EXPLORING PROKARYOTIC

ONE-COMPONENT TRANSCRIPTION FACTORS

DEVELOPING GENERALLY APPLICABLE TOOLS TO INVESTIGATE TETR FAMILY TRANSCRIPTIONAL REGULATORS

By SANG KYUN AHN, B.Sc.

McMaster University

A Thesis Submitted to the School of Graduate Studies in Partial Fulfilment of the Requirements for the Degree Doctor of Philosophy

DOCTOR OF PHILOSOPHY (2012) McMaster University
(Biochemistry and Biomedical Sciences) Hamilton, Ontario

TITLE: Development of Generally Applicable Tools to Investigate TetR Family Transcriptional Regulators

AUTHOR: Sang Kyun Ahn, B.Sc. (McMaster University)

SUPERVISOR: Professor J.R. Nodwell

NUMBER OF PAGES: xv, 179

ABSTRACT

Bacteria adapt to changes in their environment by regulating gene transcription. TetR family transcriptional regulators (TFRs) constitute one of the largest groups of bacterial transcription factors and thus, characterization of TFRs is anticipated to be crucial for a better understanding of prokaryotic physiology. Of significant importance, the majority of TFRs are predicted to respond to small-molecule signals and an emerging paradigm suggests that identifying ligands of TFRs can provide direct insight into the biochemical functions of the genes they regulate. Regulatory target genes and small-molecule ligands are unknown for all but a few TFRs and therefore, generally applicable tools for identifying these basic elements of TFRs are highly desirable. We first investigated the use of genome context as a predictive tool for identifying regulatory targets of TFRs. We find that the majority of TFRs are divergently oriented from a neighboring gene, and those with a <200 bp intergenic separation most likely regulates this neighbor. Our proposed “200 bp rule” should allow us to predict at least one regulatory target for more than half of all TFRs in the public databases. Second, we developed a biosensor mechanism amenable to high-throughput screening for identifying ligands of TFRs of unknown function. Significantly, one of our biosensors has played an integral role in characterizing the ligands of a previously uncharacterized TFR. Thus, the combined use of the tools we have developed will provide considerable benefit in understanding bacterial small-molecules responses mediated by TFRs.

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr. Justin Nodwell for providing guidance over the years. I would also like to thank my committee members Dr. Eric Brown and Dr. Lori Burrows for their insightful advice. I thank Dr. Alan Davidson at University of Toronto and his former student Zhou Yu for allowing me to participate in wonderful collaborative works. I want to also thank Dr. Mark Buttner at John Innes Centre and his former student Tung Le for their technical support. I thank all the past and present members of the Nodwell lab for sharing the most memorable period of my life. My special thanks go to the “TetR Machine”, Dr. Kapil Tahlan and Dr. Leslie Cuthbertson in particular, for being the “best” teammates to investigate the “best” family of transcriptional regulators. I want to also thank Dr. Arryn Craney who has witnessed and shared all the struggles and pleasures I have had since my first day in the lab. Finally, I thank my family for always supporting me and for being patient over the seemingly endless years of my graduate studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 3

CHAPTER 4

ABBREVIATIONS

(*S*)-DNPA	4-dihydro-9-hydroxy-1-methyl-10-oxo-3-H-naptho-[2,3-c]-pyran-3-(*S*)-acetic acid
ABC	ATP binding cassette
act	actinorhodin
AMP	adenosine monophosphate
ATP	adenosine triphosphate
BLAST	basic local alignment search tool
bp	base pair
CD	conserved domain
cpm	counts per minute
cps	counts per second
DBD	DNA-binding domain
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DTT	dithiothreitol
EC	enzyme commission
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
FAD	flavin adenine dinucleotide
His_6	six histidines
HMM	hidden Markov model

HTH	helix-turn-helix
Kb	kilo base pair
kDa	kilo dalton
Lux	bacterial luciferase
Mb	mega base pair
MFS	major facilitator superfamily
mRNA	messenger RNA
MW	molecular weight
MWCO	molecular weight cut off
NADB	NAD(P)H/NAD(P) binding
NAD	nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NADP	nicotinamide adenine dinucleotide phosphate
NADPH	reduced nicotinamide adenine dinucleotide phosphate
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
RLU	relative light unit
RNA	ribonucleic acid
RND	resistance-nodulation-division
SCO	*Streptomyces coelicolor* locus tag

SDS	sodium dodecyl sulfate
SGR	*Streptomyces griseus* locus tag
tet	tetracycline
TFR	TetR family transcriptional regulator
Tn	transposon
Tris	tris(hydroxymethyl) aminomethane
tRNA	transfer RNA

PROTEIN DATABASE CODES

PDB ID	PDB description	Reference
1BJY	TetR in complex with chlortetracycline	[56]
1QPI	TetR in complex with the operator	[60]
1JT0	QacR in complex with the operator	[110]

PUBLICATIONS

1. **Tahlan, K., Ahn, S.K., Sing, A., Bodnaruk, T.D., Willems, A.R., Davidson, A.R., and Nodwell, J.R.** 2007. Initiation of actinorhodin export in *Streptomyces coelicolor*. 2007. Mol. Microbiol. **63**:951-961.

2. **Ahn, S.K., Tahlan, K., Yu, Z., and Nodwell, J.R.** 2007. Investigation of transcription repression and small-molecule responsiveness by TetR-like transcription factors using a heterologous *Escherichia coli*-based assay. J. Bacteriol. **189**:6655-6664.

3. **Le, T.B., Fiedler, H.P., den Hengst, C.D., Ahn, S.K., Maxwell, A., and Buttner, M.J.** 2009. Coupling of the biosynthesis and export of the DNA gyrase inhibitor simocyclinone in *Streptomyces antibioticus*. Mol. Microbiol. **72**:1462-1474.

4. **Ahn, S.K., Cuthbertson, L., and Nodwell, J.R.** 2012. Genome context as a predictive tool for identifying regulatory targets of TetR family transcriptional regulators. Manuscript submitted to PLoS ONE.

5. **Cuthbertson, L., Ahn, S.K., and Nodwell, J.R.** 2012. A novel mechanism of inducible antibiotic resistance revealed using a global relational tree for one-component regulatory systems. Manuscript submitted to Chem. Biol.

CHAPTER 1: INTRODUCTION

1.1 Bacteria adapt to their environment

Bacteria are exposed to a wide variety of changes in their intracellular and extracellular environments (Figure 1.1). Environmental variations can make the living conditions far from optimal and some can be highly detrimental for bacterial survival. To overcome these circumstances and maximize their competitive advantage, bacteria must possess mechanisms for rapidly sensing stimuli and coupling them to cellular responses.

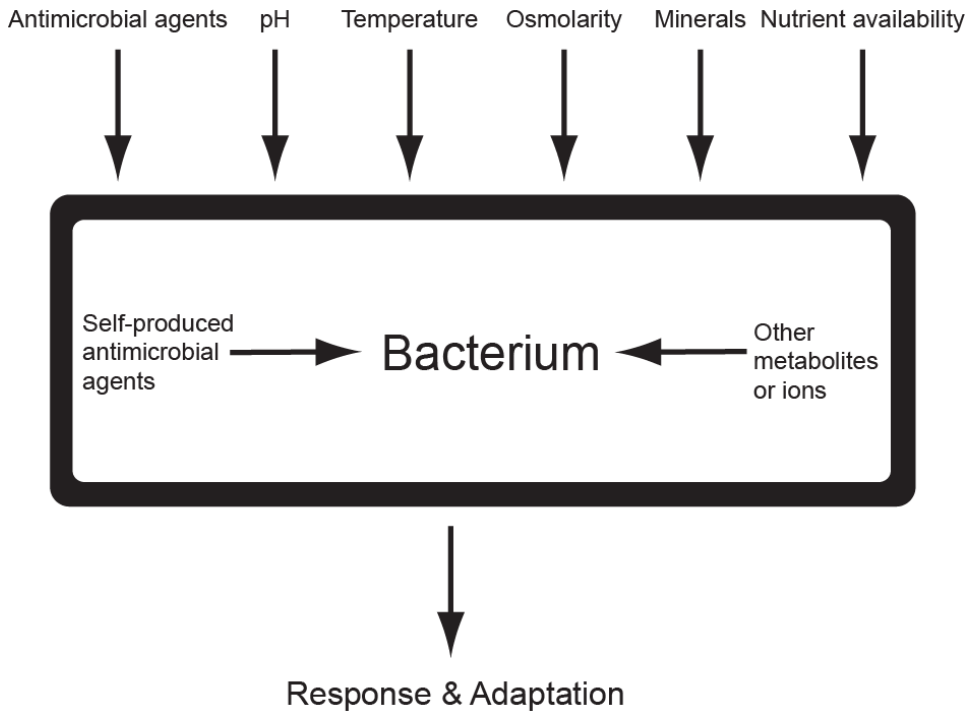

Figure 1.1. Various external and internal changes in bacterial environments.

Microorganisms have developed various response mechanisms for their growth and survival. As represented in Figure 1.2, most of these mechanisms involve the regulation of gene expression to adjust the concentration of proteins required for adapting to a particular environmental situation [1,2,3,4,5,6,7,8]. However, bacteria can also react to the signals by post-translationally modifying activity or stability of existing proteins [9,10].

For example, *Escherichia coli* increases the concentration of proteins required for the uptake and metabolism of lactose when it is grown in the presence of lactose as a sole carbon source [11]. Similarly, a shift in the surrounding temperature from 30°C to 42°C induces synthesis of more than 20 heat shock proteins in *E. coli*, including a variety of chaperones that aid in preventing protein aggregation and promoting refolding of misfolded proteins [12]. *Bacillus subtilis* can control production of the enzymes required for biosynthesis of tryptophan depending on the availability of this amino acid in the environment [13], while *Vibrio cholerae* – a non-pathogenic, free-swimming organism when living in marine and estuarine environments – secretes cholera toxin as a response to colonization of human intestinal tract and causes severe diarrhea [14].

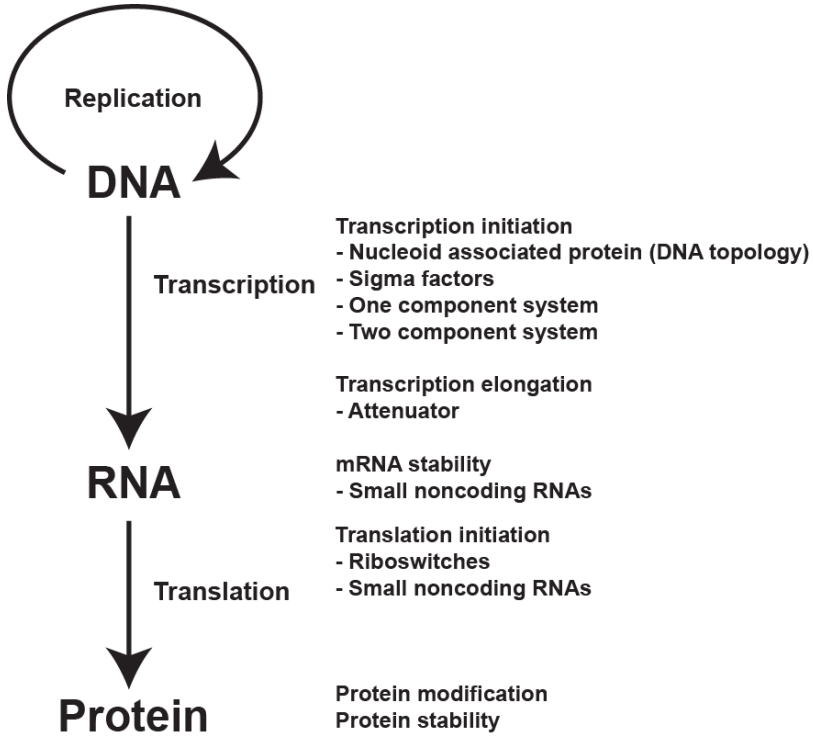

Figure 1.2. Various mechanisms of regulating gene expression or protein activity. Central dogma of molecular biology is shown on the left, while examples of the responsive mechanisms used by bacteria for environmental adaptation at different stages are indicated on the right.

1.2 Transcriptional regulation is the dominant response mechanism

Although gene expression can be controlled at various stages (Figure 1.2), regulation of transcriptional initiation is the most common mechanism bacteria use for responding to a wide range of environmental cues. In most cases these rapid, adaptive responses are triggered by DNA-binding proteins called transcriptional regulators (or transcription factors) whose ability to modulate transcription by interacting with DNA (and with transcriptional machinery) depends on the presence/absence of particular signals. These proteins can

either enhance or repress the transcription level of regulatory target genes, while a few proteins have dual activities. Transcription initiation is suppressed when repressors bind DNA to prevent RNA polymerase from either recognizing its target promoter elements or transitioning into an active transcribing state [15]. While mechanisms of transcriptional activation are more complicated, many activators function by interacting with RNA polymerase to facilitate its binding to target promoters [15].

1.2.1 Two-component regulatory system

Most transcription factors play their biological roles in “one-component” or “two-component” regulatory systems, which differ in the number of proteins required for linking input stimuli to transcriptional responses (Figure 1.3). In a typical two-component regulatory system (Figure 1.3A), two key proteins – a sensor histidine kinase and a response regulator – are involved in signal transduction [4,16,17,18,19,20]. The prototypical sensor histidine kinases are membrane-bound proteins with an extracellular domain that detects environmental signals, whereas all response regulators are present in the cytosol. Upon sensing a particular stimulus, the catalytic and ATP binding (CA) domain of a sensor kinase catalyzes autophosphorylation of a histidine residue in the dimerization and histidine phosphotransfer (DHp) domain (Figure 1.3A). This phosphoryl group is then transferred to a specific aspartate residue in the receiver domain of a response regulator, activating the protein to conduct its

biological function through binding DNA (via the output domain) to modulate transcription for a cellular response.

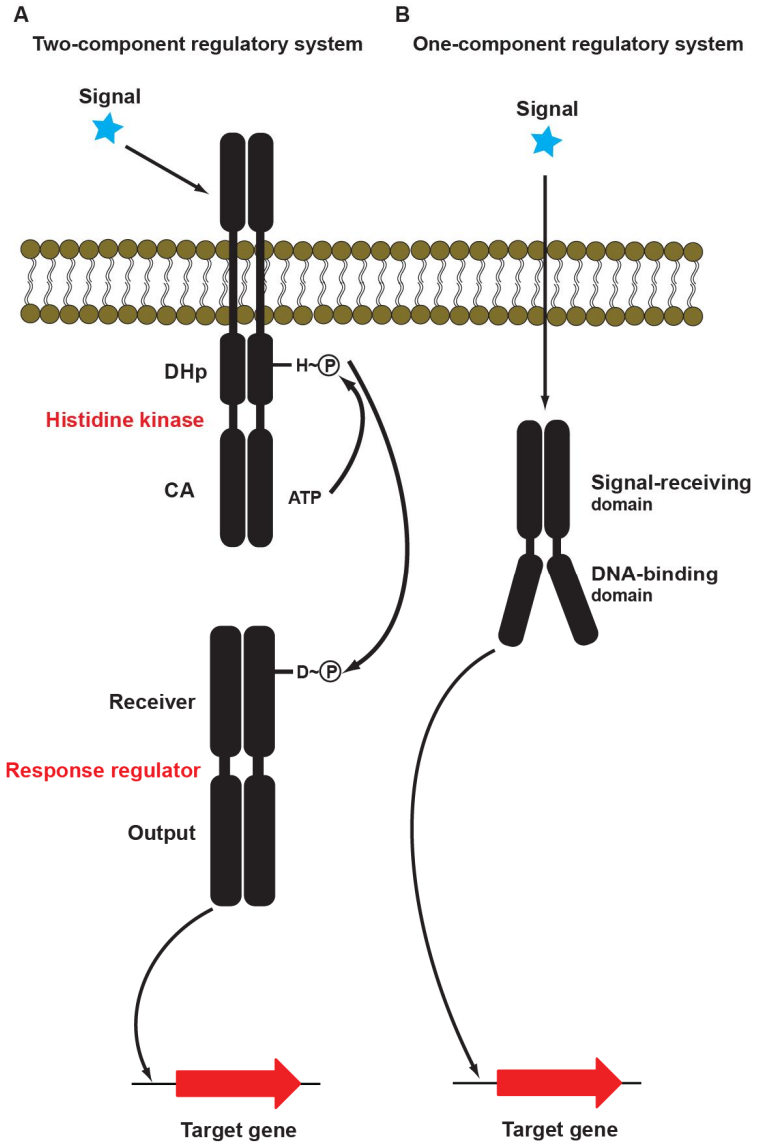

Figure 1.3. The major transcriptional regulatory systems in bacteria. (A) A two-component regulatory system functions through phosphotransfer between two proteins: a histidine sensor kinase responsible for sensing a signal (represented by a blue star) and a response regulator responsible for binding DNA to regulate transcription of its target gene(s). H – histidine; D – aspartate; Ⓟ – phosphoryl group **(B)** In one-component regulatory system, a single protein possesses the signal-receiving domain as well as the DNA-binding domain to detect and transduce an environmental stimulus.

1.2.2 One-component regulatory system

One-component regulators are mainly cytosolic and contain dual functional (signal-receiving and DNA-binding) domains (Figure 1.3B). As a result, a single protein is capable of detecting environmental signals and recognizing DNA for its regulatory activity. One-component transcriptional regulators are similar to two-component regulators in the sense that their major input signals are small molecules [21]: 93% of the known one-component regulators contain various signal-receiving domains that interact with small-molecule ligands, as do 96% of the characterized histidine kinases in two-component systems. Furthermore, many one-component regulators (84%) [21] possess helix-turn-helix (HTH) sequence motifs [3,22] for binding DNA, and this motif is also commonly found in the output domains of response regulators in two-component systems (87%) [21].

The wealth of sequenced genomes available in the public databases has accelerated identification of numerous transcriptional regulators encoded by bacterial genomes. Genome analysis conducted by Ulrich and colleagues in 2005 revealed that one-component regulatory systems are much more common in prokaryotes than two-component systems (17,000 members in one-component system vs. 4,000 members in two-component system) [21], possibly due to the fact that the more simple regulatory mechanism exhibited by one-component systems makes them better suited for rapidly mediating adaptive responses.

Furthermore, one-component regulators possess a more diverse collection of functional domains compared to two-component systems [21]. Of note, all of the known signal-receiving domains in histidine kinases and the DNA-binding domains in response regulators are conserved in various one-component regulators. For example, input and output domains of the two-component system NtrB-NtrC – PAS domain (named after period circadian protein, aryl hydrocarbon receptor, single-minded protein) and HTH motif, respectively – are conserved in the one-component regulator RocR [23,24]. On the other hand, there are a significant number of functional domains exclusively found in one-component systems.

Due to their abundance and diversity, one-component regulatory systems are considered the primordial form of prokaryotic signal transduction [21]. However, biological functions of the majority of these regulatory proteins – in addition to the identity of the small-molecule signals they detect – remain as a significant gap in our understanding of molecular biology. Therefore, investigation of transcription factors is essential for a better understanding of bacterial small-molecule responses, which are closely connected to growth and survival of microbes.

1.3 Bacterial adaptation and resistance to antibiotics

The discovery of antibiotics to treat bacterial infections is considered one of the most significant achievements in the history of medicine; however, the ability of bacteria to adapt and continue to grow in the presence of these antimicrobial small molecules at toxic concentrations has created serious clinical problems [25]. Major resistance mechanisms employed by bacteria for combating antibiotics include enhanced efflux mediated by transporter proteins, enzymatic inactivation of the drugs, and alteration or modification of the antibiotics' cellular targets [26]. Importantly, the fact that expression of the genes conferring antibiotic resistance are often drug-inducible and regulated at the level of transcription [27,28] has generated speculation that transcriptional regulators are major players involved in bacterial responses to antibiotics.

1.3.1 Tetracycline

Tetracycline has been one of the most widely used antibiotics since the 1940s due to the absence of serious side effects and broad spectrum activity [29]. Obtained from *Streptomyces aureofaciens*, this antibiotic has been effective against a wide range of organisms including both Gram-positive and Gram-negative bacteria. Its mode of action involves inhibition of protein synthesis through preventing aminoacyl-tRNA attachment to the ribosomal acceptor site [30].

Tetracycline possesses a structure called a “tetracyclic nucleus” (rings A to D in Figure 1.4A), and different functional groups can be attached to the nucleus to produce derivatives having altered drug efficiencies [29]. The members of the tetracycline group discovered between 1948 and 1963 (i.e. chlortetracycline, oxytetracycline, tetracycline, and demethylchlortetracycline, Figure 1.4B, natural molecules) are considered as “first generation”, followed by the “second generation” members that were identified between 1965 and 1972 (i.e. methacycline, doxycycline, and minocycline, Figure 1.4C, semi-synthetic molecules). More recently, a “third generation” tetracycline (glycylcycline, known as tigecycline, Figure 1.4D, semi-synthetic molecule) has been found effective against organisms resistant to the tetracyclines of the first two generations [31].

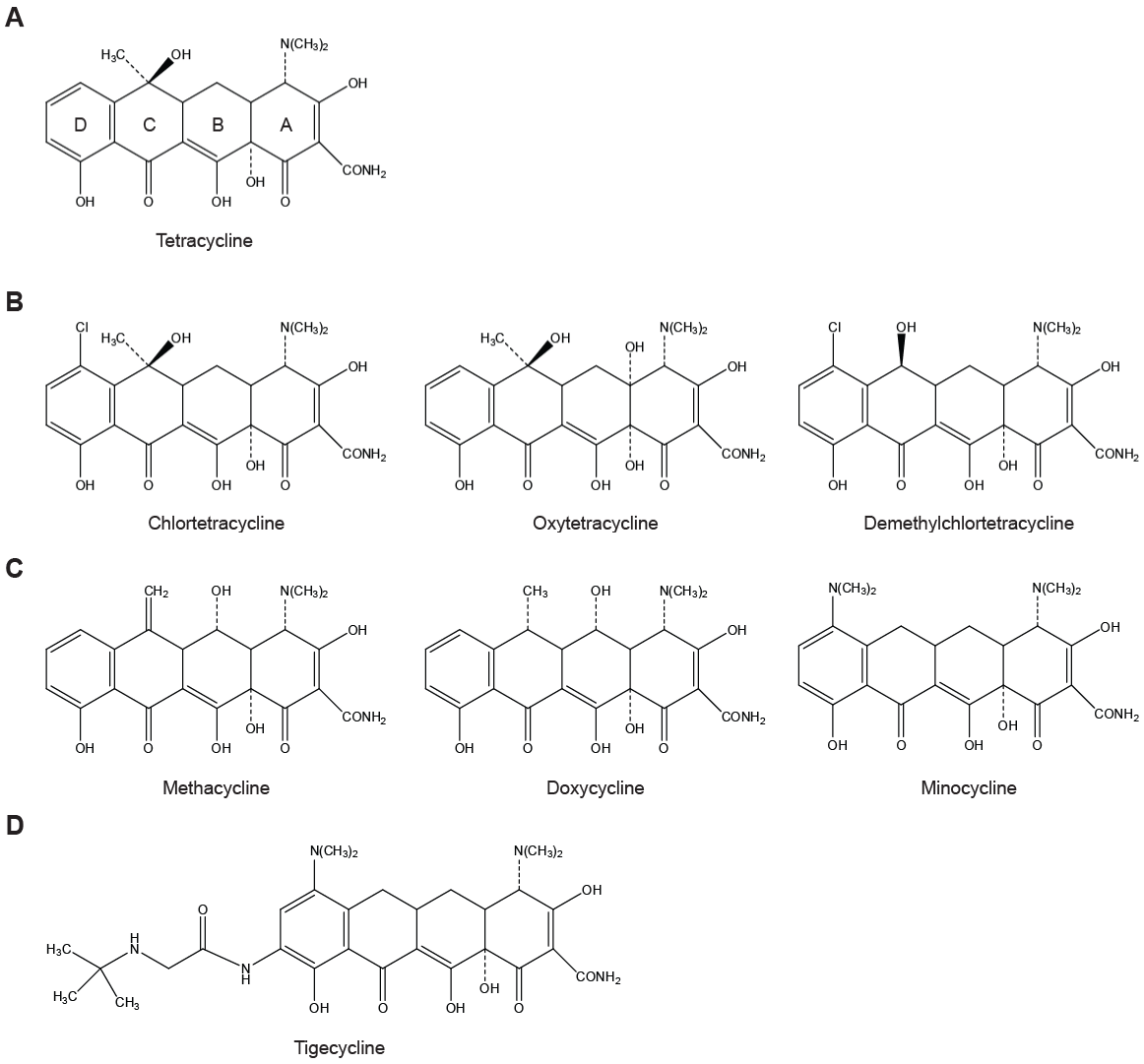

Figure 1.4. Structures of various tetracyclines. (A) Tetracycline and its "tetracyclic nucleus". **(B)** Other first generation tetracyclines. **(C)** Second generation tetracyclines. **(D)** Third generation tetracycline, tigecycline.

