UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea in Biotecnologie

ELABORATO DI LAUREA

A standard approach to tropism variation in M13 bacteriophage

Tutor: Prof. Cristiano De Pittà
Dipartimento di Biologia

Co-tutor: Dott. Davide Colaianni
Dipartimento di Biologia

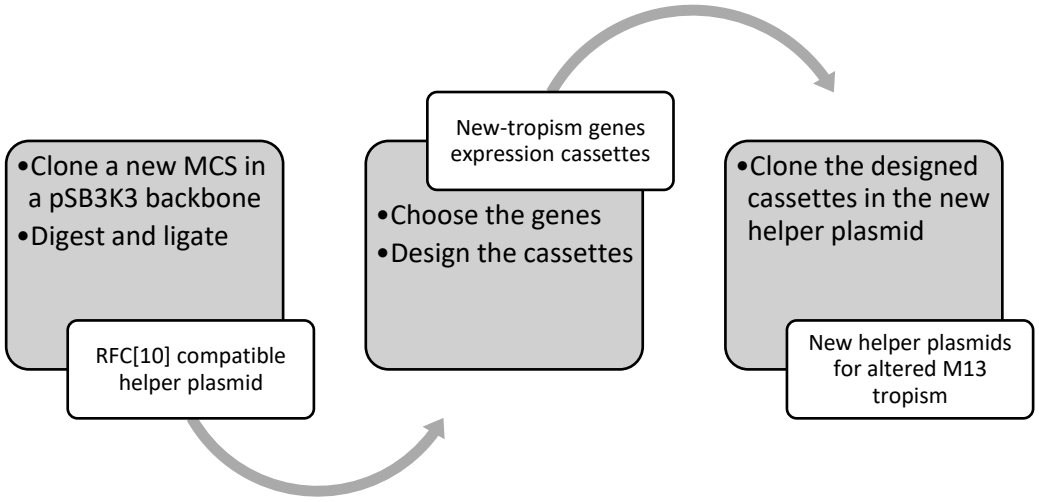
Laureando: Tommaso Varaschin

ANNO ACCADEMICO 2022/2023

INDEX

1 ABSTRACT

Antibiotic resistance is a globally recognized issue of great concern. While research for novel antibiotics races against evolution to find effective molecules, the scientific community is looking for new possible approaches to face this problem. Bacteriophages are the possible solution: although currently used only as compassionate therapy in situations where no other viable option is available, they could represent the next frontier. In this work, an approach to tropism variation based on synthetic biology is proposed, with the purpose of building a resilient platform to produce engineered M13 bacteriophages with the desired tropism determinant. Moreover, a standard helper plasmid is introduced as a fast and efficient way to produce phage particles inside packaging cells, by easily co-transforming the helper plasmid and the desired phage genome. The generation of a BioBrick RFC[10] standard-compatible helper plasmid that utilizes tropism-determinant genes as interchangeable parts allows to swap between genes to modify the tropism. In this way, it will be easier to develop new particles in order to fit for different purposes, ranging from therapeutics to protein library screenings.



Graphical abstract. Flow chart of the experimental design. In this graph are highlighted the key points of the research work that has been made in the present work

2 INTRODUCTION

2.1 THE BIOLOGICAL PROBLEM: ANTIMICROBIAL RESISTANCE

Antimicrobial resistance (AMR) is a globally recognized issue of increasing concern, occurring when microorganisms evolve to grow in the presence of substances that were previously toxic to them, particularly antibiotics[1]. As a consequence, this phenomenon is considered a significant threat to public health[1]. According to the World Health Organization (WHO), by 2050 it will cause over 10 million deaths worldwide[2]. Current widely used antibiotic therapies are facing growing obstacles in effectively contrasting bacterial infections[1,3], and research efforts in this area are becoming increasingly lengthy, complex, and costly[1,3]. Additionally, AMR has a severe impact also on the environment, with a particular interest on the agri-food sectors[4]. As reported by the Food and Agriculture Organization of the United Nations (FAO), the impact on these fields will be cause for production losses and damages to livelihoods[4], also due to the fact that AMR can spread among different hosts, thus contaminating the entire food chain[4]. As a result, it is estimated that, if not faced, AMR may cumulatively result in a 3.4 trillion USD loss in the world annual gross domestic product in ten years[4]. Thus, a collective action must take place, both on public health policies and research, as alternatives to conventional drugs needs to be found.

2.2 PHAGE THERAPY: STATE OF THE ART

The use of bacteriophages (hereinafter referred to as phages) (Figure 1) as treatment for bacterial infections has been hypothesized since their discovery, made by Félix d'Hérelle in 1917[3,5]. In the following years, *in vivo* experiments were carried out by other scientists[3], followed by the first, small-scale, and prudent clinical

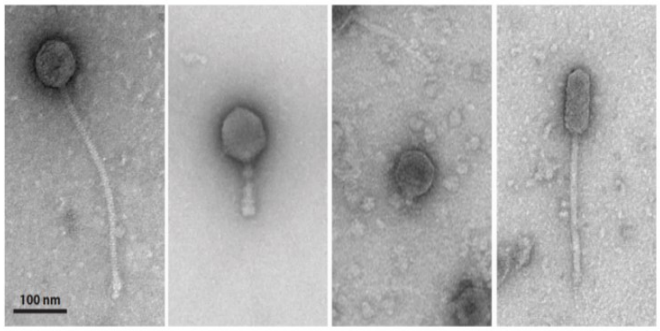


Figure 1. Bacteriophages observed at the electron microscope (EM). EM images showing four different phages, pointing out how these viruses come in different shapes and forms[18].

trials[3]. This led to a series of successful results[3], but this therapeutic approach was abandoned soon after the discovery of penicillin, made by Alexander Fleming in 1928[3]. Back to the present day, while to date their clinical application has been limited to compassionate care for individuals with multidrug-resistant infections for whom no other viable option is available[3,6], a significant part of the scientific community agrees that phages may represent the next tool for tackling this global problem[3,6]. In compassionate care, the employed bacteriophages are derived from environmental samples (*e.g.* from sewage systems) that have demonstrated *in vitro* lytic activity against the clinical isolate[3]. Therefore, up to now, phage particles are purified and administered without a precise characterization, and their potential clinical efficacy relies just on the assumption that their specific antibacterial activity *in vitro* could be reproducible also *in vivo*[3]; in addition, these therapies are usually personalized on the clinical case[3]. Yet, for this approach to become a routine in clinical practice, it requires a substantial increase in understanding what is administered, coupled with the establishment of standardized clinical protocols.

To this end, phage engineering has been proposed as a solution, allowing to thoroughly control phage particles production[5–7]. A molecular engineering approach would allow the production of safer phage particles, purified from cellular debris and potential bacterial antigens, that currently represent one of the major issues[3]. In addition, the production of engineered phage particles would make it possible to dispense a fully characterized genome of desire, containing gene editing tools and/or silencing systems that can interfere with the protein expression of pathogenic bacteria[8]. Typically, to achieve this aim, the phage genome is reduced to plasmids (hereinafter referred to as phagemids), capable of being encapsulated but lacking the capsid protein-coding genes, which are instead produced by a helper phage or a helper plasmid[7,8]. When co-transfected or co-transformed together, these two vectors enable the production of viable phage particles carrying the desired genome[7,8] (Figure 2).

