Design of temperature inducible transcription factors and cognate promoters

by

Ralph McWhinnie
B.Sc., University of Victoria, 2005

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Biochemistry and Microbiology

Design of temperature inducible transcription factors and cognate promoters

by

Ralph McWhinnie
B.Sc., University of Victoria, 2005

Supervisory Committee

Dr. Francis Nano, Supervisor
(Department of Biochemistry and Microbiology)

Dr. Martin Boulanger, Departmental Member
(Department of Biochemistry and Microbiology)

Dr. Christopher Nelson, Departmental Member
(Department of Biochemistry and Microbiology)

Dr. Diana Varela, Outside Member
(Department of Biology)

Supervisory Committee

Dr. Francis Nano, Supervisor
(Department of Biochemistry and Microbiology)

Dr. Martin Boulanger, Departmental Member
(Department of Biochemistry and Microbiology)

Dr. Christopher Nelson, Departmental Member
(Department of Biochemistry and Microbiology)

Dr. Diana Varela, Outside Member
(Department of Biology)

ABSTRACT

The ability to control expression of a gene of interest is an important tool of molecular biologists and genetic engineers. This allows the phenotype associated with the regulated gene or genetic pathway to be partially de-coupled from the genotype and expressed only under condition that lend to induction of the genetic control system employed. Such control is typically implemented through a repressor protein (Eg. TetR, LacI) which will repress transcription when bound to a promoter containing a binding site (operator) recognized specifically by that repressor. Many such repressors and their cognate promoters are well-defined and characterized in model genetic systems, such as *Escherichia coli*, and may function poorly in other bacterial species. A lack of genetic components that allow the controlled expression of heterologous

genes in less well studied bacterial species may limit their bio-industrial potential and the sophistication of engineered phenotypes. The work presented here uses random mutagenesis and selection to isolate mutants of TetR that are inducible by increased culture temperature. Induction of protein expression by temperature change can have benefits over repressors that require small-molecule inducers in bio-industrial applications as reversal of induction and reuse of growth medium are possible. The host range of these, or any, repressor protein is limited by the host range in which its cognate promoter will function. To bypass this limitation and allow use of TetR in *Francisella novicida*, a method was developed by which TetR-responsive promoters that function in this host could be selected from random DNA sequence flanking the TetR binding site (*tetO*). Many unique TetR-repressible promoters that function in *F. novicida* were recovered and tightly-regulated expression of both exogenous reporter genes and host virulence genes were demonstrated. This promoter selection technique was also applied to *E. coli*, which allowed comparison between *F. novicida*-selected promoters and those selected in an *E. coli* host. Adaption of this process for production of promoters responsive to transcription factors other than TetR would simply require the use of a different operator sequence, suggesting diverse applications for this technique. This success in promoter engineering should enable advances in synthetic biology and genetic engineering in non-model bacterial species.

Contents

List of Tables

List of Figures

List of abbreviations used

RP_C	RNAP-promoter complex, closed conformation
RP_I	RNAP-promoter complex, intermediate structure
RP_O	RNAP-promoter complex, closed conformation
Tn*10*	Transposon 10
A	Adenosine
aa	Amino acid
aa-tRNA	Aminoacyl transfer ribonucleic acid
Ap	Ampicillin
AraC	L-arabinose operon activator/repressor
ATc	Anhydrotetracycline
ATP	Adenosine triphosphate
bp	Base pair
bps	Base pairs
C	Cytosine
CAT	Chloramphenicol acetyltransferase, product of *cat* gene
cDNA	Complimentary deoxyribonucleic acid
CDS	Coding sequence
Cm	Chloramphenicol
DNA	Deoxyribonucleic acid
dNTP	2'-deoxynucleotide triphosphate

dox	Doxycycline
dsDNA	Double stranded deoxyribonucleic acid
E	Ribonucleic acid polymerase core enzyme
Eσ	Ribonucleic acid polymerase holoenzyme
EC	Elongation complex
epPCR	Error-prone polymerase chain reaction
EZDM	Easy rich defined medium
FAM	6-Fluorescin
G	Guanine
Gm	Gentamycin
GOI	Gene of interest
GTP	Guanine triphosphate
HTH	Helix-turn-helix
iNTP	Initiating nucleotide triphosphate
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kD	Kilodalton
Km	Kanamycin
LacI	Lactose repressor
LacZ	β-galactosidase, product of *lacZ* gene
LB	Lysogeny broth
MW	Molecular weight
N	Any nucleotide (A, T, C or G)
NDP	Nucleotide triphosphate
nt	Nucleotide
NTD	N-terminal domain
NTP	Nucleotide triphosphate

nts	Nucleotides
ORF	Open reading frame
PCR	Polymerase chain reaction
PE	Promoter element
ppGpp	Guanosine tetraphosphate
Ap^R	Ampicillin resistant or ampicillin resistance
R	Purine (A or G)
RBS	Ribosome binding site
RNA	Ribonucleic acid
RNAP	Ribonucleic acid polymerase
RNAPσ	Ribonucleic acid polymerase holoenzyme
rRNA	Ribosomal ribonucleic acid
RT	Reverse transcriptase
rTetR	Reverse tetracycline repressor
SOB	Super optimal broth
SOC	Super optimal broth with catabolite
ssDNA	Single stranded deoxyribonucleic acid
T	Thymine
Tc	Tetracycline
TetA	Tetracycline efflux pump, protein product of the *tetA* gene
TetR	Tetracycline repressor
TF	Transcription factor
TFBS	Transcription factor binding site
TI	Temperature inducible or temperature induction
tRNA	Transfer ribonucleic acid
TSA	Tryptic soy agar

TSB	Tryptic soy broth
TSS	Transcription start site
U	Uracil
W	Adenosine or thymine (A or T
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside
YFP	Yellow fluorescent protein
Cm^R	Chloramphenicol resistant or chloramphenicol resistance
CTD	C-terminal domain
Gm^R	Gentamycin resistant or gentamycin resistance
k_a	Affinity constant
k_d	Dissociation constant
σ	RNAP sigma subunit (sigma factor)
tetO	Tetracycline operator (TetR binding site)
$TetR^{ti}$	Temperature inducible tetracycline repressor

ACKNOWLEDGEMENTS

I would like to thank:

Members of the Nano lab, past and present. I'll never forget the good times we had, both in and out of the lab.

Special thanks to Francis Nano: microbiologist extraordinaire. Thank you for allowing me the freedom to follow up on whatever crazy idea I wanted to try, and for the occasional nudges to get me back on course. Your guidance was invaluable and your patience legendary. I can't express strongly enough how much I appreciate having been a part of your lab.

DEDICATION

Dedicated to my mother, Karen McWhinnie. From all the trips to the library to check out science books as a child, to the letters with a $50 bill enclosed as an undergraduate student, you were there the whole way. None of this would have been possible without your love and support.

Chapter 1

Introduction

The control of gene expression is vital to all living organisms. An organism's genome may contain all genes required for survival of that organism, but this does not mean all genes are required, and expression of all are certainly not required simultaneously nor in equal amounts. For an organism to thrive in a dynamic and competitive world it must carefully control expression of each gene of its genome to maintain levels of each gene product (RNAs and proteins) near that which is optimal for survival given the current environment. All life today is the product of an evolutionary process by which the fittest are not only selected based on the genes they have gained, lost or modified but, also on their ability to express those genes at the right time in the right amount. A difference in gene expression profile is the difference between a caterpillar and a butterfly, or a rose's petal and its root. A bacterium expressing the enzymes needed to uptake and catabolize lactose will be ready to compete for that energy source in its presence, but a bacterium that produces these enzymes in the absence of lactose may quickly be out-competed by peers who do not waste cellular resources producing enzymes for which no substrate is present. Instead, a successful bacterium will adapt. It will sense its environment for available nutrient sources and produce enzymes required to make use of only these, starting with the most energetically favourable substrate. Once a specific substrate is depleted, expression of

these enzymes will cease and those required to utilize the next preferred substrate will quickly be expressed. Experiments by Monod and Jacob who observed this substrate switching of *E. coli* from glucose to lactose were an important foundation for studies of transcription control that have led not only to detailed molecular understanding, but also to the ability to co-opt such genetic control systems to allow regulation of genes other than those naturally regulated [1].

Engineered tuning of gene expression in response to specific external stimuli is used to aid in the study of gene function, reduce the metabolic burden imposed on bio-industrial protein over-expression strains, and is central to the creation of bioreporters, genetic logic circuits, and other innovations emerging from the new field of synthetic biology. The research presented here may aid these fields through providing improved genetic tools for controlling gene expression through the modification of genetic control elements (promoters and transcription factors) aimed at creating elements with broader host range and the capacity to modulate gene expression in response to new environmental inputs. More specifically, we have created mutants of the tetracycline repressor protein (TetR) which can be induced to de-repress the expression of a gene of interest in response to increased temperature, rather than its wild-type chemical induction signal, tetracycline (Chapter 3).

Although TetR's natural target promoter is known to function well in *E. coli* and closely related bacteria we intend to use these repressors to control transcription in other bacterial hosts, starting with *F. novicida*. However, a hurdle to exporting transcription factor (TF) function from one organism to another is the requirement for transcription promoters that are acted upon by that TF, but are also recognized by the transcriptional machinery of the new host. Unfortunately, *E. coli* promoters function poorly in *Francisella* (an anecdotal observation that we substantiate here experimentally); therefore, the natural TetR-regulated promoter, P_{tetA}, cannot be

applied to this host. To overcome this issue we designed a selection system to identify short DNA sequences that promote transcription in *Francisella* and are amenable to tight repression by TetR (Chapter 2). This method for creating synthetic, TetR-controlled promoter sequences proved to be extremely successful for generating regulated, as well as constitutive, *Francisella* promoters. We recognised that this technique could have more general application in generating regulated promoters in a variety of bacterial species. This caused our research goals to shift to closer examination of transcription control elements produced by this method and generated *E. coli* promoters in this way. Identifying of synthetic, *tet*-controlled promoters in an *E. coli* background, in addition to *F. novicida*, demonstrates the applicability of this method for generation genetic control elements in a range of bacterial species. Continuing work in our laboratory has successfully applied these transcriptional control tools to achieve temperature-induced gene expression in *E. coli* and chemically induced expression *F. novicida* with the goal of addressing biological problems.

This introductory chapter will review the general process of transcription in bacteria including specific examination of TetR and the regulatory signals that control its activity. After that I discuss the importance of genetic tools for the control and fine-tuning of gene expression as it relates to both basic research and biotechnology applications. My hope is to convey the challenges inherent in identification and development of transcriptional control elements that possess properties suitable for a given application, and how improvements in this area can support advances in basic research, genetic engineering, and synthetic biology. This will give context for the research presented in Chapters 2 and 3 which describe my work to produce regulated promoters that function in *F. novicida* and temperature-inducible repressors, respectively. Chapter 4 will discuss how these findings may be applied to advance new and exciting concepts and tools currently being developed by the synthetic biology

community.

1.1 Transcription in prokaryotes

This section will provide general background on the machinery and processes of bacterial transcription including an overview of bacterial promoter sequences and how they interact with the transcription machinery to direct this process.

1.1.1 The RNA polymerase complex

Transcription—the polymerization of ribonucleotides into an RNA of specific sequence as directed by a DNA template—is a process catalyzed by the multi-protein complex, RNA polymerase (RNAP). Although the machinery and mechanism of transcription share significant similarities across all domains of life, RNA polymerase is notably simpler in bacteria verses eukaryotes and archaea. Bacteria possess only a single version of the core RNA polymerase which is used to transcribe mRNA, rRNA and tRNA alike, while eukaryotes have three classes of RNAP, each responsible for transcribing different classes of RNA [2]. Archaea also rely on a single RNA polymerase, but this polymerase complex is much more closely related to the eukaryotic RNAPII than to the bacterial enzyme. Eukaryotic and archaeal RNAP complexes consist of at least ten subunits, compared to only five in the bacterial RNAP holoenzyme [2].

The five subunits of the bacterial RNAP core enzyme (denoted RNAP or E) consist of four different proteins: two identical copies of the alpha (α) subunit, the two large core subunits (β and β'), and the small ω subunit, for a total subunit composition of $\alpha_2\beta\beta'\omega$, as illustrated in Figure 1.1 [3, 4]. The α subunits are composed of two distinct protein domains separated by a flexible linker region. The N-terminal domains (αNTD) are involved in assembly and stability of the β and β' subunits while the C-terminal domains (αCTD) make contact with the DNA just upstream of the classic promoter

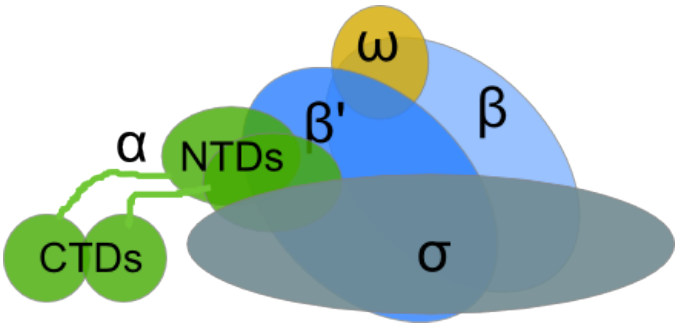


Figure 1.1: A diagram of the RNAP holoenzyme. The subunit composition and general organization of the bacterial RNA polymerase holoenzyme are depicted.

elements, and can be involved in promoter recognition through interactions with some transcription factors [5]. β' is the largest and main catalytic subunit in the complex. The cleft formed between β' and the related, slightly smaller, β subunit provides the active site responsible for catalyzing the polymerization process [6]. The ω-subunit was originally identified as a subunit of RNAP [3], but later experiments showed that its presence did not affect RNA polymerization in vitro and was presumed to be a contaminant of purification of the complex [7]. More recently, the status of ω has been restored to legitimate RNAP subunit with roles in β' stability and assembly of the core enzyme complex [8].

An additional subunit, called sigma (σ), or σ factor, is required for promoter recognition, but not catalysis of polymerization [9]. The core RNAP complex plus σ subunit forms the RNAP holoenzyme (RNAPσ or Eσ). A number of different σ factors can be present in an organism, each conferring RNAP the ability to recognize a different class of promoter sequence. These σ factors can compete with each other for a place in the holoenzyme; thus, a situation exists in which relative intracellular concentrations of the different σ variants can affect the gene expression profile of that cell on a global level [10]. The number of σ subunits encoded by a particular bacterial genome varies and can be as low as one in some *Mycoplasma* and *Ureaplasma* species [11], and as high as 30 in *Streptomyces coelicolor* [12]. *E. coli* K12 encodes seven

different σ factors [13], whereas *F. tularensis* is known to encode only two [14]. All bacteria have a main σ factor, σ^{70}, or σ^{A} in some families. σ factors other than σ^{70} typically participate in the transcription of genes involved in stress responses, so these alternative σ factors are mainly expressed in these specific stress situations. The majority of of bacterial σ factors are closely related and therefore recognize similar promoter sequence motifs, with differential promoter recognition of various σ factors provided by subtle differences in promoter architecture [10]. An exception is σ54 which is less closely related to other σ factors and recognises an unrelated promoter sequence [15]. The identity of the σ factor in Eσ should be an important consideration when measuring gene expression from specific promoters. For this reason, when considering σ^{70} promoters, measurements of promoter activity should be made in the same phase of growth, preferably when cultures are in exponential growth phase, before possible interference from alternative σ factors which may be activated due starvation stress encountered in stationary phase.

1.1.2 The process of transcription in bacteria

The study of prokaryotic transcription has used *E. coli* as a model organism almost entirely, but structural data for the complete $E\sigma^{70}$ had, until recently, only existed for E from thermophilic bacteria [6, 16, 17]. Recent structural data obtained for the *E. coli* $E\sigma^{70}$ complex has improved our understanding of this process while demonstrating the close structural similarities of the transcription machinery across bacterial species [4]. The transcription process can be considered to have five general steps: binding, isomerization, initiation, elongation, and termination. I will review the process of transcription in bacteria below with special attention on the first three steps. These steps leading up to elongation of the nascent polyribonucleotide chain can be expressed as an equilibrium where all steps are reversible until elongation is safely underway

$$\mathrm{R + P \rightleftharpoons RP_C \rightleftharpoons I_{1\text{-}3} \rightleftharpoons RP_O \overset{\mathrm{iNTP}}{\rightleftharpoons} RP_I \longrightarrow EC}$$

Figure 1.2: The steps of transcription initiation. R = RNAPσ, P = promoter, RP_C = closed promoter complex, I = intermediate, RP_O = open promoter complex, RP_I = initiation complex, EC = elongation complex. Note that all steps are reversible until formation of the elongation complex.

(Fig. 1.2).

The first step of transcription is promoter binding, where Eσ recognises specific DNA sequence motifs of the promoter and binds to the promoter DNA. The specific DNA sequences involved in this and other steps of the transcription initiation process are discussed in Section 1.1.3. Most, if not all, protein-DNA contacts are made by the σ subunit. This Eσ-promoter complex is called the closed promoter complex (RP_C) as the DNA is still closed in its helical form so has not yet begun to unwind. DNA footprinting experiments of promoters bound to Eσ trapped in RP_C reveal that Eσ^{70} protects the DNA from enzymatic digestion at base pairs from -5 to -55, relative to where transcription would begin—the transcription start site (TSS), or +1 position [18]. This RP_C structure is usually short-lived under cellular conditions as Eσ can dissociate from the DNA to revert back to the unbound Eσ and promoter (E+P) state or transcription can continue forward by unwinding of the DNA around the TSS [19].

This unwinding is induced by structural changes driven by the free energy of binding and is the beginning of the isomerization step [20]. Here the template strand moves into a channel of basic residues formed between the β and β' subunits, where the active site of polymerisation is located, while bases from -11 to -5 of the non-template strand make specific interactions with σ, replacing interactions lost between σ and the double helix form of the -10 hexamer [21–23]. These specific interactions between single stranded DNA (ssDNA) and various parts of the σ subunit and core RNAP stabilize what is now the open promoter complex (RP_O) and discourage reversal back

to RP_C. The greatest structural change in Eσ during isomerization involves closing of the "crab claw", a structure formed by parts β and β' (the "pincers") which, once template is loaded in the active site, clamp down around the DNA in a structural swing of 20°. This provides additional stabilization for the RP_O and is thought to aid in processivity during the elongation phase [24]. It should be noted that this model of the isomerization process is somewhat simplified, as biochemical analysis has identified at least two and maybe three distinct intermediate structures in this process [25]. It is unclear as to what events of isomerization occur at each of the proposed stages so they have been combined here for simplicity.

At this point the initiation stage is ready to begin. The initiating nucleotide triphosphate (iNTP), which is complementary to the +1 nucleotide (nt) of the template strand, becomes bound in the initiation pocket within the nucleic acid binding channel between the β and β' subunits. This site is close to, but distinct from, the active site that carries out the polymerization reaction as the iNTP is not added to a growing RNA chain by the same mechanism as subsequent NTPs. The iNTP is instead positioned so that its free 3'-OH group can be used as a starting point for polymerization, analogous to a primer in DNA replication [24]. This Eσ-promoter-iNTP complex is referred to as the initiation complex (RP_I). With the iNTP in place to act as a substrate for addition of the next nucleotide, polymerization can now begin. The first 9–12 nts are added to the new RNA chain without the polymerase leaving its position bound to promoter. This requires the enzyme to pull additional template DNA into the active site region while the upstream template and nascent RNA chain accumulate within RP_I [26]. This "scrunching" process can continue until the steric stress built up from the slack DNA becomes too great. At this point the initiation complex can either release the short oligonucleotide and revert back to RP_O—a process called abortive transcription—or Eσ can release from its position bound to the promoter, jettison the

σ subunit and continue with elongation of the nascent RNA [27].

This form of E bound to the DNA template and growing RNA chain is called the elongation complex (EC) or more informally as the transcription bubble, in reference to the unwound region of the helix forming a "bubble". As indicated in Figure 1.2, all steps leading up to elongation are reversible, but reversion can no longer occur once the EC is formed. During elongation the EC pulls itself along the DNA template adding additional ribonucleotides at a rate of about 40–80 nt per second in *E. coli* [28]. The energy for EC translocation comes not from NTP hydrolysis, as it does in DNA replication, but instead from the energy of binding the correct substrate in the active site by a mechanism referred to as a Brownian ratchet [29]. The process of elongation continues until the EC spontaneously dissociates, releasing the new RNA chain, or meets a transcription termination signal to force dissociation. Elongation is not continuous and may pause and backtrack many times during the transcription of a gene. Sites that are prone to pausing in *E. coli* have recently been identified and found to have a consensus sequence of $\mathtt{G}_{-10}\mathtt{G}_{-9}\mathtt{C/T}_{-1}\mathtt{G}_{+1}$ (on the coding DNA strand; positions are relative to the 3'-end of the elongating RNA chain) [30, 31]. This sequence is thought to impede translocation of the EC because it provides maximum stability to the RNA:DNA hybrid at the upstream end of the transcription bubble (to prevent separation of the nascent RNA from the DNA template and close the trailing end of the transcription bubble) and also maximum stability to the DNA duplex at the upstream end of the transcription bubble (to prevent separation of the DNA duplex to open the transcription bubble at its leading edge) [30].

Interestingly, this pause element (PE) sequence is consistent with the translational start site in prokaryotes. The `GG` in the basic Shine-Dalgarno sequence (`AGGAG` [32]) is found about seven nts upstream of the start codon (`ATG`, sometimes `GTG`, or more rarely `TTG`), which fits the consensus of the PE sequence described above. This pausing

of RNA polymerase near the translational start site has been proposed to allow a ribosome to load onto the RNA immediately behind RNA polymerase to directly couple transcription of a protein coding gene to the translation of that message [31]. Translational coupling is known to assist transcription. Ribosomal motion along the RNA is irreversible so the force of the translocating ribosome closely trailing the EC can prevent EC backtracking and "push" RNAP though pause element sequences and blockades made by DNA binding proteins [33]. Transcription elongation is also known to be aided by other RNAP transcribing the same DNA chain directly behind the leading EC [34, 35]. It is observed that promoter strength (rate of transcription initiation) is correlated with the rate of transcription elongation (rate at which ribonucleotides are added to a growing RNA chain). The mechanism for this also appears to involve the suppression of EC backtracking at pause sites as the trailing RNAP (which has not yet reached the pause site) can sterically hinder the leading EC from backtracking [34, 35].

Transcription termination signals typically punctuate genes or operons of genes to prevent transcriptional read-through past the gene (or genes) which are the target of the promoter from which transcription started. Without such stop sites, 'run-on' transcription could allow undesired transcription of adjacent genes or could produce transcripts complementary to those of neighbouring genes on the opposite strand, which can lead to degradation of both strands by nucleases that target double stranded RNAs [36]. Even in the absence of such deleterious polar effects, transcription much past the open reading frame would simply be a waste of cellular resources. In cases where the transcript is the final product (eg. rRNA and tRNA) termination of transcription at the correct site could be vital to function. In the absence of termination signals, the EC will polymerize and average of $> 10^4$ nts before the the EC dissociates by chance [37]. The need for specific transcription stop sites as the counterpart to each

promoter has occurred by two distinct mechanisms: intrinsic (Rho-independent) and Rho-dependent termination.

Intrinsic termination occurs when RNA polymerase encounters a GC-rich inverted repeat followed by a run of about eight A residues in the template strand [38]. An example of a typical intrinsic terminator sequence might be the λt_{R2} treminator, which has a sequence of GGCCTGCNNNNNNGCAGGCCAAAAATAA (template strand, inverted repeats are overlined with the A-run underlined) [37]. As transcribed into RNA the GC-rich inverted repeat forms the stem of a hairpin structure in the new RNA strand before this region of the RNA has completely cleared the exit channel of RNAP and while the polymerase is transcribing the run of A's immediately following. The transcription of A's to U's while this hairpin forms aids termination in two ways. First, it introduces a brief pause in transcription as polymerization of uracil ribonucleotides happens at a slower rate than the other ribonucleotides [39]. This allows more time for the hairpin to form before the polymerase complex has cleared the site of hairpin formation. Second, the A:U hybrid pairing between the A of the DNA template and the U at the new 3' end of the RNA transcript is inherently unstable due to exceptionally weak base-paring between adenine and uracil [40]. This allows the transcript to partially dissociate from the template in order to reduce steric limitations to formation of the adjacent hairpin [41]. The presence of this hairpin structure within the EC induces conformational changes in RNAP that weaken protein:RNA interactions and favours the dissociation of the EC [42].

The other mechanism of transcription termination is Rho-dependent termination, named for its requirement on the activity of the hexameric, ATP-dependent helicase, Rho. Rho-dependent terminators contain an *r*ho *ut*ilization (*rut*) site to which Rho binds with high affinity near the point of termination. The precise mechanism by which termination occurs was largely unknown until fairly recently when it was demonstrated

that Rho interacts with RNAP, the nascent RNA, and DNA of the termination site so that, as the EC reaches the termination site, steric stresses accumulate and induce structural changes in the EC [43]. This favours dissociation of the EC and release of the new transcript. Interestingly, *rut* sites do not appear to share any obvious conserved DNA sequence pattern across known Rho-dependent termination sites other than being C-rich and G-poor compared to surrounding sequence [44]. It is unclear how Rho is able to consistently and specifically bind and induce termination at these sites considering the lack of shared sequence identity that could act as a Rho recognition sequence. The absence of a definable Rho-dependent terminator consensus sequence creates a problem for identifying Rho-dependent termination sites in silico and makes rational design of this class of terminator impossible at this point.

1.1.3 The bacterial promoter

Promoters are the recognition and binding site for the RNAP holoenzyme. They act as beacons to direct and partition limiting units of transcription machinery to appropriate genetic locations with appropriate frequency so that the correct transcripts are produced in appropriate amounts. Generally, the relative frequency at which ribonucleotide polymersation initiates from a specific promoter—commonly referred to as promoter "strength"—is dictated by the sequence of DNA nucleotides at, and upstream of, the TSS and how they interact with Eσ and each other to influence Eσ binding, isomerization, and elongation initiation (also known as promoter escape). Essentially all (>95%) *E. coli* promoters are proximal to one or more transcription factor binding site (TFBS) [45]. Binding of a TF to its TFBS can cause an increase or decrease of transcription initiating from a promoter by a number of different mechanisms (Section 1.2.1). This section will consider promoter sequences and their role in transcription initiation in the absence of TFs.

All bacteria share similarities in general sequence motifs of promoters recognized by σ^{70} or σ^{A} containing RNAP holoenzyme, to varying extent [46–48]. This is surely a consequence of the strong similarity between the major sigma factor across all bacteria [11]. Still, sequence similarity does not necessarily equate to cross-functionality of promoters across different bacterial genera. For example, sequences known to strongly promote transcription in *E. coli* have been found to promote transcription poorly, if at all, in species such as *Bacillus subtilis* [49], *Synechococcus elongatus* [50] and *F. novicida* [51]. Despite this inability of many promoters to function across a range of hosts, large-scale examination of consensus promoter sequences have been predominantly conducted in *E. coli* [52, 53] with the findings often generalized as *the* bacterial promoter sequence. Although, considerable data also exists detailing sequences of *B. subtilis* promoters [54–56]. Attempts to identify promoter sequences of less well characterized bacteria often use bioinformatic tools trained to recognise sequence motifs identified in *E. coli*. It is interesting that this approach is effective, at least to some extent, even for bacterial species in which *E. coli* promoters do not function reliably [48, 57]. This issue of cross-species promoter sequence similarities with unpredictable cross-species activity is especially relevant to our study of synthetic promoters in *F. novicida* (Chapter 2). The following overview of bacterial promoter elements is based almost entirely on data derived entirely from *E. coli* promoters, as this is where most study has been focused. This asymmetric study of *E. coli* compared to other bacterial species is not unique to promoters; *E. coli* has historically been used to study many fundamental aspects of genetics and cell biology with the findings generalized to other bacteria, or even other domains of life. This situation has been elegantly summed-up by the eminent bacterial geneticist and physiologist Fred Neidhardt: "Not everyone is mindful of it, but all cell biologists have two organisms of interest: the one they are studying and *Escherichia coli*!" [58].

Sequences involved in promoting transcription in bacteria are composed of, in their most basic form, two AT-rich, hexameric sequences with 17 nts between them, referred to as the −35 and −10 promoter elements, respectively. Since these basic −35 and −10 elements were described, other common sequence motifs amount σ^{70} promoters have been reported upstream, downstream and between the main hexamers [59–61] (Table 1.1). These secondary promoter elements are not found in all promoters and are often present when required to compensate for deficiencies created by −10 and −35 sequences that deviate considerably from consensus [62]. This variation in sequence among promoters of different genes appears to be the result of a meticulous tuning process undertaken by evolution to define not just basal transcription levels for each gene, but also how each promoter will respond to TFs and environmental changes. Different promoter architectures can affect where the rate limiting step of transcription initiation lies and changes in intracellular conditions can differentially influence various steps of the transcription initiation process. Therefore, promoter architecture can importantly influence how expression of various genes respond to changing conditions.

The majority of sequence discrimination of promoters by Eσ results from protein-DNA interaction between the σ subunit and the −10 and −35 hexamers [4, 63, 64]. These elements are conserved to at least some extent in all σ^{70} promoters and most promoters recognized by most other sigma factors (with the exception of σ^{54}; [65]). The *E. coli* consensus for these sequences, as well as the percentage frequency of finding each consensus nucleotide at that relative position, are depicted in Figure 1.3 [63, 66, 67]. It should be noted that nucleotide sequences presented here are that of the coding (non-template) strand although both DNA strands play a role in transcription initiation [68]. What immediately stands out from the data presented in Figure 1.3 is that not all positions are equal in their relative level of conservation and that the

	−35 element					
	T	T	G	A	C	A
%	69	79	61	56	54	54

	−10 element					
	T	A	T	A	A	T
%	79	87	50	50	54	90

Figure 1.3: Consensus sequence of *E. coli* promoter −10 and −35 hexamer sequences. Numbers under each consensus nucleotide represent the percentage of promoters in which that nucleotide was found at that position. Data for −10 sequence is as reported by Mitchel et al. [62]. Data for −35 sequence is that reported by Lisser and Margalit [53].

−10 hexamer is more highly conserved than the −35 region overall. This is likely due to the central role some residues of the −10 hexamer play in multiple steps of the transcription initiation process [23]. This is also borne out experimentally as changing the nucleotide identity at highly conserved positions within the −10 hexamer away from that of consensus is typically much more detrimental to promoter function than changing nucleotide identity around less conserved positions of the −10 hexamer [69, 70]. To reflect this weighting of relative importance the −10 consensus can be depicted as $^{-12}$`TAtaaT`$^{-7}$. The uppercase letters represent highly conserved positions and superscript numbers represent consensus position of each end relative to transcription start.

Spacing between the −10 and −35 hexamers is also an important feature of promoter architecture. A 17 bp spacer is most common, but functional *E. coli* σ^{70} promoters have been identified with as few as 14 or as many as 19 bps separating the −10 and −35 hexamers [53] (Fig. 1.4B). However, varying the length of this spacer is not without consequences. A change of just ±1 bp from the consensus length has been shown to lower expression about 3-fold [71, 72]. Additionally, the spacer length may play a role differential promoter recognition by σ factors as transcription initiated by σ^S containing holoenzyme is much less sensitive to changes in spacer length than that of σ^{70} [73]. The importance of spacing between the primary hexamers is obvious when one considers that the 17 bp spacer allows for a 21 bp separation between the

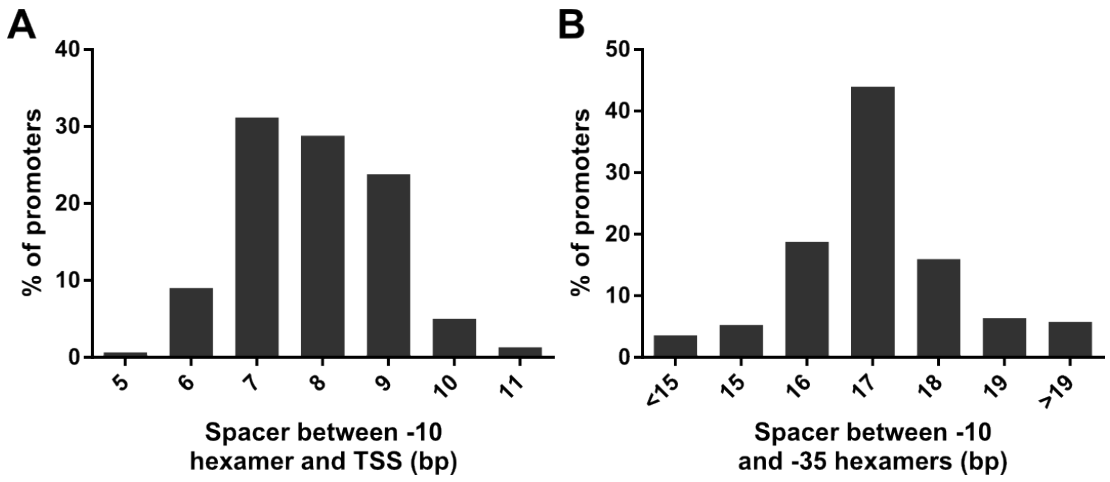


Figure 1.4: Spacing between core elements of *E. coli* promoters. A) Spacing between transcription start site and −10 hexamer. Spacing is in base pairs and represented as a percentage of total promoters analysed as reported by Mitchel et al. [62]. B) Spacing between −10 and −35 hexamers is represented as a percentage of total promoters analyzed as reported by Lisser and Margalit [53].

center of the −35 hexamer and the −11 position, where Eσ makes significant contact with the −10 element [68]. 21 base pairs make two full turns of the DNA helix in the B-form (10.5 bp per turn), which results in both promoter recognition elements facing the same side of the DNA helix where they would both be accessible to Eσ [74, 75]. The spacer region does not appear to make significant sequence-specific interactions with Eσ, which may explain why few obvious conserved sequence features are reported in this region. Exception to this is the −18 region which does interact with σ^{70} and exhibits a `T>A>C>G` preference [76, 77]. Other mutations within the spacer region have been observed to influence promoter strength [78, 79], which may be mediated though changes in DNA topography [76].

In addition to the primary −10 and −35 hexamers, secondary promoter sequence elements have been identified. These motifs are less conserved than the primary hexamers, but can still be vital for function of some promoters [80]. Examples of these are the UP element, the extended −10 motif and the discriminator. The consensus sequence of each is presented in Table 1.1. These secondary elements may compensate

Table 1.1: Promoter elements and their roles in transcription initiation. W= A or T, R= A or G.

Name	Consensus sequence	Role	Reference
UP element	$^{-57}$AWWWWWTTTTT^{-46} $^{-46}$AAAAAARNR^{-38}	RP_C formation, RP_O stability	[82–84]
−35 hexamer	$^{-35}$TTGACA^{-30}	RP_C formation	[52, 53]
Spacer	17 bps long	RP_C formation	[71, 85]
Extended −10	$^{-15}$TG^{-14}	RP_C formation	[61, 62]
−10 hexamer	$^{-12}$TATAAT^{-7}	RP_C formation, unwinding, RP_O stability	[52, 53]
Discriminator	$^{-5}$G^{-5}	RP_O formation	[86]
Transcription start	A≥G>T>C at +1	Initiation	[87]

for deficiencies in the sequence or spacing of the primary hexamers, and/or influence regulatory properties of the promoter [81]. The sequence and effect of these secondary promoter determinants are detailed below.

The extended −10 element is a region directly upstream of the −10 hexamer with a consensus of $^{-15}$TGn^{-12} [61, 62, 88]. This element is can be vital for promoters with a weak −35 hexamer [89] or increased spacing between the −10 and −35 hexamers [62]. Residues of σ^{70} interact directly with base pairs of the extended −10 motif along with the upstream most position of the classic −10 hexamer (usually a T at −12). For this reason it has been suggested that the motif $^{-15}$TGnT^{-12} be considered together with the −10 hexamer as the −15 element [62, 90]. This motif may to also aid in formation RP_O despite nucleotides of this region only contacting Eσ in double stranded form [91]. Gram positive bacteria have a more defined and conserved TGn motif at this position, which is recognized as a $^{-17}$TRTG^{-13} (R=A or G) extended −10 element in *Bacillus subtilis* [92].

Sequences between the −10 hexamer and TSS, including the TSS itself, can also be

important determinants of promoter function. Both length and sequence of this region will effect transcription initiation. The length of this region is a consequence of where transcription happens to initiate; although, since certain nucleobases are favoured at the +1 position (see Table 1.1) [53, 87], the distance from the −10 hexamer to the TSS will vary. A spacer of 7 to 9 bps is favoured in *E. coli* [93] (Fig. 1.4A). The importance of the region directly upstream of the TSS was identified early on when a common motif of $^{-5}\mathtt{C}\begin{smallmatrix}\mathtt{cc}\\\mathtt{gg}\end{smallmatrix}\mathtt{C}^{-2}$ was found in all 7 *E. coli rrn* promoters and designated the "discriminator" sequence [60]. However, this motif appears to be common to rRNA genes and not promoters in general. Wider analysis of *E. coli* promoter sequences reveals that nucleotides directly upstream of +1, in aggregate, actually have a C+G content lower than that of the genome as a whole. As promoters typically have low G+C content in this region, investigations into the effects of this area on promoter function focused on how a GC-rich stretch here might affect strand melting [94, 95]. Not surprisingly, high G+C was associated with lower rate of $\mathrm{RP_O}$ formation and reduced promoter activity in situations where isomerization is rate limiting. More recent studies have found a specific influence of the −5 nt (relative to the −10 element with the downstream most position of the hexamer defined as −7) [86], with `A` or `G` at this position greatly increasing stability of $\mathrm{RP_O}$. The presence of a C residue at −5, however, greatly diminishes $\mathrm{RP_O}$ stability. This effect has since been demonstrated to depend on an interaction between σ and the −5 position of the single stranded non-template strand, which is stabilized by a `G` residue and destabilized by `C` [21].

Another element recognised as important to function of some promoters is the UP element, an AT-rich region just upstream of the −35 hexamer (from about −40 to −60; Table 1.1) [59, 82]. Although the presence of an UP element can greatly enhance activity of certain promoters, few *E. coli* promoters are thought to employ this element. Promoters known to require an UP element for efficient activity include those for

rRNA genes, such as *rrnB* P1 [96]. A consensus sequence for this element has been reported (Table 1.1); however, as few promoters appear to include an UP element, an aggregate consensus sequence can not be easily determined through comparison of all promoters [82]. Estrem *et al.* [82] identified what they called a "functional consensus" through creation of a library of synthetic UP elements replacing the natural *rrnB* P1 UP sequence with random DNA sequence then screening for those that display strong promoter activity. Randomized sequences fused upstream of the −35 hexamer were able to maximize transcription from this promoter when they took the form of $^{-57}$`AAAWWWTWTTTTNNNAAA`$^{-40}$ (where `W`=`A` or `T`). This sequence has been divided further into proximal ands distal regions (Table 1.1) each bound separately by one of the two αCTDs (Fig. 1.1). These proximal and distal regions can function independently or in combination [83]. An UP element resembling that proposed by Estrem and colleagues [82, 83] is found in only ∼5% of *E. coli* promoters and different promoters are stimulated to different degrees by introduction of an UP-element sequence [97, 98]. Some promoters can even be negatively affected by the presence of an UP element [81]. The mechanism by which an UP element can diminish transcription at some promoters is proposed to result from reduced escape of RNAP from the promoter due to excessive binding [99], although others suggest that the UP element may, in some contexts and depending on its precise position, change overall DNA topography in a way that favors a non-productive open complex conformation [98]. The latter model is supported by data suggesting that presence of an UP element at some promoters favours transcription initiation by promoting DNA unwinding and RP_O formation, in addition to its role in enhancing binding affinity, likely through changes in DNA topography [84, 100, 101].

The preceding discussion illustrates how various promoter elements can affect the process of transcription initiation in very different ways. As such, a promoter

will evolve characteristics optimal for its role by combining sequences that conform to the consensus of various promoter elements to varying degrees. The overall rate of transcription initiation can only be as fast as the rate of the slowest step in the progression towards a competent transcription elongation complex (Fig. 1.2). The rate limiting bottleneck of a specific promoter can be placed at a different step of the transcription initiation process depending on the specific aspects of that promoter's architecture. As changing cellular condition can affect the rate of various steps of transcription initiation differently, a general mechanism emerges by which the cell can vary levels of transcription initiated from various promoters to meet the need for expression of different genes under different conditions. For example, a promoter that already binds Eσ poorly in the closed conformation, but is not limited by the rate at which RP_C is transformed to RP_O, or the rate of $RP_O \rightarrow$ EC would be disproportionally affected by changes in cellular concentration of Eσ as more free Eσ will increase the probability that Eσ will bind to that promoter to lessen the bottleneck of RP_C formation by the law of mass action. A promoter that binds Eσ strongly, but is limited by the rate at which the isomerization step occurs ($RP_C \rightarrow RP_C$), will see little change in overall rate of transcription initiation rate with increasing Eσ concentration. Specific promoter elements act to catalyze transcription initiation at different steps but sequence features which aid in one aspect may hinder another. For instance, a promoter that strongly binds free Eσ to form RP_C may have weak overall activity as stable interactions between promoter DNA and Eσ may provide energetic barriers to promoter escape. This may explain the observation that promoters which conform too closely to consensus have poor overall activity [102]. Successful promoters result from balance provided by complementary contributions for sequence motifs which reduce significant bottlenecks at any single step along the path to formation of a productive transcription elongation complex.

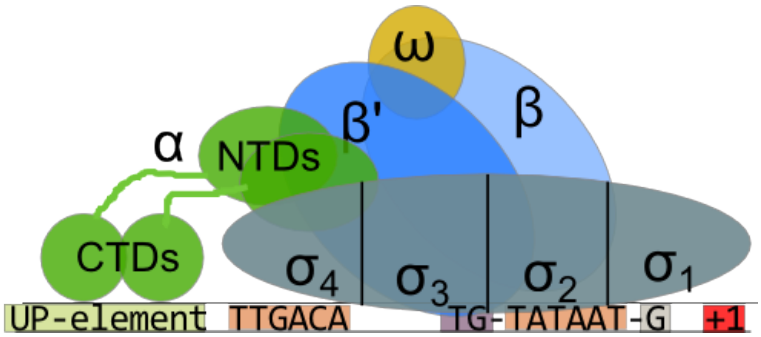


Figure 1.5: An illustration of promoter bound RNAP holoenzyme. Approximate areas of interaction between major promoter elements and Eσ^{70} are depicted.

1.2 Regulation of bacterial gene expression

As a cell encounters new surroundings it must adapt by altering its gene expression profile. Observations from as early as the late 1800s found enzyme activity within microorganisms could vary dramatically depending on growth conditions [1]. Experiments have since shown that protein expression and accumulation are modulated through changing rates of transcription, mRNA degradation, translation, or protein degradation. This section will review mechanisms by which bacteria alter RNA and protein expression levels, with focus on processes acting at the level of transcription initiation—where transcription factors act. Detailed background will be provided on a specific TF, the tetracycline repressor (TetR), as modification of this repressor, and creation of novel promoters repressed by it, are central to the work described in subsequent chapters.

1.2.1 Mechanisms of transcription regulation

An organism uses many different strategies to achieve differential expression of genes. These include alternative sigma factors, small molecule effectors, and sequence specific DNA binding proteins—activators and repressors. Alternative sigma factors are a convenient way for a cell to direct RNAP to a different set of genes by just switching

out a subunit of the RNAP holoenzyme, much like changing the bit on a multi-head screwdriver. Diverse mechanisms have evolved by which the limited pool of core RNAP is combined with different sigma factors to generate a pool of Eσ with a σ identity ratio appropriate to meet the cell's needs gene expression for its current environment [103]. A cell also has the ability to vary gene expression programs using the action of small molecules acting directly at the level of the promoter. This mechanism relies on variation of physical properties between promoters which change their response to different cellular conditions. For example, NTPs—a necessary substrate for the transcription of any gene—can also play a direct role in the changing the relative expression levels of specific subsets of genes, specifically those encoding components of the translation machinery (rRNA, tRNA and ribosomal proteins) [104]. As discussed in Section 1.1.2, promoters for these genes tend to form an unusually unstable RP_O complex due to a GC-rich discriminator motif just upstream of the +1 position. In the presence of excess initiating ribonucleotide (iNTP; usually ATP or GTP) this unstable RP_O does not severely limit the overall rate of transcription initiation as abundant iNTP can immediately react with RP_O after isomerization to push transcription into the irreversible elongation phase before RP_O collapses back into the more stable RP_C [104, 105]. However, declining [NTP] can disproportionally restrict transcription from this class of promoter as probability declines of finding the iNTP at the RNAP active site during the brief window before RP_O reverts. Most other *E. coli* transcripts are rate-limited at promoter binding or escape and are therefore not affected by declining [NTP] until levels drop much lower [106]. Guanosine 3', 5' bisphosphate (ppGpp) is another example of a small molecule effector of transcription initiation and acts to destabilize RP_O [107, 108]. ppGpp is synthesized in response to low amino acid concentration and can work synergistically with low [iNTP] to reduce expression from promoters with unstable RP_O forms. This is called the stringent response and provides

an early warning system for the cell to reduce translation capacity under conditions which will limit the production substrate for the translation machinery, mRNA.

Differential responses to changing concentrations of different small molecules between promoters have implications for heterologous protein over-expression and metabolic engineering. For instance, over-expression of a heterologous protein—especially that from a high copy number plasmid—not only uses nucleotides, amino acids and chemical energy resources; but also reduces the pool of available transcriptional and translational machinery. Reduced availability of RNAPσ will not affect expression from all promoters equally. Promoters from which transcription is rate-limited at the Eσ binding step will be negatively impacted to a greater extent by a reduction in the pool of free [Eσ]. This could greatly reduce host growth rate and provide strong selective pressure to favor the growth of individuals that have acquired mutations or genetic rearrangements which deactivate the offending transgenic parts.

Although alternative sigma factors and ribonucleotides are examples of transcription factors in the literal sense—factors that modulate transcription—this work is focused on TFs in the more classical sense: proteins that change expression of a gene though binding to specific DNA sequences (operators) in the vicinity of that gene's promoter to alter Eσ interaction with that promoter. These TFs may act at a single promoter but also include more promiscuous global regulators that act on many promoters throughout the genome. Genes regulated by a TF can be that of other TFs, producing complex regulatory networks and feedback loops. These regulatory DNA binding proteins are often modulated by an allosteric mechanism in which the presence of a small molecule or other factor will cause structural changes causing the loss or gain of competence for binding its operator, allowing dynamic control of gene regulatory activity.

Repression of gene expression by a repressor protein generally occurs by three

related, non-mutually exclusive mechanisms. The simplest is direct steric hindrance of Eσ where repressor bound to the promoter region shields the promoter from Eσ interaction (e.g. TetR [109] and LuxR [110]). A protein bound downstream of a promoter does not necessarily block transcription as it can be removed by the translocating EC. However, when bound at the promoter a repressor can stop transcription initiation by simply out-competing Eσ for DNA binding. Repression can also occur by changing the topography of DNA in the promoter region. For instance, a repressor may bind two or more operators separated by some distance along the DNA so that these multiple repressor copies interact while still bound to the DNA, pulling the two operator sites together to form a loop. This may change the three-dimensional topography of the promoter region so that it is no longer a suitable Eσ binding site (e.g. GalR [111]). The third general mechanism of repression is related to the first but more indirect. Here, a repressor binds to an operator overlapping the operator of an activator to block the activation of transcription at that promoter [112]. A repressor may combine aspects of these different mechanisms as epitomized by LacI repression of the *lac* operon. LacI binds to three operator sequences near the promoter region of LacZ: one overlapping the promoter to sterically block Eσ, one upstream of the promoter to block binding of the activator, cyclic AMP receptor protein (CRP), and one downstream of the promoter which cause DNA loops which changes local DNA topography in a way that further favours the repressed state [113].

Activation of gene expression often occurs by related but reversed mechanisms to those of repression. Presence of bound activator will increase the probability that Eσ will initiate transcription at that promoter. This can happen by direct interaction with Eσ to recruit the transcriptional machinery to the promoter, but can also occur through changes to DNA topography induced by activator binding that favours transcription initiation at nearby sites. Activation by direct interaction with RNAP often occurs

between the activator and the αCTD, as is the case for CRP activation of the *lac* operon [114] and CII activation of various bacteriophage λ genes [115]). An interesting example of indirect activation is seen in the MerR class of transcriptional activators which may bind between the −10 and −35 hexamers in a way that does not block Eσ binding but instead kinks the spacer to bring the primary hexamers into closer proximity so that they are properly recognized [116, 117]. Categorization of a TF as a repressor or activator is not absolute and many act to activate expression under one context while repressing in another. Although mechanisms of transcriptional control are varied and sometimes complex, the TF at the core of this work is the simple repressor, TetR, and its cognate operator, *tetO*. The following section will describe the properties this genetic control system in more detail.

1.2.2 Tn*10* encoded tetracycline resistance and transcriptional control elements

The tetracycline repressor used in this work is the class B variant ($TetR^B$) encoded by the transposable element, Tn*10*. The first reports of Tn*10* were as a tetracycline (Tc) resistance determinant of *Shigella flexneri* isolated in Japan in the 1950s. It was discovered as part of a multiple drug resistance plasmid (R-factor) able to mobilize to other bacterial species by conjugative transfer [118]. Studies of this R-factor revealed that the location of the tetracycline resistance marker was not always fixed and could sometimes be found as part of the host chromosome after passage [119]. Later, the genes involved in tetracycline resistance were found to be part of a mobile genetic element able to catalyse its own excision and re-insertion at a new genetic locus [120]. Tn*10* is 9,147 bp in length and contains seven genes flanked by inverted repeats of the insertion element, IS*10*, each encoding a functional transposase. Between the IS*10* elements are genes *tetA-D* and *R*, and *jemA-C* [121]. *tetA* encodes a Tc efflux pump (TetA),

which is responsible for conferring Tc resistance, while *tetR* encodes the repressor that controls transciption of *tetA* [122]. *tetC* and *tetD* turn out to not actually be involved in Tc resistance. Instead, TetD is a transcriptional activator of operons involved in resistance to redox-cycling compounds such as naladixic acid and norfloxacin [123]. TetC is a negative regulator of *tetD* transcription [124]. JemA, B and C show high similarity to a glutamate permease, an antibiotic synthesis monoxygenase, and a metallogulatory transcriptional repressor, respectively [121]. Although discovered in *Shigella* and studied extensively in *E. coli*, Tn*10* and its *tetR/tetA* Tc resistance cassette likely did not originate in a member of the *Enterobacteriaceae* family as it has a G+C content of 40%, which is consistent through the entire sequence and significantly lower than the ∼50% G+C found in *S. flexneri*, *E. coli*, and most other enteric bacteria [121].