Gram-negative bacteria are considered to possess stronger innate resistance to antimicrobial agents than Gram-positive bacteria due to the presence of double (outer and inner) membranes. However, Gram-negative bacteria susceptible to tetracyclines allow rapid influx of the drugs [32] in a partially energy-dependent manner [33]. Tetracyclines first cross the outer membrane through the porin channels – as complexes with positively charged magnesium ions (Mg^{2+}) – into the periplasm, where the complexes probably dissociate to release the uncharged forms of the antibiotics [34]. These weakly liphophilic tetracyclines are able to diffuse through the inner membrane lipid bilayers [32], and uptake of the drugs driven by the proton motive force has been demonstrated with susceptible bacteria [33], allowing the drugs to accumulate in the cytoplasm and interact with ribosomes (as tetracycline-Mg^{2+} complexes) to exert their antimicrobial activity (Of note, it is also assumed that electroneutral tetracyclines enter Gram-positive bacteria [29]).

1.3.2 TetA: efflux-mediated resistance to tetracycline

The early success of tetracyclines against the majority of commensal and pathogenic bacteria continued until the 1950s [35], highlighted by the fact that only 2% of 433 *Enterobacteriaceae* strains collected between 1917 and 1954 were resistant to tetracycline [36]. However, the number of resistant bacterial strains started to increase, with widespread use of the drug being suggested as the primary reason. Heterogeneity of tetracycline resistant determinants was first

examined in *Enterobacteriaceae* and *Pseudomonadaceae* species [37], and 29 distinct tetracycline (*tet*) resistance genes have been characterized as of 2001 [29]. Three mechanisms commonly confer resistance to tetracycline: TetA-like exporters reduce intracellular concentration of tetracyclines [38,39]; TetO-like proteins bind ribosomes to protect them against the drugs [40,41]; and TetX-like proteins are antibiotic-degrading monooxygenases [42,43].

TetA-like efflux pumps are the most frequently found (encoded by 18 of the 29 *tet* resistance genes) and the most well-characterized tetracycline resistance determinants [29]. TetA is a membrane protein with 12 transmembrane α-helices [44], belonging to the major facilitator superfamily (MFS) [45,46]. It is an antiporter that pumps Mg^{2+}-tetracycline complexes out of the cells in exchange for protons [47], and TetA is known to confer the highest resistance to tetracycline among the efflux-mediated determinants [48].

The genes required for antibiotic resistance are often present in conjugative or highly mobile elements, thus facilitating widespread resistance among bacteria [49]. Consistent with this, TetA is encoded by the transposon Tn*10* [38] and as a result, the presence of this gene has been confirmed in various Gram-negative bacteria [29].

1.3.3 TetR: transcriptional regulator of *tetA* expression

The *tetA* gene on Tn*10* is genetically coupled to *tetR* which encodes a one-component transcriptional regulator [50]. The *tetR* gene is divergently oriented to *tetA* (Figure 1.5A), and the intergenic region between the two contains two 15 bp operators (O1 and O2 in Figure 1.5B) that overlap with the promoters of both genes (P_{tetA} for *tetA*, $P_{tetR}1$ and $P_{tetR}2$ for *tetR*) [51,52,53]. In the absence of tetracycline, TetR homodimers tightly bind to these palindromic operator sequences (with an association constant value of 2×10^{11} M^{-1}), preventing RNA polymerase from recognizing and interacting with the promoters. As a result of these protein-DNA interactions, transcriptions of both *tetR* and *tetA* are negatively regulated.

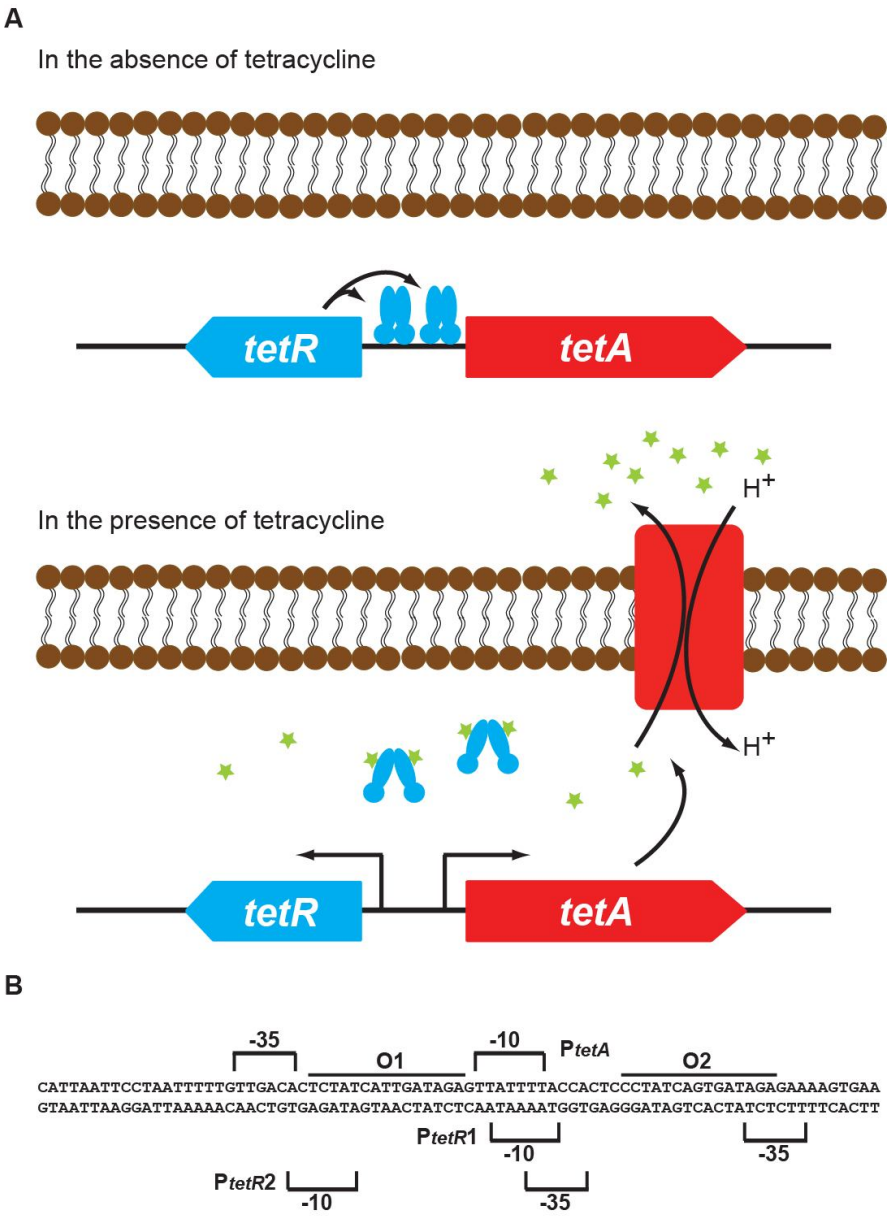

Figure 1.5. The *tetA-tetR* regulon. (A) TetR dimers recognize two operator sequences in the intergenic region and repress expressions of both *tetR* and *tetA* in the absence of tetracycline. Upon binding the tetracycline-Mg^{2+} complex, TetR dissociates from DNA, allowing the production of the TetA tetracycline/H^+ antiporter. **(B)** The intergenic sequence between *tetR* and *tetA* as well as the regulatory elements of the two genes are shown. O1 and O2 represent the operator sequences of TetR. While the *tetA* expression is driven by a single promoter (P_{tetA}), two *tetR* promoters ($P_{tetR}1$ and $P_{tetR}2$) have been identified (-10 and -35 elements for each promoter are indicated above).

Once tetracycline enters the cell, it forms a complex with Mg^{2+} and binds TetR [54,55]. This interaction causes TetR to undergo a conformational change, lowering the repressor's affinity for DNA [56]. As a result, repression is relieved and *tetA* is expressed (Figure 1.5A) [57]. The TetA efflux pump then transports tetracycline out of the cell until the cytoplasmic drug concentration drops to the point where TetR repression can be re-established. The association constant of tetracycline binding to TetR is ~10^9 M^{-1} [55], making TetR one of the most sensitive ligand-inducible systems of transcriptional regulation. The high affinity of TetR for tetracycline is the key for the resistance mechanism as the association constant of the drug to ribosome is 1000-fold lower (~10^6 M^{-1}) [58]. Therefore, repression by TetR is relieved at a lower tetracycline concentration than is required to damage the cells.

As shown in Figure 1.6A, the "Ω-shaped" TetR dimer is composed of two identical monomers, each of which contains 10 α-helices with connecting turns and loops [59]. Each monomer possesses an N-terminal DNA-binding domain constituted by three α-helices (i.e. helices α1, α2, and α3 in one monomer as well as symmetric helices α1', α2', and α'3 in the other) with the helices α2 (and α2') and α3 (and α3') forming a HTH motif (Figure 1.6A). While all three helices are required to maintain the structure of this domain, side chains of the amino acid residues in the helix α3 (and α3') are mainly responsible for sequence-specific interactions with the nucleotides or phosphate backbones of the operators [60]. Two consecutive DNA major grooves which contain the operator sequence act

as a docking site for the dimeric TetR (Figure 1.6B), with the DNA-binding domain of each monomer perpendicularly interacting with half of the palindrome present in a major groove.

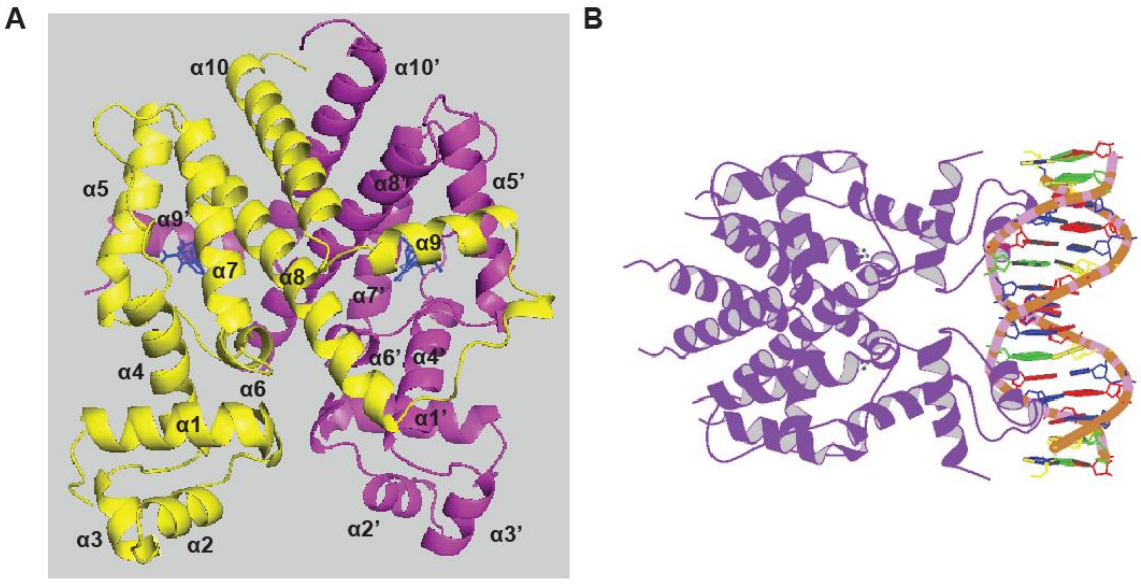

Figure 1.6. Structure of the TetR homodimer. (A) TetR-tetracycline complex. One monomer is represented in yellow while the other one is represented in purple. Two tetracycline-Mg^{2+} complexes are represented in blue. This figure was modified from the structure 1BJY in Protein Data Bank (PDB). **(B)** TetR-DNA complex (1QPI in PDB). The TetR dimer – represented in purple ribbons – binds perpendicularly to the operator DNA.

On the other hand, the helices α5 to α10 and their symmetric counterparts (i.e. α5' to α10') from a core regulatory domain responsible for dimerization and tetracycline binding, while the helix α4 acts as the linker to connect the two functional domains (i.e. DNA-binding and ligand-binding domains) [56,59]. The two monomers possess an identical ligand-binding pocket, meaning that one TetR dimer is capable of binding two tetracycline-Mg^{2+} complexes (Figure 1.6A).

Once tetracycline complexes enter the binding cavities they cause serial structural changes in TetR that result in the “pendulum-like” movement of the helix α4 (and α4’) [56]. As a consequence, the distance between the helices α3 and α3’ increases to a point where the TetR conformation is not suited for interacting with successive major grooves of the DNA, thus releasing TetR from the operator.

While this induction mechanism has been widely accepted, a recent study has suggested that the main effect of tetracycline binding is not to change the TetR structure from one defined conformation to another [61]. Instead, it is proposed that the N-terminal domains of apo-TetR are flexible and thus, they can freely access the conformation capable of binding DNA. On the other hand, interaction with tetracycline reduces this N-terminal flexibility and traps the domain conformation so it is incapable of interacting with DNA.

1.4 TetR family transcriptional regulators (TFRs)

1.4.1 Distribution of TFRs in bacteria

Bacterial transcriptional regulators are grouped in various families based on their sequence similarity and structural/functional criteria. The TetR family is one of >10 groups of transcription factors found in bacteria (Table 1.1), and it has been named after its most genetically and biochemically well investigated member, TetR. TetR family transcriptional regulators (TFRs) are characterized by and easily recognized through the high sequence conservation in their N-terminal DNA-binding domains [62]. These proteins are widely distributed in bacteria as at least one TFR is encoded by more than 80% of sequenced bacterial genomes [63], and they constitute one of the largest groups of transcription factors in the public databases [64]. The number of TFRs encoded in genome databases exceeded 20,000 distinct sequences in 2010 [65] and has continued to grow.

Table 1.1. Major families of bacterial transcriptional regulators

Family[a]	Action[b]	Structural motif	DBD position[c]	Some regulated functions	Reference
LysR	D	HTH	N	Carbon and nitrogen metabolism	[66]
NtrC/Fis	A	HTH	C	Nitrogen assimilation, aromatic amino acid synthesis, flagella, catabolic pathways, phage response, etc.	[67]
TetR	R	HTH	N	Biosynthesis of antibiotics, efflux pumps, osmotic stress, etc	[62]
AraC/XylS	A	HTH	C	Carbon metabolism, stress response and Pathogenesis	[68]
OmpR	A	Winged-helix	C	Heavy metal and virulence	[69]
LuxR	A	HTH	C	Quorum sensing, biosynthesis and metabolism, etc	[70]
GntR	R	HTH	N	General metabolism	[71]
MarR	D	HTH	C	Multiple antibiotic resistance	[72]
MerR	R	HTH	N	Resistance and detoxification	[73]
LacI	R	HTH	N	Carbon source utilization	[74]
ArsR	R	HTH	C	Metal resistance	[75]
AsnC	D	HTH	N	Amino acid biosynthesis	[76]
DeoR	R	HTH	N	Sugar metabolism	[77]
Cold shock domain	A	RNA-binding like	V	Low-temperature resistance	[78]
IclR	R	HTH	N	Carbon metabolism, efflux pumps	[79]
CRP-FNR	A/D	HTH	C	Global responses, catabolite repression	[80]

This table was modified from the reference [62].

[a] Families have been placed in the order of decreasing numbers of the members identified as of 2006 [64].

[b] Transcriptional regulatory function of the majority of the known transcription factors (D → dual repressor/activator activity; A → activator; R → repressor).

[c] N- or C-terminal position of DNA-binding domains (DBD) in the sequence regulators (N → DBD at N-terminus; C → DBD at C-terminus; V → variable).

Our analysis of 4243 TFRs encoded by the genomes of 211 bacterial and archaeal strains (including representatives of pathogenic, commensal and environmental organisms from all 30 bacterial phyla) suggests that TFRs are abundant in Actinobacteria and other soil-dwelling organisms whereas some pathogens do not encode any TFR (Cuthbertson *et al*, manuscript submitted). This trend is consistent with the observation that bacteria living in a more variable environment tend to have large genomes enriched in transcriptional regulators compared to intracellular pathogens [81].

For example, the genus *Streptomyces* (of the phylum Actinobacteria) is known for the production of clinically-useful secondary metabolites [82] and its unusual lifecycle that displays differentiated cell types [83]. Whole genomes of more than 40 streptomycetes have been sequenced to date and, possibly consistent with the organisms' complicated lifestyle, they contain large numbers of genes encoding regulatory proteins. Each of the sequenced genomes, with a few exceptions, encodes more than 100 TFRs and these numbers are high compared to the model organisms in other phyla such as *Pseudomonas aeruginosa* (40 TFRs), *B. subtilis* (18 TFRs), and *E. coli* (13 TFRs) (Figure 1.7A). In addition to their abundance, TFRs in *Streptomyces* well represent the great sequence diversity that potentially exists in the TetR family, suggested by the fact that each streptomycete has a considerable number of TFRs that are either species-specific or potential orthologs shared by multiple organisms (Figure 1.7B).

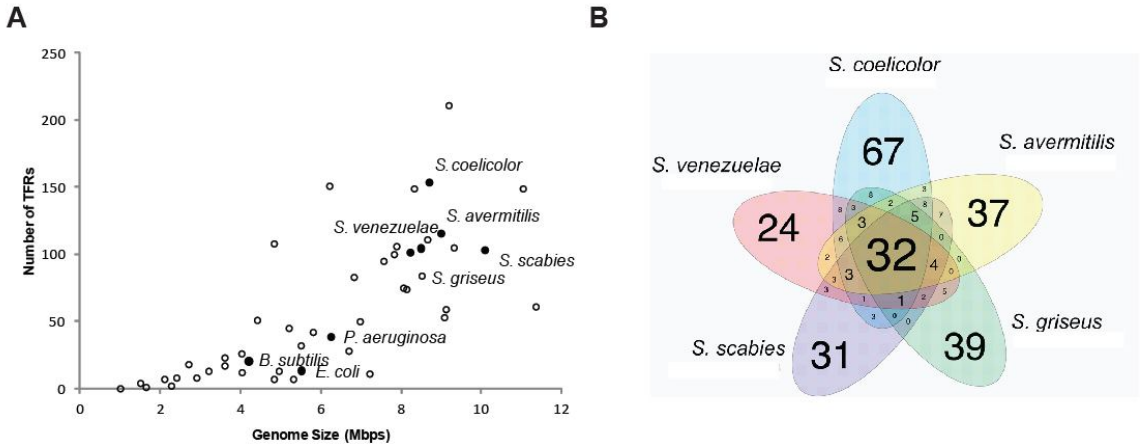

Figure 1.7. Abundance and diversity of TFRs in streptomycetes. (A) The graph shows the genome sizes of various bacteria (on x-axis) and the numbers of TFRs they encode (on y-axis). Black dots represent five streptomycetes (*Streptomyces coelicolor*, *Streptomyces avermitilis*, *Streptomyces venezuelae*, *Streptomyces griseus*, and *Streptomyces scabies*) and three other model organisms (*P. aeruginosa*, *B. subtilis*, and *E. coli*). **(B)** Amino acid sequences of the TFRs in the five aforementioned streptomycetes were analyzed to search for potential ortholgs shared by two to five organisms (Cuthbertson, unpublished). While 32 TFRs are shared by all five streptomycetes as indicated at the center of the Venn diagram, the values indicated at the corners indicate that each streptomycete possesses a considerable number of species-specific TFRs.

1.4.2 General features of TFRs

Despite the large number and ubiquity of TFR-coding genes in bacterial genomes, only a small fraction of them (<120 TFRs) have been investigated in any detail. All TFRs whose structures are available in the public databases show similar overall architecture – including all-helical, homodimeric configuration – in solution [63]. Consistent with their high sequence similarity, the N-terminal DNA-binding domains (helices α1 to α3 and their symmetric counterparts) are the most structurally conserved region among TFRs.

The majority of the characterized TFRs are repressors and these proteins bind to operator sequences to prevent transcription initiation by the RNA polymerase holoenzyme. Antibiotic resistance has been commonly viewed as the major biological function regulated by TFRs (e.g. tetracycline resistance for TetR). In fact, a large number of known TFRs modulate expression of genes encoding efflux pumps that confer resistance to various antibiotics [84,85,86,87,88,89], and regulation of an antibiotic-degrading enzyme has been also demonstrated [90]. However, the extent of prokaryotic physiology under the control of these transcription factors is not clear, as TFRs have been also implicated in the regulation of other processes including primary metabolism such as fatty acid synthesis or degradation [91,92], antibiotic biosynthesis or activation [93,94,95], biofilm formation [96], toxin production [97], and cell-cell signaling [98].

Another trend in the TetR family is their responsiveness to small-molecule ligands. Unlike the N-terminal DNA-binding domains, sequences of the C-terminal domains, which in many of the characterized TFRs interact with ligands, are highly variable, suggesting that this family can respond to a diverse range of chemical stimuli. This finding is consistent with the fact that the majority of these proteins have been linked to the physiological processes unrelated to tetracycline. Ligands have been characterized for <60 TFRs to date, and they include structurally diverse molecules such as drugs [50,89,99], metabolites [91,100,101], metals [102,103], and dyes [104,105]. It is noteworthy that some TFRs have strict ligand-binding specificities (e.g. tetracycline and structurally-related derivatives

for TetR) [29,50], while others bind a more broad spectrum of molecules (e.g. diverse cationic lipophilic drugs such as rhodamine 6G, crystal violet, and ethidium for QacR) [106].

1.4.3 Examples of known TFRs: similarities and differences to TetR

There are three key features that characterize the regulatory activity of TetR on *tetA* expression, and these "TetR rules" include 1) binding as homodimers to the operators that overlap with the target promoters; 2) repressing its own gene as well as the divergently oriented target gene; and 3) interacting directly with small-molecule signals for mediating cellular responses. While the most extensively studied TetR has been regarded as the "model" protein, how widely its regulatory mechanism is applicable to the other members of the family is not clear. Here, we describe a few of the previously characterized TFRs displaying noticeable behaviors that are consistent with and/or deviated from the "TetR rules" (Of note, ligands and/or autoregulatory functions of these TFRs will be mentioned only if they have been characterized).

SimR: a regulatory mechanism similar to TetR

SimR from *Streptomyces antibioticus* displays a highly similar regulatory activity to TetR. The divergent genetic configuration of *simR* and its regulatory target gene *simX* – encoding an MFS efflux pump that confers resistance to the antibiotic simocyclinone – resembles that of the *tetR*/*tetA* pair [85]. SimR binds to two operators in the *simR*/*simX* intergenic region, and these DNA sequence elements (each of them speculated to be 17 bp long) overlap with the target promoters (like TetR) such that the SimR-operator interactions repress expression of both *simR* and *simX*. This repression is relieved in the presence of simocyclinone and its biosynthetic intermediate [85], and structures showing the direct interactions between SimR and its ligands have been solved [107]. One unusual feature of this TFR is the role of its arginine-rich N-terminal extension residues that precede the helix α1. While SimR primarily interacts with the operator DNA through its HTH motif (i.e. helix α2 and helix α3), the positively charged arginine residues in the N-terminal extension have been suggested to assist in DNA binding through interacting with the minor groove of the DNA [108].

QacR and EthR: cooperative DNA-binding properties

Along with TetR, QacR from *Staphylococcus aureus* is one of the most extensively studied TFRs to date. Like *tetR*, *qacR* is divergently oriented to the target gene of its repressor activity, *qacA*, which encodes a membrane-bound transporter [87]. Interestingly, QacR has been shown to interact with a more diverse collection of structurally-unrelated small molecules [106] than TetR. This is possibly consistent with the fact that its target gene product QacA functions as a multidrug efflux pump [87], in contrast to TetA which is believed to be specific for tetracyclines. In addition to this ligand-binding aspect, there are two significant DNA-binding behaviors that distinguish QacR from TetR.