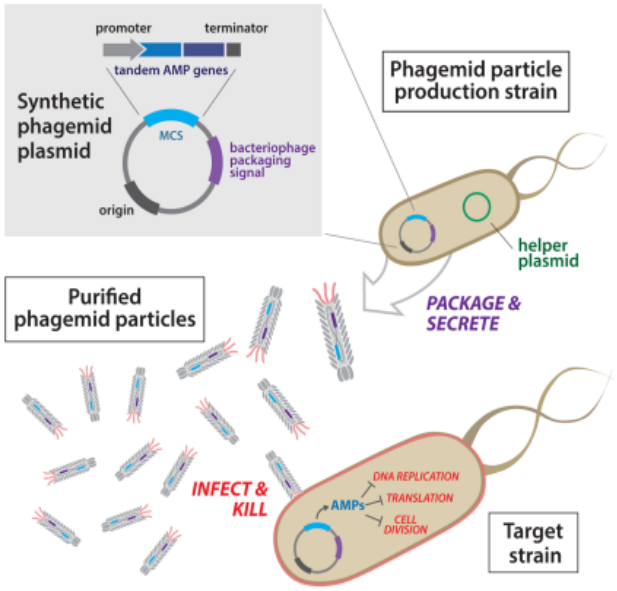


Figure 2. The packaging cycle of a phagemid. This image shows how the co-transformation of a helper plasmid with a phagemid allows the production and the subsequent purification of phage particles with the desired genome[19].

Such an approach will consequently result in the production of engineered bacteriophages that are no longer unknown and therefore suitable as clinical therapy. Nevertheless, many challenges still need to be faced, with the aim of promoting the spread of bacteriophage research, by developing innovative solutions that simplify exploration and implementation of novel and pragmatic approaches. So, the present work aims to lay the groundwork for a possible approach to bacteriophage engineering, based on the concept of DNA parts standardization, while exploring the possibility of varying the host range of bacteriophages. Phages are capable of infecting bacteria only[3], and they do so by recognizing peculiar molecules exposed on their cell surface (*i.e*. protein structures such as pili, flagella, porins, or efflux pumps)[3] (Figure 3). The recognition of these bacterial molecules relies on proteins exposed on phage surface and is highly specific, resulting in phages' ability to discriminate among the bacteria they can infect, in some cases even among different strains[3]. If on one hand it is a great advantage to have an organism able to target bacteria in such a specific manner, on the other hand it presents many limits in terms of application spectrum. These considerations lead us to explore potential strategies for engineering phage host range. Although many examples are present in the existing literature[5,6], none of them proved to be suitable for the intended purposes, mainly due to variations in the utilized bacteriophages[5,6]. Moreover, while tropism variation was explored in some head-tail phages[5,6], there are no examples for filamentous phages. Our purpose could be achieved in two opposites,

yet complementary, ways. The first one is a random approach, inspired by antibody-specific engineering, that is based on the mutagenesis of regions involved in target recognition, followed by a screening against the desired host(s)[6]. If on one hand this methodology is very time-consuming, on the other hand it can result in a broader extension of the phage tropism since many possibilities are simultaneously explored. The second method can be otherwise described as a rational approach, using synthetic biology to modulate phage host ranges by swapping tropism determinants between different phage species[5]. Although it requires a deep knowledge of the tropism determinants, this approach allows a higher confidence over the possible results. These two different methods, even if so far applied to profoundly different families of bacteriophages (*i.e*. T4, T3, and T7)[5,6], represented the guidelines of our research. In the end, keeping in mind all these considerations, the second one, with the rational design of DNA parts, was chosen, as it allows to assemble expression cassettes for alternative genes to vary the phage host range in a quicker and possibly more reliable way.

2.2.1 M13 tropism

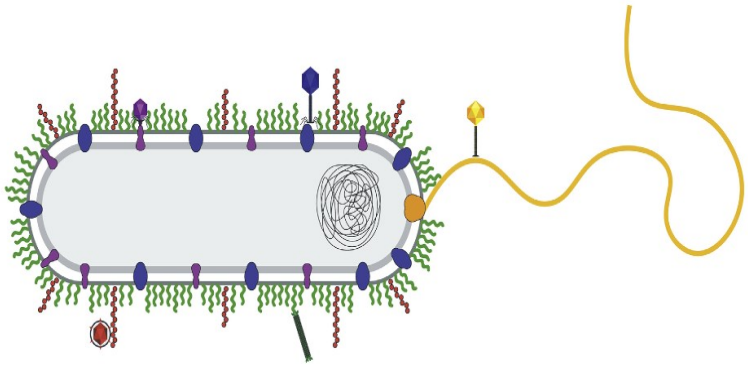


Figure 3. Examples of phage-bacteria interactions. In this image, several molecules exposed on the bacteria cell surface and capable of being recognized by proteins present on phage surface are highlighted in different colors: pili (in yellow), flagella, porins (in purple), efflux pumps (in blue), and sugar molecules of lipopolysaccharides (in red and green)[3].

Currently there are no examples of tropism variation in filamentous bacteriophages and, in this work, a possible approach in M13 bacteriophage is proposed. M13 uses a minor protein of its capsule, Gp3, to interact with the pilus (F' factor) of *Escherichia coli* [9,10]. Gp3 is expressed from *gene 3*, a 406 amino acids-long protein[9,10] that sits in three to five copies on one end of the capsid[10], being responsible not only for the interaction with the outer structures of the bacteria but also for the termination of the virion assembly[9,10]. Once correctly folded, it forms three

functionally distinguished domains (N1, N2, and the C-terminal domain CTD)[9,10]. The CTD anchors Gp3 to the phage by interacting with other phage coat proteins, whereas the N1 and N2 domains are responsible for the interaction with the bacteria[10]. The three-dimensional structures of the N1 and N2 domains have been determined by circular dichroism spectrum analysis and X-ray diffraction[11,12], whereas the CTD has not yet been characterized. Interestingly, even though N2 domain is in the middle of Gp3 amino acid sequence, the 3D reconstruction revealed that it is located on one end of the protein and is responsible for the interaction with the F' factor[10]. The initial positioning of N2 domain hides N1, which becomes available after the first contact of N2 with the pili, resulting in N1 domain interaction with the bacterial protein TolA[9]. TolA is a 421 amino acids-long protein, located in the periplasmatic space of *E. coli*[13]; it is anchored to the inner membrane with its N-terminus, spanning through the periplasm with a long α-helix ending with its C-terminal domain, also called TolAIII[13]. It has been demonstrated that TolAIII is functional in the infection process of M13 phage[9,13], due to its high affinity interaction (1-1,9μM) with the N1 domain[9]. In this work, two approaches to tropism variation are addressed: the first concerning the interaction of Gp3 with TolA and its homologues, whereas the second one focusing on M13 bacteriophage gene 3 homologues.