TetA is an integral membrane protein that acts as an efflux pump which removes Tc from the cell by proton dependent antiport [125, 126]. In the absence of Tc TetA will reduce cellular fitness by interfering with the maintenance of membrane potential [127]. Its presence may also increase sensitivity to metal ions [128] and osmotic pressure [129]. Even low levels of TetA in the cytoplasmic membrane lowers the growth rate of *E. coli* and maintenece of *tetA* expression is selected against [130, 131]. To overcome these negative effects, this antibiotic resistance cassette has evolved a mechanism to produce TetA only when necessary. TetR binds with Tc (as a complex with Mg^{2+}, $[MgTc]^{+}$) to cause an allosteric shift in the structure of the repressor, abolishing DNA binding activity, de-repressing *tetA* transcription. Further constraints are placed on this system by the antibiotic activity of the molecule TetA is responsible for evicting. Tc is a direct inhibitor of 30S ribosome function; it binds to the A-site of the ribosome and prevents interaction with amino-acyl tRNAs [132]. As such, Tc has the potential to inhibit translation of *tetA* while enabling its transcription through inactivation of

TetR. This necessitates the system quickly de-repress *tetA* and pump Tc out of the cell before [Tc] reaches a level at which translation will cease.

Multiple features of the TetA/TetR Tc^R system combine to achieve tight repression of *tetA* expression while simultaneously allowing high sensitivity of induction. TetR affinity for Tc is exceptionally high ($k_a \approx 3 \times 10^9$ M^{-1}) [133, 134], >1000-fold greater than Tc ribosome binding ($k_a \approx 10^6$ M^{-1}) [135]. Sensitivity of *tetA* induction by Tc is enhanced low [TetR] maintained in the cell. Consequently, fewer Tc molecules are required to attain a 1:1 stoichiometry with TetR than would be the case for pairing with the large pool ribosomes [136]. TetR levels are sustained by a negative feedback autoregulatory mechanism in which TetR represses transcription of its own gene in addition to that of *tetA*. *tetR* and *tetA* are expressed from divergent, overlapping promoters, P_{tetR} and P_{tetA}, *both* of which are bound and repressed by TetR (Fig. 1.6) [137]. However, TetR is more proficient in repressing P_{tetA} than is the case for its own promoter; therefore, declining TetR levels will reach a point at which some transcription from $P_{(}$tetR) is allowed while P_{tetA} is still safely repressed. This allows more TetR to accumulate until levels at which full repression of both promoters is restored [122, 138]. Another feature of TetR is an extraordinarily high affinity for its operator sequence to that of non-specific DNA, even compared to other site-specific DNA binding proteins [139]. This allows the low number of TetR molecules to be positioned at their operator sequence within the context of large excess of non-target DNA present in the cell. TetR autoregulation also provides a mechanism quick re-establishment of *tetA* repression as intracellular Tc levels diminish. High relative levels of TetR are allowed to accumulate in response to induction by Tc. This decreases sensitivity of the system to Tc for the same reason that low [TetR] increases induction sensitivity. This facilitates repression of *tetA* expression to be switched back off at a greater [Tc] than was required to induce its expression originally [140, 141].

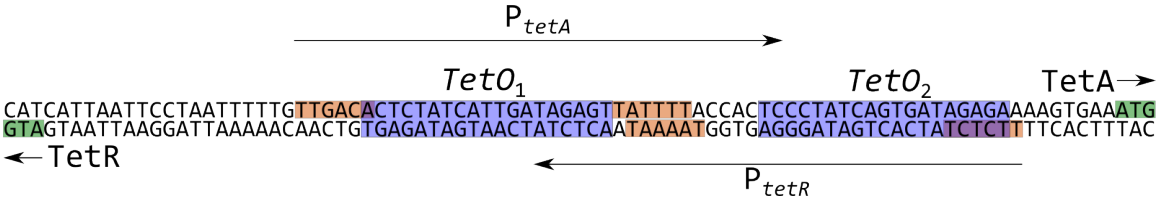


Figure 1.6: Sequence of the TetA/TetR promoter region. The position and direction of the *tetR* promoter (P_{tetR}) and *tetA* promoter (P_{tetA}) are indicated by arrows above and below the appropriate sequence. TetR binding sites (*tetO*) are highlighted in blue, promoter elements in brown, and start codons in green.

TetR functions as a homodimer of 46 kD subunits composed entirely of ten α-helical domains [142] (Fig. 1.7) [143, 144]. Each subunit consists of an N-terminal DNA binding domain (residues 1–45) and an effector binding domain responsible for dimerization and Tc binding (residues 46–208). DNA sequence discrimination and operator binding is mediated through a (helix-turn-helix) HTH motif, as is the case for many DNA binding proteins [145]. This domain was first described through X-ray crystallographic studies of the λ repressors, Cro [146] and CI [147], in the early 1980s. The HTH domain features a short "discriminator" α-helix that inserts into the major groove of the DNA helix and makes specific contacts with the nucleotide bases. Still, this recognition sequence could only account for specific interaction with about five bases, not nearly enough to account for specific binding as a recognition sequence of this length would appear thousands of times by chance in the chromosome. The solution to this problem came with the recognition of two-fold symmetry in both the operator sequence and the repressor (which was shown to be a dimer in these phage repressors, as it is in TetR). This allows the HTH motif of each subunit to make specific interactions with each side of the symmetric operator sequence, greatly increasing specificity.

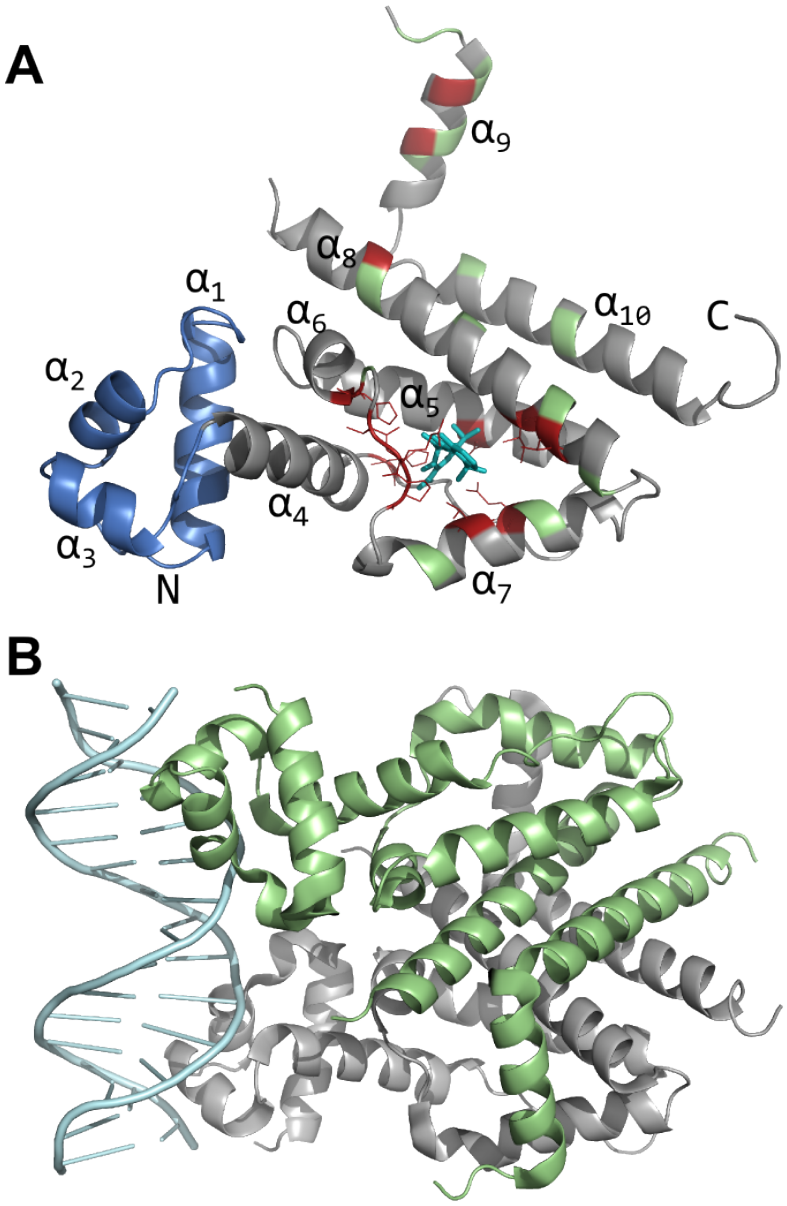


Figure 1.7: TetR protein structure. A) A single subunit of the $TetR^D$ homodimer with Tc bound. The DNA binding head (residues 1-50) is shown in blue and the effector binding domain in grey (residues 51-208). Residues involved in inducer binding are shown in red with some side chains also depicted. Residues that serve as dimerization contacts are green. X-ray crystallography data was obtained from data set PDP 2TCT [143]. B) Both subunits of the $TetR^D$ homorodimer in the uninduced state bound to *tetO*. X-ray crystallography data was obtained from data set PDP 1QP1 [144].

1.3 Importance of genetic control in the study and engineering of biological systems

The advent of recombinant DNA technologies in the 1970s opened the door to the possibility of expressing genes from a context different than that found in nature. This made possible fusion of an open reading frame of one gene to control elements of another, and researchers and bioengineers quickly exploited this technology to apply promoters and TFs to regulate the expression of genes other than their natural genetic targets [148, 149]. The first major commercial/industrial application for this technology—over-expression and purification of recombinant proteins—required gene expression regulatory tools for efficient protein production. High production levels of recombinant protein that is not needed by the cell will put significant metabolic burden on the host. Mutants that have lost the heterologous gene have a finess advantage over individuals who are forced to spend cellular resources on producing the recombinant protein. Systems were developed to repress expression of these genes so that protein production could be delayed until after the growth phase of the over-expressing culture [150, 151]. Early systems primarily employed control elements of the lactose operon (I.e. LacI/*lacO* with induction of gene expression (and recombinant protein production) by the lactose analogue, isopropyl β-D-1-thiogalactopyranoside (IPTG) [151, 152].

As genetically engineered organisms are developed with the ability to perform new and increasingly complex functions, a need has also arisen for a wider assortment of reliable genetic control systems. For instance, engineered bacteria have shown promise as bioreporters or biosensors that provide detection of a target chemical at very low concentration (e.g., an environmental pollutant such as naphthalene [153] or methyl halides [154]). Such bioreporter strains offer rapid and sensitive detection with reduced requirement for analytical equipment. To achieve this, genetic control

elements responsive to the target chemical are used to control an easily detected reporter gene. As these bioreporters rely on TFs responsive to the molecule to be detected and a promoter responsive to this TF functioning in the bioreporter host strain, processes that may improve the function of TF-regulated promoters in diverse hosts, such as that presented in Chapter 2, could be benefit further development of engineered microbes for such applications.

This idea of linking some input to expression of an unrelated, detectable phenotype (e.g., GFP fluorescence) has been expanded beyond bioreporters to create genetic circuits that act as logic gates to perform Boolean logic functions analogous to those performed by a computer [155–157]. The idea of "biological computing" is gaining popularity as a sub-discipline of synthetic biology, itself a new and ill-defined field that combines engineering principles with the tools of molecular and cell biology to create organisms with new, useful properties [158]. These systems have been used to integrate multiple input signals into output such as induction of artificial biosynthetic pathways or even output that can be stored in "memory" by instigating specific, permanent changes to DNA sequence [156]. The contents of memory could be output many generations later in response to a "print" signal (e.g. a chemical inducer) in order to produce a record indicating if the input was encountered in the past. As these circuits become more complex, an increasing number of TFs/promoter pairs are required to build additional input and output functions that do not cross-react or interfere with other elements operating the circuit or with natural host processes.

Engineered transgenic control parts also find application in basic molecular biological research. An obvious example of this is conditional gene knockout which is achieved by removing a gene of interest from its natural genetic context and instead place it under control of a TF-regulated promoter. In the repressed state the host should display a phenotype comparable to a strain with this gene completely inacti-

vated. Upon expression of the repressed gene under study, the phenotype controlled by this gene should be restored. This approach can be especially useful in study of genes essential for host viability. Complete knockout of such a gene would render the cell non-viable, but knockdown, rather than complete deletion, may allow study by changing the expression level while still supporting survival of the host [159, 160]. Conditional knockouts can also be used to examine effect of a gene expressed or repressed at various times during a biological process. Such an approach has been used to inspect aspects bacterial pathogenesis. This approach is epitomized in study of the role of type 1 pilus in urinary tract infections by an intracellular pathogenic strain of *E. coli* [161]. *tet*-control elements were employed to regulate a gene responsible for formation of the pilus. Repression of this gene did not impact early stages of infection; however, if induction was delayed past a certain time post infection, *E. coli* pathogenesis was severely compromised. This time overlapped with critical steps in biofilm establishment, which is required for host colonization, but not initial invasion. This information enabled rapid identification of the stage at which this gene product is crucial. The researchers could then move on to investigate specific molecular details of pilus interactions in host colonization [161].

Above examples illustrate the requirement for transcription control systems in a variety of applications. An important goal of the work presented in later chapters is to expand the "toolbox" of available genetic control elements—in this case, promoters and repressors—to facilitate construction of genetic circuits with improved function in a broader range of host organisms. The next chapter will describe the development and application of random DNA sequence, with a non-random *tetO* sequence, that can be subjected to a selection process by which promoters that exhibit TetR-mediated control properties are isolated and characterized. These promoters allow tight control of transcription in *F. novicida*, an organism for which few genetic regulatory tools

existed previously, and also in *E. coli*, a model organism of biological research. The following chapter presents the results of work in which mutants of the TetR were generated which allow for induction of a repressed gene by an increase in temperature. The last chapter attempts to tie together concepts and results presented up to this point with a discussion of the potential of these and other similar technologies in emerging molecular and synthetic biology applications.

Chapter 2

Identification of TetR-controlled promoters from semi-random synthetic DNA in *E. coli* and *F. novicida*

The results presented in this section are based on work previously published [51].

2.1 Introduction

The microbial world consists of an incredible range of organisms with specialized abilities and novel capabilities that have potential to be harnessed for important and valuable bio-industrial purposes. Engineering of properties into such organisms is often more difficult than would be for a thoroughly studied organism for which genetic tools (plasmids, regulatory elements, etc.) and other resources already exist. In many cases, genetic elements derived from one bacterial species will function similarly across bacteria of other families, but in the absence of cross-function with a model organism, tools specific to that host must often be developed. Effective engineering of such hosts often require suitable and reliable gene regulatory elements, such as TFs and their cognate promoters. In order for a TF to be useful in a given organism, the promoter controlled by this TF must also be recognized efficiently by the new host's transcription

machinery. If the promoter regulated by your desired TF, does not function acceptably in your host organism, there are potential work abounds to fix this. One approach to develop promoters for TF-mediated transcriptional regulation in a specific host is to modify natural promoters from the host of interest by adding one or more instances of operator sequence recognized by the target TF into the promoter. To achieve this, the DNA sequence of the operator is inserted in close proximity to the core promoter region, resulting acceptable results in many cases [162–166]. However, there can be drawbacks to TF-regulated promoters created in this manner. To successfully add TF binding to a natural promter one should have knowledge of at least the location of the -10 and -35 hexamers so that insertion of the operator can be designed to be unlikely to destroy RNAP binding motifs, but still close enough to the promoter elements so bound TF can physically interfere with RNAP recognition of that promoter (in the case where the TF is a simple repressor). A potential functional issue with this re-purposing natural promoters is the likely presence of binding sites for natural host TFs, or other undocumented control mechanisms that regulate expression of that promoter's original gene target. Such regulatory sequences are built into the vast majority of bacterial promoters which could result in inconsistent expression [167]. Furthermore, re-use of a natural control sequence would necessitate duplication of sequence with the host genome which can enhance genetic instability due to potential recombination due between these identical sequences. This can quickly lead to loss of function for an entire culture, especially if the transgenic construct applies appreciable metabolic burden or exhibits toxicity to the host [168, 169].

The work reported here avoids many of these downsides of promoter engineering by instead starting with just the TFBS (*tetO* in this case), and having it surrounded by random nucleotide sequence in a synthetic DNA fragment. Selection and screening was applied based on ability of these promoters to drive expression of an antibiotic

resistance and reporter gene. This facilitated the isolation and characterization of those fragments that happened to contain a DNA sequence that promoted transcription regulated by TetR in *F. novicida*. Bacteria of the genus *Francisella* are facultative intracellular bacterial pathogens found widely in nature [170]. Many *Francisella* biotypes infect a wide variety of animals, including humans, and are extraordinarily infectious and virulent. *F. novicida* (alternatively denoted "*F. tularensis* subspecies *novicida*") is normally non-infectious for humans but highly virulent in mice, thus is often used as a research model for human virulent *F. tularensis* species [171–173]. *F. tularensis* and *F. novicida* are closely related at the molecular level with nucleotide identity of about 98% [174, 175]. All of the molecular tools developed in one species appear to function in the other. Relatively little is known about control of transcription and the nature of promoters in *Francisella*. Analyses of genomic data from *Francisella* species have revealed that there are no complete two-component regulatory systems, only one alternative σ factor, and two distinct alleles of the α subunit of RNA polymerase [176, 177]. The presence of two divergent copies of the α subunit is unusual and may be unique to this genus [177, 178]. Both α subunit genes encode a functional protein which are both expressed in roughly equal amounts. The presence of both α variants has been confirmed in the *F. novicida* core RNAP complex as heterodimers; in fact, RNAP complex containing one of each α variant is the major, and perhaps only, form of RNAP detectable in *F. novicida* [178]. Generally, α subunits of RNAP are involved in recognition of DNA sequences upstream of the −35 promoter element, but it is not known if this interesting and unique characteristic the *Francisella* transcriptional machinery imparts function beyond that attributed to typical α homodimers found as part of other bacterial RNAP complexes.

Several studies provide evidence that promoters of antibiotic resistance cassettes functional in *E. coli* and several other bacterial species are unable to impart expression

of a resistant phenotype in *Francisella* [179–181]. For instance, a study involving transposon mutagenesis of *F. novicida* found that antibiotic resistant mutants could only be recovered in cases where the transposon had inserted downstream of a *F. novicida* promoter with an orientation that allowed the transposon's kanamycin resistance gene to be expressed from an adjacent host promoter [182]. Previous work has attempted to develop genetic control system in *Francisella*. Horzempa, et al. identified and characterized an endogenous promoter involved in expression of sugar metabolism enzymes with expression controlled by addition of glucose [183]. A TetR-regulated *Francisella* promoter was constructed by LoVullo et al. who inserted the *tet* operator sequence into the *F. tularensis groEL* promoter region and demonstrated repression by TetR [164]. Control of protein expression at the translational level has also been demonstrated in *F. novicida* and *F. tularensis* using a riboswitch responsive to theophylline [184].

Work presented in this chapter describes a method for selection of transcriptional control elements in *F. novicida*. Properties of TetR-regulated and constitutive promoter sequences isolated here are described. The strongest of these promoters had activity comparable to some of the strongest identified natural *F. tularensis* promoters. Synthetic promoters isolated in *F. tularensis* functioned in *E. coli*; however, random DNA sequences selected for promoter activity in a *E. coli* host were not able to promote transcription in *F. novicida*.

2.2 Methods

All experimental procedures presented in this section were performed by R. McWhinnie.

2.2.1 Culture conditions and transformation of bacteria

Unless otherwise indicated, *E. coli* strains were grown in modified lysogeny broth (LB; 1% tryptone, 0.5% yeast extract, 0.5% NaCl) or on LB agar, and *F. novicida* strains were grown in tryptic soy broth (TSB) or tryptic soy agar (TSA) supplemented with 0.1% L-cystine (TSBC or TSAC). Anhydrotetracycline (ATc) was used at 100 ng/ml, hygromycin (Hyg) at 150 µg/ml, chloramphenicol (Cm) at 5 µg/ml for *F. novicida* and 25 µg/ml for *E. coli*, and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) at 20 µg/ml, as required, unless otherwise indicated. Transformation of *F. novicida* was done as described previously [185]. Electroporation and chemical transformation of *E. coli* strains were performed by standard protocols [186].

2.2.2 DNA manipulations

PCR was performed using iProof High-Fidelity DNA Polymerase (Bio-Rad) for preparative PCR, or with Taq DNA Polymerase (NEB) for diagnostic PCR. Purification of DNA fragments was performed using NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel).

2.2.3 Strain and plasmid construction

Bacterial strains and plasmids used in this study are listed in Table 2.1. *E. coli* DH10B was used as the host for cloning experiments. The reporter plasmid pMP829-cat/lacZ was created by ligating the chloramphenicol acetyltransferase gene (*cat*; PCR product from pBC SK^+ template) and the *E. coli* β-galactosidase gene (*lacZ*; PCR product using BioBrick part BBa_I732017 as template, [191]) into pMP829 [190]. To create a plasmid expressing VgrG, the *lacZ* gene of pMP829-cat/lacZ was removed by digesting the plasmid with PstI and XhoI and a PCR product of the *vgrG* gene was inserted;

Table 2.1: Strains, plasmids and oligonucleotides used in this chapter

Strain or plasmid	Description	Source or reference
Strains		
F. novicida MFN245	*hsdRI hsdrII res drg*	[187]
F. novicida tetR$^+$	MFN245 *att*Tn*7*::P$_{bla}$-*tetR*$^+$ *res-aphA-1-res*; Km^R	This work
F. novicida Δ*vgrG*	MFN245 Δ*vgrG*	This work
F. novicida Δ*vgrG tetR*$^+$	MFN245 Δ*vgrG tetR*$^+$; Km^R	This work
F. novicida Δ*dotU*	MFN245 Δ*dotU*	This work
F. novicida Δ*dotU tetR*$^+$	MFN245 Δ*dotU tetR*$^+$; Km^R	This work
E. coli DH10B	F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ* ΔM15 Δ*lacX74 recA1 endA1 araD139* Δ*(ara leu)7697 galU galK* λ$^-$ *rpsL nupG*	Invitrogen
E. coli MGZ1	*E. coli* MG1655 F$^-$ λ$^-$ *ilvG rfb-50 rph-1*; chromosomally integrated Z1 cassette expresses LacI and TetR	[188]
Plasmids		
pMP720	Helper plasmid for mini-Tn*7* integration; Hyg^R	[189]
pMP749	Mini-Tn*7* *Francisella* integration vector; Ap^R Km^R	[189]
pMP749-tetR	pMP749 with *tetR* expressed from P$_{bla}$	This work
pMP823	*E. coli-Francisella* shuttle vector; P$_{bla}$; Hyg^R	[190]
pMP829	*E. coli-Francisella* shuttle vector; Hyg^R	[190]
pMP829-cat/lacZ	pMP829 with promoterless *cat* and *lacZ*	This work
pMP82-cat/vgrG	pMP829 with promoterless *cat* and *vgrG*	This work
pMP82-cat/dotU	pMP829 with promoterless *cat* and *dotU*	This work
pMP829-P$_X$-cat/lacZ	Series of plasmids recovered from *F. novicida* promoter selection	This work
pMP829-PE$_X$-cat/lacZ	Series of plasmids recovered from *E. coli* promoter selection	This work
Oligonucleotides		
BamHI-N48-tetO	`CACCTGACGTCTAAGAAGGATCC-(N×48)-TCCCTATCAGTGATAGAGA`	This work
BamHI-N30-tetOrc	`ATTACCGCCTTTGAGTGAGCGGATCC-(N×30)-TCTCTATCACTGATAGGGA`	This work
PE-cat-FAM	[FAM]-`CATTGGGATATATCAACGGTGGTATATCCA`	This work

the resulting plasmid was designated pMP829-cat/vgrG.

A *F. novicida* strain expressing TetR was created by inserting the *tetR* gene at the unique Tn*7* *att* site in the *F. novicida* chromosome. First, the *tetR* gene from Tn*10* was joined to the 0.5 kb upstream promoter region of the β-lactamase gene found in plasmid pMP823 [190] by fusion PCR [192]. This fusion product (P_{bla}-*tetR*) was ligated into the mini-Tn*7* integration vector pMP749 [190], and this kanamycin resistant (Km^R) plasmid was integrated into the *F. novicida* chromosome at the Tn*7* *att* site. As *F. novicida* readily takes up and integrates linear DNA fragments [189], chromosomal DNA extracted from *F. novicida* $tetR^+$ (Table 2.1) was used to transform *F. novicida* Δ*vgrG* and *F. novicida* Δ*dotU* to kanamycin resistance (Km^R) to introduce the Tn*7* locus with *tetR* into these strains to produce *F. novicida* Δ*vgrG* $tetR^+$ and *F. novicida* Δ*dotU* $tetR^+$.

2.2.4 Synthetic *tetO*-containing DNA libraries

Oligonucleotides BamHI-N×48-tetO and BamHI-N×30-tetOrc (Table 2.1) were annealed together at the complementary *tetO* regions and extended using Klenow fragment (NEB) to create a library of 141 bp fragments of dsDNA with AT-rich, random sequence flanking *tetO* (48 bp to one side and 30 to the other) with BamHI restriction sites at either end (Fig. 2.1). These random promoter fragments (RP-tetO) were cut with BamHI and ligated into the BamHI site upstream of the *cat* gene in pMP829-cat/lacZ so that *cat* and *lacZ* would be transcribed as a dicystronic unit. The ligation products were dialyzed against dH_2O by floating the mixture on a 0.025 µm VSWP filter (Millipore) for 2 hours to reduce the salt concentration. 15 µl of this product was used to transform *E. coli* DH10B by electroporation. After recovery in 1 ml super optimal broth (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 20 mM $MgSO_4$) with 1% glucose (SOC) for 1 hour the cells were spun down

and resuspended in 200 µl SOC and plated on LB agar containing 200 µg/ml Hyg. Incubation at 37° for 8 hours resulted in a light lawn of growth which was collected and plasmid DNA was isolated by mini-prep. This plasmid DNA was added to *F. novicida tetR*$^+$ and *E. coli* MGZ1 by chemical transformation. Hyg^R transformants were recovered for 1 hour in TSBC or SOC (for *F. novicida* or *E. coli*, respectively) containing ATc then plated on solid media containing Hyg, Cm, and ATc. Plates used for *E. coli* also contained X-gal; however, I found *F. novicida* to be sensitive to a cleavage product of X-gal [193], therefore,this indicator was not added to plates used for *F. novicida* growth. The resulting clones were picked into TSBC freezing media in 96-well plates containing Hyg [183]. After overnight growth at 37°C, clones were spotted to solid media containing Hyg, with or without ATc (*E. coli* plates also contained X-gal) then grown overnight at 37°C. *F. novicida* spot plates were assayed for LacZ expression by overlaying plates with filter paper soaked in an X-gal solution as described by Bucan et al. [193]. Colour was allowed to develop at 30° for 8 hours.

2.2.5 Chemoluminescent LacZ assay

β-galactosidase levels were determined using luminescence generated by LacZ cleavage of Galacton-Plus® (Galacto-Light Plus™ system; Applied Biosystems). Cultures were grown to mid-exponential phase in 96-well plates in TSBC with Hyg for *F. novicida*, and in EZ Rich defined medium (EZDM; Teknova) with 2% glucose and Hyg for *E. coli* MGZ1. *F. novicida* is naturally deficient in β-galactosidase activity, but *E. coli* MGZ1 harbours the full *lac* operon, but expression of this chromosomally encoded *lacZ* gene was suppressed by growth in defined medium (no lactose present) and with glucose as the carbon source (which suppresses induction of *lacZ* through inhibition of CRP). Cultures were induced with ATc for 2 hours before harvesting lysates for measurement of β-galactosidase activity, where appropriate. The OD_{600} of each culture

was measured immediately before lysis. *E. coli* cultures were lysed directly by adding 20 µl of culture to 70 µl of lysis solution (100 mM potassium phosphate, pH 7.8, 0.2% Triton X-100, 500 µg/ml polymyxin B sulfate). For harvest of *F. novicida*, culture microtiter plates were pelleted by centrifugation for 20 minutes at 4000×g and supernatants were removed before adding 70 µl of lysis solution to each well. 20 µl of lysate was added to 70 µl of reaction buffer in a white, clear bottom, 96-well microtitre plate (Griener Bio-One) followed by a 30 minute incubation at 30°. 100 µl of Accelerator-II (Applied Biosystems) was added to each well immediately before measuring luminescence for 1.0 second per well on a SpectraMax M5 plate reader (Molecular Devices). Luminosity values were normalized to culture cell density (OD_{600}) and a strain harbouring the promoterless control plasmid, pMP829-cat/lacZ, was subtracted from other readings.

2.2.6 Western blots

Cultures were grown to mid-exponential phase and ATc was added 2 hours before harvesting cells, where appropriate. 1 ml of culture was pelleted by centrifugation and resuspended in 25 µl cold dH_2O containing protease inhibitors ("cOmplete" protease inhibitor cocktail, EDTA-free; Roche) before adding 30 µl of 2×SDS loading buffer. Cultures were normalized based on cell density, separated by SDS-PAGE on a 12% polyacrylamide gel (10 µl lysate loaded per lane), transferred to nitrocellulose and blocked in Odyssey blocking buffer (LI-COR Biosciences). Primary antibodies were diluted in Odyssey blocking buffer containing 0.05% tween-20 and used at the following dilutions: rabbit anti-TetR (Abcam, ab14075; 1:1000), rabbit anti-CAT (Sigma-Aldrich, C9336; 1:1000), rabbit anti-VgrG (1:5000) (21). Primary antibody was detected using IRDye800-conjugated goat anti-rabbit (Rockland Immunochemicals) in Odyssey blocking buffer containing 0.05% tween-20 and 0.01% SDS (1:15,000) and

visualized on the Odyssey scanner (LI-COR Biosciences). Images were converted to grayscale and cleaned up using the Image Studio Lite software (LI-COR Biosciences; v5.0).

2.2.7 Mapping transcription start sites by primer extension

Cultures of *F. novicida tetR*$^+$ and *E. coli* MGZ1 harbouring promoter plasmids were grown in TSBC with Hyg (for *F. novicida*) and EZ defined medium (EZDM; Teknova) with 2% glucose and Hyg (for *E. coli*). Cultures were induced with ATc 1 hour before harvesting in mid-exponential phase. 0.5 ml of culture was added to 1 ml RNAprotect Bacteria Reagent (Qiagen) and RNA was isolated using the RNeasy mini kit (Qiagen). RNA was quantified spectrophotometrically and 6-carboxyfluorescein (FAM)-labelled cDNA was produced by reverse transcription using M-MuLV reverse transcriptase (NEB) with 5 µg of RNA as template, using the manufacturer's protocol in a reaction containing 20 U RiboLock RNase Inhibitor (Thermo Scientific) and FAM-labelled DNA oligonucleotide (PE-cat-FAM; Table 2.1). The resulting products were concentrated by ethanol precipitation, and resuspended in 10 µl HiDi formamide (Life Technologies) and 0.3 µl GeneScan 500 ROX size standards (Life Technologies). The mixture was heated to 95° for 5 minutes, cooled on ice for 1 minute, and subjected to capillary electrophoresis on an AB3730 DNA analyzer (Applied Biosystems). Data was analysed using GeneScan analysis software (Applied Biosystems).

2.2.8 Intracellular growth assay

J774A.1 murine macrophage-like cells were used to seed 96-well tissue culture treated plate at 5×10^4 cells/well in Dulbeccos Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 2 mM L-glutamine. Plates were incubated at 37° in a 5% CO_2 incubator for 12 hours to allow time for macrophages to adhere.

At this time, *F. novicida* strains to be assayed were added to the macrophages from cultures in exponential growth phase at a multiplicity of infection of 50 to 1. After one hour (t=0) wells were washed three times with phosphate buffered saline, pH 7.2 (PBS) with 10 µg/ml Gm, and fresh DMEM supplemented with 10% FBS, 2 mM L-glutamine and 2 µg/ml Gm was added to each well with or without ATc, as appropriate. Infected macrophages were lysed at various time points by washing three times with PBS before adding 0.1% deoxycholic acid in PBS. Lysates were serially diluted in PBS with 0.1% gelatin and spread on TSAC with Hyg at various dilutions to enumerate viable bacteria by plate count.

2.2.9 Creation of minimal *Francisella* promoters

Plasmids containing promoters P143, P146 and P165 were amplified by inverse PCR using 5'-phosphorylated primers situated so that polymerization extended in opposite directions around the circular template so that the entire plasmid was amplified excluding approximately 56 bp corresponding to the 5' end of the *tetO* sequence up to and including the upstream BamHI restriction site. The deleted region of each promoter was replaced by a 26 bp sequence added during the PCR process as 5' tails on the PCR primers and consisted of a randomly generated DNA sequence (generated using software found at `http://www.faculty.ucr.edu/~mmaduro/random.htm`) except for a unique PstI site, which allowed truncated promoters to be identified by restriction digest. Each resulting PCR product was ligated to itself to reform the circular plasmids which were now missing the upstream half of each promoter. Ligation products were used to transform *E. coli*, and the insert sequence verified before isolating plasmid DNA used to transform the appropriate strains of *F. novicida*.

2.2.10 Statistical analysis

Statistical analysis was performed using the Prism 5 software package (GraphPad Software, Inc.).

2.3 Results

2.3.1 Selection of synthetic promoters in *F. novicida*

A library of DNA fragments 97 bp in length (not including the flanking BamHI restriction sites) was created. Each has a 19 bp *tetO* sequence flanked by DNA of random sequence of 30 bp to one side and 48 bp to the other (Fig. 2.1). The randomized regions were designed with a 30% G+C content, slightly lower than the 32% G+C of the *F. novicida* chromosome [194]. These fragments were ligated into the BamHI site of an *F. novicida*-*E. coli* shuttle vector (pMP829-cat-lacZ) just upstream of the selective marker, *cat*. Functional plasmids produced by this reaction were selected and amplified by transformation of *E. coli* DH10B by electroporation. Plasmid DNA was then isolated from Hyg^R transformants. The pool of cells from which this plasmid DNA was recovered consisted of 1.1×10^8 individual Hyg^R *E. coli* colony forming units (cfu), as estimated by dilution plating.

A *F. novicida* strain constructed to constitutively express TetR from a chromosomal location was transformed with the preparation of selective reporter plasmid into which the library *tetO*-containing random fragments had been ligated. The TetR inducer, ATc, was added to the culture 2 hours before plating on solid media (also containing ATc) with both Cm and Hyg or just Hyg. About 1/200 the number of transformants were found on plates containing Hyg and Cm than observed on the plates with Hyg alone. This indicates that about 1 of every 200 plasmids resulting from the ligation between pMP829-cat/lacZ and the random DNA fragment library contained a fragment

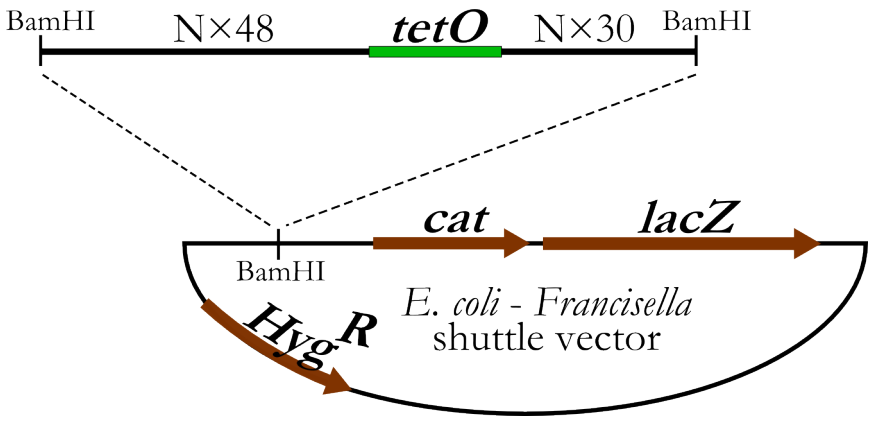


Figure 2.1: Diagram illustrating *tetO*-containing random DNA fragment and selection/reporter vector (pMP829-cat/lacZ). Semi-random DNA fragment is ligated upstream of a selectable resistance marker (*cat*) and a quantifiable reporter gene (*lacZ*). Plasmids that receive a DNA fragment carrying a sequence able to promote transcription in *F. novicida* can be selected based on Cm^R. The random fragment is asymmetrical due to off-center *tetO* placement and can insert into the vector in either direction with equal probability, creating directionality of inserted sequence with respect to the placement of *tetO* within the fragment.

capable of promoting transcription of the *cat* gene to a level sufficient to produce a Cm^R phenotype in *F. novicida*. This ligation used a vector with compatible sticky ends and roughly 50% of Hyg^R transformants result from intramolecular re-circulation of the vector. Therefore, about half of the Hyg^R transformants contain a random DNA fragment ligated into the BamHI site. Given this, and the 200 to 1 ratio of Hyg^R to Hyg^R + Cm^R phenotype observed for transformants, we can estimate that approximately 1% of random fragment inserts contain promoter activity (in the correct direction) strong enough to drive Cm resistance in *F. novicida* under the conditions present. A Cm concentration of 5 µg/ml was used for selection, which is well above the minimal inhibitory concentration determined to be approximately 1.5 µg/ml for *F. novicida* on TSA agar plates. Additionally, cells transformed with plasmid not ligated with random DNA fragments ligated upstream of the selective marker gene (pMP829-cat-lacZ) produced no colonies under these selective condition. This indicates that each Cm resistant transformant received a plasmid containing a random DNA

fragment capable of promoting transcription of the downstream *cat* gene in this host.

To visualize the relative transcriptional strength and to determine if these promoters are amenable to repression by TetR, we assayed the β-galactosidase activity in these clones produced from the *lacZ* reporter gene just downstream from *cat* on these plasmids. Interestingly, it was observed that Cm^R *F. novicida* clones (i.e. those with a functional promoter upstream of the *cat* and *lacZ* genes) were not able to form colonies on plates containing X-gal. In contrast, a clone carrying the promoterless parent plasmid (and therefore harbouring, but not expressing, the *lacZ* gene), did not display this sensitivity to X-gal. This may be due to a sensitivity of *F. novicida* to a breakdown product of LacZ cleavage of X-gal. A similar phenomenon was observed by Horzempa, et al. working with *F. tularensis* LVS [183]. Because of this sensitivity to an X-gal cleavage product, *F. novicida* clones were instead screened for TetR control of *lacZ* by exposing the cells to X-gal only after they had grown. To achieve this, spot plates were overlaid with a sheet of X-gal soaked filter paper after overnight growth. 9,216 Cm^R, *F. novicida* clones were picked from the transformation plate into 24 384-well microtitre plates containing liquid medium and grown overnight. These liquid cultures were then gridded onto new agar plates with or without ATc. Most wells produced spots of *F. novicida* growth on the solid medium, but many did not as about 15% of colonies picked by the robot did not result in bacterial cultures with obvious growth in microtitre plates after overnight growth. This is presumably because the colony picking robot failed to successfully transfer cells from the colony on the transformation plate into the liquid culture microtitre plate in these cases. Once spots of sufficient size had developed, the agar plates were overlaid with X-gal soaked filter paper. After allowing time for β-galactosidase to cleave X-gal into a blue product at 30°, clones covering a wide range of blue colour intensity were observed. This indicates that promoters isolated here represent diverse expression strength properties. Some

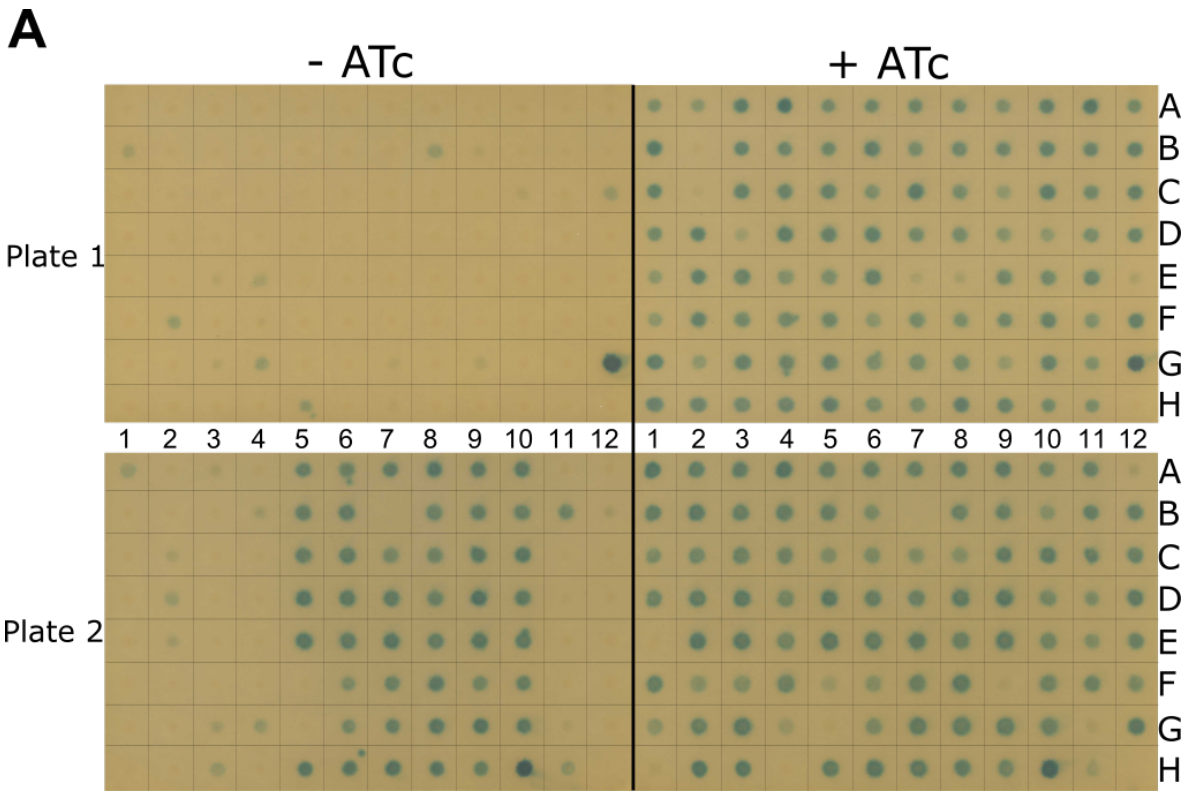


Figure 2.2: Spot plate LacZ assay for characterization of synthetic *tetO* promoters in 186 *F. novicida* clones. Clones chosen for further characterization were grown in two 96-well plates then spotted on TSA plates with and without ATc. After incubation overnight spot plates were exposed to X-gal soaked filter paper. Control of *lacZ* gene expression can be inferred from intensity of blue colour produced. In both plates, wells G12 and H12 are positive and negative promoter activity controls, respectively.

clones produced blue colonies only in the presence of ATc, others were blue under both conditions, while others produced little blue indicator under either condition (Fig. A.1).

The ~7000 Cm^R random *tetO* fragment promoter library provided more functional *F. novicida* promoters, both constitutive and *tet*-regulated, than could reasonably be subjected to further characterization. Images of each spot plate are presented in Appendix A (Fig. A.1). This pool of promoter variant strains was narrowed down to 186 individuals; a number of which will fit into two 96-well microtitre plates, including appropriate control strains. Clones were chosen over a wide range of

expression strengths, but with bias toward stronger promoters (as indicated by spots with intense blue colour in the presence of ATc on X-gal spot plate) and a bias for *tet*-controlled promoters that display tight repression of LacZ activity in the uninduced state (no obvious blue colour development of spots in the uninduced state) (Fig. A.1). Approximately two thirds of the *F. novicida* synthetic promoter clones selected for further analysis required the presence of ATc for the expression of this enzyme. The remaining random promoter clones exhibited LacZ activity with or without addition of ATc. For ATc induced promoters, those that produced little to no blue colour in the absence of ATc were favoured for selection over those that appeared to suffer from leaky expression of LacZ in the repressed state. New cultures were seeded in two 96-well plates from freezer stocks of the original 384-well microtitre plates, grown overnight, then gridded onto solid medium with and without ATc. Another X-gal spot plate assay of this smaller pool was performed. Plate 1 was composed of clones that displayed an ATc inducible, *lacZ* expression phenotype with wells G12 and H12 reserved for positive (pMP823-cat/lacZ; strong P_{bla} initiating *lacZ* expression) and negative (pMP829-cat/lacZ; promoterless) controls, respectively (Fig. 2.2, upper panel). Plate 2 contains clones with constitutive promoters in columns 5–10, with clones carrying *tet*-responsive promoters making up the rest of the plate. wells H4 and H12 contain positive and negative control strains (same clones as wells G12 and H12 of plate 1), respectively (Fig. 2.2, lower panel).

15 of these clones (10 TetR-regulated and 5 constitutive) had β-galactosidase expression levels measured quantitatively in the presence and absence of ATc (Fig. 2.3). Previously characterized *F. tularensis* LVS promoters, Pbfr and PZ12 (promoters for bacterioferritin and a tRNA gene, respectively) [57], ligated into the same genetic context as the synthetic promoters, were included as controls and as an internal reference. Expression of these *Francisella* promoters have been measured previously as

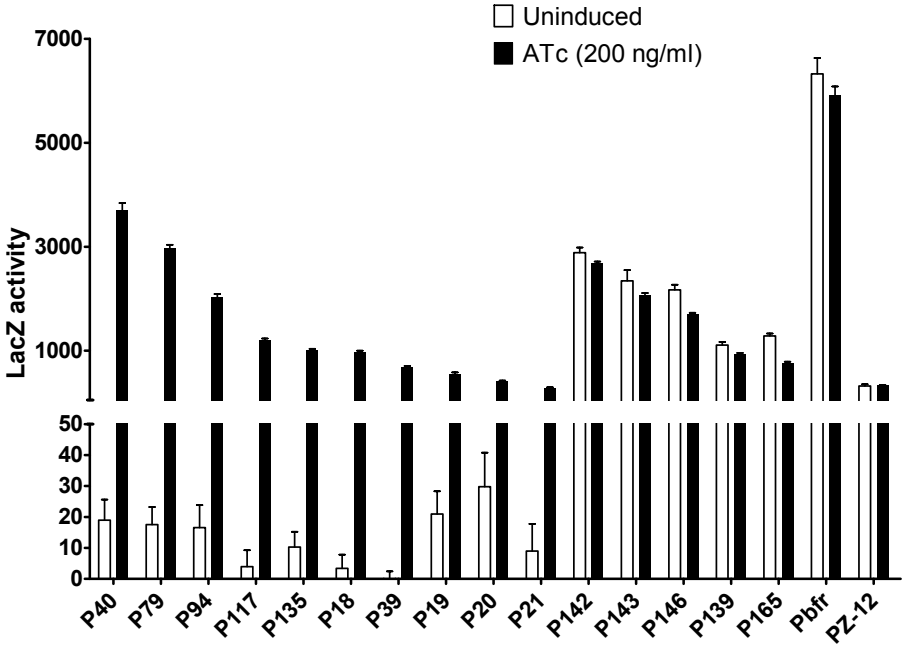


Figure 2.3: Expression from synthetic promoters in *F. novicida* as measured by LacZ assay. Values represent arbitrary β-galactosidase activity units over that of the promoterless *lacZ* reporter plasmid, pMP829-cat/lacZ. For induced condition ATc was added 3 hours before harvesting cultures. Natural *F. tularensis* promoters, Pbfr and PZ-12, are included for comparison [57]. Error bars represent S.E.M. Table 2.2 presents induction ratios for each promoter presented here and lists the position of these clones in the X-gal spot plates presented in Figures A.1 and 2.2.

part of a study that surveyed many *Francisella* promoters; therefore, knowledge of how expression strength of these reporters rank compared to our synthetic promoters, and how Pbfr and PZ12 compared to the other natural *Francisella* promoters characterized by Zaide, et al., should provide a standard by which our synthetic promoters can be compared indirectly to that larger pool of characterized *Francisella* promoters [57]. P12 and Pbfr were synthesised for inclusion in this study based on the reported sequence from nucleotide positions −80 to +20 relative to the previously identified transcription start site [57]. These 100 bp fragments were then cloned into our reporter plasmid with the same BamHI restriction sites as used for our random fragments.

LacZ assay results revealed that Pbfr was stronger than any of the 15 synthetic promoters for which expression was quantified (Fig. 2.3). However, the stronger synthetic promoters displayed activity comparable to, or greater than, the majority of natural promoters identified and characterized by Zaide et al. [57]. For comparison, the PZ-12 promoter (designated in the original paper as "12" but referred to here as PZ12 to distinguish from promoters identified in this work) was the fourth strongest natural promoter identified by Zaide et al., and about twice as strong as an average-strength promoter designated as "strong" by those researchers [57] (Fig. 2.3). Of the *tet*-regulated synthetic promoters, P40 presented the strongest expression with an induction ratio of 195 ± 68 (ratio of expression in the induced state to that in the uninduced state). P20 was the least impressive of those quantified with an induction ratio of only 14.0 ± 5.1. The limited difference seen between induced and uninduced expression is due to both weak expression in the induced state and significant leaky expression when repressed (Table 2.2). Curiously, *F. novicida* carrying synthetic constitutive promoters, as well as those with natural *F. tularensis* promoters, showed a slight decrease in β-galactosidase activity with ATc present compared levels in the absence of this inducer (Fig. 2.3). ATc does have anti-translational activity like its

analogue, tetracycline, but ribosome binding properties of ATc are reduced about 30-fold compared to Tc and would not be expected to interfere noticeably with translation at these concentrations [134].