First of all, the perfectly palindromic operator of QacR is considerably long – 28 bp in length (Figure 1.8A) – compared to the TetR operators (15 bp each, Figure 1.5B). Interestingly, this operator has two internally overlapping, weaker palindromic sequences called “nested palindromes” (Figure 1.8A) [109]. It has been observed that two sets of the QacR dimers bind to this DNA region cooperatively (Figure 1.8B), and binding of the first dimer on one nested palindrome is hypothesized to change the DNA conformation such that the second dimer can more easily access the other partially palindromic sequence [110].

.

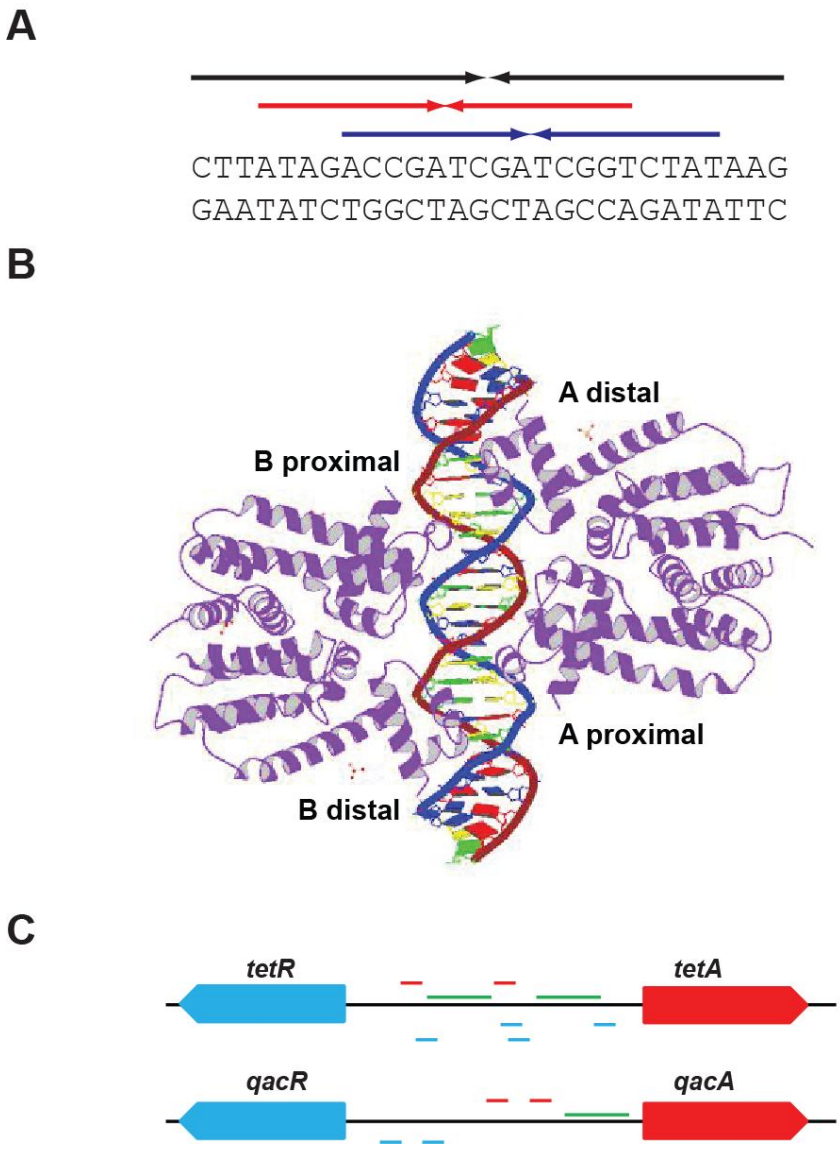

Figure 1.8. QacR and its interaction with the operator. (A) The QacR operator is a long palindromic sequence (indicated by black arrows) containing two nested partial palindromes (indicated by red and blue arrows). **(B)** Two sets of the QacR homodimers (designated as A and B) interact with the operator cooperatively. Each monomer was designated as proximal or distal depending on their locations with respect to the palindromic center. This figure was modified from the structure 1JT0 in PDB. **(C)** TetR (top) binds to two operator sequences (indicated by green lines) which overlap with the promoters of *tetR* (-10 and -35 elements indicated by blue lines) and *tetA* (red lines). On the other hand, the QacR (bottom) operator is located close to *qacA* (green line) and the QacR binding only represses activity of the *qacA* promoter (red lines) without affecting the *qacR* promoter (blue lines).

Secondly, the QacR operator is located downstream of the *qacA* promoter elements, and it overlaps with the transcriptional start site of *qacA* (Figure 1.8C) [87,110]. Therefore, this TFR has been suggested to repress transcription by blocking the transition of the RNA polymerase-promoter complex into the active transcribing state rather than by physically preventing RNA polymerase from recognizing and interacting with the promoter. Of note, there is only one QacR operator sequence in the *qacR*/*qacA* intergenic region (Figure 1.8C). It is positioned proximal to *qacA* [111] and the observation that QacR does not regulate its own expression is another major deviation from the TetR paradigm [87].

EthR from *Mycobacterium tuberculosis* is another TFR with an uncommon DNA-binding property. This TFR represses expression of the divergently transcribed *ethA* gene, encoding an enzyme required for activation of the pro-drug ethionamide [95]. Surprisingly, previous studies have indicated that eight EthR molecules (likely four EthR homodimers) interact cooperatively with the 55 bp operator in the *ethR*/*ethA* intergenic region [95], while it forms a dimer in solution just like other TFRs [112]. A structural analysis has shown that a small molecule hexadecyl octanoate is capable of interacting with EthR and lowering the protein's affinity for the target DNA [113].

AcnR and CmeR: regulation of targets transcribed in the same direction

While QacR and EthR possess unique DNA-binding properties, they are similar to TetR in the sense that both of them repress divergently oriented target genes. However, some TFRs repress genes which are transcribed in the same direction as their own genes.

For example, AcnR in *Corynebacterium glutamicum* regulates the *acn* gene – encoding the aconitase enzyme involved in the tricarboxylic acid cycle – located upstream of its own gene (Figure 1.9A) [114]. These two genes are co-transcribed and AcnR represses expression of both genes by interacting with a 16 bp palindromic operator overlapping the target promoter located upstream of *acn* (Figure 1.9A).

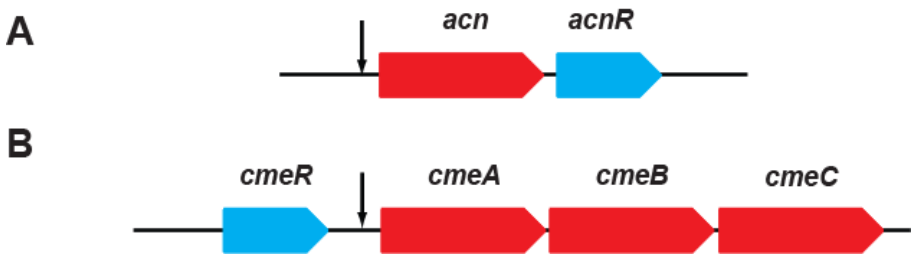

Figure 1.9. AcnR/CmeR and their target genes. The *acnR* gene is located downstream of the co-transcribed *acn* gene **(A)**, while the *cmeR* gene is located upstream of the target *cmeABC* operon **(B)**. The TFR-coding genes are represented in blue while their target genes are represented in red. Arrows indicate locations of the operators.

On the other hand, CmeR represses expression of the *cmeABC* operon that encodes the components required to form a multidrug efflux system in the resistance-nodulation-division (RND) family in *Campylobacter jejuni* [115,116]. As shown in Figure 1.9B, the *cmeR* gene is positioned upstream of the operon in the same genetic orientation. However, these four genes are not co-transcribed and this TFR regulates only *cmeABC* expression by interacting with a 16 bp palindromic operator present in the intergenic region between *cmeR* and *cmeA* (the binding site overlaps the *cmeABC* promoter) [115,117]. The CmeABC system confers bile resistance [118], and CmeR responds directly to bile salts such as taurocholate and cholate [117].

DhaS: activator with unusual ligand-detecting properties

As previously stated, the majority of the known TFRs exhibit their biological roles by repressing the expression of their target genes. However, a small number of TFRs act as activators [93,119,120] or dual repressor/activators [98,121].

DhaS from *Lactococcus lactis* is among the transcription-activating TFRs [119]; however, what makes this regulator even more unique is its indirect response to a small-molecule ligand. The DhaS-coding gene is divergently oriented to the target operon encoding subunits of dihydroxyacetone kinases (DhaK, DhaL, and DhaM) (Figure 1.10). Upstream of *dhaS* is the *dhaQ* gene that is transcribed in the opposite direction. When dihydroxyacetone is present in the

cells, this molecule first forms a complex with DhaQ, which in turn interacts directly with DhaS. DhaS – responding indirectly to the ligand through a cofactor protein DhaQ – then binds DNA upstream of *dhaK* (palindromic operator estimated to be 22 bp long) and activates expression of the *dhaKLM* operon.

Figure 1.10. Genetic configuration of *dhaS* and its surrounding genes.

AtrA: regulation of distant target genes

Although the regulatory behaviors exhibited by the aforementioned TFRs are diverse, all of them share one common characteristic with TetR: each of these TFRs regulates the target gene (or operon) that is adjacent to its own gene. However, regulatory targets are not limited to local genes for some TFRs. In fact, several TFRs have been characterized as global regulators that target multiple distant genes for controlling various physiological processes [94,100,122].

AtrA (locus tag SCO4118) is one of the most widely studied TFRs from *Streptomyces coelicolor* due to its effect on the biosynthesis of the antibiotic actinorhodin [93]. The *atrA* gene is transcribed in the opposite direction to the neighboring gene *SCO4119*, but possible local activity of this TFR has not been previously investigated (determined in Chapter 2). Instead, AtrA was first shown to activate expression of the *act*II-ORF4 gene (*SCO5085*) [93] encoding another regulatory protein that increases expression levels of multiple target genes in the

actinorhodin biosynthetic gene cluster [123]. More recently, AtrA was implicated in the regulation of an additional gene, *nagE2* (*SCO2907*), required for nital-acetylglucosamine transport [124], suggesting that the AtrA regulon might include many more distant target genes which have not yet been characterized.

1.4.4 Proposed paradigm: TFRs interact with small molecules that are functionally related to their target gene products

While our knowledge of TFRs is limited, the examples described in the previous section suggest that the "TetR rules" are not applicable to the entire family. However, we still believe that careful analysis of previously studied TFRs offers promise for the potential development of techniques or knowledge that are greatly beneficial to investigate a large number of TFRs of unknown functions.

For example, we noticed that many TFRs interact with small molecules that are structurally identical or related to the substrates used by their target gene products (Figure 1.11). In the case of TetR and TetA, tetracycline (and structurally-related derivatives) acts as the ligand/substrate for both proteins, and similar relationships have been demonstrated between other TFRs (including previously described SimR and CmeR) and their target gene products [85,89,90,99,117,125,126].

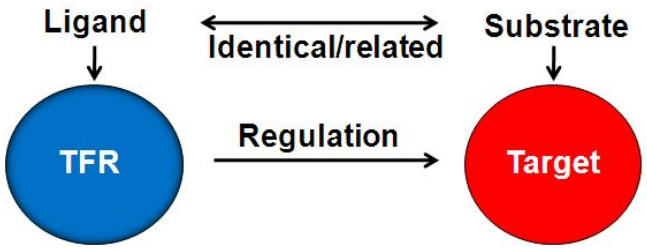

Figure 1.11. Proposed functional relationships between ligands and regulatory target gene products of TFRs.

Of significant importance, this emerging paradigm has led us to propose that identification of activating ligands provides a powerful clue concerning the biological function of uncharacterized TFRs and their target gene products. To facilitate this endeavor, our lab has recently created a relational framework, using phylogenetic methods, to classify >4200 TFRs (including 56 TFRs with known ligands and 170 TFRs whose ligands can be predicted based on the knowledge we have regarding these proteins) according to their whole sequence similarities (Cuthbertson *et al*, manuscript submitted). Our evidence suggests that this framework describes and organizes the TFR sequence diversity that exists in the current genome databases. More importantly, it provides testable predictions concerning ligands of hundreds of previously uncharacterized TFRs based on their sequence similarities to the TFRs with known or predicted ligands.

As a proof-of-principle, our framework showed that SCO7719 (later named KijR) from *S. coelicolor* is found in the same phylogenetic subgroup as KijA8 from *Actinomadura kijaniata* (Figure 1.12). The *kijA8* gene is located within the biosynthetic gene cluster for the antibiotic kijanimicin [127], and previous studies have suggested that TFRs encoded within antibiotic biosynthesis clusters interact with products of the cognate biochemical pathways [85,99,128]. We have experimentally observed the induction of KijA8 by kijanimicin and importantly, this has led us to correctly predict that KijR also binds kijanimicin based on its sequence similarity to KijA8 (Cuthbertson *et al*, manuscript submitted). Furthermore, we have subsequently shown that KijR's target gene *SCO7720* (later named *kijX*) confers resistance to kijanimicin by enzymatic degradation, further supporting the proposed functional relationship between small-molecule ligands and regulatory target gene products of TFRs.

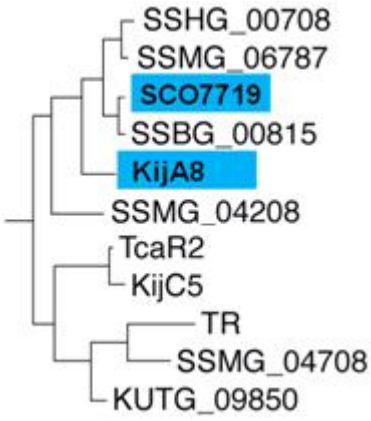

Figure 1.12. Phylogenetic group containing SCO7719 (KijR) and KijA8. SCO7719 is potentially related to TFRs with known or predicted ligands such as TcaR2 from *Micromonospora chalcea*, KijA8 and KijC5 from *A. kijaniata*. This group also includes TFRs of unknown functions such as SSHG_00708 from *Streptomyces albus*; SSMG_06787, SSMG_04208, and SSMG04708 from *Streptomyces* sp. AA4; SSBG00815 from *Streptomyces* sp. SPB74; and KUTG_09850 from *Kutzneria* sp. 744. This figure was modified from Cuthbertson *et al* (manuscript submitted).

1.5 Objectives

Despite our limited knowledge of TFRs, it is highly anticipated that investigating this family of transcription factors will provide us with a better understanding of small-molecule responses required for bacterial growth and survival. Moreover, many of the known TFRs are involved in clinically important processes such as antibiotic resistance and biosynthesis and thus, it is expected that this family will attract considerable attention from research groups in the foreseeable future. While interest in TFRs has gradually increased over the years, all but a few studies have focused on individual proteins. However, given the large number of TFRs present in bacteria, it is highly desirable to develop methodologies that are generally applicable to many members of the family.

We have suggested that activating ligands and regulatory targets of TFRs are functionally related, and we propose that this paradigm should be central in studying these transcription factors. While the number of TFRs encoded in genome databases exceeds 50,000 distinct sequences to date (Nodwell, unpublished), cognate ligands and target genes are unknown for the majority of these proteins. Therefore, the main focus of this study is to develop widely applicable tools for identifying these two basic elements of the biological roles of TFRs. We first investigate the use of genome context as a predictive tool for TFR target gene identification. Secondly, we focus on developing a whole cell-based reporter system that is amenable to high-throughput screening as a means of identifying small-molecule ligands of TFRs.

CHAPTER 2: GENOME CONTEXT AS A PREDICTIVE TOOL FOR IDENTIFYING REGULATORY TARGETS OF TETR FAMILY TRANSCRIPTIONAL REGULATORS

This chapter was adapted from:

Ahn, S.K., Cuthbertson, L., and Nodwell, J.R. 2012. Genome context as a predictive tool for identifying regulatory targets of TetR family transcriptional regulators. Manuscript submitted to PLoS ONE.

Author contributions:

All experiments were conducted by SKA. Proteins and DNA probes used for *in vitro* assays were prepared by SKA and LC.

2.1 Abstract

Although the number of TetR family transcriptional regulators (TFRs) in the public databases has grown exponentially as a result of high throughput genome sequencing, virtually nothing is known about the thousands of these proteins. Generally applicable methods for predicting their regulatory targets would assist efforts to characterize the family. Here, we investigate the chromosomal context of 372 TFRs from three *Streptomyces* species. We find that the majority (250 TFRs) are transcribed divergently from one neighboring gene, as is the case for TetR and its target *tetA*. We explore putative target gene identity and intergenic separation to see if either is predictive of a regulatory relationship. While intergenic separation is a critical factor in regulatory prediction, the identity of the putative target gene product is not. Our data suggest that those TFRs that are <200 bp from their divergently oriented neighbors will most likely regulate them. These target genes include membrane proteins (26%; 22% are predicted membrane-associated pumps), enzymes (60%), other proteins such as transcriptional regulators (1%), and proteins having no predictive function (13%). Our analysis greatly expands the diversity of biochemical functions under the control of this family of transcriptional regulators.

2.2 Introduction

Investigating transcriptional regulators and biological processes they regulate is essential for our understanding of cellular physiology. TetR family transcriptional regulators (TFRs) constitute one of the largest groups of bacterial transcriptional regulators and thus, a generally applicable tool for identifying genes under the control of TFRs would greatly accelerate our ability to assign functions to this important family of transcriptional regulators.

Like TetR, the majority of the characterized TFRs are repressors and many of them regulate genes encoding efflux pumps that confer resistance to various antimicrobial agents (Table 2.1). While TFRs involved in other processes in prokaryotic physiology have been identified (described in Chapter 1), efflux-mediated antibiotic resistance is still viewed as the major functional target of TFRs. Therefore, it is desirable to know whether this perception is accurate for the remaining TFRs of unknown functions or whether there is a greater diversity of biological and biochemical functions under TFR regulation.

Table 2.1. Examples of TFRs regulating efflux pumps

TFR	Organism	Target efflux pump (type)[a]	Orientation[b]	Reference
AcrR	*Escherichia coli*	AcrAB (RND)	D	[84]
ActR	*Streptomyces coelicolor*	ActAB (MFS, RND)	D	[123]
AmeR	*Agrobacterium tumefaciens*	AmeABC (RND)	S	[129]
ArpR	*Pseudomonas putida*	ArpABC (RND)	D	[130]
BpeR	*Brucella suis*	BpeDE (RND)	D	[131]
BpeR	*Burkholderia pseudomallei*	BpeAB-OprB (RND)	D	[132]
BreR	*Vibrio cholerae*	BreAB (RND)	D	[133]
BrtA	*Listeria monocytogenes*	MdrT (MFS)	S	[134]
CgmR	*Corynebacterium glutamicum*	CgmA (MFS)	S	[105]
CmeR	*Campylobacter jejuni*	CmeABC (RND)	S	[115]
EbrR	*Streptomyces lividans*	EbrA (SMR)	S	[135]
EbrS	*S. lividans*	EbrC (MFS)	D	[136]
EnvR	*E. coli*	AcrEF (RND)	D	[137]
		AcrAB (RND, much higher repressor activity on *acrAB*)	N/A	
EpeR	*Streptomyces clavuligerus*	EpeA (MFS)	S	[138]
FrrA	*Bradyrhizobium japonicum*	FreCAB (RND)	D	[139]
HrtR	*Lactococcus lactis*	HrtBA (ABC)	S	[140]
LanK	*Streptomyces cyanogenus*	LanJ (MFS)	D	[99]
LfrR	*Mycobacterium smegmatis*	LfrA (MFS)	S	[141]
LmrA	*Vibrio harveyi*	LmrB (MFS)	S	[142]
MepR	*P. putida*	MepA (probable linker protein for an RND system)	D	[143]
MexL	*Pseudomonas aeruginosa*	MexJK (RND)	D	[144]
MexZ	*P. aeruginosa*	MexXY (RND)	D	[145]
MtrR	*Neisseria gonorrhoeae*	MtrCDE (RND)	D	[146]
NalD	*P. aeruginosa*	MexAB-OprM (RND)	N/A	[147]
NfxB	*P. aeruginosa*	MexCD-OprJ (RND)	D	[86]
PigZ	*Serratia* sp. ATCC 39006	ZrpADBC (RND)	D	[148]
Pip	*S. coelicolor*	Pep (MFS)	S	[89]
PqrA	*S. coelicolor*	PqrB (MFS)	S	[149]
RmrR	*Rhizobium etli*	RmrAB (RND, MFS)	D	[150]
SimR	*Streptomyces antibioticus*	SimX (MFS)	D	[85]
SmeT	*Stenotrophomonas maltophilia*	SmeDEF (RND)	D	[88]
SrpR	*P. putida*	SprABC (RND)[c]	D*[d]	[151]
TetR	*E. coli*	TetA (MFS)	D	[50]
TtgR	*P. putida*	TtgABC (RND)	D	[152]
VceR	*Vibrio cholerae*	VceABC (RND)	D	[153]

[a] ABC: ATP-binding cassette; MFS: major facilitator superfamily; RND: Resistance-nodulation-division; SMR: small multidrug resistance.
[b] Most of the genes encoding TFRs are transcribed divergently from (D) or in the same direction (S) as their neighboring target genes On the other hand, a few genes are located more distantly from their targets (N/A).
[c] SrpR affects the *sprABC* expression by acting as an antirepressor for the SprS repressor.
[d] The *sprABC* operon is divergently oriented to *sprS* which is co-transcribed with its downstream gene *sprR*.

Importantly, we have noticed that the genes encoding many of the known TFRs tend to be adjacent to the genes they regulate (either as divergently transcribed genes or in operons, including the examples shown in Table 2.1), as is the case for *tetR* and *tetA*. In our view, it is highly beneficial to know if this local regulatory activity is widely conserved among TFRs of unknown functions, as the availability of genome sequences has allowed us to easily examine genomic surroundings of TFR-coding genes.

In this work, we have identified 372 genes encoding TFRs in three streptomycetes – *S. coelicolor, Streptomyces avermitilis*, and *Streptomyces griseus.* We have explored the genome context of these genes and find that most are encoded divergently to a neighboring gene. We have tested the prediction that these TFRs regulate the divergently encoded neighboring genes, and our data suggest that this is likely the case for most or all TFRs where the intergenic separation is less than 200 bp. This is true regardless of the nature of the target gene product. In addition to confirming that the TetR regulatory paradigm holds for a majority of TFRs, our analysis demonstrates a far greater diversity of putative TFR targets than previously appreciated. While no more than 25% of these proteins control the expression of membrane-associated pumps, the majority of regulatory targets encode all six classes of characterized enzymes.

2.3 Results

2.3.1 Most TFRs are divergently oriented to an adjacent gene

We searched the genomes of three *Streptomyces* – *S. coelicolor*, *S. griseus* and *S. avermitilis* – for genes encoding putative TFRs and identified 153, 104, and 115 of them respectively (total of 372 TFRs in the three streptomycetes) based on a high score for the consensus sequence (protein family PF00440) of the TFR N-terminal DNA-binding domain (TetR_N). Actinomycete chromosomes are linear and share a conserved genetic 'core' region and more variable 'arm' regions at both ends, containing primarily non-essential species-specific genes including many involved in secondary metabolism [154]. The TFR genes in these streptomycetes are distributed evenly throughout the chromosomes with a slight enrichment in the 'core' relative to the 'arm' regions. For example, *S. coelicolor* has 93 TFRs in the 'core' (4.9 Mb, approximately 19 TFRs/Mb), 27 TFRs in the left 'arm' (1.5 Mb, 18 TFRs/Mb) and 30 TFRs in the right 'arm' (2.3 Mb, 13 TFRs/Mb). In addition, *S. coelicolor* contains the SCP1 plasmid (356 kb), which includes three more TFRs.