3 MATERIALS AND METHODS

3.1 CONSTRUCTION OF A BIOBRICK-COMPATIBLE HELPER PHAGE

3.1.1 BioBrick-RFC[10] standard

The main purpose of this work was to make the phage genome, encoding the phage proteins, compatible with the BioBrick RFC[10] standard. The BioBrick standard was ideated by Tom Knight[14] with the aim to bring in biology the fundamental concepts of engineering: the ability to interchange parts, to assemble subcomponents and to rely on previously manufactured components[14]. This approach has the characteristic of leaving the key structural elements, after each assembly reaction, the same as before, so that they can be used as input to any following manipulation[14], as it happens in the engineering world. In the BioBrick standard every component, called "part", is flanked at the 5' end by *Eco*RI and *Xba*I enzymes restriction sites, forming the so-called prefix, whereas the 3' end is flanked by *Spe*I and *Pst*I enzymes cut sites, forming the suffix[14]. Both the sequences are also provided of a *Not*I enzyme cut site, useful to excise the inserted part (Figure 4). Importantly, these same restriction sites must be absent in other positions in the vector and in the standard parts, making them unique so that the assembly process can work.

```
5' --gca GAATTC GCGGCCGC T TCTAGA G --insert-- T ACTAGT A GCGGCCG CTGCAG gct--
   --cgt CTTAAG CGCCGGCG A ACATCT C ---------- A TGATCA T CGCCGGC GACGTC cga--
         EcoRI   NotI       XbaI                  SpeI     NotI    PstI
```

Figure 4. Restriction sites positioning in the BioBrick standard. The BioBrick standard is characterized by the presence of two unique sequences, the prefix and the suffix, flanking the ends of every standard-compatible part. In these two sequences, a total of four unique restriction sites (recognized by *Eco*RI, *Xba*I, *Spe*I, and *Pst*I) are present, together with two cut sites recognized by *Not*I[14].

3.1.2 Cloning of M13 genome into a pSB3K3 backbone

In order to construct a BioBrick RFC[10]-compliant helper plasmid, the genome of the M13 bacteriophage was integrated into a standard backbone, specifically pSB3K3 (addgene, Watertown, Massachusetts, USA). The starting point was the helper plasmid M13cp, previously constructed and kindly provided by Chasteen *et al*.[15], carrying the genes encoding the phage proteins and a p15a[15] replication origin (ORI). Due to the compatibility and copy number characteristics conferred by this ORI, the pSB3K3 backbone, which carries the same origin, was chosen.

Accordingly, a plasmid DNA extraction was performed on 5 mL of an *E. coli* overnight culture (grown at 37°C in agitation at 220 rpm) carrying a plasmid bearing pSB3K3 as backbone, using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) and following manufacturer's instructions. The undesired insert, located between the prefix and the suffix of the plasmid, was then excised through enzymatic digestion using *Xba*I and *Spe*I restriction enzymes; this was then followed by the ligation (further details about digestion and ligation protocols are reported in section 3.3) of the cohesive ends in order to obtain the circularized plasmid. This plasmid served as a template for the insertion of a different multiple cloning site (MCS) bearing the cut sites for *Alw*21I and *Mlu*I restriction enzymes, inserted via PCR mutagenesis. For this purpose, the following PCR primers (the restriction sites of *Alw*21I and *Mlu*I respectively are indicated as underlined) were designed on Benchling.com (San Francisco, California, USA) and then synthesized by Eurofins Genomics Italy srl (Milan, Italy) to recognize the downstream region of pSB3K3 suffix:

forward primer: 5'-TAGTGCACAGTCAGCGTAATGCTCTGCC-3'

reverse primer: 5'-ATTAACGCGTATTACCGCCTTTGAGTGAGC-3'

Mutagenesis PCR was executed on six identical samples to amplify the output, utilizing Q5 High-Fidelity DNA Polymerase (Promega, Madison, Wisconsin, USA). Each reaction was prepared in a volume of 50 µL, comprising the addition of 0.5 µL of each primer (10 mM), along with approximately 40 ng of DNA template, 10 µL of 5X Q5 Reaction Buffer, 0.75 µL of Q5 High-Fidelity DNA Polymerase and the addition of water to achieve the final volume. Employing a thermocycler, 5 PCR cycles were performed, each of which involving denaturation at 95°C for 30 seconds, annealing at 63°C for 30 seconds, and extension at 72°C for 3 minutes; subsequently, additional 30 cycles were carried out similarly to the previous ones, but with an annealing temperature of 71°C. After PCR mutagenesis, an electrophoresis on agarose gel was performed to check for the plasmid amplification. The PCR product was purified using the QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). Next, 200 ng of the purified product were digested using the aforementioned restriction enzymes and following the protocol indicated in section 3.3.2.

Concurrently, we extracted the M13 genome from the M13cp helper plasmid[14]. 200 mL of an overnight culture (grown at 37°C in agitation at 220 rpm) were subjected to plasmid extraction using the ZymoPURE II Plasmid Midiprep Kit (Zymo Research, Irvine, California, USA) following manufacturer's instructions. This construct was digested by using *Pst*I and *Mlu*I restriction enzymes in order to

segregate the region containing the ORI and the selection marker from the segment harboring the M13 genome. This latter segment was isolated by gel electrophoresis (further details are reported in section 3.3.1), with the heavier band, measuring 6.000 bp, that was excised and purified utilizing the QIAquick PCR & Gel Cleanup Kit (Qiagen, Hilden, Germany) following manufacturer's instructions.
Once obtained, the two DNA fragments, namely pSB3K3 with the new MCS, that will serve as new backbone, and M13 genome, that represents the insert, were ligated together to create a novel construct, hereinafter referred as pSB3KM13. The same ligation reaction was set up using three different molar ratios (moles of insert over moles of vector): 1:1, 2:1, and 4:1. To determine the ratios, the M13 genome was considered as the vector due to its longer length (6.000 bp) in comparison to the actual backbone (2.700 bp), which was therefore regarded as the insert. After an overnight ligation, the entire reaction volume was transformed into chemically competent *E. coli* TOP10 cells, then plated on LB agar supplemented with kanamycin (25 µg/mL) and incubated overnight at 37°C. The next day, a total of 13 colonies were selected, inoculated into liquid LB supplemented with kanamycin (25 µg/mL), and further incubated overnight at 37°C in a shaking incubator at 220 rpm. After 16 hours of incubation, the cell culture was pelleted and the plasmids extracted using the QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) following manufacturer's instructions. The samples were tested by both gel screening and Sanger sequencing. For gel screening, the construct was digested with *Alw*21I and *Eco*RI restriction enzymes to produce an expected characteristic digestion pattern comprising three bands (722, 2591, and 5489 bp). For Sanger sequencing, the selected primer was VF2 (5'-TGCCACCTGACGTCTAAGAA-3'), and the sequencing experiment was performed by Eurofins Genomics Italy srl (Milan, Italy).

