATA
Rev CI857 CDS	RP	ATCTTGCCCCCCGGGCTGCAGGAATTCGATCTAATTAAGACTAGAGGT
Rev T7Pol CDS	RP	ATCTTGCCCCCCGGGCTGCAGGAATTCGATAACCTTCTGGTGATCAA
Rev T7Lys CDS	RP	ATCTTGCCCCCCGGGCTGCAGGAATTCGATAAGCTTCTACTTATGTCTA
Rev Lac CDS	RP	ATCTTGCCCCCCGGGCTGCAGGAATTCGATATGCATCTACTTATCGCTA

Oligo nucleotides used for screening miniTn4001-Puro-1 by colony PCR		
Fwd Screen Puro	RP	CTACATGATGAATGGATT
Rev Screen Puro	RP	CTGTTTTCTTGTTCACTA

Oligonucleotides used for In situ generation of overhangs from single stranded oligonucleotides		
Oligonucleotides used for in situ generation of overhangs		
Name		Sequence
DnaA to PetM14 stitch 3	RP	CATTGTCATCGTAACATATGCGCCATTAACCTGATGTTCT
DnaA to PetM14 stitch 4	RP	AGAACATCAGGTTAATGGCGCATATGTTACGATGACAATG
DnaA to PetM14 stitch 1	RP	AAGTTCTGTTCCAGGGGCCCGTGTCACTTTCGCTTTGGCA
DnaA to PetM14 stitch 2	RP	TGCCAAAGCGAAAGTGACACGGGCCCCTGGAACAGAACTT

Oligonucleotides used for amplification of double stranded DNA fragment		
DnaA fwd	DS	GTGTCACTTTCGCTTTGGCAGC
DnaA Rev	DS	CATATGTTACGATGACAATGTTCT

Oligonucleotides used for linearization of Vector petM14		
LP1	RP	GGGCCCCTGGAACAGAACTT

LP2	RP	CGCCATTAACCTGATGTTCTGGGG
Oligo nucleotides used for colony screening of Vector petM14		
petM14-1	RP	GGGTCTGGAAGTTCTGTT
petM14-2	RP	AAGCTTATATTCCCCAGAAC
Oligonucleotides used for Assembly of two inserts by in situ generation of overhangs and de novo assembly of a promoter and RBS		
Oligonucleotides used for de novo assembly		
Name		Sequence
DuetBU-PR1-importinbeta-P1-FW	RP	AACTGAAGAACCAAGCTTAAAGCTTGCGGCCGCATAATGCTTAAGTCGAACAGAAAGTAA
DuetBU-PR2--P1-rW	RP	ATCCGCTCACAAAGCAAATAAATTTTTTATGATTTTACGATTACTTTCTGTTCGACTTAA
DuetBU-PR3	RP	TATTTGCTTTGTGAGCGGATAACAATTATAATAGATTCAATTGTGAGCGGATAACAATTT
DuetBU-PR4	RP	GGTTAATTTCTCCTCTTTAATGAATTCTGTGTGAAATTGTTATCCGCTCACAA
duet-PR5-alpha-P1fw	DS	TTAAAGAGGAGAAATTAACCATGTCCACGAACGAGAATGC
Oligonucleotides used for in situ generation of overhangs		
importin alpha-pCDF_Fw	DS	CTGGGACCTTTAACTTCTAACTCGAGTCTGGTAAAGAAAC
importin alpha-pCDF_rw	DS	GTTTCTTTACCAGACTCGAGTTAGAAGTTAAAGGTCCCAG
importin beta-P1-FW	DS	CTTTAATAAGGAGATATACCATGGAGCTGATCACCATTCT
importin beta-P1-rW	DS	AGAATGGTGATCAGCTCCATGGTATATCTCCTTATTAAAG
Oligonucleotides used for amplification of double stranded DNA fragment		
LinerpCDF_FW	DS	CTCGAGTCTGGTAAAGAAACCGC
LinearpCDFduet RW	DS	GGTATATCTCCTTATTAAAGTT
importin beta_FW	DS	ATGGAGCTGATCACCATTCTC
importin beta_RW	DS	TTAAGCTTGGTTCTTCAGTTTCCT
importin alpha_FW	DS	ATGTCCACGAACGAGAATGC
importin alpha_RW	DS	TTAGAAGTTAAAGGTCCCAGGAGC
Oligonucleotides used for linearization of Vector pCDF_Duett		
pCDFlin-1	DS	CTCGAGTCTGGTAAAGAAACCGC
pCDFlin-2	DS	GGTATATCTCCTTATTAAAGTT

Supplementary table 3 Vector backbones used in this study

Vector Name	Project	Linearization method	Source
miniTn4001-Puro-1	Fusion Library	Restriction by EcoRV	In house
PetM14	In situ generation of overhangs	PCR with LP1 and LP2	EMBL (6)
pCDF_Duett	Assembly of two inserts	PCR with pCDFlin-1 & pCDFlin-2	Novagen

Supplementary table 4 Influence of oligonucleotide concentration on the assembly of 2 DNA fragments and a de novo built sequence

Oligonucleotide concentration	Colonies obtained
550nM	2
55nM	50
5.5nM	27
0nM	26
55nM No ligase	10

Sequences used to generate the fusion library of 42 constructs.

>**Mpn638**
TAGTTTTTAGAATTTAAAGGAA
>**SP**
GAATTGATCGCCATAGGTTAAAGTAGTATAATATAAACTAATT
>**Leaderless Ldh**
TTTAGATCAAATCTCAATTGGGAACAGTAAATGATTGGTGGCTTAAACCAAATTAAAATGAAGGCA
>**EfTu**
TCAGCAATTACAAAAACAAAACAAATAAAAAATAAGGGAATTACCCCCAAGAAGACCTTTTGTGCTAACGCCAGTTTGGCA
AATCAAGTTCTGATTTTGCAATTATTTTGCTCCATATGAATTACACTACTCCAAGAATTATAAGCCTCTCTACAGCTTTATCTC
AAACTTATGTAAAATTAGAGACGTAATTCAAACAC
>**AckA**
CGTTAATAATGATGATTGAAGCTAGTACAAAAAAGACAATGATTAAGGTCTTATCATTACTTACACCAATTACCCCCATGAG
GGAAGAGATAAAGTAGAGAATAATACCTACAGCTGTCGAAACAGCAGCCACAATATATGAAGTTTTAGCACTCATTGACAG
TTATTAGAGAAAAAATGAGGACTTATTACTAATATTTTATTATTGTTACCTATTAGATAAAA

>**Ldh**
TTTAGATCAAATCTCAATTGGGAACAGTAAATGATTGGTGGCTTAAACCAAATTAAAATGAAGGCAAGCTTGAATAAATTAG
TT

>**Nothing**
ATG

>**KpnI**
GGTACCATG

>**Mpn638 CDS**
ATGACTCCTAAATTAAAGCTTAACACTAAC

>**EfTu CDS**
ATGGCAAGAGAGAAATTTGACCGATCTAAA

>**Ldh CDS**
ATGAAGAGTCTTAAAGTAGCACTCATTGGT

>**AckA CDS**
ATGAACGACAACAAAATTTTAGTAGTCAAT

>**GroEL CDS**
ATGGCAAAGGAATTAGTATTTGGCAAGAAT

>**CI857-CDS**
AGCACTAAGAAGAAGCCATTAACCCAGGAACAGTTGGAAGACGCGAGAAGGTTGAAGGCTATTTACGAAAAGAAAAAGAA
CGAATTGGGTCTAAGTCAGGAATCCGTGGCCGATAAGATGGGAATGGGACAAAGTGGGGTCGGTGCGCTATTTAACGGTATC
AATGCTCTCAACGCCTACAACGCTGCGTTGCTCGCGAAGATTCTAAAGGTGTCTGTTGAAGAATTTAGCCCCAGTATAGCTCG
CGAAATCTACGAAATGTACGAAGCTGTCTCCATGCAGCCTAGTCTGAGGTCCGAATATGAATATCCGGTGTTTTCCCACGTGC
AGGCGGGGATGTTTTCGCCAAAGTTGCGCACTTTCACCAAGGGCGATGCGGAAAGATGAGTGAGTACCACTAAAAAGGCGT

CAGATAGTGCTTTTTGATTAGAAGTGGAAGGGAACAGTATGACGGCGCCCACGGGGTCGAAGCCATCGTTTCCCGATGGGAT
GCTCATTCTCGTAGATCCAGAGCAGGCAGTCGAACCAGGCGATTTCTGTATTGCAAGGTTGGGTGGTGACGAGTTTACTTTTA
AGAAGCTAATAAGAGATAGTGGACAAGTGTTCTTACAGCCGCTCAACCCGCAGTACCCGATGATTCCCTGCAACGAGAGCTG
CTCTGTTGTGGGCAAAGTGATTGCTTCACAATGGCCGGAAGAAACATTTGGTGAACAAAAACTGATTTCCGAAGAAGACCTC
TAGTCTTAATTAGATCGAATTCCTGCAGCCCGGGGGGCAAGAT

>LacI-CDS
GTGAACGTTAAGCCTGTTACCTTATATGATGTGGCTGAATACGCGGGAGTCAGTTATCAAACCGTATCTCGCGTAGTCAACCA
GGCAAGTCACGTTTCTGCCAAAACTAGAGAAAAGGTTGAAGCGGCGATGGCTGAACTGAACTATATTCCTAACAGAGTTGCT
CAGCAATTGGCCGGCAAACAGTCTTTACTCATCGGGGTCGCTACGAGTTCCCTTGCCCTTCACGCACCGAGCCAAATAGTAG
CAGCCATCAAGTCAAGGGCCGACCAGTTAGGAGCTTCGGTTGTGGTTAGTATGGTAGAACGATCCGGCGTGGAGGCGTGTAA
AGCAGCAGTCCATAACTTACTCGCACAGAGAGTGTCAGGTCTGATTATCAATTATCCGCTCGACGATCAGGATGCTATAGCA
GTAGAAGCCGCCTGCACAAATGTACCCGCTTTATTTCTCGACGTGTCCGATCAAACTCCAATTAACTCCATTATTTTCTCCCAC
GAAGATGGGACCCGCTTAGGCGTGGAACACTTAGTTGCACTTGGACACCAACAAATAGCTTTGCTAGCAGGTCCTCTTTCAA
GCGTTTCCGCCCGTTTGCGCTTGGCCGGCTGACACAAGTATTTAACGAGAAATCAAATTCAGCCAATTGCCGAACGCGAGGG
AGACTGGTCGGCCATGTCCGGTTTTCAGCAAACAATGCAGATGTTAAATGAGGGGATTGTTCCAACCGCCATGCTCGTGGCG
AATGACCAAATGGCTCTCGGCGCCATGAGGGCGATAACGGAAAGTGGGCTAAGGGTGGGCGCGGACATTTCGGTTGTTGGG
TACGATGACACAGAAGATTCTTCGTGCTACATTCCCCCACTGACGACGATCAAACAAGATTTTCGTTTGCTCGGGCAAACTTC
GGTCGATCGCTTATTGCAGTTAAGTCAGGGGCAGGCGGTGAAGGGTAATCAGTTGCTGCCTGTTAGTTTAGTAAAACGGAAA
ACGACCTTAGCTCCTAACACTCAGACCGCGTCACCAAGAGCGCTGGCTGATAGTCTCATGCAACTGGCGCGTCAAGTTAGTC
GCCTCGAATCTGGTCAATAGCGATAAGTAGATGCATATCGAATTCCTGCAGCCCGGGGGGCAAGAT

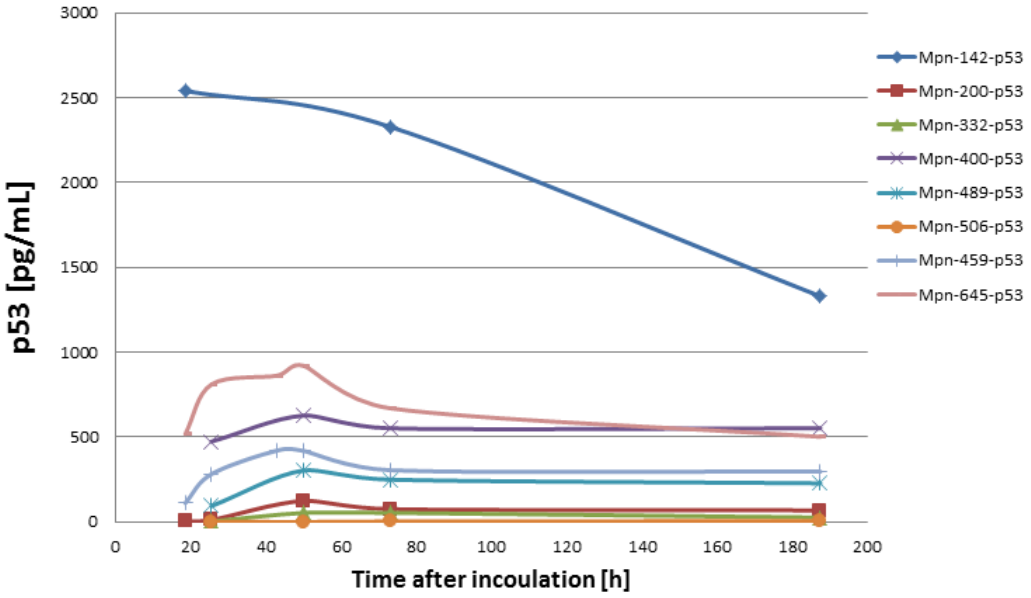
>T7Pol-CDS
AACACCATTAACATAGCGAAGAACGATTTCAGTGATATTGAACTTGCTGCGATACCATTTAACACTTTAGCCGATCATTACGG
TGAAAGATTAGCACGCGAACAATTAGCTCTGGAACACGAAAGTTATGAAATGGGTGAAGCGAGGTTCAGAAAGATGTTCGA
ACGCCAGTTAAAGGCGGGTGAAGTGGCCGATAACGCTGCCGCCAAACCGTTGATTACCACCCTCCTCCCAAAGATGATTGCG
CGCATCAACGATTGGTTCGAGGAAGTTAAAGCGAAAAGAGGTAAGCGCCCCACAGCGTTTCAATTTTTACAAGAGATTAAGC
CGGAAGCGGTGGCTTATATAACGATAAAGACCACACTAGCCTGTTTAACGTCGGCTGACAACACTACCGTGCAAGCTGTTGC
TAGTGCAATTGGTCGTGCCATTGAAGATGAGGCCCGCTTTGGACGTATAAGAGATTTGGAGGCGAAACACTTTAAGAAGAAC
GTCGAAGAGCAATTAAACAAAAGGGTGGGCCACGTGTATAAGAAAGCGTTTATGCAAGTTGTAGAAGCGGACATGTTGTCA
AAAGGGTTATTGGGTGGTGAGGCGTGGAGTTCCTGACACAAGGAAGACAGTATCCATGTGGGCGTAAGATGTATTGAGATG
CTCATCGAAAGTACCGGAATGGTGAGTCTGCACCGCCAAAATGCGGGGGTAGTCGGTCAAGATTCCGAAACCATCGAACTCG
CTCCGGAATATGCAGAGGCAATCGCCACCAGAGCTGGGGCATTAGCTGGTATTTCTCCAATGTTCCAGCCCTGTGTTGTACCG
CCGAAACCCTGAACCGGCATTACCGGCGGTGGTTACTGAGCTAACGGGAGAAGACCGCTCGCCCTGGTTCGTACACACAGTA
AGAAAGCCCTCATGAGATACGAGGATGTTTACATGCCCGAGGTTTACAAGGCCATCAATATCGCTCAGAACACGGCGTGGAA
GATAAATAAGAAAGTATTGGCCGTGGCAAACGTCATTACCAAGTGGAAGCACTGTCCAGTGGAAGATATTCCAGCTATCGAA
AGGGAAGAATTACCGATGAAGCCTGAGGACATTGACATGAACCCAGAAGCGTTGACAGCTTGGAAGAGAGCCGCTGCGGCA
GTTTACCGCAAAGACAAGGCCAGAAAATCTCGCAGAATCAGTTTGGAATTTATGCTCGAACAAGCGAACAAATTTGCCAATC
ATAAGGCGATTTGGTTTCCGTATAACATGGACTGACGCGGTCGCGTTTACGCGGTGTCCATGTTTAATCCTCAAGGCAACGAC
ATGACCAAAGGTTTGTTGACCTTAGCAAAGGGCAAACCCATAGGCAAAGAGGGCTATTACTGGCTCAAGATTCATGGCGCTA
ACTGTGCTGGGGTTGATAAGGTTCCCTTCCCCGAAAGAATTAAGTTCATCGAAGAGAATCACGAGAACATTATGGCTTGCGC
TAAGTCGCCCCTGGAAAACACTTGATGGGCGGAACAGGATTCGCCTTTTTGTTTTCTTGCCTTTTGTTTCGAATACGCGGGGG
TTCAACACCACGGTTTGTCGTACAACTGTAGTTTACCGTTGGCTTTCGACGGTAGCTGTAGTGGCATACAACACTTCTCCGCC
ATGCTGAGAGACGAAGTAGGTGGGAGGGCCGTGAACTTATTGCCCTCTGAAACAGTACAGGATATTTATGGGATCGTGGCGA
AAAAGGTAAATGAAATCCTCCAGGCTGATGCCATCAACGGCACCGACAACGAGGTCGTGACCGTTACTGACGAAAACACCG
GGGAAATATCAGAAAAAGTGAAACTCGGTACGAAGGCGCTTGCTGGCCAATGACTCGCCTATGGTGTGACACGCAGTGTTAC
AAAGAGAAGTGTAATGACCTTGGCTTATGGGTCAAAGGAGTTTGGGTTTAGGCAGCAAGTCTTAGAGGACACCATCCAACCG
GCGATTGATAGTGGTAAGGGTCTAATGTTTACCCAGCCTAACCAAGCAGCCGGCTATATGGCCAAGTTAATCTGGGAGAGTG
TTAGTGTAACGGTCGTTGCGGCGGTAGAAGCCATGAACTGGCTCAAGTCGGCCGCTAAGCTCCTCGCTGCTGAGGTAAAGGA
CAAGAAAACCGGCGAAATTTTGAGAAAAAGGTGCGCCGTTCACTGAGTCACACCCGATGGGTTTCCCGTTTGGCAAGAATAT
AAGAAACCGATTCAAACCAGACTCAATCTGATGTTCCTCGGTCAGTTTCGCCTTCAACCAACCATCAATACCAACAAAGATT
CTGAGATCGACGCGCACAAGCAGGAATCCGGGATCGCCCCCAATTTTGTCCATTCCCAGGATGGGAGTCACTTACGCAAGAC
AGTGGTGTGAGCGCACGAAAAATACGGCATTGAATCTTTTGCTTTAATCCATGACTCCTTTGGAACCATTCCCGCCGACGCCG
CTAATCTCTTTAAAGCAGTGAGAGAAACGATGGTAGATACCTATGAAAGCTGTGATGTATTAGCCGACTTCTACGATCAGTTT
GCCGATCAATTACATGAATCACAGTTAGATAAAATGCCTGCCTTACCGGCTAAGGGAAATCTTAACTTGAGGGATATACTAG
AATCCGATTTTGCCTTCGCGTAGGGATAAGTAGGAATTCGGTCCCTCTCCTGGGACCTTTTTTTTTGATCACCAGAAGGTTATC
GAATTCCTGCAGCCCGGGGGGCAAGAT

>Tet-CDS
TCCAGGCTCGACAAGTCAAAGGTGATTAACTCTGCCTTAGAACTCTTAAATGAGGTAGGTATTGAGGGTTTAACCACCAGAA
AACTCGCACAGAAACTCGGGGTGGAACAGCCGACGCTTTATTGACACGTCAAGAACAAGAGGGCACTCCTGGACGCTCTCG
CGATTGAGATGCTGGATAGGCATCATACTCACTTCTGTCCCCTCGAAGGCGAAAGTTGACAAGACTTTTTGAGAAACAACGC
TAAATCCTTTAGATGTGCCTTACTTTCGCATAGAGACGGGGCTAAGGTGCACTTAGGTACTAGGCCAACAGAAAAGCAGTAT
GAAACTTTAGAAAATCAATTGGCTTTTCTTTGCCAACAAGGTTTTAGTCTTGAAAATGCGTTGTACGCGCTCTCGGCTGTAGG
GCACTTCACGTTGGGCTGCGTCTTGGAAGACCAAGAGCACCAAGTTGCGAAGGAAGAACGTGAAACCCCGACCACCGACAG
TATGCCTCCCTTACTGCGCCAAGCGATCGAATTATTTGACCACCAAGGTGCTGAGCCGGCTTTTCTCTTCGGTCTGGAACTCA
TTATTTGTGGTTTAGAGAAGCAATTGAAGTGCGAGTCGGGAAGTGCTTACAGCAGAGCTAGGACAAAGAATAACTATGGTAG
CACCATCGAGGGCCTTTTAGACCTCCCCGATGATGACGCACCGGAGGAAGCCGGCTTGGCTGCACCCCGTTTGTCATTTCTCC
CAGCCGGTCACACCCGCAGACTTTCTACGGCGCCACCAACCGACGTATCGCTTGGGGATGAACTGCACTTGGACGGTGAAGA
CGTTGCTATGGCCCACGCTGATGCTTTGGACGATTTTGACTTAGACATGCTTGGCGATGGTGATAGCCCTGGTCCGGGTTTCA
CCCCGCACGATTCCGCCCCCTACGGGGCCTTGGATATGGCTGATTTTGAATTTGAACAGATGTTTACGGATGCTCTGGGAATC
GACGAATACGGCGGGTAGCTATAACTAGACGCGTAATCGAATTCCTGCAGCCCGGGGGGCAAGAT

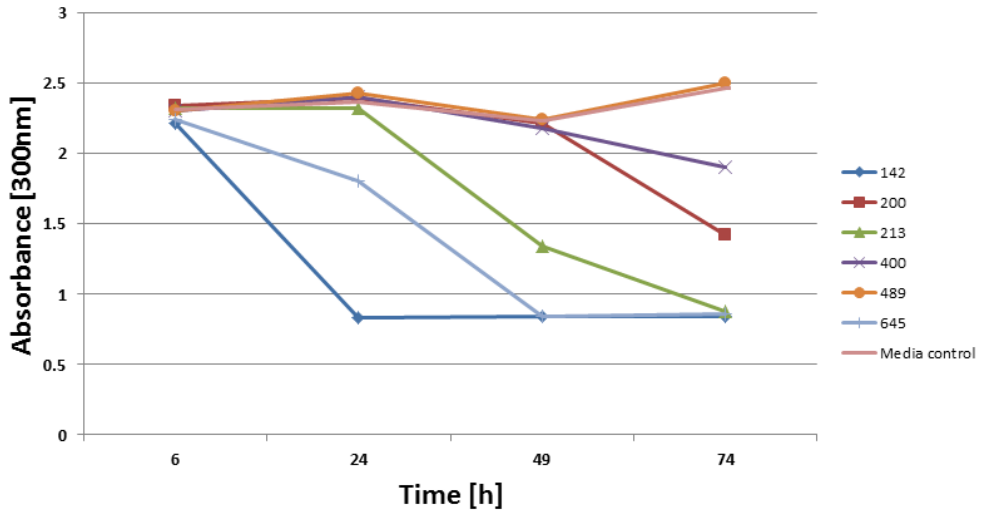
>T7Lys-CDS

GCTCGGGTGCAGTTCAAGCAACGCGAAAGTACCGATGCGATTTTTGTGCACTGTAGTGCCACAAAACCTAGTCAAAATGTGG
GGGTACGCGAGATTCGCCAGTGGCACAAGGAACAAGGCTGACTTGACGTTGGTTACCACTTCATTATCAAGCGCGATGGAAC
CGTAGAAGCTGGTCGCGATGAGATGGCAGTCGGAAGCCATGCAAAAGGGTATAACCACAATTCCATCGGGGTTTGCTTGGTT
GGTGGTATTGACGACAAAGGCAAATTCGACGCTAATTTTACCCCAGCGCAAATGCAAAGTTTACGCTCATTATTGGTCACGTT
GCTGGCGAAGTATGAAGGGGCTGGCCTCCGCGCACACCATGAAGTTGCGCCAAAAGCCTGTCCATCCTTTGACTTGAAGCGA
TGATGGGAAAAGAATGAATTAGTAACATCCGATCGCGGCGACTACAAGGACGACGACGATAAGGGATAGACATAAGTAGAA
GCTTATCGAATTCCTGCAGCCCGGGGGGCAAGAT

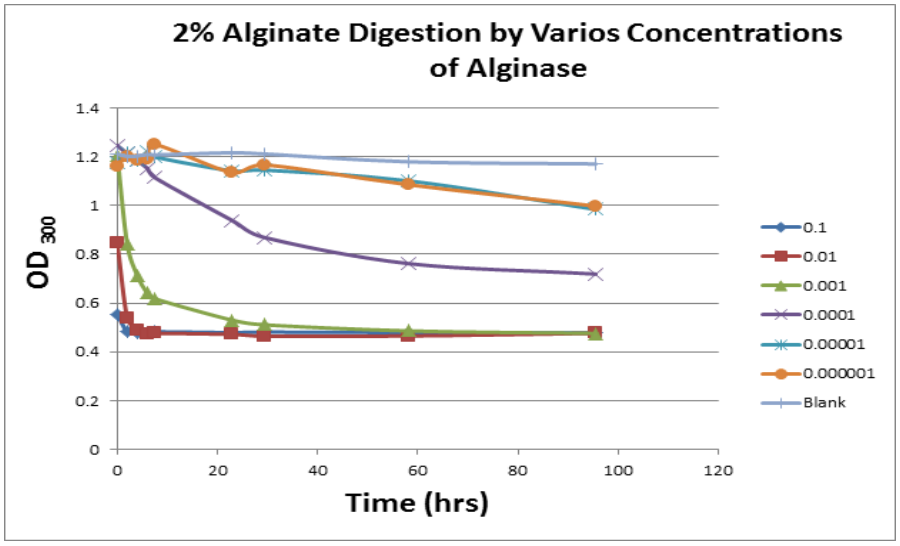
Supplementary tables Engineering *Mycoplasma pneumoniae* as therapeutic vector for lung diseases



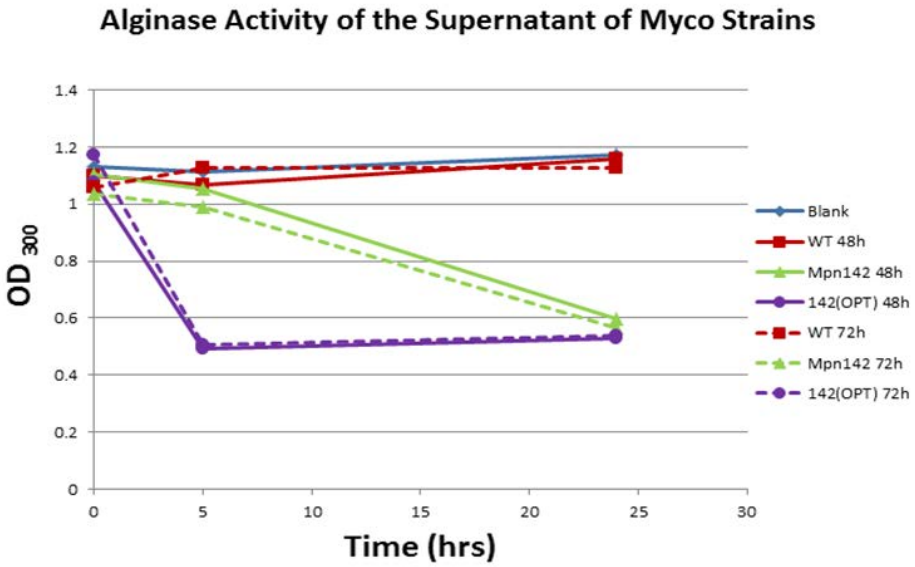
Supplementary figure 4 Results for the secretion of p53 by modified *M. pneumoniae* into Mycoplasma Hayflick media. The amounts were determined at different time points using a Roche p53 pan ELISA sandwich assay. The construct Mpn-142-p53 saturated the assay already at the 24h time point.



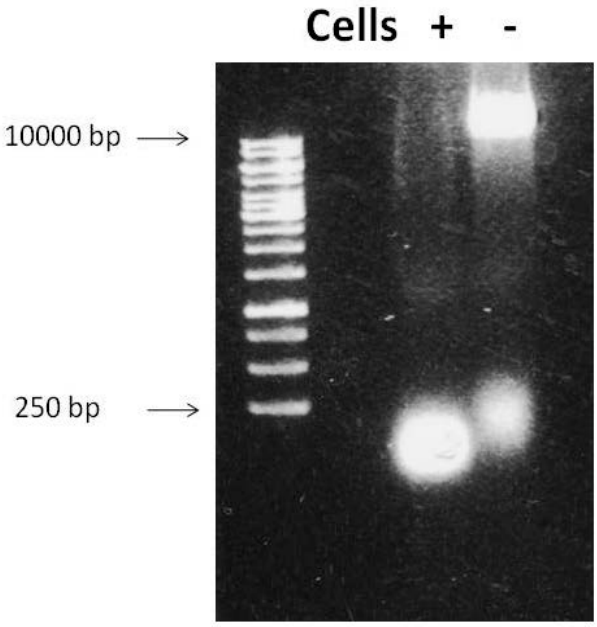
Supplementary figure 5 Results of alginate lyase activity assay in full medium from cultures secreting alginate lyase (non-adherent strain / less immunogenic strain). All cultures were normalized to the same inoculation count at day 0 and over the next 5 days samples of the media were taken. The assay is based on the degradation of polymeric alginate to shorter sugar fragments.



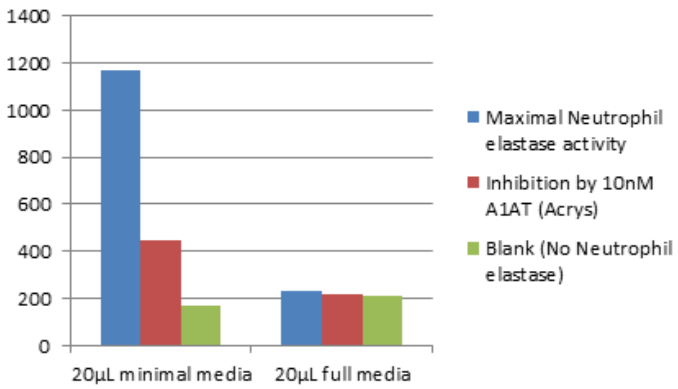
Supplementary figure 6 Standard curve of alginate lyase activity based on turbidity measurements after addition of Acidic BSA. For the standards Alginate Lyase purchased from Sigma was used (A1603)



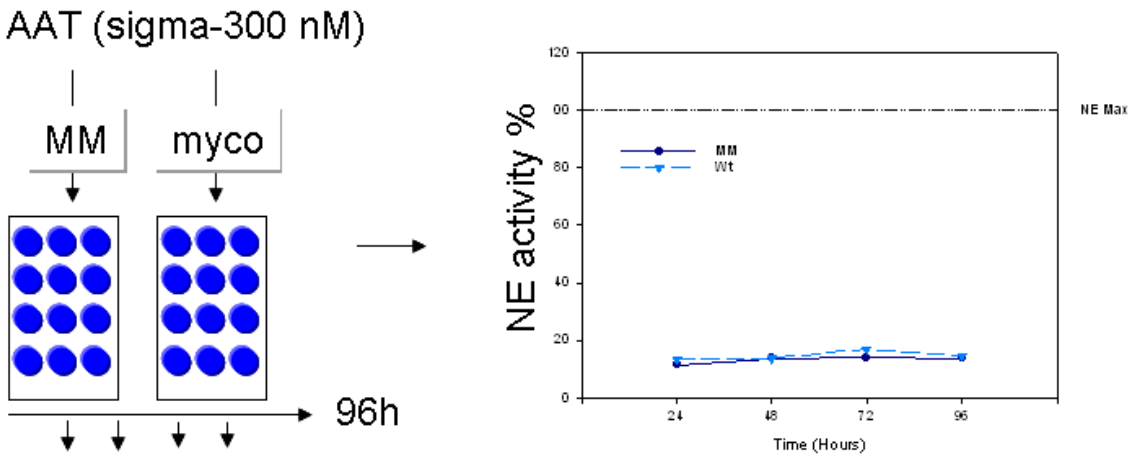
Supplementary figure 7 Quantification of alginate lyase activity from supernatant of different engineered Mycoplasma strains



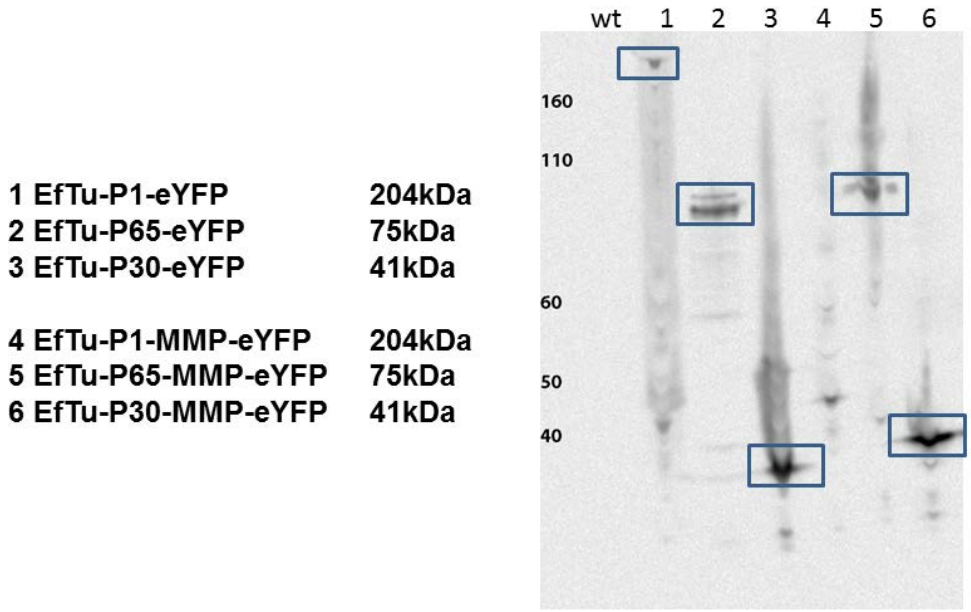
Supplementary figure 8 Comparison of the stability of λ-Phage DNA in Hayflick media in which *M. pneumoniae* cells were grown (+) or a blank of fresh media (-). The DNA was incubated for 1.5h at 37°C before it was inactivated for 20 minutes at 80°C and analyzed on agarose gel.