Ligation of the *tetO* random BamHI DNA fragments upstream of *cat* and *lacZ* in the selection vector was non-directional. As the inserted fragments were asymmetric with respect to the length of random sequence to either side of *tetO*, promoters that are the product of the selection process (fragments able to promote a Cm^R phenotype) will be found in one of two directions: forward (*tetO* found closer to downstream end of the random fragment) or reverse (*tetO* closer to upstream end of the random fragment), with respect to the direction of the downstream *cat* gene (Fig. 2.1). Another possibility is ligation of multiple random fragments to each other. Concatenation of two or more random fragments to each other should be likely given that the random fragments were present at a 3×molar excess to the plasmid during the ligation reaction. These new longer fragments will still have a BamHI overhang on both ends and therefore still able to ligate into the BamHI site of the plasmid. Random promoter fragments from the 186 clones were sequenced (clones compiled earlier into two 96-well plates and used for the X-gal spot assay in Figure 2.2 are the same clones for which sequence data was collected). Three promoters failed to produce usable sequence data so a total of 184 random promoter fragments were collected (Table A.1). Of these, 14 were found to to be exact duplicates of others within the data set. Six of the sequenced clones had not produced any discernible blue colour in even the uninduced state on the X-gal spot plates shown in Figure 2.2. These duplicates and non-expressing clones were excluded from further analysis, leaving 166 functional promoters of unique sequence. As expected, promoters were found to be the result of the random DNA fragments ligating into the BamHI site of pMP829-cat/lacZ in either the forward (*tetO* is closer the 5' end of the fragment) or reverse orientations. A

Table 2.2: Expression data for *F. novicida* TetR-regulated synthetic promoters with and without induction by ATc as measured by LacZ assay.

Promoters	X-gal spot plate position		β-galactosidase activity[‡]		Induction ratio
	Original[*]	Narrowed[†]	+ATc	−ATc	
Inducible					
P40	7–B23	1–A4	3703 ± 138	18.0 ± 6	195 ± 69
P79	14–O17	1–H8	2972 ± 66	17.5 ± 5.7	170 ± 55
P94	17–C11	1–G10	2019 ± 72	16.5 ± 7.3	122 ± 54
P117	22–B17	2–B2	1202 ± 33	3.94 ± 5.3	305 ± 411
P135	24–K16	2–D4	1011 ± 27	10.3 ± 4.9	98 ± 47
P18	1–E5	1–C1	973 ± 28	3.38 ± 4.5	288 ± 382
P39	6–I24	1–H3	676 ± 31	-1.96 ± 4.4	N/A
P19	1–H14	1–D1	555 ± 29	20.9 ± 7.3	26 ± 9.4
P20	2–A2	1–E1	416 ± 8.7	29.8 ± 11	14 ± 5.1
P21	2–A17	1–F1	284 ± 12	8.98 ± 8.8	32 ± 31
Constitutive					
P142	2–A23	2–D5	2692 ± 25	2888 ± 98	0.93±0.03
P143	2–H5	2–E5	2069 ± 41	2345 ± 205	0.88±0.08
P146	4–A19	2–H5	1692 ± 39	2170 ± 99	0.78±0.04
P139	1–K1	2–A5	925 ± 33	1108 ± 62	0.83±0.06
P165	13–K6	2–C8	762 ± 26	1284 ± 47	0.59±0.03
Natural *Francisella* promoters					
Pbfr	N/A		5923 ± 162	6327 ± 306	0.94±0.05
PZ-12	N/A		337 ± 3.4	321 ± 31	1.05±0.10

[*] Location (represented as 'plate #'–'well position') in which this clone can be found in original X-gal *F. novicida* promoter screen presented in Figure A.1.

[†] Location (represented as 'plate #'–'well position') in which this clone can be found in X-gal spot plate of the *F. novicida* narrowed down promoter pool presented in Figure 2.2

[‡] Expression LacZ reporter by luminescent LacZ assay in arbitrary β-galactosidase activity units ± S.E.M.

Table 2.3: Regulatory properties of *F. novicida* selected synthetic promoters categorized by size and orientation *tetO*-containing fragment present.

	Regulatory capabilities		
Promoter organization	TetR-responsive	Constitutive	All
Single fragment			
Forward	56	0	56
Reverse	26	25	51
Multiple fragments	39	14	53

full third of the synthetic promoters analysed here are result from fragments fusing by ligation (Table 2.2). Interestingly, all 56 promoters found in the forward orientation were susceptible to control by TetR (Table 2.3). The promoters represented by the numbers in Table 2.3 were not chosen randomly, and disproportionally composed of strong, TetR-responsive, tightly repressed (at least relative to other promoters identified by the original screen for promoter function, but that could not explain how zero constitutive promoters could be found in the forward orientation compared to 56. Instead, the difference is likely due to limitations in sequence length present when the 30 bp side of random sequence is oriented to the downstream of *tetO* (ie. the forward orientation). In this orientation, it is possible that a random fragment is found with a sequence appropriate to promote transcription of *cat* so the transformant to survives selection. However, there may be no possible sequence that can promote transcription *and also* provide enough spacing between the promoter and *tetO* to create a situation in which constitutive expression can occur. Synthetic promoters with multiple fragments yielded a greater number of TetR-responsive promoters than constitutive promoters by 39 to 14 (Table 2.3. Sequence analysis of the multiple fragment promoters did not reveal any obvious bias in orientation of the fragments that make up these larger promoters when comparing TetR-responsive to constitutive.

Expression of CAT protein levels produced by select synthetic promoters in the presence and absence of ATC were examined by western blot (Fig. 2.4). Expression of CAT driven by these various promoters appears to be consistent with the relative levels of LacZ activity that was determined previously for those clones. Physical detection of CAT protein accumulation by immuno-blot, and LacZ levels inferred from its enzymatic activity show consistency across *F. novicida* clones carrying very different promoter sequences from which they express these gene. This, together with the fact that *cat* and *lacZ* are co-transcribed and therefore must have identical transcript levels, strongly indicates that expression detected for these reporter proteins provided a reasonable indicator of transcription levels initiated from these promoters. Promoters that previously exhibited inducibility with ATc by β-galactosidase activity (P20, P39, P40, P94 and P135) all showed TetR control of CAT expression by western blot. A small amount of CAT expression in the absence of inducer was for observed for inducible promoters P39 and P40. Transcription from P142 results in CAT expression regardless of the ATc presence, as is expected from previous observations of LacZ activity when downstream of P142. P146 and P165 produced CAT in the absence of ATc, at levels in agreement with that found by LaZ assay (Fig. 2.3).

2.3.2 Promoter control of the *F. novicida* virulence factors VgrG and DotU

The genes *cat* and *lacZ* and their protein products are convenient tools for genetic selection and visible reporting of enzyme function, respectively, but both are foreign to *F. novicida* have little relevance to the biology of this bacterium. To demonstrate applicability of these synthetic promoters in expression and repression of genes native to *Francisella*, plasmids were created in which *lacZ* of pRM829-cat/lacZ was replaced by two different genes required for pathogenicity in *F. novicida*: *vgrG* and *dotU*. This resulted in plasmids pMP829-cat/vgrG and pMP829-cat/dotU, respectively

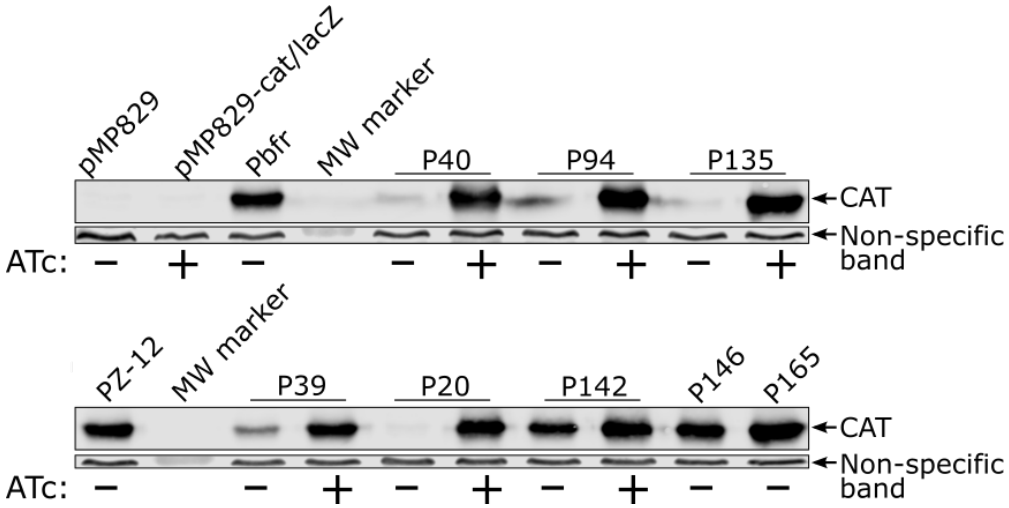


Figure 2.4: Immunoblot analysis of *tet*-controlled CAT expression from select synthetic promoters in *F. novicida tetR*$^{+}$ in the presence and absence of ATc. Natural *Francisella* promoters, PZ12 and Pbfr, are included for comparison. ATc was added to cultures 3 hours before harvest, where indicated. A non-specific band from the same blot is included to demonstrate even loading.

(Table 2.1). Two different *tet*-controlled synthetic promoters, P40 (strong) and P18 (medium-weak)(Fig. 2.3), were fused upstream of each *cat*-virulence gene operon in the exact context these promoters are found in when upstream of the *cat-lacZ* operon. This should result in plasmids which feature *tet*-control of *vgrG* or *dotU* expression together with expression of *cat* as well, which is co-transcribed with the virulence gene. The VgrG and DotU proteins are both constituents of the type VI secretion system encoded by the *Francisella* pathogenicity island (FPI) and expression of both are required for *Francisella* virulence [195]. These plasmids were introduced into a *F. novicida* hosts with the chromosomal copy of *vgrG* or *dotU* deleted, as appropriate, in strains both with and without a chromosomal copy of *tetR* (Table 2.1). Western blot analysis confirmed that P40 and P18 produced VgrG expression in a TetR-dependent manner (Fig. 2.5). Repression of P40 by TetR may not be complete as a small amount of CAT appears to be present in the absence of ATc (Fig. 2.5A). Expression of DotU from P40 and P18 followed a similar pattern to that observed for VgrG (Fig. 2.6).

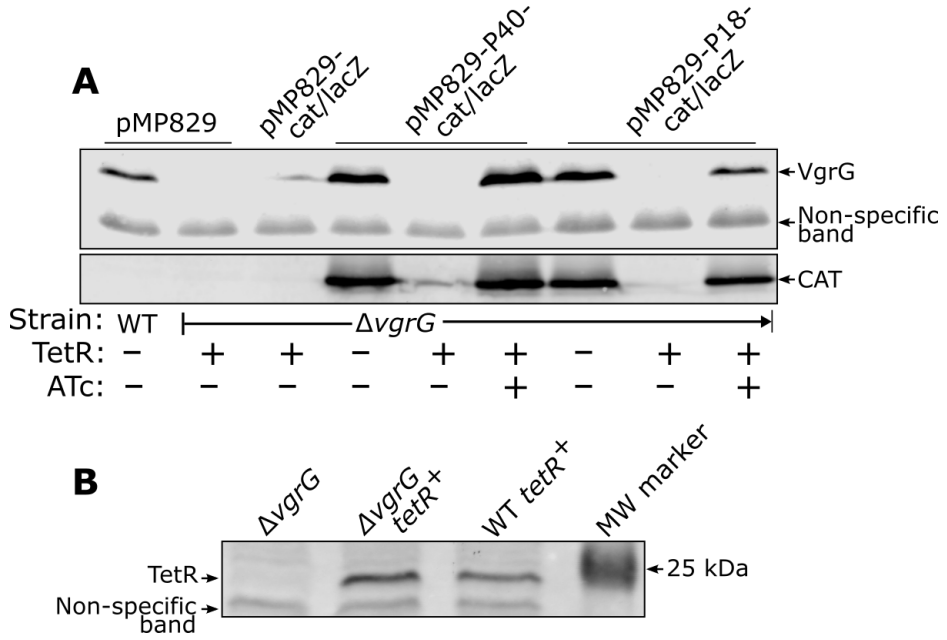


Figure 2.5: Immunoblot analysis of *tet*-controlled expression of VrgG and CAT from synthetic promoters, P40 and P18, and confirmation TetR accumulation in *F. novicida*. A) Expression of plasmid encoded VgrG and CAT in *F. novicida* Δ*vgrG* both in strains with and without *tetR* and ATc. *F. novicida* encoding its wild-type *vgrG* is included for comparison. B) Detection of TetR (23 kDa) in various *F. novicida* stains used in this work.

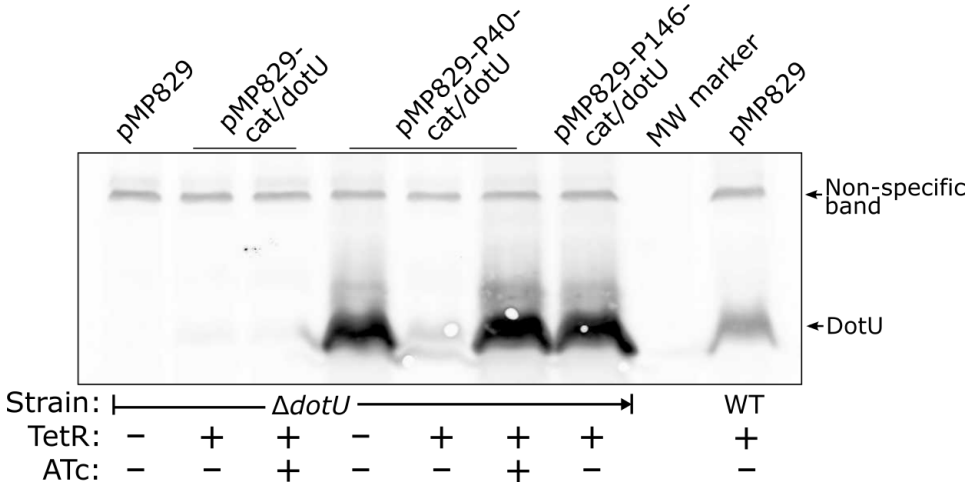


Figure 2.6: Immunoblot analysis of DotU expression from a *tet*-controlled and a constitutive synthetic promoter in *F. novicida*. The *dotU* gene was placed downstream of synthetic promoters, P40 (strong, *tet*-controlled), and P146 (strong, constitutive) in a genetic background with *dotU* deleted from the chromosome. DotU levels were probed in strains with and without a copy of *tetR*. *F. novicida* WT and a strain with a promoterless *dotU* are included as controls.

As TetR repression of CAT accumulation appears to be incomplete when expressed from P40, small amounts of VgrG may also be produced with *vgrG* expressed from the same promoter, although, any leaky expression must fall below detection limits of the VgrG immuno-blot as performed here and presented in Figure 2.5. A potentially more sensitive and biologically relevant assay for low basal VgrG concentration could be through detection of partial restoration of the intracellular growth phenotype of a *F. novicida* $\Delta vgrG$ mutant by low amounts of VgrG. In the complete absence of VgrG or DotU, pathogenicity of *F. novicida* should be abolished completely as measured by an established model of *Francisella* pathogenesis which follows growth of bacteria within cultured macrophages where pathogenesisity of different mutants will be relative to the number of viable cells of that strain recovered from within the macrophages. We found that *F. novicida* $\Delta vgrG$ $tetR^+$ carrying *vgrG* under control of P40 displayed poor growth in macrophages in the absence of ATc, but the intracellular growth phenotype was restored to near that of WT cells with ATc present (Fig. 2.7). However, intracellular growth of P40-*vgrG*, as determined by cfu recovered from macrophages, was obviously higher than the full *vgrG* knockout strain. Importantly, the number of *F. novicida* colony forming units (cfu) recovered from cells carrying a promoterless *vgrG* were almost indistinguishable from those with the full knockout, indicating that no physiologically relevant transcription is occurring in the promoterless vector. Therefore, any expression observed with the promoter present while in the repressed state must be due to incomplete repression by TetR and not due to transcription initiating weakly from a different location (e.g. within the upstream *cat* coding sequence).

In contrast to P40, P18 control of *vgrG* allowed no pathogenicity phenotype over that of a promoterless control as measured by *F. novicida* intra-macrophage growth. Induction of transcription from P18 by ATc addition allowed a wild-type like growth

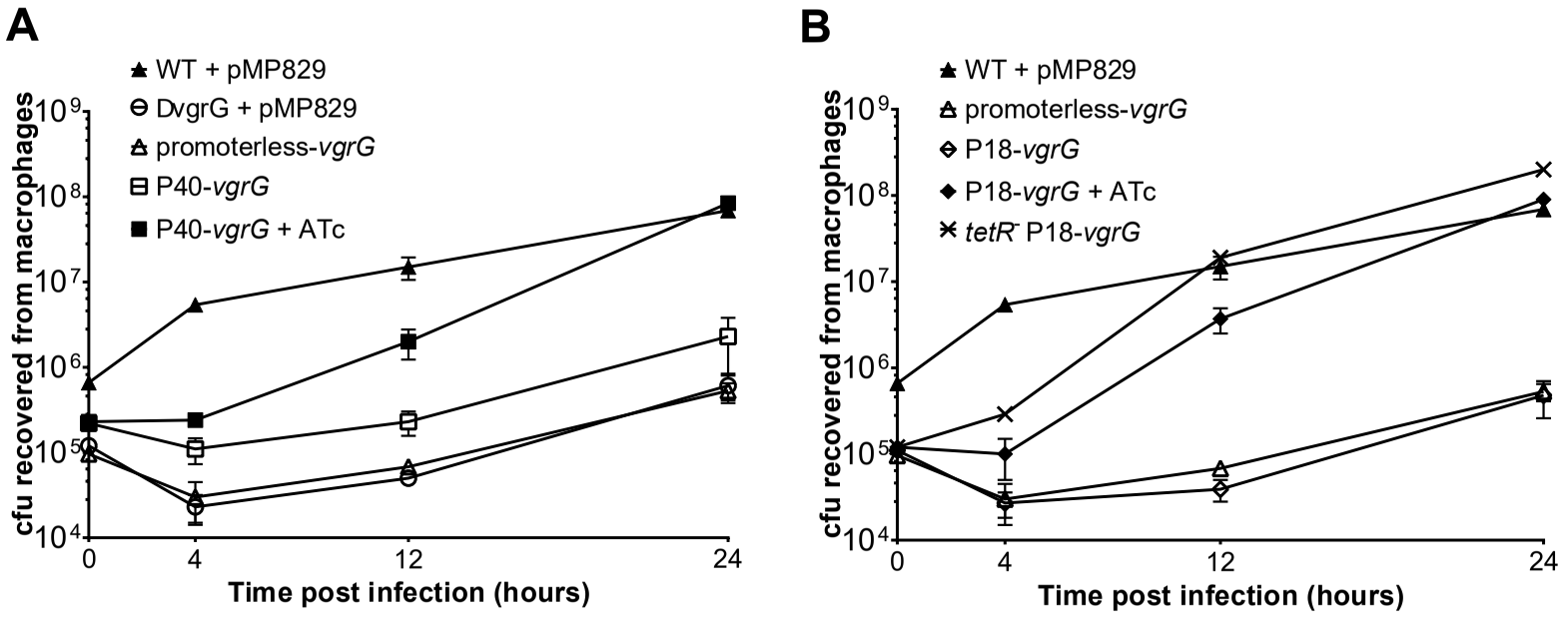


Figure 2.7: Intra-macrophage growth of *F. novicida* with *vgrG* expression controlled by synthetic promoters. A plasmid-borne copy of the virulence gene, *vgrG*, was expressed from synthetic promoters P40 (strong) or P18 (medium-weak) or a promoterless control (pMP829-cat/vgrG) and allowed to infect a macrophages-like cells in tissue culture in the presence or absence of ATc. Bacteria were recovered by gently lysing the macrophages at times indicated and enumerated by plate count. *F. novicida* strains have *vgrG* deleted ($\Delta vgrG$) except for a control strain with the chromosomal *vgrG* gene intact (WT); all express *tetR*. A) *vgrG* expressed from P40 (pMP829-P40-cat/vgrG). B) *vgrG* expressed from P18 (pMP829-P18-cat/vgrG). All data was collected in the same experiment so are directly comparable across panels A and B, but split into two graphs for increased clarity. Error bars represent S.E.M.

curve to be restored, as was the case if *tetR* was not present (Fig. 2.7). This data is consistent with findings from β-galactosidase assay (Fig. 2.3), and CAT immuno-blot (Fig. 2.5), which also demonstrates TetR mediated repression of genes downstream of either P40 or P18. However, repression of P40 was not always complete. P18, on the other hand, did not display any hints of leaky protein expression with any of the reporter gene or other we put under its control.

2.3.3 Transcription start sites and position of *tetO* in *F. novicida* promoters

Primer extension was performed on transcripts produced from 15 different synthetic promoters (10 *tet*-regulated and 5 constitutive) to determine the transcription start sites (TSSs or +1 position) of each. Promoters selected for transcript start analysis are the same set of 15 synthetic *F. novicida* promoters that represented a wide variation of expression strengths (all measured by β-galactosidase assay presented in Figure 2.3) and consisted of only a single random DNA fragment cloned into the BamHI site of the selection vector so that analysis was not complicated by multiple *tetO* sequences. Knowledge of the site at which transcription is favoured to start for each promoter allowed identification of −10 and −35 hexamers. As little information is available for *Francisella* promoters, −10 and −35 hexamers were annotated based on sequence similarity to that reported for *E. coli* σ70 promoters (Section 1.1.3). This assumption of homology to *E. coli* promoters appears to be reasonable for *Francisella* promoters as hexamers closely matching that of the canonical *E. coli* promoter were found at an appropriate distances upstream from the transcription start sites in all 15 examples (Fig. 2.8). The location of these primary promoter elements relative to *tetO* provide clues as to the general arrangement of Eσ70 and TF binding sites that lead to efficient repression by TetR vs. those that do not.

Not surprisingly, *tetO* was found close to, or overlapping with, the major promoter

elements in those promoters that exhibited *tet*-controlled expression (Fig. 2.8A), but *tetO* was located more distal to σ70-like binding sites in promoters that were not repressed by TetR (Fig. 2.8B). Furthermore, the five strongest promoters of the *tet*-controlled group were all overlapped in the −35 or −10 region by *tetO*, with the random DNA fragment having inserted in the "forward" orientation, an organization not observed in any of the five weaker *tet*-controlled promoters. This may indicate an influence of the position of *tetO*, relative to the RNAP binding site, and on the strength of these promoters in *Francisella*, independent from the regulatory properties of the *tetO*/TetR system.

In all five constitutive promoters *tetO* was found in the reverse orientation and at least 10 bp upstream of the putative −35 hexamer. Transcription start sites for each of the 5 start sites was located within the BamHI restriction site separating the random DNA fragment from sequence upstream of the *cat* gene on the plasmid (Fig. 2.8B). This uniformity is likely due to limited sequence space available to promoters form −10 and −35 hexamers that are far enough from *tetO* to not be encumbered by TetR binding of *tetO*

2.3.4 Synthetic *tet*-controlled promoters in *E. coli*

Accumulated, anecdotal evidence in the literature and work in our lab, suggests that *E. coli* promoters function poorly in *Francisella*. However, this idea has never been directly tested systematically, and little is known regarding how *Francisella* promoters generally function in *E. coli*. For the purpose of studying cross-species promoter activity, synthetic *tetO*-containing DNA fragments were selected for promoter activity in *E. coli* using a similar selection approach to that used to isolate synthetic promoters in *F. novicida* (Fig. 2.1). A preparation of random *tetO* DNA fragments ligated into the selection vector pMP829-cat/lacZ (prepared the same as described for selection

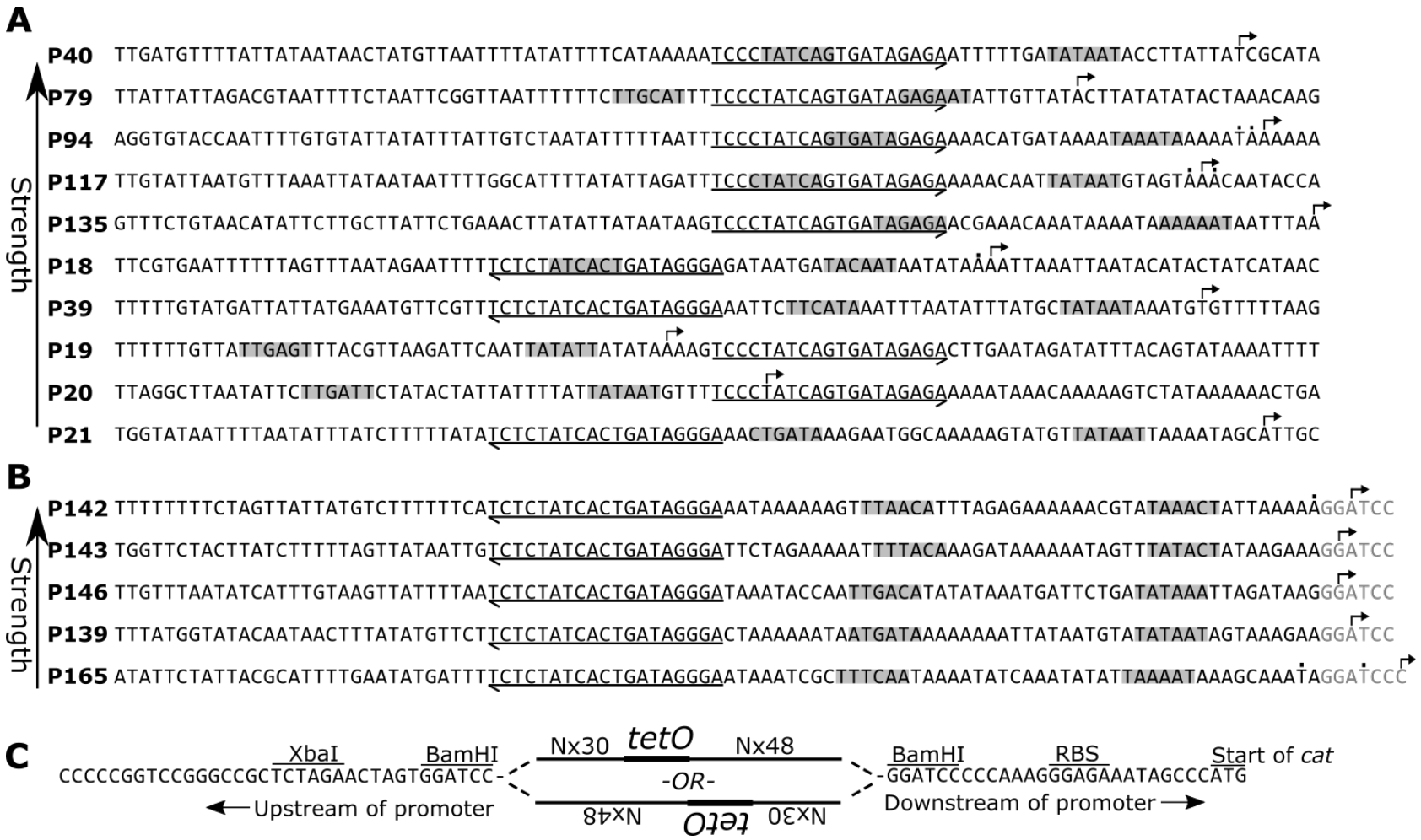


Figure 2.8: TSS mapping for 15 synthetic *tetO*-containing promoters selected for function in *F. novicida*. Bent arrows represent the transcript start (alternative start sites are marked with a dot, where detected) and this data was used to identify −10 and −35 hexamer positions (shaded). *tetO* sequences are underlined with an arrow indicating its orientation. A) *tet*-controlled synthetic *Francisella* promoters. B) Constitutive synthetic *Francisella* promoters. Some downstream sequence is also shown (greyed). C) Genetic context of promoters in pMP829-cat/lacZ. The two orientations in which the promoter fragment could insert into the selection vector are represented. The top orientation is referred to as "forward" in the text, and the bottom orientation as "reverse".

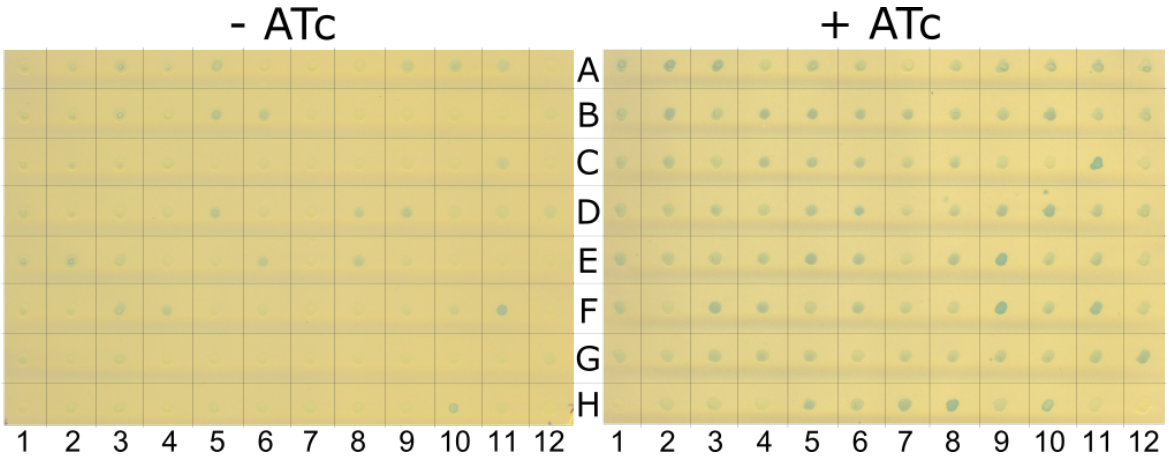


Figure 2.9: Identification of TetR responsive promoters in *E. coli* by β-galactosidase assay on agar plates. 88 Cm^R *E. coli* MGZ1 + pMP829-PEx-cat-lacZ. Cultures were spotted into LB + Hyg and allowed to grow O/N

of promoters in *F. novicida*) was used to transform *E. coli* MGZ1 (reference strain MG1655 derivative with a chromosomally expressed *tetR* gene, (2.1) to Cm^R and resulting in thousands of transformants. All of these clones are likely to have received a functional promoter sequence on the introduced random fragment as this *E. coli* strain carrying the promoterless parent plasmid was unable to form colonies on Cm under the same selection conditions. 88 of these Cm^R transformants (labelled PE-69–PE-156) were grown in liquid culture in a 96-well plate, sequenced and subjected to further analysis.

Sequencing revealed that all 88 clones had received a synthetic fragment upstream of *cat* and that 67 of those consisted of unique sequence (Table A.2). All but 1 of the 69 unique *E. coli*-selected promoters consisted of a single random fragment (PE-149, 2 fragments). This is in contrast to the *F. novicida* promoter selection where almost one third of promoters recovered consisted of two or more random fragments units that had ligated together before inserting into the selection vector (Table 2.3). The majority of these synthetic *E. coli* promoters displayed TetR repression and ATc induction as determined by X-gal spot assay plate (Fig. 2.9). Ten of these clones

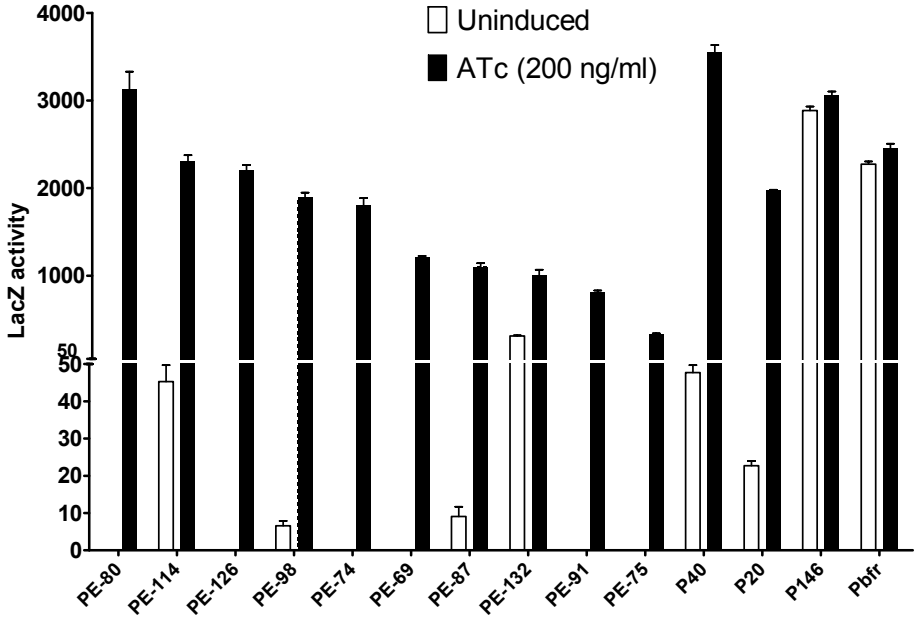


Figure 2.10: Activity of synthetic promoters in *E. coli* with and without ATc induction. Promoter activity was determined by β-galactosidase assay to measure levels of the LacZ reporter in *E. coli* MGZ1. Promoters with a PE prefix are those selected in *E. coli*. *F. novicida*-selected promoters P40, P20 and P146 and natural *Francisella* promoter, Pbfr, are also included. y-axis represents β-galactosidase activity in arbitrary units. Error bars represent S.E.M.

(6 *tet*-controlled and 4 constitutive) had expression levels quantified by LacZ assay. Included in this assay were three *F. novicida*-selected synthetic promoters (P40, P20 and P146), and the strong natural *F. novicida* promoter, Pbfr, to assess their function in *E. coli* and compare to the synthetic promoters that had been selected in *E. coli* (Fig. 2.10). *E. coli*-selected synthetic promoters behaved as expected in β-galactosidase assay as would be predicted from the X-gal spot plates. All 10 were induced by ATc and most were tightly repressed, with the exception of PE-114 and PE-132, which displayed considerable leaky expression in the absence of ATc (Fig. 2.10).

Transcription start sites were mapped for 10 *E. coli*-selected synthetic promoters (6 *tet*-controlled and 4 constitutive). Figure 2.11 shows TSS with corresponding

Figure 2.11: TSS mapping for 10 synthetic *tetO*-containing promoters selected for function in *E. coli*. Bent arrows represent the transcript start (alternative start sites are marked with a dot, where detected) and this data was used to identify −10 and −35 hexamer positions (shaded). *tetO* sequences are underlined with an arrow indicating its orientation. Promoters are ordered from strongest to weakest as measured by LacZ expression (Fig. 2.10). A) 6 *tet*-controlled *E. coli* promoters. B) 4 constitutive synthetic *E. coli* promoters.

putative −10 and −35 hexamers highlighted. As was found for the synthetic *F. novicida* promoters, all *tet*-controlled *E. coli*-selected promoters were organized with *tetO* overlapping, or in close proximity to major promoter elements or the +1 position. Unsurprisingly, the constitutive promoters all had more separation of *tetO* from the promoter elements. However, as was not seen for the the *F. novicida* constitutive promoters, *tetO* was positioned downstream of the TSS in PE-79. This promoter had a −35 hexamer that overlapped the upstream BamHI sequence, as did 3 of the 6 *tet*-controlled promoters. This is an strategy not used by any of the 15 *F. novicida*-selected promoters for which TSSs were mapped.

2.3.5 Cross-species promoter function

Interestingly, the *F. novicida*-selected promoter, P40, produced stronger expression of *lacZ* in *E. coli* than any of the 10 *E. coli*-selected promoters and was also *tet*-controlled in this host, although some leaky expression was observed here too. The other *F. novicida* promoters—both synthetic and natural—tested in *E. coli* all expressed strongly, to levels comparable with the strongest *E. coli*-selected promoters assayed here (2.10). These *F. novicida* promoters also all retained their *tet*-control or constitutive properties when used to drive transcription in *E. coli* the same as previously observed in *Francisella*. The opposite was not true, however; all *E. coli*-selected synthetic promoters were found to have little to no expression in a *F. novicida* host (Fig. 2.12).

Comparison of *F. novicida* and *E. coli* synthetic promoters on basis of sequence composition was difficult due biases and constraints introduced by the *tetO* sequence and due to the small number of promoters for which TSSs were identified. However, a simple analysis of G+C content for all unique, single fragment promoters for which sequence data was generated yielded interesting differences between *F. novicida*-selected and *E. coli*-selected promoters. The random DNA fragments from which

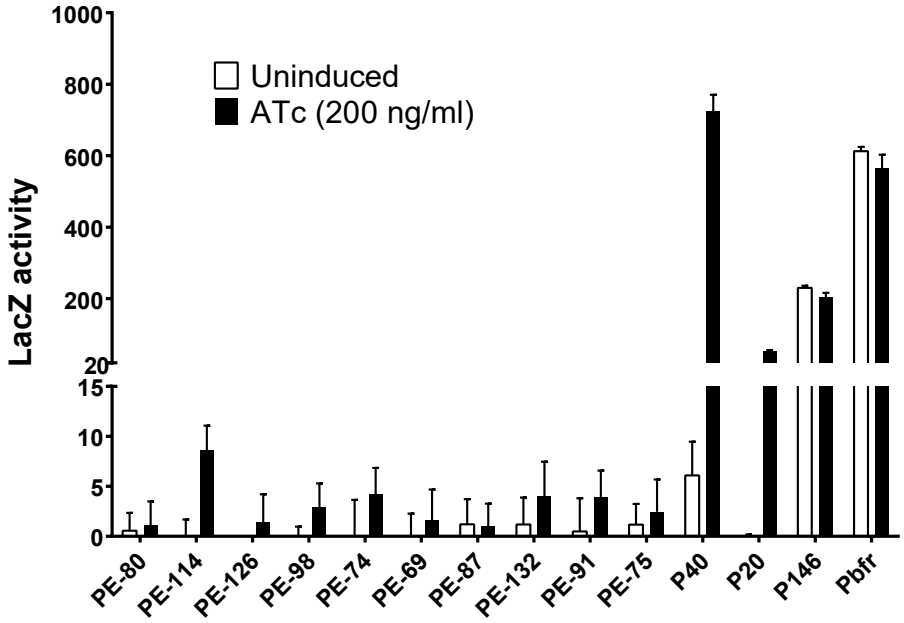


Figure 2.12: Activity of *E. coli*-selected synthetic promoters in *F. novicida*. Promoters selected in *E. coli* (indicated by "PE" prefix) were introduced into *F. novicida* $tetR^+$ and their activity was measured by LacZ β-galactosidase activity. For comparison, expression from promoters isolated in *F. novicida*, P20, P40 and P146, or the natural *Francisella* promoter, Pbfr, are shown. Values on the y-axis are arbitrary luminosity units. Error bars represent S.E.M.

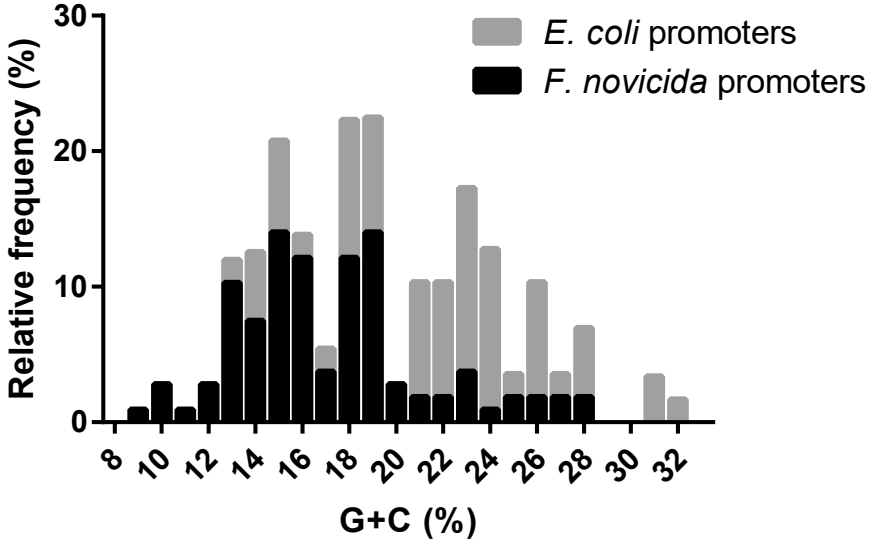


Figure 2.13: Histogram of G+C content found in 107 *F. novicida*-selected promoters vs. 59 *E. coli*-selected promoters. Mean for *F. novicida* promoters is 17.2% ± a standard error of 0.39. Mean for *E. coli* promoters is 21.7% ± a standard error of 0.47. Difference between means is statistically significant ($p < 0.0001$)

promoters were selected contained 78 randomized nucleotides (30 to one side of *tetO* and 48 to the other). The oligonucleotides from which these fragments were created had randomized positions weighted more heavily to A and T residues so that, on average, the degenerate positions should have G+C content of 30%. The mean G+C content of these 78 randomized positions for the 107 unique, single-fragment, *F. novicida*-selected promoters is 17.2%. The 59 *E. coli*-selected promoters that met this criteria had significantly greater mean G+C content at 21.7%. Figure 2.13 presents a histogram of the percentage frequencies of G+C content found in synthetic promoters from *E. coli* vs those from *F. novicida*.

2.3.6 Minimum size of *F. novicida* promoters

Data from mapping TSSs and estimation of RNAP binding sites suggest that *tetO* confers promoter repression when positioned within 5 bp of the −35 hexamer but

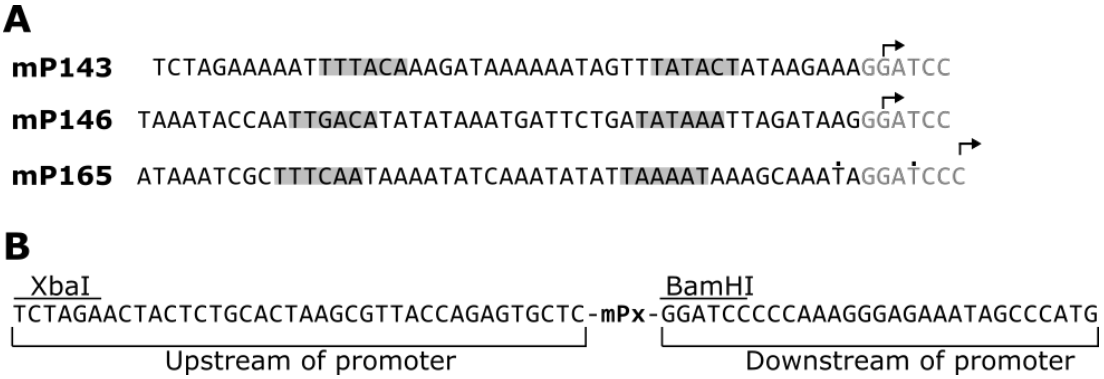


Figure 2.14: Sequences of *F. novicida* minimal promoters. A) Sequences of minimal-length promoters derived from constitutive promoters isolated in *F. novicida*. Bent arrows represent the transcript start (alternative start sites are marked with a dot, where detected) and this data was used to identify −10 and −35 hexamer positions (shaded). *tetO* sequences are underlined with an arrow indicating its orientation. Promoters are ordered from strongest to weakest as measured by LacZ expression (Fig. 2.15) B) Genetic context of the minimal *F. novicida* promoters in plasmid pMP829-cal/lacZ. 'mPx' represent the sequences shown in panel A.

does not confer TetR repression when positioned more than 10 bp from this region (Figs. 2.11 and 2.11). Taken together this implies that a region from +1 to about −50 is sufficient to encode all sequence elements necessary for transcription initiation in *F. novicida*. To test this notion the *tet* operator plus all synthetic DNA sequence upstream of *tetO* was removed from three plasmids containing constitutive *F. novicida* promoters—P143, P146 and P165. Inserted in place of the deleted sequence was 26 bps of randomly generated DNA sequence upstream of the "minimal" promoter region to serve as a neutral spacer sequence between the plasmid backbone derived sequence and the remaining promoter sequence. The sequences of these new minimal promoters, including that of the immediate genetic context, are presented in Figure 2.14.

Expression from these minimal promoters in *F. novicida*, as determined by β-galactosidase assay, was not meaningfully different from that of the corresponding full-length promoter regions from which the minimal promoters were derived (Fig. 2.15). This suggests that the deleted upstream sequence does not contribute significantly

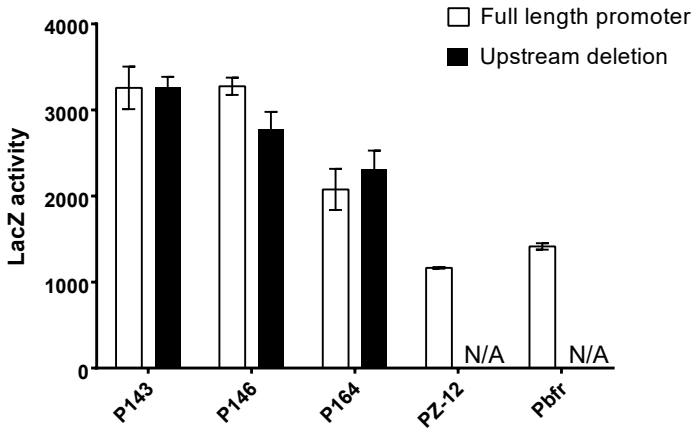


Figure 2.15: Expression of *F. novicida* minimal promoters in *F. novicida* by β-galactosidase assay. Filled bars represent expression originating from truncated versions of three constitutive, synthetic, *F. novicida* promoters identified in this study. Open bars represent expression from corresponding full-length promoter sequences. PZ-12 serves as an internal standard for expression relative to a natural promoter and has not been modified. Error bars represent S.E.M.

to promoter function. A benefit of identifying shorter synthetic promoters may arise from reduced probability of extended regions of similarity existing naturally in the chromosome, or between multiple copies of the synthetic promoter sequence present in the same host. Such similarity can contribute to genomic instability mediated by homologous recombination between similar sequences. When the three minimal promoters were used in a BLASTN query of the *F. novicida* chromosome they showed a maximum identity of only 17, 11, and 16 bp stretches of identical nucleotides for promoters mP143, mP146 and mp165, respectively. Thus, we have identified three DNA sequences that can serve as short, constitutive, *F. novicida* promoters with have low sequence identity to host chromosomal DNA sequences.

2.4 Discussion

Genetic control systems have been identified and well characterized in model organisms, such as *E. coli*. DNA sequences known to constitute promoters [196], ribosome binding sites [197], transcription stop signals [198], and protein degradation tags [199] can be assembled to produce protein of one or more genes to predicable steady-state levels. However, the function ascribed to these parts are dependent on the host cell. A DNA sequence does not encapsulate a biological function on its own. A "promoter", for example, can only promote transcription when acting together with RNAP, TFs, dNTPs, etc., provided by the cellular environment. Furthermore, a metabolic process does not occur in a vacuum and can be affected by other processes within the cell (eg. competition between two enzymes for a common substrate, or a number of promoters competing for binding a limited number of RNAP molecules). For these reasons, it is not surprising that genetic parts do not necessarily retain their function across hosts of various species, or even within the same cell under various conditions, as these will offer different intracellular environments for the DNA sequences to interact with. This is a problem for researchers and biological engineers attempting to express a heterologous gene, or gene pathway, in a less thoroughly studied organism, as promoters (or other genetic control parts) may not function in this new host.

When attempting to express a recombinant protein in a host for which characterized promoter sequences are not available, a simple solution is to barrow the sequence of a natural promoter from that host. This is common in *Francisella* as expression of recombinant DNA cassettes in this host often use the promoters of the *groE* operon, P_{groEL} [200] or *FTN1451* gene [182]. The use of native promoters is not so simple when TF-regulated transcription is desired as a suitable TF and a promoter which is controlled by the TF must both be found. Inducible transcription using host regulatory elements has been achieved in *F. novicida* through identification of a

promoter regulated by glucose (FGRp) [183]. Use of this promoter for controlled expression of a heterologous protein has many problems as repression is not complete and the inducer molecule is central to the cell's metabolism. Another attempt at inducible protein expression in *Francisella* has targeted translation initiation as the control point using a riboswitch. A riboswitch is a secondary structure in the 5'UTR of an mRNA that masks the ribosome binding site so that transcription does not occur until the riboswitch is 'flipped' by the inducing molecule, in this case theophylline, allowing transcription to begin [184]. Better results for inducible gene expression in *Francisella* were achieved though modification of P_{groEL} so that it was regulated by TetR. Lovullo et al. made a *tet*-controllable version of this promoter by inserting the *tetO* DNA sequence into the promoter so that *tetO* overlapped with the TSS causing tight repression of downstream genes [164]. This approach requires the modification of native promoters so that they contain operator sequences and has been successful in many different hosts including: *Mycobacterium tuberculosis* [162], *Borrelia burgdorferi* [163], *Synechococcus sp.* PCC 7002 [201], and *Clostridium acetobutylicum* [202].

A different approach to engineering of inducible promoters was attempted in the work presented here. Rather than start with a constitutive promoter and attempt to modify it for TF-control, I instead started with just the sequence of the operator for the TF that would be used, and let the host cell decide what form the rest of the DNA sequence would take. This method turned out to be very successful as a single transformation of *F. novicida* with the randomized *tetO* fragment produced thousands of putative synthetic *F. novicida* promoters, most of which demonstrated repression by TetR (Fig. A.1). Characterization of just a small set of these promoter sequences revealed a wide range of expression strengths and a range of regulatory properties from very tight repression by TetR to no *tet*-control at all.