3.2 DESIGN AND CLONING OF ALTERNATIVE G3 GENES

3.2.1 Bioinformatic analysis

In order to explore new tropism determinants, a series of different bioinformatic tools have been used to identify homologues and variants of M13 *gene 3*.
Bioinformatic analysis of TolA protein (UniProt ID: P19934) homologues was performed using the BLASTp tool with pre-set parameters, except for the "max target sequences" set at 250 hits and the exclusion of *Escherichia coli* as a target organism. A surface protein of *Klebsiella pneumoniae*, called "cell envelope integrity

protein TolA" (NCBI ID: MBS4159696.1), was specifically identified, sharing 100% similarity with the TolAIII domain of *E. coli* (residues 305-421)[13]. For this reason, using the online tool available on Benchling.com (San Francisco, California, USA), the sequence of M13 *gene 3* was imported and modified to delete the N2 domain, spanning aminoacidic residues from 87 to 256, making it thus able to interact only with TolA and its homologues.

In parallel, using the NCBI Taxonomy browser and looking for phylogenetically-related phages, the Pf3 bacteriophage of *Pseudomonas aeruginosa* (NCBI:txid10872) was identified as part of the *Inoviridae* family, the same as M13. The deposited genome of Pf3 phage was identified on the NCBI genome database (ID: NC_074763) and, from it, the *gene 3* sequence was imported on Benchling.com (San Francisco, California, USA) to be further processed.

Before assembling the expression cassettes for the tropism determinants genes, both on the coding sequences of the *G3* gene of M13 phage deleted of the sequence for the N2 domain, and of the *G3* gene of Pf3 phage, codon optimization was performed, utilizing two bioinformatic tools absolving different purposes. Firstly, using CODONATOR 3000, developed by the Sydney Australia iGEM Team in 2019, we harmonized the codon usage for the target specie (*E. coli*). The optimization was chosen according to the relative proportion in which the starting codon is translated into the codon of *E. coli*, based on the most similar frequency of occurrence, considering the bias in amino acid appearance. Subsequently, the IDT Codon Optimization Tool (IDT Technologies, Coralville, Iowa, USA) was used to eliminate repeated sequences and low complexity regions that could complicate the synthesis process. Using the same algorithm, a total of four forbidden restriction sites present in the two sequences were also deleted to ensure compatibility with the BioBrick standard.

Using Benchling.com online platform (San Francisco, California, USA), the expression cassettes for the tropism determinants genes were assembled joining promoters, coding sequences (CDS) and terminators. Once assembled, the sequences were synthetized by IDT Technologies (Coralville, Iowa, USA). The sequence comprising the *G3* gene of M13 deleted of the N2 domain was named *G3_M13_Cut_optimized*, whereas the *G3* gene of Pf3 bacteriophage will be referred as *G3_Pf3_optimized*.

3.2.2 Cloning

An amount of 800 ng of both *G3_M13_Cut_optimized* and *G3_Pf3_optimized* were each digested in four parallel reactions using *EcoRI* and *PstI* restriction enzymes (see section 3.3.2). After the incubation, all the samples were purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, Wisconsin, USA) following manufacturer's instructions.

At the same time, the standard helper plasmid, pSB3KM13, was also digested with the same enzymes, to open a gap between the prefix and the suffix allowing the insertion of the newly synthesized gene. The digested pSB3KM13 was run by gel electrophoresis, and the corresponding band excised and purified using the QIA quick PCR & Gel Cleanup kit (Qiagen, Hilden, Germany) following manufacturer's instructions.

The purified vector and the selected insert were then ligated together (see section 3.3.3), employing different ratios of insert over vector (5:1 and 7:1) to ensure a higher output. The resulting ligation products were transformed into *E. coli* MG1655 Z1 chemically competent cells and then plated on LB agar supplemented with kanamycin (25 µg/mL) in order to select the successfully transformed cells. After overnight incubation at 37°C, colonies containing the desired DNA sequences were identified and picked. The selected colonies were then inoculated into liquid LB medium containing kanamycin (25 µg/mL) and allowed to grow for 16 hours at 37°C in a shaking incubator at 220 rpm. Subsequently, the cell cultures were harvested and subjected to plasmid DNA extraction using QIAprep Spin Miniprep Kit (Qiagen, Hilden, Germany) to isolate the plasmids containing the desired sequences. The presence and the integrity of the sequences of interest within the plasmids were assessed by performing *Eco*RI and *Pst*I enzymatic digestion, followed by agarose gel electrophoresis in order to separate and visualize the DNA fragments.

3.3 CLONING TECHNIQUES

The protocols extensively used in the previous paragraphs, both to clone the sequences of the M13 bacteriophage into the new standard backbone and the tropism genes into the helper phage, are described below.

3.3.1 Gel electrophoresis

All electrophoresis gels were prepared with 1% m/V agarose in TAE 1X buffer. 7 µL of SYBR™ Safe DNA Gel Stain (ThermoFisher Scientific, Waltham, Massachusetts, USA) every 50 mL of gel were added to visualize the DNA bands. 6 µL of GeneRuler

1kb DNA Ladder (ThermoFisher Scientific, Waltham, Massachusetts, USA) were loaded alongside with the samples to allow an estimation of samples' lengths in base pairs (bp).

3.3.2 Enzymatic digestion

All DNA digestions were carried out using FastDigest Restriction Enzymes (ThermoFisher Scientific, Waltham, Massachusetts, USA). The total reaction volume of 20 µL was composed as it follows: 1 µL of each restriction enzyme, 2 µL of 10X FastDigest Green Buffer, 200-1000 ng of DNA, and water up to 20 µL. The reaction mix was then incubated in a thermocycler at 37°C for 15 minutes, followed by a 5-minute incubation at 80°C to inactivate the enzymes.

3.3.3 Ligation

All DNA ligations were carried out using T4 DNA Ligase (ThermoFisher Scientific, Waltham, Massachusetts, USA). The total reaction volume of 10 µL was composed as it follows: 1 µL (5 U) of T4 DNA ligase, 1 µL of 10X T4 DNA Ligase buffer, 20-100 ng of vector DNA, an amount of insert DNA in accordance with a molar ratio of insert to vector DNA ranging from 1:1 to 7:1 depending on the specific case, and water up to 10 µL. The ligation reaction was conducted in a thermocycler at 16°C overnight (not less than 16 hours). Following overnight incubation, the DNA ligase was heat-inactivated at 65°C for 20 minutes.

4 RESULTS AND DISCUSSION

4.1 A BIOBRICK RFC[10]-COMPATIBLE HELPER PLASMID

The *in vitro* production of engineered M13 phages can be achieved with the concurrent use of two constructs: a phagemid, representing the phage genome, and a helper plasmid, encoding the M13 phage capsid proteins. While the phagemid carries the intergenic sequence for encapsulation, the helper plasmid lacks this sequence, instead carrying an ORI and a selection marker[15]. Of particular interest to us was the M13cp helper plasmid, built by Chasteen *et al*.[15]. However, this construct was not compatible with the BioBrick standard[14], as it carries two *PstI* restriction sites flanking the ORI; furthermore, it lacks the BioBrick prefix and suffix. Given our intention to develop a new, standard-compatible helper plasmid, a new construct was assembled as detailed in section 3.1.2. Briefly, to eliminate the prohibited restriction sites and, at the same time, to implement the sequences of the prefix and the suffix, the region embedding the ORI and the selection marker was entirely substituted with the standard-compatible backbone pSB3K3.