Supplementary figure 9 Neutrophil elastase activity and inhibition by recombinant A1AT, in Full (modified Hayflick media with Serum) and minimal Media. While neutrophil elastase is active in minimal media it is completely inhibited in Hayflick media.



Supplementary figure 10 Neutrophil elastase protease activity in measured in presence of 300 nM of AAT pre-incubated in wt supernatant or medium alone (Minimal media MM) for 0 (T0), 24, 48, 72 and 96h. A1AT effect on NE has been evaluated at 300 nm.



Supplementary figure 11 Western blot of *M. pneumoniae* cultures transformed with 6 constructs for the conditional protein release upon exposure to MMP-9. On the left side the constructs are listed next to the expected weight of the fusion protein. Additional the expected band is marked with a square on the blot for each construct.

Supplementary table 5 List of proteins that were secreted with the highest confidence. All proteins were detected in both time points and both repeats with a p-value for secretion <0.0001. Column 1 gives the identifier of the gene. A score higher than 0.5 in column 2 indicates that the protein is predicted to be secreted by the secretome P algorithm. Column 3 shows if the Signal P servers predicted a classical signal peptide at the N-terminus. Column 4 gives the COG category and the last column indicates the function if available

Name	SecretomeP Score	Signal P prediction	COG category	Function
MPN036	0.667	N	M	Conserved hypothetical protein MPN_036
MPN083	0.916	N	S	Uncharacterized lipoprotein MPN_083
MPN084	0.891	N	S	Conserved hypothetical lipoprotein MPN_084

MPN142	0.904	Y	M	Mgp-operon protein 3 (Mgp3) (ORF-3 protein)
MPN213	0.778	Y	M	Conserved hypothetical protein MPN_213
MPN398	0.943	Y	A	Uncharacterized protein MPN_398
MPN400	0.895	N	A	Conserved hypothetical protein MPN_400
MPN489	0.892	N	M	Uncharacterized lipoprotein MPN_489
MPN592	0.911	Y	O	Conserved hypothetical lipoprotein MPN_592
MPN642	0.841	Y	M	Conserved hypothetical lipoprotein MPN_642
MPN647	0.935	Y	M	Conserved hypothetical lipoprotein MPN_647

Supplementary table 6 Overview of constructs made for the secretion of the poly M alginate lyase A1-III

Secretion constructs and strains with alginate lyase A1-III						
Name strain	Signal peptide	Signal Type	Promoter	Promoter Length [Bp]	Signal length [AA]	Cargo
Mpn142(Opt)-A1-III	Mpn142(Opt)	SpI	EfTu	150	31 AA	A1-III
Mpn142-A1-III	Mpn142	SpI	EfTu	150	31 AA	A1-III
Mpn152-A1-III	Mpn152	SpII	Mpn152	106	33 AA	A1-III
Mpn200-A1-III	Mpn200	SpII	Mpn200	197	28 AA	A1-III
Mpn213-A1-III	Mpn213	SpI	EfTu	150	34 AA	A1-III
Mpn332-A1-III	Mpn332	Neg control	EfTu	150	50AA	A1-III
Mpn400-A1-III	Mpn400	SpI	Mpn400	160	41 AA	A1-III
Mpn489-A1-III	Mpn489	SpII	Mpn489	100	34 AA	A1-III
Mpn506-A1-III	Mpn506	SpII	Mpn506	117	28 AA	A1-III
Mpn588-A1-III	Mpn588	SpI	Mpn588	100	30AA	A1-III
Mpn592-A1-III	Mpn592	SpII	Mpn592	100	36 AA	A1-III
Mpn645-A1-III	Mpn645	SpII	EfTu	150	31 AA	A1-III

Supplementary table 7 Overview of constructs made for the secretion of α-anti antitrypsin (A1AT)

Secretion constructs and strain with A1AT						
Name strain	Signal peptide	Signal Type	Promotor	Promoter Length [Bp]	Signal length [AA]	Cargo
Mpn142(Opt)-A1AT	Mpn142(Opt)	SpI	EfTu	150 Bp	31 AA	A1AT
Mpn142-A1AT	Mpn142	SpI	EfTu	150 Bp	31 AA	A1AT
Mpn200-A1AT	Mpn200	SpII	Mpn200	197 Bp	28 AA	A1AT
Mpn332-A1AT	Mpn332	Neg control	EfTu	150 Bp	50AA	A1AT
Mpn400-A1AT	Mpn400	SpI	Mpn400	160 Bp	41 AA	A1AT
Mpn489-A1AT	Mpn489	SpII	Mpn489	100 Bp	34 AA	A1AT
Mpn506-A1AT	Mpn506	SpII	Mpn506	117 Bp	28 AA	A1AT
Mpn645-A1AT	Mpn645	SpII	EfTu	150 Bp	31 AA	A1AT

Supplementary table 8 Overview of constructs made for the secretion of the tumor suppressor protein 53 (p53)

Secretion constructs and strain with p53						
Name strain	**Signal peptide**	**Signal Type**	**Promoter**	**Promoter Length [Bp]**	**Signal length [AA]**	**Cargo**
MPN142-p53	Mpn142	SpI	EfTu	150	31 AA	p53
MPN200-p53	Mpn200	SpII	Mpn200	197	28 AA	p53
Mpn332-p53	Mpn332	Neg control	EfTu	150	50AA	p53
MPN400-p53	Mpn400	SpI	Mpn400	160	41 AA	p53
MPN489-p53	Mpn489	SpII	Mpn489	100	34 AA	p53
MPN506-p53	Mpn506	SpII	Mpn506	117	28 AA	p53
MPN459-p53	Mpn459	SpI	EfTu	150	57 AA	p53
MPN645-p53	Mpn645	SpII	EfTu	150	31 AA	p53

Supplementary table 9 Table showing all data from the secretome experiment. The column SecretomeP score indicates the *in silico* prediction of the Secretome P server a score over 0.5 indicates possible secretion. The Signal P column indicates if the SignaP 4.0 server predicted a secretion signal for the Sec pathway. The ratio MS column gives the ratio obtained in the different experiments, T1-1 stands for time point 1 in the first repeat of the experiment, T2-1 is the second time point of the first repeat of the experiment. The p-value is calculated based on bimodial distribution of the ratios in each experiment and represents the likeliness of a protein to be cytosolic.

Name	SecretomeP Score	Signal P	Ratio MS T1-1	P-Value T1-1	Ratio MS T2-1	P-Value T2-1	Ratio MS T1-2	P-Value T1-2	Ratio MS T2-2	P-Value T1-2	COG category	Protein name
MPN001	0.160	N	0.507	0.323550	1.641	0.096926	0.843	0.634208	0.181	0.487571	L	dnaN
MPN001a	0.031	N									--	New ORF
MPN002	0.213	N									L	xdj1,CbpA
MPN003	0.142	N	0.512	0.317003	0.683	0.596581	1.114	0.371882	0.153	0.626948	L	gyrB
MPN004	0.066	N	1.645	0.000006	1.736	0.073287	3.669	0.000000	0.159	0.596739	L	gyrA
MPN005	0.122	N	0.216	0.705956	0.262	0.822212	0.800	0.673843	0.185	0.469932	J	serS
MPN006	0.077	N							0.136	0.703	F	tmk
MPN007	0.245	N									L	holB
MPN008	0.120	N			0.549	0.678					J	thdF
MPN009	0.116	N									L	TatD,yabD, YcfH
MPN010	0.685	N									N	
MPN011	0.640	Y									M	
MPN012	0.729	Y									M	
MPN013	0.892	N									N	
MPN014	0.142	N									L	
MPN015	0.148	N							0.086	0.877	J	rimK
MPN016	0.248	N	0.260	0.652578					0.130	0.727674	J	rimK
MPN017	0.180	N									H	mtd1
MPN018	0.869	N	0.126	0.802596	0.339	0.788089	1.552	0.079167	0.072	0.908455	I	pmd1
MPN019	0.764	N			0.409	0.754			0.077	0.898	I	msbA
MPN020	0.139	N	0.178	0.749497	0.839	0.497304	0.972	0.509451	0.174	0.522411	L	yb95, HepA, RapA,
MPN021	0.216	N	0.384	0.486049	0.359	0.778545	1.135	0.352647	0.155	0.614916	O	dnaJ
MPN022	0.329	N	0.234	0.685015	1.180	0.288883	0.904	0.576576	0.071	0.910556	O	pip

MPN023	0.187	N	0.448	0.399171	0.449	0.733123	0.789	0.683568	0.084	0.883584	J	metS
MPN024	0.168	N			0.276	0.816	0.498	0.884	0.223	0.290	K	rpoE
MPN025	0.228	N	0.255	0.659043	0.476	0.718356	0.696	0.760530	0.175	0.518587	G	tsr, fba
MPN026	0.155	N	0.393	0.473680	0.613	0.639973	0.790	0.682791			J	EngD..
MPN027	0.289	N									J	rimL
MPN028	0.909	N									M	trsB
MPN029	0.091	N	1.458	0.000097	2.266	0.010541	1.951	0.008229	0.623	0.000000	J	efp
MPN030	0.088	N									J	NusB
MPN031	0.967	N									A	
MPN032	0.129	N									O	YfkM
MPN033	0.059	N	0.246	0.670412	0.313	0.799837	0.561	0.851375	0.112	0.797461	F	upp
MPN034	0.146	N	0.402	0.461137	0.854	0.487538	1.563	0.075382	0.240	0.219755	L	polC
MPN035	0.422	Y									M	
MPN036	0.667	N	77.438	0.000000	38.000	0.000000	23.243	0.000000	3.898	0.000000	M	
MPN037	0.304	Y									M	
MPN037a	0.171	N									--	New ORF
MPN038	0.907	N									N	
MPN039	0.412	Y									M	
MPN040	0.704	N									M	
MPN041	0.847	N									M	
MPN042	0.725	Y									M	
MPN043	0.976	N			0.740	0.561	1.539	0.084			G	glpF
MPN044	0.085	N	0.089	0.836084	0.791	0.528369	1.173	0.318209	0.143	0.672194	F	tdk
MPN044a	0.616	Y									--	New ORF
MPN045	0.159	N									J	hisS
MPN046	0.105	N			0.440	0.738	0.561	0.851	0.103	0.829	J	aspS
MPN047	0.339	N	2.258	0.000000			4.870	0.000000			H	PncB
MPN047a	0.307	N									--	New ORF
MPN048	0.769	N									M	
MPN048a	0.675	N									--	New ORF
MPN049	0.749	Y									M	

MPN050	0.277	N	0.952	0.023570	0.679	0.599172	2.838	0.000002	0.237	0.233452	C	glpK
MPN051	0.133	N	2.477	0.000000	1.399	0.181816	3.721	0.000000	0.143	0.672984	C	glpD
MPN052	0.909	Y	0.296	0.605838	0.478	0.717475	1.432	0.132619	0.174	0.520839	S	
MPN053	0.446	N	0.984	0.017999	2.699	0.001330	1.659	0.046829	0.166	0.561165	G	ptsH
MPN053a	0.020	N									#N/A	#N/A
MPN054	0.306	Y									M	
MPN055	0.375	N	0.102	0.825188	0.485	0.713533			0.224	0.286640	E	potA
MPN056	0.967	N									E	potB
MPN057	0.962	N									E	potI
MPN058	0.898	Y	0.469	0.371734	1.236	0.258889	1.941	0.008805	0.118	0.777872	R	potD
MPN059	0.247	N	0.195	0.730640	0.254	0.825337	0.857	0.621522	0.140	0.686369	O	Gcp
MPN060	0.335	N			0.187	0.852	0.700	0.757			H	metX
MPN060a	0.960	N									--	New ORF
MPN061	0.048	N	0.117	0.810984	0.539	0.683407	0.957	0.524514	0.239	0.224036	U	ffh
MPN062	0.043	N	0.251	0.663291	0.466	0.723742	0.619	0.815336	0.264	0.142097	F	deoC
MPN063	0.038	N	0.855	0.049139	0.584	0.656913	1.189	0.304425	0.147	0.655379	F	deoD
MPN064	0.125	N	0.761	0.091483	0.434	0.740965	0.830	0.646429	0.177	0.506956	F	deoC
MPN065	0.096	N							0.255	0.169	F	deoA
MPN066	0.245	N	0.998	0.016062	1.310	0.221811	1.893	0.012115	0.181	0.485684	G	cdd
MPN067	0.130	N	0.453	0.392658	0.781	0.534858	0.579	0.840563	0.273	0.118360	K	nusG
MPN068	0.937	N									U	secE
MPN069	0.785	N									J	rpmG2
MPN070	0.910	N									M	
MPN071	0.172	N									R	yabC
MPN072	0.183	N									J	rnmV
MPN073	0.079	N	0.341	0.544474	0.465	0.724357	0.963	0.518113	0.221	0.298051	F	prs
MPN074	0.338	N									J	smpB
MPN075	0.384	Y									M	ywdF
MPN076	0.963	N	0.123	0.805612	0.354	0.781126	0.905	0.575099	0.069	0.916240	G	uhpT
MPN077	0.966	N			0.725	0.571					R	uhpT
MPN078	0.886	N							0.058	0.935	G	fruA

MPN079	0.167	N	0.268	0.642490	0.328	0.793042	0.497	0.884917	0.174	0.524285	G	fruK
MPN080	0.880	N									L	
MPN081	0.184	N							0.223	0.291	E	glnQ
MPN082	0.613	N	2.417	0.000000	3.118	0.000117	5.362	0.000000	0.281	0.099162	G	tklB
MPN083	0.916	N	5.466	0.000000	4.176	0.000000	2.618	0.000026	1.015	0.000000	S	
MPN084	0.891	N	34.539	0.000000	13.091	0.000000	8.007	0.000000	5.024	0.000000	S	
MPN085	0.964	N									M	
MPN086	0.954	N									M	
MPN087	0.961	N									M	
MPN088	0.517	N									S	
MPN089	0.407	N									V	hsdS
MPN090	0.964	N									M	
MPN091	0.700	N									S	
MPN091a	0.026	N									#N/A	#N/A
MPN092	0.476	N									M	
MPN093	0.873	N									M	
MPN094	0.772	N									N	
MPN095	0.960	N									E	
MPN096	0.955	N									E	
MPN097	0.891	N									M	
MPN098	0.907	N									M	
MPN099	0.898	N									M	
MPN100	0.790	N									N	
MPN101	0.918	N									M	
MPN102	0.891	N									M	
MPN103	0.963	N									S	
MPN104	0.365	N									N	
MPN104a	0.904	N									--	New ORF
MPN105	0.125	N	0.128	0.801044	0.175	0.856274	0.640	0.801233	0.143	0.671647	J	pheS
MPN106	0.124	N	0.247	0.668178	0.539	0.683467	1.114	0.372296	0.161	0.589635	J	pheT
MPN107	0.112	N									L	

MPN108	0.467	N									L	
MPN109	0.305	N									V	
MPN110	0.527	N									V	
MPN111	0.579	N									V	
MPN112	0.967	N									U	
MPN113	0.971	N									U	
MPN114	0.645	N									I	
MPN115	0.117	N	0.220	0.701022	0.754	0.552159	0.595	0.830592	0.248	0.191294	J	infC
MPN116	0.936	N									J	rpmI
MPN117	0.817	N									J	rpLT
MPN118	0.345	N					0.540	0.863			L	rnhC
MPN119	0.903	N	0.917	0.030948	1.630	0.099973	2.759	0.000005	0.290	0.081534	O	
MPN120	0.406	N	0.928	0.028394	1.033	0.374915	1.378	0.163203	0.262	0.146894	O	grpE
MPN121	0.930	N									S	mpn121
MPN122	0.129	N									L	parB
MPN123	0.130	N	0.142	0.787232	0.460	0.727332					L	parC
MPN124	0.127	N					0.776	0.695			K	hrcA
MPN125	0.214	N									L	uvrC
MPN126	0.187	N	1.380	0.000274	1.707	0.079925	3.759	0.000000	0.276	0.111934	R	ysnB
MPN127	0.378	N									N	
MPN128	0.893	N									M	
MPN129	0.932	N									M	
MPN130	0.802	N									N	
MPN131	0.942	N									M	
MPN132	0.899	N									M	
MPN132a	0.249	Y									#N/A	#N/A
MPN133	0.880	N									L	
MPN134	0.300	N	0.207	0.716758	0.520	0.694453	0.853	0.625230	0.199	0.398080	G	ugpC
MPN135	0.952	N									G	ugpA
MPN136	0.944	N									G	ugpE
MPN137	0.124	N									N	

MPN138	0.735	N									N	
MPN139	0.803	N									N	
MPN140	0.172	N	0.182	0.745599	0.362	0.777113	0.771	0.699217	0.091	0.863230	J	ytqI
MPN141	0.934	Y	3.997	0.000000	4.465	0.000000	2.004	0.005739	1.390	0.000000	M	mgpA
MPN142	0.904	Y	12.938	0.000000	7.719	0.000000	2.566	0.000044	6.064	0.000000	M	
MPN143	0.889	N									S	
MPN144	0.916	N									M	
MPN145	0.580	N									N	
MPN146	0.859	N									M	
MPN147	0.910	N									M	
MPN148	0.851	N									M	
MPN149	0.910	N									M	
MPN150	0.877	N									M	
MPN151	0.244	N									N	
MPN152	0.914	Y	12.265	0.000000			14.678	0.000000			M	
MPN153	0.156	N	0.563	0.258259	0.854	0.487802	1.198	0.295996	0.121	0.765510	L	
MPN153a	0.577	N									--	New ORF
MPN154	0.774	N	0.270	0.639070	0.301	0.805275	0.573	0.843865	0.192	0.433022	K	nusA
MPN155	0.124	N	0.358	0.522085	0.557	0.672873	0.832	0.644295	0.170	0.542630	J	infB
MPN155a	0.344	N									--	New ORF
MPN156	0.038	N									J	rbfA
MPN156a	0.019	N									--	New ORF
MPN157	0.968	N	0.040	0.874542	0.120	0.875264	0.898	0.582073	0.037	0.962219	M	
MPN158	0.439	N									H	yaaC
MPN159	0.789	N	0.233	0.685300	0.190	0.850879	0.973	0.509044	0.081	0.888712	V	tlyC
MPN160	0.963	N									M	
MPN161	0.307	N					2.090	0.003	0.086	0.878	S	
MPN162	0.842	Y									M	
MPN163	0.927	N									M	
MPN164	0.071	N	0.150	0.779357	0.679	0.599360	0.467	0.898736	0.252	0.179936	J	nusE
MPN165	0.702	N	0.142	0.787115	0.734	0.564829	0.826	0.649659	0.108	0.811651	J	rplC

MPN166	0.811	N	0.108	0.819889	0.811	0.515726	0.949	0.532806	0.119	0.771543	J	rplD
MPN167	0.623	N	0.053	0.865102	0.095	0.883533	0.798	0.675267	0.042	0.956719	J	rplW
MPN168	0.332	N	0.064	0.856418	0.254	0.825522	1.246	0.256545	0.056	0.938383	J	rplB
MPN169	0.754	N									J	rpsS
MPN169a	0.872	N									--	New ORF
MPN170	0.751	N	0.125	0.804207	1.374	0.192422	1.471	0.113075	0.069	0.914533	J	rplV
MPN171	0.256	N	0.073	0.849427	1.278	0.237514	0.691	0.763751	0.188	0.452557	J	rpsC
MPN172	0.569	N			0.523	0.693					J	rplP
MPN173	0.746	N									J	rpmC
MPN174	0.693	N									J	rpsQ
MPN175	0.035	N									J	rplN
MPN176	0.930	N									J	rplX
MPN177	0.054	N	0.103	0.823991	0.528	0.689486	0.849	0.628616	0.169	0.548886	J	rplE
MPN178	0.309	N									J	rpsN
MPN179	0.043	N									J	rpsH
MPN180	0.388	N	0.614	0.204810	1.839	0.052755	0.677	0.774615	0.295	0.071661	J	rplF
MPN181	0.284	N					1.055	0.428			J	rplR
MPN182	0.188	N	0.034	0.878446	0.671	0.604460			0.065	0.922222	J	rpsE
MPN183	0.722	N							0.216	0.319	J	rplO
MPN184	0.969	N									U	secY
MPN185	0.136	N	0.320	0.573536	0.615	0.638514	0.437	0.911251	0.249	0.187699	F	adk
MPN186	0.114	N	1.125	0.004946	5.769	0.000000	5.446	0.000000	0.078	0.896913	J	map
MPN187	0.155	N									J	infA
MPN188	0.403	N									J	rpmJ
MPN189	0.076	N									J	rpsM
MPN190	0.941	Y	0.126	0.802417					0.131	0.726648	J	rpsK
MPN191	0.049	N	0.525	0.302327	1.328	0.213436	0.781	0.690157	0.215	0.323371	K	rpoA
MPN192	0.784	N									J	rplQ
MPN193	0.125	N	0.122	0.806491	0.553	0.675086	0.732	0.731957	0.067	0.919125	P	cbiO 1
MPN194	0.070	N	0.210	0.712770	0.327	0.793755	0.815	0.660528	0.151	0.633317	P	CbiO 2
MPN195	0.890	N	0.064	0.856137			1.071	0.412933			P	cbiQ

MPN196	0.254	N									J	hisT
MPN197	0.506	N	0.185	0.742166	0.500	0.705132	0.726	0.736950	0.163	0.576224	E	pepF
MPN198	0.477	N									V	mte1
MPN198a	0.022	N									#N/A	#N/A
MPN199	0.874	Y									M	
MPN199a	0.078	N									--	New ORF
MPN200	0.911	Y	7.683	0.000000	2.517	0.003358			0.638	0.000000	M	
MPN200a	0.222	N									--	New ORF
MPN201	0.186	N									V	
MPN202	0.921	N									M	
MPN203	0.907	N									M	
MPN204	0.847	N									N	
MPN205	0.910	N									M	
MPN206	0.149	N									S	
MPN207	0.905	N	0.628	0.190898	0.545	0.680168	1.683	0.041282	0.083	0.884979	G	ptsG
MPN207a	0.955	N									--	New ORF
MPN208	0.094	N			1.661	0.092	1.169	0.322	0.203	0.382	J	rpsB
MPN208a	0.662	N									#N/A	#N/A
MPN209	0.900	N									P	mgtA
MPN210	0.052	N	0.144	0.785255	0.356	0.779919	0.692	0.763511	0.122	0.761237	U	secA
MPN211	0.223	N	0.514	0.314776	0.647	0.618845	0.787	0.684824	0.234	0.244571	L	uvrB
MPN212	0.930	N									S	
MPN213	0.778	Y	14.575	0.000000	12.941	0.000000	7.246	0.000000	2.206	0.000000	M	
MPN214	0.975	N									M	OppA??
MPN215	0.937	Y			0.427	0.744			0.038	0.961	E	oppB
MPN216	0.921	N	0.175	0.753251	0.410	0.753171	1.365	0.171362	0.044	0.953893	E	oppC
MPN216a	0.207	N									--	New ORF
MPN217	0.175	N	0.135	0.794261	0.600	0.647642	0.883	0.596892	0.109	0.809341	E	oppD
MPN218	0.245	N	0.125	0.803441	0.371	0.772752	1.003	0.479524	0.064	0.925163	E	oppF
MPN219	0.299	N	0.177	0.750195	0.295	0.807790	0.599	0.828033	0.127	0.742341	J	rplK
MPN220	0.099	N	0.190	0.736197	0.472	0.720722	0.430	0.914070	0.180	0.493038	J	rplA

MPN221	0.053	N	0.306	0.592828			0.694	0.762140	0.213	0.334644	J	pth
MPN222	0.169	N									J	yacA
MPN223	0.055	N			0.938	0.434			0.219	0.307	T	hprK
MPN224	0.955	N									U	lgt
MPN225	0.603	N			0.807	0.518					J	rpsL
MPN226	0.353	N	0.129	0.799687			1.022	0.460337	0.162	0.581441	J	rpsG
MPN227	0.068	N	1.714	0.000002	1.614	0.104630	3.643	0.000000	0.221	0.298376	J	fus
MPN228	0.875	N	0.228	0.692151	1.707	0.079906	0.365	0.936588	0.291	0.080183	J	rpsF
MPN229	0.898	N	1.762	0.000001	0.737	0.562893	0.743	0.722621	0.235	0.241296	L	ssbA
MPN230	0.092	N	0.155	0.773984	0.514	0.697356	0.902	0.578576	0.128	0.736700	J	rpsR
MPN231	0.168	N			0.284	0.813	0.521	0.873	0.195	0.419	J	rplI
MPN232	0.122	N	0.242	0.674628	0.236	0.833021	0.356	0.939148	0.258	0.158263	L	dnaB
MPN233	0.876	Y	0.439	0.411602	0.652	0.615787	1.058	0.425722	0.444	0.000437	M	
MPN233a	0.921	N									--	New ORF
MPN234	0.756	N									M	
MPN235	0.458	N									L	ung
MPN236	0.937	N									J	gatC
MPN237	0.455	N	0.798	0.072580	2.884	0.000480	1.403	0.148838	0.128	0.735672	J	gatA
MPN238	0.180	N	0.493	0.340963	0.534	0.686121	0.888	0.591389	0.169	0.548911	J	gatB
MPN239	0.167	N	0.209	0.714250	0.334	0.790456	0.991	0.491335	0.111	0.803089	K	gntR
MPN240	0.060	N	2.399	0.000000	1.760	0.068054	2.994	0.000000	0.256	0.167223	O	trxB
MPN241	0.311	N									R	WhiA
MPN242	0.951	N									U	secG
MPN243	0.098	N	0.187	0.739374	0.319	0.797237	0.952	0.528954	0.128	0.738149	K	vacB
MPN244	0.606	N	0.139	0.789955	0.366	0.775213	0.889	0.590805	0.058	0.934778	L	disA
MPN245	0.211	N	0.463	0.379858	0.492	0.709907	1.360	0.175030	0.090	0.866611	J	def
MPN246	0.117	N	0.194	0.731447	0.470	0.722086	0.848	0.630148	0.214	0.328477	F	gmk
MPN247	0.295	N	0.740	0.104255			1.581	0.069082	0.191	0.440302	T	ptc1
MPN248	0.160	N	0.118	0.810608			0.613	0.819331			T	prkC
MPN249	0.120	N									J	CpgA
MPN250	0.596	N	1.886	0.000000	0.932	0.438057	1.796	0.021985	0.104	0.824056	G	pgiB

MPN251	0.085	N					0.781	0.690	0.222	0.293	G	cfxE
MPN252	0.103	N	0.243	0.673132	0.310	0.801229	0.950	0.531583	0.094	0.856929	J	asnS
MPN253	0.963	N									I	pgsA
MPN253a	0.043	N									--	New ORF
MPN254	0.047	N			1.772	0.065			0.284	0.094	L	cinA
MPN255	0.166	N	0.404	0.459215	0.520	0.694384	0.747	0.719687	0.197	0.408875	I	ygbP
MPN256	0.494	N	0.217	0.704816	0.582	0.658484	0.423	0.916423	0.348	0.016672	R	
MPN257	0.132	N	0.275	0.633144			0.945	0.536747	0.229	0.261731	I	galE
MPN258	0.088	N	0.142	0.786747	0.761	0.547601	0.798	0.675564	0.221	0.299161	G	mglA
MPN259	0.951	N							0.152	0.631	G	rbsC1
MPN260	0.971	N	0.115	0.813531	1.838	0.052971	1.263	0.243606			G	rbsC
MPN261	0.590	N	0.468	0.373516	0.502	0.704445	1.025	0.457786	0.152	0.633089	L	topA
MPN262	0.384	N									M	
MPN263	0.093	N	0.643	0.177917	4.066	0.000000	2.716	0.000009	0.282	0.097919	O	trx
MPN264	0.179	N			0.331	0.792	0.862	0.617	0.077	0.899	R	
MPN265	0.145	N	0.414	0.445116			0.909	0.571669	0.456	0.000249	J	trpS
MPN266	0.803	N	0.535	0.289621			1.238	0.262957	0.259	0.157146	K	spxA
MPN267	0.165	N	1.471	0.000082	1.235	0.259856	1.113	0.372985	0.189	0.450465	H	ppnK
MPN268	0.857	N									G	ptsG
MPN269	0.335	N									R	ymdA
MPN270	0.973	N									M	
MPN271	0.927	Y									M	
MPN272	0.963	N									S	
MPN273	0.172	N	1.060	0.009189	0.814	0.513274	0.909	0.571975	0.313	0.045805	F	hit1
MPN274	0.973	N									P	
MPN275	0.045	N									L	yaaK
MPN276	0.960	N									S	
MPN277	0.103	N	0.278	0.629294			0.596	0.830045	0.222	0.293589	J	lysS
MPN278	0.278	N					2.207	0.001			M	yefE
MPN279	0.101	N	0.435	0.416957	0.820	0.509447	1.156	0.333697	0.240	0.218917	J	lepA
MPN279a	0.737	N									--	New ORF

MPN280	0.082	N	1.084	0.007355	1.332	0.211572	1.770	0.025634	0.205	0.369017	J	rnjA
MPN280a	0.018	N									#N/A	#N/A
MPN281	0.901	Y									M	
MPN282	0.920	N									M	
MPN283	0.912	N									N	
MPN284	0.887	Y	0.364	0.513364	0.943	0.430837	1.023	0.459815	0.315	0.043173	M	
MPN285	0.207	N									V	prrB
MPN286	0.926	N									M	
MPN287	0.295	N									N	
MPN288	0.906	Y	0.740	0.104268	1.655	0.093285	0.891	0.588427			M	
MPN289	0.839	N									V	hsdS1B
MPN290	0.121	N									V	
MPN291	0.079	N			0.202	0.846	0.686	0.768	0.122	0.760	O	
MPN292	0.256	N									J	rluD
MPN293	0.962	N									U	lsp
MPN294	0.058	N	0.433	0.419141	1.007	0.391105	0.978	0.503608	0.321	0.036636	O	pfpI
MPN295	0.818	N	0.398	0.467127	0.437	0.739382	0.888	0.591956	0.175	0.519574	K	
MPN296	0.280	N									J	rpsU
MPN297	0.737	N			0.219	0.840	0.660	0.787	0.194	0.424	D	gpsB
MPN298	0.341	N									I	acpS
MPN299	0.314	N	0.084	0.840624	0.292	0.809285	1.852	0.015633	0.040	0.958470	I	plsC
MPN300	0.296	N	0.150	0.778841							D	dyr
MPN301	0.084	N	0.245	0.671578	0.543	0.681095	0.695	0.761359	0.229	0.265168	D	scpB
MPN302	0.143	N	2.369	0.000000	4.134	0.000000	2.687	0.000012	0.247	0.195000	G	pfk
MPN303	0.197	N	0.267	0.642869	0.219	0.839621	0.736	0.728270	0.133	0.718511	G	pyk
MPN304	0.165	N									C	arcA
MPN305	0.048	N									C	arcA
MPN306	0.060	N									C	argI
MPN307	0.347	N			1.920	0.040					C	arcC
MPN308	0.963	N									E	
MPN309	0.950	N					0.491	0.887			S	p65