Given the model TetR/TetA regulatory paradigm, we predicted that most of these TFRs will regulate the expression of adjacent genes. We examined the genome context of the individual TFRs and divided them into three groups according to their orientation relative to neighboring genes. As shown in Figure 2.1A, one group is divergently oriented relative to a neighboring gene, like TetR. A second group (Figure 2.1B) is likely to be co-transcribed with an upstream or downstream neighbor. A small number of TFRs (eight in *S. coelicolor*, four each in *S. griseus* and *S. avermitilis*) have a divergent neighbor on one side and a probable co-transcribed neighbor on the other (included in the first group in Figure 2.1). The remaining TFRs have neither of these relationships with the neighbors (Figure 2.1C). TFRs oriented divergently to their neighboring genes are most common in all three streptomycetes examined and comprise 67% (250 TFRs) of the total TFRs, while 15% (55 TFRs) and 18% (67 TFRs) of the TFRs are in the second and third groups, respectively.

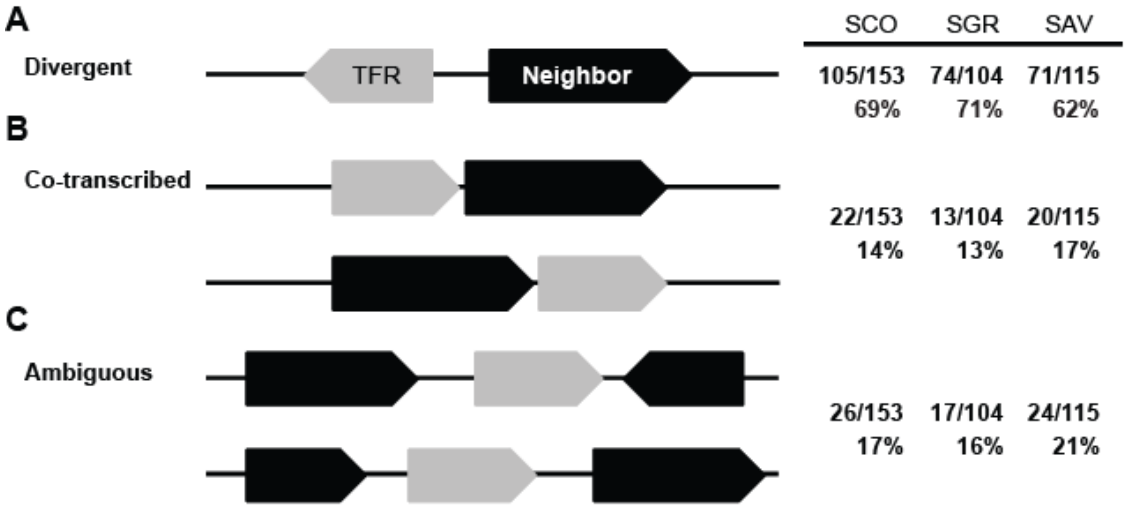

Figure 2.1. Classification of TFRs according to their relative orientation to the neighboring genes. 372 TFRs in *S. coelicolor* (SCO, 153 TFRs), *S. griseus* (SGR, 104 TFRs), and *S. avermitilis* (SAV, 115 TFRs) were divided into three groups according to their genome context to neighbors. **(A)** 250 TFRs (105 in SCO, 74 in SGR, 71 in SAV) are encoded divergently to their neighbors. **(B)** 55 TFRs (22 in SCO, 13 in SGR, 20 in SAV) are likely co-transcribed with their upstream or downstream genes and the intergenic DNAs separating them are ≤35 bp. **(C)** 67 TFRs (26 in SCO, 17 in SGR, 24 in SAV) show neither of the two aforementioned orientations.

We investigated the TFRs of four model organisms at various phylogenetic distances from *Streptomyces* (Actinobacteria, Gram-positive bacteria with ~70% GC content in their DNA) – *Mycobacterium tuberculosis* H37Rv (Actinobacteria, Gram-positive and 65.6% GC content, 49 TFRs), *Bacillus subtilis* subsp. subtilis str. 168 (Firmicutes, Gram-positive and 43.5% GC content, 18 TFRs), *P. aeruginosa* PAO1 (Gammaproteobacteria, Pseudomonadaceae, Gram-negative and 66.6% GC content, 40 TFRs), and *E. coli* str. K-12 MG1655 (Gammaproteobacteria, Enterobacteriaceae, Gram-negative and 50.8% GC content, 13 TFRs). In correlation with our analysis of the TFRs in the three

streptomycetes, the divergent orientation is most frequent in these organisms, although it is less dominant in *B. subtilis* (9 TFRs, 50%) compared to the other three organisms (32 TFRs, or 65%, in *M. tuberculsosis*; 27 TFRs, or 68%, in *P. aeruginosa*; and 10 TFRs, or 77%, in *E. coli*). This analysis suggests that in bacteria, most TFRs will be divergently oriented to their neighbors.

2.3.2 Variable features of TFRs and their divergently oriented neighbors

We investigated the relationship of the 250 TFRs having divergent neighbors from *S. coelicolor*, *S. griseus*, and *S. avermitilis*. First we explored the length of the DNA separating each TFR-encoding gene from its putative target (Table 2.2, note that the separation in bp is reported relatively to the genes' translational start sites as the transcriptional start sites are unknown in the overwhelming majority of cases).

Table 2.2. Analysis of the TFRs having divergent neighbors

TFR	Length of intergenic DNA (bp)	Divergent neighbor	Putative divergent gene product[a]
TFRs with intergenic DNA ≤200 bp			
SCO0116	116	SCO0117	NADB_Rossmann superfamily (EC 1)
SCO0155	132	SCO0154	Thioredoxin_like superfamily (EC 1)
SCO0250	63	SCO0249	NADB_Rossmann superfamily (EC 1)
SCO0253	70	SCO0252	NADB_Rossmann superfamily (EC 1)
SCO0296	127	SCO0295	MFS transporter
SCO0310	108	SCO0311	AMP-binding superfamily (EC 6)
SCO0337	121	SCO0336	Saccharop_dh family (EC 1)
SCO0430	23	SCO0429	Aldo_ket_red superfamily (EC 1)
SCO0485	148	SCO0484	NADB_Rossmann superfamily (EC 1)
SCO0508	9	SCO0507	NADB_Rossmann superfamily (EC 1)
SCO0512	65	SCO0513	PerM family transporter
SCO0520	136	SCO0519	NADB_Rossmann superfamily (EC 1)
SCO0646	111	SCO0645	NADB_Rossmann superfamily (EC 1)
SCO0669	77	SCO0668	FAD_binding_4 superfamily (EC 1)
SCO0728	39	SCO0727	Protein of unassigned function
SCO0745	199	SCO0746	Protein of unassigned function
SCO0772	96	SCO0771	Protein of unassigned function
SCO0800	114	SCO0801	CypX superfamily (EC 1)
SCO0857	139	SCO0856	Membrane protein of unassigned function
SCO0887	139	SCO0888	FMN_red superfamily (EC 1)
SCO1003	98	SCO1002	NTF2_like superfamily (EC 5)
SCO1034	102	SCO1033	ABC-type transporter
SCO1135	188	SCO1134	fer2 superfamily (EC 1)
SCO1193	101	SCO1194	MFS transporter
SCO1210	170	SCO1209	ACAD superfamily (EC 1)
SCO1339	102	SCO1338	NADB_Rossmann superfamily (EC 1)
SCO1702	89	SCO1701	ACAD superfamily (EC 1)
SCO1718	63	SCO1719	ABC-type transporter
SCO2243	90	SCO2242	Membrane protein of unassigned function
SCO2319	114	SCO2318	Glycosyltransferase_GTB_type superfamily (EC 2)
SCO2374	115	SCO2373	MFS transporter
SCO2775	97	SCO2776	ACCA superfamily (EC 6)
SCO2815	33	SCO2814	AdoMet_Mtases superfamily (EC 2)
SCO2994	106	SCO2995	ABC-type transporter
SCO3167	148	SCO3168	Peptidase_S41 superfamily (EC 3)
SCO3315	128	SCO3314	NADB_Rossmann superfamily (EC 1)
SCO3367	158	SCO3366	MFS transporter
SCO3587	0	SCO3588	Erythro_esteras superfamily (EC 3)
SCO3769	75	SCO3770	CypX superfamily (EC 1)
SCO3979	80	SCO3978	NADB_Rossmann superfamily (EC 1)
SCO4008	136	SCO4007	MFS transporter
SCO4099	139	SCO4098	LbetaH superfamily (EC 2)
SCO4167	21	SCO4168	NADB_Rossmann superfamily (EC 1)
SCO4270	115	SCO4271	NADB_Rossmann superfamily (EC 1)
SCO4303	94	SCO4304	NADB_Rossmann superfamily (EC 1)
SCO4313	0	SCO4312	HDc superfamily (EC 3)
SCO4358	96	SCO4359	ABC-type transporter

SCO4421	134	SCO4422	Glo_EDI_BRP_like superfamily (EC 4)
SCO4450	69	SCO4449	SGL family (EC 4)
SCO4454	181	SCO4455	Membrane protein of unassigned function
SCO4461	125	SCO4462	MFS transporter
SCO4480	158	SCO4481	PKc_like superfamily (EC 2)
SCO4639	64	SCO4638	AdoMet_Mtases superfamily (EC 2)
SCO4871	72	SCO4870	NADB_Rossmann superfamily (EC 1)
SCO4898	109	SCO4899	Protein of unassigned function
SCO4940	101	SCO4939	Lactamase_B superfamily (EC 3)
SCO4942	120	SCO4943	Aldo_ket_red superfamily (EC 1)
SCO4952	144	SCO4951	Aldo_ket_red superfamily (EC 1)
SCO5068	155	SCO5069	NADB_Rossmann superfamily (EC 1)
SCO5082 (ActR)	110	SCO5083	MFS transporter
SCO5209	192	SCO5208	FIG superfamily (EC 3)
SCO5238	94	SCO5237	NADB_Rossmann superfamily (EC 1)
SCO5384	93	SCO5383	ABC-type transporter
SCO5483	127	SCO5484	Membrane protein of unassigned function
SCO5517	65	SCO5516	MFS transporter
SCO5532	91	SCO5531	DMT superfamily transporter
SCO5811	187	SCO5810	MFS transporter
SCO5906	153	SCO5905	RND superfamily transporter
SCO5951	147	SCO5950	MFS transporter
SCO5956	0	SCO5955	Protein of unassigned function
SCO6121	92	SCO6122	Membrane protein of unassigned function
SCO6144	151	SCO6145	PHP family (EC 3)
SCO6265 (ScbR)	117	SCO6266 (ScbA)	hot_dog superfamily (EC 4)
SCO6350	79	SCO6351	Abhydrolase_6 family (EC 3)
SCO6694	175	SCO6693	PAP2_like superfamily (EC 3)
SCO6792	74	SCO6791	ACAD superfamily (EC 1)
SCO7222	146	SCO7223	NADB_Rossmann superfamily (EC 1)
SCO7303	193	SCO7302	2OG-FeII_Oxy superfamily (EC 1)
SCO7364	105	SCO7365	Protein of unassigned function
SCO7441	86	SCO7440	Abhydrolase_6 family (EC 3)
SCO7552	99	SCO7553	NADB_Rossmann superfamily (EC 1)
SCO7602	127	SCO7601	Amidase superfamily (EC 3)
SCO7624	53	SCO7625	NADB_Rossmann superfamily (EC 1)
SCO7645	107	SCO7644	Saccharop_dh family (EC 1)
SCO7651	30	SCO7652	NAT_SF superfamily (EC 2)
SCO7694	62	SCO7693	NADB_Rossmann superfamily (EC 1)
SCO7719	54	SCO7720	Protein of unassigned function
SCO7794	67	SCO7793	Protein of unassigned function
SCO7809	51	SCO7810	NADB_Rossmann superfamily (EC 1)
SCO7815	13	SCO7814	NADB_Rossmann superfamily (EC 1)
SCO7824	0	SCO7825	Protein of unassigned function
SCP1.242 (MmfR)	194	SCP1.243 (MmfL)	hot_dog superfamily (EC 4)
SCP1.312	180	SCP1.311	NADB_Rossmann superfamily (EC 1)
SCP1.42	180	SCP1.43	NADB_Rossmann superfamily (EC 1)
SGR0049t	95	SGR0048t	NADB_Rossmann superfamily (EC 1)
SGR0138	77	SGR0137	Protein of unassigned function
SGR0218	71	SGR0219	NADB_Rossmann superfamily (EC 1)

SGR0266	84	SGR0265	Protein of unassigned function
SGR0372	61	SGR0371	Glyco_hydro_1 superfamily (EC 3)
SGR0429	114	SGR0428	Transcriptional regulator
SGR0829	159	SGR0828	PAP2_like superfamily (EC 3)
SGR0945	113	SGR0944	PNPOx_like superfamily (EC 1)
SGR1074	121	SGR1073	ABC-type transporter
SGR1207	133	SGR1208	AANH_like superfamily (EC 1)
SGR1523	108	SGR1524	Flavin_utilizing_monoxygenases superfamily (EC 1)
SGR1633	121	SGR1632	MFS transporter
SGR1965	138	SGR1964	NADB_Rossmann superfamily (EC 1)
SGR2032	64	SGR2031	Protein of unassigned function
SGR2153	90	SGR2154	ABC-type transporter
SGR2313	123	SGR2314	FIG superfamily (EC 3)
SGR2549	126	SGR2548	Siderophore interacting protein
SGR2601	104	SGR2602	Abhydrolase_6 family (EC 3)
SGR2777	127	SGR2776	MFS transporter
SGR2879	96	SGR2880	NADB_Rossmann superfamily (EC 1)
SGR3047	162	SGR3048	SRPBCC superfamily (EC 4)
SGR3098	104	SGR3099	NADB_Rossmann superfamily (EC 1)
SGR3119	162	SGR3118	ABC-type transporter
SGR3387	156	SGR3386	MFS transporter
SGR3746	156	SGR3745	ALDH-SF superfamily (EC 1)
SGR3979	144	SGR3978	ABC-type transporter
SGR4317	184	SGR4316	Peptidase_S41 superfamily (EC 3)
SGR4423	114	SGR4424	Protein of unassigned function
SGR4539	191	SGR4538	ABC-type transporter
SGR4760	123	SGR4759	ACCA superfamily (EC 6)
SGR4763	102	SGR4762	NADB_Rossmann superfamily (EC 1)
SGR4829	158	SGR4830	RND superfamily transporter
SGR5038	124	SGR5039	Protein of unassigned function
SGR5044	70	SGR5043	NADB_Rossmann superfamily (EC 1)
SGR5084	101	SGR5083	NADB_Rossmann superfamily (EC 1)
SGR5127	127	SGR5128	MFS transporter
SGR5186	182	SGR5187	Glycosyltransferase_GTB_type superfamily (EC 2)
SGR5798	90	SGR5799	ACAD superfamily (EC 1)
SGR5901	65	SGR5900	DMT superfamily transporter
SGR6082	74	SGR6081	RND superfamily transporter
SGR6196	83	SGR6197	SMR-type transporter
SGR6294	125	SGR6293	Membrane protein of unassigned function
SGR6310	187	SGR6311	ACAD superfamily (EC 1)
SGR6406	132	SGR6407	Protein of unassigned function
SGR6435	193	SGR6436	fer2 superfamily (EC 1)
SGR6536	150	SGR6535	Protein of unassigned function
SGR6627	106	SGR6626	NADB_Rossmann superfamily (EC 1)
SGR6910	114	SGR6911	RICIN superfamily (EC 3)
SGR6935	148	SGR6934	Saccharop_dh family (EC 1)
SGR7004	164	SGR7003	Aldo_ket_red superfamily (EC 1)
SGR7090t	95	SGR7091t	NADB_Rossmann superfamily (EC 1)
SAV0146	151	SAV0145	RND superfamily transporter
SAV0151	147	SAV0152	NADB_Rossmann superfamily (EC 1)
SAV0205	90	SAV0204	Membrane protein of unassigned function

SAV0448	122	SAV0449	Protein of unassigned function
SAV0488	119	SAV0489	Protein of unassigned function
SAV0675	126	SAV0676	NADB_Rossmann superfamily (EC 1)
SAV0754	106	SAV0753	Protein of unassigned function
SAV0775	170	SAV0774	NADB_Rossmann superfamily (EC 1)
SAV0880	181	SAV0881	NADB_Rossmann superfamily (EC 1)
SAV0882	59	SAV0883	NADB_Rossmann superfamily (EC 1)
SAV0913	117	SAV0912	NADB_Rossmann superfamily (EC 1)
SAV1068	169	SAV1067	NADB_Rossmann superfamily (EC 1)
SAV1218	62	SAV1217	Lactamase_B superfamily (EC 3)
SAV1359	83	SAV1358	NADB_Rossmann superfamily (EC 1)
SAV1439	148	SAV1438	Membrane protein of unassigned function
SAV1466	185	SAV1467	NADB_Rossmann superfamily (EC 1)
SAV1468	86	SAV1469	Protein of unassigned function
SAV1778	175	SAV1777	NADB_Rossmann superfamily (EC 1)
SAV1781	68	SAV1780	PNPOx_like superfamily (EC 1)
SAV2263	159	SAV2264	MFS transporter
SAV2268	68	SAV2267	ACAD superfamily (EC 1)
SAV2270	191	SAV2269 (AvaA)	hot_dog superfamily (EC 4)
SAV2381	153	SAV2380	ABC-type transporter
SAV2454	189	SAV2455	MFS transporter
SAV2727	184	SAV2728	MFS transporter
SAV2759	134	SAV2758	Protein of unassigned function
SAV2814	149	SAV2813	MFS transporter
SAV2871	80	SAV2872	ABC-type transporter
SAV3049	188	SAV3050	FIG superfamily (EC 3)
SAV3379	38	SAV3380	NADB_Rossmann superfamily (EC 1)
SAV3546	118	SAV3545	Protein of unassigned function
SAV3619	186	SAV3620	Peptidase_S41 superfamily (EC 3)
SAV3760	148	SAV3761	Protein of unassigned function
SAV3818	145	SAV3817	Glo_EDI_BRP_like superfamily (EC 4)
SAV4017	104	SAV4018	NADB_Rossmann superfamily (EC 1)
SAV4530	186	SAV4531	Esterase_lipase superfamily (EC 3)
SAV4701	189	SAV4702	MFS transporter
SAV4782	98	SAV4781	NADB_Rossmann superfamily (EC 1)
SAV5084	98	SAV5085	NADB_Rossmann superfamily (EC 1)
SAV5174	107	SAV5173	Abhydrolase_6 family (EC 3)
SAV5279	97	SAV5278	ACCA superfamily (EC 6)
SAV5333	167	SAV5334	RND superfamily transporter
SAV5724	181	SAV5725	Membrane protein of unassigned function
SAV5796	115	SAV5797	MFS transporter
SAV6352	137	SAV6351	MFS transporter
SAV6440	57	SAV6441	ADC superfamily (EC 4)
SAV6599	77	SAV6600	ACAD superfamily (EC 1)
SAV6701	185	SAV6700	Protein of unassigned function
SAV7046	113	SAV7045	FMN_red superfamily (EC 1)
SAV7127	120	SAV7128	ACAD superfamily (EC 1)
SAV7167	42	SAV7168	4Oxalocrotonate_Tautomerase superfamily (EC 5)
SAV7427	108	SAV7426	CypX superfamily (EC 1)
SAV7471	126	SAV7472	Protein of unassigned function

TFRs with intergenic DNA >200 bp			
SCO0241	345	SCO0240	NADB_Rossmann superfamily (EC 1)
SCO3207	238	SCO3208	Peptidase_M6 superfamily (EC 3)
SCO4118 (AtrA)	425	SCO4119	Pyr_redox superfamily (EC 1)
SCO5296	308	SCO5297	DUF772 superfamily (EC 2)
SCO5418	303	SCO5419	Thioredoxin_like superfamily (EC 1)
SCO6071 (CprB)	355	SCO6070	Periplasmic binding protein
SCO6286	237	SCO6287	Thioesterase family (EC 3)
SCO6323	385	SCO6324	HAD_like superfamily (EC 3)
SCO6775	232	SCO6776	Lactamase_B superfamily (EC 3)
SCO6784	245	SCO6785	CoA_transf_3 superfamily (EC 2)
SCO7347	236	SCO7346	MFS transporter
SGR1098	303	SGR1099	PAD_porph superfamily (EC 3)
SGR1284	241	SGR1283	DMT superfamily transporter
SGR2119	452	SGR2118	Thioredoxin_like superfamily (EC 1)
SGR3022	330	SGR3021	Glo_EDI_BRP_like superfamily (EC 4)
SGR3123	397	SGR3124	TauE/SafE exporter
SGR3127	228	SGR3126	Glyco_hydro_3 superfamily (EC 3)
SGR3402	601	SGR3403	MFS transporter
SGR3731 (ArpA)	1024	SGR3730	Protein of unassigned function
SGR3905 (AtrA-g)	565	SGR3906	Pyr_redox superfamily (EC 1)
SGR4211	251	SGR4210	Abhydrolase_5 family (EC 3)
SGR4271	292	SGR4270	Peptidase_M6 superfamily (EC 3)
SGR5223	281	SGR5222	MFS transporter
SGR5234	293	SGR5233	Amidinotransf superfamily (EC 2)
SGR5269	212	SGR5270	MFS transporter
SGR5284	692	SGR5285	Transcriptional regulator
SGR5789	222	SGR5788	Cupin_2 superfamily (EC 1)
SGR5970	323	SGR5969	P-loop_NTPase superfamily (EC 3)
SGR6382	683	SGR6383	Transcriptional regulator
SGR6383	683	SGR6382	Transcriptional regulator
SGR6390	368	SGR6389	MPP_superfamily (EC 3)
SGR6440	215	SGR6441	MFS transporter
SGR6912	280	SGR6911	RICIN superfamily (EC 3)
SGR6951	268	SGR6952	Protein of unassigned function
SAV0082	205	SAV0083	Protein of unassigned function
SAV0292	1123	SAV0291	rve superfamily (EC 2)
SAV0431	371	SAV0432	Membrane protein of unassigned function
SAV0508	386	SAV0507	NADB_Rossmann superfamily (EC 1)
SAV0566	245	SAV0565	Glo_EDI_BRP_like superfamily (EC 1)
SAV0576	899	SAV0575	CypX superfamily (EC 1)
SAV0585	265	SAV0586	PP2Cc superfamily (EC 3)
SAV0669	244	SAV0668	Glo_EDI_BRP_like superfamily (EC 4)
SAV1541	255	SAV1540	fer2 superfamily (EC 1)
SAV1711	231	SAV1712	PAP2_like superfamily (EC 3)
SAV2056	229	SAV2055	NADB_Rossmann superfamily (EC 1)
SAV2098	316	SAV2097	P-loop_NTPase superfamily (EC 3)
SAV2831	236	SAV2830	Thioredoxin_like superfamily (EC 1)
SAV3699	238	SAV3700	Peptidase_M6 superfamily (EC 3)

SAV4110	424	SAV4109	Pyr_redox superfamily (EC 1)
SAV5082	237	SAV5081	ABC-type transporter
SAV5854	208	SAV5855	Glycosyltransferase_GTB_type superfamily (EC 2)
SAV7510	209	SAV7509	Protein of unassigned function

[a] For the putative enzymes, their conserved domains as well as predicted Enzyme Commission (EC) groups (EC 1: oxidoreductases; EC 2: transferases; EC 3: hydrolases; EC 4: lyases; EC 5: isomerases; EC 6: ligases) are indicated.

As shown in Figure 2.2A, the length of this DNA varies from 0 bp to 1123 bp. However, most intergenic regions (198 of 250, or 79%) are $\leq$200 bp (Figure 2.2B). A similar pattern was observed in *P. aeruginosa* and *M. tuberculosis* with 74% (20 TFRs) and 75% (24 TFRs) of their respective TFRs having divergent neighbors less than 200 bp away from the adjacent open reading frames. On the other hand, the intergenic regions in this size range are less frequent in *B. subtilis* (5 TFRs, 56%) and *E. coli* (5 TFRs, 50%) although this may be exaggerated by the smaller sample size in these organisms.