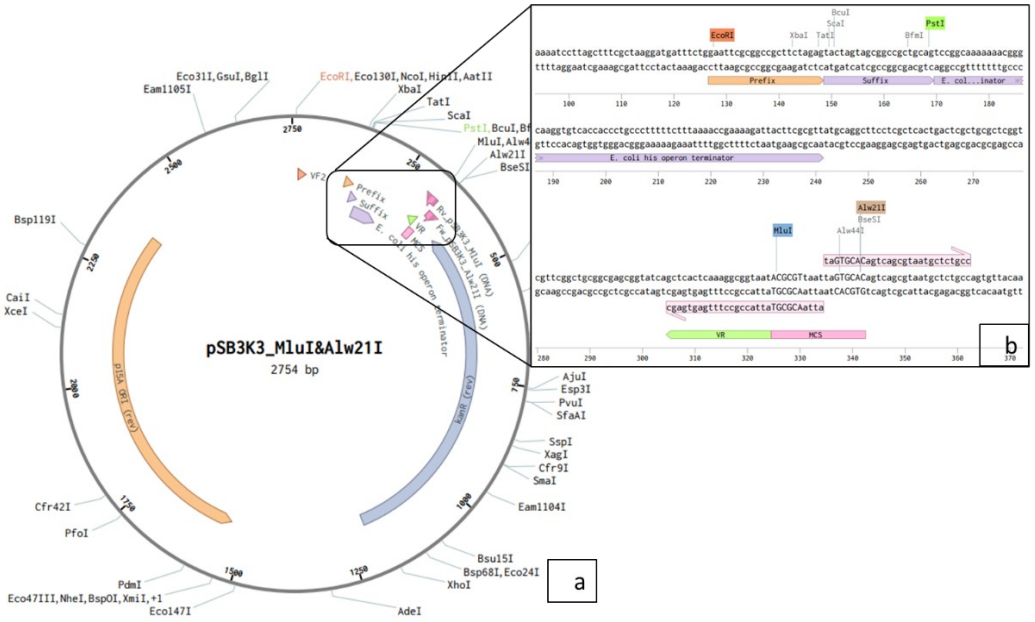


Figure 5. PCR mutagenesis of pSB3K3 backbone. (**a**) The original plasmid map of pSB3K3, with a magnification (**b**) of the downstream region of the suffix where the new MCS has been inserted by mutagenesis.

The first PCR mutagenesis on pSB3K3, required to insert a new MCS (Figures 5a and 5b), successfully performed only on one out of the six amplified samples (Figure 6).

Consequently, this unique sample was carried forward in the following procedures. The next step was to take advantage of the new MCS to clone the M13 genome isolated from the M13cp helper plasmid[15]. The resulting plasmid was named pSB3KM13 (Figure 7a). Its correct construction was confirmed through gel screening (Figure 7b) and sequencing.

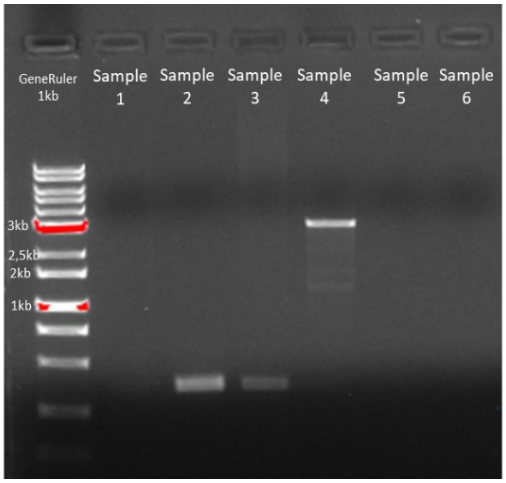


Figure 6. Electrophoresis gel performed after PCR mutagenesis. Starting from the left, 1 kb DNA ladder was loaded, followed by the six experimental samples, of which only Sample 4 showed proper amplification.

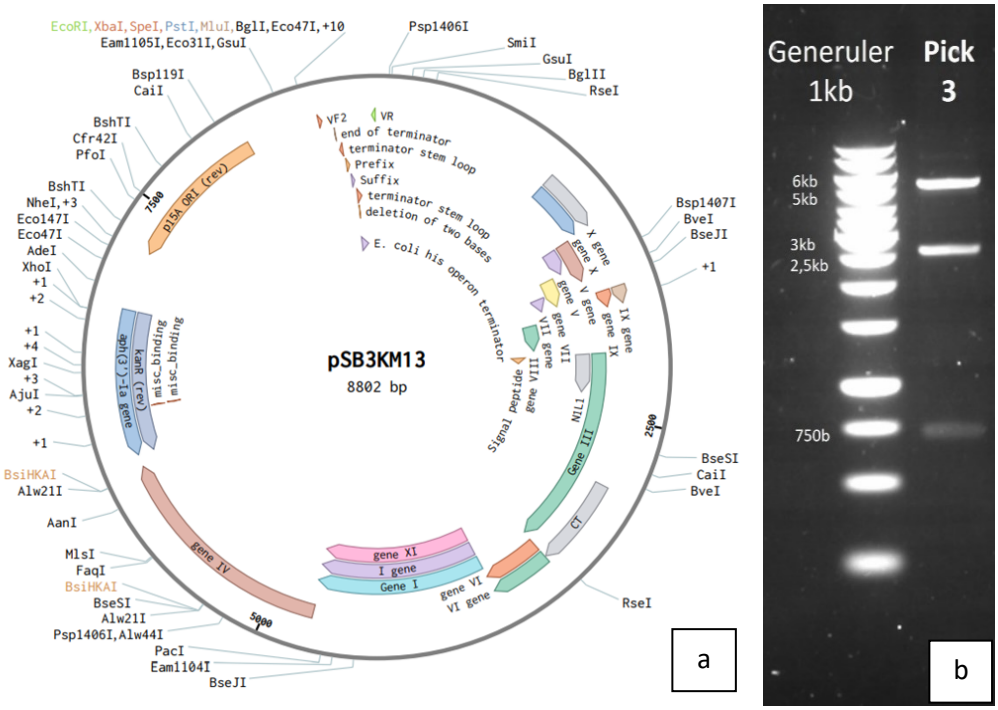


Figure 7. Construction of the helper phage. (**a**) Plasmid map of the final assembled construct, pSB3KM13. (**b**) Gel screening of pSB3KM13 double digestion with *Eco*RI and *Alw*21I restriction enzymes, from which the expected characteristic bands (722, 2591, and 5489 bp) were obtained.

4.2 NEW STANDARD PARTS FOR TROPISM VARIATION

Many benefits come with the achieved standardization of the helper plasmid, in particular the ability to easily interchange the parts between the prefix and the suffix of the vector. For this reason, we took advantage of this strategy to address the problem posed by tropism variation in bacteriophages by designing alternative genes for tropism to be inserted between the prefix and suffix regions. The design of new tropism determinants followed two paths, both based on literature research and the use of computational tools: one oriented on investigating the interaction between wild type Gp3, the protein responsible for tropism in M13, and TolA protein of *E. coli*, whereas the other focused on the research of different G3 proteins.