Use of promoters derived from mostly random DNA sequence has several advantages

compared to modifying existing promoters for inducible gene expression. First, as expression of the majority of genes in any organism is regulated, and perhaps regulated in ways not yet identified, expression from such promoters has the potential to yield inconsistent levels of expression under varying culture conditions. For example, *Francisella* P_{groES} is known to be up-regulated under certain stress conditions [200], as would be expected for a promoter controlling an operon encoding molecular chaperones GroES and GroEL, which are involved in the refolding of denatured proteins [203]. When used outside of its natural genetic context, genes expressed from this promoter are considered to be produced constitutively despite the known up-regulation of this operon in response to conditions that favour protein misfolding [204]. Synthetic promoters derived from random sequence should be considerably less likely to contain regulatory sequences as no selection has been placed on them for this activity; however, almost all natural promoters have evolved to incorporate such control sequences [167]. Reduced propensity for unwanted control give promoters derived from random sequence an advantage in situations where consistent function over a range of growth conditions is preferred.

Use of a promoter indigenous to your host requires a second copy of the promoter's sequence be introduced into the cell. Such duplications provide a stretch of sequence with perfect homology to the original promoter at a different genetic locus of this organism and may increase the probability of genomic rearrangements due to homologous recombination between such homologous sequences. Bacterial strains commonly used in the laboratory for propagation of recombinant DNA (eg. *E. coli* DH5α), are typically engineered to increase genomic stability. This can be achieved by disruption of the *recA* gene, as is the case for *E. coli* DH5α[205]. Genomic instability due to recombination may be an especially important consideration when working with *Francisella* as it is known to have a very high rate of recombination and readily

integrates linear or circular DNA into its own [206]. In cases where maintenance of a recombinant DNA sequence causes significant burden to the host without providing a survival advantage, genomic rearrangements which relieve this burden by removing or altering the offending DNA will be favoured. Mutants that have accumulated such rearrangements may quickly dominate the culture resulting in loss of sequence of interest. The random nature of the synthetic promoters described here should reduce probability of such problematic mutations, especially in a host with higher than average propensity for recombination, such as *Francisella*. To further reduce homology to sequences of the host genome, truncated versions of three constitutive *F. novicida*-selected promoters were produced by deletion of the random fragment beyond 10 nts upstream of the −35 hexamer. These minimal promoters exhibited expression levels consistent with that of their full-length counterparts (Fig. 2.15), suggesting that the upstream portion of these promoters does not play a significant role in defining promoter strength in a sequence-dependent manner. This is consistent with the finding that synthetic promoters which bind TetR upstream of the −35 motif do not display impaired function.

Synthetic promoters created here had severe sequence constraints due to the presence of *tetO* and defined length of random nucleotides. Furthermore, only a small, biased sample of promoters had expression strength quantitatively measured and transcription start sites identified. For these reasons, it is difficult to create a consensus of these sequences (the alignment would be overwhelmed by the perfect match between *tetO* regions between all promoters) but such analysis could help identify common sequence motifs specific to promoters of the host which would boost understanding of general sequence properties of *F. novicida* gene regulatory sequences. Generation of a consensus sequence for the promoters described here may have allowed identification of sequence determinants of *F. novicida*-selected promoters

that mediate differences in activity from random promoters produced in *E. coli*. The lack of explanation for my observation of complete failure of *F. novicida* to initiate transcription from an *E. coli*-selected promoter was unsatisfying. However, it was possible to compare overall G+C content of *E. coli*-selected promoters and to those produced in *F. novicida*. Promoters from *F. novicida* displayed significant skew toward lower G+C content (Fig. 2.13) compared to that of the random DNA fragments used as starting material for the selection process and that of the *F. novicida* chromosome (both 30% G+C). Promoters selected in *F. novicida* also had a significantly lower C+G content than those from *E. coli*. This difference may be a factor in promoter specificity demonstrated by *Francisella*. Four of the ten *E. coli*-selected promoters were found to have a −35 hexamer that overlapped with sequences within the BamHI recognition site upstream of the randomized DNA sequence in these fragments which already closely matched that of the *E. coli* consensus −35 (Fig. 2.11). However, none of the 15 characterized *F. novicida*-selected promoters employed this strategy to use this sequence common to all random fragments as ready-made −35 hexamers. This BamHI ligation scar and the sequence immediately upstream of it is very high in G+C (Fig. 2.8C). The inability of *Francisella* promoters to be found in close proximity to this region is consistent with the suggestion that G+C content is plays a role in defining promoters functional in *Francisella*.

The data presented here, as well as previous studies of *F. novicida* promoters [57], suggest that sequences of the core *Francisella* promoter elements are very similar to that of the *E. coli* core promoter. Still, we have demonstrated that *E. coli* promoters are unlikely to function in *F. novicida*, although the reverse does not appear to be true. This may be a necessary adaption *Francisella* has made to accommodate the low G+C content of its genome (32%). As promoters can be generally recognized as areas of low G+C content [207], and the consensus −10 and −35 promoter elements are made

up of A or T nucleotides in 10 of 12 sites, an organism with already low G+C content, such as *Francisella*, should have a genome highly enriched for σ70 promoter-like sequences. The presence of such 'accidental' promoter sequences scattered throughout the genome would likely be detrimental to the host. Therefore, it is reasonable that *Francisella* has evolved additional discriminating promoter features to reduce non-specific transcription initiation. Examination of the 15 synthetic *Francisella* promoters for which the TSS has been determined reveled a 'TGn' motif directly upstream of the −10 hexamer in 5 of the 15 promoters (33%; Fig. 2.8). Accounting for the 30% G+C content of the randomized nucleotides, this 'extended −10' motif would be expected in only 5.2% of promoters if occurring by chance alone. Other low G+C organisms, such as *Campylobacter jejuni* [208] and *Mycoplasma hyopneumoniae* [209], are known to have a highly conserved 'TGn' at this position. This motif is also known to be conserved in some *E. coli* promoters—a sub-set of promoters classified as 'extended −10 promoters' [62]. However, the hypothesis that an extended −10 promoter element is important for efficient transcription initiation in *Francisella* is not supported by the observation of an even higher proportion of this motif found in *E. coli*-selected promoters (4 of 10; Fig. 2.11). Two of these extended −10 promoters (PE98 and PE78) had function assayed in *F. novicida* but were unable to promote transcription in that host (Fig. 2.12). I was unable to identify a distinguishing sequence motif of *Francisella* promoters compared to those of *E. coli*, other than general G+C content. The use of DNA fragments with purely random sequence (containing no operator) would have allowed more meaningful analysis of synthetic promoter sequences which could help elucidate specific sequence motif that allow promoter function in *Francisella*, or any other host in which this method is employed.

Function of promoters produced here were assayed by detection of protein products of reporter genes, LacZ and CAT. These provide a convent method to analyze

transcription efficiency, but the measure is an indirect one that can be complicated by other factors. For example, the amount of protein produced will be affected not only by promoter strength, but also by premature termination of transcription, mRNA stability, translation initiation and elongation rates, and protein stability. Therefore, without direct measurement of transcript production, what is actually quantified is the sum process of transcription and translation, net of mRNA and protein degradation. Furthermore, as the LacZ detection method was an enzymatic activity assay, this indirect measure of protein concentration is another step removed from measure of transcription as enzymatic activity could misrepresent protein levels in cases of altered cellular conditions that affect the chemistry of that enzymatic reaction or by improper protein folding, etc. Of these variables, the one likely to lead to errors in estimation of promoter strength as reported here is alterations translation initiation efficiency. Secondary structure of the 5'-end of a transcript, or lack thereof, can mask or expose an RBS motif, thereby affecting Watson-Crick base pairing with 16S rRNA component of the ribosome and changing the probability of a ribosome initiating translation at a particular start codon [197]. This can significantly impact the translation of for coding sequences at the 5'-end of a transcript [210]. However, the *lacZ* gene, as well as *vgrG*, was placed as the second CDS in a co-transcribed operon behind *cat* in the reporter plasmid used to measure the strength of promoters placed upstream of these genes. These genes had their own RBS that was well removed from the 5'-end and should therefore not be significantly influenced by secondary structure. However, measurement of CAT levels—both by functional assay of Cm^R and directly by Western blot—could be altered by changes in length and sequence of the 5'-UTR, and should be interpreted with caution. Also, the anti-Shine Dalgarno sequence of *F. novicida* 16S rRNA [174] is identical to that of *E. coli*, so one would expect similar translation initiation rates in both species.

Francisella species have a poor repertoire of transcription control tools and it appears that importing controlled promoters from other species is not a viable option. In this work we demonstrate that regulated transcription control elements can be produced through a simple selection and screening process of a random DNA fragments which contain a non-random TF binding site. Using this method we were able to create a series of tightly repressed, strongly inducible promoters for *F. novicida* and *E. coli*, and produce small, defined promoters that should be ideal for genomic engineering.

Chapter 3

Creating temperature-inducible mutants of the tetracycline repressor

3.1 Introduction

In the previous chapter a method was presented by which a naturally occurring genetic switch could be modified to allow repressible transcription in an organism other than that from which the genetic elements originated. To achieve this, the natural repressor binding site, *tetO*, was kept constant while surrounding nucleotides were randomized then selected for promoter activity in *F. novicida*. In this chapter I present work that again modifies control elements of the tetracycline resistance operon to further expand its functionality. Here we attempt to engineer variants of the tetracycline repressor (TetR) to have a temperature inducible phenotype. Successful mutants would behave as wild-type below some temperature but loose that activity at higher temperatures, resulting in a system by which gene expression can be controlled through changes in culture temperature.

As discussed in previous chapters, inducible gene expression control systems are common tools in molecular and synthetic biology. These typically employ small molecules as the inducing factor; most well known of which is the lactose repressor

(LacI) and its inducing molecule, allolactose (or synthetic inducer IPTG). Other popular induction systems include AraC/arabinose, LuxR/acyl-L-homoserine and TetR/tetracycline. These systems allow one to turn on the expression of a gene by simply adding the inducing molecule to a cell culture and are of highly useful to molecular biologists. Such systems give an experimenter control over if or when a gene of interest is expressed. Any one of many well-characterized repressor-operator pair could have been chosen as the repressive system for the semi-random inducible promoters discussed in Chapter 2, and for the temperature inducible repressor mutants described in this chapter. However, the TetR/*tetO* repressor-operator combination is ideal for many reasons going beyond the tight repression and sensitive induction properties discussed in Section 1.2.2. Unlike some other commonly used bacterial repressors, TetR is not induced by common metabolites such as lactose (LacI) or arabinose (AraC), which limit their use in situations where these inducing molecules may be present. Another advantage of TetR is the great body of knowledge that has been accumulated regarding this protein. TetR is thoroughly studied and has been to used to control gene expression for many years in a wide variety of organisms, including higher eukaryotes [211, 212]. In addition to the ideal natural properties of TetR, work with this repressor has led to the development of mutants with interesting and potentially useful properties. TetR mutants have been isolated that display different binding and induction phenotypes. These include mutants that recognise different operator sequences [213], and reverse TetR (rTetR), which binds its operator only in the presence of Tc, rather than in its absence [214–216]. Such mutants could be used in conjunction with the promoters described in Chapter 2 to expand their functionality. Previous engineering of TetR-mediated transcription regulation with novel and useful properties, go beyond changes to the protein itself. New inducers of TetR have also been developed consisting of peptides [217] or oligonucleotides [218]

which complex with TetR to cause the same structural changes that lead to loss of DNA binding activity as caused by Tc and Tc analogues. As such inducers are genetically encoded, they can be expressed as tags on the end of a protein or RNA molecule. This expands possibilities for complex regulatory circuits where expression of an unrelated gene could in-turn induce a TetR regulated gene, or repress it via rTetR [219, 220].

A major drawback to using chemical inducers is that the expressed state is difficult to undo once the inducer has been added. For instance, if a gene of interest is controlled by the *lac* promoter, it can be easily induced by adding IPTG to the culture, but reversing the process would require separating the cells from the medium by centrifugation or filtration, washing the cells carefully, and replacing the medium with fresh medium free of the inducing chemical. Although feasible on lab scale, this becomes much more troublesome for larger processes, or in situations where returning back to the repressed state is desired.

In bioindustrial processes where reuse of culture medium and/or reversibility of gene expression is advantageous, induction systems that rely on non-chemical induction signals could be employed. In such a system the inducer is a change in culture environment that does not involve the presence of an inducing molecule. Common examples of non-chemical inducing stimuli include temperature, light , electric current, and magnetic field [221]. An inducing stimulus such as this can be turned off by simply returning the cells to the repressive condition (e.g. restoring the original temperature or switching off the light source). A common tool in temperature-induction is a mutant of the lambda repressor (λCI857), which looses its ability to repress transcription from the lambda promoter at temperatures above 40°C [222]. Since the serendipitous discovery of the temperature inducible λ repressor over 50 years ago, researchers have generated additional temperature inducible mutants of repressor proteins through

both random mutagenesis plus screening [223], and through rational design [224].

Currently, there are limited examples of repressors proteins that respond to non-chemical—and hence, easily reversible—stimuli. Here I present worked aimed at increasing the repertoire of temperature inducible repressors though random mutation of tetracycline repressor (TetR) followed by selection/screening for those mutants that allow gene expression only at an elevated temperature. A library of many such temperature-inducible mutants could contain individuals which switch from the repressed to induced state at various temperatures so that a repressor with appropriate properties can be selected for a particular application. Of course, a repressor protein is only useful if a promoter which is regulated by that repressor exists in the species of interest. This issue has been addressed in the work presented in chapter 2 for the case of *Francisella* and TetR controlled promoters, but those methods, more broadly applied, could provide a framework to create promoters in other host organisms that respond to potentially any transcription factor protein one chooses.

3.2 Methods

All experimental procedures presented in this section were performed by R. McWhinnie

3.2.1 Strains and culture conditions

For routine growth and genetic manipulations *E. coli* strains were grown in modified lysogeny broth (LB; 1% tryptone, 0.5% yeast extract, 0.5% NaCl) or on LB agar. Anhydrotetracycline (ATc) was used at 100 ng/ml, hygromycin (Hyg) at 150 µg/ml, chloramphenicol (Cm) at concentrations indicated, and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal) at 20 µg/ml, as required. Electroporation and chemical transformation of *E. coli* were performed by standard protocols [186].

Table 3.1: Strains, vectors, and oligonucleotides used in this chapter

Name	Description	Source
Strains		
E. coli DH10B	F$^-$ *mcrA* Δ(*mrr-hsdRMS-mcrBC*) φ80*lacZ* ΔM15 Δ*lacX74 recA1 endA1 araD139* Δ(*ara leu*)7697 *galU galK* λ$^-$ *rpsL nupG*	Invitrogen
E. coli MG1655	F$^-$ λ$^-$ *ilv*G$^-$ *rfb*-50 *rph*-1	[13]
Vectors		
pBlueScript II KS	Cloning vector, ColE1 ori (high copy), Ap^R	Stratagene
pBS-tetR	pBluescript II KS expressing *tetR* from *lac* promoter	This work
pJN105	BBR1 ori (medium copy), Gm^R	[225]
pRM100	pJN105 with P$_L$-tetO	This work
pRM101	pRM100 with P$_{BAD}$ control elements removed	This work
pRM-cat	pRM101 with *cat* under control of P$_L$-tetO	This work
pSB3K3	P15A ori (medium copy), Km^R	[191]
pSB1K3	pMB1 ori (high copy), Km^R	[191]
pSB1K3-P$_L$-tetO-yfp	pSB1K3 with *yfp* transcribed from P$_L$-tetO	This work
pSB3K3-P$_L$-tetO-yfp	pSB3K3 with *yfp* transcribed from P$_L$-tetO	This work
pSB6A1	pMB1 ori (high copy), Ap^R	[191]
pSB4A5	pSC101 ori (low copy), Ap^R	[191]
pSB6A1-tetR	pSB6A1 expressing *tetR* from medium strength promoter	This work
pSB4A5-tetR	pSB4A5 expressing *tetR* from medium strength promoter	This work
Oligonucleotides		
F-PLtetO-p105	`ctatcagtgatagagatactgagcacatcaCCGCTCTAGAACTAGTGGA`	
R-PLtetO-p105	`ggatgtcaatctctatcactgatagggaGTGGAGCTCCAATTCG`	
F-tetR	`gcgcgcggtaccAAATTAGGAATTAATGATGTCTAGATTAG`	
R-tetR	`gcgcgcctgcagGTTATGCTGCTTTTAAGACC`	

3.2.2 Plasmid construction

The *tet*-controllable, positive selection, reporter vector, pRM101-*cat* was made as follows. The broad-host range vector, pJN105 [225] (Table 3.1) was modified to add the tetracycline inducible promoter, P_L-tetO [166] by inverse PCR with primers that included the promoter sequence as 5' tails on the complimentary region of 5' phosphorylated oligos. Subsequent ligation of this PCR product resulted in a re-circularised vector that now contained the P_L-tetO sequence so that transcription initiated at this promoter would proceed into the MCS. To reduce potential read-trough transcription, a transcription terminator derived from coli phage rrnBT1 and T7TE transcription terminator sequences (BioBrick part BBa_B0015) [191] was ligated into the SacI site, after blunting the cut site by polishing with Klenow fragment (NEB), just upstream of P_L-tetO, creating plasmid pRM100. The *araC* gene and promoter it regulates (P_{BAD}) were deleted from pRM100 by inverse PCR yielding, pRM101. Insertion of the *cat* gene into the SmaI site of pRM101 produced pRM-cat, a selection vector with *cat* under control of the P_L-tetO. The fluorescent reporter plasmid pSB3K3-yfp was created via BioBrick assembly [226] of the P_L-tetO promoter (BioBrick part BBa_R0040) and *eyfp* (BioBrick part BBa_E0430) into the medium copy number vector pSB3K3 (BioBrick part BBa_J04450) [191].

3.2.3 Random mutagenesis of *tetR* and isolation of temperature inducible mutants

TetR was amplified from Tn*10* under error-prone conditions using mutagenic PCR buffer as described by Cadwell & Joyce [227]. Briefly, a mutagenic PCR buffer was made by supplementing 10x Taq polymerase buffer (Standard Taq buffer; NEB) with 1/10th the volume of 10x mutagenesis buffer (0.4 mM dTTP, 0.4 mM dCTP, 2.75 mM

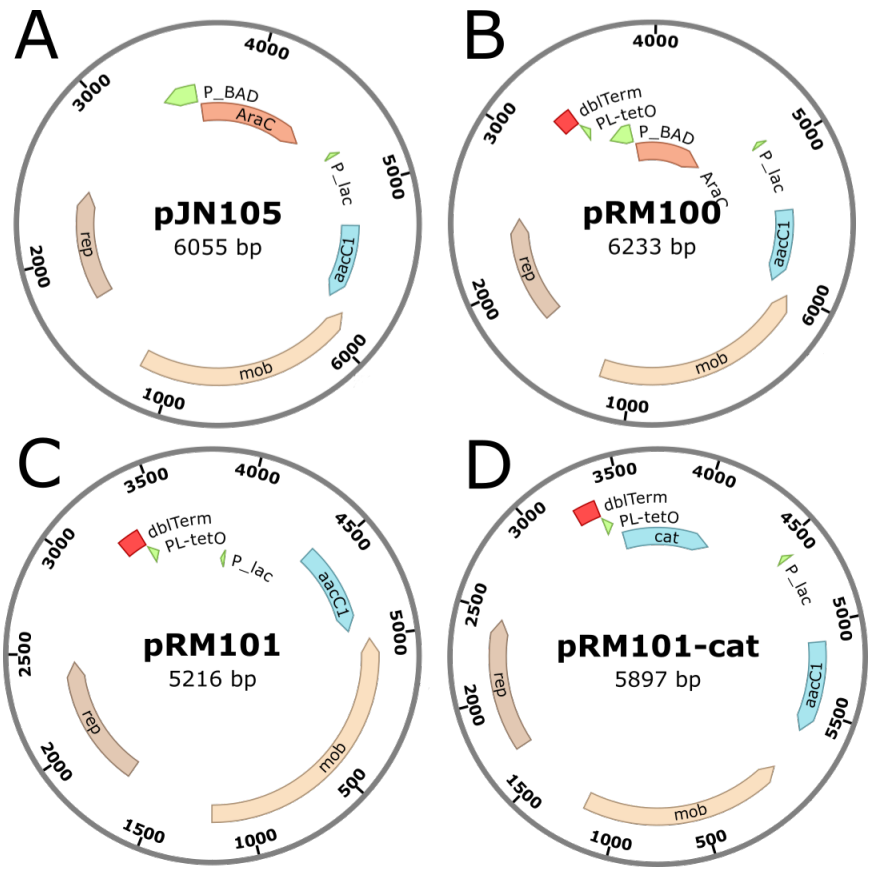


Figure 3.1: Tet-responsive selective plasmid and their parent plasmids. A) pJN105: Gm^R, broad-host range plasmid from which other reporter plasmids in this study were derived. B) pRM100: pJN105 with TetR-controlled promoter added, retains arabinose operon control elements. C) pRM101: pRM100 with arabinose operon control elements removed. D) pRM101-cat: pRM101 with *cat* gene downstream and under control of P_L-tetO. Details listed in Table 3.1.

$MgCl_2$, and 0.25 mM $MnCl_2$). This mixture was used in place of 10x PCR reaction buffer in a PCR to for 27 cycles to amplify *tetR* from Tn*10* DNA. PCR products were purified using a NucleoSpin Extract II PCR purification kit (Macherey-Nagel) then digested with KpnI and PstI, column purified again, and ligated into pBlueScript II KS that had been digested with the same enzymes. *E. coli* DH10B cells harbouring pRM101-cat were electroporated with the ligation product, recovered for 1 hour at 42° in SOC then plated on LB agar preheated to 42°C containing Gm, Ap and Cm (30µg/ml) and incubated overnight at 42°. About 10,000 clones were picked into 384-well plates containing LB freezing media (ref) containing Gm and Ap using the Genetix QPix2 colony picking robot (Molecular Devices). The 384-well plates were grown for 48 hours at room temperature then spotted onto LB agar in containing Gm, Ap, with and without Cm in Nunc Omniplates (Nalge Nunc International) using a 384-pin replicator. These plates were incubated at 30° for 24 h and any clones unable to grow on Cm at 30°, but displaying strong growth on the control plate lacking Cm, were considered putative temperature inducible mutants.

3.2.4 Western blots

Cultures were grown in LB with appropriate antibiotics and in the presence of IPTG (where indicated) and harvested at mid-exponential phase. For temperature induction of YFP expression cultures were grown at 30° then shifted to 42° two hours before harvest. In cases where temperature was not indicated cultures were grown at 37°. ATc was added two hours before harvest, where indicated. One ml of culture was pelleted by centrifugation and resuspended in 25 µl cold dH_2O containing protease inhibitors ("cOmplete" protease inhibitor cocktail, EDTA-free; Roche) before adding 30 µl of 2×SDS loading buffer. Samples were loaded based on equal culture density basis, separated by SDS-PAGE on a 12% polyacrylamide gel (10 µl lysate loaded per

lane), transferred to nitrocellulose and blocked in Odyssey blocking buffer (LI-COR Biosciences). Primary antibodies were diluted in Odyssey blocking buffer containing 0.05% tween-20 and used at the following dilutions: rabbit anti-TetR polyclonal (Abcam, ab14075; 1:1000) and mouse anti-GFP monoclonal (Abcam, ab73933; 1:5000). Anti-TetR was detected using IRDye800-conjugated goat anti-rabbit and anti-GFP was detected using IRDye700-conjugated goat anti-mouse secondary antibodies (Rockland Immunochemicals) in Odyssey blocking buffer containing 0.05% tween-20 and 0.01% SDS (1:15,000) and visualized on the Odyssey scanner (LI-COR Biosciences). Images were converted to grayscale and cleaned up using the Image Studio Lite software (LI-COR Biosciences; v5.0). Band intensity analysis was also performed using this software.

3.3 Results

3.3.1 Characterisation of selection and reporter plasmids

pRM100-cat, a Gm^R med-copy plasmid containing the Cm antibiotic marker gene, *cat*, expressed from the TetR repressed P_L-tetO promoter was constructed for use in identifying mutants of TetR with a temperature inducible phenotype. With the *tetR* mutant library provided on the compatible pBlueScript II (pBS) plasmid to control expression of Cm resistance from the P_L-tetO promoter of pRM100-cat in trans. Preliminary tests revealed that *E. coli* DH10B harbouring pRM100-cat, and pBS-$tetR^{wt}$ (pBS expressing the wild-type *tetR* gene) was unable to grow at temperatures above 39° on Cm whether or not the TetR inducer, ATc, was added to the growth medium. This unexpected growth deficiency made use of this plasmid unacceptable for selection of temperature inducible mutants as a maximum selection temperature of 42° was to be used. Fortunately, a derivative of pRM100—pRM101—which had been

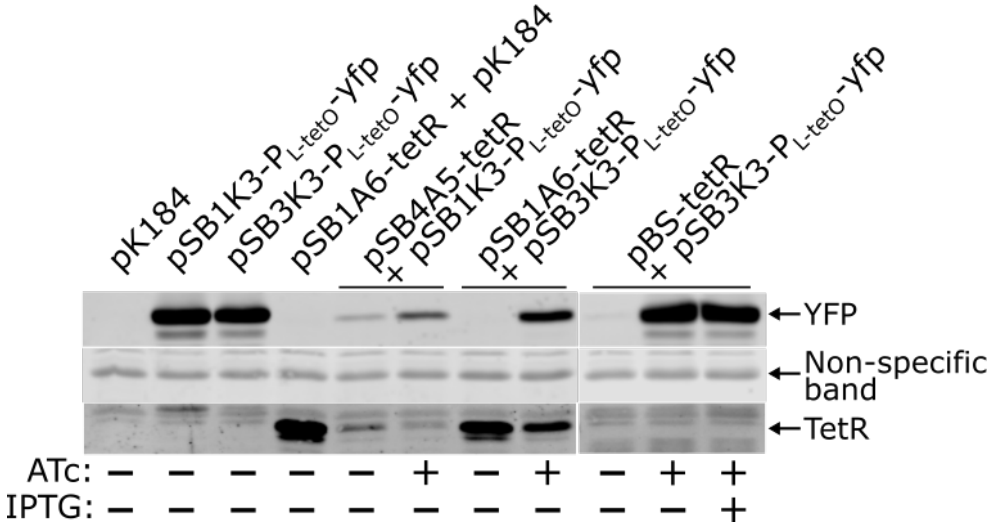


Figure 3.2: Western blot analysis of TetR and YFP expression from various vectors. *E. coli* DH10B strains harbouring the indicated plasmids were inspected for expression of YFP and TetR from various combinations of plasmids that carry separately the *yfp* gene under control of different constitutive prompters which are expected to drive the expression of *tetR* at different levels. Where indicated, ATc was added 2 hours before harvest and IPTG was included for the entire culture period in the case indicated. The three right-most lanes are from a different gel/membrane than the other, but both were processed and imaged at the same time and in an identical manner. This should make direct comparison of the two membranes possible, as is supported by the similar level of signal of the non-specific band across all lanes.

engineered for reduced complexity, when carrying the *cat* gene (pRM-cat; Fig. 3.1), generating the expected Cm resistance phenotypes in response to TetR with and without ATc, regardless of the culture temperature (30 and 42°).

3.3.2 Analysis of *tetR* mutagenesis products

Mutants of the *tetR* gene created by error-prone PCR using primers F-tetR and R-tetR (Table 3.1) and these were digested with KpnI and SpeI then ligated into pBlueScript II KS. Before A small random sample of potential *tetR* mutants were sequence to quantify the extent of mutation, uniqueness of the mutant sequences,

Table 3.2: Sequence analysis of 10 *tetR* epPCR products to assess extent of mutation

Clone	Mutations*
1	**G229A**, T249C, $\boxed{\Delta 284\text{-}285}$, A613G
2	**C157T**, **A218T**, **A365T**, **A449G**, **A484G**, **A487C**, $\boxed{\Delta 573}$
3	**A307G**, $\boxed{\text{T330G}}$, A478G
4	**T26A**, $\boxed{\text{T126A}}$, A213G, A288G, A623T
5	**T26A**, $\boxed{\text{T126A}}$, A213G, A288G, A623T
6	T72A, $\boxed{\text{A463T}}$, A513G
7	**A14T**, **A136G**, T268C, **A284T**, **A332G**, **T338C**, T372C, **A383G**, **A385G**, A612G
8	A19G, **T21A**, A165G, $\boxed{\text{C514T}}$, T526A
9	T126C, **C222A**, A333G, **A416G**, A426G, **A467G**, **A515G**, **T556C**, $\boxed{\text{T609A}}$
10	G57A, A288T, $\boxed{\text{C346T}}$, A449G, T534C

* Mutations in bold result in an amino acid substitution; non-bold, non-boxed nucleotide substitutions are silent. Boxes indicate the creation of a stop codons; in the case of deletions the box indicates the creation of a stop codon shortly following the frame-shift.

and percentage of full-length mutants generated by the error-prone PCR. *E. coli* DH10B without a selective reporter plasmid was transformed with the error-prone PCR product of *tetR* ligated into pBlueScript II KS. Ten white clones were picked at random and sequenced; the changes from wild type are shown in Table 3.2. Nine of the ten were unique sequences with three to ten single base-pair changes per gene within the coding region. Two had deletions of one and two base-pairs both resulting in a frame-shift and truncated protein. The average number of SNPs in the nine unique sequences was 5.7 with only one predicted to produce full length protein due to non-sense mutations. Transitions were favoured to transversions (35 to 14) by the epPCR conditions employed and no G to C or G to T transversions were observed in any of the ten sequenced mutagenesis products.

Table 3.3: Tabulation of mutations accumulated by epPCR.

Mutation	Our results*	Results of Rasila et al. [228]†
Transitions		
A:T → G:C	28 (56%)	48 (38.1%)
A → G	21 (42%)	
T → C	7 (14%)	
G:C → A:T	5 (10%)	24 (19.0%)
G → A	2 (4%)	
C → T	3 (6%)	
Transversions		
A:T → T:A	13 (26%)	35 (27.8%)
A → T	7 (14%)	
T → A	6 (12%)	
A:T → C:G	2 (4%)	5 (4.0%)
A → C	1 (2%)	
T → G	1 (2%)	
G:C → T:A	1 (2%)	8 (6.3%)
G → T	0	
C → A	1 (2%)	
G:C → C:G	0	0
Deletions		
Δ1 bp	1 (2%)	6 (4.8%)
Δ2 bps	1 (2%)	0

* Out of 50 total mutation in 9 epPCR products of *tetR*, not selected for TetR activity. 5562 total bps sequenced

† Out of 126 total mutations in LacZ. 16510 total bps sequenced.

3.3.3 Isolation of temperature inducible TetR mutants

Cells possessing the P_L-tetO controlled *cat* gene were electroporated with the mutant library of *tetO* ligated into pBlueScript II. After recovery in SOC at 37° for 30 minutes the temperature was bumped up to 42° for 30 minutes. The transformation culture was spread on LB agar containing Ap, Gm, and Cm (50 ğ/ml) and preheated to 42°. The agar trays were immediately returned to 42° and incubated O/N. Over 10,000 transformants were picked using a robot into 384-well plates containing 50 ĭ of LB freezing media (LB with 36 mM K_2HPO_4, 13.2 mM KH_2PO_4, 1.7 mM sodium citrate, 0.4 mM $MgSO_4$, 6.8 mM $(NH_4)_2SO_4$, and 4.4% glycerol (vol/vol)) + Ap + Gm and grown O/N at 30°.

Previous experiments showed that empty vector should make up a very low proportion of these Cm^R, and therefore *tetR*$^-$, transformants, but the sequence of epPCR products suggests a high proportion will have acquired a premature stop codon which should cause complete loss of function. Most full length *tetR* mutants with mutations that render TetR non-functional at 42° will likely be non-functional at a lower temperature. To isolate clones that produce a repressor protein that is non-functional at 42°, but retains function at a lower temperature, clones unable to repress *cat* expression at 42° were spotted from 384-well plates onto LB agar containing Ap, Gm and Cm at 30°. Of the 10,000+ *tetR*$^-$ clones identified at 42°, 72 clones were not able to grow on Cm at this temperature, indicating a $TetR^+$ phenotype at 30° (*tetR*^{ti}).

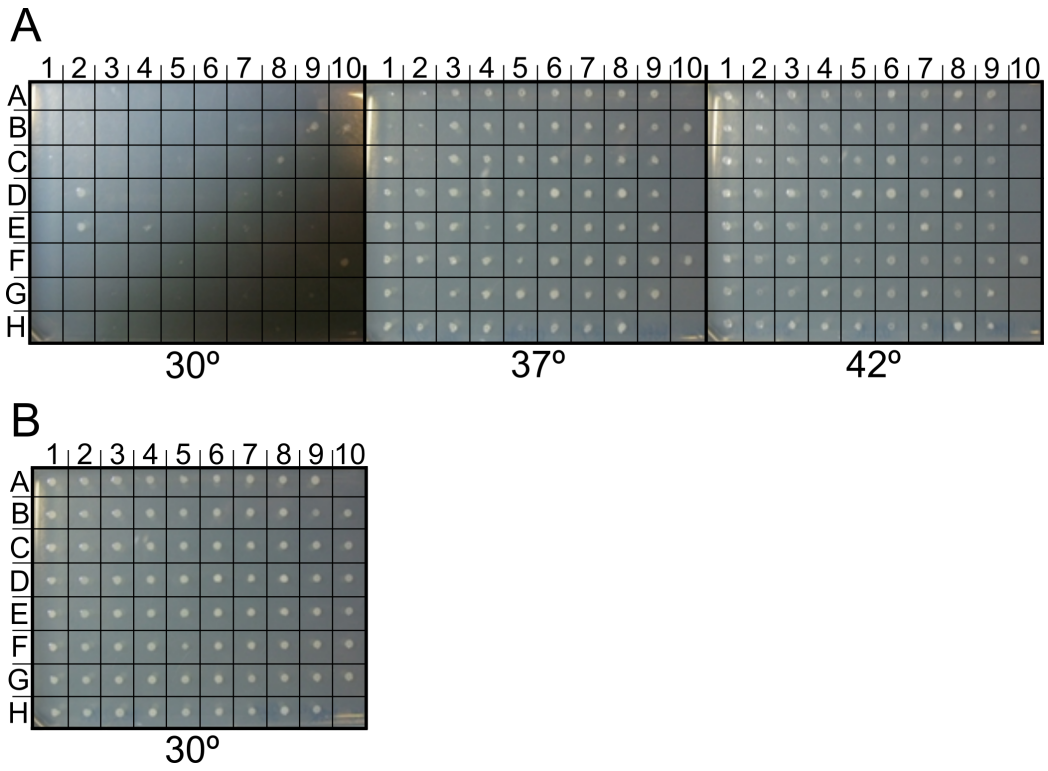


Figure 3.3: Agar spot assay for survival on chloramphenicol of clones with temperature inducible expression of CAT activity. 72 clones that previously exhibited a ti Cm^R phenotype, along with controls, were spotted onto solid medium and grown at 30, 37, and 42°. A) Growth on plates containing 50 µg/ml Cm at various temperatures. B) Control plate without Cm. Images of control plates lacking Cm at temperatures above 30° have been lost, but data from these plates was recorded at the time and is presented in Table 3.4, which also indicates the layout of clones in this figure.

Table 3.4: Survival of putative *tetR*^{ti} clones on Cm at various temperatures. First column gives a label for clones or, in the case of control spots, indicates the genotype. Second column indicates the well position of the clone, as shown in Figure 3.3.

		30°		37°		42°		
Label	Well	0	50	0	50	0	50 µg/ml	[Cm]
tetR^{ti}-1	A1	+++	–	+++	–	+++	+++	
tetR^{ti}-2	B1	+++	–	+++	–	+++	+++	
tetR^{ti}-3	C1	+++	–	+++	+++	+++	+++	
tetR^{ti}-4	D1	+++	–	+++	+++	+++	+++	
tetR^{ti}-5	E1	+++	–	+++	+++	+++	+++	
tetR^{ti}-6	F1	+++	–	+++	+++	+++	+++	
tetR^{ti}-7	G1	+++	–	+++	+++	+++	+++	
tetR^{ti}-8	H1	+++	–	+++	+++	+++	+++	
tetR^{ti}-9	A2	+++	–	+++	+	+++	+++	
tetR^{ti}-10	B2	+++	+	+++	+	+++	+++	
tetR^{ti}-11	C2	+++	–	+++	–	+++	+++	
tetR^{ti}-12	D2	+++	+++	+++	+++	+++	+++	
tetR^{ti}-13	E2	+++	+++	+++	+++	+++	+++	
tetR^{ti}-14	F2	+++	–	+++	+++	+++	+++	
tetR^{ti}-15	G2	+++	–	+++	–	+++	++	
tetR^{ti}-16	H2	+++	–	+++	+++	+++	+++	
tetR^{ti}-17	A3	+++	–	+++	+++	+++	+++	
tetR^{ti}-18	B3	+++	–	+++	+++	+++	+++	
tetR^{ti}-19	C3	+++	–	+++	+++	+++	+++	
tetR^{ti}-20	D3	+++	–	+++	++	+++	+++	
tetR^{ti}-21	E3	+++	–	+++	+++	+++	+++	
tetR^{ti}-22	F3	+++	–	+++	+++	+++	+++	
tetR^{ti}-23	G3	+++	+	+++	+++	+++	+++	
tetR^{ti}-24	H3	+++	+	+++	+++	+++	+++	
tetR^{ti}-25	A4	+++	–	+++	+++	+++	+++	
tetR^{ti}-26	B4	+++	–	+++	++	+++	+++	
tetR^{ti}-27	C4	+++	–	+++	+++	+++	+++	
tetR^{ti}-28	D4	+++	–	+++	+++	+++	+++	
tetR^{ti}-29	E4	+++	+	+++	+	+++	+++	
tetR^{ti}-30	F4	+++	–	+++	+++	+++	+++	
tetR^{ti}-31	G4	+++	–	+++	+++	+++	+++	
tetR^{ti}-32	H4	+++	–	+++	+++	+++	+++	
tetR^{ti}-33	A5	+++	–	+++	+++	+++	+++	
tetR^{ti}-34	B5	+++	–	+++	+++	+++	+++	

Continued on next page...

Table 3.4 – continued from previous page

		30°		37°		42°		
Label	Well	0	50	0	50	0	50 µg/ml	[Cm]
tetR^{ti}-35	C5	+++	+	+++	+++	+++	+++	
tetR^{ti}-36	D5	+++	–	+++	+++	+++	+++	
tetR^{ti}-37	E5	+++	–	+++	+++	+++	+++	
tetR^{ti}-38	F5	++	+	+	++	++	+++	
tetR^{ti}-39	G5	+++	+	++	+++	+++	+++	
tetR^{ti}-40	H5	+++	–	+++	+	+++	+++	
tetR^{ti}-41	A6	+++	–	+++	+++	+++	+++	
tetR^{ti}-42	B6	+++	–	+++	+++	+++	+++	
tetR^{ti}-43	C6	+++	–	+++	+++	+++	+++	
tetR^{ti}-44	D6	+++	–	+++	+++	+++	+++	
tetR^{ti}-45	E6	+++	–	+++	+++	+++	+++	
tetR^{ti}-46	F6	+++	–	+++	+++	+++	+++	
tetR^{ti}-47	G6	+++	–	+++	+++	+++	+++	
tetR^{ti}-48	H6	+++	–	+++	+++	+++	+++	
tetR^{ti}-49	A7	+++	–	+++	+++	+++	+++	
tetR^{ti}-50	B7	+++	–	+++	+++	+++	+++	
tetR^{ti}-51	C7	+++	+	+++	+++	+++	+++	
tetR^{ti}-52	D7	+++	+	+++	+++	+++	+++	
tetR^{ti}-53	E7	+++	+	+++	+++	+++	+++	
tetR^{ti}-54	F7	+++	–	+++	+++	+++	+++	
tetR^{ti}-55	G7	+++	+	+++	+++	+++	+++	
tetR^{ti}-56	H7	+++	–	+++	+++	+++	+++	
tetR^{ti}-57	A8	+++	–	+++	+++	+++	+++	
tetR^{ti}-58	B8	+++	–	+++	+++	+++	+++	
tetR^{ti}-59	C8	+++	++	+++	+++	+++	+++	
tetR^{ti}-60	D8	+++	+	+++	+++	+++	+++	
tetR^{ti}-61	E8	+++	–	+++	+++	+++	+++	
tetR^{ti}-62	F8	+++	–	+++	+++	+++	+++	
tetR^{ti}-63	G8	+++	–	+++	+++	+++	+++	
tetR^{ti}-64	H8	+++	+	+++	+++	+++	+++	
tetR^{ti}-65	A9	+++	–	+++	+++	+++	+++	
tetR^{ti}-66	B9	+++	+++	++	+++	+++	+++	
tetR^{ti}-67	C9	+++	+	+++	+++	+++	+++	
tetR^{ti}-68	D9	+++	–	+++	+++	+++	+++	
tetR^{ti}-69	E9	+++	+	+++	+++	+++	+++	
tetR^{ti}-70	F9	+++	–	+++	+++	++	+++	
tetR^{ti}-71	G9	+++	+	+++	+++	++	+++	
tetR^{ti}-72	H9	+++	–	+++	–	+++	+++	

Continued on next page...

Table 3.4 – continued from previous page

		30°		37°		42°		
Label	Well	0	50	0	50	0	50 µg/ml	[Cm]
Empty	A10	N/A						
cat^{+}, *tetR*$^{-}$	B10	+++	+++	+++	+++	+++	+++	
cat^{+}, *tetR*$^{\mathrm{wt}}$	C10	+++	–	+++	–	+++	–	
cat^{-}, *tetR*$^{-}$	D10	+++	–	+++	–	+++	–	
cat^{-}, *tetR*$^{\mathrm{wt}}$	E10	+++	–	+++	–	+++	–	
cat^{+}, *tetR*$^{-}$	F10	+++	+++	+++	+++	+++	+++	
cat^{+}, *tetR*$^{\mathrm{wt}}$	G10	+++	–	+++	–	+++	–	
Empty	H10	N/A						

3.3.4 More detailed characterisation of TetR$^{\mathrm{ti}}$ mutants by Cm survival assay

The original pool of 72 clones carrying putative *tetR*$^{\mathrm{ti}}$ mutants (Fig. 3.3 and Table 3.4) was narrowed down to a set carrying the 44 most promising *tetR*$^{\mathrm{ti}}$ mutants and these were analysed with a more rigorous Cm survival assay. The clones were selected based on low growth on Cm in the repressed state (30°) as these should contain *tetR*$^{\mathrm{ti}}$ mutants able to most tightly repress gene expression at low temperature. This assay was performed as that described in Section 3.3.3, this time at more temperatures (30°, 33°, 35°, 37°and 39°C) and at both a low and high Cm concentration (50 and 20 µg, respectively). These 44 *tetR*$^{\mathrm{ti}}$ mutants, along with *tetR*$^{\mathrm{wt}}$ and *tetR*$^{-}$ controls, were grown in duplicate in 96-well format. Cultures were diluted back to equal density in a fresh plate before spotting the cultures onto solid medium with and without Cm. After 30 hours growth at the indicated temperature, these plates were observed and imaged (Fig. 3.4. None of the clones harbouring epPCR-derived TetR were able to grow at 30°, but these clones all displayed some amount of Cm$^{\mathrm{R}}$ phenotype at 33° and above, consistent with the original temperature sensitivity assay, with the only the *tetR*$^{-}$ control able to grow on Cm at 30°. However, for the *tetR*$^{\mathrm{ti}}$ mutants, variability between duplicate spots of the same clone on the same plate was significant

in some cases. The spot survival data presented here show a clear positive correlation between growth temperature and chloramphenicol resistance for the majority of clones analysed.

Table 3.5: Survival of select *tetR*^{ti} clones on 50 µg/mL at various temperatures. Level of growth is an average of two spots as shown in Figure 3.4 indicated by '+' signs with '–' for no detectable growth

		Growth on Cm at 50 µg/mL				
Strains	Wells	30°	33°	35°	37°	39°
tetR^{ti}-1	A1, A7	–	–	–	–	+
tetR^{ti}-2	B2, B7	–	–	–	–	–
tetR^{ti}-4	C1, C7	–	–	–	+	+++
tetR^{ti}-5	D1, D7	–	–	–	–	+
tetR^{ti}-6	E1, E7	–	–	+	+	+++
tetR^{ti}-7	F1, F7	–	–	++	+	++
tetR^{ti}-8	G1, G7	–	–	–	+	+++
tetR^{ti}-9	H1, H7	–	–	–	–	–
tetR^{ti}-10	A2, A8	–	+	–	+	+
tetR^{ti}-11	B2, B8	+	–	–	–	–
tetR^{ti}-14	C2, C8	–	+	+	+	+
tetR^{ti}-15	D2, D8	–	–	–	–	–
tetR^{ti}-16	E2, E8	–	+	+	+	+
tetR^{ti}-18	F2, F8	–	+	+++	+++	+++
tetR^{ti}-19	G2, G8	–	+	++	+	+
tetR^{ti}-20	H2, H8	–	–	+	+	–
tetR^{ti}-21	A3, A9	–	+	+++	+++	+++
tetR^{ti}-24	B3, B9	–	+	+++	+++	+++
tetR^{ti}-25	C3, C9	+	–	+++	+++	+++
tetR^{ti}-27	D3, D9	–	++	+++	+++	++
tetR^{ti}-29	E3, E9	–	–	+	++	++
tetR^{ti}-31	F3, F9	–	–	+	++	++
tetR^{ti}-32	G3, G9	–	+++	+++	+++	+++
tetR^{ti}-33	H3, H9	+	–	+	+	+++
tetR^{ti}-35	A4, A10	–	–	–	+	++
tetR^{ti}-36	B4, B10	–	–	–	–	++
tetR^{ti}-39	C4, C10	–	–	+	++	+++
tetR^{ti}-40	D4, D10	–	+	+++	+++	+++
tetR^{ti}-41	E4, E10	–	–	–	–	+++

Continued on next page...

Table 3.5 – continued from previous page

Strains	Wells	Growth on Cm at 50 µg/mL				
		30°	33°	35°	37°	39°
tetR^{ti}-42	F4, F10	–	–	+	+	+++
tetR^{ti}-43	G4, G10	+	+	–	–	++
tetR^{ti}-45	H4, H10	+	+	+	+++	+++
tetR^{ti}-47	A5, A11	–	+	+++	+++	+++
tetR^{ti}-48	B5, B11	–	+++	+++	+++	+++
tetR^{ti}-49	C5, C11	+	+	++	+	+
tetR^{ti}-53	D5, D11	–	–	–	–	++
tetR^{ti}-54	E5, E11	–	+	++	+++	++
tetR^{ti}-56	F5, F11	–	–	+	+	++
tetR^{ti}-58	G5, G11	+	+	++	++	++
tetR^{ti}-61	H5, H11	–	+	–	+	+++
tetR^{ti}-62	A6, A12	–	++	+++	+++	+++
tetR^{ti}-63	B6, B12	–	+++	+++	+++	+++
tetR^{ti}-68	C6, C12	–	+	++	+++	++
tetR^{ti}-72	D6, D12	–	–	–	–	++
cat^{-}, *tetR*$^{-}$	E6, E12	–	–	–	–	–
cat^{-}, *tetR*^{wt}	F6, F12	–	–	–	–	–
cat^{+}, *tetR*$^{-}$	G5, G12	+++	+++	+++	+++	+++
cat^{+}, *tetR*^{wt}	H6, H12	–	–	–	–	–

3.3.5 Temperature induction of select TetR^{ti} mutants by western blot

A western blot of TetR expressed from three different plasmids, alone and in combination with different plasmids expressing YFP from P_L-tetO suggest that the level of repression and induction achieved with ATc is influenced by the relative levels of repressor and target promoter (Fig. 3.2). Expression of TetR from pSB4A5-tetR and pSB6A1-tetR (which have pSC101 and pMB1 origin of replications, respectively), was detected as a band at the size expected size of TetR (23 KDa). Substantially more TetR was expressed from pSB6A1-tetR than from pSB4A5-tetR, which is expected based on the higher copy number the latter plasmid. No TetR-reactive band was observed for *E. coli* harbouring plasmid pBS-tetR, suggesting that expression is below

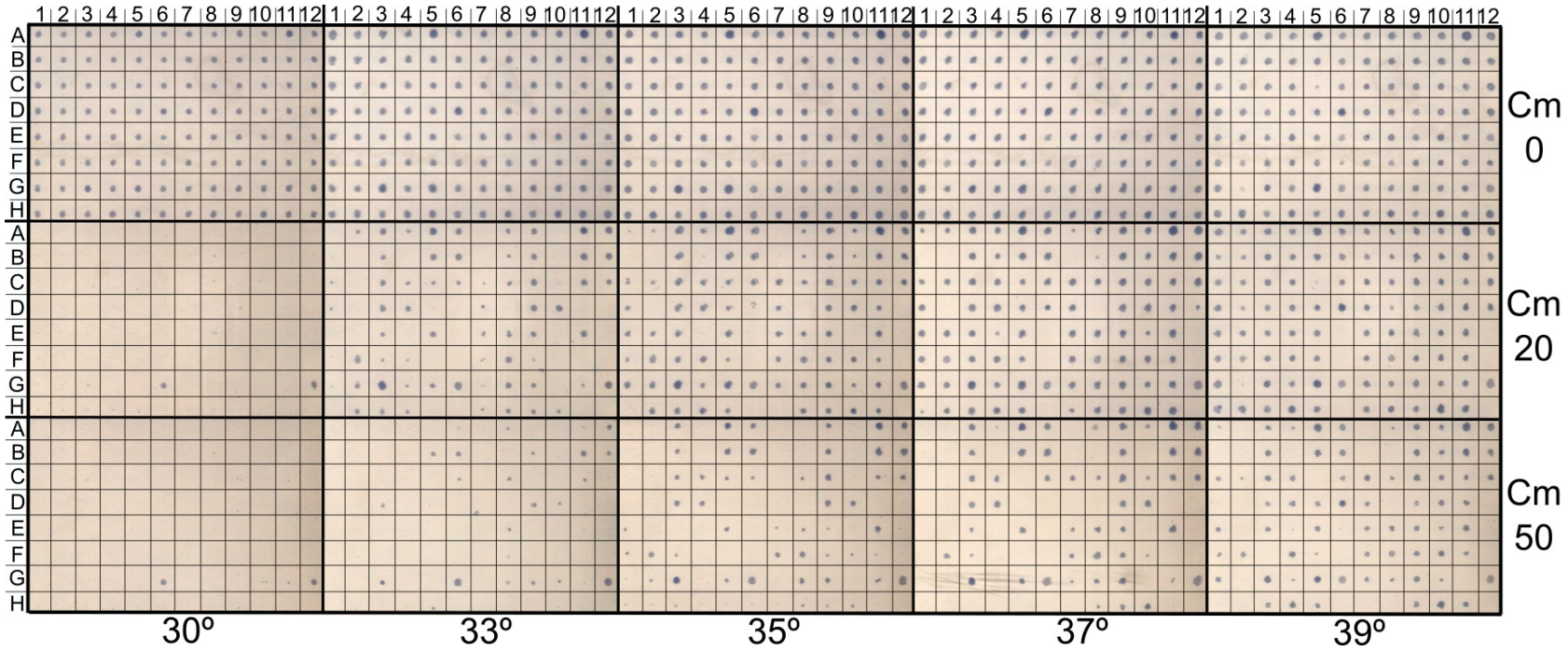


Figure 3.4: Further characterisation of select clones with a temperature-dependent Cm^R phenotype. Spot plate of 44 clones with promising $TetR^{ti}$ phenotype as determined by Cm resistance at high temperature but not at low temperature where spotted on solid media with and without Cm at various temperatures in duplicate with appropriate controls. See Table 3.5 for description of plate layout and tabulated growth data for each clone. Image is inverted for improved visual clarity of bacterial spots.

the detection level of the blot. This is curious considering that pBlueScript II is expected to be maintained at a much higher copy number than either the other *tetR* plasmids compared to and P_{lac} should be stronger than the weak promoter driving *tetR* expression on these other plasmids. YFP expression from pSB1K3-P_L-tetO-yfp and pSB3K3-P_L-tetO-yfp was similar and strong when no TetR was present; however, repression/induction of the three different TetR/P_L-tetO-yfp systems varied greatly. pSB4A5-tetR + pSB1K3-P_L-tetO-yfp gave a significant level of leaky YFP expression with only a minor increase upon induction with ATc. pSB1A6-tetR + pSB3K3-P_L-tetO-yfp produced no detectable leaky YFP expression and gave a strong band upon induction, although the induced levels of expression were still a fraction that in the absence of TetR. pBS-tetR + pSB3K3-P_L-tetO-yfp showed a very faint band of leaky expression in the repressed state but was completely induced upon addition of ATc. Based on these results, subsequent assays to characterise each $TetR^{ti}$ mutant in more detail was performed with *tetR*^{ti} mutants on the pBlueScript II vector with pSB3K3-P_L-tetO-yfp as the reporter. This analysis of $TetR^{ti}$ mutants 4, 15, 22, and 34 showed robust expression of YFP at 42°, but little to no detectable YFP in cultures maintained at 30°(Fig. 3.5).