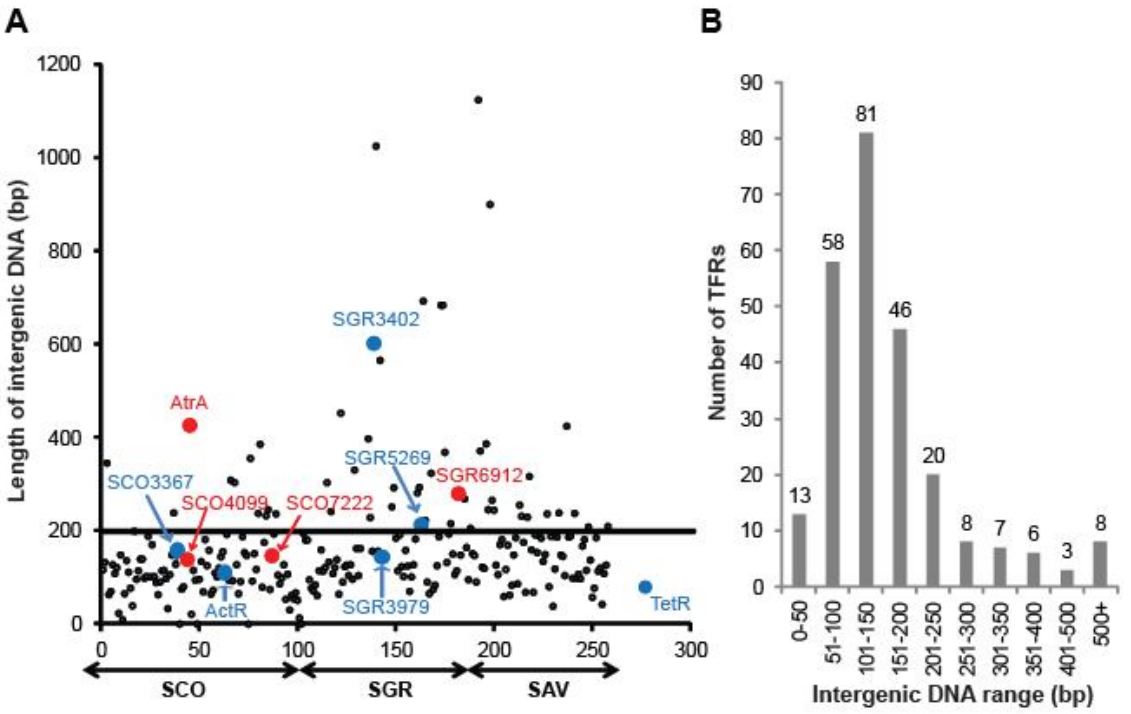

Figure 2.2. Length of intergenic DNAs between TFRs and their divergent neighbors. (A) Each of the 250 TFRs having divergent neighbors in *S. coelicolor* (SCO), *S. griseus* (SGR), and *S. avermitilis* (SAV) is represented as a dot with the value on y-axis indicating the length of the intergenic sequence between its own gene and divergent gene. On x-axis, the TFRs are placed in the order of their gene annotations along the length of the linear chromosomes (the host streptomycete is stated below). The larger colored dots correspond to the TFRs investigated in this study (see Table 2.4 and text for details. A model TFR, TetR, is shown on the graph as a reference). Blue dots indicate the TFRs whose divergent neighbors encode putative membrane transporters, while the TFRs represented by red dots are adjacent to genes encoding putative enzymes. **(B)** The TFRs having divergent neighbors are grouped according to the range of their intergenic DNA length.

We next analyzed the protein products encoded by the divergent neighboring genes using protein BLAST and Conserved Domain Search (CD-Search, discussed in [155]) (Table 2.2). As shown in Figure 2.3A, the predicted gene products include putative enzymes (154 of 250, or 62%), membrane proteins (61, or 24%), and other proteins such as transcriptional regulators (6, or 2%). The function of 29, or 12%, of the putative targets could not be predicted as they lack any known motif and/or have no BLAST hit with proteins of known function.

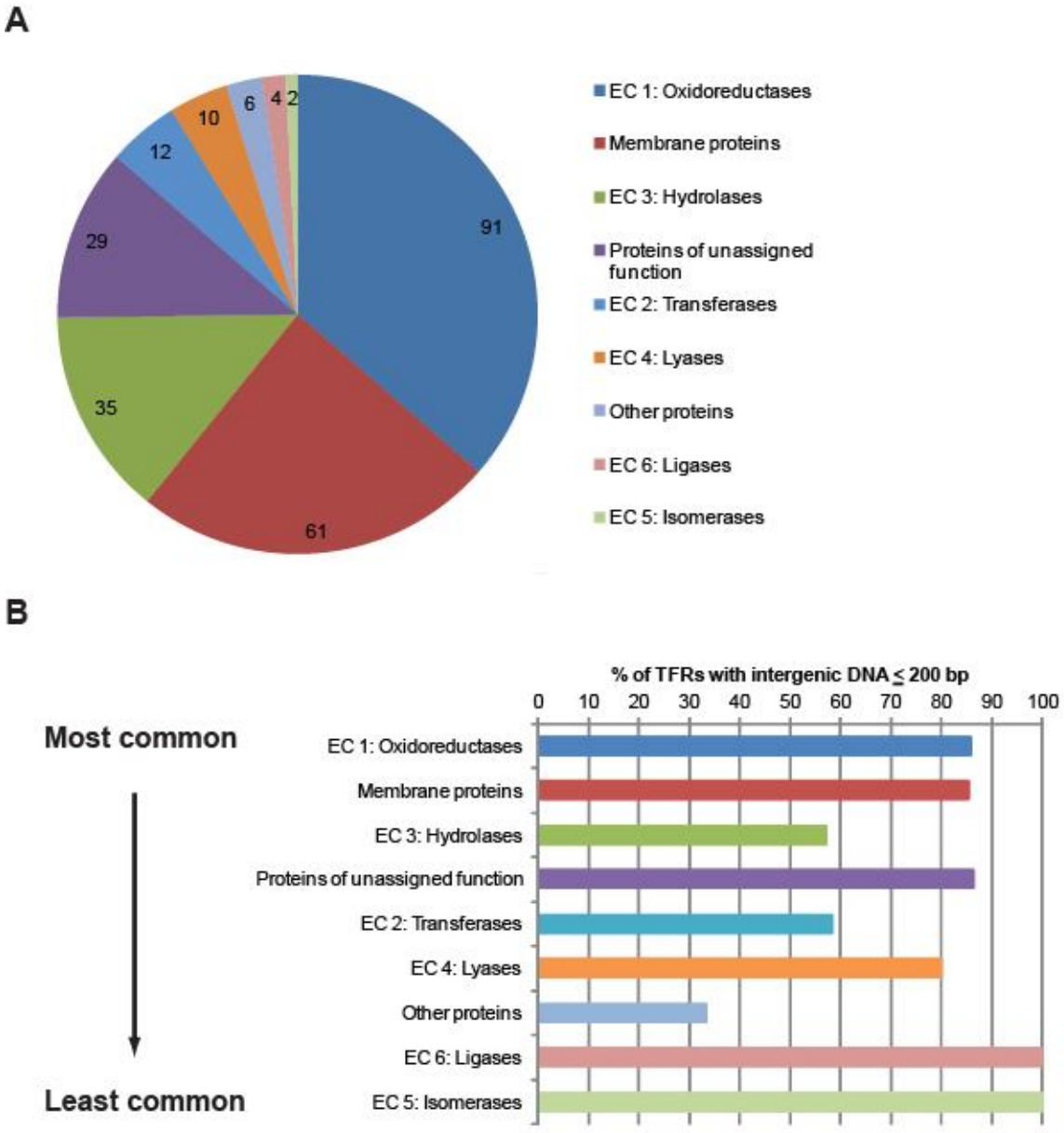

Figure 2.3. Diverse product types encoded by the divergent neighboring genes. (A) The number of TFRs adjacent to each type of divergent gene products – sorted by enzymes in six groups (EC 1 to EC 6), membrane proteins, other proteins (e.g. transcriptional regulators), and proteins of unassigned function. **(B)** TFRs were grouped according to their divergent gene type and the percentage of TFRs having intergenic DNAs ≤200 bp for each group is shown.

The predicted enzymes were further divided based on two criteria: the Enzyme Commission (EC) number to indicate the type of the chemical reactions they are predicted to catalyze [156] as well as any conserved domain they possess. As demonstrated in Figure 2.3A and Table 2.3, our analysis revealed that the 154 putative enzymes include members in all six known EC groups (i.e. EC 1 to EC 6). For example, 91 of the 154 putative enzymes are predicted to be oxidoreductases (EC 1). 51 of these have a conserved sequence of the Rossmann fold (NADB_Rossmann, cl09931, in Table 2.3), which is characterized by the $Gx_{1-2}GxxG$ motif [157] and is one of the most common folds in the Protein Data Bank (PDB) [158]. A large number of proteins containing the Rossmann fold bind to nucleotide cofactors such as FAD and NAD(P) and function as oxidoreductases such as lactate dehydrogenases and flavodoxins [159]. On the other hand, eight proteins are grouped in the acyl-CoA dehydrogenase superfamily (ACAD, cl09933, in Table 2.3), known to be involved in a broad spectrum of primary and secondary metabolic processes such as the β-oxidation of fatty acids [160] and antibiotic biosynthesis [161].

Table 2.3. Types of protein products encoded by the divergent neighboring genes

Type of gene product		References
EC 1: Oxidoreductases[a]		
Family of conserved domain	Number of TFRs	
NADB_Rossmann superfamily[b] (cl09931[c])	51	[159]
ACAD superfamily[b] (cl09933[c])	8	[162,163]
Aldo_ket_red superfamily[b] (cl00470[c])	4	[164]
CypX superfamily[b] (cl12078[c])	4	[165]
Thioredoxin_like superfamily[b] (cl00388[c])	4	[166]
Saccharop_dh (pfam03435)[d]	3	[167]
Pyr_redox superfamily[b] (cl15766)	3	[168]
fer2 superfamily[b] (cl00159) + Fer2_2 superfamily (cl08334[c])	3	[169]
FMN_red superfamily[b] (cl00438[c])	2	[170]
PNPOx_like superfamily[b] (cl00381[c])	2	[171]
Flavin_utilizing_monoxygenases superfamily[b] (cl07892[c])	1	[172]
ALDH-SF superfamily[b] (cl11961[c])	1	[173]
Cupin_2 superfamily[b] (cl09118[c])	1	[174]
FAD_binding_4 superfamily[b] (cl14794[c])	1	[175]
Glo_EDI_BRP_like superfamily[b] (cl14632[c])	1	[176]
2OG-FeII_Oxy superfamily[b] (cl15773[c])	1	[177]
AANH_like superfamily[b] (cl00292[c])	1	[178]
Number of TFRs divergent to the EC 1 genes	**91**	
EC 2: Transferases[a]		
Family of conserved domain	Number of TFRs	
Glycosyltransferase_GTB_type superfamily[b] (cl10013[c])	3	[179]
AdoMet_Mtases superfamily[b] (cl15754[c])	2	[180]
PKc_like superfamily[b] (cl09925[c])	1	[181]
CoA_transf_3 superfamily[b] (cl15643[c])	1	[182]
LbetaH superfamily[b] (cl00160[c])	1	[183,184]
NAT_SF superfamily[b] (cl00357[c])	1	[185]
Amidinotransf superfamily[b] (cl12043[c])	1	[186]
DUF772 superfamily[b] (cl15789[c])	1	[187]
rve superfamily[b] (cl01316[c])	1	[188]
Number of TFRs divergent to the EC 2 genes	**12**	
EC 3: Hydrolases[a]		
Family of conserved domain	Number of TFRs	
Abhydrolase_6 (pfam12697)[d]	4	[189]
Peptidase_S41 superfamily[b] (cl02526[c])	3	[190]
Peptidase_M6 superfamily[b] (cl11525[c])	3	[191]
Lactamase_B superfamily[b] (cl00446[c])	3	[192]
FIG superfamily[b] (cl00289[c])	3	[193]
PAP2_like superfamily[b] (cl00474[c])	3	[194]
P-loop_NTPase superfamily[b] (cl09099[c])	2	[195]

HAD_like superfamily[b] (cl11391[c])	1	[196]
Glyco_hydro_1 superfamily[b] (cl01046[c])	1	[197]
Erythro_esteras superfamily[b] (cl10069[c])	1	[198,199]
Glyco_hydro_3 superfamily[b] (cl07971[c])	1	[200]
MPP_superfamily[b] (cl13995[c])	1	[201]
RICIN superfamily[b] (cl15820[c])	1	[202]
Amidase superfamily[b] (cl11426[c])	1	[203]
PAD_porph superfamily[b] (cl01113[c])	1	[204]
Esterase_lipase superfamily[b] (cl12031[c])	1	[205]
PP2Cc superfamily[b] (cl00120[c])	1	[206]
Thioesterase (pfam00975)[d]	1	[207]
Abhydrolase_5 (pfam12695)[d]	1	[208]
HDc superfamily[b] (cl00076[c])	1	[209]
PHP (pfam02811)	1	[201]
Number of TFRs divergent to the EC 3 genes	**35**	

EC 4: Lyases[a]

Family of conserved domain	Number of TFRs	
Glo_EDI_BRP_like superfamily[b] (cl14632[c])	4	[210]
hot_dog superfamily[b] (cl00509[c])	3	[211,212,213]
ADC superfamily[b] (cl01919[c])	1	[214]
SRPBCC superfamily[b] (cl14643[c])	1	[215]
SGL (pfam08450)[d]	1	[216]
Number of TFRs divergent to the EC 4 genes	**10**	

EC 5: Isomerases[a]

Family of conserved domain	Number of TFRs	
4Oxalocrotonate_Tautomerase superfamily[b] (cl00235[c])	1	[217]
NTF2_like superfamily[b] (cl09109[c])	1	[218]
Number of TFRs divergent to the EC 5 genes	**2**	

EC 6: Ligases[a]

Family of conserved domain	Number of TFRs	
ACCA superfamily[b] (cl15772[c])	3	[219]
AMP-binding superfamily[b] (cl15778[c])	1	[220]
Number of TFRs divergent to the EC 6 genes	**4**	

Membrane proteins

Family of conserved domain	Number of TFRs	
MFS transporters	26	[221]
ABC-type transporters	14	[222]
RND superfamily transporters	5	[223]
DMT superfamily transporters	3	[224]
SMR-type transporters	1	[225]

PerM family of transporters	1	[226]
TauE/SafE exporter	1	[227]
Other membrane proteins of unassigned function	10	
Number of TFRs divergent to the genes encoding membrane proteins	**61**	
	Number of TFRs	
Proteins of other functions (e.g. transcription regulators)	**6**	
	Number of TFRs	
Proteins of unassigned functions	**29**	

[a] The EC groups indicate the type of reactions putative enzymes are predicted to catalyze.
[b] The conserved domain superfamilies (as named in NCBI) were used to further group putative enzymes in each EC group.
[c] An accession number for a superfamily domain has the prefix "cl" for "cluster".
[d] If a putative enzyme lacks a superfamily conserved domain, the protein family (pfam) motif was indicated instead.

Among the membrane proteins encoded by the putative target genes, 84% (51 of 61) are predicted to be transporters while the remainders contain putative transmembrane segments but lack any other predictive sequence motif (Table 2.3). While 26 of the transporters are predicted to belong to the major facilitator superfamily (MFS), the others belong to families such as the ATP-binding cassette (ABC) or resistance-nodulation-division (RND) transporter families.

Certain gene types such as EC 1 oxidoreductases (36%) and membrane proteins (24%) were found more frequently than others (e.g. EC 6 ligases, 2%, and EC 5 isomerases, 1%) (Figure 2.3A). There was no obvious correlation between the length of the intergenic DNA and the type of divergent gene product (Figure 2.3B).

While two of the best characterized TFRs – TetR and QacR – are divergently oriented to their target efflux pumps [50,87], our analysis suggests that there is a much greater diversity in the predicted functions of the possible targets regulated by TFRs: most of these putative target genes do not encode export proteins.

2.3.3 *In vitro* analysis of selected TFRs having divergent neighboring genes

To determine whether the length of the intergenic DNA or the putative function of the neighboring gene correlates with regulation by an adjacent TFR, we selected eight previously uncharacterized TFRs from *S. coelicolor* and *S. griseus* for molecular genetic analysis (Figure 2.2A and Table 2.4). We chose TFRs divergent to putative transporters (three MFS and one ABC-type transporters) or enzymes (two EC 1 oxidoreductases, one EC 2 transferase, and one EC 3 hydrolase) with intergenic DNAs of varying lengths (139 bp to 601 bp). In addition, ActR (SCO5082) from *S. coelicolor* was used as a well-characterized control that regulates the divergently transcribed *actAB* operon encoding putative efflux pumps for an antibiotic actinorhodin [123,128]. The coding sequences of these proteins were amplified, subcloned, and expressed in *E. coli* such that they could be purified via His_6-tags.

Table 2.4. Nine TFRs investigated in this study and their divergent neighbors

TFR of interest	Predicted divergent gene product	Length of intergenic DNA
TFRs whose divergent genes encode putative transporters		
ActR (SCO5082)	ActA (SCO5083, MFS)	110 bp
SGR3979	SGR3977/SGR3978 (ABC)	144 bp
SCO3367	SCO3366 (MFS)	158 bp
SGR5269	SGR5270 (MFS)	212 bp
SGR3402	SGR3403 (MFS)	601 bp
TFRs whose divergent genes encode putative enzymes		
SCO4099	SCO4098 (Acyltransferase, EC 2)	139 bp
SCO7222	SCO7223 (Monooxygenase, EC 1)	146 bp
SGR6912	SGR6911 (Glycosyl hydrolase, EC 3)	280 bp
AtrA (SCO4118)	SCO4119 (NADH dehydrogenase, EC 1)	425 bp

We conducted electrophoretic mobility shift assays (EMSAs) to determine whether the nine TFRs bound their respective intergenic DNAs. As shown in Figure 2.4A and 2.4B, ActR (intergenic DNA=110 bp) and SCO4099 (139 bp) formed tight complexes with their cognate intergenic sequences. Although the numbers of protein-DNA complexes – consistent with the number of discrete binding sites – detected for ActR (three complexes) and SCO4099 (one complex) were different, mobility shifts were observed at protein concentrations as low as 0.2 nM and 6.25 nM, respectively. Similar observations were made with SGR3979 (144 bp, Figure 2.4C), SCO7222 (146 bp, Figure 2.4D), SCO3367 (158 bp, Figure 2.4E), and SGR5269 (212 bp, Figure 2.4F), all of which have intergenic sequences close to or smaller than 200 bp.

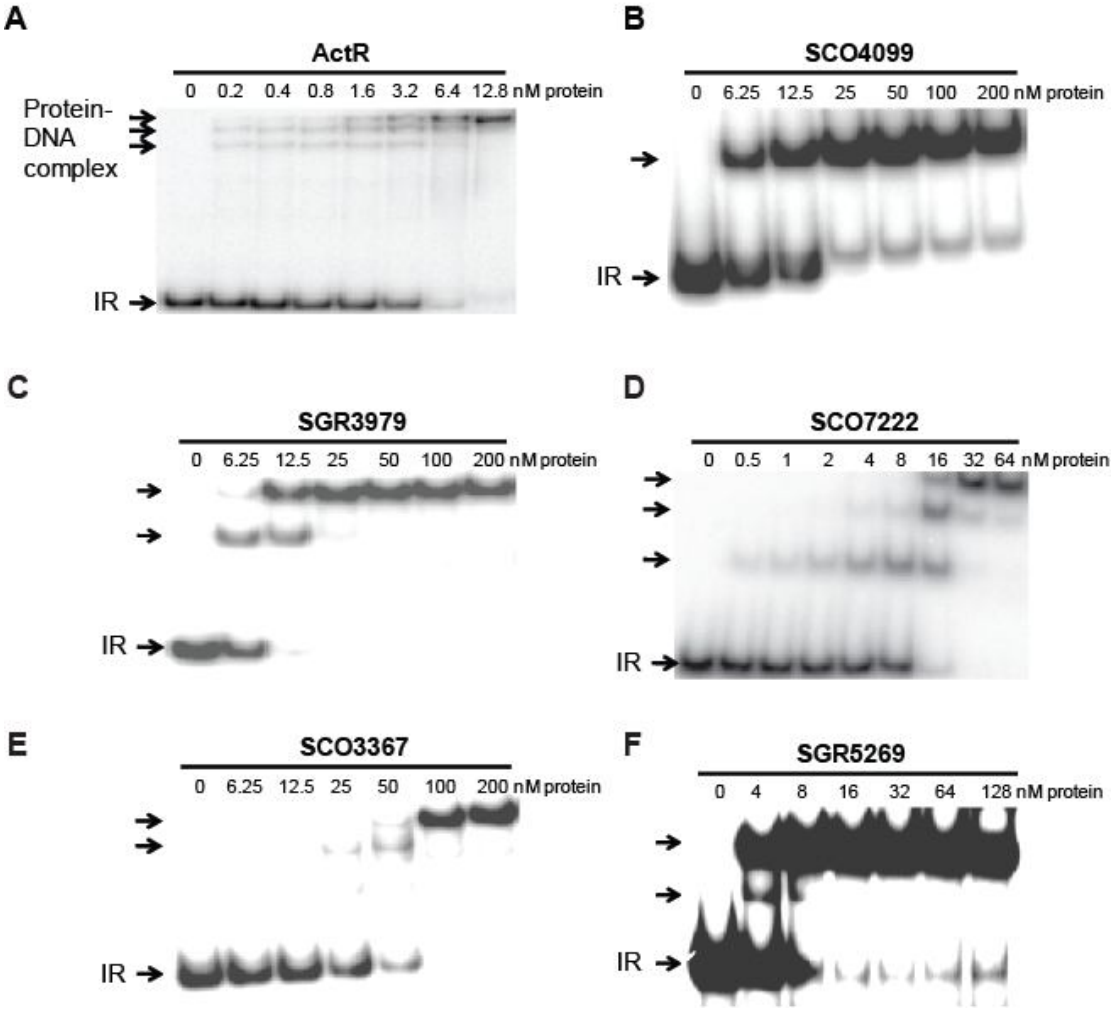

Figure 2.4. ActR, SCO4099, SGR3979, SCO3367, and SGR5269 bind the intergenic DNAs between their own genes and divergent neighbors. EMSA shows the interaction between ActR **(A)**, SCO4099 **(B)**, SGR3979 **(C)**, SCO7222 **(D)**, SCO3367 **(E)**, or SGR5269 **(F)** and the entire sequence of its cognate intergenic region (IR). The indicated concentrations of a TFR were incubated with a ^{32}P-labeled DNA fragment containing the intergenic sequence between the TFR-encoding gene and its divergent neighboring gene. Unbound DNA fragment is indicated by the bottom arrow, while the shifts representing protein-DNA complexes are indicated by the upper arrows.

The intergenic sequences of SGR6912 (280 bp), AtrA (SCO4118, 425 bp), and SGR3402 (601 bp) are much longer than 200 bp, so they were divided into multiple, overlapping probes (two probes for SGR6912, three for AtrA, and four for SGR3402) for *in vitro* assays. Similar to our observations with the TFRs with shorter intergenic sequences, both SGR6912 (Figure 2.5A and 2.5B) and AtrA (Figure 2.5C to 2.5E) bound their cognate intergenic sequence fragments. SGR6912 clearly bound more tightly to the probe closer to its own gene ($IGR_{SGR6912}$, shift observed at 6.25 nM, Figure 2.5A) than to the other probe closer to its divergent neighboring gene *SGR6911* ($IGR_{SGR6911}$, shift at 100 nM, Figure 2.5B). Similarly, a very high concentration of AtrA, 100 nM, was required for the formation of a detectable complex with the DNA probe closest to the divergent *SCO4119* gene (Figure 2.5C). In contrast, the two probes closer to the *atrA* gene itself formed complexes with 12.5 nM or 25 nM of the protein (Figure 2.5D and 2.5E). Finally, SGR3402 – with the longest intergenic sequence – did not interact with any of the four probes even at the highest protein concentration tested (400 nM, Figure 2.6). While it is possible that SGR3402 binds within the open reading frame(s) of *SGR3402* and/or *SGR3403*, these results imply that SGR3402 does not regulate *SGR3403* expression.