MPN310	0.335	N	0.146	0.782911	0.583	0.657609	0.530	0.868158	0.103	0.828530	M	hmw2
MPN311	0.529	N	0.171	0.756940	0.055	0.895720	0.418	0.918556	0.137	0.699415	D	
MPN312	0.692	N			0.270	0.819					S	
MPN313	0.958	N									S	
MPN314	0.080	N	1.545	0.000029	1.205	0.275302	1.822	0.018779	0.227	0.271757	D	mraZ
MPN315	0.108	N	0.168	0.760684	0.469	0.722344	0.797	0.676234	0.310	0.049347	D	mraW
MPN316	0.295	N	0.132	0.796581							D	ftsA
MPN316a	0.593	N									--	New ORF
MPN317	0.107	N	0.113	0.814684	0.486	0.713243	1.165	0.325324	0.199	0.397028	D	ftsZ
MPN318	0.974	N									E	
MPN319	0.971	N									E	
MPN320	0.459	N	0.495	0.338120	0.693	0.590257	1.343	0.185748	0.219	0.306526	F	thyA
MPN321	0.240	N	0.176	0.752243	1.864	0.048607	0.659	0.787970	0.265	0.139031	F	dhfr
MPN322	0.382	N	0.376	0.496681	1.109	0.329331	1.036	0.446447	0.306	0.055052	F	nrdF
MPN323	0.289	N	0.297	0.604374	0.694	0.589788	0.569	0.846615	0.272	0.119659	F	nrdI
MPN324	0.322	N	0.370	0.505963	1.527	0.132547	0.985	0.497342	0.355	0.013392	F	nrdE
MPN325	0.102	N									J	rplU
MPN326	0.097	N			0.884	0.469			0.126	0.745	J	ysxB
MPN327	0.519	N							0.049	0.948	J	rpl27
MPN328	0.077	N	0.510	0.320511	2.245	0.011537	1.661	0.046471	0.141	0.679668	L	nfo
MPN329	0.144	N									K	fur
MPN330	0.121	N									S	
MPN331	0.236	N	0.487	0.348793	0.780	0.535066	1.072	0.411389	0.246	0.199794	O	tig
MPN332	0.132	N	0.092	0.833514	0.411	0.752588	0.989	0.492777	0.148	0.648708	O	lon
MPN333	0.926	N									R	
MPN334	0.329	N									R	bcrA
MPN335	0.950	N									R	
MPN336	0.277	N	0.090	0.835172					0.103	0.829451	H	nadD
MPN337	0.362	N	0.121	0.807329			1.063	0.420096	0.062	0.928904	A	
MPN338	0.773	N	0.100	0.826639	0.264	0.821423	0.919	0.561590	0.028	0.970678	A	
MPN339	0.956	N									A	

MPN340	0.285	N									L	
MPN341	0.183	N									L	
MPN342	0.428	N			0.705	0.583			0.206	0.367	V	hsdM
MPN343	0.259	N									V	
MPN344	0.771	N							1.621	0.000	N	
MPN345	0.455	N									V	hsdR
MPN346	0.545	N									V	
MPN347	0.197	N									V	hsdR
MPN347a	0.924	N									--	New ORF
MPN348	0.128	N									H	mthfs
MPN349	0.091	N	0.259	0.653591	0.357	0.779565	0.627	0.809994	0.119	0.774494	R	ymdB
MPN350	0.969	N									I	plsY
MPN351	0.103	N									V	trmK
MPN352	0.211	N	0.110	0.817395	0.176	0.855779	0.686	0.767851	0.091	0.864957	K	sigA
MPN353	0.184	N	0.234	0.684728					0.345	0.018672	L	dnaG
MPN354	0.130	N	0.234	0.685030	0.565	0.668173			0.265	0.140421	J	glyS
MPN355	0.143	N					2.226	0.001	0.156	0.610	J	yacO
MPN356	0.297	N	0.706	0.127061	3.416	0.000016	1.831	0.017867	0.159	0.595531	J	cysS
MPN357	0.182	N	0.578	0.241518	0.930	0.439430	1.583	0.068407	0.175	0.515712	L	lig
MPN358	0.642	N									A	
MPN359	0.902	N									A	
MPN359a	0.107	N									--	New ORF
MPN360	0.854	N							0.067	0.918	J	rpmE
MPN361	0.092	N	0.748	0.099476	0.533	0.686546	0.696	0.759939	0.163	0.579213	J	prfA
MPN362	0.446	N	0.122	0.806569					0.085	0.879591	J	hemK
MPN363	0.963	Y									M	
MPN364	0.905	N									M	
MPN365	0.149	N									V	
MPN366	0.859	N									M	
MPN367	0.883	N									M	
MPN368	0.837	N			2.894	0.000					N	

MPN369	0.928	Y									M	
MPN370	0.939	N									M	
MPN371	0.554	N									S	
MPN372	0.696	N	3.276	0.000000	2.533	0.003112	3.750	0.000000	0.221	0.295862	V	ptxA
MPN373	0.902	N									M	
MPN374	0.856	N									M	
MPN375	0.894	N									M	
MPN376	0.914	Y	2.848	0.000000	2.670	0.001549	1.223	0.275521	0.761	0.000000	A	
MPN376a	0.049	N									--	New ORF
MPN377	0.672	N									A	
MPN378	0.195	N	0.714	0.121080	2.438	0.004885	2.181	0.001499	0.132	0.719630	L	dnaE
MPN379	0.324	N	0.131	0.797516			1.048	0.434922	0.199	0.396763	L	polA
MPN380	0.205	N	0.313	0.583223	0.344	0.785707	1.075	0.408503	0.120	0.770488	L	fpg
MPN381	0.716	N	0.443	0.406709	0.982	0.406407	0.778	0.692705	0.179	0.496358	R	yidA
MPN382	0.112	N									H	coaE
MPN383	0.141	N	0.855	0.049064	1.546	0.125810	1.583	0.068404	0.288	0.084803	R	yidA
MPN384	0.268	N	1.311	0.000641	1.484	0.147717	4.280	0.000000	0.203	0.381312	J	leuS
MPN385	0.933	N									A	
MPN386	0.126	N	0.480	0.357867	1.295	0.229360	0.855	0.623450	0.360	0.011516	F	yaaF
MPN387	0.507	N	0.302	0.597037	0.205	0.845147	0.603	0.825713	0.164	0.573682	S	
MPN388	0.847	N									S	
MPN389	0.193	N	2.241	0.000000	2.488	0.003862	4.015	0.000000	0.357	0.012757	O	lplA
MPN390	0.065	N	2.203	0.000000	4.936	0.000000	3.849	0.000000	0.234	0.241957	C	pdhD
MPN391	0.110	N	1.995	0.000000	1.980	0.032478	2.269	0.000725	0.328	0.030831	C	pdhC
MPN392	0.055	N	1.418	0.000167	1.257	0.248110	1.602	0.062529	0.205	0.372381	C	pdhB
MPN393	0.199	N	1.143	0.004101	1.225	0.264740	1.254	0.249977	0.243	0.208653	C	pdhA
MPN394	0.073	N	1.777	0.000001	2.050	0.025080	2.849	0.000002	0.190	0.444967	C	nox
MPN394a	0.079	N									#N/A	#N/A
MPN395	0.047	N	0.181	0.746710	0.848	0.491529	0.749	0.717927	0.308	0.052108	F	apt
MPN396	0.935	N	0.139	0.790287	0.595	0.650564			0.129	0.731321	U	secD
MPN397	0.160	N	0.185	0.741412	0.295	0.807882	1.093	0.391182	0.146	0.660159	T	spoT

MPN398	0.943	Y	2.006	0.000000	3.864	0.000001	10.157	0.000000	0.565	0.000001	A	
MPN399	0.946	N									A	
MPN400	0.895	N	77.134	0.000000	43.785	0.000000	22.047	0.000000	9.311	0.000000	A	
MPN401	0.195	N	0.282	0.623457	0.366	0.774913	0.706	0.752126	0.172	0.532183	K	greA
MPN402	0.349	N	0.388	0.480044	0.465	0.724439	1.064	0.419655	0.156	0.612323	J	proS
MPN403	0.806	N									S	
MPN404	0.550	N									M	
MPN405	0.932	Y									A	
MPN406	0.042	N									I	acpA
MPN407	0.888	N	0.984	0.017961			0.807	0.667286	1.212	0.000000	I	
MPN407a	0.075	N									--	New ORF
MPN408	0.868	Y	0.367	0.510050	0.611	0.641184	0.950	0.531233	0.160	0.591948	M	
MPN408a	0.189	N									--	New ORF
MPN409	0.917	N									M	
MPN410	0.823	N									N	
MPN410a	0.054	N									--	New ORF
MPN411	0.899	Y									M	
MPN412	0.928	N									M	
MPN413	0.877	N									S	
MPN414	0.900	N									M	
MPN415	0.872	N			0.764	0.545					P	hatA
MPN416	0.120	N									P	hatB
MPN417	0.967	N									P	hatC
MPN418	0.073	N									J	alaS
MPN419	0.171	N	0.661	0.161913	1.060	0.358346	1.374	0.166229	0.108	0.811600	L	ruvX
MPN420	0.154	N					0.766	0.703			I	glpQ
MPN421	0.976	N									U	cglT
MPN422	0.420	N	0.359	0.520247			1.042	0.441261	0.116	0.782847	J	mnmA
MPN423	0.045	N									S	
MPN424	0.629	N									K	ylxM
MPN425	0.208	N	0.404	0.459046	0.783	0.533577	1.259	0.246800	0.247	0.197020	U	ftsY

MPN426	0.184	N	0.264	0.647148	0.283	0.813336	0.788	0.684075	0.158	0.604012	D	smc
MPN427	0.478	N	0.727	0.112834	2.405	0.005693	1.429	0.134223	0.237	0.230878	R	yidA
MPN428	0.091	N	0.477	0.360964	0.701	0.585287	0.580	0.839713	0.136	0.704527	G	pta
MPN429	0.084	N	1.691	0.000003	1.926	0.039391	2.937	0.000001	0.215	0.322694	G	pgk
MPN429a	0.624	N									--	New ORF
MPN430	0.054	N	0.382	0.488888	0.683	0.596718	0.736	0.728633	0.095	0.851797	I	gap
MPN431	0.960	N									P	cbiQ
MPN432	0.314	N									P	cbiO 1
MPN433	0.110	N									P	cbiO 2
MPN434	0.149	N	0.325	0.566610	0.454	0.730570	0.475	0.895205	0.152	0.632242	O	dnaK
MPN435	0.954	N									A	
MPN436	0.876	Y	0.178	0.749050	14.648	0.000000	1.021	0.460963	0.351	0.015228	M	
MPN437	0.850	N									M	
MPN438	0.900	N									S	
MPN439	0.842	Y									M	
MPN440	0.847	N									M	
MPN441	0.367	Y									S	
MPN442	0.962	Y									M	
MPN443	0.356	N			1.707	0.080					J	
MPN444	0.897	Y	1.806	0.000000	18.910	0.000000	1.889	0.012429	1.608	0.000000	M	
MPN445	0.701	N	0.093	0.832446			0.694	0.762022	0.144	0.666139	I	lip3
MPN446	0.122	N	0.093	0.832860	0.431	0.742538	1.489	0.104954	0.114	0.791812	J	rpsD
MPN447	0.829	N	0.398	0.466317	0.215	0.841055	0.472	0.896520	0.223	0.288106	M	HWM1
MPN448	0.967	N									H	fbp
MPN449	0.902	N							0.146	0.661	A	
MPN450	0.068	N									L	holA
MPN451	0.966	N									L	comE3
MPN452	0.928	N	0.232	0.687603	0.207	0.844206	0.473	0.896029	0.123	0.756836	M	hmw3
MPN453	0.928	Y									M	P3
MPN454	0.821	Y	2.957	0.000000	8.262	0.000000	11.288	0.000000			S	
MPN455	0.970	N									I	ctaD

MPN456	0.891	Y	0.357	0.523606	0.565	0.668138	0.968	0.513885	0.244	0.205222	E	
MPN457	0.890	N									S	
MPN458	0.814	N									S	
MPN459	0.898	N	9.987	0.000000					0.807	0.000000	E	
MPN460	0.917	N									P	ktrB
MPN461	0.071	N	0.113	0.815138			1.106	0.379024			P	ktrA
MPN462	0.917	N									M	
MPN463	0.908	N									S	
MPN464	0.896	N									M	
MPN464a	0.796	N									--	New ORF
MPN465	0.142	N									M	
MPN465a	0.019	N									--	New ORF
MPN466	0.809	N									M	
MPN467	0.854	Y									M	
MPN468	0.939	N									M	
MPN469	0.947	N									A	
MPN470	0.110	N	0.289	0.614986	0.640	0.623286	0.547	0.858740	0.139	0.690586	E	pepX
MPN471	0.537	N									J	rpmG
MPN472	0.078	N	0.974	0.019635	1.109	0.329162	3.488	0.000000	0.234	0.244391	I	degV
MPN473	0.288	N									I	lip2
MPN474	0.554	N	0.178	0.749451	0.169	0.858512	0.464	0.900093	0.159	0.595733	N	
MPN474a	0.103	N	0.103	0.824269	0.192	0.850034	0.391	0.928231	0.089	0.869359	--	New ORF
MPN475	0.291	N									T	engA
MPN476	0.082	N					0.366	0.936	0.543	0.000	F	cmk
MPN477	0.936	N									A	
MPN478	0.246	N	0.422	0.434205					0.249	0.189743	F	YebC
MPN479	0.482	N	1.336	0.000475	2.317	0.008470	1.475	0.111352	0.209	0.351003	I	acpH
MPN480	0.659	N	0.462	0.380838	0.192	0.849954	0.723	0.739068	0.058	0.934604	J	valS
MPN480a	0.018	N									--	New ORF
MPN481	0.127	N	0.554	0.267697			1.637	0.052593	0.143	0.673734	T	engB
MPN482	0.940	N									A	

MPN483	0.229	N									I	yibD
MPN484	0.639	N									N	
MPN485	0.874	N									S	
MPN486	0.159	N									S	
MPN487	0.407	N	1.038	0.011210							O	nifS
MPN488	0.062	N									O	nifU
MPN488a	0.179	N									--	New ORF
MPN489	0.892	N	7.862	0.000000	9.315	0.000000	10.759	0.000000	1.758	0.000000	M	
MPN489a	0.372	N									#N/A	#N/A
MPN490	0.085	N									L	recA
MPN491	0.934	N	3.164	0.000000	8.022	0.000000	1.512	0.094810	1.141	0.000000	L	mnuA
MPN492	0.116	N									G	ulaE
MPN493	0.068	N	0.751	0.097275	1.197	0.279790			0.375	0.006970	G	ulaD
MPN494	0.077	N					1.337	0.190			G	ulaC
MPN495	0.042	N									G	ulaB
MPN496	0.959	N									G	ulaA
MPN497	0.034	N									G	ulaG
MPN498	0.777	N									G	araD
MPN499	0.126	N									S	
MPN500	0.919	N									M	
MPN501	0.705	N									N	
MPN502	0.925	N									M	
MPN503	0.910	N									M	
MPN504	0.667	N							1.621	0.000	N	
MPN505	0.928	Y			4.285	0.000					M	
MPN506	0.911	Y	6.314	0.000000			8.656	0.000000	5.262	0.000000	M	
MPN506a	0.968	N									#N/A	#N/A
MPN507	0.605	N									V	
MPN508	0.659	N									U	
MPN509	0.564	Y			25.218	0.000	17.334	0.000			S	
MPN510	0.251	N									S	

MPN511	0.689	N									S	
MPN512	0.864	Y									M	
MPN513	0.123	N									V	
MPN514	0.486	N									S	
MPN515	0.100	N	0.194	0.731720	0.466	0.724102	1.031	0.451301	0.126	0.744427	K	rpoC
MPN516	0.208	N	0.231	0.687892	0.595	0.650656	0.934	0.547125	0.162	0.582573	K	rpoB
MPN517	0.102	N	0.215	0.707176	0.514	0.697729	0.872	0.606828	0.115	0.787077	H	yhdA
MPN518	0.176	N	0.696	0.134062	2.138	0.017904	1.091	0.393121	0.166	0.564737	T	
MPN519	0.442	N	0.037	0.876744					0.078	0.897464	I	lip3
MPN520	0.500	N	0.277	0.629851	0.318	0.797836	1.118	0.367627	0.124	0.754423	J	ileS
MPN520a	0.057	N									--	New ORF
MPN521	0.455	N									J	ygl3
MPN522	0.142	N									J	trmB
MPN523	0.880	N	0.390	0.478354	0.972	0.412747	1.114	0.371788	0.757	0.000000	M	
MPN524	0.411	N									N	
MPN525	0.203	N									L	dnaB
MPN526	0.175	N					0.917	0.564			K	
MPN527	0.963	N									A	
MPN528	0.170	N	0.286	0.618146	0.481	0.715640	0.610	0.821375	0.173	0.527139	C	ppa
MPN529	0.740	N	0.303	0.595864			0.656	0.790158			D	ihf
MPN530	0.112	N	0.293	0.609078	1.619	0.103323	0.708	0.750778	0.310	0.049741	S	
MPN531	0.071	N	0.286	0.618202	0.488	0.712099	0.636	0.804034	0.330	0.029108	O	clpB
MPN532	0.150	N	0.102	0.824541			0.639	0.801995	0.097	0.846860	I	licA
MPN533	0.112	N	1.107	0.005893	1.426	0.170396	1.250	0.253538	0.199	0.396809	C	ackA
MPN534	0.117	N									S	
MPN535	0.116	N									L	ruvA
MPN536	0.114	N									L	ruvB
MPN537	0.237	N									L	mucB
MPN538	0.230	N	0.185	0.741354	0.567	0.667006	0.723	0.739399	0.153	0.627117	J	rplJ
MPN539	0.027	N	0.218	0.703312	0.267	0.820138	0.620	0.814460	0.169	0.545580	J	rplL
MPN540	0.607	N									J	rpmF

MPN541	0.925	N									J	rpsT
MPN542	0.890	N			5.355	0.000					A	
MPN543	0.246	N									J	fmt
MPN544	0.192	N									S	
MPN545	0.817	N	0.271	0.638648	0.466	0.724046	0.495	0.885736	0.211	0.344433	K	rnc
MPN546	0.103	N	0.224	0.696913	0.274	0.816978	0.852	0.626352	0.058	0.934790	I	plsX
MPN547	0.160	N	0.159	0.770017	0.531	0.687771	0.870	0.609185	0.062	0.928455	C	dak
MPN548	0.204	N									J	truC
MPN549	0.129	N	0.278	0.628738	0.359	0.778281	0.634	0.805009	0.168	0.551298	L	recJ
MPN550	0.065	N									H	thiI
MPN551	0.495	N									L	yqaJ
MPN552	0.256	N	0.273	0.634979					0.281	0.100015	S	
MPN553	0.143	N	0.177	0.750745	0.256	0.824483	0.651	0.793786	0.072	0.910174	J	
MPN554	0.060	N									D	ywpH
MPN555	0.044	N	0.670	0.154733	1.369	0.194796	0.676	0.775664	0.399	0.002899	O	
MPN556	0.156	N	0.795	0.074151	1.292	0.230565	1.693	0.039262	0.182	0.484705	J	argS
MPN557	0.246	N			1.939	0.038					D	gidA
MPN558	0.160	N	0.358	0.521362			0.935	0.546359			D	gidB
MPN559	0.225	Y									S	
MPN560	0.081	N	0.328	0.563163	1.138	0.312772	0.758	0.710294	0.161	0.585941	C	arcA
MPN561	0.066	N									F	udk
MPN562	0.159	N			11.000	0.000	1.482	0.108			H	outB
MPN563	0.164	N							0.110	0.805	J	obg
MPN564	0.053	N									C	adh
MPN565	0.944	N									A	
MPN566	0.111	N	0.400	0.464145							I	glpQ
MPN567	0.882	N	2.434	0.000000	0.467	0.723645	2.473	0.000113	0.152	0.629511	M	p2
MPN568	0.210	N									D	spg
MPN569	0.072	N									O	ybeY
MPN570	0.879	N									A	
MPN571	0.833	N									U	

MPN572	0.284	N	2.062	0.000000	4.923	0.000000	5.085	0.000000	0.202	0.383679	O	pepA
MPN573	0.063	N	2.511	0.000000	2.234	0.012073	3.497	0.000000	0.409	0.001925	O	groEL
MPN574	0.839	N	2.162	0.000000	1.936	0.038002	2.286	0.000625	0.270	0.124774	O	groES
MPN575	0.929	N									A	
MPN576	0.322	N	2.378	0.000000	0.921	0.444769	1.565	0.074471	0.185	0.469279	E	glyA
MPN577	0.768	N									S	
MPN578	0.109	N									S	
MPN579	0.953	N									A	
MPN580	0.841	N									O	
MPN581	0.888	N									O	
MPN582	0.847	Y									O	
MPN583	0.925	N									S	
MPN584	0.697	N									M	
MPN585	0.843	Y									M	
MPN586	0.882	N									O	
MPN587	0.378	Y									M	
MPN588	0.913	Y	8.130	0.000000	9.520	0.000000	2.911	0.000001			M	
MPN589	0.905	N									S	
MPN590	0.919	Y									M	
MPN591	0.907	N	0.693	0.136618	1.290	0.231512	0.889	0.590903			M	
MPN592	0.911	Y	35.123	0.000000	32.685	0.000000	24.927	0.000000	3.198	0.000000	O	
MPN592a	0.017	N									--	New ORF
MPN593	0.949	N									M	
MPN594	0.897	N									M	
MPN595	0.065	N									G	lacA
MPN596	0.240	N									D	erzA
MPN597	0.064	N					0.726	0.736			C	atpC
MPN598	0.071	N	0.399	0.465214	0.200	0.847054	0.563	0.849773	0.174	0.523392	C	atpD
MPN599	0.188	N									C	atpG
MPN600	0.091	N	0.170	0.758145	0.244	0.829477	0.669	0.780620	0.094	0.854723	C	atpA
MPN601	0.102	N									C	atpH

MPN602	0.839	N			0.399	0.759	2.007	0.006	0.168	0.554	C	atpF
MPN603	0.941	N									C	atpE
MPN604	0.966	N					1.175	0.317			C	atpB
MPN605	0.972	N									A	mpn65
MPN606	0.125	N	1.076	0.007960	2.359	0.007013	1.739	0.030485	0.219	0.307702	G	eno
MPN607	0.810	N	0.767	0.088596	0.700	0.586346	2.062	0.003769	0.177	0.507138	O	pmsR
MPN608	0.115	N									P	phoU
MPN609	0.153	N					0.643	0.799			P	pstB
MPN610	0.950	N									P	pstA
MPN611	0.886	N	0.368	0.507409	0.848	0.491788	1.692	0.039431	0.226	0.274662	P	pstS
MPN612	0.723	N									S	
MPN613	0.565	N									S	
MPN614	0.190	N									S	
MPN615	0.138	N									V	hsdS
MPN616	0.776	N	0.035	0.878120	0.251	0.826650	0.897	0.582702	0.072	0.910085	J	rpsI
MPN617	0.793	N									J	rplM
MPN618	0.453	N	0.267	0.642866	0.505	0.702794	0.629	0.808965	0.175	0.519552	L	dnaX
MPN619	0.169	N	0.891	0.037902	1.014	0.386709	1.337	0.189770	0.162	0.584556	L	uvrA
MPN620	0.220	N	0.061	0.858464			0.751	0.715991			S	
MPN621	0.094	N	1.042	0.010843	1.167	0.296329	1.911	0.010787	0.195	0.416404	J	rnj
MPN622	0.885	N									J	rpsO
MPN623	0.096	N									K	deaD
MPN624	0.778	N									J	rpmB
MPN625	0.088	N	2.018	0.000000	35.463	0.000000	1.679	0.042208	0.223	0.287057	O	osmC
MPN626	0.391	N									K	
MPN627	0.076	N	0.498	0.335275	0.518	0.695023	0.760	0.708530	0.250	0.186921	G	ptsI
MPN628	0.132	N	0.598	0.220725	0.424	0.746055	0.825	0.650757	0.176	0.515383	G	pgm
MPN629	0.165	N	1.156	0.003605	1.376	0.191668	2.408	0.000209	0.144	0.666525	G	tim
MPN630	0.943	N									M	yfiB
MPN631	0.070	N	0.617	0.202099	2.159	0.016466	1.171	0.319596	0.558	0.000001	J	tsf
MPN632	0.082	Y	0.973	0.019756	1.473	0.151915	1.160	0.330058	0.239	0.226119	F	pyrH

MPN633	0.239	N									O	
MPN634	0.658	N									T	ywaC
MPN634a	0.972	N									--	New ORF
MPN635	0.190	N									S	
MPN636	0.141	N					1.014	0.469			J	frr
MPN637	0.969	Y									I	cdsA
MPN638	0.207	N	0.202	0.722451	0.373	0.771943	0.586	0.836544	0.152	0.630670	V	hsdS
MPN639	0.872	N									M	
MPN640	0.908	Y									M	
MPN641	0.845	Y									M	
MPN642	0.841	Y	5.010	0.000000	4.490	0.000000	12.181	0.000000	0.564	0.000001	M	
MPN643	0.731	N			4.812	0.000	4.790	0.000	0.318	0.041	M	
MPN644	0.816	Y	26.513	0.000000	16.679	0.000000					M	
MPN645	0.818	Y	51.404	0.000000	43.889	0.000000					M	
MPN646	0.807	Y			7.884	0.000					M	
MPN647	0.935	Y	12.347	0.000000	26.750	0.000000	25.225	0.000000	1.170	0.000000	M	
MPN648	0.416	N									S	
MPN649	0.314	N									M	
MPN650	0.807	Y									M	
MPN651	0.963	N									G	mtlA
MPN652	0.201	N									G	mtlD
MPN653	0.103	N									G	mtlF
MPN654	0.929	Y									M	
MPN655	0.900	N									N	
MPN655a	0.106	N									--	New ORF
MPN656	0.186	N									J	rbgA
MPN657	0.944	N									A	
MPN658	0.572	N							0.103	0.829	J	rplS
MPN659	0.228	N									J	trmD
MPN660	0.058	N	0.113	0.815360	0.533	0.686949	1.133	0.354607	0.096	0.849067	J	rpsP
MPN661	0.966	N									V	

MPN662	0.906	N									E	pilB
MPN663	0.152	N	0.101	0.825576			0.806	0.668556	0.080	0.890985	J	ytpR
MPN664	0.148	N	1.623	0.000009	2.575	0.002522	3.959	0.000000	0.238	0.228457	I	degV
MPN665	0.067	N	0.481	0.356024	0.676	0.600902	0.697	0.759399	0.230	0.260956	J	tuf
MPN666	0.974	N									R	
MPN667	0.157	N	0.185	0.741581	0.188	0.851312	0.584	0.837319	0.140	0.684361	M	gtaB
MPN668	0.061	N	2.550	0.000000	2.845	0.000599	1.767	0.026073	0.502	0.000024	O	osmC
MPN669	0.062	N	0.119	0.809410	0.318	0.797501	0.686	0.768258	0.107	0.815023	J	tyrS
MPN670	0.902	N	0.112	0.816135	0.562	0.670318	0.854	0.623873	0.098	0.843022	S	
MPN671	0.322	N	0.176	0.751218	0.346	0.784596	0.917	0.564131	0.125	0.749545	O	ftsH
MPN671a	0.068	N									--	New ORF
MPN672	0.066	N									F	hpt
MPN672a	0.904	N									--	New ORF
MPN673	0.096	N	0.891	0.037709	6.292	0.000000	1.232	0.267916	0.169	0.549641	I	ygbB
MPN674	0.187	N	0.359	0.520451	0.587	0.655178	0.518	0.874454	0.365	0.009551	C	ldh
MPN675	0.078	N									N	
MPN676	0.434	Y									S	
MPN677	0.174	N									R	ywfO
MPN678	0.134	N	1.182	0.002748	1.321	0.216766	2.689	0.000012	0.129	0.731874	J	gltX
MPN679	0.229	N									J	ksgA
MPN680	0.964	N									U	yidC
MPN681	0.198	N									J	rnpA
MPN682	0.607	N									J	rpmH
MPN683	0.587	N	0.167	0.761248					0.081	0.889072	P	devA
MPN684	0.928	N	0.493	0.340958	2.013	0.028875	0.910	0.570107	0.317	0.041501	P	mpn684
MPN685	0.255	N	0.063	0.857074							P	cysA
MPN686	0.107	N	0.186	0.740652	0.346	0.784894	0.691	0.764263	0.287	0.087771	L	dnaA
MPN687	0.921	N									S	
MPN688	0.154	N	0.223	0.698484	0.444	0.735487	0.562	0.850697	0.242	0.213837	D	soj

Sequences

Sequences “Development of self-replicating plasmids for M. pneumoniae”