3.3.6 Sequence analysis of temperature inducible repressors

The sequence of 39 *tetR*^{ti} mutants were determined and found all to be unique. The number of aa changes per mutant ranged from 1 to 10 with an average of 4.3 over the 206 residues of the TetR protein (Table 3.6). The position found to be mutated in the largest number of $TetR^{ti}$ mutants was Trp75 with substitutions in 12 of the 39 sequences. A W75R mutation was sufficient for a TI phenotype as the only mutant with a single aa change, $TetR^{ti}$-15, carried this substitution. Of the 12 substitutions at W75, 10 were W75R, with W75C and W75G also represented. The high number

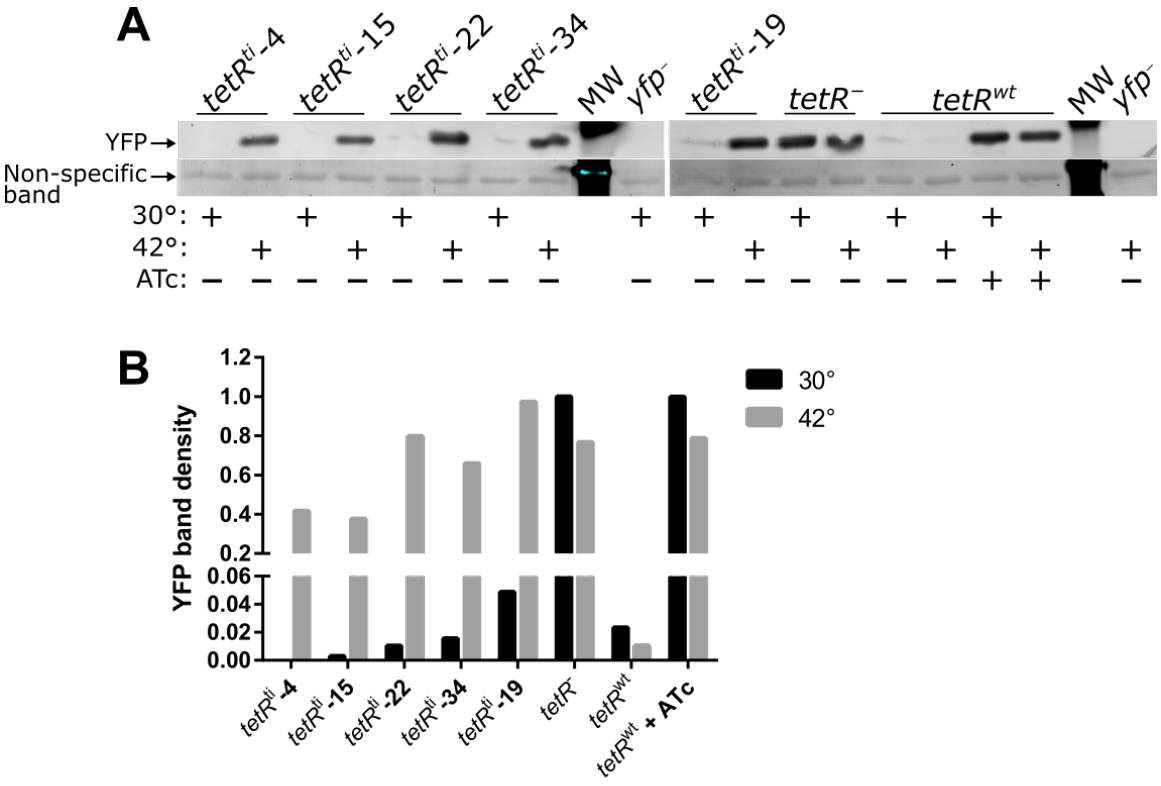


Figure 3.5: Western blot analysis of YFP expression from select $tetR^{\mathrm{ti}}$ clones at 30° and 42°. A) Western blot for TetR-controlled expression of YFP from five $tetR^{\mathrm{ti}}$ mutants along with $tetR^{-}$ and $tetR^{\mathrm{wt}}$ controls. ATc induced YFP expression from the $tetR^{\mathrm{wt}}$ strain are included for comparison to temperature induction from the mutant strains. B) YFP expression data from the blot in part (A) represented as intensity of the YFP bands divided by the intensity of the non-specific band indicated. The intensity values and induction ratios are tabulated in Table 3.6

Table 3.6: Expression levels and induction ratios of TetR^{ti} controlled YFP after temperature induction at 42° as determined by intensity of YFP immunoblot bands in Figure 3.5.

	Expression			Induction ratio	
Clone tested	30°	42°	42° normalized†	Original	Normalized†
tetR^{ti}-4	0.000049	0.417	0.544	8364	10907
tetR^{ti}-15	0.0030	0.377	0.492	124	162
tetR^{ti}-22	0.0103	0.798	1.04	77.2	101
tetR^{ti}-34	0.0156	0.659	0.859	42.2	55.0
tetR^{ti}-19	0.0487	0.973	1.27	20.0	26.1
tetR^{-}	1.00	0.767	1.00	0.767	1.00
tetR^{wt}	0.023	0.010	0.0135	0.446	0.581
tetR^{wt} + ATc	0.999	0.787	1.03	0.788	1.03

† Normalized to make up for the reduced expression from lambda left promoter. "Normalized" values define *tetR*$^{-}$ expression at 42° as 1.00 and scale other expression values at 42° accordingly.

of W75R mutations is consistent with a strong selection for this specific substitution rather than it being an artefact of the epPCR process, resulting from a mutation established within the first couple rounds of epPCR then faithfully carried through the remainder of the reaction so to be present in a large proportion of PCR products. The W75R mutation was the only substitution found to be sufficient to produce TetR^{ti} phenotype (as seen in TetR^{ti}-15; DNA sequences are presented in Table A.2). Furthermore, the Trp codon (`TTG`) was mutated to the Arg codon `CGG` in 7 of the 12 mutants and to `AGG` in the remaining 3, which is consistent with the ratio of T $\rightarrow$ C and T $\rightarrow$ A mutants found when analyzing epPCR mutation without selection (Table 3.2). The single W75C and W75G substitutions resulted from G $\rightarrow$ T and T $\rightarrow$ G SNPs, respectively, which were both found to be rare using this mutagenesis procedure. The next most commonly substituted positions were N47, K98, C144 and I194, each of which were seen mutated in four different TetR^{ti} variants (Fig. 3.7). Amino acid substitutions were observed in 96 of the 206 positions found all throughout the protein sequence, although fewer changes were observed in the N-terminal end

of the protein which acts as the DNA binding head which contained an average of 0.30 substitutions per residue (residues 1–50; 15 substitutions) over the 39 sequences. The remainder of the protein received an average of 0.87 substitutions per residue (Fig. 3.6).

Due to the presence of multiple aa substitutions in all but one of the sequenced $TetR^{ti}$ variants it is difficult to determine the contribution any single mutant makes to the TI phenotype. However, information may be gained by comparison between $TetR^{ti}$-15 (W75R) and $TetR^{ti}$-2 (W75R, T160I) or $TetR^{ti}$-8 (W75R, H66Y), as the latter two variants share the W75R substitution with $TetR^{ti}$-15 but contain one addition change each. The T160I substitution of $TetR^{ti}$-2 is in a disordered region between helices α_8 and α_9, which shares almost no conserved residues between any of the 7 highly related TetR classes (A–E, G and H; [143, 230]). $TetR^{ti}$-2 shows a very similar temperature induced Cm^R phenotype to that of $TetR^{ti}$-15 in each of the Cm survival assays performed where clones carrying either of these mutant repressors showed only a Cm^R phenotype at 42° (Fig. 3.3, Table 3.4, or at 39°, but only at the lower Cm concentration of 20 µg/mL (Fig. 3.4, Tables 3.5 and A.3). $TetR^{ti}$-8, however, allows induction of a Cm^R phenotype at lower temperature than $TetR^{ti}$-2 or -15 with observable growth at 33° at 20 µg/mL Cm or 37° at 50 µg/mL Cm.

3.4 Discussion

In this work we attempted to isolate a mutant of the tetracycline repressor that would allow gene expression from its promoter at elevated temperature. A genetic system was created in which the *cat* gene, providing chloramphenicol resistance, was transcribed from a TetR-controlled promoter so that the host cell could not survive in the presence of Cm when a plasmid expressing a functional *tetR* gene was also

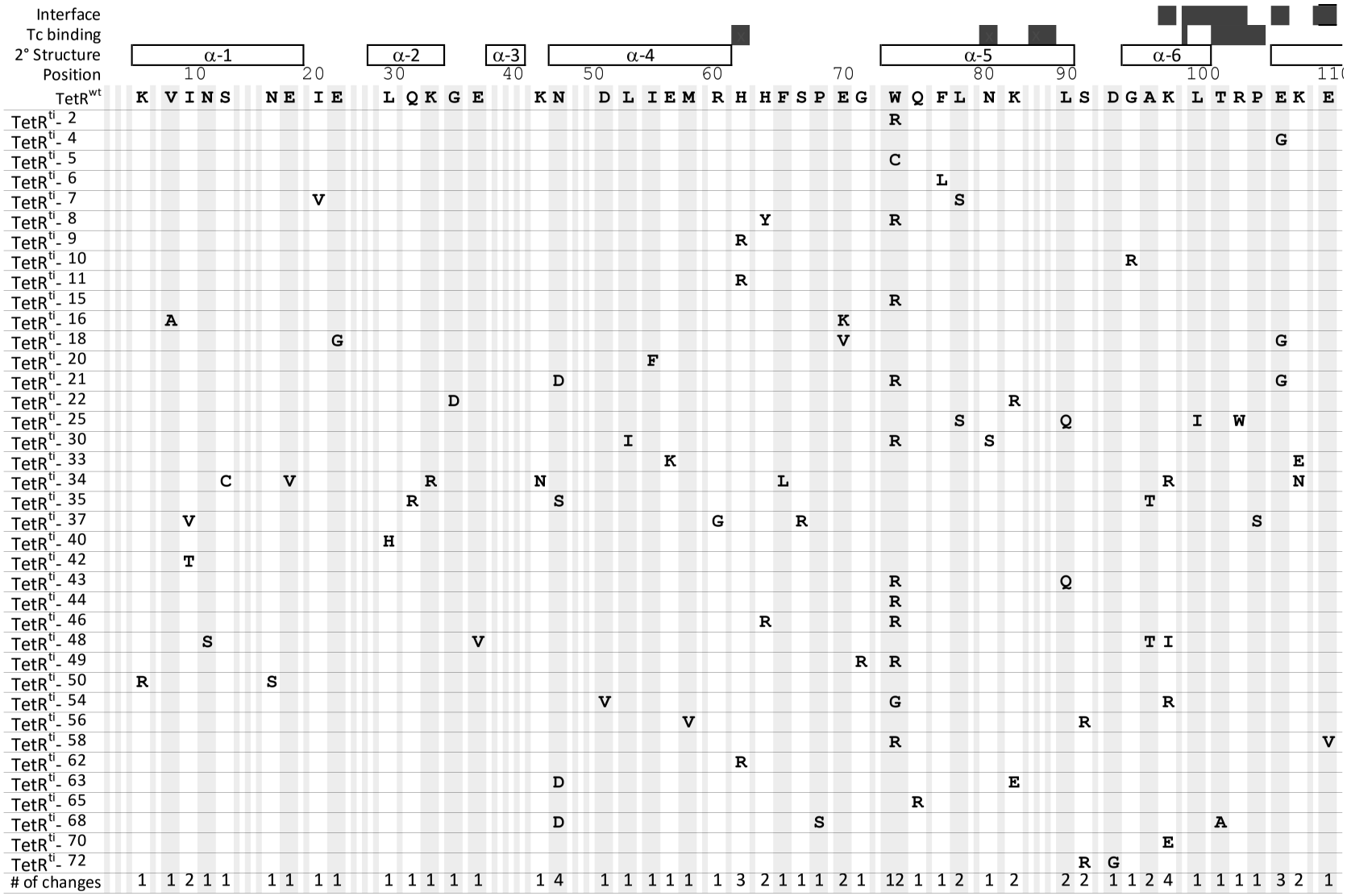
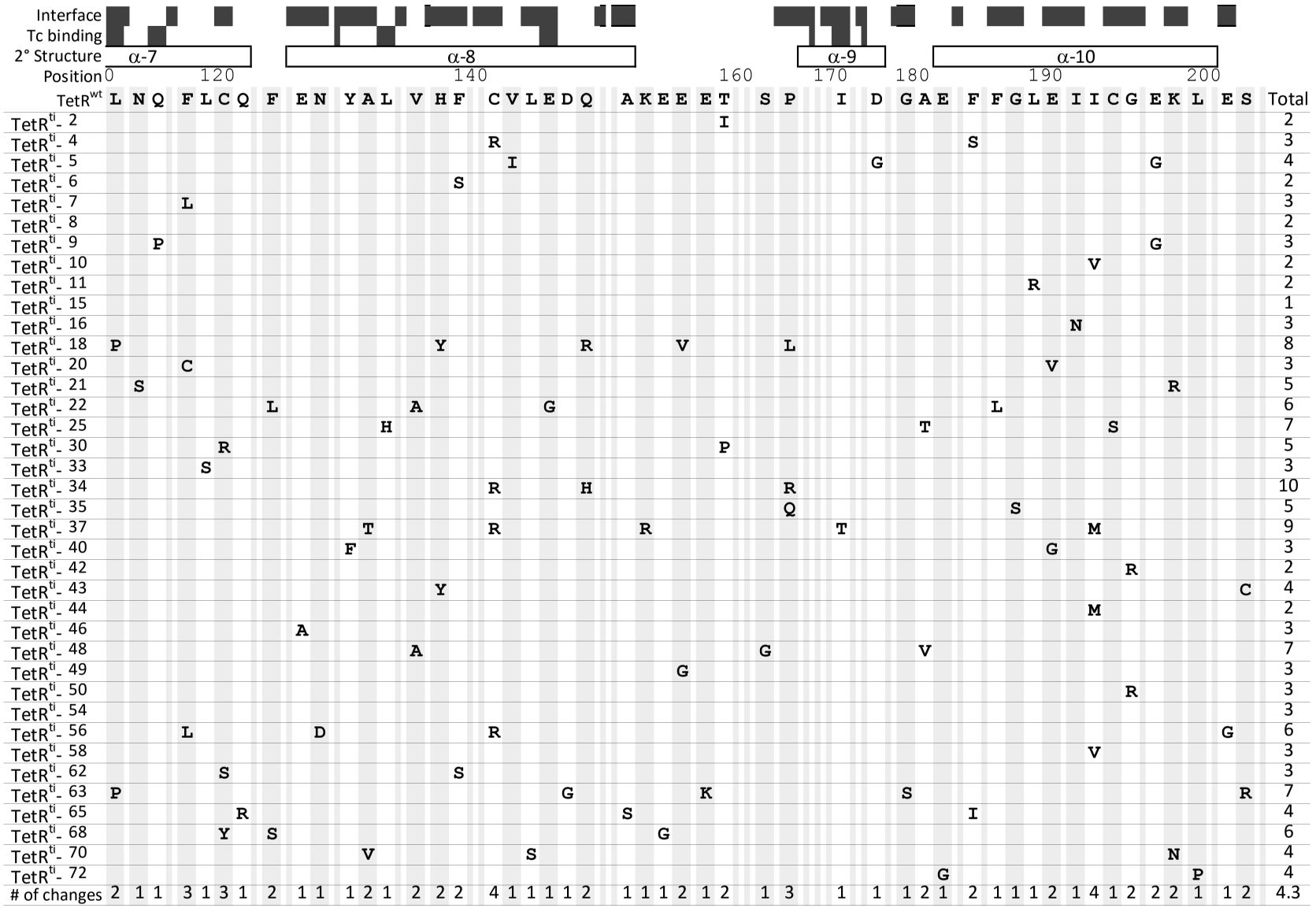


Figure 3.6: Amino acid changes in the 39 sequenced $TetR^{ti}$ variants. Sequence of $TetR^{wt}$ is shown above (only positions where changes were found). Black shading and white rectangle above sequences indicate amino acids involved in dimerization interface, Tc binding, and α-helices of the TetR protein. Numbers at bottom indicate the number of aa changes observed at a specific position among the 39 mutants.



	TetR^{wt}	L	N	Q	F	L	C	Q	F	E	N	Y	A	L	V	H	F	C	V	L	E	D	Q	A	K	E	E	E	T	S	P	I	D	G	A	E	F	F	G	L	E	I	I	C	G	E	K	L	E	S	Total
TetR^{ti}- 2																													I																							2
TetR^{ti}- 4																		R																				S																3
TetR^{ti}- 5																			I															G														G					4
TetR^{ti}- 6																	S																																					2
TetR^{ti}- 7					L																																																3
TetR^{ti}- 8																																																					2
TetR^{ti}- 9			P																																																	3
TetR^{ti}- 10																																																					2
TetR^{ti}- 11																																																					2
TetR^{ti}- 15																																																					1
TetR^{ti}- 16																																																					3
TetR^{ti}- 18	P															Y						R				V						L																				8
TetR^{ti}- 20					C																																																3
TetR^{ti}- 21		S																																																		5
TetR^{ti}- 22									L						A						G																															6
TetR^{ti}- 25														H																							T															7
TetR^{ti}- 30							R																						P																							5
TetR^{ti}- 33						S																																														3
TetR^{ti}- 34																		R				H								R																						10
TetR^{ti}- 35																														Q							S														5
TetR^{ti}- 37												T						R						R								T											M							9
TetR^{ti}- 40											F																																	G							3
TetR^{ti}- 42																																														R						2
TetR^{ti}- 43																Y																																		C		4
TetR^{ti}- 44																																													M						2
TetR^{ti}- 46										A																																									3
TetR^{ti}- 48															A														G							V															7
TetR^{ti}- 49																										G																										3
TetR^{ti}- 50																																													R						3
TetR^{ti}- 54																																																				3
TetR^{ti}- 56					L					D								R																																G		6
TetR^{ti}- 58																																													M						3
TetR^{ti}- 62							S										S																																		3
TetR^{ti}- 63	P																				G							K							S														R	7
TetR^{ti}- 65								R																S														I												4
TetR^{ti}- 68							Y		S																G																									6
TetR^{ti}- 70													V							S																												N			4
TetR^{ti}- 72																																						G										P			4
# of changes	2	1	1		3	1	3	1		2		1	1		1	2	1		2	2	2		4	1	1	1	1	2		1	1	1	2	1	2		1	3		1		1		1	2	1		2		1	1	1	2	1	4	1	2		2	2	1		1	2	4.3

Figure 3.6a: Amino acid changes in the 39 sequenced TetR^{ti} variants, continued

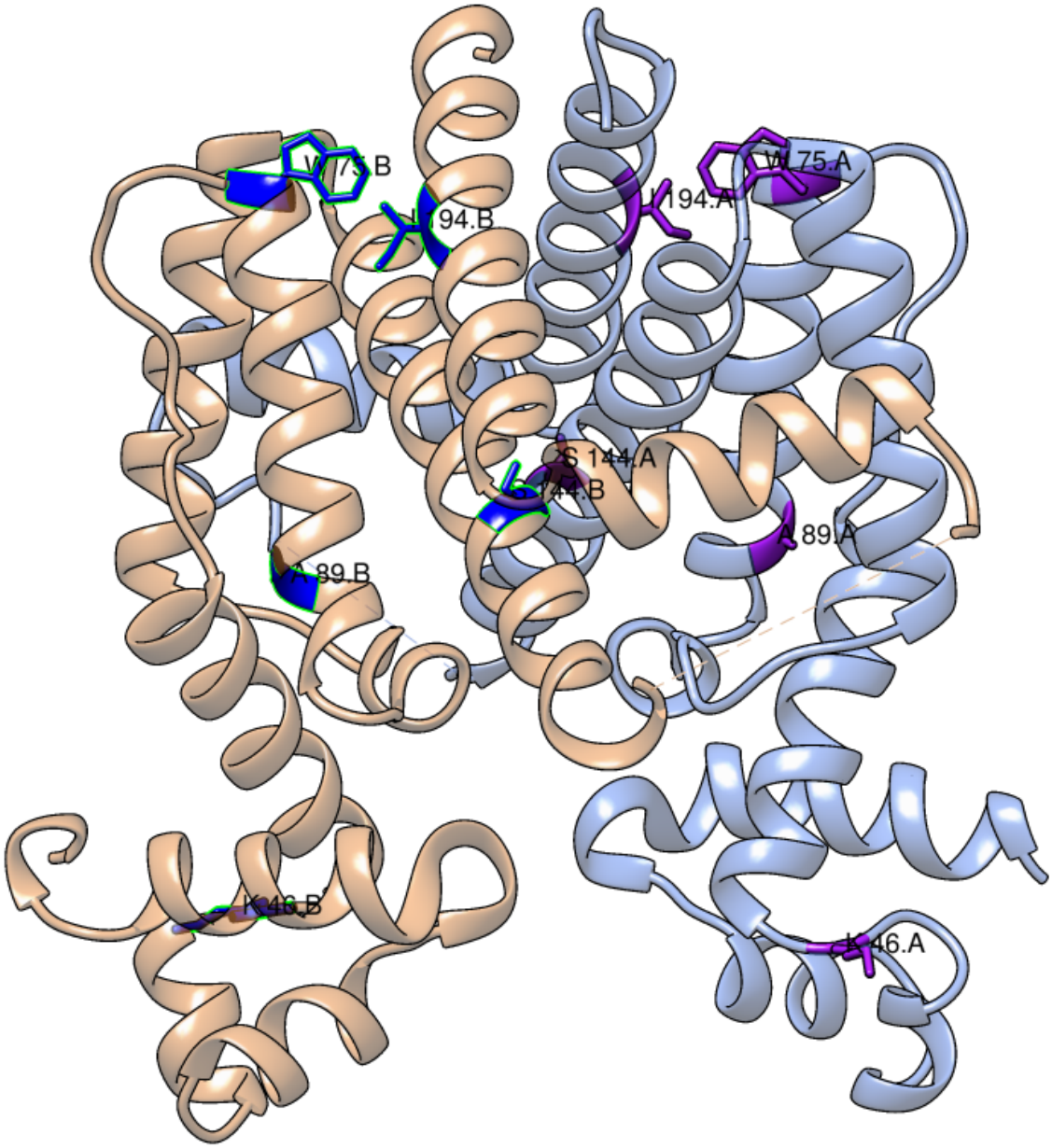


Figure 3.7: Common amino acid substitutions found in the TetR^{ti} mutants mapped onto the 3D structure of TetR. Residues N46, W75, K98, S144 and I194 (all of which are found mutated in four or more of the 39 sequenced TetR^{ti} mutants) are highlighted on both chains of the TetR structure (PDB 2ns7; [229]).

present. When plasmids containing random mutants of the *tetR* gene were introduced, they would be expected to result mostly in Cm^R clones because of the high likelihood of missense mutation or amino acid substitutions that abolish repressor function. However, our hope was that some products of *tetR* mutation would still encode a functional protein, at least at lower temperature, but accumulated mutations would result in a less stable protein structure that loses its ability to maintain its structure—and therefore its repressor activity—at elevated temperature. It turned out that many clones harbouring random mutants TetR were unable to repress expression of *cat* at 42°, but when grown at 30°, a small but significant number of these same clones were unable to grow on Cm, suggesting that TetR was functional in repressing the *cat* gene only at the lower temperature.

Finding a mutant that has the ability to be induced by temperature is not such an obscure or specific property. As all proteins will cease to be active at some temperature; presumably any repressor is "inducible" by temperature at some temperature value. Here we just set up selection and screening criteria that isolated mutants with reduced stability such that the induction temperature overlaps with viable growth temperatures. In order to create a temperature inducible mutant of an existing repressor protein what is required is a mutation, or combination of multiple mutations, that lower the stability of the protein so that it will no longer function to block transcription at a temperature lower than the maximal survival temperature of the organism, but is still able to function at a temperature high enough to useful for culturing this organism. This is likely how so many $TetR^{ti}$ mutants could be identified with relative ease.

In this case, finding a less stable variant has the obvious benefit of allowing control of gene expression by temperature, but much more effort has gone into engineering proteins for quite the opposite property: increased thermostability. Enzymes with improved thermostability can have uses in industrial processes that would otherwise

be too hot for mesophilic enzymes to be employed [231]. This has spurred research into the molecular details of thermostability in order to better predict changes that would increase thermostability and allow more successful rational design of engineered thermotolerant enzymes. Few examples exist of proteins engineered intentionally for increased thermosensitivity.

The epPCR method used to create our random mutant library of TetR variants employed a mutagenic PCR buffer that reduces the fidelity of DNA replication by Taq DNA polymerase. This is achieved by increasing the concentration of DNA polymerase co-factor, Mg^{2+} and adding some Mn^{2+}, another divalent cation that can take the place of Mg^2 in interactions between negative charged residues on the polymerase and the phosphates of incoming dNTPs. Increased concentrations of dTTP and dCTP are also included to counteract for bias of misincorporated bases inherent in this method [227]. However, analysis of PCR products created by this technique still have an over-representation of A $\rightarrow$ G transitions and an under-representation of A:T $\rightarrow$ G:C and G:C $\rightarrow$ T:A transversions. Not a single G:C $\rightarrow$ C:G transversion was not observed during this test, but was found at very low rate among the sequences of the $TetR^{ti}$ mutants. The bias observed here was very similar to that observed by Rasila et al. in a study that compared various random mutagenesis methods [228]. The near lack of G:C $\rightarrow$ C:G SNPs limited the overall sequence space that could be tested. For instance, the Trp75 residue was changed to Arg in 10 of the 39 $TetR^{ti}$ mutants sequenced, but never to Ser (`TGG` $\rightarrow$ `TCG`) or Leu (`TGG` $\rightarrow$ `TTG`), changes that just require a single nucleotide change in the Trp codon. Of course, these substitutions could be absent because they fail to cause a temperature inducible phenotype or abolish repressor function completely, but their absence from the mutants recovered here is consistent with the low rate at which the necessary SNPs are created. The most likely mutation for the trp codon is to a stop codon (`TAG` or `TGA`). As 21 of 50

aa substitutions observed in the epPCR products to which no selection was applied were G $\rightarrow$ A substitution, and only 13 of 50 were T $\rightarrow$ A or C, a mutation in the Trp codon produced by the epPCR process employed here would be almost 4$\times$more likely to produce a stop codon than it is to produce an Arg codon. This bias could be reduced by the use of mutagenic Taq polymerase variants, such as Mutazyme II, which provides a more balanced mutagenic profile [228] than the epPCR procedure used here. Reduced mutational bias could be especially valuable in situations where more complex mutant phenotypes are desired.

TetR^{ti}-15 with its W75R mutation was found to be sufficient for temperature induction, but accumulation of additional substitutions resulted in mutants that allowed more robust temperature induction at lower temperatures. TetR^{ti}-2, which also contained the W75R substitution, with an added substitution in a non-conserved loop unlikely to contribute to protein function, was comparable phenotypically to TetR^{ti}-15, as would be expected if its TI phenotype was due only to W75R. This observation supports the notion that the W75R TetR^{ti} phenotype observed in TetR^{ti}-15 is sufficient for the phenotype and does not require a secondary mutation elsewhere in the host genome. Such observations also vouches for the reproducibility of the Cm survival assay used to compare the TI phenotype of these two similar mutants. Inspection of the protein structure of the TetR dimer reveals a probable mechanism for loss of protein stability upon substitution of Arg for Trp at position 75. W75 is found within α_5, which is bundled tight with with α_8 and α_{10} (including the $\alpha_{8'}$ and $\alpha_{10'}$ of the other subunit) to form the rigid, hydrophobic core of the protein. Changing this large, aromatic residue to a charged residue should greatly destabilize the hydrophobic packing in this region. Furthermore, positions 74–77 make up a helix capping box, which is a motif that signals the end of the helix. Ser74 is the 'N-cap' at the N-terminal end of α_5 and makes specific H-bonds with Asp77 which help interrupt

the H-bonding between atoms of the backbone that stabilize the α-helix structure, which causes an end of the helix. This major substitution in neighbouring position 75 could weaken the H-bonding of the N-cap residues and destabilize the overall protein structure [232].

Interestingly, the impact of substitutions at W75 have been analyzed previously, as part of a study that examined structural changes induced by binding of TetR to Tc [233]. As tryptophan fluorescence can vary with changes in the surrounding protein structure this can be used as a tool to study protein dynamics. As TetR has two Trp residues, W43 and W75, one was mutated to Phe so to not interfere with fluorescence from the other. The authors noticed a drop in the temperature at which TetR denatures (T_{max})—from 64.2° to 50.0°C—upon introduction of the W75F mutation [233, 234]. This provides some context for temperature inducibility of a W75R mutant, which displays loss of function starting around 39°, which is about 11° lower than that observed for substitution with Phe, another non-polar, aromatic residue similar to Trp. However, that work was done with TetR in vitro and did not involve a cellular system, which may make comparison meaningless between T_{max} and the induction temperature observed here.

One factor that may affect temperature induction by $TetR^{ti}$ mutants not addressed here is concentration of TetR protein present in the cell. Increase of repressor concentration over that which is needed to sufficiently occupy all operators can have negative effects due to increased opportunity to interfere with host function via off-target binding and can also reduce the sensitivity of induction [136]. The importance of regulating repressor concentration at low levels to the overall process is evidenced by the evolution of a negative feedback system in many natural gene induction systems, including TetR, whereby the repressor controls not only transcription of its target gene, but also of itself. This allows for the repressor to shut off its own expression

once its concentration is sufficient to do so; more repressor will be made once levels drop below that at which it can maintain repression of itself. Influence of repressor concentration on sensitivity of induction can de explained in terms of the equilibrium between non-induced repressor [R] and repressor bound to inducer [RI], which will theoretically, never go fully to [RI] no matter what the concentration of inducer molecule added. More total repressor present ([R] + [RI]) will result in a larger pool of [R], which may still be sufficiently large to repress transcription, at least somewhat, even in the presence of a large excess of inducer. A similar principle should apply to repressors induced by temperature. As thermal denaturation should resemble a stochastic process between native folded and denatured protein, a larger pool of total repressor will mean more remaining in the native state at a given temperature.

The $TetR^{ti}$ mutants produced here have accumulated many changes in the nucleotide and protein sequence which could have consequences for transcription, translation, and stability of both transcript and protein. Such changes could potentially affect the steady-state TetR concentration in the cell which could change the sensitivity of the repressor to induction, by temperature or otherwise, for the reasons just described. Furthermore, unfolding of unstable $TetR^{ti}$ mutants will be effected by the heat-shock response of the cell, which may use chaperones to refold denatured proteins or could act to degrade the repressors at high temperature. The repressor concentration could also affect the heat-shock response. If induction suddenly creates a large pool of unfolded protein this may be poorly tolerated and could potentially lead to aggregation and overwhelm the host's protein turnover machinery. This could influence the temperature induction phenotype (T_{max}) for the $TetR^{ti}$ mutant under study and have implications for host fitness in general. Further study or application of these, or any, temperature induced repressor should take these issues into account by characterizing repressor temperature induction phenotype in context of its concentration if mutants

are to be directly compared.

Recent work has demonstrated transcription induction systems that respond to light, which has been coined as "optogenetics" [235]. Such systems employ cyanobacterial phytochromes that activate a two-component signaling system in response to light of a certain wavelength range. This results in phosphorylation of an activating transcription factor which switches on transcription at its cognate promoter [236]. Optogenetic systems allow induction of gene expression in response to a non-chemical inducer, which is reversible and preserves potential growth medium reuse, as is the case with temperature induction. However, optogenetics may have some advantages over heat induction. Light of most wavelengths will not significantly affect the growth properties of many cell types typically used as hosts for such induction systems, which is not necessarily the case for a shift in temperature. Also, multiple optogenetic activator systems can operate independently in a single host as the different systems can be designed to respond to distinct light wavelengths [237]. Potential pitfalls of such systems include inability of light to penetrate through large, dense cultures, as may be encountered in industrial applications. Also, use with cyanobacteria or other pigmented host organisms may be difficult or impossible.

Here I have demonstrated the identification of many TetR mutants that allow expression of a target gene to be induced by increased temperature. Many of these TetR^{ti} mutants were were able to tightly control expression of the *cat* gene in *E. coli* so that survival on chloramphenicol could be controlled by ambient temperature. A range of mutant repressors able to induce gene expression over a range of different temperatures were identified. Such repressors may have application where induced gene expression is required but addition of a chemical inducer molecule is not acceptable.

Chapter 4

Conclusions and applications

As is the case with many research projects, the objectives of this work took a few turns over the course of it's execution. The original focus was the creation of temperature-inducible (TI) repressor proteins for the control of lethal proteins in a pathogen. Strains harbouring this "kill-switch" could grow normally under the temperature at which the mutant repressor is induced. If this temperature allows induction near core body temperature, but not at that of the extremities, such strains could have potential as a live vaccine. They may simulate normal infection at the skin, provoking a rigorous immune response, but not allow systemic infection; the kill-switch would be flipped as the pathogen migrated further into warmer, internal areas of the host. Not long after beginning this work a grant was secured for creating temperature inducible protein production platforms in cyanobacteria. Such organisms could produce recombinant proteins using CO_2 and sunlight as substrate. The ability to induce protein expression using a temperature shift, rather than addition of a chemical compound, could allow reuse of growth medium, reducing water use and cost of such a system. These projects, although unrelated in application, have in common the requirement for temperature-inducible gene expression in non-model bacterial species. It quickly became apparent

that if the goals of either project were to be met it would be necessary to identify or create promoters that could respond to a TI repressor while functioning in the specific host of interest.

There were multiple ways such promoters could be found or produced. One could attempt to identify a natural promoter with it's cognate repressor then attempt to isolate TI mutants of that repressor. This option is not attractive as new promoter–repressor pairs must be identified for each host, requiring again engineering a TI mutant of the repressor anew for each different host to be used. And as this control system is duplicated from the host, potential interaction with normal cell metabolism may occur. Another option would be to alter a natural promoter from the organism in which temperature-regulated gene expression is to be added so that this promoter is controlled by the TI repressor mutant one wishes to use. This approach potentially allows one TI repressor to be used in any host organism, as long as a promoter from that host could be modified to respond that repressor. In this case only one repressor would need to be engineered for temperature-inducibility, greatly reducing the difficulty of porting TI gene expression to different organisms. However, this approach still suffers from the problem of potential interference and genetic instability resulting from duplication of host control sequences. Furthermore, if a desire to reduce transcription levels arose, the process would have to be started all over again with a promoter of different strength. Preliminary attempts to create TetR-repressed versions of natural promoters of *Francisella* (which, as an intracellular pathogen, was an ideal host for proof-of-principle of the "kill-switch" project), were not successful, presumably because insertion of *tetO* disrupted sequence determinants necessary for promoter function. This indicated that perhaps engineering promoters for control by a heterologous TF may be more difficult than suggested in the literature, which would obviously be bias towards successful cases.

Putting aside these attempts to create *Francisella* promoters repressed by TetR (a repressor we had by this point had some success identifying TI mutants of) through rational design, I instead devised a method to simply select for appropriate *tet*-controlled promoter sequences from a library of DNA fragments that consisted of *tetO* flanked by randomized sequence. Inserted upstream of a antibiotic resistance gene, any such fragment that resulted in a resistance phenotype must promote expression of the resistance marker. The resultant promoters could then be screened for control by TetR by measuring expression of a downstream reporter gene in the presence and absence of an inducing molecule. Promoters that display control should also be controlled by mutants of TetR that are inducible by temperature as these bind to the same operator. Such an approach does not require a priori knowledge of promoter architecture in the host of interest and has the potential to generate many promoters with different expression strengths providing options in cases where various expression levels are required. A method was developed by which promoters could be identified from random DNA sequence based on their ability to drive expression of an antibiotic resistance gene, selecting against the vast majority of sequences that do not promote transcription. This approach was more successful than expected and resulted in more *tet*-controlled than could be characterized, as well as many promoters that were unresponsive to TetR. In fact, the success of this method was so exciting that the downstream applications (the TI kill-switch in pathogenic bacteria and TI protein expression in cyanobacteria) were mostly forgotten in favour of studying the action of these synthetic promoters in *Francisella* and applying this method to produce *tet*-controlled promoters in *E. coli*.

Perhaps the most compelling reason to develop synthetic DNA parts for use in genetic engineering applications rather than re-purposing DNA sequences "borrowed" from the chromosome of that host, relates to the issue of genetic stability. This

is becoming more of a problem as the idea of "biobricks" has grown in popularity, and people are building gene circuits in which the same part (eg. a promoter or terminator) is used many times in constructing a genetic pathway. These multiple copies of 100% homologous sequence can undergo a recombination event with greatly increased probability compared to unrelated sequence. A study of this phenomenon in *E. coli* K12 found that a constitutively expressed GFP reporter gene would be completely lost from a plasmid within an average of 30 generations when flanked by identical DNA sequence of about 140 bp each, but was maintained faithfully when lacking these homologous flanks [168]. This result, and others like it [169], demonstrate the potential instability created by repeated sequences in genetic circuit design and highlight the potential benefit of using unique control sequences (such as promoters) in synthetic DNA constructs, in contrast to current common practice.

The tools and techniques presented have been demonstrated in specific bacterial hosts (ie. *F. novicida* and *E. coli*) but are likely applicable to a wide variety of genetically tractable organisms. As synthetic biology and industrial biotechnology expand in sophistication and find new application, new synthetic systems for gene expression control will be required. Complex gene circuits can require that many transcriptional regulatory systems (ie. TFs with cognate promoters) operate simultaneously within the same cell (eg. biological computing endeavors [157]). This may be possible for thoroughly studied model organisms for which a large number of TF/promoter pairs have been identified and characterized, as is the case for *E. coli*. For most other microbes with potential biotechnological value, engineering of such species is limited by lack of available genetic tools, including promoter/TF pairs for precise control of gene expression. The process described in Chapter 2, should allow for creation of promoters that respond to most any TF for which the operator sequence is known and adapted for function in any host. The ability to easily create a series of promoters

that contain the operator of one's choice could allow temperature inducible repressors presented in Chapter 2 to be used in organisms for which no *tet*-controlled promoters previously existed, or for creation of a series of optogenetic-regulated promoters of synthetic sequence which provide a wider range of expression strengths and greater genomic stability than repeated use of a couple previously characterized suitable promoters many times within a complex genetic circuit. Such systems could aid in the creation and optimization of engineered microbial hosts for increased production of valuable metabolites, pharmaceuticals, biofuels or other potential outputs of industrial biotechnology.

References

1. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* **3,** 318–56 (1961).
2. Reeve, J. N. Archaeal chromatin and transcription. *Molecular Microbiology* **48,** 587–98 (2003).
3. Burgess, R. R. Separation and characterization of the subunits of ribonucleic acid polymerase. *The Journal of Biological Chemistry* **244,** 6168–76 (1969).
4. Murakami, K. S. X-ray crystal structure of *Escherichia coli* RNA polymerase σ^{70} holoenzyme. *The Journal of Biological Chemistry* **288,** 9126–34 (2013).
5. Blatter, E. E., Ross, W., Tang, H., Gourse, R. L. & Ebright, R. H. Domain organization of RNA polymerase α subunit: C-terminal 85 amino acids constitute a domain capable of dimerization and DNA binding. *Cell* **78,** 889–96 (1994).
6. Murakami, K. S., Masuda, S. & Darst, S. A. Structural basis of transcription initiation: RNA polymerase holoenzyme at 4 Å resolution. *Science* **296,** 1280–4 (2002).
7. Nüsslein, C & Heyden, B. Chromatography of RNA polymerase from *Escherichia coli* on single stranded DNA-agarose columns. *Biochemical and Biophysical Research Communications* **47,** 282–9 (1972).
8. Ghosh, P., Ishihama, A. & Chatterji, D. *Escherichia coli* RNA polymerase subunit ω and its N-terminal domain bind full-length β' to facilitate incorporation into the $\alpha_2\beta$ subassembly. *European Journal of Biochemistry* **268,** 4621–7 (2001).
9. Burgess, R. R., Travers, A. A., Dunn, J. J. & Bautz, E. K. Factor stimulating transcription by RNA polymerase. *Nature* **221,** 43–6 (1969).
10. Gruber, T. M. & Gross, C. A. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annual Review of Microbiology* **57,** 441–66 (2003).
11. Mittenhuber, G. An inventory of genes encoding RNA polymerase sigma factors in 31 completely sequenced eubacterial genomes. *Journal of Molecular Microbiology and Biotechnology* **4,** 77–91 (2002).
12. Bentley, S. D. *et al.* Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **417,** 141–7 (2002).
13. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453–62 (1997).
14. Meibom, K. L. *et al.* The heat-shock protein ClpB of *F. tularensis* is involved in stress tolerance and is required for multiplication in target organs of infected mice. *Molecular Microbiology* **67,** 1384–401 (2008).
15. Buck, M & Cannon, W. Specific binding of the transcription factor sigma-54 to promoter DNA. *Nature* **358,** 422–4 (1992).

16. Vassylyev, D. G. *et al.* Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6 Å resolution. *Nature* **417,** 712–9 (2002).

17. Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. & Darst, S. A. Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science* **296,** 1285–90 (2002).

18. Craig, M. L., Suh, W.-C. & Record, M. T. HO$^{\bullet}$ and DNase I probing of Eσ^{70} RNA polymerase-λP_R promoter open complexes: Mg^{2+} binding and its structural consequences at the transcription start site. *Biochemistry* **34,** 15624–32 (1995).

19. Gries, T. J., Kontur, W. S., Capp, M. W., Saecker, R. M. & Record, M. T. One-step DNA melting in the RNA polymerase cleft opens the initiation bubble to form an unstable open complex. *Proceedings of the National Academy of Sciences* **107,** 10418–23 (2010).

20. Ruff, E. F., Record, M. T. & Artsimovitch, I. Initial events in bacterial transcription initiation. *Biomolecules* **5,** 1035–62 (2015).

21. Haugen, S. P., Ross, W., Manrique, M. & Gourse, R. L. Fine structure of the promoter-σ region 1.2 interaction. *Proceedings of the National Academy of Sciences* **105,** 3292–7 (2008).

22. Kontur, W. S., Saecker, R. M., Davis, C. A., Capp, M. W. & Record, M. T. Solute probes of conformational changes in open complex (RP_O) formation by *Escherichia coli* RNA polymerase at the λP_R promoter: evidence for unmasking of the active site in the isomerization step and for large-scale coupled folding in the subsequent conversion to RP_O. *Biochemistry* **45,** 2161–77 (2006).

23. Roberts, C. W. & Roberts, J. W. Base-specific recognition of the nontemplate strand of promoter DNA by *E. coli* RNA polymerase. *Cell* **86,** 495–501 (1996).

24. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell* **98,** 811–24 (1999).

25. Davis, C. A., Bingman, C. A., Landick, R., Record, M. T. & Saecker, R. M. Real-time footprinting of DNA in the first kinetically significant intermediate in open complex formation by *Escherichia coli* RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **104,** 7833–8 (2007).

26. Kapanidis, A. N. *et al.* Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science* **314,** 1144–7 (2006).

27. Revyakin, A., Liu, C., Ebright, R. H. & Strick, T. R. Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science* **314,** 1139–43 (2006).

28. Vogel, U & Jensen, K. F. The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *Journal of Bacteriology* **176,** 2807–13 (1994).

29. Bar-Nahum, G. *et al.* A ratchet mechanism of transcription elongation and its control. *Cell* **120,** 183–93 (2005).

30. Vvedenskaya, I. O. *et al.* Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science* **344,** 1285–9 (2014).

31. Larson, M. H. *et al.* A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344,** 1042–7 (2014).

32. Shine, J. & Dalgarno, L. The 3'-Terminal Sequence of *Escherichia coli* 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proceedings of the National Academy of Sciences* **71,** 1342–6 (1974).

33. Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328,** 504–8 (2010).

34. Epshtein, V. & Nudler, E. Cooperation between RNA polymerase molecules in transcription elongation. *Science* **300,** 801–5 (2003).

35. Epshtein, V., Toulmé, F., Rahmouni, A. R., Borukhov, S. & Nudler, E. Transcription through the roadblocks: the role of RNA polymerase cooperation. *The EMBO journal* **22,** 4719–27 (2003).

36. Green, P. J., Pines, O & Inouye, M. The role of antisense RNA in gene regulation. *Annual Review of Biochemistry* **55,** 569–97 (1986).

37. Peters, J. M., Vangeloff, A. D. & Landick, R. Bacterial transcription terminators: The RNA 3-end chronicles. *Journal of Molecular Biology* **412,** 793–813 (2011).

38. D'Aubenton Carafa, Y., Brody, E. & Thermes, C. Prediction of rho-independent *Escherichia coli* transcription terminators. *Journal of Molecular Biology* **216,** 835–58 (1990).

39. Kingston, R., Nierman, W. & Chamberlin, M. A direct effect of guanosine tetraphosphate on pausing of *Escherichia coli* RNA polymerase during RNA chain elongation. *Journal of Biological Chemistry* **256,** 2787–97 (1981).

40. Farnham, P. J. & Platt, T. A model for transcription termination suggested by studies on the *trp* attenuator in vitro using base analogs. *Cell* **20,** 739–48 (1980).

41. Gusarov, I & Nudler, E. The mechanism of intrinsic transcription termination. *Molecular Cell* **3,** 495–504 (1999).

42. Epshtein, V., Cardinale, C. J., Ruckenstein, A. E., Borukhov, S. & Nudler, E. An allosteric path to transcription termination. *Molecular Cell* **28,** 991–1001 (2007).

43. Epshtein, V., Dutta, D., Wade, J. & Nudler, E. An allosteric mechanism of Rho-dependent transcription termination. *Nature* **463,** 245–9 (2010).

44. Ciampi, M. S. Rho-dependent terminators and transcription termination. *Microbiology* **152,** 2515–28 (2006).

45. Collado-Vides, J. *et al.* Bioinformatics resources for the study of gene regulation in bacteria. *Journal of Bacteriology* **191,** 23–31 (2009).

46. Towsey, M., Timms, P., Hogan, J. & Mathews, S. A. The cross-species prediction of bacterial promoters using a support vector machine. *Computational biology and chemistry* **32,** 359–66 (2008).

47. Warren, R. L. *et al.* Transcription of foreign DNA in *Escherichia coli*. *Genome research* **18,** 1798–805 (2008).

48. Hudson, M. C. & Stewart, G. C. Differential utilization of *Staphylococcus aureus* promoter sequences by *Escherichia coli* and *Bacillus subtilis*. *Gene* **48,** 93–100 (1986).