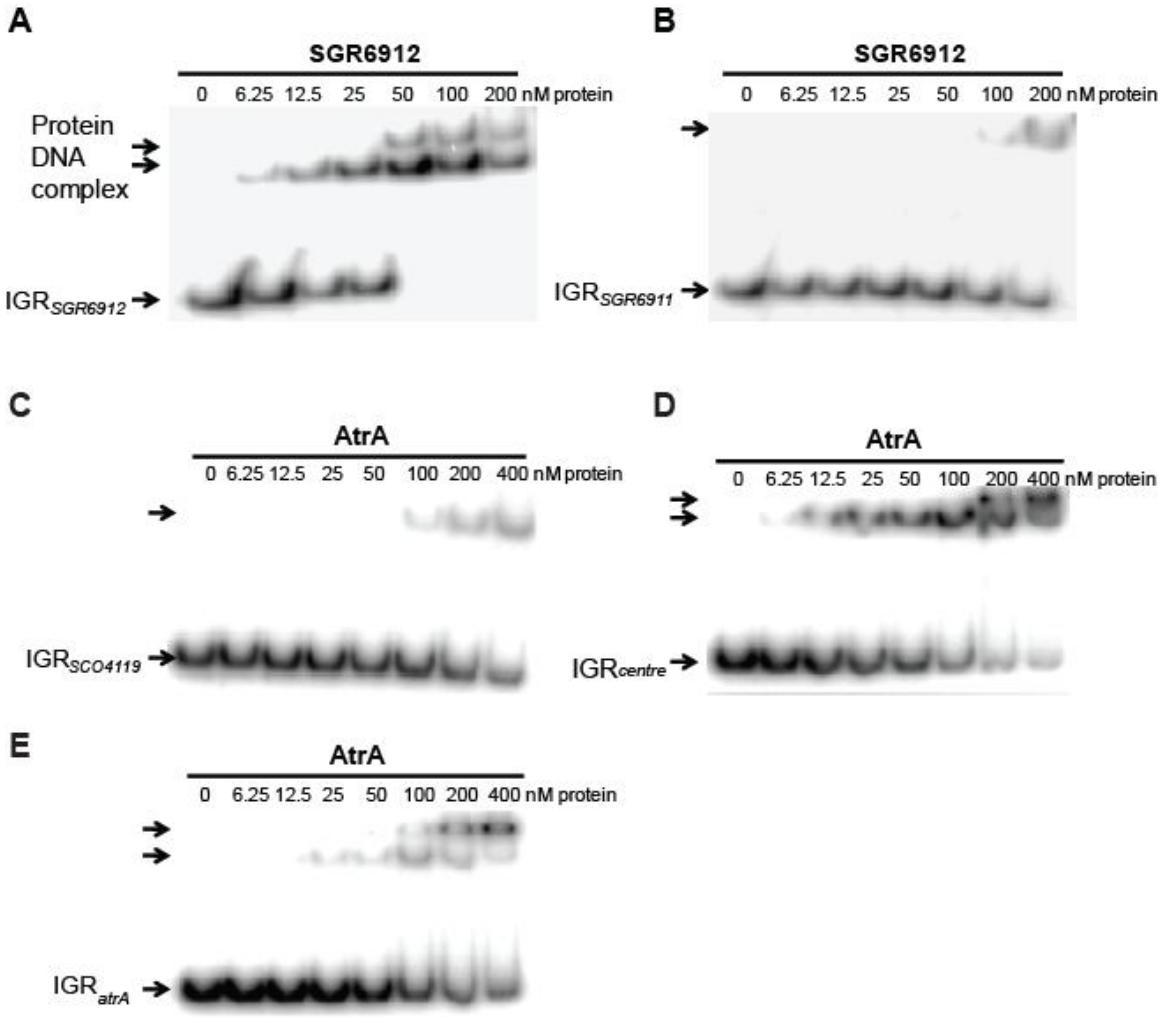

Figure 2.5. SGR6912 and AtrA bind the intergenic DNAs between their own genes and divergent neighbors. EMSA shows the interaction between SGR6912 or AtrA and the partial sequences of its cognate intergenic region. The indicated concentrations of a TFR were incubated with a ^{32}P-labeled DNA fragment containing the partial intergenic sequences between the TFR-encoding gene and its divergent neighboring gene. **(A and B)** The IGR$_{6912}$ probe contains the 200 bp sequence upstream of the *SGR6912* translational start, while the IGR$_{6911}$ probe contains the 200 bp sequence upstream of the *SGR6911* translational start site (partial *SGR6911*/*SGR6912* intergenic sequences). **(C to E)** Three probes were prepared for AtrA (IGR$_{SCO4119}$ → the 180 bp sequence from the SCO4119 translational start site; IGR$_{centre}$ → the central 180 bp region between the *atrA* and *SCO4119* translational start sites; IGR$_{atrA}$ → the 180 bp sequence from the *atA* translational start site).

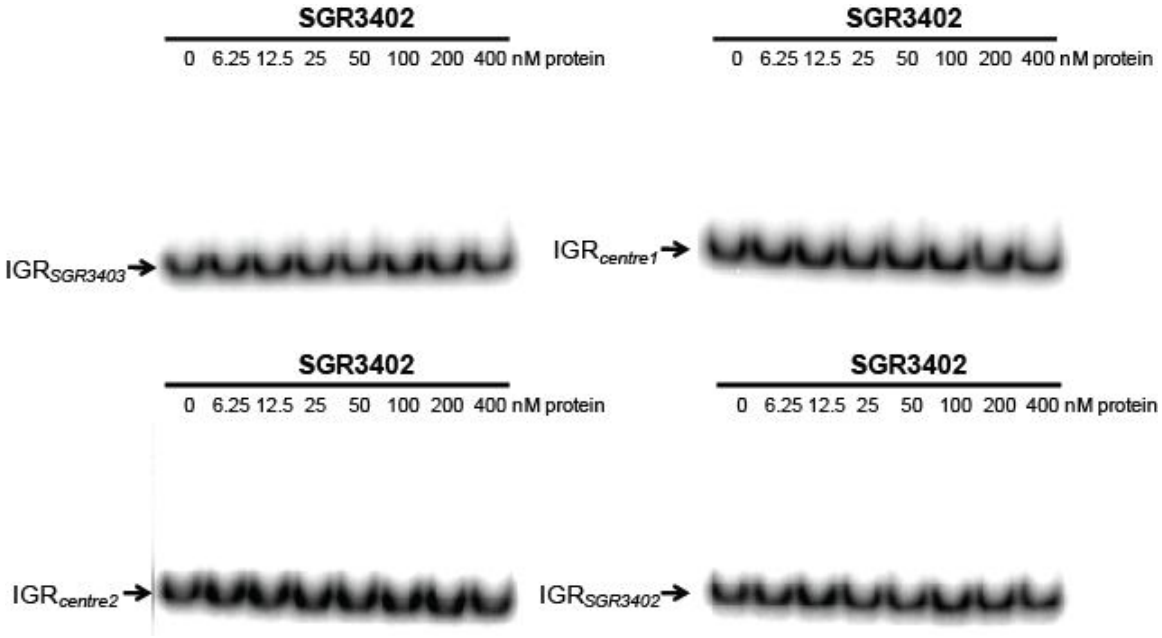

Figure 2.6. SGR3402 does not bind the intergenic DNA between *SGR3402* and *SGR3403*. Four probes for SGR3402 (IGR$_{SGR3403}$, 180 bp; IGR$_{centre1}$, 180 bp; IGR$_{centre2}$, 190 bp; and IGR$_{SGR3402}$, 148 bp, partially cover the *SGR3402*/*SGR3403* intergenic regions in the order of the increasing distance to the *SGR3403* translational start site) were prepared and incubated with the indicated concentrations of SGR3402.

We mapped the binding sites of the eight TFRs that bound the cognate intergenic sequences through DNase I footprint assays on both DNA strands, observing protected regions ranging from 15-48 bp (Figure 2.7 and 2.8). Importantly, these protected regions were observed at differing locations relative to their neighboring genes (summarized in Figure 2.9). For example, SGR3979 protected a single region (Figure 2.7A) located close to the divergent gene *SGR3978*, while SCO7222 bound three discrete regions (Figure 2.7B). Two of the SCO7222 binding sites are located closer to its own gene, while the remaining site is more adjacent to *SCO7223*. For SGR6912, the assays were

conducted with both probes of the $IGR_{SGR6911}$ and $IGR_{SGR6912}$ sequences. Protection by SGR6912 was only observed with the $IGR_{SGR6912}$ probe (Figure 2.7C), indicating that the operators of this TFR are positioned more proximal to *SGR6912*.

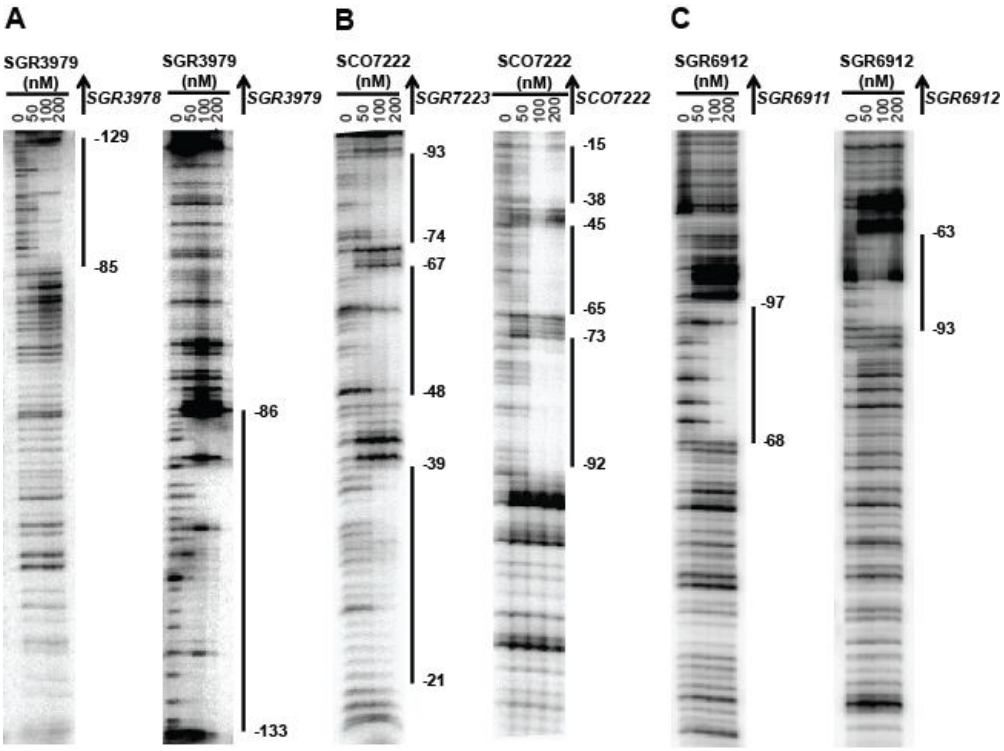

Figure 2.7. SGR3979, SCO7222, and SGR6912 display different DNA protection patterns on the cognate intergenic DNAs. DNaseI footprinting assays showing the protection patterns of SGR3979 on the entire *SGR3978*/*SGR3979* intergenic sequence **(A)**; SCO7222 on the entire *SCO7222*/*SCO7223* intergenic sequence **(B)**; and SGR6912 on the IGR_{6912} sequence containing the partial *SGR6911*/*SGR6912* intergenic region (same as Figure 2.5A) **(C)**. In the presence of the indicated concentrations of a TFR, a DNA fragment containing the entire or partial intergenic sequence between the TFR-encoding gene and its divergent neighboring gene was exposed to DNaseI. For the left gel for each TFR, the primer that was extended toward its divergent neighboring gene was labeled at 5′-end to prepare the probe, while the other primer extended toward its own gene was labeled to prepare the probe for the right gel. The regions protected by the TFRs are indicated by solid vertical lines. The numbers beside each line indicate the start and end positions of the protected region relative to the translational start site of the TFR-encoding gene.

SGR5269 behaved similarly to SGR6912 and only bound a single region (Figure 2.8A) adjacent to its own gene, while SCO3367 had two binding sites (Figure 2.8B) – one closer to *SCO3367* and the other one closer to *SCO3366*. Both ActR (Figure 2.8C) and SCO4099 (Figure 2.8D) bound near or at the centre of their respective intergenic sequences although SCO4099 protected a much smaller region than ActR. No footprint was obtained with AtrA (data not shown) although the previous EMSA experiments indicated that this TFR can bind its intergenic sequence (Figure 2.5C to 2.5E). Of note, most of the DNA protection patterns exhibited by these TFRs, except SCO3367, are dissimilar to what has been reported with TetR, which binds two distinct regions containing the *tetR*-proximal and *tetA*-proximal operators (Figure 1.5B) in order to regulate both genes [53].

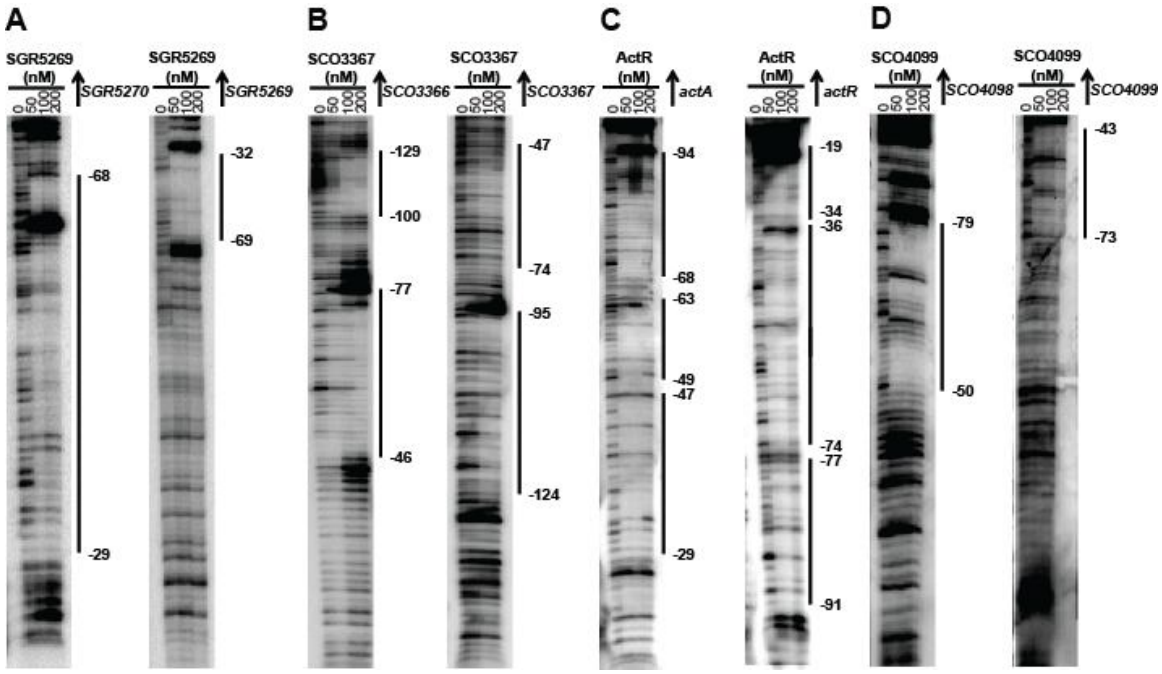

Figure 2.8. SGR5269, SCO3367, ActR, and SCO4099 show different protection patterns on their cognate intergenic sequences. A DNA fragment containing the entire sequence of the *SGR5269*/*SGR5270*, *SCO3366*/*SCO3367*, *actR*/*actA*, or *SCO4098*/*SCO4099* intergenic region was exposed to DNase I in the presence of the indicated concentrations of the cognate TFR SGR5269 **(A)**, SCO3367 **(B)**, ActR **(C)**, or SCO4099 **(D)**. Two sequencing gels are shown for each TFR. For the left gel of each TFR, the primer that was extended toward the divergent neighboring gene was labeled at 5′-end to prepare the probe, while the other primer extended toward its own gene was labeled for the right gel. The regions protected by the TFRs are indicated by solid vertical lines. The numbers beside the lines indicate the start and end positions of the protected regions relative to the translational start site of the TFR-encoding gene.

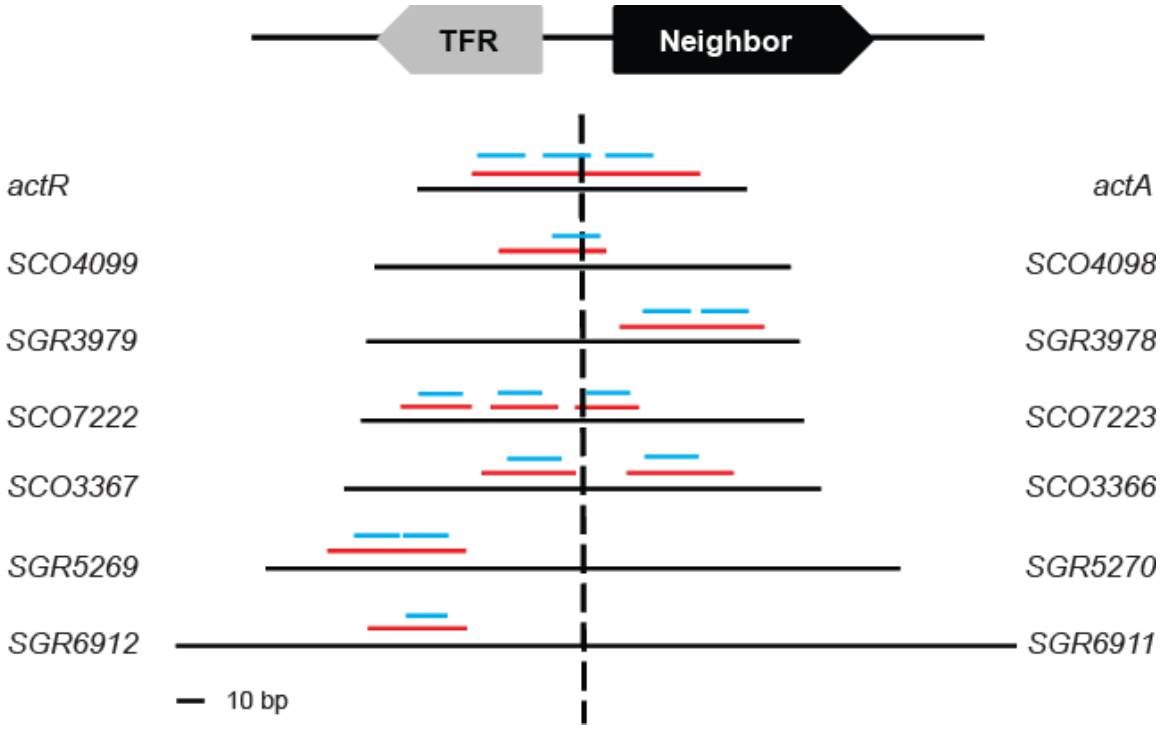

Figure 2.9. Seven TFRs bind different regions in the intergenic DNAs relative to their divergent neighbors. Solid black horizontal lines represent the intergenic DNAs, while red horizontal lines indicate the regions protected by the TFRs, on one or both strands of the DNAs. Putative operators of the TFRs were identified through sequence analysis of the protected regions (Table 2.5), and their positions are indicated by blue horizontal lines. These lines are oriented such that all of the TFR-encoding genes are located on the left side while their divergent neighboring genes are located on the right side. Dashed vertical line represents the center of the intergenic DNAs.

The operators of the known TFRs typically consist of 15 to 20 bp, which are either perfectly or partially palindromic. This is consistent with the fact that TFRs are known to dimerize in solution and each monomer likely interacts with half of the palindromic site. Taking advantage of this information, candidate operator sequences were identified within the regions protected by the seven TFRs with successful footprints (Table 2.5), and they correlated well with the

numbers of protein-DNA complexes observed by EMSA. For example, three perfect repeats of the consensus TGGAACGNCGTTCCA (Table 2.5) were found within the regions bound by SCO7222 (Figure 2.9) in the *SCO7222*/*SCO7223* intergenic region, consistent with the three protein-DNA complexes this TFR formed with the intergenic sequence (Figure 2.4D). Similarly, the region protected by ActR had three weaker palindromes (Figure 2.9) containing a direct or inverted sequence of CCACCGTT (Table 2.5), correlating well with the three shifts detected (Figure 2.4A).

Table 2.5. Putative operator sequences of the seven TFRs with successful footprints

TFR	Number of shifts observed by EMSA	Putative operator sequence[a]
ActR	3	GAACGGGCCACCGTTT
		CGCGACCACCGTTCCAT
		AGAACGGTGGTCGTTCG
SGR3979	2	TGCGTAATGCTTACGCA
		CGCGTATGGCATACGCA
SCO3367	2	ACTTGACGCCCGGCTAGT
		ACTTGCCGGGCGGCAAGT
SGR5269	2	TTGCGCAGTGGGCAA
		TTGCCCAGTGTGCAT
SCO4099	1	CACCTGTCGCACTAGTG
SCO7222	3	TGGAACGTCGTTCCA
		TGGAACGACGTTCCA
		TGGAACGCCGTTCCA
SGR6912	2	ACTAACCACTTAGT

[a] The palindromic nucleotides are italicized, while the repeated nucleotides are underlined.

The only exception was SGR6912, for which only one palindrome (Table 2.5) was identified within the protected region (Figure 2.9) in contrast to the two shifts detected with the $IGR_{SGR6912}$ probe by EMSA (Figure 2.5A). Interestingly, no effect was observed when this sequence was used in competition with the $IGR_{SGR6912}$ probe (data not shown), suggesting that this sequence does not

contain all the nucleotides required for efficiently interacting with SGR6912. The actual operator might therefore consist of an extended sequence (at 5'- and/or 3'-ends) capable of binding two protein dimers, possibly in a cooperative manner. Of note, only part of this putative operator sequence is conserved in the $IGR_{SGR6911}$ probe (missing the first three nucleotides of the putative operator shown in Table 2.5), within the region it overlaps with $IGR_{SGR6912}$. This might explain the considerably lower affinity SGR6912 has for this probe (Figure 2.5B) compared to the $IGR_{SGR6912}$ probe (Figure 2.5A). Therefore, the lack of protection by SGR6912 on $IGR_{SGR6911}$ observed in footprinting assays is likely due to the weakness of this interaction and/or the fact that the putative SGR6912 binding site is interrupted at end of the $IGR_{SGR6911}$ probe (where optimal resolution of protected region was not possible).

More importantly, TFRs having shorter intergenic sequences (i.e. ActR, SCO4099, SGR3979, SCO7222, and SCO3367) tended to bind the operators located proximal to both the TFR gene and the putative target, or to bind proximally to the putative target (Figure 2.9). In contrast, it is evident that the two TFRs with larger intergenic sequences (i.e. SGR5269 and SGR6912) bind the operator sequences that are distal from the divergent genes (Figure 2.9). These results, combined with the observation that SGR3402 did not interact with its intergenic sequence (Figure 2.6), suggest that the length of an intergenic sequence might be predictive of a regulatory relationship between a TFR and a divergently oriented gene.

2.3.4 Regulatory activity of the TFRs on their divergent neighbors

To biologically assess the regulatory activity of the nine TFRs on their neighboring genes, we used a *lux*-based system [228] to create *Streptomyces*-based transcriptional reporters. Two reporter plasmids were constructed for each TFR: a "promoter only" construct where expression of the *lux* operon is driven by the promoter of the divergently transcribed neighboring gene ("Without TFR" in Figure 2.10) and a second reporter where the TFR gene was included in *cis* ("With TFR" in Figure 2.10).

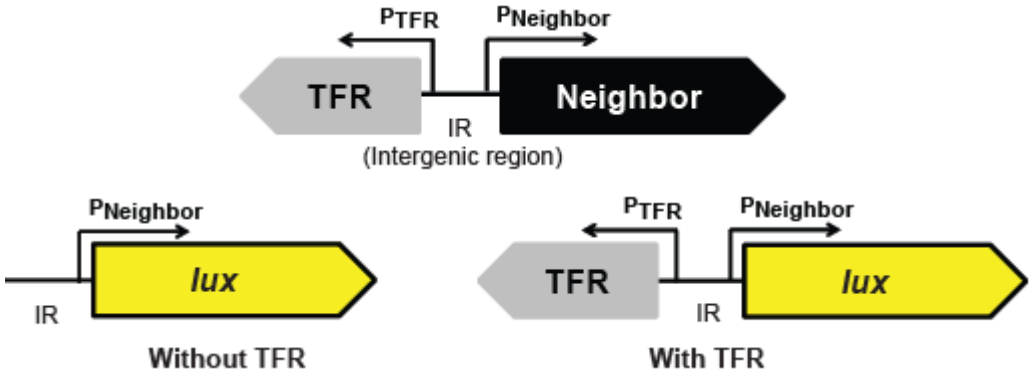

Figure 2.10. *Streptomyces*-based reporter construction. Two reporter plasmids were constructed for each TFR. For these plasmids, expression of the *lux* operon (would be driven by the promoter of the divergently transcribed neighboring gene ($P_{Neighbor}$) in the absence (Without TFR) or in the presence (With TFR) of the TFR-encoding gene in *cis*. The TFR expression would be driven by the natural promoter P_{TFR}.