4.2.1 Exploit TolA homologues

The first path was based on the fundamental work of Donald A. Marvin, who suggested that M13 mutants lacking the N2 domain of Gp3 can infect bacteria without pilus with a 100-fold higher efficiency[16], since N2 masks the interaction of N1 with the bacterial co-receptor TolA. Building on this, we designed a gene with a deleted N2 domain, drawing from the findings of Nilsson and colleagues, who precisely delineated the amino acid regions corresponding to each individual domain[10]. Therefore, an alternative gene for tropism that could ideally enable *E. coli* infection in the absence of a pilus has been proposed. This gene, named "*G3_cut*," was obtained by *in silico* deletion of the region between aminoacidic residues 87 and 256, corresponding to the N2 domain[10]. However, considering this protein's ability to interact with the TolA protein of *E. coli*, specifically with the TolAIII domain[9,10], we hypothesized that this same modified gene could also interact with homologous proteins of TolA. The bioinformatic analysis of TolA homologues, thoroughly explained in section 3.2.1, identified a surface protein of *Klebsiella pneumoniae* that shared high similarity with *E. coli* TolA. Specifically, 100% similarity was found in the C-terminal region, which is directly implied in the interaction with the bacteriophage[9,13]. This suggests that such modification could potentially allow the engineered bacteriophage's entry into *K. pneumoniae* as well. This specie was of particular interest as a Gram-negative bacterium, which represents, according to the World Health Organization (WHO), a threat for multiple antibiotic resistances[3].

4.2.2 Exploit G3 homologues

The second path that was taken stems from the research of phylogenetically closed bacteriophages. Since proteins of evolutionary-related phages share a

certain degree of similarity, we hypothesized that they could be able to interact between each other regardless of their origin, still resulting in the correct assembly of the phage capsid. M13 bacteriophage is part of the *Inoviridae* family[16], that includes other filamentous phages; among them, the Pf3 bacteriophage was identified as capable of infecting *Pseudomonas aeruginosa*. Given the relevance of this pathogen in the AMR phenomenon[3], it was considered as an interesting possibility in attempting tropism variation. As in all filamentous phages the tropism determinant protein is annotated as "Gene 3 product"[10], the coding sequence was identified by analyzing the deposited genome in the NCBI databank (ID: NC_074763). This led to the generation of the "*Gp3_Pf3*" sequence.

4.3 *IN SILICO* DESIGN OF EXPRESSION CASSETTES

Once we identified the putative genes for expanding M13 tropism in *K. pneumoniae* and in *P. aeruginosa*, it was necessary to design appropriate expression cassettes that could be inserted into the previously obtained standard-compatible helper plasmid. To accomplish this, preliminary considerations were made. Firstly, we noticed that the metabolic burden imposed by this plasmid is significant, as it expresses a total of 12 proteins (11 from the phage capsid, plus the selection marker), which significantly prolongs colony growth (Figure 8). Secondly, our goal was to preferentially express alternative tropism genes and consequently incorporate them into the phage, as opposed to the wild-type gene product 3; for these reasons, a precise choice of the promoter and of the ribosome binding site (RBS) was taken.

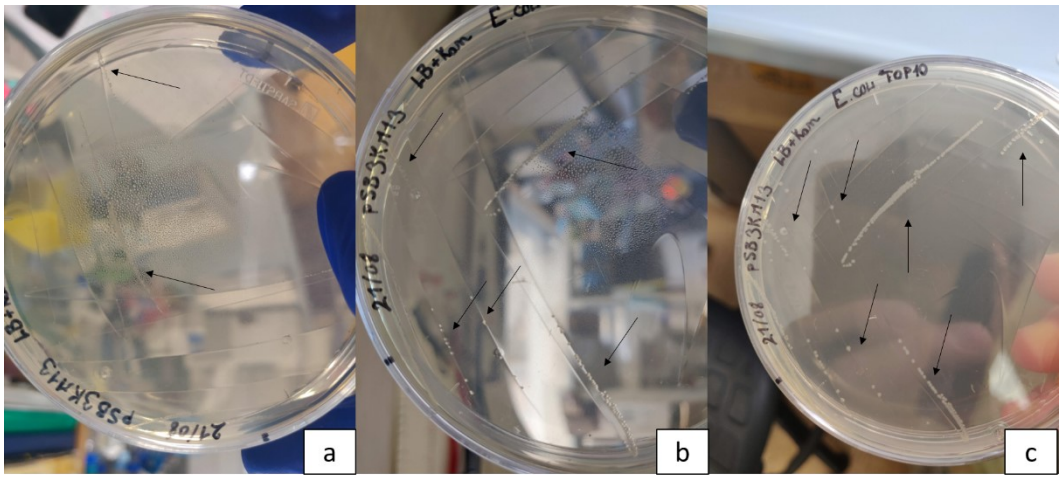


Figure 8. Metabolic burden of pSB3KM13. In the figure, *E. coli* TOP10 colonies bearing pSB3KM13 after different incubation times are reported. (**a**) After overnight incubation (16 hours), few and small colonies are visible on the plate. (**b**) After 24 hours the colonies were bigger and in a higher number. (**c**) After 36 hours, the colonies were clearly visible and could be easily isolated.

Given all these premises, for the construction of the expression cassettes, a strong but inducible promoter was chosen. This choice allowed the metabolic burden to be, if necessary, limited; at the same time, when fully induced, such promoter enables the high expression of the alternative gene, so that it could be preferentially incorporated into the phage particles instead of the wild-type Gp3 protein. The selected promoter was *pL-lac0-1*[17], widely used and compatible with the BioBrick standard, derived from the hybridization of the strong *pL* promoter of phage lambda with the *LacO* site of the Lac operon (see part BBa_J428041 in the Registry of Standard Biological Parts) (Appendix). This promoter, in fact, combines the strength of the *pL* promoter, with the possibility to regulate it, derived from the *LacO* binding site[17]. Indeed, if the cell expresses the *LacI* gene at high levels, it becomes feasible to induce the expression of the gene regulated by this promoter through the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to the culture medium. Moreover, by employing varying concentrations of the inducer, it is possible to finely adjust the expression level of the synthetic *gene 3*, thus creating a controllable system that is in line with the requirement of promoting cell vitality. Similarly, the choice of the ribosome binding site (RBS) fell on a strong element, in order to ensure that, upon transcript production, it is efficiently translated by ribosomes. This decision aligns with the aim of favoring the production and incorporation of this gene product over the wild-type gene 3 product (Gp3) of phage M13. Hence, the RBS chosen was compatible with the BioBrick standard and deposited in the Registry of Standard Biological Parts as BBa_B0034 (Appendix).

Finally, the terminator was selected based on its strength and efficiency, with a preference for an intrinsic terminator, independent from the expression of cellular proteins. Therefore, the T1 terminator of *E. coli* was opted for (see part BBa_B0010 in the Registry of Standard Biological Parts) (Appendix).