49. Henkin, T. M. & Sonenshein, A. L. Mutations of the *Escherichia coli lacUV5* promoter resulting in increased expression in *Bacillus subtilis*. *MGG Molecular & General Genetics* **209,** 467–74 (1987).

50. Huang, H.-H., Camsund, D., Lindblad, P. & Heidorn, T. Design and characterization of molecular tools for a Synthetic Biology approach towards developing cyanobacterial biotechnology. *Nucleic Acids Research* **38,** 2577–93 (2010).

51. McWhinnie, R. L. & Nano, F. E. Synthetic Promoters Functional in *Francisella novicida* and *Escherichia coli*. *Applied and Environmental Microbiology* **80,** 226–34 (2014).

52. Hawley, D. K. & McClure, W. R. Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Research* **11,** 2237–55 (1983).

53. Lisser, S & Margalit, H. Compilation of *E. coli* mRNA promoter sequences. *Nucleic Acids Research* **21,** 1507–16 (1993).

54. Helmann, J. D. Compilation and analysis of *Bacillus subtilis* σ^{A}-dependent promoter sequences: evidence for extended contact between RNA polymerase and upstream promoter DNA. *Nucleic Acids Research* **23,** 2351–60 (1995).

55. Voskuil, M. The −16 region of *Bacillus subtilis* and other gram-positive bacterial promoters. *Nucleic Acids Research* **26,** 3584–90 (1998).

56. Nicolas, P. *et al.* Condition-dependent transcriptome reveals high-level regulatory architecture in *Bacillus subtilis*. *Science* **335,** 1103–6 (2012).

57. Zaide, G. *et al.* Identification and characterization of novel and potent transcription promoters of *Francisella tularensis*. *Applied and Environmental Microbiology* **77,** 1608–18 (2011).

58. Neidhardt, F. C. & Curtiss, R. *Escherichia coli and Salmonella: cellular and molecular biology* 3rd ed., 1–2 (ASM Press, Washington, DC: 2004).

59. Ross, W *et al.* A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262,** 1407–13 (1993).

60. Travers, A. A. Promoter sequence for stringent control of bacterial ribonucleic acid synthesis. *Journal of Bacteriology* **141,** 973–6 (1980).

61. Burr, T., Mitchell, J., Kolb, A., Minchin, S. & Busby, S. DNA sequence elements located immediately upstream of the −10 hexamer in *Escherichia coli* promoters: a systematic study. *Nucleic Acids Research* **28,** 1864–70 (2000).

62. Mitchell, J. E., Zheng, D., Busby, S. J. W. & Minchin, S. D. Identification and analysis of 'extended m10' promoters in *Escherichia coli*. *Nucleic Acids Research* **31,** 4689–95 (2003).

63. Pribnow, D. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proceedings of the National Academy of Sciences* **72,** 784–8 (1975).

64. Simpson, R. B. Contacts between *Escherichia coli* RNA polymerase and thymines in the *lac* UV5 promoter. *Proceedings of the National Academy of Sciences* **76,** 3233–7 (1979).

65. Barrios, H, Valderrama, B & Morett, E. Compilation and analysis of σ54-dependent promoter sequences. *Nucleic acids research* **27,** 4305–13 (1999).

66. Schaller, H, Gray, C & Herrmann, K. Nucleotide sequence of an RNA polymerase binding site from the DNA of bacteriophage fd. *Proceedings of the National Academy of Sciences of the United States of America* **72,** 737–41 (1975).

67. Seeburg, P. H., Nusslein, C. & Schaller, H. Interaction of RNA Polymerase with Promoters from Bacteriophage fd. *European Journal of Biochemistry* **74,** 107–13 (1977).

68. Mekler, V. & Severinov, K. Cooperativity and interaction energy threshold effects in recognition of the −10 promoter element by bacterial RNA polymerase. *Nucleic acids research* **41,** 7276–85 (2013).

69. Berman, M. L. & Landy, A. Promoter mutations in the transfer RNA gene *tyrT* of *Escherichia coli*. *Proceedings of the National Academy of Sciences* **76,** 4303–7 (1979).

70. A computer aided oligonucleotide analysis provides a model sequence for RNA polymerase-promoter recognition in *E.coli*. *Nucleic Acids Research* **5,** 3759–74 (1978).

71. Warne, S. E. & DeHaseth, P. L. Promoter recognition by *Escherichia coli* RNA polymerase. Effects of single base pair deletions and insertions in the spacer DNA separating the −10 and −35 regions are dependent on spacer DNA sequence. *Biochemistry* **32,** 6134–40 (1993).

72. Mulligan, M. E., Brosius, J & McClure, W. R. Characterization *in vitro* of the effect of spacer length on the activity of *Escherichia coli* RNA polymerase at the TAC promoter. *The Journal of Biological Chemistry* **260,** 3529–38 (1985).

73. Typas, A. & Hengge, R. Role of the spacer between the −35 and −10 regions in σ^S promoter selectivity in *Escherichia coli*. *Molecular microbiology* **59,** 1037–51 (2006).

74. Siebenlist, U. *E. coli* RNA polymerase interacts homologously with two different promoters. *Cell* **20,** 269–81 (1980).

75. Dombroski, A. J., Johnson, B. D., Lonetto, M. & Gross, C. A. The sigma subunit of *Escherichia coli* RNA polymerase senses promoter spacing. *Proceedings of the National Academy of Sciences* **93,** 8858–62 (1996).

76. Hook-Barnard, I. G. & Hinton, D. M. The promoter spacer influences transcription initiation via σ^{70} region 1.1 of *Escherichia coli* RNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 737–42 (2009).

77. Singh, S. S., Typas, A., Hengge, R. & Grainger, D. C. *Escherichia coli* σ70 senses sequence and conformation of the promoter spacer region. *Nucleic Acids Research* **39,** 5109–18 (2011).

78. Beutel, B. A. & Record, M. *E. coli* promoter spacer regions contain nonrandom sequences which correlate to spacer length. *Nucleic Acids Research* **18,** 3597–603 (1990).

79. Jensen, P. R. & Hammer, K. The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. **64,** 82–7 (1998).

80. Hook-Barnard, I. G. & Hinton, D. M. Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters. *Gene Regulation and Systems Biology* **1,** 275–93 (2007).

81. Miroslavova, N. S. & Busby, S. J. W. Investigations of the modular structure of bacterial promoters. *Biochemical Society Symposium* **73,** 1–10 (2006).

82. Estrem, S. T., Gaal, T., Ross, W. & Gourse, R. L. Identification of an UP element consensus sequence for bacterial promoters. *Proceedings of the National Academy of Sciences* **95,** 9761–6 (1998).

83. Estrem, S. T. *et al.* Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. *Genes & Development* **13,** 2134–47 (1999).

84. Ross, W. & Gourse, R. L. Sequence-independent upstream DNA-αCTD interactions strongly stimulate *Escherichia coli* RNA polymerase-*lacUV5* promoter association. *Proceedings of the National Academy of Sciences* **102,** 291–6 (2005).

85. Essential structure of *E. coli* promoter effect of spacer length between the two consensus sequences on promoter function. *Nucleic Acids Research* **11,** 5855–64 (1983).

86. Haugen, S. P. *et al.* rRNA promoter regulation by nonoptimal binding of σ region 1.2: an additional recognition element for RNA polymerase. *Cell* **125,** 1069–82 (2006).

87. O'Neill, M. *Escherichia coli* promoters. I. Consensus as it relates to spacing class, specificity, repeat substructure, and three-dimensional organization. *Journal of Biological Chemistry* **264,** 5522–30 (1989).

88. Barne, K. A., Bown, J. A., Busby, S. J. & Minchin, S. D. Region 2.5 of the *Escherichia coli* RNA polymerase σ^{70} subunit is responsible for the recognition of the 'extended −10' motif at promoters. *The EMBO Journal* **16,** 4034–40 (1997).

89. Kumar, A *et al.* The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter. *Journal of molecular biology* **232,** 406–18 (1993).

90. Djordjevic, M. Redefining *Escherichia coli* σ^{70} promoter elements: −15 motif as a complement of the −10 motif. *Journal of Bacteriology* **193,** 6305–14 (2011).

91. Cook, V. M. & Dehaseth, P. L. Strand opening-deficient *Escherichia coli* RNA polymerase facilitates investigation of closed complexes with promoter DNA: effects of DNA sequence and temperature. *The Journal of Biological Chemistry* **282,** 21319–26 (2007).

92. Voskuil, M. I. & Chambliss, G. H. The TRTGn Motif Stabilizes the Transcription Initiation Open Complex. *Journal of Molecular Biology* **322,** 521–32 (2002).

93. Harley, C. B. & Reynolds, R. P. Analysis of *E. coli* promoter sequences. *Nucleic Acids Research* **15,** 2343–61 (1987).

94. Lamond, A. I. & Travers, A. A. Genetically separable functional elements mediate the optimal expression and stringent regulation of a bacterial tRNA gene. *Cell* **40,** 319–26 (1985).

95. Jung, Y. H. & Lee, Y. *Escherichia coli rnpB* promoter mutants altered in stringent response. *Biochemical and Biophysical Research Communications* **230,** 582–6 (1997).

96. Rao, L *et al.* Factor independent activation of *rrnB* P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. *Journal of Molecular Biology* **235,** 1421–35 (1994).

97. Ross, W., Aiyar, S. E., Salomon, J. & Gourse, R. L. *Escherichia coli* promoters with UP elements of different strengths: modular structure of bacterial promoters. *Journal of Bacteriology* **180,** 5375–83 (1998).

98. Tagami, H. & Aiba, H. An inactive open complex mediated by an UP element at *Escherichia coli* promoters. *Proceedings of the National Academy of Sciences* **96,** 7202–7 (1999).

99. Strainic, M. G., Sullivan, J. J., Velevis, A & DeHaseth, P. L. Promoter recognition by *Escherichia coli* RNA polymerase: effects of the UP element on open complex formation and promoter clearance. *Biochemistry* **37,** 18074–80 (1998).

100. Davis, C. A., Capp, M. W., Record, M. T. & Saecker, R. M. The effects of upstream DNA on open complex formation by *Escherichia coli* RNA polymerase. *Proceedings of the National Academy of Sciences* **102,** 285–90 (2005).

101. Cellai, S. *et al.* Upstream promoter sequences and αCTD mediate stable DNA wrapping within the RNA polymerase-promoter open complex. *EMBO reports* **8,** 271–8 (2007).

102. Vo, N. V., Hsu, L. M., Kane, C. M. & Chamberlin, M. J. *In vitro* studies of transcript initiation by *Escherichia coli* RNA polymerase. 3. Influences of individual DNA elements within the promoter recognition region on abortive initiation and promoter escape. *Biochemistry* **42,** 3798–811 (2003).

103. Österberg, S., del Peso-Santos, T. & Shingler, V. Regulation of alternative sigma factor use. *Annual Review of Microbiology* **65,** 37–55 (2011).

104. Gaal, T, Bartlett, M. S., Ross, W, Turnbough, C. L. & Gourse, R. L. Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science* **278,** 2092–7 (1997).

105. Schneider, D. A., Gaal, T. & Gourse, R. L. NTP-sensing by rRNA promoters in *Escherichia coli* is direct. *Proceedings of the National Academy of Sciences* **99,** 8602–7 (2002).

106. Reppas, N. B., Wade, J. T., Church, G. M. & Struhl, K. The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate limiting. *Molecular Cell* **24,** 747–57 (2006).

107. Barker, M. M., Gaal, T, Josaitis, C. A. & Gourse, R. L. Mechanism of regulation of transcription initiation by ppGpp. I. Effects of ppGpp on transcription initiation *in vivo* and *in vitro*. *Journal of Molecular Biology* **305,** 673–88 (2001).

108. Barker, M. M., Gaal, T & Gourse, R. L. Mechanism of regulation of transcription initiation by ppGpp. II. Models for positive control based on properties of RNAP mutants and competition for RNAP. *Journal of Molecular Biology* **305,** 689–702 (2001).

109. Unger, B., Becker, J. & Hillen, W. Nucleotide sequence of the gene, protein purification and characterization of the pSC101-encoded tetracycline resistance-gene-repressor. *Gene* **31,** 103–8 (1984).

110. Devine, J. H., Shadel, G. S. & Baldwin, T. O. Identification of the operator of the *lux* regulon from the *Vibrio fischeri* strain ATCC7744. *Proceedings of the National Academy of Sciences* **86,** 5688–92 (1989).

111. Choy, H. E. & Adhya, S. Control of *gal* transcription through DNA looping: inhibition of the initial transcribing complex. *Proceedings of the National Academy of Sciences* **89,** 11264–8 (1992).

112. Shin, M *et al.* Repression of *deoP2* in *Escherichia coli* by CytR: conversion of a transcription activator into a repressor. *The EMBO Journal* **20,** 5392–9 (2001).

113. Oehler, S, Eismann, E. R., Krämer, H & Müller-Hill, B. The three operators of the *lac* operon cooperate in repression. *The EMBO journal* **9,** 973–9 (1990).

114. Ebright, R. H. Transcription activation at Class I CAP-dependent promoters. *Molecular Microbiology* **8,** 797–802 (1993).

115. Jain, D. *et al.* Crystal structure of bacteriophage lambda cII and its DNA complex. *Molecular cell* **19,** 259–69 (2005).

116. Heldwein, E. E. & Brennan, R. G. Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature* **409,** 378–82 (2001).

117. Brown, N. L., Stoyanov, J. V., Kidd, S. P. & Hobman, J. L. The MerR family of transcriptional regulators. *FEMS Microbiology Reviews* **27,** 145–63 (2003).

118. Nakaya, R., Nakamura, A. & Murata, Y. Resistance transfer agents in *Shigella*. *Biochemical and Biophysical Research Communications* **3,** 654–9 (1960).

119. Debnau, E. & Stocker, B. A. Genetics of Plasmids in *Salmonella typhimurium*. *Nature* **204,** 1112–3 (1964).

120. Cohen, S. N. Transposable genetic elements and plasmid evolution. *Nature* **263,** 731–8 (1976).

121. Chalmers, R., Sewitz, S., Lipkow, K. & Crellin, P. Complete Nucleotide Sequence of Tn*10*. *Journal of Bacteriology* **182,** 2970–2 (2000).

122. Beck, C. F., Mutzel, R, Barbé, J & Müller, W. A multifunctional gene (*tetR*) controls Tn*10*-encoded tetracycline resistance. *Journal of Bacteriology* **150,** 633–42 (1982).

123. Griffith, K. L., Becker, S. M. & Wolf, R. E. Characterization of TetD as a transcriptional activator of a subset of genes of the *Escherichia coli* SoxS/MarA/Rob regulon. *Molecular Microbiology* **56,** 1103–17 (2005).

124. Pepe, C. M., Suzuki, C, Laurie, C & Simons, R. W. Regulation of the "*tetCD*" genes of transposon Tn*10*. *Journal of Molecular Biology* **270,** 14–25 (1997).

125. McMurry, L., Petrucci, R. E. & Levy, S. B. Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **77,** 3974–7 (1980).

126. Yamaguchi, A, Udagawa, T & Sawai, T. Transport of divalent cations with tetracycline as mediated by the transposon Tn*10*-encoded tetracycline resistance protein. *Journal of Biological Chemistry* **265,** 4809–13 (1990).

127. Eckert, B & Beck, C. F. Overproduction of transposon Tn*10*-encoded tetracycline resistance protein results in cell death and loss of membrane potential. *Journal of Bacteriology* **171,** 3557–9 (1989).

128. Griffith, J. K., Cuellar, D. H., Fordyce, C. A., Hutchings, K. G. & Mondragon, A. A. Structure and function of the class C tetracycline/H+ antiporter: three independent groups of phenotypes are conferred by TetA(C). *Molecular Membrane Biology* **11,** 271–7 (1994).

129. Stavropoulos, T. Expression of the *tetA*(C) tetracycline efflux pump in *Escherichia coli* confers osmotic sensitivity. *FEMS Microbiology Letters* **190,** 147–50 (2000).

130. Nguyen, T., Phan, Q., Duong, L., Bertrand, K. & Lenski, R. Effects of carriage and expression of the Tn*10* tetracycline-resistance operon on the fitness of *Escherichia coli* K12. *Molecular Biology and Evolution* **6,** 213–25 (1989).

131. Maloy, S. R. & Nunn, W. D. Selection for loss of tetracycline resistance by *Escherichia coli*. *Journal of Bacteriology* **145,** 1110–1 (1981).

132. Sarkar, S & Thach, R. E. Inhibition of formylmethionyl-transfer RNA binding to ribosomes by tetracycline. *Proceedings of the National Academy of Sciences* **60,** 1479–86 (1968).

133. Takahashi, M., Altschmied, L. & Hillen, W. Kinetic and equilibrium characterization of the Tet repressor-tetracycline complex by fluorescence measurements. *Journal of Molecular Biology* **187,** 341–8 (1986).

134. Lederer, T *et al.* Tetracycline analogs affecting binding to Tn*10*-Encoded Tet repressor trigger the same mechanism of induction. *Biochemistry* **35,** 7439–46 (1996).

135. Epe, B & Woolley, P. The binding of 6-demethylchlortetracycline to 70S, 50S and 30S ribosomal particles: a quantitative study by fluorescence anisotropy. *The EMBO journal* **3,** 121–6 (1984).

136. Georgi, C., Buerger, J., Hillen, W. & Berens, C. Promoter strength driving TetR determines the regulatory properties of Tet-controlled expression systems. *PloS One* **7,** e41620 (2012).

137. Bertrand, K. P., Postle, K., Wray, L. V. & Reznikoff, W. S. Overlapping divergent promoters control expression of Tn*10* tetracycline resistance. *Gene* **23,** 149–56 (1983).

138. Bertrand, K. P., Postle, K., Wray, L. V. & Reznikoff, W. S. Overlapping divergent promoters control expression of Tn*10* tetracycline resistance. *Gene* **23,** 149–56 (1983).

139. Berens, C & Hillen, W. Gene regulation by tetracyclines. Constraints of resistance regulation in bacteria shape TetR for application in eukaryotes. *European Journal of Biochemistry* **270,** 3109–21 (2003).

140. Saenger, W, Orth, P, Kisker, C, Hillen, W & Hinrichs, W. The tetracycline repressor—A paradigm for a biological switch. *Angewandte Chemie (International ed. in English)* **39,** 2042–52 (2000).

141. Batchelor, E., Silhavy, T. J. & Goulian, M. Continuous control in bacterial regulatory circuits. *Journal of Bacteriology* **186,** 7618–25 (2004).

142. Hinrichs, W *et al.* Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. *Science* **264,** 418–20 (1994).

143. Kisker, C, Hinrichs, W, Tovar, K, Hillen, W & Saenger, W. The complex formed between Tet repressor and tetracycline-Mg^{2+} reveals mechanism of antibiotic resistance. *Journal of molecular biology* **247,** 260–80 (1995).

144. Orth, P, Schnappinger, D, Hillen, W, Saenger, W & Hinrichs, W. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nature Structural Biology* **7,** 215–9 (2000).

145. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. The many faces of the helix-turn-helix domain: Transcription regulation and beyond. *FEMS Microbiology Reviews* **29,** 231–62 (2005).

146. Structure of the cro repressor from bacteriophage λ and its interaction with DNA. *Nature* **290,** 754–8 (1981).

147. The operator-binding domain of λrepressor: structure and DNA recognition. *Nature* **298,** 443–7 (1982).

148. Casadaban, M. J. Fusion of the *Escherichia coli lac* genes to the *ara* promoter: a general technique using bacteriophage Mu-1 insertions. *Proceedings of the National Academy of Sciences* **72,** 809–13 (1975).

149. Polisky, B., Bishop, R. J. & Gelfand, D. H. A plasmid cloning vehicle allowing regulated expression of eukaryotic DNA in bacteria. *Proceedings of the National Academy of Sciences* **73,** 3900–04 (1976).

150. Sninsky, J. J., Uhlin, B. E., Gustafsson, P. & Cohen, S. N. Construction and characterization of a novel two-plasmid system for accomplishing temperature-regulated, amplified expression of cloned adventitious genes in *Escherichia coli*. *Gene* **16,** 275–86 (1981).

151. Dubendorf, J. W. & Studier, F. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with *lac* repressor. *Journal of Molecular Biology* **219,** 45–59 (1991).

152. Terpe, K. Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology* **72,** 211–22 (2006).

153. King, J. M. *et al.* Rapid, sensitive bioluminescent reporter technology for naphthalene exposure and biodegradation. *Science* **249,** 778–81 (1990).

154. Farhan Ul Haque, M., Nadalig, T., Bringel, F., Schaller, H. & Vuilleumier, S. Fluorescence-based bacterial bioreporter for specific detection of methyl halide emissions in the environment. *Applied and Environmental Microbiology* **79,** 6561–7 (2013).

155. Anderson, J. C., Voigt, C. A. & Arkin, A. P. Environmental signal integration by a modular AND gate. *Molecular Systems Biology* **3** (2007).

156. Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology* **31,** 448–52 (2013).

157. Moe-Behrens, G. H. The biological microprocessor, or how to build a computer with biological parts. *Computational and Structural Biotechnology Journal* **7,** 1–18 (2013).

158. Serrano, L. Synthetic biology: promises and challenges. *Molecular Systems Biology* **3,** 158 (2007).

159. Fan, F *et al.* Regulated ectopic expression and allelic-replacement mutagenesis as a method for gene essentiality testing in *Staphylococcus aureus*. *Plasmid* **46,** 71–5 (2001).

160. Ji, Y *et al.* Identification of critical staphylococcal genes using conditional phenotypes generated by antisense RNA. *Science* **293,** 2266–9 (2001).

161. Wright, K. J., Seed, P. C. & Hultgren, S. J. Development of intracellular bacterial communities of uropathogenic *Escherichia coli* depends on type 1 pili. *Cellular Microbiology* **9,** 2230–41 (2007).

162. Ehrt, S. *et al.* Controlling gene expression in mycobacteria with anhydrotetracycline and Tet repressor. *Nucleic Acids Research* **33,** e21 (2005).

163. Whetstine, C. R., Slusser, J. G. & Zückert, W. R. Development of a single-plasmid-based regulatable gene expression system for *Borrelia burgdorferi*. *Applied and Environmental Microbiology* **75,** 6553–8 (2009).

164. Lovullo, E. D., Miller, C. N., Pavelka, M. S. & Kawula, T. H. TetR-based gene regulation systems for *Francisella tularensis*. *Applied and Environmental Microbiology* **78,** 6883–9 (2012).

165. Geissendörfer, M & Hillen, W. Regulated expression of heterologous genes in *Bacillus subtilis* using the Tn*10* encoded *tet* regulatory elements. *Applied Microbiology and Biotechnology* **33,** 657–63 (1990).

166. Lutz, R & Bujard, H. 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 Research* **25,** 1203–10 (1997).

167. Collado-Vides, J, Magasanik, B & Gralla, J. D. Control site location and transcriptional regulation in *Escherichia coli*. *Microbiological Reviews* **55,** 371–94 (1991).

168. Sleight, S. C., Bartley, B. A., Lieviant, J. A. & Sauro, H. M. Designing and engineering evolutionary robust genetic circuits. *Journal of Biological Engineering* **4,** 1–12 (2010).

169. Canton, B., Labno, A. & Endy, D. Refinement and standardization of synthetic biological parts and devices. *Nature Biotechnology* **26,** 787–93 (2008).

170. Ellis, J., Oyston, P. C. F., Green, M. & Titball, R. W. Tularemia. *Clinical Microbiology Reviews* **15,** 631–46 (2002).

171. De Bruin, O. M. *et al.* The biochemical properties of the *Francisella* pathogenicity island (FPI)-encoded proteins IglA, IglB, IglC, PdpB and DotU suggest roles in type VI secretion. *Microbiology* **157,** 3483–91 (2011).

172. Kieffer, T. L., Cowley, S., Nano, F. E. & Elkins, K. L. *Francisella novicida* LPS has greater immunobiological activity in mice than *F. tularensis* LPS, and contributes to *F. novicida* murine pathogenesis. *Microbes and Infection* **5,** 397–403 (2003).

173. Tempel, R., Lai, X.-H., Crosa, L., Kozlowicz, B. & Heffron, F. Attenuated *Francisella novicida* transposon mutants protect mice against wild-type challenge. *Infection and Immunity* **74,** 5095–105 (2006).

174. Rohmer, L. *et al.* Comparison of *Francisella tularensis* genomes reveals evolutionary events associated with the emergence of human pathogenic strains. *Genome Biology* **8,** 1–16 (2007).

175. Champion, M. D. *et al.* Comparative genomic characterization of *Francisella tularensis* strains belonging to low and high virulence subspecies. *PLoS Pathogens* **5,** e1000459 (2009).

176. Mohapatra, N. P. *et al.* Identification of an orphan response regulator required for the virulence of *Francisella spp.* and transcription of pathogenicity island genes. *Infection and Immunity* **75,** 3305–14 (2007).

177. Charity, J. C. *et al.* Twin RNA polymerase-associated proteins control virulence gene expression in *Francisella tularensis*. *PLoS Pathogens* **3,** e84 (2007).

178. Mukhamedyarov, D., Makarova, K. S., Severinov, K. & Kuznedelov, K. *Francisella* RNA polymerase contains a heterodimer of non-identical α subunits. *BMC Molecular Biology* **12,** 1 (2011).

179. LoVullo, E. D., Sherrill, L. A., Perez, L. L. & Pavelka, M. S. Genetic tools for highly pathogenic *Francisella tularensis* subsp. *tularensis*. *Microbiology* **152,** 3425–35 (2006).

180. Maier, T., Pechous, R. & Casey, M. In vivo Himar1-based transposon mutagenesis of *Francisella tularensis*. *Applied and Environmental Microbiology* **72,** 1878–85 (2006).

181. Qin, A. & Mann, B. J. Identification of transposon insertion mutants of *Francisella tularensis tularensis* strain Schu S4 deficient in intracellular replication in the hepatic cell line HepG2. *BMC Microbiology* **6,** 69 (2006).

182. Gallagher, L. A. *et al.* A comprehensive transposon mutant library of *Francisella novicida*, a bioweapon surrogate. *Proceedings of the National Academy of Sciences* **104,** 1009–14 (2007).

183. Horzempa, J., Tarwacki, D. M., Carlson, P. E., Robinson, C. M. & Nau, G. J. Characterization and application of a glucose-repressible promoter in *Francisella tularensis*. *Applied and Environmental Microbiology* **74,** 2161–70 (2008).

184. Reynoso, C. M. K., Miller, M. A., Bina, J. E., Gallivan, J. P. & Weiss, D. S. Riboswitches for intracellular study of genes involved in *Francisella pathogenesis*. *mBio* **3,** e00253–12 (2012).

185. Ludu, J. S. *et al.* The *Francisella* pathogenicity island protein PdpD is required for full virulence and associates with homologues of the type VI secretion system. *Journal of Bacteriology* **190,** 4584–95 (2008).

186. Green, M. R. & Sambrook, J. *Molecular Cloning: A Laboratory Manual.* Fourth ed., 1890 (Cold Spring Harbor Laboratory Press, 2012).

187. Gallagher, L. A., McKevitt, M., Ramage, E. R. & Manoil, C. Genetic dissection of the *Francisella novicida* restriction barrier. *Journal of Bacteriology* **190,** 7830–7 (2008).

188. Cox, R. S., Surette, M. G. & Elowitz, M. B. Programming gene expression with combinatorial promoters. *Molecular Systems Biology* **3,** 145 (2007).

189. LoVullo, E. D., Molins-Schneekloth, C. R., Schweizer, H. P. & Pavelka, M. S. Single-copy chromosomal integration systems for *Francisella tularensis*. *Microbiology* **155,** 1152–63 (2009).

190. LoVullo, E. D., Sherrill, L. A. & Pavelka, M. S. Improved shuttle vectors for *Francisella tularensis* genetics. *FEMS Microbiology Letters* **291,** 95–102 (2009).

191. *Registry of Standard Biological Parts* <`parts.igem.org`> (visited on 11/20/2014).

192. Shevchuk, N. A. *et al.* Construction of long DNA molecules using long PCR-based fusion of several fragments simultaneously. *Nucleic Acids Research* **32,** e19 (2004).

193. Buchan, B. W., McLendon, M. K. & Jones, B. D. Identification of differentially regulated *Francisella tularensis* genes by use of a newly developed Tn*5*-based transposon delivery system. *Applied and Environmental Microbiology* **74,** 2637–45 (2008).

194. Larsson, P. *et al.* The complete genome sequence of *Francisella tularensis*, the causative agent of tularemia. *Nature genetics* **37,** 153–9 (2005).

195. Bröms, J. E., Meyer, L., Lavander, M., Larsson, P. & Sjöstedt, A. DotU and VgrG, core components of type VI secretion systems, are essential for *Francisella* LVS pathogenicity. *PLoS One* **7,** e34639 (2012).

196. Davis, J. H., Rubin, A. J. & Sauer, R. T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Research* **39,** 1131–41 (2011).

197. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nature Biotechnology* **27,** 946–50 (2009).

198. Chen, Y.-J. *et al.* Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nature Methods* **10,** 659–64 (2013).

199. Cameron, D. E. & Collins, J. J. Tunable protein degradation in bacteria. *Nature Biotechnology* **32,** 1276–81 (2014).

200. Ericsson, M, Golovliov, I, Sandström, G, Tärnvik, A & Sjöstedt, A. Characterization of the nucleotide sequence of the *groE* operon encoding heat shock proteins chaperone-60 and -10 of *Francisella tularensis* and determination of the T-cell response to the proteins in individuals vaccinated with *F. tularensis*. *Infection and Immunity* **65,** 1824–9 (1997).

201. Markley, A. L., Begemann, M. B., Clarke, R. E., Gordon, G. C. & Pfleger, B. F. Synthetic biology toolbox for controlling gene expression in the cyanobacterium *Synechococcus sp.* strain PCC 7002. *ACS synthetic biology* **4,** 595–603 (2015).

202. Dong, H., Tao, W., Zhang, Y. & Li, Y. Development of an anhydrotetracycline-inducible gene expression system for solvent-producing *Clostridium acetobutylicum*: A useful tool for strain engineering. *Metabolic engineering* **14,** 59–67 (2011).

203. Horwich, A. L., Farr, G. W. & Fenton, W. A. GroEL-GroES-mediated protein folding. *Chemical Reviews* **106,** 1917–30 (2006).

204. Ericsson, M, Tärnvik, A, Kuoppa, K, Sandström, G & Sjöstedt, A. Increased synthesis of DnaK, GroEL, and GroES homologs by *Francisella tularensis* LVS in response to heat and hydrogen peroxide. *Infection and Immunity* **62,** 178–83 (1994).

205. Kawashima, H, Horii, T, Ogawa, T & Ogawa, H. Functional domains of *Escherichia coli* recA protein deduced from the mutational sites in the gene. *Molecular & general genetics* **193,** 288–92 (1984).

206. Anthony, L. S. D., Gu, M., Cowley, S. C., Leung, W. W. S. & Nano, F. E. Transformation and allelic replacement in *Francisella* spp. *Journal of General Microbiology* **137,** 2697–703 (1991).

207. Kanhere, A. & Bansal, M. A novel method for prokaryotic promoter prediction based on DNA stability. *BMC Bioinformatics* **6,** 1 (2005).

208. Wosten, M. M. S. M., Boeve, M., Koot, M. G. A., van Nuenen, A. C. & van der Zeijst, B. A. M. Identification of *Campylobacter jejuni* promoter sequences. *Journal of Bacteriology* **180,** 594–9 (1998).

209. Weber, S. d. S., Sant'Anna, F. H. & Schrank, I. S. Unveiling *Mycoplasma hyopneumoniae* promoters: sequence definition and genomic distribution. *DNA Research* **19,** 103–15 (2012).

210. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods* **10,** 354–60 (2013).

211. Gatz, C, Frohberg, C & Wendenburg, R. Stringent repression and homogeneous de-repression by tetracycline of a modified CaMV 35S promoter in intact transgenic tobacco plants. *The Plant Journal* **2,** 397–404 (1992).

212. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences* **89,** 5547–51 (1992).

213. Helbl, V, Tiebel, B & Hillen, W. Stepwise selection of TetR variants recognizing *tet* operator 6C with high affinity and specificity. *Journal of Molecular Biology* **276,** 319–24 (1998).

214. Henssler, E. M., Bertram, R, Wisshak, S & Hillen, W. Tet repressor mutants with altered effector binding and allostery. *The FEBS Journal* **272,** 4487–96 (2005).

215. Kamionka, A, Bogdanska-Urbaniak, J, Scholz, O & Hillen, W. Two mutations in the tetracycline repressor change the inducer anhydrotetracycline to a corepressor. *Nucleic Acids Research* **32,** 842–7 (2004).

216. Volzing, K., Biliouris, K. & Kaznessis, Y. N. proTeOn and proTeOff, new protein devices that inducibly activate bacterial gene expression. *ACS Chemical Biology* **6,** 1107–16 (2011).

217. Klotzsche, M, Berens, C & Hillen, W. A peptide triggers allostery in *tet* repressor by binding to a unique site. *The Journal of Biological Chemistry* **280,** 24591–9 (2005).

218. Hunsicker, A. *et al.* An RNA aptamer that induces transcription. *Chemistry & Biology* **16,** 173–80 (2009).

219. Goeke, D. *et al.* Short peptides act as inducers, anti-inducers and corepressors of Tet repressor. *Journal of Molecular Biology* **416,** 33–45 (2012).

220. Meitert, J., Aram, R., Wiesemann, K., Weigand, J. E. & Suess, B. Monitoring the expression level of coding and non-coding RNAs using a TetR inducing aptamer tag. *Bioorganic & Medicinal Chemistry* **21,** 6233–8 (2013).

221. Levskaya, A. *et al.* Synthetic biology: engineering *Escherichia coli* to see light. *Nature* **438,** 441–2 (2005).

222. Sussman, R & Jacob, F. [On a thermosensitive repression system in the *Escherichia coli* lambda bacteriophage]. *Comptes rendus hebdomadaires des séances de l'Académie des sciences* **254,** 1517–9 (1962).

223. Yabuta, M, Onai-Miura, S & Ohsuye, K. Thermo-inducible expression of a recombinant fusion protein by *Escherichia coli lac* repressor mutants. *Journal of biotechnology* **39,** 67–73 (1995).

224. Nauta, A *et al.* Design of thermolabile bacteriophage repressor mutants by comparative molecular modeling. *Nature biotechnology* **15,** 980–3 (1997).

225. Newman, J. R. & Fuqua, C. Broad-host-range expression vectors that carry the L-arabinose-inducible *Escherichia coli araBAD* promoter and the *araC* regulator. *Gene* **227,** 197–203 (1999).

226. Shetty, R. P., Endy, D. & Knight, T. F. Engineering BioBrick vectors from BioBrick parts. *Journal of biological engineering* **2,** 1–12 (2008).

227. Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *Genome Research* **2,** 28–33 (1992).

228. Rasila, T. S., Pajunen, M. I. & Savilahti, H. Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment. *Analytical biochemistry* **388,** 71–80 (2009).

229. Luckner, S. R., Klotzsche, M., Berens, C., Hillen, W. & Muller, Y. A. How an agonist peptide mimics the antibiotic tetracycline to induce Tet-repressor. *Journal of Molecular Biology* **368,** 780–90 (2007).

230. Schnappinger, D. & Hillen, W. Tetracyclines: antibiotic action, uptake, and resistance mechanisms. *Archives of Microbiology* **165,** 359–69 (1996).

231. Wijma, H. J., Floor, R. J. & Janssen, D. B. Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Current opinion in structural biology* **23,** 588–94 (2013).

232. Harper, E. T. & Rose, G. D. Helix stop signals in proteins and peptides: The capping box. *Biochemistry* **32,** 7605–9 (1993).

233. Kędracka-Krok, S. & Wasylewski, Z. The influence of the tryptophan point mutation on the stability of tetracycline repressor (TetR). *Thermochimica Acta* **458,** 138–147 (2007).

234. Kedracka-Krok, S., Gorecki, A., Bonarek, P. & Wasylewski, Z. Kinetic and thermodynamic studies of tet repressor-tetracycline interaction. *Biochemistry* **44,** 1037–46 (2005).

235. Camsund, D., Lindblad, P. & Jaramillo, A. Genetically engineered light sensors for control of bacterial gene expression. *Biotechnology journal* **6,** 826–36 (2011).

236. Ohlendorf, R., Vidavski, R. R., Eldar, A., Moffat, K. & Möglich, A. From dusk till dawn: one-plasmid systems for light-regulated gene expression. *Journal of molecular biology* **416,** 534–42 (2012).

237. Tabor, J. J., Levskaya, A. & Voigt, C. A. Multichromatic control of gene expression in *Escherichia coli*. *Journal of molecular biology* **405,** 315–24 (2011).