To avoid possible interference from chromosomally encoded TFRs acting in *trans*, we introduced each of the reporters into a sequenced heterologous host. To choose an appropriate host for each reporter we used protein BLAST to identify a streptomycete that did not possess any TFR with 40% or greater protein sequence identity (Table 2.6). For each TFR, we introduced the two reporter constructs separately into a selected host and monitored growth and bioluminescence as a function of time. The only exception was AtrA, which occurs in all streptomycetes (Cuthbertson, unpublished data), and its reporters were introduced into the natural host *S. coelicolor*.

Table 2.6. Selected heterologous *Streptomyces* host for each TFR

TFR	Host	Top BLAST hit[a]
ActR	*S. venezuelae* ATCC 10712	SVEN3777 – 27(45)
SGR3979	*S. coelicolor* M145	SCO4358 – 39(52)
SCO3367	*S. albus* J1074	SSHG_05469 – 26(46)
SGR5269	*S. coelicolor* M145	SCO2374 – 36(50)
SGR3402	*S. coelicolor* M145	None
SCO4099	*S. sviceus* ATCC 29083	SSEG_10996 – 38(52)
SCO7222	*S. venezuelae* ATCC 10712	SVEN6489 – 38(53)
SGR6912	*S. coelicolor* M145	None
AtrA	*S. coelicolor* M145	AtrA – 100(100)

[a] Proteins in the selected hosts with at least 75% query coverage are indicated. The amino acid sequence identity and similarity (in the bracket) are shown.

As shown in Figure 2.11 and 2.12, luminescence from the “promoter only” constructs was greater than that of the promoterless vector control (3-fold to 197-fold at t=8 h) while growth rate was unchanged (data not shown). This suggested that the promoters of the divergent neighboring genes were active in the heterologous species.

Compared to the “promoter only” constructs, three different outcomes occurred when the cognate TFRs were expressed in *cis*. As expected, in the presence of ActR, luminescence from P_{actAB} was reduced 23-fold at t=8 h (Figure 2.11A). This is consistent with the previous studies showing that ActR represses the *actAB* promoter [123,128] and it validates our reporter system. Similar results were observed when SCO7222 (72-fold reduction, Figure 2.11B), SGR3979 (47-fold reduction, Figure 2.11C), SCO3367 (83-fold reduction, Figure 2.11D), and SCO4099 (33-fold reduction, Figure 2.11E) were expressed in *cis*.

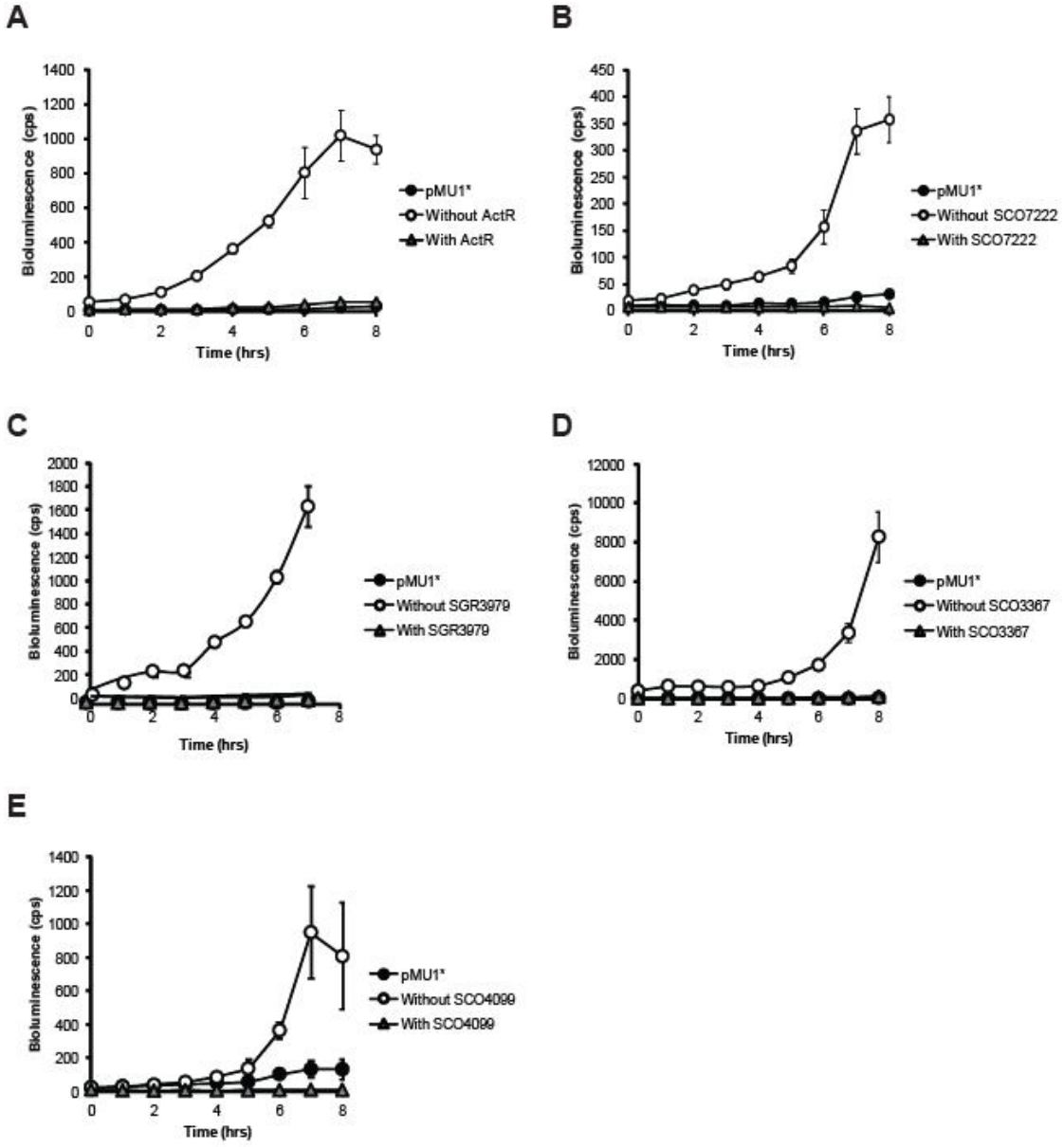

Figure 2.11. Five TFRs repress expression of their divergent neighbors. Two reporter plasmids constructed for ActR (**A**), SCO7222 (**B**), SGR3979 (**C**), SCO3367 (**D**), or SCO4099 (**E**) (as described in Figure 2.10) were introduced separately into a heterologous *Streptomyces* host for comparing their bioluminescence production as a function of time. Average bioluminescence values, measured in counts per second (cps), as well as +/- standard deviation of the values were obtained from at least three independent readings. Compared to the "Without TFR" constructs, reduced bioluminescence production was observed when the cognate TFRs were expressed in *cis*.

In contrast, AtrA appeared to enhance the *lux* expression by 4-fold compared to the “promoter only” construct (Figure 2.12A). These data suggest a role for AtrA in activating expression of its divergent neighboring gene *SCO4119* (encoding a putative NADH dehydrogenase), and it is consistent with the previously documented effect of AtrA as a transcriptional activator. In previous work this protein was shown to positively regulate the expression of *act*II-ORF4, which in turn activates the expression of genes involved in the biosynthesis of actinorhodin [93].

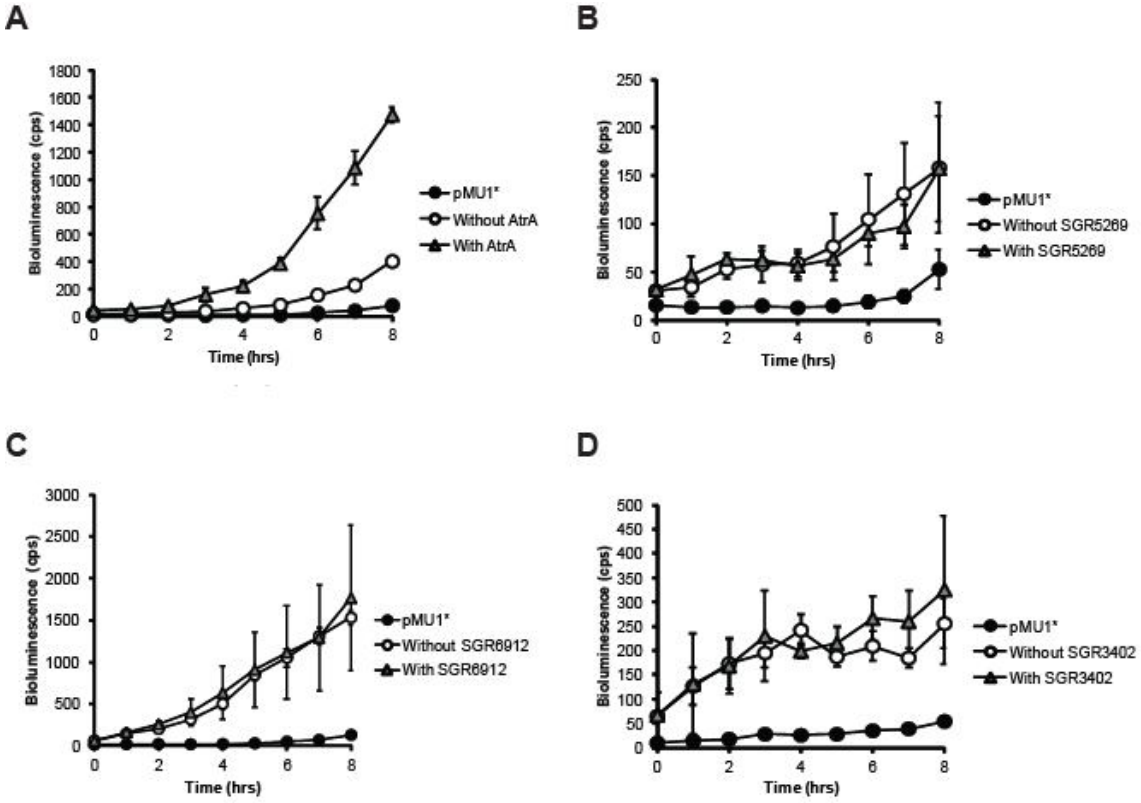

Figure 2.12. AtrA has a positive effect on the expression of its divergent neighbor while SGR5269, SGR6912, and SGR3402 have no effect. Two reporter plasmids constructed for AtrA (**A**), SGR5269 (**B**), SGR6912 (**C**), or SGR3402 **(D)** were introduced separately into a heterologous *Streptomyces* host for comparing their bioluminescence production as a function of time. Average bioluminescence values, measured in cps, as well as +/- standard deviation of the values were obtained from at least three independent readings. Compared to the "Without TFR" construct, increased bioluminescence (AtrA) or no change (SGR5269, SGR6912, and SGR3402) was observed when the cognate TFR was expressed in *cis*.

On the other hand, expression of SGR5269, SGR6912, and SGR3402 had no effect on luminescence as compared to their cognate "promoter only" constructs (Figure 2.12B to 2.12D). One possibility for this observation is lack of the TFR expression in the heterologous host under the conditions tested. To rule out this possibility, we constructed reporters where *lux* expression is driven by the promoter of the TFR itself. Luminescence from each of these reporters was above that of the vector control (data not shown). It could be speculated that these TFRs require ligands or co-regulator proteins to elicit activity and that these are not present; however, this is unlikely for SGR5269 and SGR6912 as they tightly bound their target DNAs *in vitro* without any addition of cofactor (Figure 2.4F and Figure 2.5A). Another possibility is that ligands of these TFRs are present in the selected host and they prevent the TFRs from binding the operators although this is unlikely for SGR3402 as it did not bind DNA *in vitro* without the presence of any added ligand (Figure 2.6). We have not ruled out these possibilities, however, the most likely explanation is that SGR5269, SGR6912, and SGR3402 do not regulate their divergent neighboring genes – *SGR5270*, *SGR6911*, and *SGR3403*, respectively. Therefore, the interactions of SGR5269 and SGR6912 with their intergenic DNA sequences *in vitro* likely indicate that these TFRs are autoregulatory and do not act as repressors or activators of the promoters of their divergent neighbors.

These reporter assays underscore the correlation between the length of the intergenic sequence and the regulatory activity of TFRs observed in our *in vitro* data (Figure 2.4 to 2.8). All five of the TFRs (ActR, SCO4099, SGR3979, SCO7222, and SCO3367) with intergenic sequences <200 bp repressed the promoters of their divergently oriented neighboring genes, like TetR. On the other hand, three (SGR5269, SGR6912, and SGR3402) of the four TFRs with the intergenic sequences >200 bp had no regulatory activity on their divergently transcribed neighboring genes while the fourth TFR, AtrA, activated expression.

No correlation was observed between the biochemical activity of the divergent gene product and the regulatory role of the adjacent TFR. ActR, SGR3979, and SCO3367 control expression of the genes encoding putative export pumps while SCO4099, SCO7222, and AtrA control expression of the genes encoding putative enzymes. These data support the idea that physiological processes under the regulation of TFRs are not limited to transporter-associated export of antibiotics but instead include a great diversity of enzymatic functions.

2.4 Discussion

The majority of the genes encoding TFRs (67%) in *S. coelicolor*, *S. griseus*, and *S. avermitilis* are transcribed divergently from an adjacent gene. The lengths of the intergenic DNA sequences separating the two genes are highly variable; however, in most cases the separation is less than 200 bp. Our data suggest that those TFRs having intergenic DNAs <200 bp are, in most cases, likely to be repressors of the divergent genes. As evidence for this, we have confirmed that ActR is a repressor of *actAB* and also demonstrated repression of *SCO4098*, *SGR3978*, *SCO7223*, and *SCO3366* by their cognate divergent TFR – SCO4099, SGR3979, SCO7222, and SCO3367, respectively. Consistent with our analysis, many previously characterized TFRs obey this "200 bp" rule, including EbrS (intergenic DNA=65 bp), EthR (75 bp), TetR (81 bp), SimR (138 bp), DesT (158 bp), QacR (177 bp), XdhR (188 bp), and LanK (190 bp) [50,85,87,95,99,136,229,230]. The prediction that such TFRs with intergenic sequences <200 bp will regulate adjacent genes is important because it means that at least one transcriptional target gene can be identified for more than half of all TFRs in the public databases, encompassing at least 25,000 distinct genes.

The regulatory prediction is less reliable for TFRs that are separated from divergent neighboring genes by >200bp; however, it is worth pointing out that our data do not rule out a classical, TetR-like regulatory relationship for these proteins and indeed, exceptions are known. For example, AtuR in *P. aeruginosa*, BpeR in *Burkholderia pseudomallei*, and Mce3R in *M. tuberculosis* are all TetR-

like repressors of divergent neighbors where the intergenic sequences are 280 bp, 409 bp, and 898 bp, respectively [231,232,233].

Surprisingly, while most previously characterized TFRs control the expression of export pumps, we find that most of the divergent genes encode putative enzymes: membrane-associated export proteins such as MFS (e.g. ActA and SCO3367) and ABC pumps (e.g. SGR3978) constitute less than 25% of the divergent gene products for TFRs that obey the “200 bp” rule. Importantly, the TFRs are likely, in most cases, repressors of the divergent enzyme-encoding genes. The variety of protein products of these genes is enormous and encompasses all known classifications of enzymes such as EC 1 oxidoreductases (e.g. SCO7223) and EC 2 transferases (e.g. SCO4098).

It is likely that some of these enzymes are involved in resistance mechanisms for antibiotics or other toxic molecules; however, we suggest that in many cases the biological roles are metabolic in nature. Indeed, it is important to note that catalytic mechanisms of the putative targets in Table 2.2 and Table 2.3 are predicted based on the limited information obtained solely from their sequences, but their biochemical and biological roles are completely unknown. As described in Chapter 1, an emerging paradigm suggests that in many cases small-molecule ligands of TFRs are related or identical to the substrates of their target gene products. Thus, we strongly believe that identifying ligands for TFRs of unknown functions promises to provide important biochemical and biological insights into these target genes.

2.5 Conclusion

Given the vast wealth of genome sequence information currently available, analysis of genome context can be a valuable method for predicting regulatory targets of TFRs. In this study, we have demonstrated that regulatory relationships are very likely when TFRs are separated from their divergently oriented neighboring genes by less than 200 bp. This “200 bp” rule is particularly useful due to the fact that more than half of the TFRs in various model organisms follow the rule and thus, we believe this trend might be observed in the majority of the organisms whose genomes encode multiple TFRs.

2.6 Materials and Methods

2.6.1 Genomic and bioinformatic analysis of TFRs

TFRs were identified using protein BLAST (blast.ncbi.nlm.nih.gov) with the consensus sequence for Hidden Markov Model (HMM) Pfam PF00440 (TetR_N). The genome context of individual TFRs was analyzed at StrepDB (streptomyces.org.uk) and National Center for Biotechnology Information (NCBI, www.ncbi.nlm.nih.gov), and each TFR was placed in three groups depending on their orientation to neighboring genes. TFRs divergently oriented to their immediate neighboring genes – regardless of the length of intergenic sequences between them – were placed in the first group. The second group contains TFRs that are predicted to be co-transcribed with their upstream and/or downstream genes when separated 35 bp or less, while the members in the last group lack

the aforementioned relationships with the adjacent genes. The protein products of the divergent neighboring genes were analyzed using protein BLAST as well as NCBI CD-Search to predict their functions.

2.6.2 Bacterial strains and culture conditions

Bacterial strains used in this study are described in Table 2.7. *E. coli* TOP10 was used during general cloning process, while *E. coli* BL21(DE3) was used for protein overexpression. All *E. coli* strains were grown using Luria Broth (LB) or LB agar medium, containing the appropriate antibiotics when required (50 μg/ml kanamycin and 50 μg/ml apramycin). *Streptomyces* cultures were grown using mannitol soy (MS), R2YE, R5, and maltose yeast extract with tap water (MYMTap) media [85,234].

Table 2.7. Strains used in this work

Strain	Description	Source or reference
Escherichia coli		
BL21(DE3)	F$^-$ *dcm ompT hsdS*(r_B^- m_B^-) *gal met* λ(DE3)	Novagen
Top10	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ψ80*lacZ*ΔM15 *nupG recA1 araD*139 Δ(*ara-leu*)7697 *galE*15 *galK*16 *rpsL*(Str^R) *endA1* λ$^-$	Invitrogen
ET12567/pUZ8002	ET12567 containing helper plasmid pUZ8002	[234]
Streptomyces		
S. coelicolor	M145 prototroph, SCP1- SCP2-	The John Innes Centre
S. venezuelae	ATCC 10712 prototroph	The John Innes Centre
S. sviceus	ATCC 29083 prototroph	Broad Institute
S. albus	J1074 prototroph	Broad Institute

2.6.3 Plasmids, primers, and sequencing

The plasmids and primers used in this study are listed in Table 2.8 and Table 2.9 respectively. Oligonucleotide primers used in this study were obtained from the Institute for Molecular Biology and Biotechnology (MOBIX) facility at McMaster University or from Sigma Aldrich. Polymerase chain reactions (PCR) were carried out using Vent DNA polymerase (New England Biolabs). DNA sequencing was carried out by the MOBIX facility.

Table 2.8. Plasmids used in this work

Plasmid	Description (selection marker)[a]	Background	Reference
pET28a	Vector for His_6-tagged protein overexpression (Kan^r)		Novagen
pET28a-ActR	ActR-overexpressing vector for protein purification (Kan^r)	pET28a	[128]
pET28a-SCO7222	SCO7222-overexpressing vector for protein purification (Kan^r)	pET28a	[235]
pET28a-SCOSCO3367	SCO3367-overexpressing vector for protein purification (Kan^r)	pET28a	This work
pET28a-SCO4099	SCO4099-overexpressing vector for protein purification (Kan^r)	pET28a	This work
pET28a-AtrA	AtrA-overexpressing vector for protein purification (Kan^r)	pET28a	This work
pET28a-SGR3979	SGR3979-overexpressing vector for protein purification (Kan^r)	pET28a	This work
pET28a-SGR5269	SGR5269-overexpressing vector for protein purification (Kan^r)	pET28a	This work
pET28a-SGR6912	SGR6912-overexpressing vector for protein purification (Kan^r)	pET28a	This work
pET28a-SGR3402	SGR3402-overexpressing vector for protein purification (Kan^r)	pET28a	This work
pMU1*	Promoterless *luxCDABE* operon, flanked by transcriptional terminators, preceded by in-frame stop codon and ribosome binding site, *aac(3)IV* ($Apra^r$)	pRT801	[228]
pMU1*-IGR(ActR)	The *actR/actA* intergenic region fused to *luxCDABE* ($Apra^r$)	pMU1*	This work
pMU1*-ActR-IGR	The *actR* gene and the intergenic region fused to *luxCDABE* ($Apra^r$)	pMU1*	This work
pMU1*-IGR(SCO7222)	The *SCO7222/SCO7223* intergenic region fused to *luxCDABE* ($Apra^r$)	pMU1*	This work
pMU1*-SCO7222-IGR	The *SCO7222* gene and the intergenic region fused to *luxCDABE* ($Apra^r$)	pMU1*	This work

pMU1*-IGR(SCO3367)	The *SCO336/SCO3367* intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work
pMU1*-SCO3367-IGR	The *SCO3367* gene and the intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work
pMU1*-IGR(SCO4099)	The *SCO4098/SCO4099* intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work
pMU1*-SCO4099-IGR	The *SCO4099* gene and the intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work
pMU1*-IGR(AtrA)	The *atrA/SCO4119* intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work
pMU1*-AtrA-IGR	The *atrA* gene and the intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work
pMU1*-IGR(SGR3979)	The *SGR3978/SGR3979* intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work
pMU1*-SGR3979-IGR	The *SGR3979* gene and the intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work
pMU1*-IGR(SGR5269)	The *SGR5269/SGR5270* intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work
pMU1*-SGR5269-IGR	The *SGR5269* gene and the intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work
pMU1*-IGR(SGR6912)	The *SGR6911/SGR6912* intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work
pMU1*-SGR6912-IGR	The *SGR6912* gene and the intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work
pMU1*-IGR(SGR3402)	The *SGR3402/SGR3403* intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work
pMU1*-SGR3402-IGR	The *SGR3402* gene and the intergenic region fused to *luxCDABE* (Apra^r)	pMU1*	This work