The two sequences, "*G3_Pf3*" and "*G3_cut*", designed as described above (Sections 4.2.1 and 4.2.2), were codon optimized before being *in vitro* synthesized. Codon optimization refers to the modification of base triplets that encode the same amino acid, in a way that optimizes the genetic code according to the characteristic frequency of codon usage in the target organism, and it is used in genetic engineering when expressing heterologous proteins in an organism to achieve ideal expression efficiency. In the case under consideration, codon optimization was performed at two levels, as further explained in the Materials and methods section: the first aimed to harmonize the genetic code to favor the expression of the encoded protein in *E. coli*, while the second aimed to optimize the sequence

for the *in vitro* synthesis, thus avoiding the presence of repeated sequences and hairpins.

In the end, to assemble the final cassettes, all the aforementioned elements were put together in the proper order, with some minor adjustments. For both constructs, the standard prefix and suffix were added at the 5' and 3' ends, respectively. The suffix was then followed by the *pL-lacO-1* sequence accompanied by the RBS. A start codon (ATG), coding for methionine, and a double stop codon (TAA) were added to the cassette expressing the Gp3_cut gene, whereas, to the cassette expressing the Gp3_Pf3, just a double stop codon (TAA) was added. Both the coding sequences were then assembled to be followed by the T1 terminator and the suffix and were finally named *G3_M13_Cut_optimized* and *G3_Pf3_optimized*, respectively (Appendix).

4.4 HELPER PLASMID EXPRESSING NEW TROPISM GENES

These expression cassettes were cloned into the helper plasmid pSB3KM13, previously obtained as described in section 4.1, and the resulting construct was then transformed into the *E. coli* MG1655 Z1 cell line, which is characterized by the constitutive expression at high levels of the regulatory genes *LacI*, *AraC*, and *TetR*. Consequently, this allowed the utilization of the regulated cassettes under suitable promoters in presence of the correct inducer, in this case IPTG. The correct insertion of the tropism expression cassette was assessed via gel electrophoresis, but unfortunately none of the samples resulted positive in our first experiment. However, once the helper plasmid expressing new tropism gene will be obtained, in order to produce phage particles with altered tropism it will be sufficient to co-transform it with an M13 phagemid (*e.g*. pTZ19R). The phagemid, due to the presence of the encapsidation sequence, can be inserted into the assembled phage particles. Consequently, it will be possible to generate phages carrying a desired genome with modified tropism.

In conclusion, our modifications propose a versatile platform for the simple and consistent production of phage particles, with the potential to easily modify tropism through interchange of standardized genetic components. Tropism-determining genes will become standardized parts, readily exchangeable among them, thanks to the new helper plasmid pSB3KM13. This platform, if combined with a phagemid of desire can allow the production of engineered bacteriophages carrying a genome of desire, able to infect specific bacteria of interest.

4.5 FUTURE PERSPECTIVES

In order to apply this work to the clinical practice, the consequent logical step will be the purification of bacteriophages with altered tropism and the assessment of their ability to target the desired cells. A possible experimental setup could see the production of phage particles bearing a phagemid carrying an adequate reporter gene (*e.g*. *RFP*) under a strong promoter suitable for the target bacterial species. This would allow to easily demonstrate the phage infection, as the bacteria will show the phenotype given by the reporter gene. Once the ability to transduce bacteria will be demonstrated, the reporter gene could be replaced with the desired system to interfere with the genes involved in AMR.

It is remarkable to notice that the helper plasmid pSB3KM13 we developed can represent a versatile platform to easily alter the tropism and to rapidly test new proposed variants. This strategy could specifically gain importance when considering non-characterized pathogens and new strains. It is important to point out that, even if the two proposed genes will work as intended, such an approach could be further explored by combining methods of rational design and random mutation complemented with a massive screening of the generated proteins. Indeed, thanks to the possibilities offered by the advancements in DNA synthesis technologies and their increasingly affordable prices, it will be possible to rapidly generate gene libraries for tropism determinant genes. These libraries can be obtained through random mutations in controlled regions, which can subsequently be easily screened using this platform. By utilizing the helper plasmid, the genes for tropism can be readily inserted and cloned into packaging cells to create phage libraries that express a diverse range of tropism determinants. These libraries can then be swiftly screened against the target pathogen to identify the clone expressing the right protein, which once isolate will be used to produce phage particles carrying the desired genome that showed efficacy in treating the specific bacterial infection. In the context of an ongoing battle against evolution in order to effectively fight antibiotic resistance, this platform could serve as a promising starting point. We think that the standardization of parts combined with the capability of new generation gene-synthesis can lead to an easy constant update of the platform here developed, making research in this field easier and less time-consuming than before.

5 BIBLIOGRAPHY

1. Dadgostar, P. Antimicrobial Resistance: Implications and Costs. *Infect. Drug Resist.* 12, 3903–3910 (2019).
2. WHO. Antimicrobial resistance SEVENTY-SECOND WORLD HEALTH ASSEMBLY, WHA72.5, Agenda Item 11.8, 28 May 2019.
3. Kortright, K. E., Chan, B. K., Koff, J. L. & Turner, P. E. Phage Therapy: A Renewed Approach to Combat Antibiotic-Resistant Bacteria. *Cell Host Microbe* 25, 219–232 (2019).
4. FAO. What is it? | Antimicrobial Resistance. *Food and Agriculture Organization (FAO) website visited on 2 July 2023*.
5. Ando, H., Lemire, S., Pires, D. P. & Lu, T. K. Engineering Modular Viral Scaffolds for Targeted Bacterial Population Editing. *Cell Syst.* 1, 187–196 (2015).
6. Yehl, K. *et al.* Engineering Phage Host-Range and Suppressing Bacterial Resistance through Phage Tail Fiber Mutagenesis. *Cell* 179, 459-469.e9 (2019).
7. Krom, R. J., Bhargava, P., Lobritz, M. A. & Collins, J. J. Engineered Phagemids for Nonlytic, Targeted Antibacterial Therapies. *Nano Lett.* 15, 4808–4813 (2015).
8. Citorik, R. J., Mimee, M. & Lu, T. K. Sequence-specific antimicrobials using efficiently delivered RNA-guided nucleases. *Nat. Biotechnol.* 32, 1141–1145 (2014).
9. Karlsson, F., Borrebaeck, C. A. K., Nilsson, N. & Malmborg-Hager, A.-C. The Mechanism of Bacterial Infection by Filamentous Phages Involves Molecular Interactions between TolA and Phage Protein 3 Domains. *J. Bacteriol.* 185, 2628–2634 (2003).
10. Nilsson, N., Malmborg, A.-C. & Borrebaeck, C. A. K. The Phage Infection Process: a Functional Role for the Distal Linker Region of Bacteriophage Protein 3. *J. Virol.* 74, 4229–4235 (2000).
11. Holliger, P., Riechmann, L. & Williams, R. L. Crystal Structure of the Two N-terminal Domains of g3p from Filamentous Phage fd at 1.9 AÊ : Evidence for Conformational Lability. *J. Mol. Biol.* 288, 649–657 (1999).
12. Lubkowski, J., Hennecke, F., Plückthun, A. & Wlodawer, A. The structural basis of phage display elucidated by the crystal structure of the N-terminal domains of g3p. *Nat. Struct. Biol.* 5, 140–147 (1998).
13. Levengood, S. K., Beyer, W. F. & Webster, R. E. TolA: a membrane protein involved in colicin uptake contains an extended helical region. *Proc. Natl. Acad. Sci.* 88, 5939–5943 (1991).
14. Knight, T. Idempotent Vector Design for Standard Assembly of Biobricks: *DSpaceMIT Commun. Synth. Biol.* (2003).
15. Chasteen, L., Ayriss, J., Pavlik, P. & Bradbury, A. R. M. Eliminating helper phage from phage display. *Nucleic Acids Res.* 34, e145–e145 (2006).
16. Marvin, D. Filamentous phage structure, infection and assembly. *Curr. Opin. Struct. Biol.* 8, 150–158 (1998).
17. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338 (2000).
18. Hatfull GF, Dedrick RM, Schooley RT. Phage Therapy for Antibiotic-Resistant Bacterial Infections. *Annu Rev Med*. 73, 197-211. (2022)
19. Russell J. Krom, Prerna Bhargava, Michael A. Lobritz, and James J. Collins Engineered Phagemids for Nonlytic, Targeted Antibacterial Therapies, *Nano Letters* 15 (7), 4808-4813. (2015)