Appendix A

Additional information

Figure A.1: Screen for *F. novicida* promoters by X-gal assay, plates 1–4

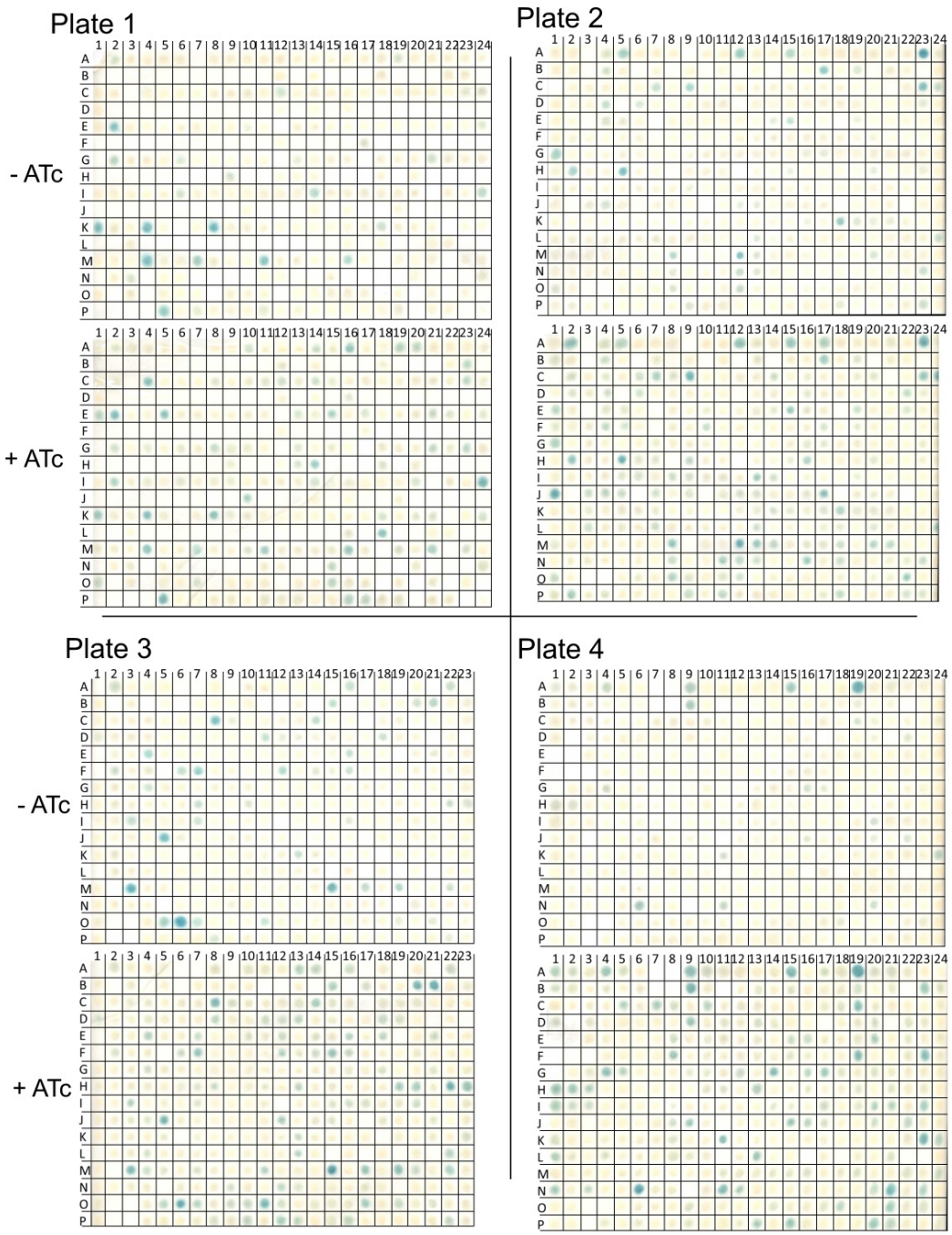


Figure A.1a: Screen for *F. novicida* promoters by X-gal assay, plates 5–8

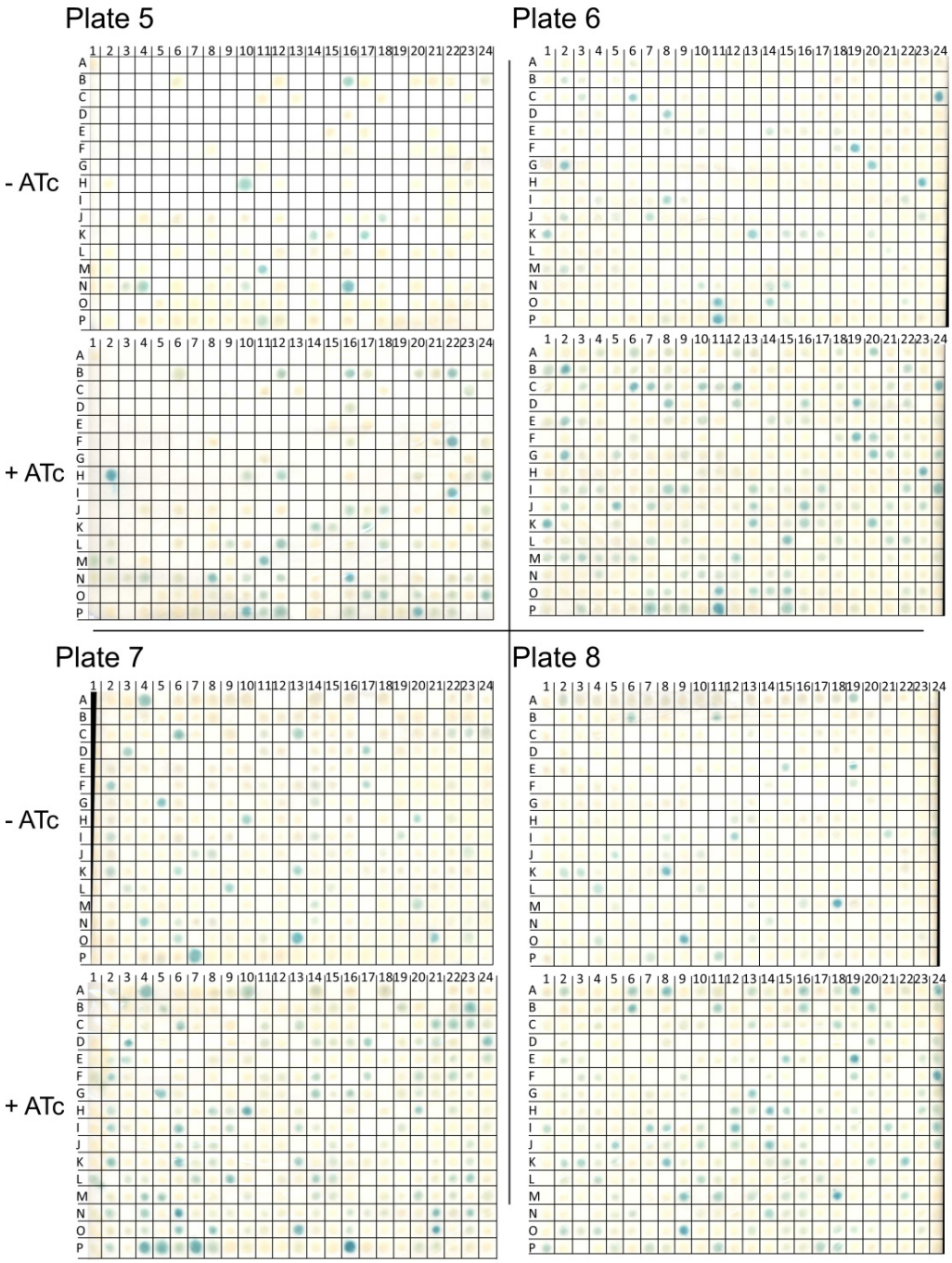


Figure A.1b: Screen for *F. novicida* promoters by X-gal assay, plates 9–12

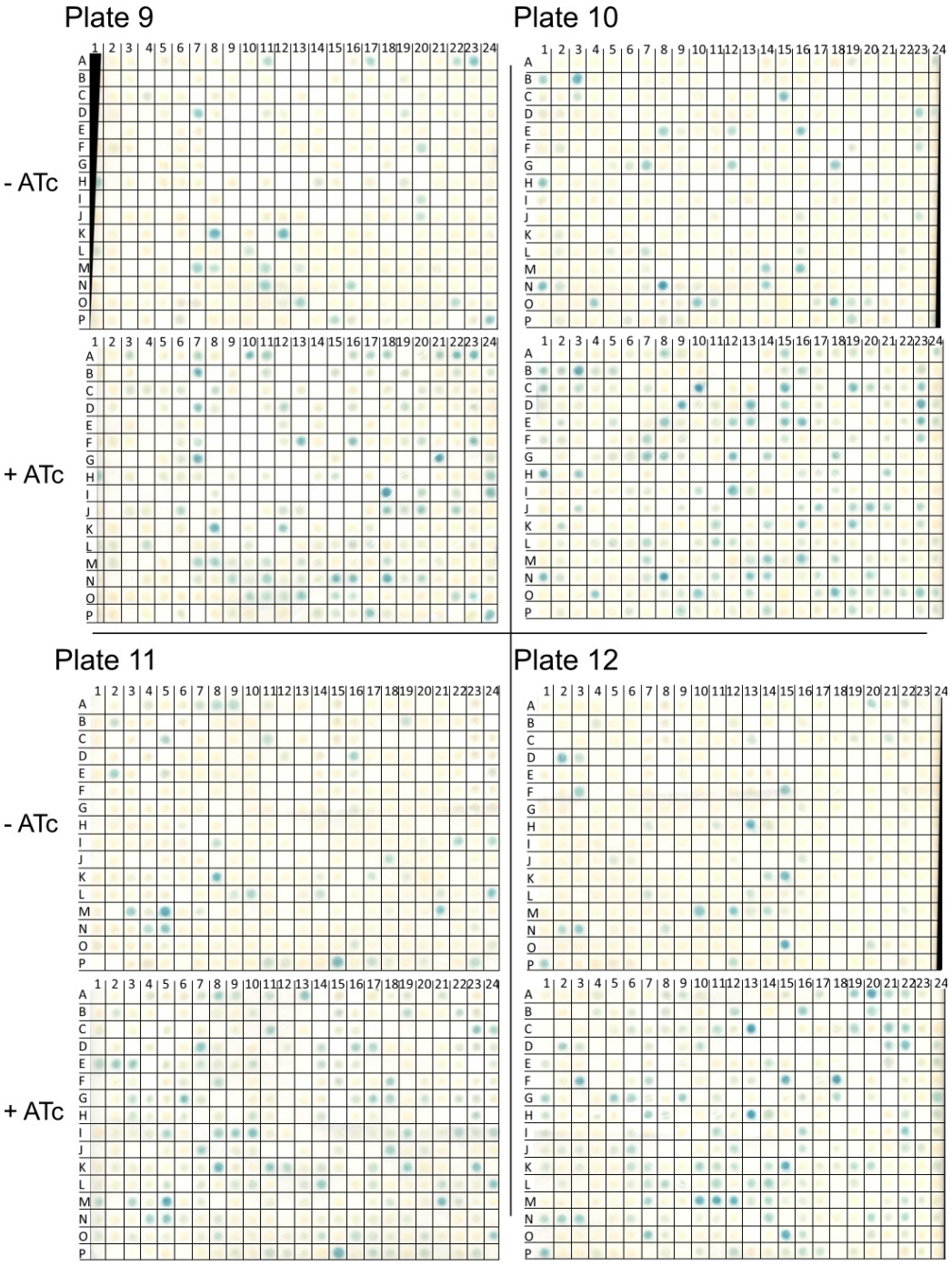


Figure A.1c: Screen for *F. novicida* promoters by X-gal assay, plates 13-16

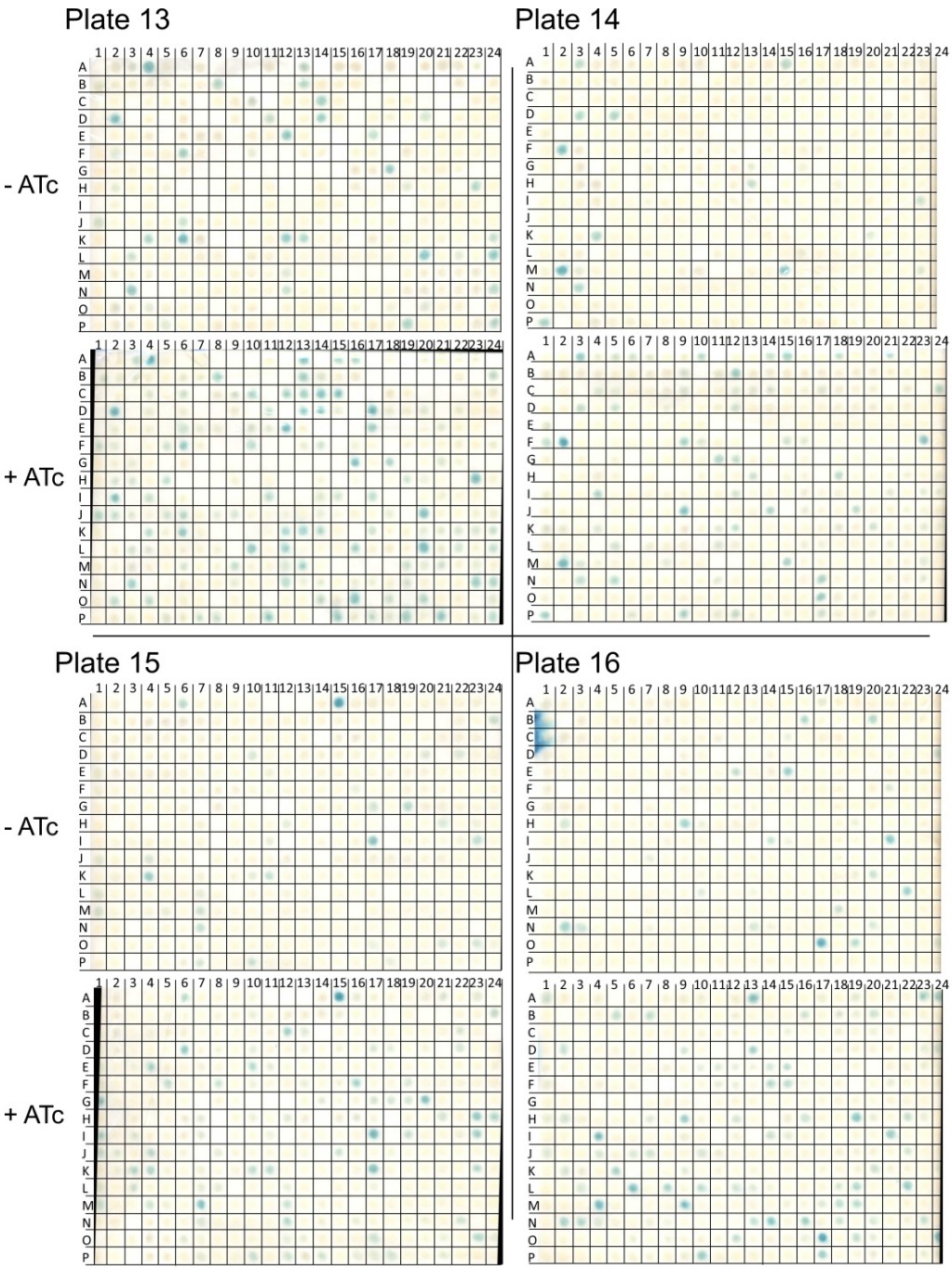


Figure A.1d: Screen for *F. novicida* promoters by X-gal assay, plates 17–20

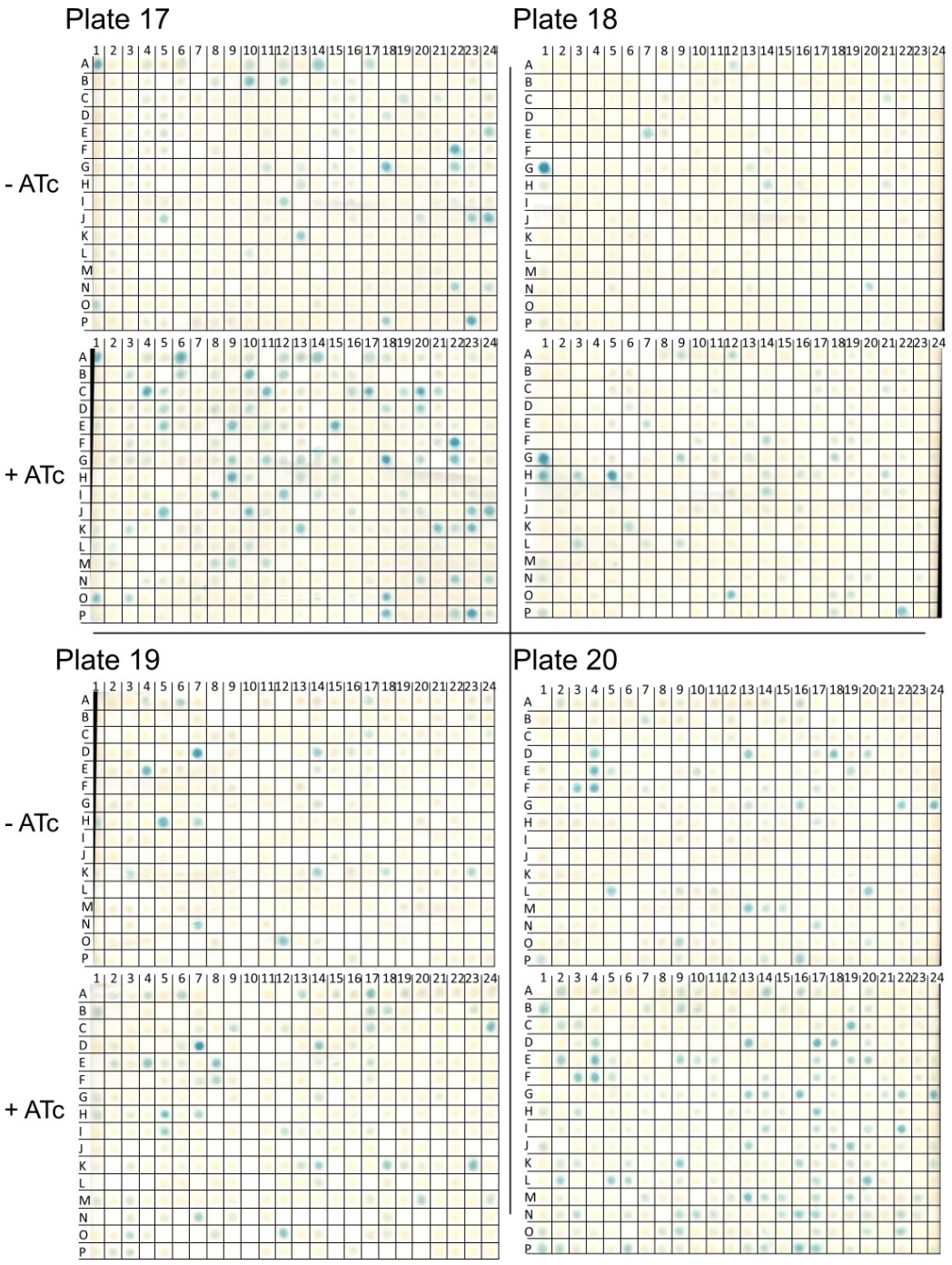


Figure A.1e: Screen for *F. novicida* promoters by X-gal assay, plates 21–24

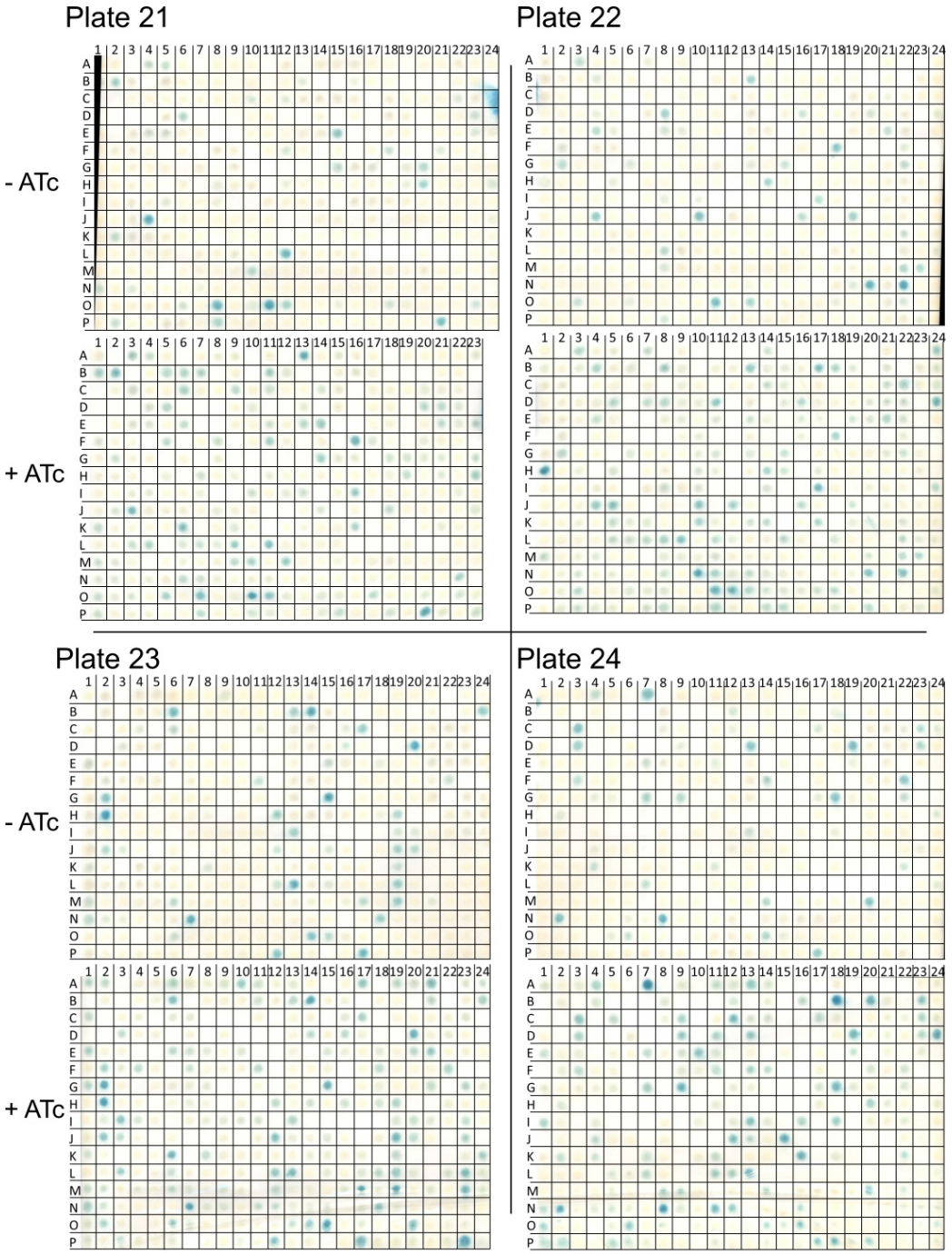


Table A.1: Sequences of synthetic *F. novicida* promoters of clones presented in Figure 2.2

Name	Spot	Sequence (5'–3' with respect to downstream *cat* gene)
P1	2–A11	TTATTGTATTCTTTTTTTATAGTTTAATTATCTCTATCACTGATAGGGAAAAATGTGATAGGATATAAGAAGTGAAAAAACTTACTAAGAAAAACTA
P2	2–B11	TATTGAAAGGTTACTTCTTAGGTAGACTTATCTCTATCACTGATAGGGATTAAATTATTGACAAAACTAGCCTTTAGGGTATTATTATATAACTAAA
P3	2–C11	AAAGTTGATTTTAAATTTATTATAATTTTTAAATTTATTGTGATATAATCCCTATCAGTGATAGAGATCTTTAACGAAGATAAATGAACAAAACTTAGGATCCACATT AGCTTAGATGATAAGGTCTGGGTATTAAAGGTTTTCTCTTTCATCCCTATCAGTGATAGAGAAATAAGAATCAATAACTAAATTAATTGTTA
P4	2–D11	CGCTCTAGAACTAGTGGATCCTATATGATTGTTAGATACACTTTTTTGTTTTCTCTATCACTGATAGGGAAAAAATAGAAAATTACAACAAAATTGAGTATAATCGAT AAGCAAAAAA
P5	2–E11	Duplicate of P4
P6	2–F11	TTATTGTATTCTTTTTTTATAGTTTAATTATCTCTATCACTGATAGGGAAAAATGTGATAGGATATAAGAAGTGAAAAAACTTACTAAGAAAAACTA
P7	2–G11	AGACTGCCTTTTAAGTAAGTTTAGTCTTAAAGCTATTTTGTTTTGTTCTCCCTATCAGTGATAGAGATAATAGATTTAATATAAAAACTACATGTATGGATCCTTATT CTCTGAATTAGATATATGGTACAATACTGAATATTTAATAAAAATGAGTTCTAA
P8	2–H11	TTTAAAAGTTTTGCATAAAAGTGTGATACATCTCTATCACTGATAGGGAAATAAAAATGGTTGACAAAAAACATTATGAATTCGTATACTTAAAAAT
P9	2–A12	TACTTTAATTTTTGTAGCATTTATTAAAATTTAATTTGTGATATTATATCCCTATCAGTGATAGAGAATAAATTATTTAATGATACTTTCTAATAGAGGATCCATTAC CTTCTCAATATGTTTCTTAGTTGTCTCTCTATCACTGATAGGGAGAATGTACTAGTAGAAATACACTGTATAAACAAATTATGTAAAAAAAT
P10	2–B12	CTTTATATTTATGAACATAATTATTATATTTCTCTATCACTGATAGGAAATTATCCAAAATATGATACAATAATAGAGTCACAATTAATAAAAAACG
P11	2–C12	ACTTAATTATAATGAAAACATAATAATTTTACATTTAATAATAATTATTCCCTATCAGTGATAGAGAAAACTAACTAATTAATGTATACTTAAAAAA
P12	2–D12	AATTATGTGAGTTTATATATTATATTATATGTCAGATTTGAACGGTCTTCCCTATCAGTGATAGAGATGCTATAATTATGGCCAACCAAATATAAAAGGATCCCCCAA AGAGGAGAAATAGCCCATGGAGAAAAAAATCACTGGATATACCACCGTTGAAAAATCTGACATATAATATAATATATAAACTCACATAATTGGATCCACTAGTTCTAG AGCG
P13	2–E12	GTTCCCCTTAAAAGAATAATCATTTATTTAAGATTTAAATATGAAACTTCCCTATCAGTGATAGAGAATTTATGATAATATGAACTAACAATAATGAGGATCCTTATT TATTTATTTGTTAAATGTTCTTTAATCTCTATCACTGATAGGAATAAATAGATATAAAATGATAAATAATAACTAACATTTAAACATTTACA
P14	2–F12	Poor sequence data
P15	2–G12	TTTCAATTAATTTTGATTTAAATATTTTTTTCTGATATACTTTCTTAATCCCTATCAGTGATAGAGAACAATTTATAACTAACATATAACAAGATTAGGATCCCAGTG TAAATAGCTGTATTTATGTATAATTTTGTTTTTTTTATTACATTCCCTATCAGTGATAGAGAATAGAGTTAATAAAAACGGCTAAAAGACCA
P16	1–A1	GATGAGAGTTTATCAGTTAAATTATTTGGATATCTTTTTGTTTATAATTCCCTATCAGTGATAGAGAAACACAATAATCCTACAGTATTACTTTACAGGATCCGTACT ATTTATGTTTATTTGAAGAATATTTGATATTATAATATTGTGATCCCTATCAGTGATAGAGAACAAAATTAACACTAATATAAAAGAAAAAT
P17	1–B1	CGTTAAAAAAATATTTTTTTTTTTATTATTTTTAYGTATAATTATAGAGTTCCTATMAGTGATAGAGAAGAAAAAATGAACAAAACATAGTACTTCAA
P18	1–C1	TTCGTGAATTTTTTAGTTTAATAGAATTTTTCTCTATCACTGATAGGGAGATAATGATACAATAATATAAAATTAAATTAATACATACTATCATAAC
P19	1–D1	CTTTTTTGTTATTGAGTTTACGTTAAGATTCAATTATATTATATAAAAGTCCCTATCAGTGATAGAGACTTGAATAGATATTTACAGTATAAAATTTT
P20	1–E1	TTAGGCTTAATATTCTTGATTCTATACTATTATTTTATTATAATGTTTTCCCTATCAGTGATAGAGAAAAATAAACAAAAAGTCTATAAAAAACTGA
P21	1–F1	TGGTATAATTTTAATATTTATCTTTTTATATCTCTATCACTGATAGGGAAACTGATAAAGAATGGCAAAAAGTATGTTATAATTAAAATAGCATTGC
P22	1–G1	ATAATTTTCGTATTTTTTTTGTATAATTTTTTTTTAGTATTGTGTTATGATCCCTATCAGTGATAGAGAATAAATGCATACAAAAATCCTATTAAATAT
P23	1–H1	ATTTTATTGATTTTGTTCTTATGTTCATTTACAATATAATGTTGTATATCCCTATCAGTGATAGAGAATTACAAAATTAAATAATAGTTTTAATCGTGGATCC
P24	1–A2	ATTTTACATTAATTTTATTGATATCGTTTATCTCTATCACTGATAGGGACTAATAACATGAACAAACCTCCTTATATTGAATGTTTATTTTAAATTCGGAAAATAATT TATGTTTTAATTCCCTATCAGTGATAGAGAATTGAATAAAAAAAATTCCATGACTACTAAGGATCCGCATCAAAATGAAATTAATTACGTTAATAGGATTAAACATAA AACATTCCCTATCAGTGATAGAGAGTATATAATATCTATCAAAAATATAAAGAA

Continued on next page...

Table A.1 – continued from previous page

Name	Spot	Sequence (5'–3' with respect to downstream *cat* gene)
P25	1–B2	TTTTAGGCTTTCATTATATAATGTAATTTGTCTCTATCACTGATAGGGAACTATTTACAATTATATTTACTAATGGAAACGAGAAAAATGAAAATACGGGATCCTGGG GACAATAACCTGAATTTTTTTTACTGCATTTTCTTAAAAGTTGTTTCCCTATCAGTGATAGAGATTACTAAAAGAATATTATTGTCAATTCAATGGATCCAATCTTTAT TAAATATAGTGTTAGGTTATACTTATCAATTAACCTTATTCCCTATCAGTGATAGAGATTAACTATAAGTAATGAGATAGAAAAAAAT
P26	1–C2	ATATCTTTTCTATCACTGATTAATTAACTGTCTCTATCACTGATAGGGACTATTTAAGTAATAACGTAGATGATACAATACGATAAACAAAATATAT
P27	1–D2	AATTTGTATTGAATTACCAATTTATTAGTATATTATATTAAATATCTGTCCCTATCAGTGATAGAGAAAAACCAACAAATACGTTTAAGAATATAAT
P28	1–E2	AGAAATATATTATTGTACTGAATAATCGTTATCAAGTGGTATAAATTTTCCCTATCAGTGATAGAGAAGCATAAAAAAAGTACGCAATACAAAAAAAGGATCCTGTTT CCGACTCGCTTCTTGATATTTTTAAAATTTTTTAATAAAATAATCCCTATCAGTGATAGAGAAAAATATATTAGATGAAAATTTATAAGAGT
P29	1–F2	ACTAAATTGTAATTTAATATTGCTTACATATTGATATTCATATATAATTTCCTATCAGTGATAGAGAAACCTAAACAAAAATACTATAAATTAACCT
P30	1–G2	TCAAATCGTTTATATACATCTGGATCTTGTTGTAGAAATAGTGTACAATCCCTATCAGTGATAGAGATGCTAAAACTAACAATAAGTATTTAAAGTT
P31	1–H2	TTTTTTTTATATTTTTTTAATGTTTAAACTAACGTGTAGTAATGAATTTCCCTATCAGTGATAGAGAAAATGGTATATAATAAATATTATAAAAAAA
P32	1–A3	Duplicate of 31
P33	1–B3	AAAACACTTTTACGGTATTTTGTTGAATTGTCTCTATCACTGATAGGGAATTAAATGATTTCTAAACTATGATTAGAAATTTTAGTTTTGATAAGTTGGATCCTTGTA TTTTTCCAAAATATAGTTATTATTATCTCTATCACTGATAGGGATAAGACTTAACAATAGATAAAGATATGAATAAAAATGTTATAAAAAAT
P34	1–C3	TGGCTATATGTTTAGTCACTAATCACTTTATCTCTATCACTGATAGGGATTAGGGTTTTAATTACAAAGAATTATGATTATAATAATATAAATATAG
P35	1–D3	GTTTAACTTTTATGGTAATACTTTTTTTTTTCTCTATCACTGATAGGGAATACATAAATCAATGATATTATATATATATAAAAAAAATTAAATACCA
P36	1–E3	TCTTCTTATTATTTTATAATGTACCGATTTCTCTATCACTGATAGGGAAAGCTTTGAAAAAGAATTATAATTTCGGTATAATCTAAAAATTATTAA
P37	1–F3	TCATTTTTTTTATGGTAAGTTATTAAGTTGTCTCTATCACTGATAGGGATTGAATGTTTACATAAAATTTAAAAAGAATTGAAAAATCTTATAATAAGGATCCTATGT TGGATTTTCTTTATTCTTATGATATTCTCTATCACTGATAGGGAACAACAAATTTATTTTACCAAGCACTCGCTAAGACATACCAATTGTAA
P38	1–G3	ATCAGCGAAAATTGTAAATTATTTAATGTTTCTCTATCACTGATAGGGATTGCTTCTTAATGATAAACAAAGTAAAATAGTATGAGTATTATTTTAT
P39	1–H3	TTTTTGTATGATTATTATGAAATGTTCGTTTCTCTATCACTGATAGGGAAATTCTTCATAAATTTAATATTTATGCTATAATAAATGTGTTTTTAAG
P40	1–A4	TTGATGTTTTATTATAATAACTATGTTAATTTTATATTTTCATAAAAATCCCTATCAGTGATAGAGAATTTTTGATATAATACCTTATTATCGCATA
P41	1–B4	CTTTTAATTTGAATATGTTATAGTGAATATTCTCTATCACTGATAGGGACGTAATAAAGTATGGACAATACCCATAAACATATATTACTTTAATTAAGGATCCTATTT ATTTTATTAGTTAATTGACAAACAGTGATTATTTATCTTAATATCCCTATCAGTGATAGAGAAAGAAGATATTATCTGTTAAAATATAAAAA
P42	1–C4	GCGTTTAATTTTTCGATATTTTTTATTTATTCTCTATCACTGATAGGGACTATATGTTATAATTCAAGGCAATGAACGCAAATACCAGATATGACAA
P43	1–D4	ATGATATACATACTGCTATTTGTCGGCAACTATTATAAGAAACGTTTGTCCCTATCAGTGATAGAGATAAAACAAGTACATAGAAGGTTGAATTAATGGATCCATGAC TGTAATATATTGTATACTTTATATTTCTCTATCACTGATAGGGAAAAAACAGCCAATTTAAAATAATTACTACAACCTATAGTGTAAAGAAA
P44	1–E4	ATGATATACATACTGCTATTTGTCGGCAACTATTATAAGAAACGTTTGTCCCTATCAGTGATAGAGATAAAACAAGTACATAGAAGGTTGAATTAATGGATCCATGAC TGTAATATATTGTATACTTTATATTTCTCTATCACTGATAGGGAAAAAACAGCCAATTTAAAATAATTACTACAACCTATAGTGTAAAGAAA
P45	1–F4	AGGTATTAGATTTTGATTCGTAAAATTTGTAACATTTACTTATAAAAATTCCTATCAGTGATAGAGAACTTTGTATTATAATATATTAAAAAAAAAG
P46	1–G4	TTTCTTTTATTTGGGATTGTTTTTTCAGATTTTTTTTAGTAAAATGATTTCCTATCAGTGATAGAGAAAAAAATAAGATATTAAAAAGAAAAAATT
P47	1–H4	TTTTGGTTTTTTTATGATGGATTTTATTTATTTATGATGGTATTATATTCCCTATCAGTGATAGAGAAGCCAATAAAAAAATATTTAATTTAAAATA
P48	1–A5	ATTTTTTCGATTGTTATCCTTTAAAAACTCCGTCAATTTTGATATAGTTCCCTATCAGTGATAGAGAATAGATTCAATCCCAACAAAAAACCCAATT
P49	1–B5	CGTTAGTTAAGATATTAGGGTTTTTTTTTGCAATATAAATTTTGTATAATCCCTATCAGTGATAGAGATTTCTATTGGAAATCTTCAAATATAATTAT
P50	1–C5	TATTTAATTTGTGTTTTGTTATAAAATTTACTATAATATTAATTTAATTCCCTATCAGTGATAGAGAATATATAACATAAGAATTACAGATATACCT
P51	1–D5	TATCTTTTTTTTTTTGTATAAAGTAGTAGGTCTCTATCACTGATAGGGATCTAATGGAATTTCCAAAAGGATATTTCATCTTAACGTGTAATTATATGGATCCAATATG TTGTATTTTGCGTATCTATTAGCTATAATTCAGATTAATCTGTCCCTATCAGTGATAGAGAATTCAAAAATAGAAAGTTATAACAATATTC
P52	1–E5	TCTATAGATTTTTTTGTTTGACTCTAGTTTTCTCTATCACTGATAGGGAACATTATTGAAATTTAAGGAACAATATCATATTATAATAAGTAAAACT
P53	1–F5	TTTTGGTTTTTTTATGATGGATTTTATTTATTTATGATGGTATTATATTCCCTATCAGTGATAGAGAAGCCAATAAAAAAATATTTAATTTAAAATA

Continued on next page...

Table A.1 – continued from previous page

Name	Spot	Sequence (5'–3' with respect to downstream *cat* gene)
P54	1–G5	GGATCCTATATTTGGATCATTTGTTTTTATGTTTAGTCTCTATCACTGATAGGGAAAAATAAAAATTATAAAATTAAAGAATTTTGAGTAAAAAATCTATAATGGATC CAAACTGATCATTAACCAGACTAATAGTTTTTATTATTTTGAGTTTTTGTCCCTATCAGTGATAGAGACTCTTAAAATAAAAAATAATATACTATAAAGGATCCCATA TAAAAATTGACATTATCCAATTATATTCATATAATTAAAGAATTCCCTATCAGTGATAGAGAAAACGAAAAAACTATCCAAAACTAATACTA
P55	1–H5	GGATCCTTTTTATTTATGTTTAAAACTTGTTTATAATTTAAACTTATACTATAATTCCTATCAGTGATAGAGAACTAAAAAGAATATTAAATAGATCTTTAAA
P56	1–A6	CAATTTTATTGTATATTACATTTTTTACTGATATAATATTTTATGAATTCCCTATCAGTGATAGAGAAAATATAAATGGAAAAGTAAATTCCTGCAT
P57	1–B6	AGTTGTTTATTGCTTAATCTTATCAAGATATCTTTCAATGTATATTGTTCCCTATCAGTGATAGAGATCTAATAAAAGTAATCCTTAAAAAGAAATAGGATCCAAAAA TCTTAGAACTTAATTGATATTTAATAGAATTAATGTTATTATGTCCCTATCAGTGATAGAGAAGTAATTAAGATTAATCGAAAAATATAATT
P58	1–C6	GTTATATACATAGTTTCTTGCTTAATAATAATATTTGATATAAAATTTCCCTATCAGTGATAGAGAAAATTTATTAATAACTATAATAATAATACA
P59	1–D6	AATTTGCAAATATTATTCTTGATTATATAATGATTTTTAGGTATCTTATCCCTATCAGTGATAGAGAAAAATAAAAAAAACAAATATTAAATTTTGT
P60	1–E6	TAATTCAATTTTATTTTTTCATAAAAAATTTAGTCATTAGGTATAAATTCCCTATCAGTGATAGAGATTATACAATAACAATAAAAAAATTAAATTAGGATCCGCTTC TATTTGTATTCAATATTTCTCTGTAGTATAATAGGTACTATTATCCCTATCAGTGATAGAGACTAAATAACATTAACCACAGTACTAAAAAA
P61	1–F6	TAGGAAAATATCTTTTGTTGAAATTTTTTGAACATATTGTATACTTTGTTCCTATCAGTGATAGAGAGAATTCATGAACAAAATAAAAGTTCAGTAT
P62	1–G6	TTCTGAAAGCATGAATAGGATTATTAAATTTCTCTATCACTGATAGGGATTAGAATTTAAAAATGACACCCGATCACATCTTCACACGAAAAATAATGGATCCGTCTA TTTTTTTCTATATAATATTTATTAATCTCTATCACTGATAGGGACCCACGAATACAAAACTAAGATAATTAATAAACAAATAATATTTATTT
P63	1–H6	ACTTTGTTATAAACCTGTTGATTAATTTTTTTTTATTCAATATACTTAATCCCTATCAGTGATAGAGAAAAAGATTAAATAACATACATAAAAAGTAAGGATCCTGCGA TTCGTCTCGGTCATTATATTTGAATTCTCTATCACTGATAGGGACCTAAACATTGAACAAGATTTAAAAATTTTAAAAATTATAGTATGAAA
P64	1–A7	AGAAATATTAATAATTATTTTTGGAAAGATTCTCTATCACTGATAGGGAGATCAACACAACACCTTTGCACTAACATTATAAAATAATAACTTATACGGATCCATTTC TTATTTTGCTGTATCGATACTAGAATCTCTATCACTGATAGGGATATTCAAGCAGGAAAGTTAAAAACAAGAAAAAAATAAAAAACACTAGAGGATCCTATAAATTAT GTTATACTCAAATTGTTAATTCTCTATCACTGATAGGGAGAAAATTAAAACATTAAAACCTACACCCGTAATAAATCAGCATTTAAT
P65	1–B7	AAATATTAGTATTTTTTTGGATTTAAATTCTCTATCACTGATAGGGAAAATAATATTAAATAATTTAAAAACTTGCTACAATAGATGATAATAAT
P66	1–C7	TAAGGAATTAATTTTGTAATATACTGTATATTATTTGCTTTTGTGGTTTCCCTATCAGTGATAGAGAAAATATCATACTATATAAACATTAATATTGGATCCCTTTGT TTTTTGTGTTAGACTATGAATGGTTCTCTATCACTGATAGGGATAATTAAACAAAATACAAATTACAAAGAAGCAAAACATAATAGATTAA
P67	1–D7	TAAAATTGTATTTTTCATTATTTTATAGCTATTTTTTTGTTATAATATTCCCTATCAGTGATAGAGAAAATACGCCACAATAATATAAACATAAAAT
P68	1–E7	TTAATAGTAATTCTTTTTAAATAGCGTTATTCTCTATCACTGATAGGGATACAATAAATACATTCCAATCCAGTCATTTGAACCACATCTACCAAAAGGATCCTTATT TATTAAAAATTTCAGTTTTTATGTATGGATTCTATAGTATAATCCCTATCAGTGATAGAGACAAAATAACATAAAAATAAAACTTAAACTAGGATCCATGTTTCTTCC AGTAAAATTTGTTTTATTGTTATTCTCAATTGCATAGTTCCTATCAGTGATAGAGATCCAATATTTTAAACGAATTTTAATGACAT
P69	1–F7	CAAATACGCTTTTAACAAAAATATGGAAGTAAATGATTAGGTATCGGATCCTTTTTTTCTTGACTTTTTTTTGATTCCTTTCTAACATGACTTAATGATTCCCTATCA GTGATAGAGAAAAAGAAGTATAATACAAAAAAAGTTAAAT
P70	1–G7	TKATGCATTTATCRTTTTAGTCTTTTTGTTTCTCTATCACTGATAGGGATTGAAAATTTAAGAAATATCATATATATTGTCGATATAATAGAAATAG
P71	1–H7	TTTCACTATTTTATGTTTCTTGTTACTGTGTCTCTATCACTGATAGGGATACTCTTGACGCATGAATCTAAATTTAATATAATTATAATGATGAAAGG
P72	1–A8	ATATCGACAGTTATTTAAATTTTAAAATACTCTCTATCACTGATAGGGACTAGTAAATAATAACAAGATCTTATATTATATATATCTAGACTACTAA
P73	1–B8	CAAGAACACTTTGAATTTTTCATTGTTAAGTCTCTATCACTGATAGGGAAAAATACCCAATTAAATTATATACACATAATAAAATGGTGTATAATATGGATCTATAAT ATGATTTATCCCTATCAGTGATAGAGAAACTGTTATCGACGAACTAAAAATATGTTA
P74	1–C8	TTAAGAGTTTGGCTTTCTATATTTGTTAATCTCTATCACTGATAGGGATAACATTTTATAAAACTCGCAAATGTAGAATAAAACAAAGTAGATATTGGATCCGTGTAT TATATTGCTTTCGTTTTAATCTTTTGTATAATTAATTTAGTTTCCCTATCAGTGATAGAGAAAATAACCTAACATAAATCATAAAAAATAA
P75	1–D8	GTTGTTTCTTTACTAGTTTCCTCAACAATACAGATGTTATATTTTTTTTCCCTATCAGTGATAGAGAAATAACAATAACTAAAAATTTAATATCATAGGATCCTCTTT GTTTTATTTTGTGTTATTATTATATTCTCTATCACTGATAGGGAGTATTTTTTGCTTTAAATAAGTATGAATAATTTATTAAACCGATAAAA
P76	1–E8	GTTATTGGCAATTACATCAATATTTGATATATTTATACTTTATAAAATTCCCTATCAGTGATAGAGAATTTAATTAAACAAGGAAAGTATAATGTAT

Continued on next page...

Table A.1 – continued from previous page

Name	Spot	Sequence (5'–3' with respect to downstream *cat* gene)
P77	1–F8	AGTTACTTCAATTTTTTACAATTAAATATTTCTCTATCACTGATAGGGATAGAAAAAAAAAAAATACTTTAGCTATAATAATTTTATCTCATAATAA
P78	1–G8	ATTTAAATCTGTTTGTATGTTACAATAAATGTTTTTCGAACTCTATTTTCCCTATCAGTGATAGAGATTTAATGTAAAATTTGATATAATGGACAAA
P79	1–H8	TTATTATTAGACGTAATTTTCTAATTCGGTTAATTTTTTCTTGCATTTTCCCTATCAGTGATAGAGAATATTGTTATACTTATATATACTAAACAAG
P80	1–A9	TTGACTTTGTCATTTTTATATTACATAATTTCTCTATCACTGATAGGGATACTCAACACTTAAACCTTTAAAAGTCGAGCGAGATTCAAATTATGGA
P81	1–B9	CTGTTCAGTTTTTTTTTATATTTTATAAAATCTCTATCACTGATAGGGACAAAAAGTAAATAATTACTTTTAATAAATTATAAATACTTTTATCATAGGATCCAATTA TGTTATTATCAAATAATATTTACTTTCTCTATCACTGATAGGGAAGGACTATCTGATTGAAAATAAATTACTATTACCCCAAAGAGGAGAAATAGCCC
P82	1–C9	ACTTTTAGGTATAAATTTTATTGCATTAAAATGATTCAGATATTAATATCCCTATCAGTGATAGAGACTCGAATCAATAAGAAAAATAAGATAGTAA
P83	1–D9	TTTAACATATTGATTTTGAATGATTTTTTTTTATATAGTAATAATTTAATCCCTATCAGTGATAGAGATTAATGAAAAGATTTGAAAATAGAAAAGTT
P84	1–E9	TTTTTTATCTTGTTGCCATTGGTTTTTTGTGTTAAAATTTTAAAATTATCCCTATCAGTGATAGAGAATAAAAACACTTAAACACGTATAACAAAGC
P85	1–F9	TTTTGTTTATAATATCGCTTGAGTTATCTGATAAATTAATTTTATTTTTCCCTATCAGTGATAGAGAAATAGTATATAATATTAAATATAGATACGG
P86	1–G9	TGTCTTATTTTTATGTTATATATAGTATTTTCTCTATCACTGATAGGAATTAATGGATAAATTTTTGATAAAATATGATATAATAAACAATGACTAA
P87	1–H9	TAATCTTTATTTTGTGTTATTTATTTTTTAAATTATATGTTATAATCGTCCCTATCAGTGATAGAGATACATATTTTACAATAGTAAAATCTAACGA
P88	1–A10	TTTTGATTTACAAATAAGATTTTCTGATTTATACTTATTAGTGTTTATTCCCTATCAGTGATAGAGATATTGAACATAACAATAATAAGACAATAAA
P89	1–B10	TAGTTTTAATCTATATTATTAGTATTTTTATAGACAAATTATTTATAGTCCCTATCAGTGATAGAGATTCTATGATATAATCCAGTAGAAATATAA
P90	1–C10	ATCATGCGGAAGCAAATAATTGAATAGACTATATTTTTTTTTTAAGTTTTTCCTATCAGTGATAGAGAAAATAAGATAAAATATTAAATAAACTTGAA
P91	1–D10	TCTTCATTATAATATACGCATTTATTATTTTCTCTATCACTGATAGGGAATATAATAATATTCTATGATAAAATAATAATAATTATAAAAGAAGCCG
P92	1–E10	TTTGTATCTAGACCTAAAATATTTGAATTGTCTCTATCACTGATAGGGACACAGTGAAAAAAAATCAAAAAAAAAAAAGTAGAAGTGAGTCAAATAGGATCCATTTTAG TCCATATTTTTTACATATCCTCTTGAGTATTTTGCTATATGTCCCTATCAGTGATAGAGAATTTCCTATATACTGGTAAGAATACGAAAA
P93	1–F10	GTATTTTAATTTTTTTGGTTACGTTTTTAATTTTTATGTTATAATAGTTCCCTATCAGTGATAGAGAATTAAAATATTCTATCGATCTATACTAAAA
P94	1–G10	AGGTGTACCAATTTTGTGTATTATATTTATTGTCTAATATTTTTAATTTCCCTATCAGTGATAGAGAAAACATGATAAAATAAATAAAAATAAAAAA
P95	1–H10	TCTCCAAGATTTTTATTTTTTTAAGAATACTCTCTATCACTGATAGGGAATTGATGTTATAATATATTGCTAATAATATATCATGTATTTACTAATA
P96	1–A11	ATGGATTGTCTTATTCGTGTGATAATTAATTACAATTTTTAAAGATTTTCCCTATCAGTGATAGAGAAAGTGTGCTAATATAACTTACGAAAAAAAG
P97	1–B11	TTCATTTTATTTACATTGTGGTTTCCTTGTTCTCTATCACTGATAGGGATTTTAGAAAATGAATGGTTACCGACGTACTCAGATAAATATAAAGACTGGATCCATAAT TAATTAGTGCAAATTAACTTTGATGCTTTTATTCTCTTTACTATCCCTATCAGTGATAGAGATTCCGTTCGATAAGAATAAAAAATAGTATCGGATCCTGTTGTATAA GTCTAGAATATTTTATATAATCTCTATCACTGATAGGGAAAAAAAAGTCACTTGTAAATCAGACTAGGCTGTCATTAACATACAAAA
P98	1–C11	TATACATTATTTACATTTTTAATTTTTTTCTATATACTTTATTTTTTTTCCCTATCAGTGATAGAGAAAAAAAAAACAATCTAAAATAAGTAATTTT
P99	1–D11	CTTCACATTTTTATATACTTTTCAATGATCTCTCTATCACTGATAGGGATTGTAATAACAGATAAAATTTAATAAAATAACTTATAGACTAATAGAA
P100	1–E11	GATCTGTGCCTATATACATAGGTTTAATTCATGTATTATCTTATAGATTCCCTATCAGTGATAGAGAGAACACTTATAAAATTAATATCCGGTATAAGGATCCTTTAT TTTTCAATTACTTTATGTTATTATTTCTCTATCACTGATAGGGAATTGAAAAAATTATTTTCCCCCAAAGAGGAGAAATAGCCC
P101	1–F11	AATTTTTATGTTAATTTTGTAAAATTTATTATATAATTTATTATACTTTCCCTATCAGTGATAGAGAAACATAAAACGAAAATAGAATTATCCAAGA
P102	1–G11	ATTTTTTAAGTTATGTTATAATGTATTGATTCTCTATCACTGATAGGGAATAAATAAAAAATACAAACTATAATTTACAATTAAAACAATTCCTAAT
P103	1–H11	CCCGTTTCTCCATCATGATTTGCTTAGTTTGAATAAATTATATATAATTCCCTATCAGTGATAGAGAATGTTATTATTTCTAAAACATATTTAGTAC
P104	1–A12	TTTAATAGTTAGATTTTTGATTTTTTAATGTGCCGAAAACATTGCTTTCCTATCAGTGATAGAGATAAAATATGTAAAATGTTAAATTTTAACAC
P105	1–B12	Duplicate of 85
P106	1–C12	TTGACTTGTTTTATATATTGCTTTATAATAACTTTTTTTGTTTTCTATTCCCTATCAGTGATAGAGAATTATGATAACTGATACCGTAACACAAAAAGGATCC
P107	1–D12	ATTAATTTATGTTGTAATTATTTAACTATTATATATAATCTAGTAATATCCCTATCAGTGATAGAGAATATATGTCCAAATAATATAAATAAGGAAA
P108	2–A1	Duplicate of 106

Continued on next page...

Table A.1 – continued from previous page

Name	Spot	Sequence (5'–3' with respect to downstream *cat* gene)
P109	2–B1	TTGTATTAATATAATTTTTAATTTAACGTATCTCTATCACTGATAGGGAATTATTATTTGTAAATTTTCTGACATTTTATAAAAAGCAAACCTTAGTGGATCCTATTT GAAATGCTTGTATTTTTTTTCAATATTGATGTACCATAACTTGTTCCCTATCAGTGATAGAGACACCTCCTAATTAAATACGAATAAATTTTT
P110	2–C1	ATTATAGTCTGATTAAATTGTAATTTAACCTTTTTTGCATTAGTGTATTCCCTATCAGTGATAGAGATGCTATATAAAATTTAAAAAAATCCTAATTGGATCCTTATA ATATTGTAATTATTTTTGTTAATTATAAAGTAATTTGTTAAAATCCCTATCAGTGATAGAGATGACTATGTAAGACAAAGTCAATTTGATTT
P111	2–D1	TTTAGTCTTTATCATGGCGTTAATTAATGTTTTAACGTTATAAAGTTATCCCTATCAGTGATAGAGATAAATATGTTATAATGTCTCTTTAAAAAGT
P112	2–E1	ATATTTTGCGTTGAAATCCCTGTATACTTTAGCATATTATATCTTAAATCCCTATCAGTGATAGAGATGAAGTAGAGTAGTGTATGTACAAATTTAA
P113	2–F1	GATAATGACCTTAAGATATGTTTAATTTTATCTCTATCACTGATAGGGAGAAAAAAACCAAATAAAAAATATAAAAAGCAAAAATGCTAATTACATAGGATCCCAATC GATTTTTATTTATTGTTCTTTTTATTATGATATAATGTTTTTATCCCTATCAGTGATAGAGAACGCAAATATAAACAAATATGATATTTAAT
P114	2–G1	CCATGATTTGGTTTTTTTTACATGGATAGAAATAAACTTATTATTAATTCCCTATCAGTGATAGAGAAAGTTATGTTAAAATAAGAACACACGCTAA
P115	2–H1	TTCCAAATAATTGTACTTTGTGATAGTATTTCTCTATCACTGATAGGGATGTGGGACTTAATTAACACTACACATTGTAGTAAAAATTAAACCCCAAAGAGGA
P116	2–A2	AGATTAAGGCTTTTTTTTATTGATTATTGTAACTTTTAAAGTTATTATTCCCTATCAGTGATAGAGAAAAAAATAAGTGAAAAAACTTTTATATAAC
P117	2–B2	TTGTATTAATGTTTAAATTATAATAATTTTGGCATTTTATATTAGATTTCCCTATCAGTGATAGAGAAAAACAATTATAATGTAGTAAACAATACCA
P118	2–C2	AAAACGATTTGCTTATATAATTTTTATTTATTTGTGTTTAATAGTTTATCCCTATCAGTGATAGAGAAAGACAAATGTATTAAACTAATACTTTATAGGATCCTTTCT AACATTTATGTGCTATAATGTATTGTTTCTATCACTGATAGGGATAACTGATAACCTGAATATCAAAGATTCGAATATAAAGACCTAGTATC
P119	2–D2	Duplicate of 106
P120	2–E2	TATATTTTCTATTTATTATTAACGTAAGTGCAATTTTTTTAAATATTCCCTATCAGTGATAGAGAAAAATTCATAAATAATAAGAAATAAAAACAAGTTTAACTATTG ACACTTATTTTATTTTTTTTTTATAATGGTTATATCCCTATCAGTGATAGAGATCAAATACAAAAAAAAATATGTACAATTTA
P121	2–F2	TTTTTTTATTTATCTTCTTCAATTTTTAGTTCTCTATCACTGATAGGGATATAAAAAATTATAAAAATTGAATATAATAGATATGACCAATATAAAA
P122	2–G2	Duplicate of 116
P123	2–H2	AAATTTCTTAGATGTTTTTATTGCACTTACATATTTTGTTCATTATAATCCCTATCAGTGATAGAGAGAAAACTTTTTAGTGCAAACTTGAAAATAAGGATCCATGTA TTTATAAATGTTTAATTTAATTGTTTCAGTATTATATGCCGATTCCCTATCAGTGATAGAGAAATGTTATATTGTATTAAACAAAGAAAGAA
P124	2–A3	TCTATGGACTCGGTATGTTTTAAATTTTCGTTCTCTATCACTGATAGGGAAAGATGAAGTAGATTATCTCTTACAATCAATATTTACATATATAAAAAAGGATCCTAT TTTTATATAATTTAGCTATACTTTTTAACTGTTTGCTGGTATAATTCCCTATCAGTGATAGAGAATTGCTAACGATAATAAAAAAACATATATA
P125	2–B3	CTGGCACTAAATATTTTTTTTATTATTTCTTTTACGGATTTAGTATATTCCCTATCAGTGATAGAGAAAATAATGATATAATAAAACGATAACTACA
P126	2–C3	AAAGATTCTTGAATTATGTTGATATTACTGTTATATTAAACATATTAGTCCCTATCAGTGATAGAGACATGCATATTATTTAATAGAATATTAACCT
P127	2–D3	TAATTTATTTTACATTTCAATGATTTTATGACATAATTATCTTCTGATTCCCTATCAGTGATAGAGAAAGTATTATAATAAATACAATTAAAACAAAAGACCGCTCAC TCAAAGGCC"
P128	2–E3	Duplicate of 123
P129	2–F3	CTTGATTTTTGATTTATTTATGTATGTTACCTCTCTATCACTGATAGGGATAAACCGAATATAGTGATCTAACAATATGCAAATAACAATGAAAAACAGGATCCTTAT TTCTTTTTGTTTATCTATCATATCTATGTTATTATTTAAGTGATTCCCTATCAGTGATAGAGACCTAATTGACTACAAAACTAATTTTTGAAC
P130	2–G3	Duplicate of 120
P131	2–H3	Duplicate of 106
P132	2–A4	CATTTTAGTTTGACATACAATTTATCATATATTATAATGCTGTTATTATCCCTATCAGTGATAGAGATATAGCATTATTAAAAACTAAACAAACTGT
P133	2–B4	TTATTCTATTTTGTTTATAATATTTTTCTATCTCTATCACTGATAGGAATTTGATAGAAAATTTAAAAGAACTATATTATAATAGAAAAAATACAT
P134	2–C4	TTATATTGTTAGTATATGTATTATTGAATTTCTCTATCACTGATAGGGAAAATTATTTATGAAAAACTTTACAAATACATTGCTATACAAACTATTTGGATCCATCTT TTCGTTATGGTATACTTATATACTGTCTCTATCACTGATAGGGAAACCTATGCTTAATAACCCAATATCAAAATTTGTATTGTTATAATTAT
P135	2–D4	GTTTCTGTAACATATTCTTGCTTATTCTGWAACTTATATTATAATAAGTCCCTATCAGTGATAGAGAACGAAACAAATAAAATAAAAAATAATTTAA
P136	2–E4	TGTATACATTTCTTTGATTGTATTATAATTTTAATTTTTAATGCTCGTTCCCTATCAGTGATAGAGACTTTATGATAATATAAAGAAAACGAAAGAA

Continued on next page...