>Ori_1 *M. pneumoniae* M129, complete genome
TATCAGGGTTGCTATTGAGGGTTTCACGGGCAAACTCATTGTCCACAATCACAAAAAGGACATTATTTTTTAAAGCAA
ACCGTTTAAGATTCTTAATGTACTTGTCATAAAAACCTAAAGTAGTTTCGTATTGCTTTTTTAAAAGAAGTTTAAAAG
CACTAAATTGTTCCATTACTCTTCTATTACCTGATCAAGAATCATCGTTAAGATCCGCGAAATCGGAATGTTAACCGC
TTTAGCATAACGCTTAACCTTATTCCAAACACGGGTACTCGTATAAAAGTGCACCCCGGTGGTTAAGCGTTCTTGCCG
TTCTAAGGGAATAATACCGGGCACTTTCGCCTTTGGTGGTGGTAGATTTTCTTCTTCCTGCAATTTATTAGATGGGCGA
GTAGTCTGCTTAACTTGTGGCTGTTCAACCGCTGGTTGTGGTTCTGGTTGGGGAACAGGTTCCACAATTGGTTGCGTA
GCCTCCACCTGAATATTAGATGTAGCTACTGGTTGTTGATTAAAGTTAGTGGTTGTAGGTTGTTGAAACTCATTTAAG
TGCACTTCTGCAGTTTGGTGAACTGGTTGGGTTGGAACCTGATTATTAAAGTTAGTTGGTTCTGATTGCGGTTGCACAT
TAACTGGTTGGGTTGAAGTTTGTGGTTCTTGTTTGGGAAACTGGTTAAAGGATTTCACAAACTCATCGCCAAACTTGA
GTAATGGTCCAAAATCAAACTGGGGCATGTCCTTAGAAAAGAACTTTCTGTCTTGACTAGATTGATTTGCTTTAGTTT
GGTTAGTTAGATCATGTTGGGGTTGGACAACCGGTTCTAGTTTTAATTGTTGTTCTACCTTTTTAACTTCTTGAACAGT
TTCTGCATCATCTTCATCCTGAGCTAGGTCAAAGAAGAAGCGTTTCTTTTTACTGATGGCCATGGTTGACTAGGTTAA
GAATTTCTTTAGTAATCTCTAAATATTCAGTTTGGTACTTCTTGTTGGATGTTGGTGATACCAAAGAGATAGGTAACTT
TTCATAACCAACAGCAGCAGATGACTTAGATGTTAAAGATACAAAGTGTTCTGAAAAAGCAACGTTGTTCTTGTGCG
CTTTTGATTTCACCAACTCAATCACATCATTGTGTAATCTGGTGCGCATATTGACCTTGGTTGGTACCAAAATGGTTTT
TAAGTTGGGGTTCTTTTCCTTAAAAGTATCAATCGTTTCGACAATCCGCATTAAACCCAACATAGAGTATTGATCAGG
TTCAAACGGGATCACAATCACATCAGAGAGGCTCATGGCTGTGGAGACCAAGGTAGCCATGTTAGGCGGTGTGTCTA
GTAAAACAAAGTCATACCTGCGGGTGAGTTGTTTAACTATTTCCGCAATATCAGAAGCCTTGTACTTTTTGCGAGCAA
TATCAATGTCCGCAAAATTTAACTCAAAGTTACAAGGCAAAATGTCTAGTCCTTCATAAACGGGCAACAAACAGTCA
TCAATTTCAATTGAACTGTTGGCGCCGTTAAACTTAGGCACTTTTAACAAAATATCAATTAAGGTGTTATTTAATCTTT
CTGGATTTTGTCCAAAAGAGGCTGAAACGTTGCCCTGACCATCAAGGTCTAAAATAACTTTCCGTTGTTCAGGGCACA
ACTTTACAAGCGAACCAGCAACGTTAGTGGCCATGGTTGTTTTTAAAACGCCACCTTTGTTATTAACAAAAGAAATAA
TCATATGCGCGCGTATAAGTTTGCTTAAATTCTAATATTTATTTAGTATTATTCCGGTATTATTTACCGACGAAATTAA
TACCATCAGGGTATTAAGATGCTACCAACGTGGTATTAAAATGTGCCCAACCGCGAAAAAGAAAGTGGTATATAGGA
AAATGGAAAACAAATTTGTTTGCCGTTTACTATTAAATACTAATCTTCTATATAGTATAGAGAAACTTTTTCTTTAACA
TAATATTATCTTAATATTATTTACCTACTAATAGCTTAATATTATTAGTATTTATTTAGTATTATGCTAATACTATGCAG
ATATTATCTTAATATTATCTATAGTATTAGGCTAATATTATTCTTAATATTTATTAAGGTACTAAAGCATTACCTATAG
GTGATATTATGACAATACTAAAGTGGTTAGTATTATTAGGGTATTATTCAAAGTATTCTCCAACACTATTCCCTTAGCT
CTAACTTAATATCCCTTTAATACCAAGATGGTTAAACAATGAAGAAAATTAGTTGTTTGTTCTGGCCCTTTACTATATA
TGTACATAGTACATATATAAGAAAGGCTACCATGACATTATATATAATAGGCATCCAGAAATAAAAGCCAAATTAAT
CTATAATTAAGTATTTATACTTAATTATAGTTATTTTATTAACAAAATTAGTGACATTTGGCCATCCATAATAAATAAA
ATCAAATAAACAAAACAAAAAGCAAAAAAACTATAAAAAGGGCATGAAAGTTTTGATTAATAAGAATGAGTTGAAC
AAAATCCTCAAAAAACTCAACAATGTAATCGTATCTAACAATAAGATGAAACCATACCACTCTTATTTATTAATAGA
GGCTACAGAAAAGGAAATTAACTTCTATGCTAACAACGAGTACTTTTCTGCTAAATGTACCTTAGCCGAAAACATTG
ATGTACTTGAAGAAGGTGAAGTAATTGTTAAAGGCAA
>Ori_3 *M. pneumoniae* M129, complete genome
GGTGTTATTTAATCTTTCTGGATTTTGTCCAAAAGAGGCTGAAACGTTGCCCTGACCATCAAGGTCTAAAATAACTTT
CCGTTGTTCAGGGCACAACTTTACAAGCGAACCAGCAACGTTAGTGGCCATGGTTGTTTTTAAAACGCCACCTTTGTT
ATTAACAAAAGAAATAATCATATGCGCGCGTATAAGTTTGCTTAAATTCTAATATTTATTTAGTATTATTCCGGTATTA
TTTACCGACGAAATTAATACCATCAGGGTATTAAGATGCTACCAACGTGGTATTAAAATGTGCCCAACCGCGAAAAA
GAAAGTGGTATATAGGAAAATGGAAAACAAATTTGTTTGCCGTTTACTATTAAATACTAATCTTCTATATAGTATAGA
GAAACTTTTTCTTTAACATAATATTATCTTAATATTATTTACCTACTAATAGCTTAATATTATTAGTATTTATTTAGTAT
TATGCTAATACTATGCAGATATTATCTTAATATTATCTATAGTATTAGGCTAATATTATTCTTAATATTTATTAAGGTA
CTAAAGCATTACCTATAGGTGATATTATGACAATACTAAAGTGGTTAGTATTATTAGGGTATTATTCAAAGTATTCTC
CAACACTATTCCCTTAGCTCTAACTTAATATCCCTTTAATACCAAGATGGTTAAACAATGAAGAAAATTAGTTGTTTG
TTCTGGCCCTTTACTATATATGTACATAGTACATATATAAGAAAGGCTACCATGACATTATATATAATAGGCATCCAG
AAATAAAAGCCAAATTAATCTATAATTAAGTATTTATACTTAATTATAGTTATTTTATTAACAAAATTAGTGACATTT
GGCCATCCATAATAAATAAAATCAAATAAACAAAACAAAAAGCAAAAAAACTATAAAAAGGGCATGAAAGTTTTGA
TTAATAAGAATGAGTTGAACAAAATCCTCAAAAAACTCAACAATGTAATCGTATCTAACAATAAGATGAAACCATAC
CACTCTTATTTATTAATAGAGGCTACAGAAAAGGAAATTAACTTCTATGCTAACAACGAGTACTTTTCTGCTAAATGT
ACCTTAGCCGAAAACATTGATGTACTTGAAGAAGGTGAAGTAATTGTTAAAGGCAAAATCTTTAGCGAACTCATTAA
CGGAATTAAAGAAGACATCATTACTATTCAAGAGAAAGATCAAACTCTTTTAGTCAAAACAAAAAAAACAAACATTA
ACCTTAACACGATTGATAAGAAAGAATTCCCCAGAATCCGTTTCAACCAAAACGTTGATTTGAAGGAATTTGATGAA
CTTAAAATCCAACACAGCCTTTTAACTAAAGGACTTAAAAAGATTGCCCATGCTGTTTCTACATTTAGAGAATCCACT
AGAAAATTCAACGGGGTTAACTTCAACGGTTCCAATGGTAAACAAATCTTTTTAGAGGCATCGGATTCTTATAAGCTC
TCTGTTTATGAAATCAAACAAAAAACCGATCCATTTAATTTCATTGTCGAAACTAATCTTTTGAGCTTCATCAATTCTT
TTAACCCTGAAGGTGGTGATTTAATCAGTATCTTCTTCCGCAAAGAACACAAGGATGATTTAAGTACCGAATTACTGA
TTAAGTTAGATAACTTCTTAATTAACTACACCTCAATTAACGAAAGCTTTCCGCGGGTAATGCAGTTGTTTGACTTTG
AACCAGAAACCAAAGTAACCATTCAAAAAAACGAACTTAAAGATGCATTACAAAGAATCTTGACGTTAGCTCAAAA
CGAGCGCTTCTTTCTCTGTGACATGCAAGTAACCAACTCCCACCTCAAAATTAATTCCAACGTTCAAAACATTGGTGC
ATCCTTAGAAGAAGTTACTTGCCTTAAGTTCGAAGGTCACAAACTCAACATCGCTGTTAATGCGCTTTCCCTCTTGGA
ACACATTGACTCATTTGATACTGATGAAATTGAGCTTTACTTCCAGGGCAGTAACAAATACTTTTTAATTAGTTCGAA
CAACGAACCTGAACTTAAAGAAATCCTAGTTCCTTCCAAGTAACTAATGACGCTTTATGACTTACTAGAGCTGCCCCA
AACTGCTACACTACAGGAAATCAAAACAGCTTACAAACGCTTAGCTAAACGTTATCACCCTGACATCAATAAACAAG

GGGCAGATACCTTTGTCAAGATTAATAACGCCTATGCCGTTTTAAGCGATACCACCCAAAAAGCGGAGTATGACGCT
ATGCTGCGCTTTAGTGAATTTGAAGACCGAGTTAAACGCCTGGACTTCTCGATTAAGTGACACGAACAGTTTATGGAG
GAATTACAGTTCCACCACAACTGGGACTTTGACTTTATCCGCAACAGGGAATACACCCAGCCAACCCCTACCAACAA
TAAGTATTCTTCCTTTTTAGACAAGGATGTATCCTTGGCCTTTTACCAACTTTACTCCAAGGGCAAGTTGGACTTTGAC
TTGGAAGACACACTTCTTAGACGTCATTCGATCAAGCAAGCCTTCTTAAAGGGCAAGAAGCTTAACGATGTTTTAAA
AGAGCAGTACAACTATTTAGGTTGATTAGAGGCCAAGCGCTATTTCAACATTGATGTCGAAATCGAATTAACCCCCA
AAGAAGTGCGTGAAGGTGGGGTGGTCAACCTCCCTTTAAAGATCAAAGTCATTTCCAATAACTACCCTGGACAAATG
TGGTACGAACTCAACAAGAACTATTCCTTCCGGCTTCTTTGAGACATTAAAAACGGTGAGGTGGCAGAGTTCTTTGGC
AAGGGTAATCGGGCCTTGGGTTGACGCGGTGATCTGATTGTGCGAATGCGGATTGTCGACAAAATTAAGAAGCGCTT
ACGCATCTTTTCCAGTCATTTTGAACAGGACAAAACCAAGTTGTGATTCCTCGTGCCGCAGGACAAACAAGACAATC
CTAATAAGTGGGTTTTTGACTATAAAACACACGAGTTTATTGTTTAAAATCGCTTAATTTTTAATGTGGTATAATTGTT
TGGATTCGCCATAAATTTTGCTTATGGCTTTTAACTTTCCAACGTGCACATTAAATGGAAGACAATTAAATGGAAGAC
AATAACAAAACGCAAGCTTACGATTCCAGTAGCATTAAGATTCTTGAAGGCTTAGAGGCCGTGCGTAAGCGCCCCGG
CATGTACATCGGTTCCACCGGCGAAGAAGGGTTGCACCACATGATCTGGGAAATCATTGATAACTCAATTGACGAGG
CCATGGGTGGCTTTGCGAGTACTGTCAAACTAACCCTTAA

>Ori_4 *M. pneumoniae* M129, complete genome

GGTGTTATTTAATCTTTCTGGATTTTGTCCAAAAGAGGCTGAAACGTTGCCCTGACCATCAAGGTCTAAAATAACTTT
CCGTTGTTCAGGGCACAACTTTACAAGCGAACCAGCAACGTTAGTGGCCATGGTTGTTTTTAAAACGCCACCTTTGTT
ATTAACAAAAGAAATAATCATATGCGCGCGTATAAGTTTGCTTAAATTCTAATATTTATTTAGTATTATTCCGGTATTA
TTTACCGACGAAATTAATACCATCAGGGTATTAAGATGCTACCAACGTGGTATTAAAATGTGCCCAACCGCGAAAAA
GAAAGTGGTATATAGGAAAATGGAAAACAAATTTGTTTGCCGTTTACTATTAAATACTAATCTTCTATATAGTATAGA
GAAACTTTTTCTTTAACATAATATTATCTTAATATTATTTACCTACTAATAGCTTAATATTATTAGTATTTATTTAGTAT
TATGCTAATACTATGCAGATATTATCTTAATATTATCTATAGTATTAGGCTAATATTATTCTTAATATTTATTAAGGTA
CTAAAGCATTACCTATAGGTGATATTATGACAATACTAAAGTGGTTAGTATTATTAGGGTATTATTCAAAGTATTCTC
CAACACTATTCCCTTAGCTCTAACTTAATATCCCTTTAATACCAAGATGGTTAAACAATGAAGAAAATTAGTTGTTTG
TTCTGGCCCTTTACTATATATGTACATAGTACATATATAAGAAAGGCTACCATGACATTATATATAATAGGCATCCAG
AAATAAAAGCCAAATTAATCTATAATTAAGTATTTATACTTAATTATAGTTATTTTATTAACAAAATTAGTGACATTT
GGCCATCCATAATAAATAAAATCAAATAAACAAAACAAAAAGCAAAAAAACTATAAAAAGGGCATGAAAGTTTTGA
TTAATAAGAATGAGTTGAACAAAATCCTCAAAAAACTCAACAATGTAATCGTATCTAACAATAAGATGAAACCATAC
CACTCTTATTTATTAATAGAGGCTACAGAAAAGGAAATTAACTTCTATGCTAACAACGAGTACTTTTCTGCTAAATGT
ACCTTAGCCGAAAACATTGATGTACTTGAAGAAGGTGAAGTAATTGTTAAAGGCAA

>Ori_2 *M. pneumoniae* M129, complete genome

ACTGCCTTACAAACGCCCTTGAACAAAACATCGTAATGTTTAGCCTCAGGGTCTTTGTAGTGCCGTCTTTTAGGGTGC
TTCTTATTTAACTTAACGATTAGCTTTGACGCTTTTTTACTTTGACAAAAATCGGGTCATTGCAAACAATAATCGAGGC
TTCGTTTATTTTTAGCATTTAAAGACATTGAAACATGTCAGGTTTAACAACGGGACAGGCCAGTTCAATTAAACACTT
AAAGAATTATAAAAGGACTAAATTACCAAGGAAGTGATAATGTCTTGGAGGGAGTCGTCCTTCTCAATCATTTTAGC
AACCCTTTTAACTGCAGCCACCACTGTAGAGTGGTTCCTTTTAAAAATTTTGCCTATTTGTTGGAACTGCATCTTGTAC
TTTTCCCTTAAAATATAGTTACAAACGTCACGAACACGAACTAACTCTGCTTTACGGTTCTCAGACAAAACACTGTCC
ACCGGTACATTAAAGCGGCGGCAAACGTTTTCTATTAATAAATAAGGATCAAAGCTCTTTTTGTGGAACTTCTCAAAT
TCCTCAAACAGGATTTCCTTTAGATTGTCGTTGTTAATCAGGTTTTGCTTAGTTGTTTTGACATAAAACAGCAACTTAG
TCGCAATCCCGTTTAAAGAACGAACATCGTTACCAGATATTTGCGCTGCCTTTTCGCGTGCTTCATTGGTAATTTGGAT
GTTGGGATCCTTTTCTTTGAGCTTAACGGCCAAAATTTCACACAGGGAGCTCAAATTGTGTTTTTCAATCTTTAACAAG
AGTCCCGATTTAAAGCGTGATACCATCCGTGGATCGATGTTGGCCAACTCATCGGGGGTTTTATCGGACACAAACAC
AATTTGCTTGCCCTTACTTACTAAGTTATTAAAGATGGAAAAGAGGATCTCTAAGGTTTTTTCCTTTTTCCCGAAGATC
TGAGTATCGTCAATTAAGAGTAAATCCAATGACTCGTAAGCAGCCTTGAGCTTTTCAATGCCTTGATCCTTTTCCTTTT
GGTAAAAGGCATTGACAACCTCCTGGGCAAAGTCACTCGATACCACATACTTTACCTTAGCGTTGGGAAAGTGGAAG
AACTTATCATTCCCAATCGCTTGGAGTAAGTGGGTTTTACCCAGACCAGTCTCCCCATAGATAAATAAGGGGGAGAA
ATCAGCGTCCTGAGTTTCGCTAATCCGTACACCCGCTTCATAAACCCGGCGGTTGCCATCACCAACCACAAAGTTTTT
AAAGGTAAAGTTTTTGCTTAGCCCCGAACTTTGGTAAAGCGTTTCGCGGTTATCTTCCTCTAGTTTGGCTAAATTGATA
AAGAAGTCCTGTTCGTTGACAAACCGAACTGCTCTAATGTCATCGTACAGTTTGCTAGCTAACCCGACAATATCAGGG
TTGCTATTGAGGGTTTCACGGGCAAACTCATTGTCCACAATCACAAAAAGGACATTATTTTTTAAAGCAAACCGTTTA
AGATTCTTAATGTACTTGTCATAAAAACCTAAAGTAGTTTCGTATTGCTTTTTTAAAAGAAGTTTAAAAGCACTAAAT
TGTTCCATTACTCTTCTATTACCTGATCAAGAATCATCGTTAAGATCCGCGAAATCGGAATGTTAACCGCTTTAGCAT
AACGCTTAACCTTATTCCAAACACGGGTACTCGTATAAAAGTGCACCCCGGTGGTTAAGCGTTCTTGCCGTTCTAAGG
GAATAATACCGGGCACTTTCGCCTTTGGTGGTGGTAGATTTTCTTCTTCCTGCAATTTATTAGATGGGCGAGTAGTCTG
CTTAACTTGTGGCTGTTCAACCGCTGGTTGTGGTTCTGGTTGGGGAACAGGTTCCACAATTGGTTGCGTAGCCTCCAC
CTGAATATTAGATGTAGCTACTGGTTGTTGATTAAAGTTAGTGGTTGTAGGTTGTTGAAACTCATTTAAGTGCACTTCT
GCAGTTTGGTGAACTGGTTGGGTTGGAACCTGATTATTAAAGTTAGTTGGTTCTGATTGCGGTTGCACATTAACTGGT
TGGGTTGAAGTTTGTGGTTCTTGTTTGGGAAACTGGTTAAAGGATTTCACAAACTCATCGCCAAACTTGAGTAATGGT
CCAAAATCAAACTGGGGCATGTCCTTAGAAAAGAACTTTCTGTCTTGACTAGATTGATTTGCTTTAGTTTGGTTAGTT
AGATCATGTTGGGGTTGGACAACCGGTTCTAGTTTTAATTGTTGTTCTACCTTTTTAACTTCTTGAACAGTTTCTGCAT
CATCTTCATCCTGAGCTAGGTCAAAGAAGAAGCGTTTCTTTTTACTGATGGCCATGGTTGACTAGGTTAAGAATTTCT
TTAGTAATCTCTAAATATTCAGTTTGGTACTTCTTGTTGGATGTTGGT

>Ori_5_extra_sequence *M. pneumoniae* M129, complete genome

AATCGCTTAATTTTTAATGTGGTATAATTGTTTGGATTCGCCATAAATTTTGCTTATGGCTTTTAACTTTCCAACGTGC
ACATTAAATGGAAGACAATTAA

Sequences “A synthetic “cloning platform” to supply genetic tools in Mycoplasma pneumoniae”

Sequence of Platform version 1.0

LOCUS Standalone_Platform_in_pJ251with_TNp_new_overhang 13251 bp DNA circular UNA 30-Jul-2010
DEFINITION
FEATURES Location/Qualifiers
CDS 14..808
/label=Kanamycin
Misc._feature 993..1084
/note="stuff from the Dna20 Low copy plasmid"
/label="Some Blue stuff pTR or rrnB1 B2 T1 txn terminator"
Terminator 1249..1277
/label="Terminator 1"
Promoter 1284..1339
/label="Synthetic Tet Promotor"
RBS 1340..1376
/label="RBS binding site"
CDS 1377..4043
/label="T7 Polymerase"
Misc._signal 4044..4056
/label=Stop
Terminator 4063..4088
/label="Terminator 2"
Promoter 4095..4396
/label="Mpn376 Promoter"
RBS 4397..4432
/label="Tet RBS"
CDS 4433..5437
/label="Tet Repressor"
Misc._signal 5438..5450
/label="Stop 2"
RBS 5451..5491
/label="Lac RBS"
CDS 5492..6580
/label="Lac repressor gene"
Misc._signal 6581..6593
/label="Stop 3"
Promoter 6600..6656
/label="Short Mpn376"
RBS 6657..6730
/label="CI 857 RBS"
CDS 6731..7471
/label="CI 857"
Misc._feature 7442..7471
/label="PFt c- Myc Tag"
Misc._signal 7472..7484
/label="Stop 4"
RBS 7485..7518
/label="RBS Lysozyme"
CDS 7519..7998
/label="T7 Lysozyme"
Misc._feature 7972..7998
/label="Flag Tag"
Misc._signal 7999..8011
/label="Stop 5"
Terminator 8018..8043
/label="Eva Terminator 2"
Misc._recombina 8086..8119
/label="Lox P1"
Misc._recombina 8162..8209
/label="FRT wt"
Promoter 8216..8546
/label="Spiralin Promoter"
CDS 8547..9143
/label=Puromycin
Misc._signal 9144..9156
/label="Stop 7"
Promoter 9163..9208
/label="T7 Lac Promoter"
RBS 9209..9240
/label="RBS mCherry"
CDS 9241..9948
/label=mCherry
Misc._signal 9949..9961
/label="Stop 8"
Misc._recombina 9968..10001

/label="Lox P wt"
Misc._recombina 10044..10091
/label="FRT wt"
Terminator 10098..10145
/label="T7 Terminator"
Primer 10105..10176
/label="Fwd Primer Tnp module Avr"
Misc._recombina 10155..11864
/label="Tnp module"
Misc._recombina 10329..10354
/label="IS256 site1"
CDS 10454..11626
/label=Tnp
Misc._recombina 11737..11762
/label="IS256 site2"
Primer complement(11843..11914)
/label="Primer Tnp module Avr"
Misc._feature 11942..12058
/label="Some Blue stuff pTF3 and bla txn terminator"
Misc._feature 12249..12362
/label="rpn txn terminator"
Replication_ori 12425..13251
/label="p15a Ori"
ORIGIN
1 GCTTTTAGAA AAACTCATCG AGCATCAAAT GAAACTGCAA TTTATTCATA TCAGGATTAT
61 CAATACCATA TTTTTGAAAA AGCCGTTTCT GTAATGAAGG AGAAAACTCA CCGAGGCAGT
121 TCCATAGGAT GGCAAGATCC TGGTATCGGT CTGCGATTCC GACTCGTCCA ACATCAATAC
181 AACCTATTAA TTTCCCCTCG TCAAAAATAA GGTTATCAAG TGAGAAATCA CCATGAGTGA
241 CGACTGAATC CGGTGAGAAT GGCAAAAGTT TATGCATTTC TTTCCAGACT TGTTCAACAG
301 GCCAGCCATT ACGCTCGTCA TCAAAATCAC TCGCATCAAC CAAACCGTTA TTCATTCGTG
361 ATTGCGCCTG AGCGAGGCGA AATACGCGAT CGCTGTTAAA AGGACAATTA CAAACAGGAA
421 TCGAGTGCAA CCGGCGCAGG AACACTGCCA GCGCATCAAC AATATTTTCA CCTGAATCAG
481 GATATTCTTC TAATACCTGG AACGCTGTTT TTCCGGGGAT CGCAGTGGTG AGTAACCATG
541 CATCATCAGG AGTACGGATA AAATGCTTGA TGGTCGGAAG TGGCATAAAT TCCGTCAGCC
601 AGTTTAGTCT GACCATCTCA TCTGTAACAT CATTGGCAAC GCTACCTTTG CCATGTTTCA
661 GAAACAACTC TGGCGCATCG GGCTTCCCAT ACAAGCGATA GATTGTCGCA CCTGATTGCC
721 CGACATTATC GCGAGCCCAT TTATACCCAT ATAAATCAGC ATCCATGTTG GAATTTAATC
781 GCGGCCTCGA CGTTTCCCGT TGAATATGGC TCATATTCTT CCTTTTTCAA TATTATTGAA
841 GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA
901 AACAAATAGG GGTCAGTGTT ACAACCAATT AACCAATTCT GAACATTATC GCGAGCCCAT
961 TTATACCTGA ATATGGCTCA TAACACCCCT TGTTTGCCTG GCGGCAGTAG CGCGGTGGTC
1021 CCACCTGACC CCATGCCGAA CTCAGAAGTG AAACGCCGTA GCGCCGATGG TAGTGTGGGG
1081 ACTCCCCATG CGAGAGTAGG GAACTGCCAG GCATCAAATA AAACGAAAGG CTCAGTCGAA
1141 AGACTGGGCC TTTCGCCCGG GCTAATTAGG GGGTGTCGCC CTTATTCGAC TCTATAGGGA
1201 AGTTCCTATT CTCTAGAAAG TATAGGAACT TCTGAAGGGG GGCAGCTGGC CCTCCGCTCC
1261 ACGGAGGGCT TTTTTTTGCA TGCTCAAAAA GAATTGATCT CTATCATTGA TAGATATAAT
1321 ATAAACTAAT TGGAGGATTG AGCTCGACCA TTCAACATCC AAGGAGGATA CTGCAGATGC
1381 ACCATCATCA CCACCATAAC ACCATTAACA TAGCGAAGAA CGATTTCAGT GATATTGAAC
1441 TTGCTGCGAT ACCATTTAAC ACTTTAGCCG ATCATTACGG TGAAAGATTA GCACGCGAAC
1501 AATTAGCTCT GGAACACGAA AGTTATGAAA TGGGTGAAGC GAGGTTCAGA AAGATGTTCG
1561 AACGCCAGTT AAAGGCGGGT GAAGTGGCCG ATAACGCTGC CGCCAAACCG TTGATTACCA
1621 CCCTCCTCCC AAAGATGATT GCGCGCATCA ACGATTGGTT CGAGGAAGTT AAAGCGAAAA
1681 GAGGTAAGCG CCCCACAGCG TTTCAATTTT TACAAGAGAT TAAGCCGGAA GCGGTGGCTT
1741 ATATAACGAT AAAGACCACA CTAGCCTGTT TAACGTCGGC TGACAACACT ACCGTGCAAG
1801 CTGTTGCTAG TGCAATTGGT CGTGCCATTG AAGATGAGGC CCGCTTTGGA CGTATAAGAG
1861 ATTTGGAGGC GAAACACTTT AAGAAGAACG TCGAAGAGCA ATTAAACAAA AGGGTGGGCC
1921 ACGTGTATAA GAAAGCGTTT ATGCAAGTTG TAGAAGCGGA CATGTTGTCA AAAGGGTTAT
1981 TGGGTGGTGA GGCGTGGAGT TCCTGACACA AGGAAGACAG TATCCATGTG GGCGTAAGAT
2041 GTATTGAGAT GCTCATCGAA AGTACCGGAA TGGTGAGTCT GCACCGCCAA AATGCGGGGG
2101 TAGTCGGTCA AGATTCCGAA ACCATCGAAC TCGCTCCGGA ATATGCAGAG GCAATCGCCA
2161 CCAGAGCTGG GGCATTAGCT GGTATTTCTC CAATGTTCCA GCCCTGTGTT GTACCGCCGA
2221 AACCCTGAAC CGGCATTACC GGCGGTGGTT ACTGAGCTAA CGGGAGAAGA CCGCTCGCCC
2281 TGGTTCGTAC ACACAGTAAG AAAGCCCTCA TGAGATACGA GGATGTTTAC ATGCCCGAGG
2341 TTTACAAGGC CATCAATATC GCTCAGAACA CGGCGTGGAA GATAAATAAG AAAGTATTGG
2401 CCGTGGCAAA CGTCATTACC AAGTGGAAGC ACTGTCCAGT GGAAGATATT CCAGCTATCG
2461 AAAGGGAAGA ATTACCGATG AAGCCTGAGG ACATTGACAT GAACCCAGAA GCGTTGACAG
2521 CTTGGAAGAG AGCCGCTGCG GCAGTTTACC GCAAAGACAA GGCCAGAAAA TCTCGCAGAA
2581 TCAGTTTGGA ATTTATGCTC GAACAAGCGA ACAAATTTGC CAATCATAAG GCGATTTGGT
2641 TTCCGTATAA CATGGACTGA CGCGGTCGCG TTTACGCGGT GTCCATGTTT AATCCTCAAG
2701 GCAACGACAT GACCAAAGGT TTGTTGACCT TAGCAAAGGG CAAACCCATA GGCAAAGAGG
2761 GCTATTACTG GCTCAAGATT CATGGCGCTA ACTGTGCTGG GGTTGATAAG GTTCCCTTCC
2821 CCGAAAGAAT TAAGTTCATC GAAGAGAATC ACGAGAACAT TATGGCTTGC GCTAAGTCGC
2881 CCCTGGAAAA CACTTGATGG GCGGAACAGG ATTCGCCTTT TTGTTTTCTT GCCTTTTGTT
2941 TCGAATACGC GGGGGTTCAA CACCACGGTT TGTCGTACAA CTGTAGTTTA CCGTTGGCTT
3001 TCGACGGTAG CTGTAGTGGC ATACAACACT TCTCCGCCAT GCTGAGAGAC GAAGTAGGTG
3061 GGAGGGCCGT GAACTTATTG CCCTCTGAAA CAGTACAGGA TATTTATGGG ATCGTGGCGA

3121 AAAAGGTAAA TGAAATCCTC CAGGCTGATG CCATCAACGG CACCGACAAC GAGGTCGTGA
3181 CCGTTACTGA CGAAAACACC GGGGAAATAT CAGAAAAAGT GAAACTCGGT ACGAAGGCGC
3241 TTGCTGGCCA ATGACTCGCC TATGGTGTGA CACGCAGTGT TACAAAGAGA AGTGTAATGA
3301 CCTTGGCTTA TGGGTCAAAG GAGTTTGGGT TTAGGCAGCA AGTCTTAGAG GACACCATCC
3361 AACCGGCGAT TGATAGTGGT AAGGGTCTAA TGTTTACCCA GCCTAACCAA GCAGCCGGCT
3421 ATATGGCCAA GTTAATCTGG GAGAGTGTTA GTGTAACGGT CGTTGCGGCG GTAGAAGCCA
3481 TGAACTGGCT CAAGTCGGCC GCTAAGCTCC TCGCTGCTGA GGTAAAGGAC AAGAAAACCG
3541 GCGAAATTTT GAGAAAAAGG TGCGCCGTTC ACTGAGTCAC ACCCGATGGG TTTCCCGTTT
3601 GGCAAGAATA TAAGAAACCG ATTCAAACCA GACTCAATCT GATGTTCCTC GGTCAGTTTC
3661 GCCTTCAACC AACCATCAAT ACCAACAAAG ATTCTGAGAT CGACGCGCAC AAGCAGGAAT
3721 CCGGGATCGC CCCCAATTTT GTCCATTCCC AGGATGGGAG TCACTTACGC AAGACAGTGG
3781 TGTGAGCGCA CGAAAAATAC GGCATTGAAT CTTTTGCTTT AATCCATGAC TCCTTTGGAA
3841 CCATTCCCGC CGACGCCGCT AATCTCTTTA AAGCAGTGAG AGAAACGATG GTAGATACCT
3901 ATGAAAGCTG TGATGTATTA GCCGACTTCT ACGATCAGTT TGCCGATCAA TTACATGAAT
3961 CACAGTTAGA TAAAATGCCT GCCTTACCGG CTAAGGGAAA TCTTAACTTG AGGGATATAC
4021 TAGAATCCGA TTTTGCCTTC GCGTAGGGAT AAGTAGGAAT TCGGTCCCTC TCCTGGGACC
4081 TTTTTTTTTG ATCACCAGAA GGTTATGGGT TCAAGTCCTA TTGGGCGCGC CAAATTATTG
4141 ATTCCAACCA ACTGCTCTCA GTTGGTTTTT TTATTTTAGC TACGCTTATA CACATTACTT
4201 TAACTATATT GAGATAAGCA GAGCAAAGTA ACTCTTCATT TAGAGGGTAA AATGAATGAA
4261 ATGCTTGGTG AAATACCACG CCATAAAAAG TTTTTTGAGT TCAAACCGTG CGTCTTTTGT
4321 CAATTTCGTT CAGCATTTGC TTTAATACAC GCGCAGAAGG ATTGTCAAAA ATATTTTTAA
4381 AGTGCTAGAA TAAAGCGTTA ACAACAGCGT ATATTCTGAG GAGGAAGGAT CCATGTCCAG
4441 GCTCGACAAG TCAAAGGTGA TTAACTCTGC CTTAGAACTC TTAAATGAGG TAGGTATTGA
4501 GGGTTTAACC ACCAGAAAAC TCGCACAGAA ACTCGGGGTG GAACAGCCGA CGCTTTATTG
4561 ACACGTCAAG AACAAGAGGG CACTCCTGGA CGCTCTCGCG ATTGAGATGC TGGATAGGCA
4621 TCATACTCAC TTCTGTCCCC TCGAAGGCGA AAGTTGACAA GACTTTTTGA GAAACAACGC
4681 TAAATCCTTT AGATGTGCCT TACTTTCGCA TAGAGACGGG GCTAAGGTGC ACTTAGGTAC
4741 TAGGCCAACA GAAAAGCAGT ATGAAACTTT AGAAAATCAA TTGGCTTTTC TTTGCCAACA
4801 AGGTTTTAGT CTTGAAAATG CGTTGTACGC GCTCTCGGCT GTAGGGCACT TCACGTTGGG
4861 CTGCGTCTTG GAAGACCAAG AGCACCAAGT TGCGAAGGAA GAACGTGAAA CCCCGACCAC
4921 CGACAGTATG CCTCCCTTAC TGCGCCAAGC GATCGAATTA TTTGACCACC AAGGTGCTGA
4981 GCCGGCTTTT CTCTTCGGTC TGGAACTCAT TATTTGTGGT TTAGAGAAGC AATTGAAGTG
5041 CGAGTCGGGA AGTGCTTACA GCAGAGCTAG GACAAAGAAT AACTATGGTA GCACCATCGA
5101 GGGCCTTTTA GACCTCCCCG ATGATGACGC ACCGGAGGAA GCCGGCTTGG CTGCACCCCG
5161 TTTGTCATTT CTCCCAGCCG GTCACACCCG CAGACTTTCT ACGGCGCCAC CAACCGACGT
5221 ATCGCTTGGG GATGAACTGC ACTTGGACGG TGAAGACGTT GCTATGGCCC ACGCTGATGC
5281 TTTGGACGAT TTTGACTTAG ACATGCTTGG CGATGGTGAT AGCCCTGGTC CGGGTTTCAC
5341 CCCGCACGAT TCCGCCCCCT ACGGGGCCTT GGATATGGCT GATTTTGAAT TTGAACAGAT
5401 GTTTACGGAT GCTCTGGGAA TCGACGAATA CGGCGGGTAG CTATAACTAG ACGCGTAAAC
5461 AAACTAGGAA ATATTAAGGA GGAATAGATC TATGGTGAAC GTTAAGCCTG TTACCTTATA
5521 TGATGTGGCT GAATACGCGG GAGTCAGTTA TCAAACCGTA TCTCGCGTAG TCAACCAGGC
5581 AAGTCACGTT TCTGCCAAAA CTAGAGAAAA GGTTGAAGCG GCGATGGCTG AACTGAACTA
5641 TATTCCTAAC AGAGTTGCTC AGCAATTGGC CGGCAAACAG TCTTTACTCA TCGGGGTCGC
5701 TACGAGTTCC CTTGCCCTTC ACGCACCGAG CCAAATAGTA GCAGCCATCA AGTCAAGGGC
5761 CGACCAGTTA GGAGCTTCGG TTGTGGTTAG TATGGTAGAA CGATCCGGCG TGGAGGCGTG
5821 TAAAGCAGCA GTCCATAACT TACTCGCACA GAGAGTGTCA GGTCTGATTA TCAATTATCC
5881 GCTCGACGAT CAGGATGCTA TAGCAGTAGA AGCCGCCTGC ACAAATGTAC CCGCTTTATT
5941 TCTCGACGTG TCCGATCAAA CTCCAATTAA CTCCATTATT TTCTCCCACG AAGATGGGAC
6001 CCGCTTAGGC GTGGAACACT TAGTTGCACT TGGACACCAA CAAATAGCTT TGCTAGCAGG
6061 TCCTCTTTCA AGCGTTTCCG CCCGTTTGCG CTTGGCCGGC TGACACAAGT ATTTAACGAG
6121 AAATCAAATT CAGCCAATTG CCGAACGCGA GGGAGACTGG TCGGCCATGT CCGGTTTTCA
6181 GCAAACAATG CAGATGTTAA ATGAGGGGAT TGTTCCAACC GCCATGCTCG TGGCGAATGA
6241 CCAAATGGCT CTCGGCGCCA TGAGGGCGAT AACGGAAAGT GGGCTAAGGG TGGGCGCGGA
6301 CATTTCGGTT GTTGGGTACG ATGACACAGA AGATTCTTCG TGCTACATTC CCCCACTGAC
6361 GACGATCAAA CAAGATTTTC GTTTGCTCGG GCAAACTTCG GTCGATCGCT TATTGCAGTT
6421 AAGTCAGGGG CAGGCGGTGA AGGGTAATCA GTTGCTGCCT GTTAGTTTAG TAAAACGGAA
6481 AACGACCTTA GCTCCTAACA CTCAGACCGC GTCACCAAGA GCGCTGGCTG ATAGTCTCAT
6541 GCAACTGGCG CGTCAAGTTA GTCGCCTCGA ATCTGGTCAA TAGCGATAAG TAGATGCATG
6601 AAGGATTGTC AAAAATATTT TTAAAGTGCT AGAATAAAGC ATGAAACGTT TACAGTCCGC
6661 GGCAAAAGGG ACAACAGCTT AGGCCCTCCA GCGCGGACAC TCAAATTCTA CTATAAGGAG
6721 GAGCGGCCGC ATGAGCACTA AGAAGAAGCC ATTAACCCAG GAACAGTTGG AAGACGCGAG
6781 AAGGTTGAAG GCTATTTACG AAAAGAAAAA GAACGAATTG GGTCTAAGTC AGGAATCCGT
6841 GGCCGATAAG ATGGGAATGG GACAAAGTGG GGTCGGTGCG CTATTTAACG GTATCAATGC
6901 TCTCAACGCC TACAACGCTG CGTTGCTCGC GAAGATTCTA AAGGTGTCTG TTGAAGAATT
6961 TAGCCCCAGT ATAGCTCGCG AAATCTACGA AATGTACGAA GCTGTCTCCA TGCAGCCTAG
7021 TCTGAGGTCC GAATATGAAT ATCCGGTGTT TTCCCACGTG CAGGCGGGGA TGTTTTCGCC
7081 AAAGTTGCGC ACTTTCACCA AGGGCGATGC GGAAAGATGA GTGAGTACCA CTAAAAAGGC
7141 GTCAGATAGT GCTTTTTGAT TAGAAGTGGA AGGGAACAGT ATGACGGCGC CCACGGGGTC
7201 GAAGCCATCG TTTCCCGATG GGATGCTCAT TCTCGTAGAT CCAGAGCAGG CAGTCGAACC
7261 AGGCGATTTC TGTATTGCAA GGTTGGGTGG TGACGAGTTT ACTTTTAAGA AGCTAATAAG
7321 AGATAGTGGA CAAGTGTTCT TACAGCCGCT CAACCCGCAG TACCCGATGA TTCCCTGCAA
7381 CGAGAGCTGC TCTGTTGTGG GCAAAGTGAT TGCTTCACAA TGGCCGGAAG AAACATTTGG
7441 TGAACAAAAA CTGATTTCCG AAGAAGACCT CTAGTCTTAA TTAGAGTACT ACGACGCAGA
7501 AGGAGGAGGA GAAGGCCTAT GGCTCGGGTG CAGTTCAAGC AACGCGAAAG TACCGATGCG
7561 ATTTTTGTGC ACTGTAGTGC CACAAAACCT AGTCAAAATG TGGGGGTACG CGAGATTCGC
7621 CAGTGGCACA AGGAACAAGG CTGACTTGAC GTTGGTTACC ACTTCATTAT CAAGCGCGAT