6 APPENDIX

>pL-lacO-1 Protomoter BBa_J428041
AATTGTGAGCGGATAACAATTGACATTGTGAGCGGATAACAAGATACTGAGCACA

>RBS BBa_B0034
AAAGAGGAGAAA

>T1 Terminator BBa_B0010
CCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGG
TGAACGCTCTC

>G3_M13_CUT_OPTIMIZED
GTCAGAATTCGCGGCCGCTTCTAGAGAATTGTGAGCGGATAACAATTGACATTGTGAGCGGATAACAAG
ATACTGAGCACATACTAGAGAAAGAGGAGAAATACTAGATGGTGAAAAAGTTACTTTTTGCCATTCCGC
TGGTCGTACCATTTTATTCACATTCGGCAGAGACTGTTGAGAGTTGCTTGGCAAAGCCACATACTGAAA
ATTCTTTTACAAATGTATGGAAAGACGACAAGACCCTGGATCGCTACGCAAATTATGAAGGATGCTTAT
GGAATGCGACCGGCGTCGTAGTCTGCACTGGTGATGAGACCCAATGCTACGGAACGTGGGTGCCAATCG
GATTAGCGATTCCCGAGAACGAAAGTGAAGGTGGTGGTTCTGAGGGTGGAGGGTCCGGGGGTGGATCTG
GTTCAGGCGATTTCGACTATGAAAAAATGGCAAATGCAAATAAGGGCGCTATGACGGAGAATGCAGACG
AAAATGCTCTGCAATCGGATGCAAAGGGGAAGTTGGACTCTGTCGCAACTGATTATGGCGCGGCAATTG
ATGGTTTCATCGGCGACGTGAGTGGCCTTGCAAATGGAAATGGTGCCACAGGGGACTTTGCTGGCTCAA
ATTCGCAGATGGCACAAGTTGGTGACGGAGATAATTCTCCCTTAATGAATAACTTTCGTCAATATCTTC
CCTCTCTGCCACAGTCGGTTGAGTGCCGTCCGTTTGTTTTCAGTGCCGGGAAGCCTTATGAATTTAGCA
TCGACTGCGACAAGATCAATTTGTTTCGTGGTGTGTTCGCCTTTCTGCTGTATGTGGCGACATTCATGT
ACGTGTTTAGCACGTTCGCCAATATCTTGCGTAACAAAGAGTCTTAATAATACTAGAGCCAGGCATCAA
ATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTC
TACTAGTAGCGGCCGCTGCAGGTCA

>G3_PF3_OPTIMIZED
GTCAGAATTCGCGGCCGCTTCTAGAGAATTGTGAGCGGATAACAATTGACATTGTGAGCGGATAACAAG
ATACTGAGCACATACTAGAGAAAGAGGAGAAATACTAGATGGGATTACACTATCTTTTCGGGTTCTGCT
TGGCGCTGTTCTCGTTTTCCGCCATCGCCGCCGGACCAGTCAGTACGGAAGTGGCCGCCGGTACAACGA
CCTACCGTGTTACGAATACTACCGTCCGCACTCCACCAAACGTGACTCTTAGTCCCGTCCGTGATATCA
CACCATATGTAGAGAAAATCCCAAATAAGGGACTGGCACAGGCGGCACAAGGTCGCCTTATTGTGGCGC
AACGCGCCGCATCAGTCCCTGTTACAGGATTTTTCAATGTGTCTGGCGCTGTGGTCAAGAGTGGCGCGA
AATCGTTCCTGCGCTCGGCCGGACGCGCTAGTGGTATTGGCTTGGGCCTTGCCGCACTTTTAGAAGCCG
CAGATTGGGTTTTTGACGAGGAAGGAGAGATTGTAAAACCATTGCCTGGCGGTGGCTCCCCGGTGTTGA
TGCCCCGCCCCGTAATCCTGAACGAGTACACTGTTACTGGCTCCGCGGGGCAATGGAGCATCAGCAAAG
AATACGAACCCGATCCTCGCAGTGTTCCAGGGTGGTATTCGTACAATGGAAACCCAGTCTGGGTCTCGG
CGGTTGAAGATGTTGGCTTTACATGGCGCTACTGGTATTTCGCAGACGTATTAATGGATGGGCAAGGTC
GCCCCAATTATCTGGTCGCGTATAGCGACTCAGGCCCTAACGAATACTGGCAGGACGTAGGAGGGTATA
GTCTGGACTCTCTTCCGACTGAACCTGAATTTGTCCCTTTAACAGATGCGGAGTTGGAAGCCGGGATTG
ATCAGTACTACGAGCCTGATCCCGACGACTGGCGCAACCTGTTCCCGTATATTGAACCGGACTCGTTCA
CAATTGAGACTCCTATCCCCTCACTGGACTTGTCTCCTGTGGTCTCCAGCTCGACTAACAACCAAACCG
GAAAAGTTACTGTTACAGAAACAACAACATCCGTAGATTTTGAGGTATCCGACAACAACAGTAGTCAGC
CTTCCATCAGTGTGAATGAAACAACGACAGAAAATGTCTATGTGGACGGTGATCTGGTATCCTCAGAAA
CGAATACGACTGTCACTAATCCACCTTCGTCCGGGACAAGTACGCCCCCGAGTAGCGGTTCCGGGTCGG
ATTTCCAATTACCATCGTTCTGTTCGTGGGCAACTGCTGTCTGCGATTGGTTTGATTGGACACAAGAAC
CTATCGACGAAGAACCGGACCTGAGTGGTATCATTTCTGACATTGATGATCTTGAACGTACCAAAGATA
TCAGCTTCGGCTCTAAAAGTTGCCCTGCCCCAATTGCCTTGGATATCGAGTTTCTTGACATGTCGGTTG
ATTTGTCCTTTGAGTGGTTTTGTGAACTGGCTGGTATTATTTACTTTATGGTAATGGCGTCCGCGTATG
TGTTGGCGGCCTACATTACCTTAGGGGTTGTACGCGGATAGTAATACTAGAGCCAGGCATCAAATAAAA
CGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACTAG
TAGCGGCCGCTGCAGGTCA