Table A.1 – continued from previous page

Name	Spot	Sequence (5'–3' with respect to downstream *cat* gene)
P137	2–F4	TATTCATTGTAATTTATTTATTTATTTGTTAAGATGCTTTAGGATTTTTCCCTATCAGTGATAGAGAAATTAAGGTAGAATAAACAAAAAAATTTAT
P138	2–G4	Poor sequence data
P139	2–A5	TTTATGGTATACAATAACTTTATATGTTCTTCTCTATCACTGATAGGGACTAAAAAATAATGATAAAAAAAATTATAATGTATATAATAGTAAAGAA
P140	2–B5	TTTTTTAGGAACTATTTATATTATGTTTTTATTTGGTTGATATTAGATGTCCCTATCAGTGATAGAGAACCGAAATAAATATTAAAAAACAACCCTTTGGATCCAAAA TTGATATTATTATTCAAATCTGTTACAATTAATATGTACGACATTCCCTATCAGTGATAGAGAAATAAAACATAAACCAAGTGTAATGCATTA
P141	2–C5	TTAATAGTTATTATTAATTTTGTTATAAAATCTCTATCACTGATAGGGACAAACAGCTAATAAATAAAGTTAAAAAAATATTTATAATAAAAATTTGGATCCTCATAA CTTTTTTCACCTAAATTGTTTTTTTCTCTATCACTGATAGGGATATAATGGATTTACAACCAAAAACATTTATATTAATATTAGTAAAGAT
P142	2–D5	TTTTTTTTCTAGTTATTATGTCTTTTTTCATCTCTATCACTGATAGGGAAATAAAAAAGTTTAACATTTAGAGAAAAAACGTATAAACTATTAAAAA
P143	2–E5	TGGTTCTACTTATCTTTTTAGTTATAATTGTCTCTATCACTGATAGGGATTCTAGAAAAATTTTACAAAGATAAAAAATAGTTTATACTATAAGAAA
P144	2–F5	TTTGAATAATTGTTAAATATTTTTGGATATTCTCTATCACTGATAGGGACAAATAGTACAATATACTATAAAATGATTAATTAACAGAATTGCAATT
P145	2–G5	TTTGTTCTGGGCTTTACATTCTTAGTTGTCTCTCTATCACTGATAGGGATGACACTTAAGATTTAATTATTACACAAAACTTATTCCTTGATATTTGTGGATCCGTTT AGTTTATGTTAAAGAATCTTTTTTACTGGCTTGAAATTAATTATTCCCTATCAGTGATAGAGAGACTAAACAATTAAATAAATAATAGAAGAAGGATCCGATTTTTCG CATATTTTGTATTTTAGCTGTTCTCTATCACTGATAGGGACTTTACTATTCAATAAAACATTTATAATATTTAAATCAACTAATTATGA
P146	2–H5	TTGTTTAATATCATTTGTAAGTTATTTTAATCTCTATCACTGATAGGGATAAATACCAATTGACATATATAAATGATTCTGATATAAATTAGATAAG
P147	2–A6	ATTTTTATTTTCCTTATAGATGTACATTGGTCTCTATCACTGATAGGGATAACAAAAAAAAAAAACTAATAATATATAATAAAATCATTTTTAAATGGATCCTGTTTG TATATAATTAGTTTTAAATGTCACTTTTAAATTAGGTTTCTTCCCTATCAGTGATAGAGAAATTACTCAAAAGACAACGTATTTAAATCT
P148	2–B6	CAATGATTAAACAGTTAATATCATAAAGTTTCTCTATCACTGATAGGGAAAAAAAAACTACAAAACAATAGTATTTAAAGGTAAAATAGATTTACAAA
P149	2–C6	Duplicate of 148
P150	2–D6	CCACTTAATTCAATCCATTGTTCATTCATTTCTCTATCACTGATAGGGATATAAATGACTATTGCAAAATAAAATAATTTAAATTAAAATGAGTAAA
P151	2–E6	ATTTATAATAGTTTAATTTTCGGTTGATGTTCTCTATCACTGATAGGGACATTATGATTAAATAATAAATTCAAATAGATTTACCTATGCGATTAAGGGATCCTAGCT AAGACGAGGGTCTTTGTTTTTTATTAACTTTAGTTTTATTTAATCCCTATCAGTGATAGAGAGAGGTTTTACGTAAATAATGTTTGACATATGGATCCTGCTATTTTA AGATTGCTAACACATTTTATTTGATTGGTAATTTAATATCCCTATCAGTGATAGAGATAGTATAATAAAATAAAATATTTAGAAGCA
P152	2–F6	AATTTAATTTATGAAAATTTGGTCTTTGTGTCTCTATCACTATAGGGAACTTTACAACAAAAAAAGAATAAACTAAAATATAATTATGCAAAAATA
P153	2–G6	TTACAGTTATTTTATAATATATGGATATGATCTCTATCACTGATAGGGAAACACATGCGAAATTATAAGGAAAATAATTACCTAATGAAAAAATATTGGATCCGTTTA TGATTTTTGATATGATATTTGGTTAGTTCTTATACATAATGTTTCCCTATCAGTGATAGAGAATAAATAACAAAGGATACGAAATCGAAAAA
P154	2–H6	CATATTTTTCCGTCTATTTATCAGTATGTTTCTCTATCACTGATAGGGATACATACTATATAAACGACTATAAAGTAGAAATTTATCAAATAATTGTGGATCCCATGC GTTAATAAGTGTATAATATTACTAACATTTTGACAGGCGGTCATCCCTATCAGTGATAGAGATTAGATCTCTCGAATTTAAGAAAAAACAAA
P155	2–A7	CTTATTTTATTAATGTGATGTATGCCTGATTCTCTATCACTGATAGGGAAAAAAATGATTATAATCTTACGATTATGATTCAACAAGAAAAAAAGACGGATCCTGTTT TTTTGTATCTTAAGACTTATCTTATTCTCTATCACTGATAGGGAAAATATAACTAATAATTGTTTTATAATAAAATCTTAAATAACAGCAAAGGATCCCATCTGTGGG CTTAGCATTATCGATTTTTCTCTATCACTGATAGGGATACAAAATTTCTTGAAAAATAATTATGTTGATTATAAAATTTAAAAAA
P156	2–B7	Poor sequence data
P157	2–C7	TAAAGTGAATTCAATGGTTATTATTAGGATTCTCTATCACTGATAGGGACATAATAAGTATTTAAAAATATTATGGATTGTCCTATAATATATATAT
P158	2–D7	CTATGATTGAGTTTCGAATTTTTAAATATATCTCTATCACTGATAGGAATTTTAGTAATTGTAATTTTTAGAAATTTTAGTTATCATATTTCTCACT
P159	2–E7	AGAATTTACTCTTCATTTATTTATTTTTCATCTCTATCACTGATAGGGATAAAAAAATAATTTAAAAACCATATAAAAGATGATATTATAGCACTAC
P160	2–F7	AATGATTATCAAAGTTTTGTATCTTTAATTTCTCTATCACTGATAGGGATATAATTGTTTACAACTAGATTAAATTAAGATATAATTATATACTAAC
P161	2–G7	TTTCTTTAACAATAGTTTTATAGTGTGTTTTCTCTATCACTGATAGGGAATTGACTAAGCATTAAAAAAAATGTATAATATAACGAACCACTAATCC
P162	2–H7	GGCTCAGGCAGTGCAATTTTTGTAACCGTATCTCTATCACTGATAGGGACAACTTAGGCAAATAGGACGAAATAAATATATATCAACACTATGAGATGGATCCATATT CATTGGCAATTTTATATTTGTTTCATCTCTATCACTGATAGGGAGAATAAATTTCCAATTAATAAATTTTTGCGATATAATAAAACACAATA

Continued on next page...

Table A.1 – continued from previous page

Name	Spot	Sequence (5'–3' with respect to downstream *cat* gene)
P163	2–A8	TATATATGAGAGATTACTTAATTTGTATATTCTCTATCACTGATAGGGAAAAAAACCCATAGATTTTTTCTCAAACTATCTTTATAATAAGATAATT
P164	2–B8	TTTGACTTTGGGACTTATTTTAAACCCCTATCTCTATCACTGATAGGGACTAAATTATTTGACAACTAATATTAGCCTAGTATATACTTGTTGTGGATCCAATTTTTTTATAATAATTTTGTAGTTTTAATATAATAGGTGTTGGAATTTCCCTATCAGTGATAGAGAATCAAAATATAAAACAGTTTAGAAATATCG
P165	2–C8	ATATTCTATTACGCATTTTGAATATGATTTTCTCTATCACTGATAGGGAATAAATCGCTTTCAATAAAATATCAAATATATTAAAATAAAGCAAATA
P166	2–D8	CAATATTTGACGTTTAATTTTGGAATTATCTATTAATTATATATAATTTCCCTATCAGTGATAGAGATACTTAGTAACAAGAACTCATAACATACAAGGATCCTATGGGTTTAGTTATATTTTTTAGATGAATTTTTATTTTACCGTTTATTCCCTATCAGTGATAGAGAAAATATTATTTGACTATACGTATCCTATTTAGGTATAATAGTGTATTAGAATTTCCCTATCAGTGATAGAGAAACACTAATACTTAGCTAATATTTAACCAAGGATCCTTATAGTGTTCGATTGCTTGTCGTATTTAATCTCTATCACTGATAGGGAGCACTTTAAAAAATCCAATTATTATAAATAACTAGTTGGATGTAAAGA
P167	2–E8	CTATGATTGAGTTTCGAATTTTTAAATATATCTCTATCACTGATAGGAATTTTAGTAATTGTAATTTTTAGAAATTTTAGTTATCATATTTCTCACT
P168	2–F8	TTATTGTGTGAAGCATTACTTCATAGGCTTTCTCTATCACTGATAGGGATGACTTTAACTTTCTGACATAATAAAGAAAAAAGTAAATATTCAGTTAGGATCCTAGATATTAATGTAATTTTAGTATATAATTTACTCAATAAATCTACTATCCCTATCAGTGATAGAGAAAATTAACATAATTGAAAGATAATAACTAA
P169	2–G8	CTATATTCCTATAGTTAATTGTTTTCCAGATCTCTATCACTGATAGGGAATAATTTTTACTTGACAAAAATACAGAAAACATAATAAAATAATA
P170	2–H8	TTAGGAAGTGGTATGATTTCTGCAGATGCTTCTCTATCACTGATAGGGACAAAACTTCTAGAATTTTAAGATATATTTTGATACACTAGAAACAAAA
P171	2–A9	Duplicate of P146
P172	2–B9	Duplicate of P155
P173	2–C9	TAACATTTCAAGCAATGTCTTTCGCATGTTTCTYTWTCACTGATAGGGAAAAAAAAAAAAAACTGAATCACTCAATGTTTCATTTGAAAAAAATTGAAGGATCCCTTATAATGAATTTAATTAATTGTTTATTTTATAATTTATTTATTTTTTCCCTATCAGTGATAGAGAAACAAATACAATAACAGAAAAAAAAACCAA
P174	2–D9	GTTTTTATTAATGGTCTTGAACTTTAATTTTCTCTATCACTGATAGGGAATAAAAAAATTATTGACTAACAGCAAAAACTCAGTATAATTATTGTAA
P175	2–E9	ATATTTAAAAATATAATGAGCTATTTAATTTCTCTATCACTGATAGGGACACACTGGTATGTTTAGTGATAAATGTTTCACAAATAACTAGATATTTGGATCCTTTTAATATATAGTAGAATACATTAAATATTTCTAAGATGAGTGTTCATCCCTATCAGTGATAGAGATAACAACGATAAATACAAAGAACAAATAAA
P176	2–F9	TACGTGTGCATAGTATATAAATTTTTATTATCTCTATCACTGATAGGGATAAAATTTGATAAAAAACATTTGTAATCTTATATAATAATGATTATTA
P177	2–G9	TTTTTGAATAGATTTTAGTTATTAATATGATCTCTATCACTGATAGGGAACATAAAAAACATTGTGATTCTCATTTATATATGTTTAAATAAACCAA
P178	2–H9	GTTTGATTAGTAAATGAATTTATTTTGCATTCTCTATCACTGATAGGGATAATAAAACAATCGATACTTTGCTTTAAAAGCCTTTCTAATATTTAATGGATCCTATACAATTGTGCTCTATAAATAATTAGATTATTGTTGTTTATTTAACTCCCTATCAGTGATAGAGAGTAATAATACCAAAAAAAAAAATAGAAACAG
P179	2–A10	TTTTTTCTCCCAAGTATTTTTTTTTTGTCAATCTCTATCACTGATAGGGATAATAAAAATTGAACCAAATCAATTATATGATATATTATAAGGTTATA
P180	2–B10	TTCCAAAGTGGATATTAAAAAGATCTTAGTTCTCTATCACTGATAGGGAGATAATCTATAGATTTATGATAAAAATCATACTAAAATATTATAATGA
P181	2–C10	TAACATTTCAAGCAATGTCTTTCGCATGTTTYTYTATCACTGATAGGGAAAAAAAAAAAAACTGAATCACTCAATGTTTCATTTGAAAAAAATTGAAGGATCCCTTATAATGAATTTAATTAATTGTTTATTTTATAATTTATTTATTTTTTCCCTATCAGTGATAGAGAAACAAATACAATAACAGAAAAAAAAACCAA
P182	2–D10	Duplicate of P179
P183	2–E10	Duplicate of P179
P184	2–F10	TTAGCTTTTTCTATTATACTTGAATGTATGTCTCTATCACTGATAGGGATAAACAACCTAATTAAAAATTGGCGGTAAAACTATTGGAAATCAAGAAGGATCCTGCTATCATAAACAGACATTACTGTTCTATTAGGGCGAAGTTTTATTTTCTCTATCACTGAATAGGGAAAAATAAGCAAAAAAATTAATTGGAACTATATATTCATCTATCAAGAA
P185	2–G10	Duplicate of P170

Table A.2: Sequences of synthetic *E. coli* promoters of clones presented in Figure 2.9

Name	Spot	Sequence (5'–3' with respect to downstream *cat* gene)
PE-69	A1	TGATTTAATAATCTTCTAATATTAGAGTTGTCTCTATCACTGATAGGGATTAGACAAATAGAAATAAAAATAGCAACTAGTAAAAATTATATCACTG
PE-70	A2	Duplicate of PE-69
PE-71	A3	ATTTGTCCTTTTTATTTGTTTATTTGGTAATCTCTATCACTGATAGGGAATCATGGTGAAACGTAAAAAGCAAAGTGATATCATTAAAGAATTAAAA
PE-72	A4	TTTTGTAAGTTTATTATTCGTTTCTATTTATCTCTATCACTGATAGGGAGAAAGATATATATTCAAATAGATTTTATTTACAAAACTGAAAAGTATA
PE-73	A5	GATTTTTATTTGTGTGTTTACGTTAATTACTTGTTTAATTTGATATTATCCCTATCATGATAGAGATACGTAATATAAAAGCCTGAATCTTGATATA
PE-74	A6	CTGAATTTAAGTTTGTTATGCTTATTGGATTCTCTATCACTGATAGGGAAAATTTAACTTGTTAATCATTATACCTTTGACAAATAAAACTATAAAA
PE-75	A7	TTTAATTTTTTTATATATTTATAGCTACAATAATGTAACATCCCTATCAGTGATAGAGAAAAGTACGAAAAAAGCCTATTAATAGTGAC
PE-76	A8	ATATTGACATTTGTTGAAATTGATTTATTTTCTCTATCACTGATAGGGATATCTATACACAGTATAAGATCGAAAAAATACTCGACAATATCAATAA
PE-77	A9	TGTTTTACTTAAATACATCTAGGGTGTAAAGTATTTGTCTACATAATATCCCTATCAGTGATAGAGAACATCCAAAATATCGGTATAGCCGAAAAGC
PE-78	A10	ATGTTTATCTGTAATTTTTATTGAAATATGTCTCTATCACTGATAGGGAAAACCTAAGGTATTGTATTTCTGGTGATCAGTATGATATTATAGACAG
PE-79	A11	AAGATTGTTTAGGTATAGTAAGTCTGTTTTTATTGGAATAAATACAGATCCCTATCAGTGATAGAGAAAATCAAATTTTTTAGCAATAAAAAAGTAG
PE-80	A12	Duplicate of PE-76
PE-81	B1	TTGATAAATTGTTTTTTACGGATGGTATGCGTTATTGAAAATATCTATTCCCTATCAGTGATAGAGAATATTATAATCGTTAACACCTCTAATAGTA
PE-82	B2	TTTGTATATAATAAGTTATAATGTATCATTTCTCTATCACTGATAGGGAAATCAAGAAAGAATCATAATAAAAGGGTGATTTATTAATCCAAAAAAA
PE-83	B3	TCTGATAAATATTTATTACACTTCTTGATTTAGTCAATAAATAACTTGTCCCTATCAGTGATAGAGAAAAAACTATAAATCAAAAAACTCTAATACT
PE-84	B4	TATAAATTTTCATGTAAATTAAGTATCTTTTAGTACAGTTTATTTTGTCCCTATCAGTGATAGAATAAACTACTCTAGAAATGAACAAACTATAT
PE-85	B5	TTCTTCTTTTGATTTTGTTTTTTACTTCTATCTCTATCACTGATAGGGACAATCAAAAGTAAATTCTAAAAATAAAAATTAGTTGATATAATGAATC
PE-86	B6	Duplicate of PE-78
PE-87	B7	TTACGTAAATTATTGTTATTACGTTTATCTCATCATATTTCTCTTGATTCCCTATCAGTGATAGAGATACTTAAAAAAAAAGCTATAATGATTAAAT
PE-88	B8	TTAAGCCTTGTTTATTTTTTCTTTTTTAGATCTCTATCACTGATAGGGATTTCTTGTTATATTGACAAAAAAATCGATCGATCAACCATGGCATGGT
PE-89	B9	Duplicate of PE-69
PE-90	B10	Duplicate of PE-82
PE-91	B11	GATTCTTTGTAGATGTTATATTTTGCTGTTTCTCTATCACTGATAGGGATTCAAAGAAAACCTAATAAATTAAAAATTGTTAACTCTCACTACGAAA
PE-92	B12	TATTAAATATTATCTGTTTATGCTATATTCTATTTTTTTTTTTAGTTAGTCCCTATCAGTGATAGAGATAAAGAAAAAAAAAAAAAAAACCCTTAAAAG
PE-93	C1	CTTTTAGGTTTAATATGAAATAAATAGATTTCTCTATCACTGATAGGGAATAGTGTGTTATGTTCTTACAATATTGCAATATAAATTCTTCTTACGT
PE-94	C2	TTTATTATGATTGTTAAATTAACTCAAAAAATGTAACTTATAGTTTTTTCCCTATCAGTGATAGAGATAGAAATAGTAGAATTAGTAATAAAAGAAA
PE-95	C3	TTTGGATATACAGGAGTTTCTACTCGTGTTAAAGTTTTTTGGTATTTTTCCCTATCAGTGATAGAGAACCAATGAAATATTTAAAAGAGTATAAAAA
PE-96	C4	TTTTTAAGAGCTCAATCTTAAATTTCATGCTAAAATCTGTAGAACTTATCCTATCAGTGATAGAGAATAAACAATTATTACGCATAATGAATGACA

Continued on next page...

Table A.2 – continued from previous page

Name	Spot	Sequence (5'–3' with respect to downstream *cat* gene)
PE-97	C5	TAGTCTTATTACTATGTACAATGTAGGGTTTCTCTATCACTGATAGGGAGAATGGTTTTATTTAGATATTGTAATACAAACAAATCTGATTATTAAC
PE-98	C6	TTAATTTTTTTTATATTATATTAGATAATTTCTCTATCACTGATAGGGAACAAAAACATAAAGGAATAACAAATATCTATGTTATAAACTGAAACAA
PE-99	C7	Duplicate of PE-93
PE-100	C8	TTTTTATTGAATTTACTTATTCACAAGTTGTAAGTTTTCTTTATACTATCCCTATCAGTGATAGAGAAGCCAATTCATTACATGTTTTTTAACGAGG
PE-101	C9	TTAAAGATTTCGAGTAGAAGAATATATATATAGATGATATTTTTATGTTCCCTATCAGTGATAGAGACCTAACAAAAGAAAGCACGAAAGTATCCAA
PE-102	C10	AAATTAACTTTACTATGATTTTTTTTAGGATCTCTATCACTGATAGGGAATATCCCAAGCATTTAAATTAATATTTAAACTAACAATTAAAAA
PE-103	C11	GTGTGTGTTTAGGATGTAATATTATTGACAAGTCGCACTAGAAAATAATCCCTATCAGTGATAGAGACAAAAGCATAGTAAGGATAAGTCTAGAACT
PE-104	C12	AAACAGGATTAATTATTCTGATAGTATGTTACAAAAGTTATCTTTTACTCCCTATCAGTGATAGAGAACAAAACAAATTTCCTTTAGTTTTCACAAA
PE-105	D1	CTATAATTTCATCTGAGGAATTTTTTTTCATCTCTATCACTGATAGGGAAAAACTAACGAAAAAGGTACAACAAATGCTAAAATTATCAAAACAACT
PE-106	D2	Duplicate of PE-91
PE-107	D3	Duplicate of PE-100
PE-108	D4	ATTGGTATCCAAAGCAAAGAGTGTTAATCTGATTATTCTTTTTTAGATTCCCTATCAGTGATAGAGAAATATTAGAGAAATCTATAGTACCTTAAAT
PE-109	D5	TTAAGATTTTAGATTATAGAATAGTGTTTGTAATAGAGAATTGTAACTTCCCTATCAGTGATAGAGAAATTCAATTATTTCAAAACCACCACATATG
PE-110	D6	AAATGTTTAACATCTCTACTTATTGCTTTATCTCTATCACTGATAGGGATAAAATACGTCAGACAAATAATCCTTTAAAAGCAAGCATTCAATTCGA
PE-111	D7	TTTTTTGTTTTCATATCATTTCGAACGTGCTAGTTTATGTATTCATTTTCCCTATCAGTGATAGAGAATTGTGATTTACAGTAAAATTAATAAAAAA
PE-112	D8	ATGTTATTCTGTGATGTATATTCCGTATTTGTTCATATGTTGTATTAGTCCCTATCAGTGATAGAGACATACCTAACATCGAGCACCTAAAAACATA
PE-113	D9	TTTAAGTTATGTGGTTTCAGAGCACTTATATCTCTATCACTGATAGGGAGAACATAAAATCGATAGCTAGAATTAAATAGTAAATATATAATTTTTA
PE-114	D10	Duplicate of PE-82
PE-115	D11	TAATTTTATTTTTTATTATTATTTTGTGGAGCTACTGAGTATAATATTTCCCTATCAGTGATAGAGATACGAATATTGAACTCAAATAAAAATTCTT
PE-116	D12	TCAGATTTTTCATCTATATTATATTTTAGTCATTATCTATTAGTTTGTTCCCTATCAGTGATAGAGATTTACGATTAAAATGCAGTCTAAGAAAAAA
PE-117	E1	TAAGTTGTATTGTATAACTAGGTTATTCGGAAAGTTTTACTTATAATCTCCCTATCAGTGATAGAGACGAAAACATTCCAAAGATTATATTTATGTA
PE-118	E2	TATAAGATTCAATGTTAGTATAGTATAATATCTCTATCACTGATAGGGAATCATAGTATTAGACGCTTTAAAATATTCACGGTTCCATTATAATAAA
PE-119	E3	ATATCTTTTTTTTTTTTTAGTAAATTGTTATCTCTATCACTGATAGGGATTTATTTACCTTATAATATTTAAAAACTTAATATGACAAGACAAGCTA
PE-120	E4	Duplicate of PE-69
PE-121	E5	Duplicate of PE-82
PE-122	E6	Duplicate of PE-77
PE-123	E7	TCTGAGTTTGCTTCTTTTATTTCTTTTTGCCTTATAATTAAGGATGATTCCCTATCAGTGATAGAGAAGTGCACTCAAAAATATAAATAGATACAAA
PE-124	E8	AATGTTAATGTTTTCATGTTGTTATAAAATTCTCTATCACTGATAGGGAATAGAATTAAGTTTAAATATCTAAATAAATAAATATAATAAAAACCAT
PE-125	E9	Duplicate of PE-82

Continued on next page...

Table A.2 – continued from previous page

Name	Spot	Sequence (5'–3' with respect to downstream *cat* gene)
PE-126	E10	Duplicate of PE-100
PE-127	E11	Duplicate of PE-93
PE-128	E12	CCTGTAGTATCAAAGTATGATAATGTTTTATCTCTATCACTGATAGGGAATAATAGAACATTAACAAAATCAGACAAATCAGTATGTCTTACGTAA
PE-129	F1	Duplicate of PE-91
PE-130	F2	CTATTTTGCAAATGATGTTCTATTGCTTTAAATAATTGTTTATTTTCTTCCCTATCAGTGATAGAGAATAAACTTAATTAAGTCTTTTATACAAAAA
PE-131	F3	GCAATAATATTTTATTATATATCTCGTATATCTCTATCACTGATAGGGACGAAAAAAATCAAGGGATATAGTAGTAATAATTTACTATAATTAGTTA
PE-132	F4	Duplicate of PE-77
PE-133	F5	TTTTCGTTTAACTTTTTGTATTCATTTTTACATTATAAAATTGCAATTTCCCTATCAGTGATAGAGAATTAAAAAAGAACTTTTTAATTAATTGAAT
PE-134	F6	Duplicate of PE-96
PE-135	F7	Duplicate of PE-93
PE-136	F8	ATAATTATTGGTTTTGTCTTGTTTTTTTTGTCTCTATCACTGATAGGGATTATAGTGATCAAAACAGTATAGAGTATTAGAAATATTACAATCCATC
PE-137	F9	Duplicate of PE-82
PE-138	F10	TTGCTAACTACAAATGGTACTATTAGTTAATGCGGACTACTTAAACTATCCCTATCAGTGATAGAGAAAAATAAATTTAAAAAATAATAGGTTAATC
PE-139	F11	Duplicate of PE-73
PE-140	F12	TTAAAATCATGACGTATTATCCGTCTAATTTCTCTATCACTGATAGGGACATTTGCTTTTAACACTATTAGAATCTCATATAACAAAAACAACCTGC
PE-141	G1	AATTAAGTAGGATTTGTTTTACTTCTTTTTAATGTATGCTATTAATTCTCCCTATCAGTGATAGAGAAATTCTTAAAAAAAAGTTAAGAAAATCAAA
PE-142	G2	TTATTCTTAAAAGTTGTTGCTTCCTTAAAAGACAAATATCTATTAAATTCCCTATCAGTGATAGAGAAAGCTCATCGTCCTCAACATTGATACAAGA
PE-143	G3	ATTTATACATCGATCGTATTCTTATTTTTTATAAATAATGTATTTGTGTCCCTATCAGTGATAGAGATAAAATTATATTATCATATATCAAAAATAA
PE-144	G4	Duplicate of PE-100
PE-145	G5	Duplicate of PE-100
PE-146	G6	ATTTTGTTAACTCCTTTTTTATTTATTATGTGTAAACGTTGCATTCTTTCCCTATCAGTGATAGAGAATGTTAAAGAACTCTCCCATTAACGTAATA
PE-147	G7	Duplicate of PE-100
PE-148	G8	Duplicate of PE-100
PE-149	G9	AAAGTAATTTTATTTACTTTCTTTTTCTTATCTCTATCACTGATAGGGATCGAAGATGACCTATATTAAAATCACAGAAAAAAGACAACGACAAACT GGATCCTACTATTCTCGGTCTTAGTAAAGAATCATAGGTTGAATGTCCGATTTTTCCCTATCAGTGATAGAGATATTAGGATATTATTGGTAGTATG CAATTA
PE-150	G10	Duplicate of PE-100
PE-151	G11	Duplicate of PE-149
PE-152	G12	AAAGAATCATTTCTATTTTGAAAATATTTGTCTCTATCACTGATAGGGATTAAAATAGTAATTTGAACCAATAATTATCAGAAGGGGTTAACACTTC

Continued on next page...

Table A.2 – continued from previous page

Name	Spot	Sequence (5'–3' with respect to downstream *cat* gene)
PE-153	H1	AATTTTAATTTTTCTGTAAAAATAGAATTGTCTCTATCACTGATAGGGAATGAAAGGAAGTAAAAATTTAATATAGTGCAATCTAATGAACTATTAG
PE-154	H2	TTTAGTTTCTTTTTTAATTGTTAAAATAAATCTCTATCACTGATACGGAAAGAGACATAATCTAAATTGAGTAATAAATTTTTTTCAACCATTAAAA
PE-155	H3	CATAAATCTTGATTATATTATCAACGATTTAATTTTTTATATTTGCTGTCCCTATCAGTGATAGAGAAATTAAAGAATACAATCACTAAAACACAAA
PE-156	H4	AGCCATTCAATCTTTTAATCCATTTTAATACTTTTTCATAAAATCAGTTCCCTATCAGTGATAGAGAATTAAAATACGAATCAGCGAATATAATATC

Table A.3: Survival of select *tetR*^{ti} clones on 20 µg/mL at various temperatures. Level of growth is an average of two spots as shown in Figure 3.4 indicated by '+' signs with '–' for no detectable growth

		Growth on Cm at 20 µg/mL				
Strains	Wells	30°	33°	35°	37°	39°
tetR^{ti}-1	A1, A7	–	–	++	+++	+++
tetR^{ti}-2	B1, B7	–	–	–	–	+++
tetR^{ti}-4	C1, C7	–	++	+++	+++	+++
tetR^{ti}-5	D1, D7	–	+++	+++	+++	+++
tetR^{ti}-6	E1, E7	–	++	+++	+++	+++
tetR^{ti}-7	F1, F7	–	+	+++	+++	+++
tetR^{ti}-8	G1, G7	–	++	+++	+++	+++
tetR^{ti}-9	H1, H7	–	+	+	++	+++
tetR^{ti}-10	A2, A8	–	+++	++	+++	+++
tetR^{ti}-11	B2, B8	+	+	+	+	+++
tetR^{ti}-14	C2, C8	+	+	+++	+++	+++
tetR^{ti}-15	D2, D8	–	–	–	–	++
tetR^{ti}-16	E2, E8	+	+	+++	+++	+++
tetR^{ti}-18	F2, F8	+	+++	+++	+++	+++
tetR^{ti}-19	G2, G8	–	+++	+++	+++	++
tetR^{ti}-20	H2, H8	+	+++	+++	+++	+++
tetR^{ti}-21	A3, A9	–	+++	+++	+++	+++
tetR^{ti}-24	B3, B9	–	+++	+++	+++	+++
tetR^{ti}-25	C3, C9	–	+++	+++	+++	+++
tetR^{ti}-27	D3, D9	+	+++	+++	+++	+++
tetR^{ti}-29	E3, E9	+	+++	+++	+++	+++
tetR^{ti}-31	F3, F9	–	++	+++	+++	+++
tetR^{ti}-32	G3, G9	+	+++	+++	+++	+++
tetR^{ti}-33	H3, H9	–	+++	+++	+++	+++
tetR^{ti}-35	A4, A10	–	++	++	+++	+++
tetR^{ti}-36	B4, B10	–	–	+	+++	+++
tetR^{ti}-39	C4, C10	–	++	++	+++	+++
tetR^{ti}-40	D4, D10	–	+++	+++	+++	+++
tetR^{ti}-41	E4, E10	–	+	+	++	+++
tetR^{ti}-42	F4, F10	–	+	+++	+++	+++
tetR^{ti}-43	G4, G10	–	++	+++	+++	+++
tetR^{ti}-45	H4, H10	+	+++	+++	+++	+++
tetR^{ti}-47	A5, A11	–	+++	+++	+++	+++
tetR^{ti}-48	B5, B11	–	+++	+++	+++	+++

Continued on next page...

Table A.3 – continued from previous page

Strains	Wells	Growth on Cm at 20 µg/mL				
		30°	33°	35°	37°	39°
tetR^{ti}-49	C5, C11	–	+++	+++	+++	++
tetR^{ti}-53	D5, D11	–	++	+++	+++	+++
tetR^{ti}-54	E5, E11	–	+++	+++	+++	+++
tetR^{ti}-56	F5, F11	–	+	+++	+++	+++
tetR^{ti}-58	G5, G11	–	++	+++	+++	+++
tetR^{ti}-61	H5, H11	+	+	+++	+++	+++
tetR^{ti}-62	A6, A12	–	+++	+++	+++	+++
tetR^{ti}-63	B6, B12	–	+++	+++	+++	+++
tetR^{ti}-68	C6, C12	–	+++	+++	+++	+++
tetR^{ti}-72	D6, D12	–	+	+	+++	+++
cat^{-}, *tetR*$^{-}$	E6, E12	–	–	–	–	–
cat^{-}, *tetR*^{wt}	F6, F12	–	–	–	–	–
cat^{+}, *tetR*$^{-}$	G5, G12	+++	+++	+++	+++	+++
cat^{+}, *tetR*^{wt}	H6, H12	–	–	–	–	–

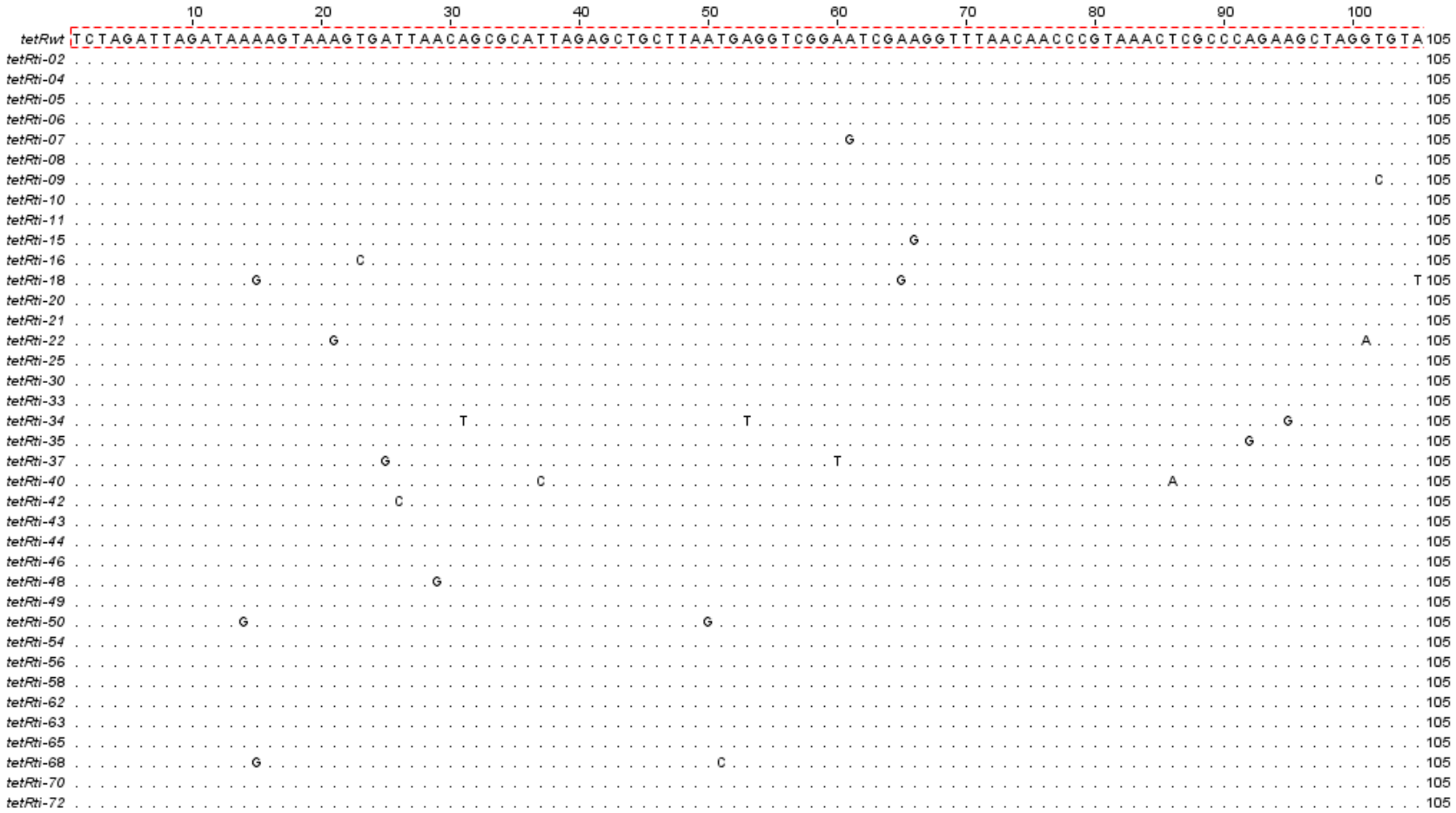


Figure A.2: DNA sequence changes in the 39 sequenced $TetR^{ti}$ variants

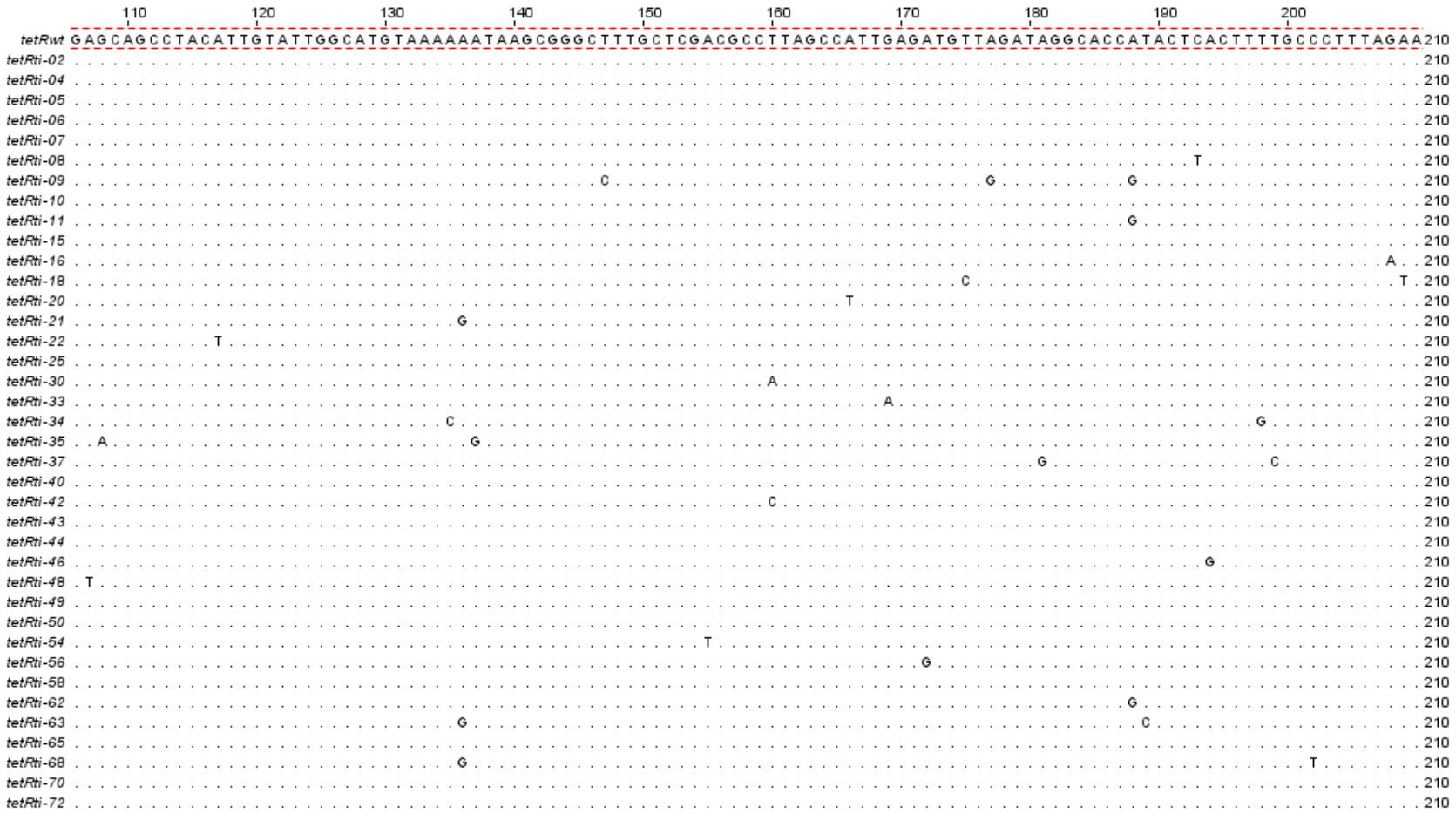


Figure A.2a: DNA sequence changes in the 39 sequenced TetR^{ti} variants, continued

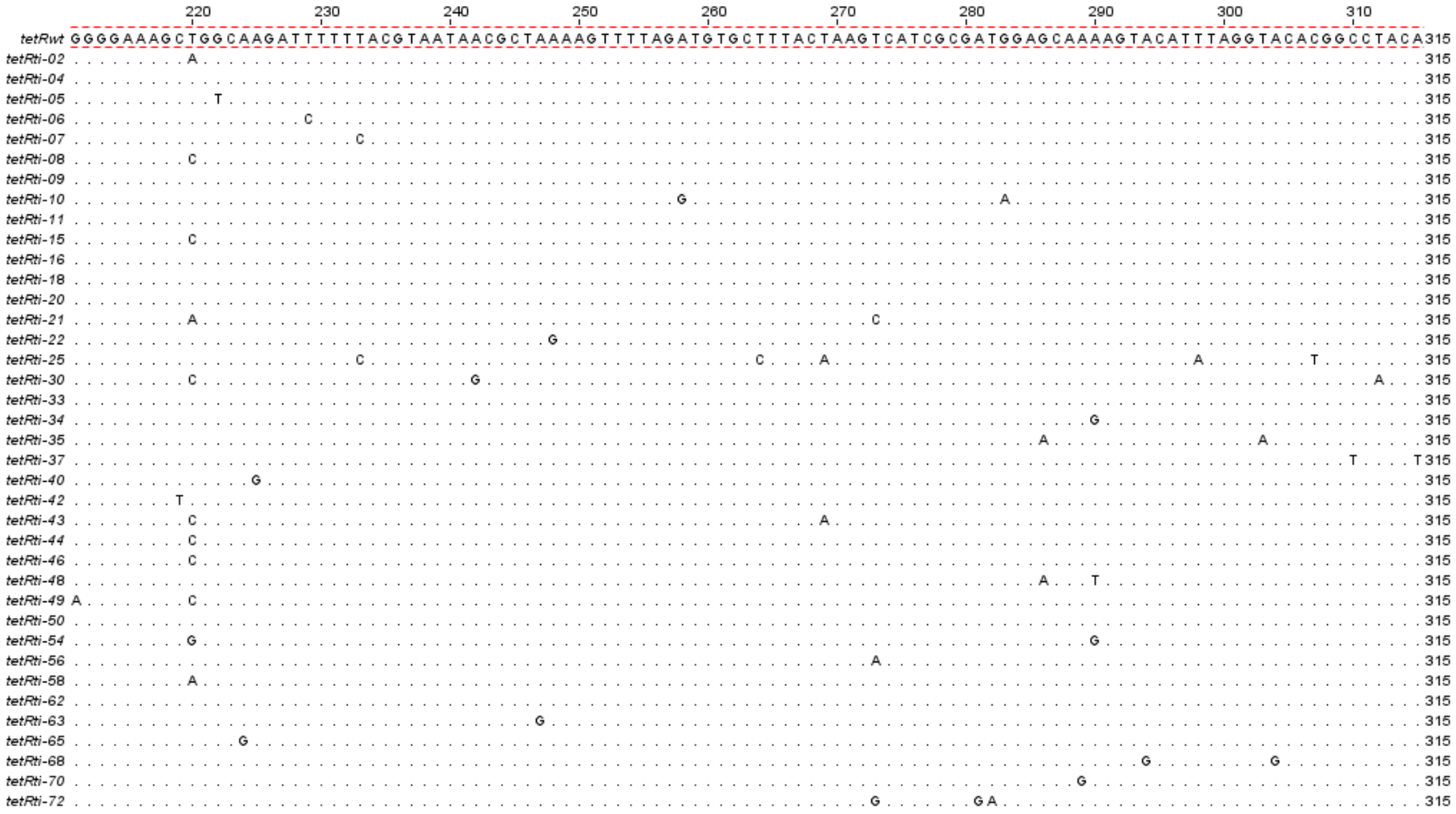


Figure A.2b: DNA sequence changes in the 39 sequenced TetR^{ti} variants, continued

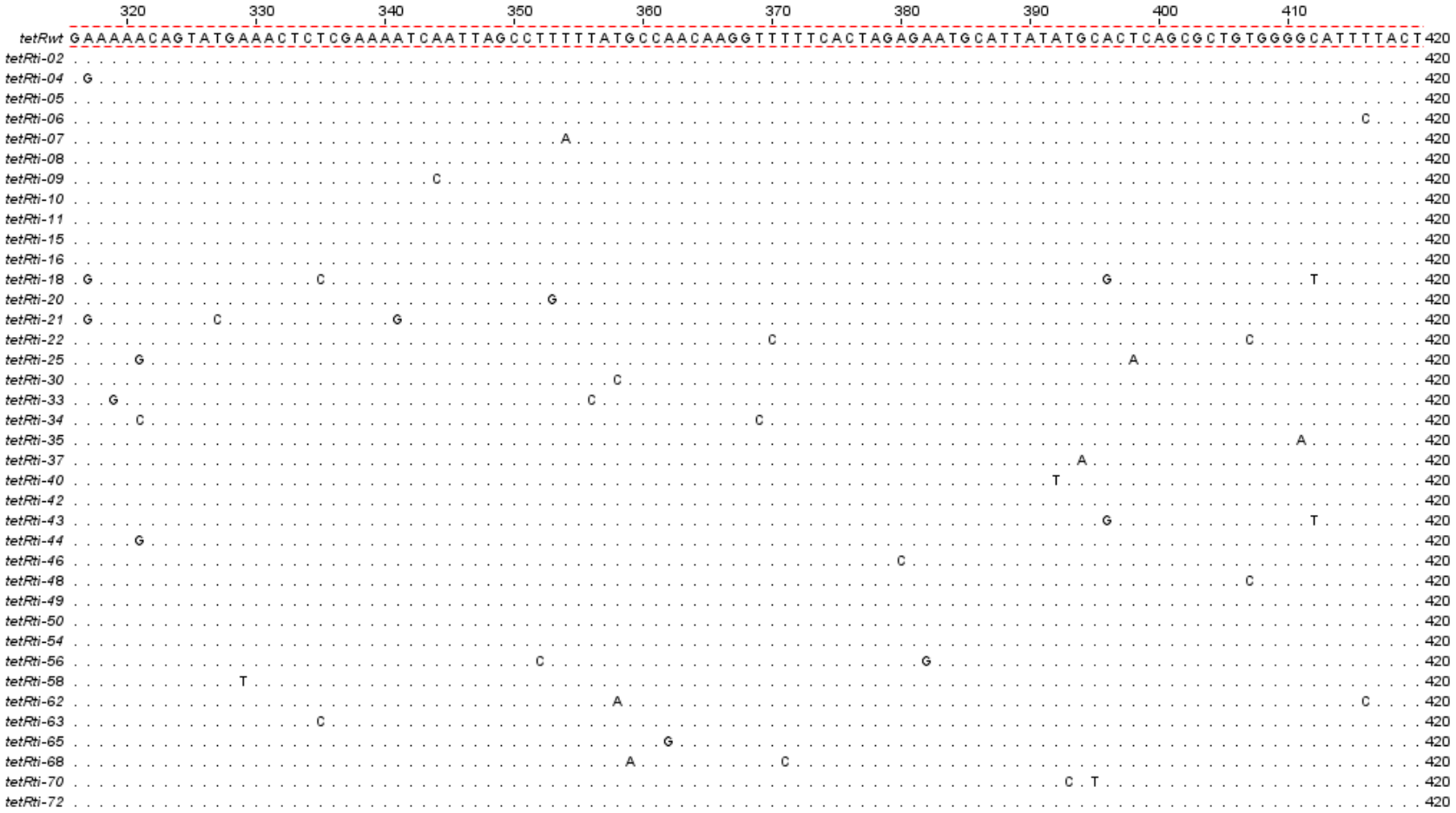


Figure A.2c: DNA sequence changes in the 39 sequenced TetR^{ti} variants, continued

Figure A.2d: DNA sequence changes in the 39 sequenced TetR^{ti} variants, continued

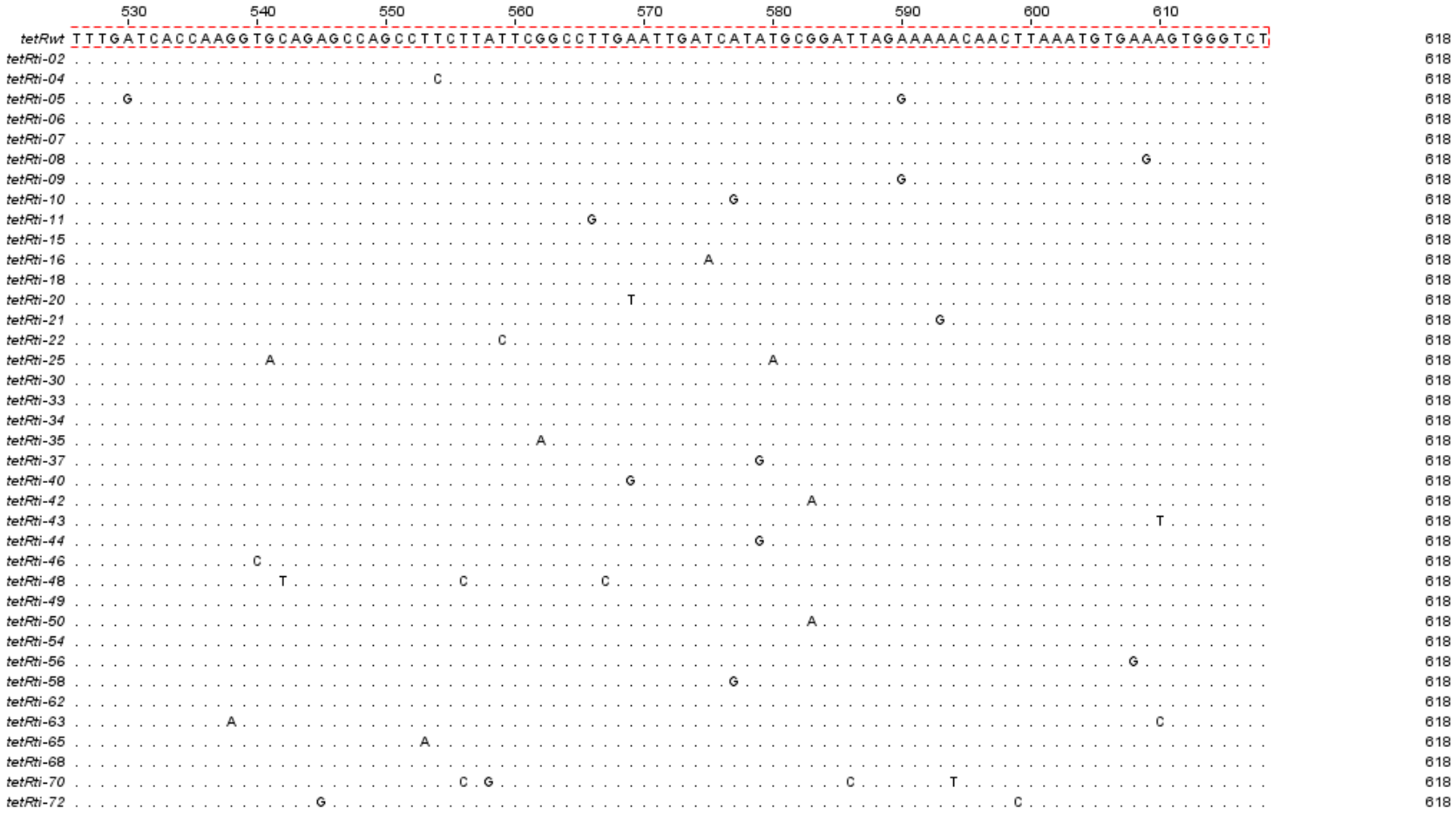


Figure A.2e: DNA sequence changes in the 39 sequenced TetR^{ti} variants, continued