7681 GGAACCGTAG AAGCTGGTCG CGATGAGATG GCAGTCGGAA GCCATGCAAA AGGGTATAAC
7741 CACAATTCCA TCGGGGTTTG CTTGGTTGGT GGTATTGACG ACAAAGGCAA ATTCGACGCT
7801 AATTTTACCC CAGCGCAAAT GCAAAGTTTA CGCTCATTAT TGGTCACGTT GCTGGCGAAG
7861 TATGAAGGGG CTGGCCTCCG CGCACACCAT GAAGTTGCGC CAAAAGCCTG TCCATCCTTT
7921 GACTTGAAGC GATGATGGGA AAAGAATGAA TTAGTAACAT CCGATCGCGG CGACTACAAG
7981 GACGACGACG ATAAGGGATA GACATAAGTA GAAGCTTGCC GGGTCAGCTC CCGGCTTTTT
8041 TTTCATATGG GGGGTGAAAA CGTCAAACTT TCGGCTACGA CCGGTATAAC TTCGTATAGT
8101 ATAGTATATA CGAACGGTAC TCGAGGCAGA AAGAAGTACC CACGGTTATA TTCTCACTAG
8161 TGAAGTTCCT ATTCCGAAGT TCCTATTCTC TAGAAAGTAT AGGAACTTCG GTACCCCGCC
8221 AGAATTAAAA GTTAGTGAAC AAGAAAACAG TGAAGCACCA GTTTCTGAAC CAAAAGAAGA
8281 CGAAAAAACA AAAAAAGATT AAGCAATTTA TTTGGAAAAT CTTTTTTTGT TTTTTTAAGA
8341 AATATTTATT GTTTTTTTTA AAAAATTATG TACAATTGCT ACTATAAGGG AAAGAAAAAA
8401 AGAAAGATAT AAATTGTATA AAGTAGGGTT AGAAGCAATT AATAATTATT ATTAATGTTA
8461 TTTTTCTCTT ATATATTCAA TGTAATTTTA ATTACATTTG CTTTTAATAA AAACACTACT
8521 TAATAGAGAA AGGAAATATA GATACTATGA CCGAGTATAA ACCCACCGTA CGCCTGGCGA
8581 CCAGGGACGA CGTTCCTCGT GCGGTGAGGA CACTCGCTGC CGCATTTGCC GACTATCCAG
8641 CCACGAGACA CACCGTTGAC CCAGATCGCC ACATTGAAAG AGTGACCGAG TTGCAAGAAC
8701 TTTTCCTCAC CAGAGTGGGG CTTGACATTG GAAAGGTGTG GGTAGCTGAC GACGGGGCTG
8761 CCGTTGCCGT CTGGACCACG CCTGAATCCG TGGAAGCTGG TGCAGTTTTT GCAGAGATTG
8821 GGCCCAGAAT GGCCGAGCTT AGCGGGTCCA GATTAGCAGC CCAACAACAG ATGGAGGGCT
8881 TACTCGCTCC ACATAGACCA AAAGAACCCG CGTGGTTTTT AGCCACAGTC GGCGTTTCGC
8941 CCGATCACCA AGGTAAGGGT CTTGGGTCCG CTGTGGTTCT TCCTGGAGTA GAGGCTGCCG
9001 AAAGGGCTGG CGTGCCCGCC TTTTTGGAAA CTAGCGCTCC ACGCAACCTA CCATTTTATG
9061 AACGACTAGG CTTTACCGTC ACGGCGGATG TGGAGGTGCC AGAAGGTCCC CGAACCTGGT
9121 GTATGACTAG GAAACCAGGG GCTTAGAGGT AAGTAGCTTA AGTAATACGA CTCACTATAG
9181 GGGAATTGTG AGCGGATAAC AATTCCCCTA TAAGTAATAA ACTAGATAAG GAGGATTAGA
9241 ATGGTTTCCA AGGGCGAAGA GGATAACATG GCTATCATTA AGGAGTTCAT GAGATTCAAG
9301 GTTCACATGG AAGGTAGTGT GAACGGGCAC GAGTTCGAAA TTGAGGGGGA AGGCGAAGGG
9361 CGTCCTTATG AAGGAACCCA GACTGCCAAA CTTAAAGTAA CGAAGGGAGG CCCTTTGCCA
9421 TTTGCTTGAG ACATACTGAG TCCCCAGTTT ATGTATGGTT CAAAGGCTTA CGTCAAGCAC
9481 CCCGCTGATA TTCCAGACTA CCTCAAGTTG AGTTTTCCGG AGGGTTTCAA ATGAGAACGT
9541 GTTATGAACT TCGAAGATGG TGGCGTCGTG ACTGTTACCC AGGACAGTTC TTTACAAGAT
9601 GGGGAGTTCA TTTACAAGGT TAAGTTGCGA GGCACTAATT TTCCATCCGA TGGCCCGGTA
9661 ATGCAAAAGA AAACAATGGG GTGGGAAGCC TCGTCCGAGA GAATGTACCC AGAAGACGGT
9721 GCGTTAAAGG GGGAAATTAA GCAAAGATTG AAGTTAAAAG ATGGGGGTCA TTATGATGCC
9781 GAGGTCAAGA CCACGTATAA GGCTAAGAAG CCAGTCCAAC TCCCAGGAGC GTATAACGTC
9841 AACATCAAAC TCGACATCAC GTCCCATAAC GAGGACTATA CTATTGTAGA ACAGTACGAA
9901 CGCGCGGAAG GTCGTCACTC CACTGGTGGG ATGGATGAAC TCTACAAGTA GTGCTAACTA
9961 GCCCGGGATA ACTTCGTATA GCATACATTA TACGAAGTTA TGATATCATT CCGTGATGCT
10021 TCATACGTAT GCTATGTCCA TGGGAAGTTC CTATTCCGAA GTTCCTATTC TTCAAATAGT
10081 ATAGGAACTT CGTCGACCTA GCATAACCCC TTGGGGCCTC TAAACGGGTC TTGAGGGGTT
10141 TTTTGGTTTA AACCTTCTGT GGATAACCGT ATTACCGCCT TTGAGTGAGC TGATACCGCT
10201 CGCCGCAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA
10261 ATACGCAAAC CGCCTCTCCC CGCGCGTTGG CCGATTCATT AATGCACGCT AGCATTTAAA
10321 TGGTACCCTT TTACACAATT ATACGGACTT TATCCTGCAG GGGCCCAATT GTGTAAAAGT
10381 AAAAAGGCCA TATAACAGTC CTTTTACGGT ACAATGTTTT TAACGACAAA AACATACCCA
10441 GGAGGACTTT TACATGACCC AAGTACATTT TACACTGAAA AGCGAAGAGA TTCAAAGCAT
10501 TATTGAATAT TCTGTAAAGG ATGACGTTTC TAAAAATATT TTAACAACGG TATTTAATCA
10561 ACTAATGGAA AATCAACGAA CAGAATATAT TCAAGCAAAA GAATATGAAC GAACAGAAAA
10621 CCGACAAAGT CAACGAAATG GCTATTATGA GCGCAGCTTT ACGACACGTG TAGGCACGCT
10681 AGAATTAAAA GTACCCAGAA CACGTGATGG CCATTTTTCA CCCACAGTGT TTGAACGTTA
10741 TCAACGAAAC GAAAAAGCCC TCATGGCTTC AATGTTGGAA ATGTATGTAT CAGGCGTTTC
10801 AACTCGTAAA GTATCAAAAA TTGTGGAAGA ACTTTGTGGT AAATCCGTCT CTAAGTCCTT
10861 CGTTTCTAGC TTAACAGAAC AGCTAGAACC TATGGTTAAC GAGTGACAGA ATCGTTTATT
10921 ATCAGAAAAA AATTATCCTT ACTTAATGAC CGATGTACTC TATATAAAAG TACGAGAAGA
10981 AAATCGAGTA CTCTCAAAAA GCTGTCATAT AGCGATTGGA ATAACCAAAG ATGGCGACCG
11041 TGAAATTATC GGCTTCATGA TTCAAAGTGG CGAAAGCGAA GAGACCTGGA CAACATTTTT
11101 TGAATACCTA AAAGAACGCG GTTTACAAGG TACGGAACTC GTTATTTCTG ATGCGCACAA
11161 AGGATTAGTC TCTGCCATTA GAAAATCCTT CACCAACGTA AGTTGGCAAA GATGCCAAGT
11221 TCACTTCCTA AGAAATATCT TTACCACCAT TCCTAAAAAA AATTCAAAAT CTTTCAGAGA
11281 AGCTGTTAAA GGAATTTTTA AGTTCACAGA TATTAACTTA GCGCGTGAGG CTAAAAATCG
11341 ATTGATTCAT GATTATATCG ATCAACCAAA ATATTCAAAA GCTTGCGCAT CATTGGATGA
11401 TGGATTCGAA GACGCCTTTC AATATACCGT ACAAGGAAAT TCCCACAATC GACTAAAGAG
11461 TACCAATCTA ATTGAACGAC TGAATCAAGA AGTACGCAGA AGAGAAAAGA TTATTCGCAT
11521 CTTCCCCAAT CAAACATCAG CCAATCGCTT AATTGGAGCC GTTCTTATGG ACCTACATGA
11581 TGAATGGATT TATTCTTCAA GAAAATACAT CAATTTTGAT AAGTAGAAAT GGTAAAAACA
11641 TTGTATAGCA TTTTACACAG GAGTCTGGAC TTGACTCACT TCCTTTATTA TTTTTCATTT
11701 TTTTGACCTC GAGGGGGGGC CCACCATACA GCTGACGATA AAGTCCGTAT AATTGTGTAA
11761 AAACCCATAG CTTTGGACAC ACACTAGTAC GGATCCACCC GCAATTACTG TGAGTTAGCT
11821 CACTCATTAG GCACCCCAGG CTTTACACTT TATACTTCCG GCTCGAAAAT GAAGGGAAGT
11881 TCCTATACTT TCTAGAGAAT AGGAACTTCT ATAGGGAGTC GAATAAGGGC GACACAAAAG
11941 GTATTCTAAA TGCATAATAA ATACTGATAA CATCTTATAG TTTGTATTAT ATTTTGTATT
12001 ATCGTTGACA TGTATAATTT TGATATCAAA AACTGATTTT CCCTTTATTA TTTTCGAGAT
12061 TTATTTTCTT AATTCTCTTT AACAAACTAG AAATATTGTA TATACAAAAA ATCATAAATA
12121 ATAGATGAAT AGTTTAATTA TAGGTGTTCA TCAATCGAAA AAGCAACGTA TCTTATTTAA
12181 AGTGCGTTGC TTTTTTCTCA TTTATAAGGT TAAATAATTC TCATATATCA AGCAAAGTGA

```
   12241 CAGGCGCCCT TAAATATTCT GACAAATGCT CTTTCCCTAA ACTCCCCCCA TAAAAAAACC
   12301 CGCCGAAGCG GGTTTTTACG TTATTTGCGG ATTAACGATT ACTCGTTATC AGAACCGCCC
   12361 AGGATGCCTG GCAGTTCCCT ACTCTCGCCG CTGCGCTCGG TCGTTCGGCT GCGGGACCTC
   12421 AGCGCTAGCG GAGTGTATAC TGGCTTACTA TGTTGGCACT GATGAGGGTG TCAGTGAAGT
   12481 GCTTCATGTG GCAGGAGAAA AAAGGCTGCA CCGGTGCGTC AGCAGAATAT GTGATACAGG
   12541 ATATATTCCG CTTCCTCGCT CACTGACTCG CTACGCTCGG TCGTTCGACT GCGGCGAGCG
   12601 GAAATGGCTT ACGAACGGGG CGGAGATTTC CTGGAAGATG CCAGGAAGAT ACTTAACAGG
   12661 GAAGTGAGAG GGCCGCGGCA AAGCCGTTTT TCCATAGGCT CCGCCCCCCT GACAAGCATC
   12721 ACGAAATCTG ACGCTCAAAT CAGTGGTGGC GAAACCCGAC AGGACTATAA AGATACCAGG
   12781 CGTTTCCCCC TGGCGGCTCC CTCGTGCGCT CTCCTGTTCC TGCCTTTCGG TTTACCGGTG
   12841 TCATTCCGCT GTTATGGCCG CGTTTGTCTC ATTCCACGCC TGACACTCAG TTCCGGGTAG
   12901 GCAGTTCGCT CCAAGCTGGA CTGTATGCAC GAACCCCCCG TTCAGTCCGA CCGCTGCGCC
   12961 TTATCCGGTA ACTATCGTCT TGAGTCCAAC CCGGAAAGAC ATGCAAAAGC ACCACTGGCA
   13021 GCAGCCACTG GTAATTGATT TAGAGGAGTT AGTCTTGAAG TCATGCGCCG GTTAAGGCTA
   13081 AACTGAAAGG ACAAGTTTTG GTGACTGCGC TCCTCCAAGC CAGTTACCTC GGTTCAAAGA
   13141 GTTGGTAGCT CAGAGAACCT TCGAAAAACC GCCCTGCAAG GCGGTTTTTT CGTTTTCAGA
   13201 GCAAGAGATT ACGCGCAGAC CAAAACGATC TCAAGAAGAT CATCTTATTA A
//
```

Complete sequence Version 2.0 of the Platform

```
LOCUS       Complete_New_Standalone_EfTu     19916 bp    DNA     linear   UNA 30-Jul-2010
DEFINITION
FEATURES             Location/Qualifiers
     Gene            172..846
                     /label=TRP1
     Replication_ori 1491..1759
                     /label=ARS4/CEN5
     Promoter        2161..2294
                     /label="Spectinomycin Promoter"
     CDS             2295..3305
                     /label=Spectinomycin
     ORF             3347..3508
                     /label="COS ORF"
     Gene            4677..5432
                     /label=repE
     Gene            6011..7186
                     /label=sopA
     Gene            7186..8157
                     /label=sopB
     Replication_ori 8995..9284
                     /label="ori T 1"
     Promoter        9281..9630
                     /label="Tet Ptom order"
     Misc._feature   9285..9626
                     /label="Tet promoter as described by Breton et al. ordered from Genscript"
     Promoter        9381..9530
                     /label="Tet Ptomoter by Breton"
     Misc._signal    9516..9522
                     /label="SD Kamionka"
     CDS             9531..12197
                     /label="T7 Polymerase"
     Misc._signal    12198..12210
                     /label=Stop
     Terminator      12217..12242
                     /label="Terminator 2"
     Promoter        12273..12480
                     /label="EfTu Promotor including ATG"
     CDS             12481..13482
                     /label="Tet Repressor"
     Misc._signal    13483..13495
                     /label="Stop 2"
     RBS             13496..13536
                     /label="Lac RBS"
     CDS             13537..14625
                     /label="Lac repressor gene"
     Misc._signal    14626..14638
                     /label="Stop 3"
     Promoter        14645..14701
                     /label="Short Mpn376"
     RBS             14702..14775
                     /label="CI 857 RBS"
     CDS             14776..15516
                     /label="CI 857"
```

```
     Misc._feature   15487..15516
                     /label="PFt c- Myc Tag"
     Misc._signal    15517..15529
                     /label="Stop 4"
     RBS             15530..15563
                     /label="RBS Lysozyme"
     CDS             15564..16043
                     /label="T7 Lysozyme"
     Misc._feature   16017..16043
                     /label="Flag Tag"
     Misc._signal    16044..16056
                     /label="Stop 5"
     Terminator      16063..16088
                     /label="Eva Terminator 2"
     Misc._recombina 16131..16164
                     /label="Lox P1"
     Misc._recombina 16207..16254
                     /label="FRT wt"
     Promoter        16261..16591
                     /label="Spiralin Promoter"
     CDS             16592..17188
                     /label=Puromycin
     Misc._signal    17189..17201
                     /label="Stop 7"
     Promoter        17208..17253
                     /label="T7 Lac Promoter"
     RBS             17254..17285
                     /label="RBS mCherry"
     CDS             17286..17993
                     /label=mCherry
     Misc._signal    17994..18006
                     /label="Stop 8"
     Misc._recombina 18013..18046
                     /label="Lox P wt"
     Misc._recombina 18089..18136
                     /label="FRT wt"
     Terminator      18143..18190
                     /label="T7 Terminator"
     Misc._recombina 18381..18406
                     /label="IS 256"
     ORF             18998..19678
                     /label=Transposase
     Misc._recombina 19789..19814
                     /label="IS 256"
ORIGIN
        1 CCTCGCCGCA GTTAATTAAA GTCAGTGAGC GAGGAAGCGC GTAACTATAA CGGTCCTAAG
       61 GTAGCGAATC CTGATGCGGT ATTTTCTCCT TACGCATCTG TGCGGTATTT CACACCGCAT
      121 AGATCGGCAA GTGCACAAAC AATACTTAAA TAAATACTAC TCAGTAATAA CCTATTTCTT
      181 AGCATTTTTG ACGAAATTTG CTATTTTGTT AGAGTCTTTT ACACCATTTG TCTCCACACC
      241 TCCGCTTACA TCAACACCAA TAACGCCATT TAATCTAAGC GCATCACCAA CATTTTCTGG
      301 CGTCAGTCCA CCAGCTAACA TAAAATGTAA GCTTTCGGGG CTCTCTTGCC TTCCAACCCA
      361 GTCAGAAATC GAGTTCCAAT CCAAAAGTTC ACCTGTCCCA CCTGCTTCTG AATCAAACAA
      421 GGGAATAAAC GAATGAGGTT TCTGTGAAGC TGCACTGAGT AGTATGTTGC AGTCTTTTGG
      481 AAATACGAGT CTTTTAATAA CTGGCAAACC GAGGAACTCT TGGTATTCTT GCCACGACTC
      541 ATCTCCATGC AGTTGGACGA TATCAATGCC GTAATCATTG ACCAGAGCCA AAACATCCTC
      601 CTTAAGTTGA TTACGAAACA CGCCAACCAA GTATTTCGGA GTGCCTGAAC TATTTTTATA
      661 TGCTTTTACA AGACTTGAAA TTTTCCTTGC AATAACCGGG TCAATTGTTC TCTTTCTATT
      721 GGGCACACAT ATAATACCCA GCAAGTCAGC ATCGGAATCT AGAGCACATT CTGCGGCCTC
      781 TGTGCTCTGC AAGCCGCAAA CTTTCACCAA TGGACCAGAA CTACCTGTGA AATTAATAAC
      841 AGACATACTC CAAGCTGCCT TTGTGTGCTT AATCACGTAT ACTCACGTGC TCAATAGTCA
      901 CCAATGCCCT CCCTCTTGGC CCTCTCCTTT TCTTTTTTCG ACCGAATTAA TTCTTAATCG
      961 GCAAAAAAAG AAAAGCTCCG GATCAAGATT GTACGTAAGG TGACAAGCTA TTTTTCAATA
     1021 AAGAATATCT TCCACTACTG CCATCTGGCG TCATAACTGC AAAGTACACA TATATTACGA
     1081 TGCTGTTCTA TTAAATGCTT CCTATATTAT ATATATAGTA ATGTCGTGAT CTATGGTGCA
     1141 CTCTCAGTAC AATCTGCTCT GATGCCGCAT AGTTAAGCCA GCCCCGACAC CCGCCAACAC
     1201 CCGCTGACGC GCCCTGACGG GCTTGTCTGC TCCCGGCATC CGCTTACAGA CAAGCTGTGA
     1261 CCGTCTCCGG GAGCTGCATG TGTCAGAGGT TTTCACCGTC ATCACCGAAA CGCGCGAGAC
     1321 GAAAGGGCCT CGTGATACGC CTATTTTTAT AGGTTAATGT CATGATAATA ATGGTTTCTT
     1381 AGACGGATCG CTTGCCTGTA ACTTACACGC GCCTCGTATC TTTTAATGAT GGAATAATTT
     1441 GGGAATTTAC TCTGTGTTTA TTTATTTTTA TGTTTTGTAT TTGGATTTTA GAAAGTAAAT
     1501 AAAGAAGGTA GAAGAGTTAC GGAATGAAGA AAAAAAAATA AACAAAGGTT TAAAAAATTT
     1561 CAACAAAAAG CGTACTTTAC ATATATATTT ATTAGACAAG AAAAGCAGAT TAAATAGATA
     1621 TACATTCGAT TAACGATAAG TAAAATGTAA AATCACAGGA TTTTCGTGTG TGGTCTTCTA
     1681 CACAGACAAG GTGAAACAAT TCGGCATTAA TACCTGAGAG CAGGAAGAGC AAGATAAAAG
     1741 GTAGTATTTG TTGGCGATCC CCCTAGAGTC TTTTACATCT TCGGAAAACA AAAACTATTT
     1801 TTTCTTTAAT TTCTTTTTTT ACTTTCTATT TTTAATTTAT ATATTTATAT TAAAAAATTT
```

1861 AAATTATAAT TATTTTTATA GCACGTGATG AAAAGGACCC AGGTGGCACT TTTCGGGGAA
1921 ATGTGCGCGG AACCCCTATT TGTTTATTTT TCTAAATACA TTCAAATATG TATCCGCTCA
1981 TGAGACAATA ACCCTGATAA ATGCTTCAAT AATATTGAAA AAGGAAGAGT ATGAGTATTC
2041 AACATTTCCG TGTCGCCCTT ATTCCCTTTT TTGCGGCATT TTGCCTTCCT GTTTTTGCTC
2101 ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGACGC GTAGTCTAGA
2161 CCAGCCAGGA CAGAAATGCC TCGACTTCGC TGCTACCCAA GGTTGCCGGG TGACGCACAC
2221 CGTGGAAACG GATGAAGGCA CGAACCCAGT GGACATAAGC CTGTTCGGTT CGTAAGCTGT
2281 AATGCAAGTA GCGTATGCGC TCACGCAACT GGTCCAGAAC CTTGACCGAA CGCAGCGGTG
2341 GTAACGGCGC AGTGGCGGTT TTCATGGCTT GTTATGACTG TTTTTTTGGG GTACAGTCTA
2401 TGCCTCGGGC ATCCAAGCAG CAAGCGCGTT ACGCCGTGGG TCGATGTTTG ATGTTATGGA
2461 GCAGCAACGA TGTTACGCAG CAGGGCAGTC GCCCTAAAAC AAAGTTAAAC ATTATGAGGG
2521 AAGCGGTGAT CGCCGAAGTA TCGACTCAAC TATCAGAGGT AGTTGGCGCC ATCGAGCGCC
2581 ATCTCGAACC GACGTTGCTG GCCGTACATT TGTACGGCTC CGCAGTGGAT GGCGGCCTGA
2641 AGCCACACAG TGATATTGAT TTGCTGGTTA CGGTGACCGT AAGGCTTGAT GAAACAACGC
2701 GGCGAGCTTT GATCAACGAC CTTTTGGAAA CTTCGGCTTC CCCTGGAGAG AGCGAGATTC
2761 TCCGCGCTGT AGAAGTCACC ATTGTTGTGC ACGACGACAT CATTCCGTGG CGTTATCCAG
2821 CTAAGCGCGA ACTGCAATTT GGAGAATGGC AGCGCAATGA CATTCTTGCA GGTATCTTCG
2881 AGCCAGCCAC GATCGACATT GATCTGGCTA TCTTGCTGAC AAAAGCAAGA GAACATAGCG
2941 TTGCCTTGGT AGGTCCAGCG GCGGAGGAAC TCTTTGATCC GGTTCCTGAA CAGGATCTAT
3001 TTGAGGCGCT AAATGAAACC TTAACGCTAT GGAACTCGCC GCCCGACTGG GCTGGCGATG
3061 AGCGAAATGT AGTGCTTACG TTGTCCCGCA TTTGGTACAG CGCAGTAACC GGCAAAATCG
3121 CGCCGAAGGA TGTCGCTGCC GGCTGGGCAA TGGAGCGCCT GCCGGCCCAG TATCAGCCCG
3181 TCATACTTGA AGCTAGACAG GCTTATCTTG GACAAGAAGA AGATCGCTTG GCCTCGCGCG
3241 CAGATCAGTT GGAAGAATTT GTCCACTACG TGAAAGGCGA GATCACCAAG GTAGTCGGCA
3301 AATAACCCTC GAGCATTCAA GGCGCCTTGA TTATTTGACG TGGTTTGATG GCCTCCACGC
3361 ACGTTGTGAT ATGTAGATGA TAATCATTAT CACTTTACGG GTCCTTTCCG GTGATCCGAC
3421 AGGTTACGGG GCGGCGACCT CGCGGGTTTT CGCTATTTAT GAAAATTTTC CGGTTTAAGG
3481 CGTTTCCGTT CTTCTTCGTC ATAACTTAAT GTTTTTATTT AAAATACCTC GCGAGTGGCA
3541 ACACTGAAAA TACCCATGGA GCGGCGTAAC CGTCGCACAG GAAGGACAGA GAAAGCGCGG
3601 ATCTGGGAAG TGACGGACAG AACGGTCAGG ACCTGGATTG GGGAGGCGGT TGCCGCCGCT
3661 GCTGCTGACG GTGTGACGTT CTCTGTTCCG GTCACACCAC ATACGTTCCG CCATTCCTAT
3721 GCGATGCACA TGCTGTATGC CGGTATACCG CTGAAAGTTC TGCAAAGCCT GATGGGACAT
3781 AAGTCCATCA GTTCAACGGA GGTCTACACG AAGGTTTTTG CGCTGGATGT GGCTGCCCGG
3841 CACCGGGTGC AGTTTGCGAT GCCGGAGTCT GATGCGGTTG CGATGCTGAA ACAATTATCC
3901 TGAGAATAAA TGCCTTGGCC TTTATATGGA AATGTGGAAC TGAGTGGATA TGCTGTTTTT
3961 GTCTGTTAAA CAGAGAAGCT GGCTGTTATC CACTGAGAAG CGAACGAAAC AGTCGGGAAA
4021 ATCTCCCATT ATCGTAGAGA TCCGCATTAT TAATCTCAGG AGCCTGTGTA GCGTTTATAG
4081 GAAGTAGTGT TCTGTCATGA TGCCTGCAAG CGGTAACGAA AACGATTTGA ATATGCCTTC
4141 AGGAACAATA GAAATCTTCG TGCGGTGTTA CGTTGAAGTG GAGCGGATTA TGTCAGCAAT
4201 GGACAGAACA ACCTAATGAA CACAGAACCA TGATGTGGTC TGTCCTTTTA CAGCCAGTAG
4261 TGCTCGCCGC AGTCGAGCGA CAGGGCGAAG CCCTCGAGTG AGCGAGGAAG CACCAGGGAA
4321 CAGCACTTAT ATATTCTGCT TACACACGAT GCCTGAAAAA ACTTCCCTTG GGGTTATCCA
4381 CTTATCCACG GGGATATTTT TATAATTATT TTTTTTATAG TTTTTAGATC TTCTTTTTTA
4441 GAGCGCCTTG TAGGCCTTTA TCCATGCTGG TTCTAGAGAA GGTGTTGTGA CAAATTGCCC
4501 TTTCAGTGTG ACAAATCACC CTCAAATGAC AGTCCTGTCT GTGACAAATT GCCCTTAACC
4561 CTGTGACAAA TTGCCCTCAG AAGAAGCTGT TTTTTCACAA AGTTATCCCT GCTTATTGAC
4621 TCTTTTTTAT TTAGTGTGAC AATCTAAAAA CTTGGCACAC TTCACATGGA TCTGTCATGG
4681 CGGAAACAGC GGTTATCAAT CACAAGAAAC GTAAAAATAG CCCGCGAATC GTCCAGTCAA
4741 ACGACCTCAC TGAGGCGGCA TATAGTCTCT CCCGGGATCA AAAACGTATG CTGTATCTGT
4801 TCGTTGACCA GATCAGAAAA TCTGATGGCA CCCTACAGGA ACATGACGGT ATCTGCGAGA
4861 TCCATGTTGC TAAATATGCT GAAATATTCG GATTGACCTC TGCGGAAGCC AGTAAGGATA
4921 TACGGCAGGC ATTGAAGAGT TTCGCGGGGA AGGAAGTGGT TTTTTATCGC CCTGAAGAGG
4981 ATGCCGGCGA TGAAAAAGGC TATGAATCTT TTCCTTGGTT TATCAAACGT GCGCACAGTC
5041 CATCCAGAGG GCTTTACAGT GTACATATCA ACCCATATCT CATTCCCTTC TTTATCGGGT
5101 TACAGAACCG GTTTACGCAG TTTCGGCTTA GTGAAACAAA AGAAATCACC AATCCGTATG
5161 CCATGCGTTT ATACGAATCC CTGTGTCAGT ATCGTAAGCC GGATGGCTCA GGCATCGTCT
5221 CTCTGAAAAT CGACTGGATC ATAGAGCGTT ACCAGCTGCC TCAAAGTTAC CAGCGTATGC
5281 CTGACTTCCG CCGCCGCTTC CTGCAGGTCT GTGTTAATGA GATCAACAGC AGAACTCCAA
5341 TGCGCCTCTC ATACATTGAG AAAAAGAAAG GCCGCCAGAC GACTCATATC GTATTTTCCT
5401 TCCGCGATAT CACTTCCATG ACGACAGGAT AGTCTGAGGG TTATCTGTCA CAGATTTGGG
5461 GGTGGTTCGT CACATTTGTT CTGACCTACT GAGGGTAATT TGTCACAGTT TTGCTGTTTC
5521 CTTCAGCCTG CATGGATTTT CTCATACTTT TTGAACTGTA ATTTTTAAGG AAGCCAAATT
5581 TGAGGGCAGT TTGTCACAGT TGATTTCCTT CTCTTTCCCT TCGTCATGTG ACCTGATATC
5641 GGGGGTTAGT TTGTCATCAT TGATGAGGGT TGATTATCAC AGTTTATTAC TCTGAATTGG
5701 CTATCCGCGT GTGTACCTCT ACCTGGAGTT TTTCCCACGG TGGATATTTC TTCTTGCGCT
5761 GAGCGTAAGA GCTATCTGAC AGAACAGTTC TTCTTTGCTT CCTCGCCAGT TCGCTCGCTA
5821 TGCTCGGTTA CACGGCTGCG GCGAGCGCTA GTGATAATAA GTGACTGAGG TATGTGCTCT
5881 TCTTATCTCC TTTTGTAGTG TTGCTCTTAT TTTAAACAAC TTTGCGGTTT TTTGATGACT
5941 TTGCGATTTT GTTGTTGCTT TGCAGTAAAT TGCAAGATTT AATAAAAAAA CGCAAAGCAA
6001 TGATTAAAGG ATGTTCAGAA TGAAACTCAT GGAAACACTT AACCAGTGCA TAAACGCTGG
6061 TCATGAAATG ACGAAGGCTA TCGCCATTGC ACAGTTTAAT GATGACAGCC CGGAGGCGAG
6121 GAAAATAACC CGGCGCTGGA GAATAGGTGA AGCAGCGGAT TTAGTTGGGG TTTCTTCTCA
6181 GGCTATCAGA GATGCCGAGA AAGCAGGGCG ACTACCGCAC CCGGATATGG AAATTCGAGG
6241 ACGGGTTGAG CAACGTGTTG GTTATACAAT TGAACAAATT AATCATATGC GTGATGTGTT
6301 TGGTACGCGA TTGCGACGTG CTGAAGACGT ATTTCCACCG GTGATCGGGG TTGCTGCCCA
6361 TAAAGGTGGC GTTTACAAAA CCTCAGTTTC TGTTCATCTT GCTCAGGATC TGGCTCTGAA

6421 GGGGCTACGT GTTTTGCTCG TGGAAGGTAA CGACCCCCAG GGAACAGCCT CAATGTATCA
6481 CGGATGGGTA CCAGATCTTC ATATTCATGC AGAAGACACT CTCCTGCCTT TCTATCTTGG
6541 GGAAAAGGAC GATGTCACTT ATGCAATAAA GCCCACTTGC TGGCCGGGGC TTGACATTAT
6601 TCCTTCCTGT CTGGCTCTGC ACCGTATTGA AACTGAGTTA ATGGGCAAAT TTGATGAAGG
6661 TAAACTGCCC ACCGATCCAC ACCTGATGCT CCGACTGGCC ATTGAAACTG TTGCTCATGA
6721 CTATGATGTC ATAGTTATTG ACAGCGCGCC TAACCTGGGT ATCGGCACGA TTAATGTCGT
6781 ATGTGCTGCT GATGTGCTGA TTGTTCCCAC GCCTGCTGAG TTGTTTGACT ACACCTCCGC
6841 ACTGCAGTTT TTCGATATGC TTCGTGATCT GCTCAAGAAC GTTGATCTTA AAGGGTTCGA
6901 GCCTGATGTA CGTATTTTGC TTACCAAATA CAGCAATAGT AATGGCTCTC AGTCCCCGTG
6961 GATGGAGGAG CAAATTCGGG ATGCCTGGGG AAGCATGGTT CTAAAAAATG TTGTACGTGA
7021 AACGGATGAA GTTGGTAAAG GTCAGATCCG GATGAGAACT GTTTTTGAAC AGGCCATTGA
7081 TCAACGCTCC TCAACTGGTG CCTGGAGAAA TGCTCTTTCT ATTTGGGAAC CTGTCTGCAA
7141 TGAAATTTTC GATCGCCTGA TTAAACCACG CTGGGAGATT AGATAATGAA GCGTGCGCCT
7201 GTTATTCCAA AACATACGCT CAATACTCAA CCGGTTGAAG ATACTTCGTT ATCGACACCA
7261 GCTGCCCCGA TGGTGGATTC GTTAATTGCG CGCGTAGGAG TAATGGCTCG CGGTAATGCC
7321 ATTACTTTGC CTGTATGTGG TCGGGATGTG AAGTTTACTC TTGAAGTGCT CCGGGGTGAT
7381 AGTGTTGAGA AGACCTCTCG GGTATGGCCA GGTAATGAAC GTGACCAGGA GCTGCTTACT
7441 GAGGACGCAC TGGATGATCT CATCCCTTCT TTTCTACTGA CTGGTCAACA GACACCGGCG
7501 TTCGGTCGAA GAGTATCTGG TGTCATAGAA ATTGCCGATG GGAGTCGCCG TCGTAAAGCT
7561 GCTGCACTTA CCGAAAGTGA TTATCGTGTT CTGGTTGGCG AGCTGGATGA TGAGCAGATG
7621 GCTGCATTAT CCAGATTGGG TAACGATTAT CGCCCAACAA GTGCTTATGA ACGTGGTCAG
7681 CGTTATGCAA GCCGATTGCA GAATGAATTT GCTGGAAATA TTTCTGCGCT GGCTGATGCG
7741 GAAAATATTT CACGTAAGAT TATTACCCGC TGTATCAACA CCGCCAAATT GCCTAAATCA
7801 GTTGTTGCTC TTTTTTCTCA CCCCGGTGAA CTATCTGCCC GGTCAGGTGA TGCACTTCAA
7861 AAAGCCTTTA CAGATAAAGA GGAATTACTT AAGCAGCAGG CATCTAACCT TCATGAGCAG
7921 AAAAAAGCTG GGGTGATATT TGAAGCTGAA GAAGTTATCA CTCTTTTAAC TTCTGTGCTT
7981 AAAACGTCAT CTGCATCAAG AACTAGTTTA AGCTCACGAC ATCAGTTTGC TCCTGGAGCG
8041 ACAGTATTGT ATAAGGGCGA TAAAATGGTG CTTAACCTGG ACAGGTCTCG TGTTCCAACT
8101 GAGTGTATAG AGAAAATTGA GGCCATTCTT AAGGAACTTG AAAAGCCAGC ACCCTGATGC
8161 GACCACGTTT TAGTCTACGT TTATCTGTCT TTACTTAATG TCCTTTGCTA CAGGCCAGAA
8221 AGCATAACTG GCCTGAATAT TCTCTCTGGG CCCACTGTTC CACTTGTATC GTCGGACTGA
8281 TAATCAGACT GGGACCACGG TCCCACTCGT ATCGTCGGTC TGATTATTAG TCTGGGACCA
8341 CGGTCCCACT CGTATCGTCG GTCTGATTAT TAGTCTGGGA CCACGGTCCC ACTCGTATCG
8401 TCGGTCTGAT AATCAGACTG GGACCACGGT CCCACTCGTA TCGTCGGTCT GATTATTAGT
8461 CTGGGACCAT GGTCCCACTC GTATCGTCGG TCTGATTATT AGTCTGGGAC CACGGTCCCA
8521 CTCGTATCGT CGGTCTGATT ATTAGTCTGG AACCACGGTC CCACTCGTAT CGTCAGTCTG
8581 ATTATTAGTC TGGGACCACG GTCCCACTCG TATCGTCGGT CTGATTATTA GTCTGGGACC
8641 ACGATCCCAC TCGTGTTGTC GGTCTGATTA TCGGTCTGGG ACCACGGTCC CACTTGTATT
8701 GTCGATCAGA CTATCAGCGT GAGACTACGA TTCCATCAAT GCCTGTCAAG GGCAAGTATT
8761 GACATGTCGT CGTAACCTGT AGAACGGAGT AACCTCGGTG TGCGGTTGTA TGCCTGCTGT
8821 GGATTGCTGC TGTGTCCTGC TTATCCACAA CATTTTGCGC ACGGTTATGT GGACAAAATA
8881 CCTGGTTACC CAGGCCGTGC CGGCACGTTA ACCGGGCTGC ATCCGATGCA AGTGTGTCGC
8941 TGTCGACGGC CTCCTCACCC GGTCACGTGA GCTCATTTAA CCCACTCCAC AAAAAGGCTC
9001 AACAGGTTGG TGGTTCTCAC CACCAAAAGC ACCACACCCC ACGCAAAAAC AAGTTTTTGC
9061 TGATTTTTCT TTATAAATAG AGTGTTATGA AAAATTAGTT TCTCTTACTC TCTTTATGAT
9121 ATTTAAAAAA GCGGTGTCGG CGCGGCTACA ACAACGCGCC GACACCGTTT TGTAGGGGTG
9181 GTACTGACTA TTTTTATAAA AAACATTATT TTATATTAGG GGTGCTGCTA GCGGCGCGGT
9241 GTGTTTTTTT ATAGGATACC GCTAGGGGCG CTGCTAGCGG TGCGTCCCTG TTTGCATTAT
9301 GAATTAGTTA CGCTAGGGAT AACAGGGTAA TATAGAACCC GAACGACCGA GCGCAGCGGC
9361 GGCCGCGCTG ATACCGCCGC CAAAAACTAA AAAAAATATT GACACTCTAT CATTGATAGA
9421 GTATAATTAA CGGGATCCTC TATCATTGAT AGAGGGATCC CGCCAAGCTT GGGATCCCCA
9481 GCTTGTTGAT ACACTAATGC TTTTATATAG GGAAAAGGTG GTGAACTACT ATGCACCATC
9541 ATCACCACCA TAACACCATT AACATAGCGA AGAACGATTT CAGTGATATT GAACTTGCTG
9601 CGATACCATT TAACACTTTA GCCGATCATT ACGGTGAAAG ATTAGCACGC GAACAATTAG
9661 CTCTGGAACA CGAAAGTTAT GAAATGGGTG AAGCGAGGTT CAGAAAGATG TTCGAACGCC
9721 AGTTAAAGGC GGGTGAAGTG GCCGATAACG CTGCCGCCAA ACCGTTGATT ACCACCCTCC
9781 TCCCAAAGAT GATTGCGCGC ATCAACGATT GGTTCGAGGA AGTTAAAGCG AAAAGAGGTA
9841 AGCGCCCCAC AGCGTTTCAA TTTTTACAAG AGATTAAGCC GGAAGCGGTG GCTTATATAA
9901 CGATAAAGAC CACACTAGCC TGTTTAACGT CGGCTGACAA CACTACCGTG CAAGCTGTTG
9961 CTAGTGCAAT TGGTCGTGCC ATTGAAGATG AGGCCCGCTT TGGACGTATA AGAGATTTGG
10021 AGGCGAAACA CTTTAAGAAG AACGTCGAAG AGCAATTAAA CAAAAGGGTG GGCCACGTGT
10081 ATAAGAAAGC GTTTATGCAA GTTGTAGAAG CGGACATGTT GTCAAAAGGG TTATTGGGTG
10141 GTGAGGCGTG GAGTTCCTGA CACAAGGAAG ACAGTATCCA TGTGGGCGTA AGATGTATTG
10201 AGATGCTCAT CGAAAGTACC GGAATGGTGA GTCTGCACCG CCAAAATGCG GGGGTAGTCG
10261 GTCAAGATTC CGAAACCATC GAACTCGCTC CGGAATATGC AGAGGCAATC GCCACCAGAG
10321 CTGGGGCATT AGCTGGTATT TCTCCAATGT TCCAGCCCTG TGTTGTACCG CCGAAACCCT
10381 GAACCGGCAT TACCGGCGGT GGTTACTGAG CTAACGGGAG AAGACCGCTC GCCCTGGTTC
10441 GTACACACAG TAAGAAAGCC CTCATGAGAT ACGAGGATGT TTACATGCCC GAGGTTTACA
10501 AGGCCATCAA TATCGCTCAG AACACGGCGT GGAAGATAAA TAAGAAAGTA TTGGCCGTGG
10561 CAAACGTCAT TACCAAGTGG AAGCACTGTC CAGTGGAAGA TATTCCAGCT ATCGAAAGGG
10621 AAGAATTACC GATGAAGCCT GAGGACATTG ACATGAACCC AGAAGCGTTG ACAGCTTGGA
10681 AGAGAGCCGC TGCGGCAGTT TACCGCAAAG ACAAGGCCAG AAAATCTCGC AGAATCAGTT
10741 TGGAATTTAT GCTCGAACAA GCGAACAAAT TTGCCAATCA TAAGGCGATT TGGTTTCCGT
10801 ATAACATGGA CTGACGCGGT CGCGTTTACG CGGTGTCCAT GTTTAATCCT CAAGGCAACG
10861 ACATGACCAA AGGTTTGTTG ACCTTAGCAA AGGGCAAACC CATAGGCAAA GAGGGCTATT
10921 ACTGGCTCAA GATTCATGGC GCTAACTGTG CTGGGGTTGA TAAGGTTCCC TTCCCCGAAA

10981 GAATTAAGTT CATCGAAGAG AATCACGAGA ACATTATGGC TTGCGCTAAG TCGCCCCTGG
11041 AAAACACTTG ATGGGCGGAA CAGGATTCGC CTTTTTGTTT TCTTGCCTTT TGTTTCGAAT
11101 ACGCGGGGGT TCAACACCAC GGTTTGTCGT ACAACTGTAG TTTACCGTTG GCTTTCGACG
11161 GTAGCTGTAG TGGCATACAA CACTTCTCCG CCATGCTGAG AGACGAAGTA GGTGGGAGGG
11221 CCGTGAACTT ATTGCCCTCT GAAACAGTAC AGGATATTTA TGGGATCGTG GCGAAAAAGG
11281 TAAATGAAAT CCTCCAGGCT GATGCCATCA ACGGCACCGA CAACGAGGTC GTGACCGTTA
11341 CTGACGAAAA CACCGGGGAA ATATCAGAAA AAGTGAAACT CGGTACGAAG GCGCTTGCTG
11401 GCCAATGACT CGCCTATGGT GTGACACGCA GTGTTACAAA GAGAAGTGTA ATGACCTTGG
11461 CTTATGGGTC AAAGGAGTTT GGGTTTAGGC AGCAAGTCTT AGAGGACACC ATCCAACCGG
11521 CGATTGATAG TGGTAAGGGT CTAATGTTTA CCCAGCCTAA CCAAGCAGCC GGCTATATGG
11581 CCAAGTTAAT CTGGGAGAGT GTTAGTGTAA CGGTCGTTGC GGCGGTAGAA GCCATGAACT
11641 GGCTCAAGTC GGCCGCTAAG CTCCTCGCTG CTGAGGTAAA GGACAAGAAA ACCGGCGAAA
11701 TTTTGAGAAA AAGGTGCGCC GTTCACTGAG TCACACCCGA TGGGTTTCCC GTTTGGCAAG
11761 AATATAAGAA ACCGATTCAA ACCAGACTCA ATCTGATGTT CCTCGGTCAG TTTCGCCTTC
11821 AACCAACCAT CAATACCAAC AAAGATTCTG AGATCGACGC GCACAAGCAG GAATCCGGGA
11881 TCGCCCCCAA TTTTGTCCAT TCCCAGGATG GGAGTCACTT ACGCAAGACA GTGGTGTGAG
11941 CGCACGAAAA ATACGGCATT GAATCTTTTG CTTTAATCCA TGACTCCTTT GGAACCATTC
12001 CCGCCGACGC CGCTAATCTC TTTAAAGCAG TGAGAGAAAC GATGGTAGAT ACCTATGAAA
12061 GCTGTGATGT ATTAGCCGAC TTCTACGATC AGTTTGCCGA TCAATTACAT GAATCACAGT
12121 TAGATAAAAT GCCTGCCTTA CCGGCTAAGG GAAATCTTAA CTTGAGGGAT ATACTAGAAT
12181 CCGATTTTGC CTTCGCGTAG GGATAAGTAG GAATTCGGTC CCTCTCCTGG GACCTTTTTT
12241 TTTGATCACC AGAAGGTTAT GGGTTCAAGT CCCAATGTCA GCAATTACAA AAACAAAACA
12301 AATAAAAAAT AAGGGAATTA CCCCCAAGAA GACCTTTTGT GCTAACGCCA GTTTGGCAAA
12361 TCAAGTTCTG ATTTTGCAAT TATTTTGCTC CATATGAATT ACACTACTCC AAGAATTATA
12421 AGCCTCTCTA CAGCTTTATC TCAAACTTAT GTAAAATTAG AGACGTAATT CAAACACATG
12481 TCCAGGCTCG ACAAGTCAAA GGTGATTAAC TCTGCCTTAG AACTCTTAAA TGAGGTAGGT
12541 ATTGAGGGTT TAACCACCAG AAAACTCGCA CAGAAACTCG GGGTGGAACA GCCGACGCTT
12601 TATTGACACG TCAAGAACAA GAGGGCACTC CTGGACGCTC TCGCGATTGA GATGCTGGAT
12661 AGGCATCATA CTCACTTCTG TCCCCTCGAA GGCGAAAGTT GACAAGACTT TTTGAGAAAC
12721 AACGCTAAAT CCTTTAGATG TGCCTTACTT TCGCATAGAG ACGGGGCTAA GGTGCACTTA
12781 GGTACTAGGC CAACAGAAAA GCAGTATGAA ACTTTAGAAA ATCAATTGGC TTTTCTTTGC
12841 CAACAAGGTT TTAGTCTTGA AAATGCGTTG TACGCGCTCT CGGCTGTAGG GCACTTCACG
12901 TTGGGCTGCG TCTTGGAAGA CCAAGAGCAC CAAGTTGCGA AGGAAGAACG TGAAACCCCG
12961 ACCACCGACA GTATGCCTCC CTTACTGCGC CAAGCGATCG AATTATTTGA CCACCAAGGT
13021 GCTGAGCCGG CTTTTCTCTT CGGTCTGGAA CTCATTATTT GTGGTTTAGA GAAGCAATTG
13081 AAGTGCGAGT CGGGAAGTGC TTACAGCAGA GCTAGGACAA AGAATAACTA TGGTAGCACC
13141 ATCGAGGGCC TTTTAGACCT CCCCGATGAT GACGCACCGG AGGAAGCCGG CTTGGCTGCA
13201 CCCCGTTTGT CATTTCTCCC AGCCGGTCAC ACCCGCAGAC TTTCTACGGC GCCACCAACC
13261 GACGTATCGC TTGGGGATGA ACTGCACTTG GACGGTGAAG ACGTTGCTAT GGCCCACGCT
13321 GATGCTTTGG ACGATTTTGA CTTAGACATG CTTGGCGATG GTGATAGCCC TGGTCCGGGT
13381 TTCACCCCGC ACGATTCCGC CCCCTACGGG GCCTTGGATA TGGCTGATTT TGAATTTGAA
13441 CAGATGTTTA CGGATGCTCT GGGAATCGAC GAATACGGCG GGTAGCTATA ACTAGACGCG
13501 TAAACAAACT AGGAAATATT AAGGAGGAAT AGATCTATGG TGAACGTTAA GCCTGTTACC
13561 TTATATGATG TGGCTGAATA CGCGGGAGTC AGTTATCAAA CCGTATCTCG CGTAGTCAAC
13621 CAGGCAAGTC ACGTTTCTGC CAAAACTAGA GAAAAGGTTG AAGCGGCGAT GGCTGAACTG
13681 AACTATATTC CTAACAGAGT TGCTCAGCAA TTGGCCGGCA AACAGTCTTT ACTCATCGGG
13741 GTCGCTACGA GTTCCCTTGC CCTTCACGCA CCGAGCCAAA TAGTAGCAGC CATCAAGTCA
13801 AGGGCCGACC AGTTAGGAGC TTCGGTTGTG GTTAGTATGG TAGAACGATC CGGCGTGGAG
13861 GCGTGTAAAG CAGCAGTCCA TAACTTACTC GCACAGAGAG TGTCAGGTCT GATTATCAAT
13921 TATCCGCTCG ACGATCAGGA TGCTATAGCA GTAGAAGCCG CCTGCACAAA TGTACCCGCT
13981 TTATTTCTCG ACGTGTCCGA TCAAACTCCA ATTAACTCCA TTATTTTCTC CCACGAAGAT
14041 GGGACCCGCT TAGGCGTGGA ACACTTAGTT GCACTTGGAC ACCAACAAAT AGCTTTGCTA
14101 GCAGGTCCTC TTTCAAGCGT TTCCGCCCGT TTGCGCTTGG CCGGCTGACA CAAGTATTTA
14161 ACGAGAAATC AAATTCAGCC AATTGCCGAA CGCGAGGGAG ACTGGTCGGC CATGTCCGGT
14221 TTTCAGCAAA CAATGCAGAT GTTAAATGAG GGGATTGTTC CAACCGCCAT GCTCGTGGCG
14281 AATGACCAAA TGGCTCTCGG CGCCATGAGG GCGATAACGG AAAGTGGGCT AAGGGTGGGC
14341 GCGGACATTT CGGTTGTTGG GTACGATGAC ACAGAAGATT CTTCGTGCTA CATTCCCCCA
14401 CTGACGACGA TCAAACAAGA TTTTCGTTTG CTCGGGCAAA CTTCGGTCGA TCGCTTATTG
14461 CAGTTAAGTC AGGGGCAGGC GGTGAAGGGT AATCAGTTGC TGCCTGTTAG TTTAGTAAAA
14521 CGGAAAACGA CCTTAGCTCC TAACACTCAG ACCGCGTCAC CAAGAGCGCT GGCTGATAGT
14581 CTCATGCAAC TGGCGCGTCA AGTTAGTCGC CTCGAATCTG GTCAATAGCG ATAAGTAGAT
14641 GCATGAAGGA TTGTCAAAAA TATTTTTAAA GTGCTAGAAT AAAGCATGAA ACGTTTACAG
14701 TCCGCGGCAA AAGGGACAAC AGCTTAGGCC CTCCAGCGCG GACACTCAAA TTCTACTATA
14761 AGGAGGAGCG GCCGCATGAG CACTAAGAAG AAGCCATTAA CCCAGGAACA GTTGGAAGAC
14821 GCGAGAAGGT TGAAGGCTAT TTACGAAAAG AAAAAGAACG AATTGGGTCT AAGTCAGGAA
14881 TCCGTGGCCG ATAAGATGGG AATGGGACAA AGTGGGGTCG GTGCGCTATT TAACGGTATC
14941 AATGCTCTCA ACGCCTACAA CGCTGCGTTG CTCGCGAAGA TTCTAAAGGT GTCTGTTGAA
15001 GAATTTAGCC CCAGTATAGC TCGCGAAATC TACGAAATGT ACGAAGCTGT CTCCATGCAG
15061 CCTAGTCTGA GGTCCGAATA TGAATATCCG GTGTTTTCCC ACGTGCAGGC GGGGATGTTT
15121 TCGCCAAAGT TGCGCACTTT CACCAAGGGC GATGCGGAAA GATGAGTGAG TACCACTAAA
15181 AAGGCGTCAG ATAGTGCTTT TTGATTAGAA GTGGAAGGGA ACAGTATGAC GGCGCCCACG
15241 GGGTCGAAGC CATCGTTTCC CGATGGGATG CTCATTCTCG TAGATCCAGA GCAGGCAGTC
15301 GAACCAGGCG ATTTCTGTAT TGCAAGGTTG GGTGGTGACG AGTTTACTTT TAAGAAGCTA
15361 ATAAGAGATA GTGGACAAGT GTTCTTACAG CCGCTCAACC CGCAGTACCC GATGATTCCC
15421 TGCAACGAGA GCTGCTCTGT TGTGGGCAAA GTGATTGCTT CACAATGGCC GGAAGAAACA
15481 TTTGGTGAAC AAAAACTGAT TTCCGAAGAA GACCTCTAGT CTTAATTAGA GTACTACGAC

15541 GCAGAAGGAG GAGGAGAAGG CCTATGGCTC GGGTGCAGTT CAAGCAACGC GAAAGTACCG
15601 ATGCGATTTT TGTGCACTGT AGTGCCACAA AACCTAGTCA AAATGTGGGG GTACGCGAGA
15661 TTCGCCAGTG GCACAAGGAA CAAGGCTGAC TTGACGTTGG TTACCACTTC ATTATCAAGC
15721 GCGATGGAAC CGTAGAAGCT GGTCGCGATG AGATGGCAGT CGGAAGCCAT GCAAAAGGGT
15781 ATAACCACAA TTCCATCGGG GTTTGCTTGG TTGGTGGTAT TGACGACAAA GGCAAATTCG
15841 ACGCTAATTT TACCCCAGCG CAAATGCAAA GTTTACGCTC ATTATTGGTC ACGTTGCTGG
15901 CGAAGTATGA AGGGGCTGGC CTCCGCGCAC ACCATGAAGT TGCGCCAAAA GCCTGTCCAT
15961 CCTTTGACTT GAAGCGATGA TGGGAAAAGA ATGAATTAGT AACATCCGAT CGCGGCGACT
16021 ACAAGGACGA CGACGATAAG GGATAGACAT AAGTAGAAGC TTGCCGGGTC AGCTCCCGGC
16081 TTTTTTTTCA TATGGGGGGT GAAAACGTCA AACTTTCGGC TACGACCGGT ATAACTTCGT
16141 ATAGTATAGT ATATACGAAC GGTACTCGAG GCAGAAAGAA GTACCCACGG TTATATTCTC
16201 ACTAGTGAAG TTCCTATTCC GAAGTTCCTA TTCTCTAGAA AGTATAGGAA CTTCGGTACC
16261 CCGCCAGAAT TAAAAGTTAG TGAACAAGAA AACAGTGAAG CACCAGTTTC TGAACCAAAA
16321 GAAGACGAAA AAACAAAAAA AGATTAAGCA ATTTATTTGG AAAATCTTTT TTTGTTTTTT
16381 TAAGAAATAT TTATTGTTTT TTTTAAAAAA TTATGTACAA TTGCTACTAT AAGGGAAAGA
16441 AAAAAAGAAA GATATAAATT GTATAAAGTA GGGTTAGAAG CAATTAATAA TTATTATTAA
16501 TGTTATTTTT CTCTTATATA TTCAATGTAA TTTTAATTAC ATTTGCTTTT AATAAAAACA
16561 CTACTTAATA GAGAAAGGAA ATATAGATAC TATGACCGAG TATAAACCCA CCGTACGCCT
16621 GGCGACCAGG GACGACGTTC CTCGTGCGGT GAGGACACTC GCTGCCGCAT TTGCCGACTA
16681 TCCAGCCACG AGACACACCG TTGACCCAGA TCGCCACATT GAAAGAGTGA CCGAGTTGCA
16741 AGAACTTTTC CTCACCAGAG TGGGGCTTGA CATTGGAAAG GTGTGGGTAG CTGACGACGG
16801 GGCTGCCGTT GCCGTCTGGA CCACGCCTGA ATCCGTGGAA GCTGGTGCAG TTTTTGCAGA
16861 GATTGGGCCC AGAATGGCCG AGCTTAGCGG GTCCAGATTA GCAGCCCAAC AACAGATGGA
16921 GGGCTTACTC GCTCCACATA GACCAAAAGA ACCCGCGTGG TTTTTAGCCA CAGTCGGCGT
16981 TTCGCCCGAT CACCAAGGTA AGGGTCTTGG GTCCGCTGTG GTTCTTCCTG GAGTAGAGGC
17041 TGCCGAAAGG GCTGGCGTGC CCGCCTTTTT GGAAACTAGC GCTCCACGCA ACCTACCATT
17101 TTATGAACGA CTAGGCTTTA CCGTCACGGC GGATGTGGAG GTGCCAGAAG GTCCCCGAAC
17161 CTGGTGTATG ACTAGGAAAC CAGGGGCTTA GAGGTAAGTA GCTTAAGTAA TACGACTCAC
17221 TATAGGGGAA TTGTGAGCGG ATAACAATTC CCCTATAAGT AATAAACTAG ATAAGGAGGA
17281 TTAGAATGGT TTCCAAGGGC GAAGAGGATA ACATGGCTAT CATTAAGGAG TTCATGAGAT
17341 TCAAGGTTCA CATGGAAGGT AGTGTGAACG GGCACGAGTT CGAAATTGAG GGGGAAGGCG
17401 AAGGGCGTCC TTATGAAGGA ACCCAGACTG CCAAACTTAA AGTAACGAAG GGAGGCCCTT
17461 TGCCATTTGC TTGAGACATA CTGAGTCCCC AGTTTATGTA TGGTTCAAAG GCTTACGTCA
17521 AGCACCCCGC TGATATTCCA GACTACCTCA AGTTGAGTTT TCCGGAGGGT TTCAAATGAG
17581 AACGTGTTAT GAACTTCGAA GATGGTGGCG TCGTGACTGT TACCCAGGAC AGTTCTTTAC
17641 AAGATGGGGA GTTCATTTAC AAGGTTAAGT TGCGAGGCAC TAATTTTCCA TCCGATGGCC
17701 CGGTAATGCA AAAGAAAACA ATGGGGTGGG AAGCCTCGTC CGAGAGAATG TACCCAGAAG
17761 ACGGTGCGTT AAAGGGGGAA ATTAAGCAAA GATTGAAGTT AAAAGATGGG GGTCATTATG
17821 ATGCCGAGGT CAAGACCACG TATAAGGCTA AGAAGCCAGT CCAACTCCCA GGAGCGTATA
17881 ACGTCAACAT CAAACTCGAC ATCACGTCCC ATAACGAGGA CTATACTATT GTAGAACAGT
17941 ACGAACGCGC GGAAGGTCGT CACTCCACTG GTGGGATGGA TGAACTCTAC AAGTAGTGCT
18001 AACTAGCCCG GGATAACTTC GTATAGCATA CATTATACGA AGTTATGATA TCATTCCGTG
18061 ATGCTTCATA CGTATGCTAT GTCCATGGGA AGTTCCTATT CCGAAGTTCC TATTCTTCAA
18121 ATAGTATAGG AACTTCGTCG ACCTAGCATA ACCCCTTGGG GCCTCTAAAC GGGTCTTGAG
18181 GGGTTTTTTG GTTTAAACCT TCTGTGGATA ACCGTATTAC CGCCTTTGAG TGAGCTGATA
18241 CCGCTCGCCG CAGCCGAACG ACCGAGCGCA GCGAGTCAGT GAGCGAGGAA GCGGAAGAGC
18301 GCCCAATACG CAAACCGCCT CTCCCCGCGC GTTGGCCGAT TCATTAATGC ACGCTAGCAT
18361 TTAAATGGCC CAAATGGCCC TTTTACACAA TTATACGGAC TTTATCATGC AGGGGCCCAA
18421 TTGTGTAAAA GTAAAAAGGC CATATAACAG TCCTTTTACG GTACAATGTT TTTAACGACA
18481 AAAACATACC CAGGAGGACT TTTACATGAC CCAAGTACAT TTTACACTGA AAAGCGAAGA
18541 GATTCAAAGC ATTATTGAAT ATTCTGTAAA GGATGACGTT TCTAAAAATA TTTTAACAAC
18601 GGTATTTAAT CAACTAATGG AAAATCAACG AACAGAATAT ATTCAAGCAA AAGAATATGA
18661 ACGAACAGAA AACCGACAAA GTCAACGAAA TGGCTATTAT GAGCGCAGCT TTACGACACG
18721 TGTAGGCACG CTAGAATTAA AAGTACCCAG AACACGTGAT GGCCATTTTT CACCCACAGT
18781 GTTTGAACGT TATCAACGAA ACGAAAAAGC CCTCATGGCT TCAATGTTGG AAATGTATGT
18841 ATCAGGCGTT TCAACTCGTA AAGTATCAAA AATTGTGGAA GAACTTTGTG GTAAATCCGT
18901 CTCTAAGTCC TTCGTTTCTA GCTTAACAGA ACAGCTAGAA CCTATGGTTA ACGAGTGACA
18961 GAATCGTTTA TTATCAGAAA AAAATTATCC TTACTTAATG ACCGATGTAC TCTATATAAA
19021 AGTACGAGAA GAAAATCGAG TACTCTCAAA AAGCTGTCAT ATAGCGATTG GAATAACCAA
19081 AGATGGCGAC CGTGAAATTA TCGGCTTCAT GATTCAAAGT GGCGAAAGCG AAGAGACCTG
19141 GACAACATTT TTTGAATACC TAAAAGAACG CGGTTTACAA GGTACGGAAC TCGTTATTTC
19201 TGATGCGCAC AAAGGATTAG TCTCTGCCAT TAGAAAATCC TTCACCAACG TAAGTTGGCA
19261 AAGATGCCAA GTTCACTTCC TAAGAAATAT CTTTACCACC ATTCCTAAAA AAAATTCAAA
19321 ATCTTTCAGA GAAGCTGTTA AAGGAATTTT TAAGTTCACA GATATTAACT TAGCGCGTGA
19381 GGCTAAAAAT CGATTGATTC ATGATTATAT CGATCAACCA AAATATTCAA AAGCTTGCGC
19441 ATCATTGGAT GATGGATTCG AAGACGCCTT TCAATATACC GTACAAGGAA ATTCCCACAA
19501 TCGACTAAAG AGTACCAATC TAATTGAACG ACTGAATCAA GAAGTACGCA GAAGAGAAAA
19561 GATTATTCGC ATCTTCCCCA ATCAAACATC AGCCAATCGC TTAATTGGAG CCGTTCTTAT
19621 GGACCTACAT GATGAATGGA TTTATTCTTC AAGAAAATAC ATCAATTTTG ATAAGTAGAA
19681 ATGGTAAAAA CATTGTATAG CATTTTACAC AGGAGTCTGG ACTTGACTCA CTTCCTTTAT
19741 TATTTTTCAT TTTTTTGACC TCGAGGGGGG GCCCACCATA CAGCTGACGA TAAAGTCCGT
19801 ATAATTGTGT AAAAACCCAT AGCTTTGGAC ACACACTAGT ACGGATCCAC CCGCAATTAC
19861 TGTGAGTTAG CTCACTCATT AGGCACCCCA GGCTTTACAC TTTATACTTC CGGCTC
//

Complete sequence Construct 4

LOCUS Construct_4_pJ251_Reporter_T7Pol_LacI_T7Lys 11507 bp DNA circular UNA 30-Jul-2010
DEFINITION
FEATURES Location/Qualifiers
Misc._recombina 219..220
/label="Lox P1"
Misc._recombina 263..310
/label="FRT wt"
Promoter 317..647
/label="Spiralin Promoter"
CDS 648..1244
/label=Puromycin
Misc._signal 1245..1257
/label="Stop 7"
Misc._recombina 1445..1470
/note="IS256 site according to Kostas annotation in the
pmTjCoor file"
/label="IS256 Site 1"
/label="IS256 Site 1"
CDS 1570..2742
/gene="tnp"
/codon_start=1
/transl_table=11
/product="transposase"
/protein_id="ACR82773.1"
/db_xref="GI:239583490"
/translation="MTQVHFTLKSEEIQSIIEYSVKDDVSKNILTTVFNQLMENQRTE
YIQAKEYERTENRQSQRNGYYERSFTTRVGTLELKVPRTRDGHFSPTVFERYQRNEKA
LMASMLEMYVSGVSTRKVSKIVEELCGKSVSKSFVSSLTEQLEPMVNEWQNRLLSEKN
YPYLMTDVLYIKVREENRVLSKSCHIAIGITKDGDREIIGFMIQSGESEETWTTFFEY
LKERGLQGTELVISDAHKGLVSAIRKSFTNVSWQRCQVHFLRNIFTTIPKKNSKSFRE
AVKGIFKFTDINLAREAKNRLIHDYIDQPKYSKACASLDDGFEDAFQYTVQGNSHNRL
KSTNLIERLNQEVRRREKIIRIFPNQTSANRLIGAVLMDLHDEWIYSSRKYINFDK"
Misc._recombina 2853..2878
/note="IS 256 Sitye according to Kostas annotation in
pTmJcor"
/label="IS256 Site 2"
/label="IS256 Site 2"
Terminator complement(3012..3059)
/label="T7 Terminator"
/label="T7 Terminator"
Misc._recombina complement(3066..3113)
/label="FRT wt"
/label="FRT wt"
Misc._recombina complement(3156..3189)
/label="Lox P wt"
/label="Lox P wt"
Misc._signal complement(3196..3208)
/label="Stop 8"
/label="Stop 8"
CDS complement(3209..3916)
/label=mCherry
RBS complement(3917..3948)
/label="RBS mCherry"
Protein_binding complement(3953..3973)
/label=O1
Protein_binding complement(3975..3994)
/label="T7 Pol binding site"
Protein_binding complement(4039..4058)
/label=Osym
Misc._feature 4400..4513
/label="rpn txn terminator"
Replication_ori 4576..5402
/label="p15a Ori"
CDS 5416..6210
/label=Kanamycin
Promoter 6741..6824
/product="Ldh 84 Bp"
/label="Ldh 84 Bp"
/label="Ldh Promoter"
CDS 6825..6854
/product="Ldh CDS"
/label="Ldh CDS"
CDS 6855..9561

/product="T7 Polymerase"
/label="T7 Polymerase"
Promoter 9579..9621
/product="Synthetic Promotor Eva"
/label="Synthetic Promotor Eva"
/label="Synthetic Promoter"
CDS 9622..9651
/product="GroEL CDS"
/label="GroEL CDS"
CDS 9652..10754
/product="Lac Repressor"
/label="Lac Repressor"
Promoter 10757..10981
/product="AckA Promotor"
/label="AckA Promotor"
/label="AckA Promoter"
CDS 10982..11011
/product="AckA CDS"
/label="AckA CDS"
CDS 11012..11507
/product="T7 Lysozyme"
/label="T7 Lysozyme"
ORIGIN
1 TGATCCCCCT GGCCAGTGCA CGTCTGCTGT CAGATAAAGT CTCCCGTGAA CTTTACCCGG
61 TGGTGCATAT CGGGGATGAA AGCTGGCGCA TGATGACCAC CGATATGGCC AGTGTGCCGG
121 TCTCCGTTAT CGGGGAAGAA GTGGCTGATC TCAGCCACCG CGAAAATGAC ATCAAAAACG
181 CCATTAACCT GATGTTCTGG GGAATATAAC CCGGGAAGTA CTCGAGGCAG AAAGAAGTAC
241 CCACGGTTAT ATTCTCACTA GTGAAGTTCC TATTCCGAAG TTCCTATTCT CTAGAAAGTA
301 TAGGAACTTC GGTACCCCGC CAGAATTAAA AGTTAGTGAA CAAGAAAACA GTGAAGCACC
361 AGTTTCTGAA CCAAAAGAAG ACGAAAAAAC AAAAAAAGAT TAAGCAATTT ATTTGGAAAA
421 TCTTTTTTTG TTTTTTTAAG AAATATTTAT TGTTTTTTTT AAAAAATTAT GTACAATTGC
481 TACTATAAGG GAAAGAAAAA AAGAAAGATA TAAATTGTAT AAAGTAGGGT TAGAAGCAAT
541 TAATAATTAT TATTAATGTT ATTTTTCTCT TATATATTCA ATGTAATTTT AATTACATTT
601 GCTTTTAATA AAAACACTAC TTAATAGAGA AAGGAAATAT AGATACTATG ACCGAGTATA
661 AACCCACCGT ACGCCTGGCG ACCAGGGACG ACGTTCCTCG TGCGGTGAGG ACACTCGCTG
721 CCGCATTTGC CGACTATCCA GCCACGAGAC ACACCGTTGA CCCAGATCGC CACATTGAAA
781 GAGTGACCGA GTTGCAAGAA CTTTTCCTCA CCAGAGTGGG GCTTGACATT GGAAAGGTGT
841 GGGTAGCTGA CGACGGGGCT GCCGTTGCCG TCTGGACCAC GCCTGAATCC GTGGAAGCTG
901 GTGCAGTTTT TGCAGAGATT GGGCCCAGAA TGGCCGAGCT TAGCGGGTCC AGATTAGCAG
961 CCCAACAACA GATGGAGGGC TTACTCGCTC CACATAGACC AAAAGAACCC GCGTGGTTTT
1021 TAGCCACAGT CGGCGTTTCG CCCGATCACC AAGGTAAGGG TCTTGGGTCC GCTGTGGTTC
1081 TTCCTGGAGT AGAGGCTGCC GAAAGGGCTG GCGTGCCCGC CTTTTTGGAA ACTAGCGCTC
1141 CACGCAACCT ACCATTTTAT GAACGACTAG GCTTTACCGT CACGGCGGAT GTGGAGGTGC
1201 CAGAAGGTCC CCGAACCTGG TGTATGACTA GGAAACCAGG GGCTTAGAGG TAAGTAGCTT
1261 AAGTTCTGTG GATAACCGTA TTACCGCCTT TGAGTGAGCT GATACCGCTC GCCGCAGCCG
1321 AACGACCGAG CGCAGCGAGT CAGTGAGCGA GGAAGCGGAA GAGCGCCCAA TACGCAAACC
1381 GCCTCTCCCC GCGCGTTGGC CGATTCATTA ATGCACGCTA GCATTTAAAT GGCCCAAATG
1441 GCCCTTTTAC ACAATTATAC GGACTTTATC ATGCAGGGGC CCAATTGTGT AAAAGTAAAA
1501 AGGCCATATA ACAGTCCTTT TACGGTACAA TGTTTTTAAC GACAAAAACA TACCCAGGAG
1561 GACTTTTACA TGACCCAAGT ACATTTTACA CTGAAAAGCG AAGAGATTCA AAGCATTATT
1621 GAATATTCTG TAAAGGATGA CGTTTCTAAA AATATTTTAA CAACGGTATT TAATCAACTA
1681 ATGGAAAATC AACGAACAGA ATATATTCAA GCAAAAGAAT ATGAACGAAC AGAAAACCGA
1741 CAAAGTCAAC GAAATGGCTA TTATGAGCGC AGCTTTACGA CACGTGTAGG CACGCTAGAA
1801 TTAAAAGTAC CCAGAACACG TGATGGCCAT TTTTCACCCA CAGTGTTTGA ACGTTATCAA
1861 CGAAACGAAA AAGCCCTCAT GGCTTCAATG TTGGAAATGT ATGTATCAGG CGTTTCAACT
1921 CGTAAAGTAT CAAAAATTGT GGAAGAACTT TGTGGTAAAT CCGTCTCTAA GTCCTTCGTT
1981 TCTAGCTTAA CAGAACAGCT AGAACCTATG GTTAACGAGT GACAGAATCG TTTATTATCA
2041 GAAAAAAATT ATCCTTACTT AATGACCGAT GTACTCTATA TAAAAGTACG AGAAGAAAAT
2101 CGAGTACTCT CAAAAAGCTG TCATATAGCG ATTGGAATAA CCAAAGATGG CGACCGTGAA
2161 ATTATCGGCT TCATGATTCA AAGTGGCGAA AGCGAAGAGA CCTGGACAAC ATTTTTTGAA
2221 TACCTAAAAG AACGCGGTTT ACAAGGTACG GAACTCGTTA TTTCTGATGC GCACAAAGGA
2281 TTAGTCTCTG CCATTAGAAA ATCCTTCACC AACGTAAGTT GGCAAAGATG CCAAGTTCAC
2341 TTCCTAAGAA ATATCTTTAC CACCATTCCT AAAAAAAATT CAAAATCTTT CAGAGAAGCT
2401 GTTAAAGGAA TTTTTAAGTT CACAGATATT AACTTAGCGC GTGAGGCTAA AAATCGATTG
2461 ATTCATGATT ATATCGATCA ACCAAAATAT TCAAAAGCTT GCGCATCATT GGATGATGGA
2521 TTCGAAGACG CCTTTCAATA TACCGTACAA GGAAATTCCC ACAATCGACT AAAGAGTACC
2581 AATCTAATTG AACGACTGAA TCAAGAAGTA CGCAGAAGAG AAAAGATTAT TCGCATCTTC
2641 CCCAATCAAA CATCAGCCAA TCGCTTAATT GGAGCCGTTC TTATGGACCT ACATGATGAA
2701 TGGATTTATT CTTCAAGAAA ATACATCAAT TTTGATAAGT AGAAATGGTA AAAACATTGT
2761 ATAGCATTTT ACACAGGAGT CTGGACTTGA CTCACTTCCT TTATTATTTT TCATTTTTTT
2821 GACCTCGAGG GGGGGCCCAC CATACAGCTG ACGATAAAGT CCGTATAATT GTGTAAAAAC
2881 CCATAGCTTT GGACACACAC TAGTACGGAT CCACCCGCAA TTACTGTGAG TTAGCTCACT
2941 CATTAGGCAC CCCAGGCTTT ACACTTTATA CTTCCGGCTC GAACTTCCCT TCATTTTCCT
3001 AGGGTTTAAA CCAAAAAACC CCTCAAGACC CGTTTAGAGG CCCCAAGGGG TTATGCTAGG
3061 TCGACGAAGT TCCTATACTA TTTGAAGAAT AGGAACTTCG GAATAGGAAC TTCCCATGGA
3121 CATAGCATAC GTATGAAGCA TCACGGAATG ATATCATAAC TTCGTATAAT GTATGCTATA

3181 CGAAGTTATC CCGGGCTAGT TAGCACTACT TGTAGAGTTC ATCCATCCCA CCAGTGGAGT
3241 GACGACCTTC CGCGCGTTCG TACTGTTCTA CAATAGTATA GTCCTCGTTA TGGGACGTGA
3301 TGTCGAGTTT GATGTTGACG TTATACGCTC CTGGGAGTTG GACTGGCTTC TTAGCCTTAT
3361 ACGTGGTCTT GACCTCGGCA TCATAATGAC CCCCATCTTT TAACTTCAAT CTTTGCTTAA
3421 TTTCCCCCTT TAACGCACCG TCTTCTGGGT ACATTCTCTC GGACGAGGCT TCCCACCCCA
3481 TTGTTTTCTT TTGCATTACC GGGCCATCGG ATGGAAAATT AGTGCCTCGC AACTTAACCT
3541 TGTAAATGAA CTCCCCATCT TGTAAAGAAC TGTCCTGGGT AACAGTCACG ACGCCACCAT
3601 CTTCGAAGTT CATAACACGT TCTCATTTGA AACCCTCCGG AAAACTCAAC TTGAGGTAGT
3661 CTGGAATATC AGCGGGGTGC TTGACGTAAG CCTTTGAACC ATACATAAAC TGGGGACTCA
3721 GTATGTCTCA AGCAAATGGC AAAGGGCCTC CCTTCGTTAC TTTAAGTTTG GCAGTCTGGG
3781 TTCCTTCATA AGGACGCCCT TCGCCTTCCC CCTCAATTTC GAACTCGTGC CCGTTCACAC
3841 TACCTTCCAT GTGAACCTTG AATCTCATGA ACTCCTTAAT GATAGCCATG TTATCCTCTT
3901 CGCCCTTGGA AACCATTCTA ATCCTCCTTA TCTAGTTTAT TACTTATAGG GGAATTGTTA
3961 TCCGCTCACA ATTCCCCTAT AGTGAGTCGT ATTACTTAAG CTACTTACCT CTAAGCCCCT
4021 GGTTTCCTAG TCATACACaa ttgtgagcgc tcacaattTC TATAGGGAGT CGAATAAGGG
4081 CGACACAAAA GGTATTCTAA ATGCATAATA AATACTGATA ACATCTTATA GTTTGTATTA
4141 TATTTTGTAT TATCGTTGAC ATGTATAATT TTGATATCAA AAACTGATTT TCCCTTTATT
4201 ATTTTCGAGA TTTATTTTCT TAATTCTCTT TAACAAACTA GAAATATTGT ATATACAAAA
4261 AATCATAAAT AATAGATGAA TAGTTTAATT ATAGGTGTTC ATCAATCGAA AAAGCAACGT
4321 ATCTTATTTA AAGTGCGTTG CTTTTTTCTC ATTTATAAGG TTAAATAATT CTCATATATC
4381 AAGCAAAGTG ACAGGCGCCC TTAAATATTC TGACAAATGC TCTTTCCCTA AACTCCCCCC
4441 ATAAAAAAAC CCGCCGAAGC GGGTTTTTAC GTTATTTGCG GATTAACGAT TACTCGTTAT
4501 CAGAACCGCC CAGGATGCCT GGCAGTTCCC TACTCTCGCC GCTGCGCTCG GTCGTTCGGC
4561 TGCGGGACCT CAGCGCTAGC GGAGTGTATA CTGGCTTACT ATGTTGGCAC TGATGAGGGT
4621 GTCAGTGAAG TGCTTCATGT GGCAGGAGAA AAAAGGCTGC ACCGGTGCGT CAGCAGAATA
4681 TGTGATACAG GATATATTCC GCTTCCTCGC TCACTGACTC GCTACGCTCG GTCGTTCGAC
4741 TGCGGCGAGC GGAAATGGCT TACGAACGGG GCGGAGATTT CCTGGAAGAT GCCAGGAAGA
4801 TACTTAACAG GGAAGTGAGA GGGCCGCGGC AAAGCCGTTT TTCCATAGGC TCCGCCCCCC
4861 TGACAAGCAT CACGAAATCT GACGCTCAAA TCAGTGGTGG CGAAACCCGA CAGGACTATA
4921 AAGATACCAG GCGTTTCCCC CTGGCGGCTC CCTCGTGCGC TCTCCTGTTC CTGCCTTTCG
4981 GTTTACCGGT GTCATTCCGC TGTTATGGCC GCGTTTGTCT CATTCCACGC CTGACACTCA
5041 GTTCCGGGTA GGCAGTTCGC TCCAAGCTGG ACTGTATGCA CGAACCCCCC GTTCAGTCCG
5101 ACCGCTGCGC CTTATCCGGT AACTATCGTC TTGAGTCCAA CCCGGAAAGA CATGCAAAAG
5161 CACCACTGGC AGCAGCCACT GGTAATTGAT TTAGAGGAGT TAGTCTTGAA GTCATGCGCC
5221 GGTTAAGGCT AAACTGAAAG GACAAGTTTT GGTGACTGCG CTCCTCCAAG CCAGTTACCT
5281 CGGTTCAAAG AGTTGGTAGC TCAGAGAACC TTCGAAAAAC CGCCCTGCAA GGCGGTTTTT
5341 TCGTTTTCAG AGCAAGAGAT TACGCGCAGA CCAAAACGAT CTCAAGAAGA TCATCTTATT
5401 AAGCTTTTAG AAAAACTCAT CGAGCATCAA ATGAAACTGC AATTTATTCA TATCAGGATT
5461 ATCAATACCA TATTTTTGAA AAAGCCGTTT CTGTAATGAA GGAGAAAACT CACCGAGGCA
5521 GTTCCATAGG ATGGCAAGAT CCTGGTATCG GTCTGCGATT CCGACTCGTC CAACATCAAT
5581 ACAACCTATT AATTTCCCCT CGTCAAAAAT AAGGTTATCA AGTGAGAAAT CACCATGAGT
5641 GACGACTGAA TCCGGTGAGA ATGGCAAAAG TTTATGCATT TCTTTCCAGA CTTGTTCAAC
5701 AGGCCAGCCA TTACGCTCGT CATCAAAATC ACTCGCATCA ACCAAACCGT TATTCATTCG
5761 TGATTGCGCC TGAGCGAGGC GAAATACGCG ATCGCTGTTA AAAGGACAAT TACAAACAGG
5821 AATCGAGTGC AACCGGCGCA GGAACACTGC CAGCGCATCA ACAATATTTT CACCTGAATC
5881 AGGATATTCT TCTAATACCT GGAACGCTGT TTTTCCGGGG ATCGCAGTGG TGAGTAACCA
5941 TGCATCATCA GGAGTACGGA TAAAATGCTT GATGGTCGGA AGTGGCATAA ATTCCGTCAG
6001 CCAGTTTAGT CTGACCATCT CATCTGTAAC ATCATTGGCA ACGCTACCTT TGCCATGTTT
6061 CAGAAACAAC TCTGGCGCAT CGGGCTTCCC ATACAAGCGA TAGATTGTCG CACCTGATTG
6121 CCCGACATTA TCGCGAGCCC ATTTATACCC ATATAAATCA GCATCCATGT TGGAATTTAA
6181 TCGCGGCCTC GACGTTTCCC GTTGAATATG GCTCATATTC TTCCTTTTTC AATATTATTG
6241 AAGCATTTAT CAGGGTTATT GTCTCATGAG CGGATACATA TTTGAATGTA TTTAGAAAAA
6301 TAAACAAATA GGGGTCAGTG TTACAACCAA TTAACCAATT CTGAACATTA TCGCGAGCCC
6361 ATTTATACCT GAATATGGCT CATAACACCC CTTGTTTGCC TGGCGGCAGT AGCGCGGTGG
6421 TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT GGTAGTGTGG
6481 GGACTCCCCA TGCGAGAGTA GGGAACTGCC AGGCATCAAA TAAAACGAAA GGCTCAGTCG
6541 AAAGACTGGG CCTTTCGCCC GGGCTAATTA GGGGGTGTCG CCCTTATTCG ACTCTATAGG
6601 CGGTATCGAT GATATCCCCG GGATGCAGTT TAAGGTTTAC ACCTATAAAA GAGAGAGCCG
6661 TTATCGTCTG TTTGTGGATG TACAGAGTGA TATTATGACA CGCCCGGGCG ACGGATGGCG
6721 ACGGTATCGA TAAGCTTGAT TTTAGATCAA ATCTCAATTG GGAACAGTAA ATGATTGGTG
6781 GCTTAAACCA AATTAAAATG AAGGCAAGCT TGAATAAATT AGTTATGAAG AGTCTTAAAG
6841 TAGCACTCAT TGGTAACACC ATTAACATAG CGAAGAACGA TTTCAGTGAT ATTGAACTTG
6901 CTGCGATACC ATTTAACACT TTAGCCGATC ATTACGGTGA AAGATTAGCA CGCGAACAAT
6961 TAGCTCTGGA ACACGAAAGT TATGAAATGG GTGAAGCGAG GTTCAGAAAG ATGTTCGAAC
7021 GCCAGTTAAA GGCGGGTGAA GTGGCCGATA ACGCTGCCGC CAAACCGTTG ATTACCACCC
7081 TCCTCCCAAA GATGATTGCG CGCATCAACG ATTGGTTCGA GGAAGTTAAA GCGAAAAGAG
7141 GTAAGCGCCC CACAGCGTTT CAATTTTTAC AAGAGATTAA GCCGGAAGCG GTGGCTTATA
7201 TAACGATAAA GACCACACTA GCCTGTTTAA CGTCGGCTGA CAACACTACC GTGCAAGCTG
7261 TTGCTAGTGC AATTGGTCGT GCCATTGAAG ATGAGGCCCG CTTTGGACGT ATAAGAGATT
7321 TGGAGGCGAA ACACTTTAAG AAGAACGTCG AAGAGCAATT AAACAAAAGG GTGGGCCACG
7381 TGTATAAGAA AGCGTTTATG CAAGTTGTAG AAGCGGACAT GTTGTCAAAA GGGTTATTGG
7441 GTGGTGAGGC GTGGAGTTCC TGACACAAGG AAGACAGTAT CCATGTGGGC GTAAGATGTA
7501 TTGAGATGCT CATCGAAAGT ACCGGAATGG TGAGTCTGCA CCGCCAAAAT GCGGGGGTAG
7561 TCGGTCAAGA TTCCGAAACC ATCGAACTCG CTCCGGAATA TGCAGAGGCA ATCGCCACCA
7621 GAGCTGGGGC ATTAGCTGGT ATTTCTCCAA TGTTCCAGCC CTGTGTTGTA CCGCCGAAAC
7681 CCTGAACCGG CATTACCGGC GGTGGTTACT GAGCTAACGG GAGAAGACCG CTCGCCCTGG

7741 TTCGTACACA CAGTAAGAAA GCCCTCATGA GATACGAGGA TGTTTACATG CCCGAGGTTT
7801 ACAAGGCCAT CAATATCGCT CAGAACACGG CGTGGAAGAT AAATAAGAAA GTATTGGCCG
7861 TGGCAAACGT CATTACCAAG TGGAAGCACT GTCCAGTGGA AGATATTCCA GCTATCGAAA
7921 GGGAAGAATT ACCGATGAAG CCTGAGGACA TTGACATGAA CCCAGAAGCG TTGACAGCTT
7981 GGAAGAGAGC CGCTGCGGCA GTTTACCGCA AAGACAAGGC CAGAAAATCT CGCAGAATCA
8041 GTTTGGAATT TATGCTCGAA CAAGCGAACA AATTTGCCAA TCATAAGGCG ATTTGGTTTC
8101 CGTATAACAT GGACTGACGC GGTCGCGTTT ACGCGGTGTC CATGTTTAAT CCTCAAGGCA
8161 ACGACATGAC CAAAGGTTTG TTGACCTTAG CAAAGGGCAA ACCCATAGGC AAAGAGGGCT
8221 ATTACTGGCT CAAGATTCAT GGCGCTAACT GTGCTGGGGT TGATAAGGTT CCCTTCCCCG
8281 AAAGAATTAA GTTCATCGAA GAGAATCACG AGAACATTAT GGCTTGCGCT AAGTCGCCCC
8341 TGGAAAACAC TTGATGGGCG GAACAGGATT CGCCTTTTTG TTTTCTTGCC TTTTGTTTCG
8401 AATACGCGGG GGTTCAACAC CACGGTTTGT CGTACAACTG TAGTTTACCG TTGGCTTTCG
8461 ACGGTAGCTG TAGTGGCATA CAACACTTCT CCGCCATGCT GAGAGACGAA GTAGGTGGGA
8521 GGGCCGTGAA CTTATTGCCC TCTGAAACAG TACAGGATAT TTATGGGATC GTGGCGAAAA
8581 AGGTAAATGA AATCCTCCAG GCTGATGCCA TCAACGGCAC CGACAACGAG GTCGTGACCG
8641 TTACTGACGA AAACACCGGG GAAATATCAG AAAAAGTGAA ACTCGGTACG AAGGCGCTTG
8701 CTGGCCAATG ACTCGCCTAT GGTGTGACAC GCAGTGTTAC AAAGAGAAGT GTAATGACCT
8761 TGGCTTATGG GTCAAAGGAG TTTGGGTTTA GGCAGCAAGT CTTAGAGGAC ACCATCCAAC
8821 CGGCGATTGA TAGTGGTAAG GGTCTAATGT TTACCCAGCC TAACCAAGCA GCCGGCTATA
8881 TGGCCAAGTT AATCTGGGAG AGTGTTAGTG TAACGGTCGT TGCGGCGGTA GAAGCCATGA
8941 ACTGGCTCAA GTCGGCCGCT AAGCTCCTCG CTGCTGAGGT AAAGGACAAG AAAACCGGCG
9001 AAATTTTGAG AAAAAGGTGC GCCGTTCACT GAGTCACACC CGATGGGTTT CCCGTTTGGC
9061 AAGAATATAA GAAACCGATT CAAACCAGAC TCAATCTGAT GTTCCTCGGT CAGTTTCGCC
9121 TTCAACCAAC CATCAATACC AACAAAGATT CTGAGATCGA CGCGCACAAG CAGGAATCCG
9181 GGATCGCCCC CAATTTTGTC CATTCCCAGG ATGGGAGTCA CTTACGCAAG ACAGTGGTGT
9241 GAGCGCACGA AAAATACGGC ATTGAATCTT TTGCTTTAAT CCATGACTCC TTTGGAACCA
9301 TTCCCGCCGA CGCCGCTAAT CTCTTTAAAG CAGTGAGAGA AACGATGGTA GATACCTATG
9361 AAAGCTGTGA TGTATTAGCC GACTTCTACG ATCAGTTTGC CGATCAATTA CATGAATCAC
9421 AGTTAGATAA AATGCCTGCC TTACCGGCTA AGGGAAATCT TAACTTGAGG GATATACTAG
9481 AATCCGATTT TGCCTTCGCG TAGGGATAAG TAGGAATTCG GTCCCTCTCC TGGGACCTTT
9541 TTTTTTGATC ACCAGAAGGT TATCGAATTC CTGCAGATGA ATTGATCGCC ATAGGTTAAA
9601 GTAGTATAAT ATAAACTAAT TATGGCAAAG GAATTAGTAT TTGGCAAGAA TGTGAACGTT
9661 AAGCCTGTTA CCTTATATGA TGTGGCTGAA TACGCGGGAG TCAGTTATCA AACCGTATCT
9721 CGCGTAGTCA ACCAGGCAAG TCACGTTTCT GCCAAAACTA GAGAAAAGGT TGAAGCGGCG
9781 ATGGCTGAAC TGAACTATAT TCCTAACAGA GTTGCTCAGC AATTGGCCGG CAAACAGTCT
9841 TTACTCATCG GGGTCGCTAC GAGTTCCCTT GCCCTTCACG CACCGAGCCA AATAGTAGCA
9901 GCCATCAAGT CAAGGGCCGA CCAGTTAGGA GCTTCGGTTG TGGTTAGTAT GGTAGAACGA
9961 TCCGGCGTGG AGGCGTGTAA AGCAGCAGTC CATAACTTAC TCGCACAGAG AGTGTCAGGT
10021 CTGATTATCA ATTATCCGCT CGACGATCAG GATGCTATAG CAGTAGAAGC CGCCTGCACA
10081 AATGTACCCG CTTTATTTCT CGACGTGTCC GATCAAACTC CAATTAACTC CATTATTTTC
10141 TCCCACGAAG ATGGGACCCG CTTAGGCGTG GAACACTTAG TTGCACTTGG ACACCAACAA
10201 ATAGCTTTGC TAGCAGGTCC TCTTTCAAGC GTTTCCGCCC GTTTGCGCTT GGCCGGCTGA
10261 CACAAGTATT TAACGAGAAA TCAAATTCAG CCAATTGCCG AACGCGAGGG AGACTGGTCG
10321 GCCATGTCCG GTTTTCAGCA AACAATGCAG ATGTTAAATG AGGGGATTGT TCCAACCGCC
10381 ATGCTCGTGG CGAATGACCA AATGGCTCTC GGCGCCATGA GGGCGATAAC GGAAAGTGGG
10441 CTAAGGGTGG GCGCGGACAT TTCGGTTGTT GGGTACGATG ACACAGAAGA TTCTTCGTGC
10501 TACATTCCCC CACTGACGAC GATCAAACAA GATTTTCGTT TGCTCGGGCA AACTTCGGTC
10561 GATCGCTTAT TGCAGTTAAG TCAGGGGCAG GCGGTGAAGG GTAATCAGTT GCTGCCTGTT
10621 AGTTTAGTAA AACGGAAAAC GACCTTAGCT CCTAACACTC AGACCGCGTC ACCAAGAGCG
10681 CTGGCTGATA GTCTCATGCA ACTGGCGCGT CAAGTTAGTC GCCTCGAATC TGGTCAATAG
10741 CGATAAGTAG ATGCATCGTT AATAATGATG ATTGAAGCTA GTACAAAAAA GACAATGATT
10801 AAGGTCTTAT CATTACTTAC ACCAATTACC CCCATGAGGG AAGAGATAAA GTAGAGAATA
10861 ATACCTACAG CTGTCGAAAC AGCAGCCACA ATATATGAAG TTTTAGCACT CATTGACAGT
10921 TATTAGAGAA AAAATGAGGA CTTATTACTA ATATTTTATT ATTGTTACCT ATTAGATAAA
10981 AATGAACGAC AACAAAATTT TAGTAGTCAA TGCTCGGGTG CAGTTCAAGC AACGCGAAAG
11041 TACCGATGCG ATTTTTGTGC ACTGTAGTGC CACAAAACCT AGTCAAAATG TGGGGGTACG
11101 CGAGATTCGC CAGTGGCACA AGGAACAAGG CTGACTTGAC GTTGGTTACC ACTTCATTAT
11161 CAAGCGCGAT GGAACCGTAG AAGCTGGTCG CGATGAGATG GCAGTCGGAA GCCATGCAAA
11221 AGGGTATAAC CACAATTCCA TCGGGGTTTG CTTGGTTGGT GGTATTGACG ACAAAGGCAA
11281 ATTCGACGCT AATTTTACCC CAGCGCAAAT GCAAAGTTTA CGCTCATTAT TGGTCACGTT
11341 GCTGGCGAAG TATGAAGGGG CTGGCCTCCG CGCACACCAT GAAGTTGCGC CAAAAGCCTG
11401 TCCATCCTTT GACTTGAAGC GATGATGGGA AAAGAATGAA TTAGTAACAT CCGATCGCGG
11461 CGACTACAAG GACGACGACG ATAAGGGATA GACATAAGTA GAAGCTT
//

Sequences “Engineering Mycoplasma pneumoniae as therapeutic vector for lung diseases”

>EfTu_Mpn213 *M. pneumoniae* M129, complete genome .

AAGCTTGATATCGAATTCGAAGACCTTTTGTGCTAACGCCAGTTTGGCAAATCAAGTTCTGATTTTGCAATTATTTTGC

TCCATATGAATTACACTACTCCAAGAATTATAAGCCTCTCTACAGCTTTATCTCAAACTTATGTAAAATTAGAGACGT
AATTCAAACACATGAAGCTTAGTGCTATTATCTCCCTATCAGTCGCTGGTACTGTGGGAACAACTGCGGTGGTAGTAC
CTACAACTATAACGCTTGTAAATAAGACCCACCAAGGcAGCCTGCAGCCCGGGGGGCAAGA
>EfTu_Mpn459 *M. pneumoniae* M129, complete genome
AAGCTTGATATCGAATTCGAAGACCTTTTGTGCTAACGCCAGTTTGGCAAATCAAGTTCTGATTTTGCAATTATTTTGC
TCCATATGAATTACACTACTCCAAGAATTATAAGCCTCTCTACAGCTTTATCTCAAACTTATGTAAAATTAGAGACGT
AATTCAAACACATGGCTTTCATGCCATGTTTTTCATATAGCAATAAACATCATATCAACAACATTTATGAATGATCCA
TGAAATTTCAGCGAAAATACTGGGGTTTGTTATCAACCTTAGGGGTAAGTTCAGCAGTAGCACTCAGTGCCTGCGCTG
CCCAAGCACGAGATGTTTATGTTACTGGcAGCCTGCAGCCCGGGGGGCAAGA
>EfTu_Mpn142 *M. pneumoniae* M129, complete genome
.AAGCTTGATATCGAATTCGAAGACCTTTTGTGCTAACGCCAGTTTGGCAAATCAAGTTCTGATTTTGCAATTATTTTG
CTCCATATGAATTACACTACTCCAAGAATTATAAGCCTCTCTACAGCTTTATCTCAAACTTATGTAAAATTAGAGACG
TAATTCAAACACATGAAATCGAAGCTAAAGTTAAAACGTTATTTACTGTTTTTACCACTTTTACCGCTAGGGACGTTG
TCACTAGCCAACACCTACCTCCTCCAAGGcAGCCTGCAGCCCGGGGGGCAAGA
>EfTu_Mpn645 *M. pneumoniae* M129, complete genome
.AAGCTTGATATCGAATTCGAAGACCTTTTGTGCTAACGCCAGTTTGGCAAATCAAGTTCTGATTTTGCAATTATTTTG
CTCCATATGAATTACACTACTCCAAGAATTATAAGCCTCTCTACAGCTTTATCTCAAACTTATGTAAAATTAGAGACG
TAATTCAAACACATGAAACTGAAACTTAAATTTCTATTAATTTCTCTTTTAGGTTCTAGTTTGTTGTTAAGCGCTTGTT
CTTCAGCAGCTACTCAAGTAATTTCTGGcAGCCTGCAGCCCGGGGGGCAAGA
>EfTu_Mpn332_Neg_Control *M. pneumoniae* M129, complete genome
.AAGCTTGATATCGAATTCGAAGACCTTTTGTGCTAACGCCAGTTTGGCAAATCAAGTTCTGATTTTGCAATTATTTTG
CTCCATATGAATTACACTACTCCAAGAATTATAAGCCTCTCTACAGCTTTATCTCAAACTTATGTAAAATTAGAGACG
TAATTCAAACACATGCCAGCTGTAAAAAAACCACAAATTCTCGTTGTGCGTAACCAAGTTATCTTTCCGTACAACGGC
TTTGAACTTGATGTCGGTCGTGAACGCTCGAAAAAACTAATTAAAGCACTCAAAAACCTCAAAACCAAACGTTTGGT
GTTAGTCGGcAGCCTGCAGCCCGGGGGGCAAGA
>EfTu_Mpn142OptGAAGACCTTTTGTGCTAACGCCAGTTTGGCAAATCAAGTTCTGATTTTGCAATTATTTTGCTCCATA
TGAATTACACTACTCCAAGAATTATAAGCCTCTCTACAGCTTTATCTCAAACTTATGTAAAATTAGAGACGTAATTCA
AACACATGAAGTCCAAGTTGAAACTCAAACGCTATTTACTCTTTCTCCCCTTGTTACCACTCGGTACCTTGAGTTTAGC
TAACACTTACTTGTTACAA
>Fragment_Mpn200_Sig_F_-_Mpn200_Sig_R *M. pneumoniae* M129, complete genome
.AAGCTTGATATCGAATTCGACAGTAGTTTAAACTGATTCTTTACCTCTAAATAAAAATACTAAAGCCCATTAAAGCC
GAACCACTAATTTTGCCCTAGCTTGTAAAGCAGTTAAATATTTTTTAAGATATCTTCACCTCTTTTTTGCGTTTAGTGA
TGAAATTTAAGTATGGTGCCATTGTTTTCAGTGGTCTTTTAGGAGTCTCTGCCATTTTAGCTGCTTGTGGTACACGCGG
TAAAGGCAGCCTGCAGGCACAC
>Fragment_Mpn400_Sig_R_-_Mpn400_Sig_F *M. pneumoniae* M129, complete genome
.GTGTGCCTGCAGGCTGCCATTAACGATTAGAACTGCGGAAAAAGCATTTATTGCAGCACCAAAGAGACCAAGCGCTC
AGAAACCACCCTTTTTGAGGCGCTTTAGCGTCTTTTTGTCCTTGATTTTGAAATTTAATTTCATCTTTTTTTAATAGTTTA
AAATTGCAGAACTAAAACAGTTCTTACTCAACTAAATACTTGAACTACCTCAACGACAACGGGAAATTTAATTGAAC
ACCTATTAAAGATATAGTGTTCTTCTCTATTTTAGATTTAGTAATCCCTAAAGGAGAAAATTTACGCGAATTCGATAT
CAAGCTT
>Fragment_Mpn489_Sig_R_-_Mpn489_Sig_F *M. pneumoniae* M129, complete genome
.GTGTGCCTGCAGGCTGCCATTGGAGGTATTGAGTGCAGAACACGCCGCTAGCACCACCCCTAGGCCAATTCCGGTAA
ATGTAAACGCAACTAAAGGTCATCGCTTTAACTTATAACCCATAAATTGTGTGTCCGTTAAAGAATAATAGTTTGCTA
ATTATCTCATGAATGGCACTAATTTAGCCATTGTTTATCCAGCTAATAAAATAGGTGAAGGTGAAGAATTCGATATCA
AGCTT
>Fragment_Mpn506_Sig_F_-_Mpn506_Sig_R *M. pneumoniae* M129, complete genome
.AAGCTTGATATCGAATTCGATTAAATTTTCATCTTAAAAGCTTTTATTTTTACCACACAGGATCTAATAAATTTCAAG
AGTTAAACAAGTACAAACTTTGTTAATATTTCTCCAAATTTCGGGTTTTTAGTGTTTATGAAATTTAAGTATGGTGCCA
TCTTTTTTAGTGGCTTTCTAGGTCTTTCTGCCATTTTGGCTGCCTGTGGTACAAAGGGTAAAGGCAGCCTGCAGGCACA
C

>Fragment_Mpn_152_fwd_Mpn152_Rev *M. pneumoniae* M129, complete genome

.AAGCTTGATATCGAATTCGCTTTTAAAAATACTTTTACTTCAGTAACTCAAACTAGTAAATTTCAAGGGTTAAACATC
TACAAACTTTGTTAATATTTCTCTAAATTTCGGGTTTCTAGTGTTTATGAAATTTAAGTACGGCGCTATTGTTTTTAGT
GGTCTTTTGGGCGTTTCGGCCATTTTAGCTGCCTGTGGGGCAAGGGGCAAATTTAACCAAGTTGATGGCAGCCTGCAG
GCACAC

>Fragment_Mpn588_Sig_R_-_Mpn588_Sig_F *M. pneumoniae* M129, complete genome

.GTGTGCCTGCAGGCTGCCAAAGTTTGGCTGGGTTGCAGCACAGGCACTAAGAATAGTGCTGGTACCCAGCAGTGTTA
AGAAACCCAATAACTTGAATTGCAGTCTCATAGACTAACTATTTTAATACAGATAGCCAAGGAGCTAAACCATTATA
GCCCGGTTTATCTAGTAAATAAAGAGACACTTAAACCATCAATGATTAATTGAAGAATTCGATATCAAGCTT

>Fragment_Mpn592_Sig_R_-_Mpn592_Sig_F *M. pneumoniae* M129, complete genome

.GTGTGCCTGCAGGCTGCCATTTTTGTGATTAGTGTTAGCTACTGTTAGCGTAGCAGAGCAAGCAGTTAGAAATGCTTG
ACTGGCGAACAGTGTGAGAAAACCAAAACCCTTGAGTTTAAACCCCATAGTGTGCTTTAATTTTAAAAAAGAAAGTT
TGAAAGGTGAACCCTTAATTAAAGGTGAACTTAGTGCAGTACTAAAGTCGCACTTCTTCTAAAGGTCTGTTGAATTCG
ATATCAAGCTT

>Alginate_lyase_A1-III_coding_sequence

ConstructCCCGGcAGCCTGCAGGCACACCCCTTCGACCAGGCCGTCGTGAAAGACCCCACGGCCTCGTATGTCGACGTC
AAGGCGCGTCGTACCTTCTTGCAGAGCGGGCAGCTCGATGACCGCCTCAAGGCAGCGCTGCCCAAGGAGTACGACTG
CACGACCGAGGCAACGCCTAACCCGCAGCAAGGCGAGATGGTCATTCCGCGCCGTTATCTTTCCGGCAATCACGGCC
CGGTGAATCCGGACTACGAACCGGTGGTGACGCTTTATCGTGATTTCGAGAAAATTTCCGCCACGCTTGGAAATCTCT
ACGTTGCGACGGGCAAGCCGGTGTACGCCACTTGTCTGCTGAACATGCTGGACAAGTGGGCCAAGGCCGACGCGTTG
CTCAACTACGACCCCAAGTCGCAGTCGTGGTACCAAGTCGAATGGTCGGCGGCGACCGCCGCCTTTGCCCTGTCGAC
GATGATGGCCGAGCCGAACGTCGACACAGCCCAGCGCGAGCGTGTGGTGAAGTGGCTCAACCGTGTGGCGCGCCATC
AGACGAGCTTTCCGGGGGGGCGACACGAGTTGCTGCAACAATCACTCGTATTGGCGCGGTCAGGAAGCGACCATCATC
GGCGTGATCAGCAAGGACGATGAACTCTTCCGTTGGGGGCTGGGCCGTTATGTGCAGGCGATGGGGCTGATCAATGA
AGACGGCAGCTTCGTGCATGAAATGACGCGTCACGAACAGTCCTTGCACTATCAGAACTACGCCATGCTGCCGCTGA
CGATGATTGCCGAAACGGCATCACGTCAGGGTATCGATCTGTACGCGTACAAAGAGAACGGTCGCGATATTCACTCA
GCGCGCAAGTTTGTCTTTGCTGCGGTGAAGAACCCGGACCTCATCAAGAAGTACGCTTCCGAGCCGCAGGATACGCG
TGCCTTCAAGCCGGGGCGGGGCGACCTGAACTGGATCGAATACCAGCGCGCGCGCTTTGGTTTTGCCGATGAGCTGG
GCTTCATGACGGTGCCGATCTTCGATCCGCGTACCGGTGGTTCGGGCACTTTGCTGGCGTACAAACCGGGCGCCGGTT
GGTCCCATCCCCAATTTGAAAAATAACCCGGG

>p53

ACCAAGGCAGCCTGCAGGAGGAGCCGCAGTCAGATCCTAGCGTCGAGCCCCCTCTGAGTCAGGAAACATTTTCAGAC
CTATGGAAACTACTTCCTGAAAACAACGTTCTGTCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCG
GACGATATTGAACAATGGTTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAGAATGCCAGAGGCTGCTCCCCCCGT
GGCCCCTGCACCAGCAGCTCCTACACCGGCGGCCCCTGCACCAGCCCCCTCCTGGCCCCTGTCATCTTCTGTCCCTTCC
CAGAAAACCTACCAGGGCAGCTACGGTTTCCGTCTGGGCTTCTTGCATTCTGGGACAGCCAAGTCTGTGACTTGCACG
TACTCCCCTGCCCTCAACAAGATGTTTTGCCAACTGGCCAAGACCTGCCCTGTGCAGCTGTGGGTTGATTCCACACCC
CCGCCCGGCACCCGCGTCCGCGCCATGGCCATCTACAAGCAGTCACAGCACATGACGGAGGTTGTGAGGCGCTGCCC
CCACCATGAGCGCTGCTCAGATAGCGATGGTCTGGCCCCTCCTCAGCATCTTATCCGAGTGGAAGGAAATTTGCGTGT
GGAGTATTTGGATGACAGAAACACTTTTCGACATAGTGTGGTGGTGCCCTATGAGCCGCCTGAGGTTGGCTCTGACTG
TACCACCATCCACTACAACTACATGTGTAACAGTTCCTGCATGGGCGGCATGAACCGGAGGCCCATCCTCACCATCAT
CACACTGGAAGACTCCAGTGGTAATCTACTGGGACGGAACAGCTTTGAGGTGCGTGTTTGTGCCTGTCCTGGGAGAG
ACCGGCGCACAGAGGAAGAGAATCTCCGCAAGAAAGGGGAGCCTCACCACGAGCTGCCCCCAGGGAGCACTAAGCG
AGCACTGCCCAACAACACCAGCTCCTCTCCCCAGCCAAAGAAGAAACCACTGGATGGAGAATATTTCACCCTTCAGA
TCCGTGGGCGTGAGCGCTTCGAGATGTTCCGAGAGCTGAATGAGGCCTTGGAACTCAAGGATGCCCAGGCTGGGAAG
GAGCCAGGGGGGAGCAGGGCTCACTCCAGCCACCTGAAGTCCAAAAAGGGTCAGTCTACCTCCCGCCATAAAAAAC
TCATGTTCAAGACAGAAGGGCCTGACTCAGACTGACCCGGGGGGGGGCAAGA

>Alpha-1-Anti-trypsin_M.pneumoniae_cdon_adapted

cctgcagGAAGACCCCCAAGGGGACGCGGCTCAAAAGACTGACACTTCCCACCATGACCAAGACCACCCTACCTTTAAC
AAGATTACCCCCAACTTAGCGGAATTTGCTTTTAGTCTATACCGGCAATTAGCCCACCAATCCAACAGTACCAATATT
TTTTTTTCCCCTGTGAGTATTGCGACTGCTTTCGCCATGCTCTCCTTAGGCACCAAGGCGGATACGCACGACGAGATTT

TAGAAGGTTTAAATTTTAACCTCACAGAAATCCCGGAAGCGCAAATTCATGAGGGTTTTCAAGAGCTCCTCCGAACC
CTCAACCAACCTGATAGTCAACTCCAATTAACCACCGGTAATGGTTTATTTCTCAGTGAAGGTTTAAAGCTAGTGGAC
AAGTTCCTGGAAGACGTCAAAAAGCTGTACCACTCAGAGGCTTTTACGGTTAACTTTGGCGATACCGAGGAAGCTAA
GAAACAAATTAACGACTACGTGGAAAAGGGAACGCAGGGCAAAATCGTGGACCTGGTTAAGGAATTGGATCGTGAT
ACTGTCTTCGCGTTAGTGAATTACATTTTTTTTCAAAGGTAAATGGGAACGTCCCTTCGAGGTTAAGGATACCGAAGAG
GAAGATTTTCACGTGGATCAAGTGACCACCGTGAAGGTGCCAATGATGAAGCGACTTGGTATGTTCAACATTCAACA
CTGCAAGAAGTTATCCAGTTGGGTGTTATTAATGAAATACTTAGGTAATGCTACCGCTATTTTTTTTTTACCAGACGA
AGGCAAACTACAACACTTAGAGAATGAGCTCACCCACGACATTATTACCAAGTTTTTAGAGAATGAGGATCGTAGAA
GCGCTAGTCTTCATCTTCCCAAATTATCCATCACGGGGACCTATGACTTAAAGAGTGTTTTAGGACAGTTAGGTATTA
CGAAGGTTTTTAGTAATGGCGCGGATCTCTCCGGCGTTACTGAAGAAGCCCCCTTAAAGCTCTCCAAGGCTGTGCATA
AGGCGGTGTTAACCATTGATGAAAAAGGCACGGAGGCGGCGGGCGCTATGTTCCTTGAAGCTATACCCGTGAGCATT
CCCCCCGAAGTTAAGTTTAACAAACCCTTCGTTTTTCTTATGATCGAGCAGAATACCAAGAGCCCCCTCTTTATGGGG
AAAGTGGTGAATCCCACCCAGAAAGGCGCCGGTTGGTCCCATCCCCAATTTGAAAAATAACCCGGG

>P30_Nterm_120bp *M. pneumoniae* M129, complete genome.
ATGAAGTTACCACCTCGAAGAAAGCTTAAACTGTTTTTATTAGCCTGAATGCTAGTGCTGTTCAGCGCTTTAATAGTG
CTTGCAACCTTAATTTTGGTACAGCACAACAATACCGAACTGACAGAAGTTAAGAGTGAATTGAGTCCCCTTAACGTT
GTTTTACACGCAGAAGAGGATACAGTACAAATTCAGGGCAAGCCGATTACTGAGCAAGCATGGTTTATTCCTACAGT
TGCTGGTTGCTTTGGTTTTAGTGCCCTAGCCATCATCTTGGGCCTTGCTATAGGACTGCCAATTGTGAAGCGCAAGGA
AAAACGCTTATTGGAGGAAAAGGAACGCCAAGAACAGTTAGCGGAACAG
>MMP_cleavage_site TCGGGTAGTGGCCCATTAGGCTTAAGTGGCTCAGGCCCTCTTGGGCTGGGCAGTGGTAGC
>eYFP
ATGGTTAGTAAAGGTGAAGAATTGTTTACGGGTGTTGTTCCGATCTTGGTTGAATTGGATGGGGACGTGAATGGTCAT
AAGTTTTCCGTCTCTGGTGAGGGAGAAGGGGATGCGACTTACGGGAAACTTACTTTAAAATTCATTTGTACCACCGGT
AAGTTGCCCGTGCCCTGACCAACCCTGGTGACAACGTTTGGGTATGGTCTGCAATGTTTCGCTCGCTACCCTGACCAC
ATGAAGCAACACGACTTCTTTAAATCGGCTATGCCTGAAGGTTACGTTCAAGAACGAACCATTTTTTTTAAGGATGAT
GGGAACTATAAGACTCGAGCGGAAGTAAAATTCGAAGGCGACACCCTCGTGAATCGCATCGAGTTAAAGGGGATTG
ACTTCAAGGAAGATGGTAACATTCTCGGACATAAGCTCGAATACAACTATAATAGTCACAACGTCTACATTATGGCG
GACAAGCAGAAGAACGGGATCAAGGTCAACTTTAAAATTAGGCACAACATCGAAGACGGTTCGGTTCAACTCGCCG
ATCACTACCAACAAAACACCCCGATTGGTGATGGTCCCGTGCTTCTACCGGATAATCACTACCTGAGTTACCAAAGCG
CCTTGTCGAAAGATCCTAACGAAAAGCGCGACCACATGGTTTTACTGGAATTTGTAACGGCTGCTGGAATTACGCTCG
GGATGGATGAATTATATAAATCGGGGTTGAGAAGCACCGGGTCCCGCtaa
>P65 *M. pneumoniae* M129, complete genome.
ATGGATATAAATAAACCAGGTTGAAATCAGTCTGACCAACAAGCTACTGCATACGATCCCAATCAGCAGCAGTACTA
TGGTGATGGTAGTACTTATTACGACCCCGACCAAGCAGTTGATCCCAACCAAGCGTATTACCCCGATCCGAATACCTA
TCCTGATGCAGCGGCTTATTACGGCTATGGTCAGGATGGTCAAGCATATCCACAAGACTATGCTCAAGATCCCAACC
AAGCGTATTATGCCGATCCCAATGCTTATCAGGACCCAAACGCTTACACCGATCCTAACGCCTATGTAGATCCCAATG
CTTATCAGGACCCAAACGCTTATGTAGATCCTAATAACTACACCGATCCCAATGCTTATTACGGCTATGGTCAGGATG
GTCAAGCATATCCACAAGACTATGCTCAAGATCCCAACCAAGCGTATTATGCCGATCCCAATGCTTATCAGGACCCA
AACGCTTACACCGATCCTTACTATGTTACTAGTACCGATCCCAACGCTTACTATGGTCAGGTCGATAATGTACCTGCT
TTAGAGGCGAGTGATCTCGCTTATGAGGTAACACCGCAAGAGCAAGCAGCGGAACAAGAGCTGTTTAGCGAACCGG
AAACAAAGGTAATCCGTGAAATCCACGAATTTCCCTTTGAGAAGATTCGCTCTTACTTCCAAACGGACTTTGACAGTT
ATAACAGCCGTTTAACCCAGCTCAAAGACAAGTTAGATAACGCCATCTTTTCGATGCGCAAAGCGATTGATACTGTC
AAGGAAAACAGCGCTAACTTGCAAATTATGAAGCAGAACTTTGAACGCCAACTCAAAGAACAGCAAACCCAACGGC
TTACAAGTAATACTGACGCGGAAAAGATTGGTGCTAAGATCAACCAATTAGAGGAACGCATGCAACGTTTATCGCGC
ACTATGGAAAGTGTGGAGTGAACCAAAAAGGAGCCACGTCAAGAACAGTTCGATCCACGCTTTGTCGATCCACGTAA
CTTTAATAACTATGTAAACAACACTGACACAATGATGTCAATGTTTGAAAAGGTATTAATGATGAACTTACTGCGCAG
CACCACTCCAGTCCAACCTCCAGTGCAGTACTTTACTCCTCAACCCCTTACAGCATCGCCACGTCCAGTGTATGAAGA
ACCGATCTCCGCTTCGTTCCGCCGTCGTGGTTACCGTGGTGATGAATTTTACGAA

>eFTu_promotor

GAAGACCTTTTGTGCTAACGCCAGTTTGGCAAATCAAGTTCTGATTTTGCAATTATTTTGCTCCATATGAATTACACTA
CTCCAAGAATTATAAGCCTCTCTACAGCTTTATCTCAAACTTATGTAAAATTAGAGACGTAATTCAAACAC

>P1 *M. pneumoniae* M129, complete genome.

ATGCACCAAACCAAAAAAACTGCCTTGTCCAAGTCCACTTGGATTCTCATCCTCACCGCCACCGCCTCCCTCGCGACG
GGACTCACCGTAGTGGGACACTTCACAAGTACCACCACGACGCTCAAGCGCCAGCAATTTAGCTACACCCGCCCTGA
CGAGGTCGCGCTGCGCCACACCAATGCCATCAACCCGCGCTTAACCCCGTGAACGTATCGTAACACGAGCTTTTCCTC
CCTCCCCCTCACGGGTGAAAATCCCGGGGCGTGGGCCTTAGTGCGCGACAACAGCGCTAAGGGCATCACTGCCGGCA
GTGGCAGTCAACAAACCACGTATGATCCCACCCGAACCGAAGCGGCTTTGACCGCATCAACCACCTTTGCGTTACGC
CGGTATGACCTCGCCGGGCGCGCCTTATACGACCTCGATTTTTCGAAGTTAAACCCGCAAACGCCCACGCGCGACCA
AACCGGGCAGATCACCTTTAACCCCTTTGGCGGCTTTGGTTTGAGTGGGGCTGCACCCCAACAGTGAAACGAGGTCA
AAAACAAGGTCCCCGTCGAGGTGGCGCAAGACCCCTCCAATCCCTACCGGTTTGCCGTTTTACTCGTGCCGCGCAGCG
TGGTGTACTATGAGCAGTTGCAAAGGGGGTTGGGCTTACCACAGCAGCGAACCGAGAGTGGTCAAAATACTTCCACC
ACCGGGGCAATGTTTGGCTTGAAGGTGAAGAACGCCGAGGCGGACACCGCGAAGAGCAATGAAAAACTCCAGGGCG
CTGAGGCCACTGGTTCTTCAACCACATCTGGATCTGGCCAATCCACCCAACGTGGGGGTTCGTCAGGGGACACCAAA
GTCAAGGCTTTAAAAATAGAGGTGAAAAAGAAATCGGACTCGGAGGACAATGGTCAGCTGCAGTTAGAAAAAAATG
ATCTCGCCAACGCTCCCATTAAGCGGAGCGAGGAGTCGGGTCAGTCCGTCCAACTCAAGGCGGACGATTTTGGTACT
GCCCTTTCCAGTTCGGGATCAGGCGGCAACTCCAATCCCGGTTCCCCCACCCCCTGAAGGCCGTGGCTTGCGACTGAG
CAAATTCACAAGGACCTCCCCAAATGATCCGCCTCGATCCTGATTCTGTACGATGCGCCTTATGCGCGCAACCGTACC
GCCATTGACCGCGTTGATCACTTGGATCCCAAGGCCATGACCGCGAACTATCCGCCCAGTTGAAGAACGCCCAAGTG
AAACCACCACGGTTTGTGGGACTGAAAGGCGCGCGATGTTTTGCTCCAAACCACCGGGTTCTTCAACCCGCGCCGCC
ACCCCGAGTGGTTTGATGGCGGGCAGACGGTCGCGGATAACGAAAAGACCGGGTTTGATGTGGATAACTCTGAAAAC
ACCAAGCAGGGCTTTCAAAAGGAAGCTGACTCCGACAAGTCGGCCCCGATCGCCCTCCCGTTTGAAGCGTACTTCGC
CAACATTGGCAACCTCACCTGGTTCGGGCAAGCGCTTTTGGTGTTTGGTGGCAATGGCCATGTTACCAAGTCGGCCCA
CACCGCGCCTTTGAGTATAGGTGTCTTTAGGGTGCGCTATAATGCAACTGGTACCAGTGCTACTGTAACTGGTTGACC
ATATGCCTTACTGTTCTCAGGCATGGTCAACAAACAAACTGACGGGTTAAAGGATCTACCCTTTAACAATAACCGCTG
GTTTGAATATGTACCACGGATGGCAGTTGCTGGCGCTAAGTTCGTTGGTAGGGAACTCGTTTTAGCGGGTACCATTAC
CATGGGTGATACCGCTACCGTACCTCGCTTACTGTACGATGAACTTGAAAGCAACCTGAACTTAGTAGCGCAAGGCC
AAGGTCTTTTACGCGAAGACTTGCAACTCTTCACACCCTACGGATGAGCCAATCGTCCGGATTTACCAATCGGGGCTT
GAAGTAGTAGTAGTAGTAGTCACAACGCACCCTACTACTTCCACAATAACCCCGATTGACAAGACCGTCCAATC
CAAAATGTGGTTGATGCCTTTATTAAGCCCTGAGAGGACAAGAACGGTAAGGATGATGCCAAATACATCTACCCTTA
CCGTTACAGTGGCATGTGAGCTTGACAGGTATACAACTGGTCCAATAAGCTCACTGACCAACCATTAAGTGCTGACTT
TGTCAATGAGAATGCTTACCAACCAAACTCCTTGTTTGCTGCTATTCTCAATCCGGAATTGTTAGCAGCTCTTCCCGAC
AAGGTTAAATACGGTAAGGAAAACGAGTTTGCTGCTAACGAGTACGAGCGCTTTAACCAGAAGTTAACGGTAGCTCC
TACCCAAGGAACAAACTGATCCCACTTCTCCCCCACGCTTTCCCGTTTCTCCACCGGGTTCAACCTTGTGGGGTCGGT
GCTCGACCAGGTGTTGGATTATGTGCCCTGGATTGGGAATGGGTACAGGTATGGCAATAACCACCGGGGCGTGGATG
ATATAACCGCGCCTCAAACCAGCGCGGGGTCGTCCAGCGGAATTAGTACGAACACAAGTGGTTCGCGTTCCTTTCTCC
CGACGTTTTCCAACATCGGCGTCGGCCTCAAAGCGAATGTCCAAGCCACCCTCGGGGGCAGTCAGACGATGATTACA
GGCGGTTCGCCTCGAAGAACCCTCGACCAAGCCAACCTCCAGCTCTGAACGGGGGCGGGGTGAAGGAATGATAAGG
CTTCAAGTGGACAAAGTGACGAAAACCACACCAAGTTCACGAGCGCTACGGGGATGGACCAGCAGGGACAATCAGG
TACCTCCGCGGGGAATCCCGACTCGTTAAAGCAGGATAATATTAGTAAGAGTGGGGATAGTTTAACCACGCAGGACG
GCAATGCGATCGATCAACAAGAGGCCACCAACTACACCAACCTCCCCCCCAACCTCACCCCCACCGCTGATTGACCG
AACGCGCTGTCATTCACCAACAAGAACAACGCGCAGCGCGCCCAGCTCTTCCTCCGCGGCTTGTTGGGCAGCATCCC
GGTGTTGGTGAATCGAAGTGGGTCCGATTCCAACAAATTCCAAGCCACCGACCAAAAATGGTCCTACACCGACTTAC
ATTCGGACCAAACCAAACTGAACCTCCCCGCTTACGGTGAGGTGAATGGGTTGTTGAATCCGGCGTTGGTGGAAACC
TATTTTGGGAACACGCGAGCGGGTGGTTCGGGGTCCAACACGACCAGTTCACCCGGTATCGGTTTTAAAATTCCCGA
ACAAAATAATGATTCCAAAGCCACCCTGATCACCCCCGGGTTGGCTTGAACGCCCCAGGACGTCGGTAACCTCGTTG
TCAGTGGCACCACGGTGAGCTTCCAGCTCGGCGGGTGGCTGGTCACCTTCACGGACTTTGTCAAACCCCGCGCGGGTT
ACCTCGGTCTCCAGTTAACGGGCTTGGATGCAAGTGATGCGACGCAGCGCGCCCTCATTTGGGCCCCCCGGCCCTGA
GCGGCCTTTCGTGGCAGTTGGGTCAACCGGTTGGGCCGCGTGGAGAGTGTGTGGGATTTGAAGGGGGTGTGGGCGGA
TCAAGCTCAGTCCGACTCGCAAGGATCTACCACCACCGCAACAAGGAACGCCTTACCGGAGCACCCGAATGCTTTGG

CCTTTCAGGTGAGTGTGGTGGAAGCGAGTGCTTACAAGCCAAACACGAGCTCCGGCCAAACCCAATCCACTAACAGT
TCCCCCTACCTGCACTTGGTGAAGCCTAAGAAAGTTACCCAATCCGACAAGTTAGACGACGATCTTAAAAACCTGTTG
GACCCCAACCAGGTTCGCACCAAGCTGCGCCAAAGCTTTGGTACAGACCATTCCACCCAGCCCCAGCCCCAATCGCT
CAAAACAACGACACCGGTATTTGGGACGAGTAGTGGTAACCTCAGTAGTGTGCTTAGTGGTGGGGGTGCTGGAGGGG
GTTCTTCAGGCTCAGGTCAATCTGGCGTGGATCTCTCCCCCGTTGAAAAAGTGAGTGGGTGGCTTGTGGGGCAGTTAC
CAAGCACGAGTGACGGAAACACCTCCTCCACCAACAACCTCGCGCCTAATACTAATACGGGGAATGATGTGGTGGGG
GTTGGTCGACTTTCTGAAAGCAACGCCGCAAAGATGAATGACGATGTTGATGGTATTGTACGCACCCCACTCGCTGA
ACTGTTAGATGGGGAAGGACAAACAGCTGACACTGGTCCACAAAGCGTGAAGTTCAAGTCTCCTGACCAAATTGACT
TCAACCGCTTGTTTACCCACCCAGTCACCGATCTGTTTGATCCGGTAACTATGTTGGTGTATGACCAGTACATACCGCT
GTTTATTGATATCCCAGCAAGTGTGAACCCTAAAATGGTTCGTTTAAAGGTCTTGAGCTTTGACACCAACGAACAGAG
CTTAGGTCTCCGCTTAGAGTTCTTTAAACCTGATCAAGATACCCAACCAAACAACAACGTTCAGGTCAATCCGAATAA
CGGTGACTTCTTACCACTGTTAACGGCCTCCAGTCAAGGTCCCCAAACCTTGTTTAGTCCGTTTAACCAGTGACCTGA
TTACGTGTTGCCGTTAGCGATCACTGTACCTATTGTTGTGATTGTGCTCAGTGTTACCTTAGGACTTGCCATTGGAATC
CCAATGCACAAGAACAAACAGGCCTTGAAGGCTGGGTTTGCGCTATCAAACCAAAAGGTTGATGTGTTGACCAAAGC
GGTTGGTAGTGTCTTTAAGGAAATCATTAACCGCACAGGTATCAGTCAAGCGCCAAAACGCTTGAAACAAACCAGTG
CGGCTAAACCAGGAGCACCCCGCCCACCAGTACCACCAAAGCCAGGGGCTCCTAAGCCACCAGTGCAACCACCTAA
AAAACCCGCT