[a] Kan, kanamycin resistance and Apra, apramycin resistance

Table 2.9. Primers used in this work

Name	Sequence (5' → 3')[a]	Purpose
pETActR-F	GGGCATATGTCCCGAAGCGAGG	Forward primer for ActR over-expression
pETActR-R	GGATCCTCATGACTCCGCGGGG	Reverse primer for ActR over-expression
pET7222-F	GGCATATGGCATCCAGGT	Forward primer for SCO7222 over-expression
pET7222-R	GGATCCTACCGGCCGGGC	Reverse primer for SCO7222 over-expression
pET3367-F	CCGGAATTCAACGGCACCAAGCAGCAGCGA	Forward primer to amplify the *SCO3367* gene to construct pET28a-SCO3367
pET3367-R	CGCGAAGCTTTGCTGTGACGGCGCCCCCTA	Reverse primer to amplify the *SCO3367* gene to construct pET28a-SCO3367
pET4099-F	CCGGAATTCATGACGGATTCGACGGATTCGA	Forward primer to amplify the *SCO4099* gene to construct pET28a-SCO4099
pET4099-R	TTTCGCTCGAGTCACGCGCCCTCCGTCCGGGGGA	Reverse primer to amplify the *SCO4099* gene to construct pET28a-SCO4099
pET3979-F	CCGGAATTCGTGACGGAGAAGAACGGCT	Forward primer to amplify the *SGR3979* gene to construct pET28a-SGR3979
pET3979-R	TTTCGCTCGAGTCAAGCGGACTTGACCTCGCCGT	Reverse primer to amplify the *SGR3979* gene to construct pET28a-SGR3979
pET5269-F	CCGGAATTCATGGACAGCAGCACCGCGGGA	Forward primer to amplify the *SGR5269* gene to construct pET28a-SGR5269
pET5269-R	TTTCGCTCGAGTCACGCGAACCCCCGCCCCAT	Reverse primer to amplify the *SGR5269* gene to construct pET28a-SGR5269
pET6912-F	CCGGAATTCGTGAGCAACCAGTCAGTGAGCA	Forward primer to amplify the *SGR6912* gene to construct pET28a-SGR6912
pET6912-R	TTTCGCTCGAGTCAGCCTGCGGTGCGCGGGCCGA	Reverse primer to amplify the *SGR6912* gene to construct pET28a-SGR6912
pET3402-F	AAAAACATATGAGCGCGCGGGACACCGGG	Forward primer to

		amplify the *SGR3402* gene to construct pET28a-SGR3402
pET3402-R	AAAAAGGATCCTCAGCGACGGCGGCCGGC	Reverse primer to amplify the *SGR3402* gene to construct pET28a-SGR3402
ActIR-1	GTGCTCCTCATCGTATGGCATGAACG	Forward primer to amplify the intergenic region between *actR* and *actA*
ActIR-2	GGCGTCCCCCGGGTCCTC	Reverse primer to amplify the intergenic region between *actR* and *actA*
7222IR-F	GCGCTCATCCTAGACC	Forward primer to amplify the intergenic region between *SCO7222* and *SCO7223*
7222IR-R	GGTGCTTCCCTCCTGA	Reverse primer to amplify the intergenic region between *SCO7222* and *SCO7223*
3367IR-F	CTCTCCTACTCCCCCAGC	Forward primer to amplify the intergenic region between *SCO3366* and *SCO3367*
3367IR-R	CGTGCCGCCCATCCTCCT	Reverse primer to amplify the intergenic region between *SCO3366* and *SCO3367*
4099IR-F	CGGGTGGAGCACGGTGGGAT	Forward primer to amplify the intergenic region between SCO*4098* and *SCO4099*
4099IR-R	CGGCCGTCCACTATCCGTTTGA	Reverse primer to amplify the intergenic region between SCO*4098* and *SCO4099*
AtrAIR1-F	AGACAATCCCCCGGTAATGA	Forward primer to amplify the IGR_{atrA} sequence, partial intergenic region between *atrA* and *SCO4199*
AtrAIR1-R	CGGAACGTGCCCCTTCTGAA	Reverse primer to amplify the IGR_{atrA} sequence, partial intergenic region between *atrA* and *SCO4199*
AtrAIR2-F	CTACACACCGTGGGAAAGAA	Forward primer to amplify the IGR_{centre}

		sequence, partial intergenic region between *atrA* and *SCO4199*
AtrAIR2-R	AGCCGCGGGAACCAACGTCG	Reverse primer to amplify the IGR_{centre} sequence, partial intergenic region between *atrA* and *SCO4199*
AtrAIR3-F	CACTCACCCACGCTGACC	Forward primer to amplify the $IGR_{SCO4119}$ sequence, partial intergenic region between *atrA* and *SCO4199*
AtrAIR3-R	CCCGACGTAGCCACCGCC	Reverse primer to amplify the $IGR_{SCO4119}$ sequence, partial intergenic region between *atrA* and *SCO4199*
3979IR-F	GGTTCGGATCTCCTTTTGCGT	Forward primer to amplify the intergenic region between *SGR3978* and *SGR3979*
3979IR-R	CGCATCCCCTTCGCCCACCGT	Reverse primer to amplify the intergenic region between *SGR3978* and *SGR3979*
5269IR-F	GGGTGCGGCGGCGGGGGCCCCG	Forward primer to amplify the intergenic region between *SGR5269* and *SGR5270*
5269IR-R	GCGGGCATTATCCCCCGGA	Reverse primer to amplify the intergenic region between *SGR5269* and *SGR5270*
6912IR1-F	GGCGAGGACGAGTGCGGCGG	Forward primer to amplify the $IGR_{SGR6911}$ sequence, partial intergenic region between *SGR6911* and *SGR6912*
6912IR1-R	ACCACTTAGTGGGTGTGCTTTCATGCC	Reverse primer to amplify the $IGR_{SGR6911}$ sequence, partial intergenic region between *SGR6911* and *SGR6912*
6912IR2-F	GGTTCGTACCTGCTGACAGG	Forward primer to amplify the $IGR_{SGR6912}$ sequence, partial

		intergenic region between *SGR6911* and *SGR6912*
6912IR2-R	CGGTTCACCTTAAGCCTGCC	Reverseprimer to amplify the $IGR_{SGR6912}$ sequence, partial intergenic region between *SGR6911* and *SGR6912*
3402IR1-F	GCGGTATGGCCCCGCCGG	Forward primer to amplify the $IGR_{SGR3403}$ sequence, partial intergenic region between *SGR3402* and *SGR3403*
3402IR1-R	CCGGGCCTCCGGTCTTCG	Reverse primer to amplify the $IGR_{SGR3403}$ sequence, partial intergenic region between *SGR3402* and *SGR3403*
3402IR2-F	GGGTCGAAGACCGGAGGC	Forward primer to amplify the $IGR_{centre1}$ sequence, partial intergenic region between *SGR3402* and *SGR3403*
3402IR2-R	GCCGGATACCCACCCCCG	Reverse primer to amplify the $IGR_{centre1}$ sequence, partial intergenic region between *SGR3402* and *SGR3403*
3402IR3-F	GGGGCTGTCATCGGCCGC	Forward primer to amplify the $IGR_{centre2}$ sequence, partial intergenic region between *SGR3402* and *SGR3403*
3402IR3-R	AACGATGGCCGGCGGCGG	Reverse primer to amplify the $IGR_{centre2}$ sequence, partial intergenic region between *SGR3402* and *SGR3403*
3402IR4-F	GCCGCCGGCCATCGTTTT	Forward primer to amplify the $IGR_{SGR3402}$ sequence, partial intergenic region between *SGR3402* and *SGR3403*
3402IR4-R	GGCCGGAGATCCCTTCTG	Reverse primer to amplify the $IGR_{SGR3402}$

		sequence, partial intergenic region between *SGR3402* and *SGR3403*
ActRIGR-F	AAAAAAGATATCGTGCTCCTCATCGTATGGCATGAACGGG	Forward primer to construct pMU1*-IGR(ActR)
ActRIGR-R	AAAAAAGGATCCGGCGTCCCCCGGGTCCTCGACTATT	Reverse primer to construct pMU1*-IGR(ActR) and pMU1*-ActR-IGR
ActR-F	AAAAAAGATATCTCATGACTCCGCGGGGGGCGATCC	Forward primer to construct pMU1*-ActR-IGR
SCO7222IGR-F	AAAAAAGGATCCGCGCTCATCCTAGACCGCTT	Forward primer to construct pMU1*-IGR(SCO7222)
SCO7222IGR-R	AAAAAAGGTACCGGTGCTTCCCTCCTGATGGC	Reverse primer to construct pMU1*-IGR(SCO7222) and pMU1*-SCO7222-IGR
SCO7222-F	AAAAAAGATATCGAAGGCGGTCGGCCGGGACC	Forward primer to construct pMU1*-SCO7222-IGR
SCO3367IGR-F	AAAAAAGATATCCGTGCCGCCCATCCTCCTAC	Forward primer to construct pMU1*-IGR(SCO3367)
SCO3367IGR-R	AAAAAAGGATCCCTCTCCTACTCCCCCAGCTA	Reverse primer to construct pMU1*-IGR(SCO3367) and pMU1*-SCO3367-IGR
SCO3367-F	AAAAAAGATATCCTACGCGCCCGCGGCTCC	Forward primer to construct pMU1*-SCO3367-IGR
SCO4099IGR-F	AAAAAAGATATCCGGCCGTCCACTATCCGTTTGATCG	Forward primer to construct pMU1*-IGR(SCO4099)
SCO4099IGR-R	AAAAAAGGTACCCGGGTGGAGCACGGTGGGAT	Reverse primer to construct pMU1*-IGR(SCO4099) and pMU1*-SCO4099-IGR
SCO4099-F	AAAAAAGATATCTCACGCGCCCTCCGTCCGGG	Forward primer to construct pMU1*-SCO4099-IGR
AtrAIGR-F	AAAAAAGATATCAGACAATCCCCCGGTAATGACGTCTCCCC	Forward primer to construct pMU1*-IGR(AtrA)
AtrAIGR-R	AAAAAAGGATCCCGACGTAGCCACCGCCGA	Reverse primer to construct pMU1*-IGR(AtrA) and pMU1*-AtrA-IGR
AtrA-F	AAAAAAGATATCTCACACCGGCCGCGACCGCAG	Forward primer to construct pMU1*-AtrA-

		IGR
SGR3979IGR-F	AAAAAAGATATCCGCATCCCCTTCGCCCACCG	Forward primer to construct pMU1*-IGR(SGR3979)
SGR3979IGR-R	AAAAAAGGATCCGGTTCGGATCTCCTTTTGCG	Reverse primer to construct pMU1*-IGR(SGR3979) and pMU1*-SGR3979-IGR
SGR3979-F	AAAAAAGATATCTCAAGCGGACTTGACCTCGC	Forward primer to construct pMU1*-SGR3979-IGR
SGR5269IGR-F	AAAAAAGATATCGCGGGCATTATCCCCCGGAGGCGAG	Forward primer to construct pMU1*-IGR(SGR5269)
SGR5269IGR-R	AAAAAAGGATCCGGGTGCGGCGGCGGGGGC	Reverse primer to construct pMU1*-IGR(SGR5269) and pMU1*-SGR5269-IGR
SGR5269-F	AAAAAAGATATCACGCGAACCCCCGCCCCATCTCGTCCAGG	Forward primer to construct pMU1*-SGR5269-IGR
SGR6912IGR-F	AAAAAAGATATCCGGTTCACCTTAAGCCTGCCGGAGCCC	Forward primer to construct pMU1*-IGR(SGR6912)
SGR6912IGR-R	AAAAAAGGATCCGGCGAGGACGAGTGCGGCGG	Reverse primer to construct pMU1*-IGR(SGR6912) and pMU1*-SGR6912-IGR
SGR6912-F	AAAAAAGATATCTCAGCCTGCGGTGCGCGGGCCG	Forward primer to construct pMU1*-SGR6912-IGR
SGR3402IGR-F	AAAAAAGATATCGGCCGGAGATCCCTTCTGTCGAGTCCGC	Forward primer to construct pMU1*-IGR(SGR3402)
SGR3402IGR-R	AAAAAAGGATCCGCGGTATGGCCCCGCCGGCC	Reverse primer to construct pMU1*-IGR(SGR3402) and pMU1*-SGR3402-IGR
SGR3402-F	AAAAAAGATATCTCAGCGACGGCGGCCGGCGACCC	Forward primer to construct pMU1*-SGR3402-IGR

[a] Restriction endonuclease recognition sequences introduced by these oligonucleotides are underlined.

2.6.4 Expression and Purification of His_6-tagged TFRs

S. coelicolor and *S. griseus* chromosomal DNAs were used as templates to PCR amplify the DNA fragments containing the *actR*, *SCO7222*, *SCO3367*, *SCO4099, atrA* (*SCO4118*), *SGR3402, SGR3979, SGR5269,* and *SGR6912* open reading frames which were introduced separately into pET28a, giving pET28a-ActR, pET28a-SCO7222, pET28a-SCO3367, pET28a-SCO4099, pET28a-AtrA, pET28a-SGR3402, pET28a-SGR3979, pET28a-SGR5269, and pET28a-SGR6912, respectively.

E. coli BL21(DE3) cultures containing individual vectors were grown at 37°C to an OD_{600} of 0.4-0.6 and TFR expression was induced through addition of 1 mM isopropyl β-D-thiogalactopyranoside for 3 to 5 hours at 37°C. Cells were collected by centrifugation at 2,700xg for 15 min at 4°C in the Sorvall SLA-3000 rotor and lysed using the BugBuster reagent (Novagen). The lysate was cleared by centrifugation at 17,200xg for 30 min at 4°C in the Sorvall SS-34 rotor and filtered through a 0.45 μm filter to remove smaller debris and insoluble protein. 4 mL of QIAGEN Ni-NTA agarose solution was added to the filtered lysate and the mixture was allowed to incubate for 1 h at 4°C with gentle shaking. The column was washed with 50 mM Tris, pH 7.9, 0.5 M NaCl, 20 mM imidazole and eluted in 50 mM Tris, pH 7.9, 0.5 M NaCl, 1 M imidazole. Elution fractions were monitored by SDS-PAGE. Fractions containing a TFR were pooled and exchanged into buffer C (20 mM Tris, pH 7.9, 0.5 M NaCl, 20% v/v glycerol). The

desalted protein was concentrated using an Amicon Ultra Centrifugal Filter (10,000 MWCO; Millipore).

2.6.5 EMSAs

S. coelicolor and *S. griseus* chromosomal DNA templates were used in PCR reactions to isolate double-stranded DNA fragments containing the intergenic sequences – between *actR* (*SCO5082*) and *actA* (*SCO5083*); *SCO7222* and *SCO7223*; *SCO3366* and *SCO3367*; *SGR3978* and *SGR3979*; *SGR5269* and *SGR5270* – which served as the probes for ActR, SCO7222, SCO3367, SGR3979, and SGR5269, respectively in the assays. The probes for AtrA, SGR3402, and SGR6912 were prepared by obtaining the DNA fragments (148 bp to 200 bp in lengths) containing different regions within their intergenic sequences – between *atrA* (*SCO4118*) and *SCO4119*; *SGR3402* and *SGR3403*; *SGR6911* and *SGR6912* respectively – with partially overlapped ends. The DNA fragments were 5′-end labeled using [γ-^{32}P] ATP (PerkinElmer) and T4 polynucleotide kinase (New England Biolabs).

A labeled probe (1 ng), varying amounts of a purified protein, and 90 ng of salmon sperm DNA (Sigma-Aldrich) were used in 15 µl reactions containing 1x EMSA reaction buffer (10 mM Tris-Cl (pH 7.8), 150 mM NaCl, 2 mM DTT and 10% glycerol). Reactions were incubated at 30°C for 10 minutes and were fractionated on 12% non-denaturing polyacrylamide gels containing 1.5% glycerol. The gels

were exposed using a phosphor screen (Amersham) and bands were detected using a PhosphorImager (Molecular Dynamics).

2.6.6 DNase I footprinting assays

The same pairs of primers to amplify the intergenic sequences in the previous EMSAs were used for DNase I footprinting. The probes in the assays were prepared by PCR using one unlabeled primer and one 5′-end labeled primer (using [γ-^{32}P] ATP and T4 polynucleotide kinase). 150,000 cpm of a labeled DNA probe, varying amounts of a purified protein, and 90 ng of salmon sperm DNA (Sigma-Aldrich) were used in 40 µl reactions containing 1x EMSA reaction buffer. After the reactions were incubated at 30°C for 10 minutes, 10 µl DNase I solution (1 U in 10 mM $CaCl_2$) was added. The incubation was continued for 60 seconds at room temperature and reactions were stopped by adding 140 µl DNase I stop solution (200 mM NaCl, 30 mM EDTA, and 1% SDS). The digested samples were then precipitated with ethanol and resuspended in 5 µl Stop Solution (from Thermosequenase Cycle Sequencing Kit (USB): 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). Samples were heated at 80°C for 3 minutes, cooled on ice, and separated on 8% polyacrylamide/7 M urea sequencing gels. Dried gels were exposed using a phosphor screen (Bio-Rad) and bands were detected using a PhosphorImager (Molecular Dynamics). Sequencing ladders were prepared using Thermosequenase Cycle Sequencing Kit (USB).

2.6.7 Construction of *lux*-based reporter plasmids and bioluminescence measurements

Two reporter plasmids were constructed for each TFR of interest (Table 2.8). For the first, a DNA fragment containing the intergenic sequence between a TFR of interest and its divergent neighbor gene was cloned into pMU1* [228] in an orientation such that *lux* expression was driven by the promoter of the divergent neighbor (Figure 2.10). The second construct had a DNA fragment containing the TFR gene as well as its intergenic sequence introduced to pMU1* in the same orientation as the first. In this construct, the TFR gene was transcribed by its natural promoter in the opposite direction to the *lux* operon (Figure 2.10).

Host organisms for the reporters were designated by using protein BLAST to identify a streptomycete that does not possess any possible ortholog of the selected TFR (at least 40% identity in the amino acid sequence with at least 75% query coverage). 2 x 10^7 colony forming units of the *Streptomyces* reporter spores were inoculated and grown for 16 hours to 20 hours. The overnight grown cells were then subcultured to set the starting OD (OD_{450} for *S. coelicolor* and OD_{600} for the other streptomycetes) at 0.05 (t=0), and the cultures were measured for bioluminescence and OD every hour using VICTOR^{TM} X Light 2030 luminescence reader (PerkinElmer) and Epoch microplate spectrophotometer (BioTek), respectively.

CHAPTER 3: DEVELOPMENT OF TWO-PLASMID WHOLE CELL-BASED REPORTER SYSTEM FOR INVESTIGATING TETR FAMILY TRANSCRIPTIONAL REGULATORS

This chapter was adapted from:

Ahn, S.K., Tahlan, K., Yu, Z., and Nodwell, J.R. 2007. Investigation of transcription repression and small-molecule responsiveness by TetR-like transcription factors using a heterologous *Escherichia coli*-based assay. J. Bacteriol. **189**:6655-6664.

Also, partially adapted from:

Tahlan, K., Ahn, S.K., Sing, A., Bodnaruk, T.D., Willems, A.R., Davidson, A.R., and Nodwell, J.R. 2007. Initiation of actinorhodin export in *Streptomyces coelicolor*. 2007. Mol. Microbiol. **63**:951-961.

Author Contributions:

The TetR biosensor was constructed by TDB. The ActR biosensors were constructed by AS and SKA. Concentrations of tetracycline derivatives required for half maximum induction of TetR were measured by AS. The culture supernatant of *Streptomyces coelicolor* was prepared by AS and KT. All other experiments were conducted by SKA.

3.1 Abstract

SCO7222 and ActR are TetR family transcriptional regulators (TFRs) encoded in the *Streptomyces coelicolor* genome. We have developed *Escherichia coli*-based biosensors that accurately report the interactions of the two TFRs with their cognate operator DNAs and induction of ActR by its cognate ligand(s). We find that SCO7222 and ActR bind to the operator sequences (previously predicted in Chapter 2) in the intergenic regions between their genes and their putative target genes *SCO7223* and *actA*, respectively. The operators recognized by these TFRs are related such that O^{7223} (an operator for SCO7222) is bound tightly by both SCO7222 and ActR. In contrast, O^{act} (an operator for ActR) is bound tightly by ActR and more weakly by SCO7222. We demonstrate ligand specificity of these proteins by showing that while TetR (but not ActR or SCO7222) interacts with tetracyclines, ActR (but not TetR or SCO7222) responds to the *S. coelicolor* culture supernatant. Through operator-targeted mutagenesis, we find that at least two nucleotide changes in O^{7223} are often required to disrupt its interaction with SCO7222, while ActR is more sensitive to changes on O^{act}. Most importantly, the interaction of each protein with the wild-type and mutant operator sequences in *E. coli*-based reporters and *in vitro* assays correlates perfectly. We believe that our biosensor system should be broadly applicable to TFRs.

3.2 Introduction

In the previous chapter, we have described the use of genome context as a predictive tool to identify regulatory targets of TetR family transcriptional regulators (TFRs). In order to assign biochemical functions to these target gene products, simple techniques that can facilitate identification and analysis of the small-molecule ligands that interact with TFRs are desirable. The relational framework we have described (Cuthbertson *et al*, manuscript submitted) is one way of predicting ligands of many TFRs; however, no such prediction is possible for all members of the family.

While *Streptomyces*-based reporters described in Chapter 2 (Figure 2.10) are one approach for ligand discovery, streptomycetes are not always the most ideal choice for constructing TFR reporters. Most of their genomes are rich in TFR-coding genes and therefore, potential nonspecific activity by these proteins on the reporters in *trans* can be a concern. In addition, slow growth and the complicated lifecycle exhibited by *Streptomyces* [83] make these organisms less attractive for reporter strain preparation. Therefore, methods that can be applied to TFRs in a genome scale using facile model prokaryotes such as *Escherichia coli* would also be beneficial. Importantly, it is highly desirable to develop systems that are amenable to high-throughput screening, which is a valuable technique when virtually nothing is predicted or known regarding the identity of small-molecule signals.

In this study, we have developed an *E. coli*-based biosensor mechanism as the first attempt at such a reporter system. This two-plasmid biosensor mechanism (Figure 3.1) is based on a synthetic promoter consisting of the -10 and -35 sequence elements, separated by a putative binding site for a TFR. These promoters are cloned upstream of the *lux* operon. The gene encoding the cognate TFR is cloned into a second plasmid such that cells containing the *lux* plasmid are spontaneously bioluminescent while those containing both are not, as the TFR-operator interaction should provide a physical barrier to prevent RNA polymerase from recognizing and interacting with the promoter.

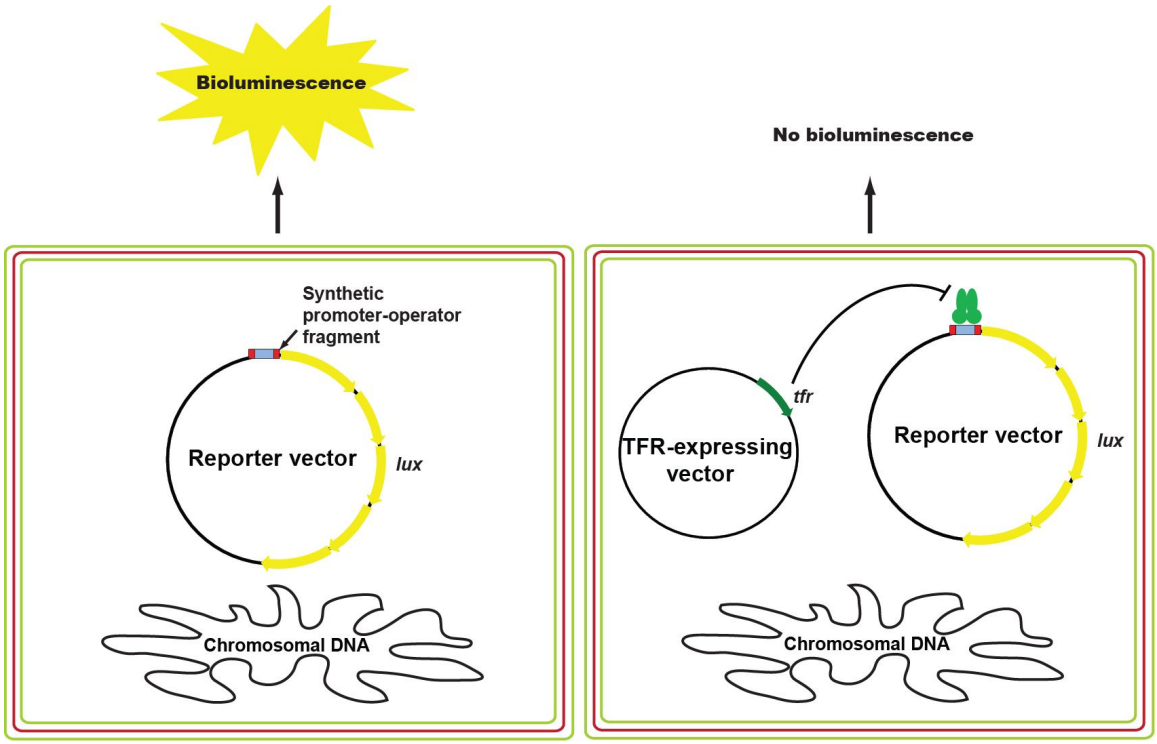

Figure 3.1. Schematic diagram of a two-plasmid TFR biosensor. A reporter vector is constructed by inserting a synthetic promoter fragment – containing the -10 and -35 elements (red boxes) surrounding a putative binding site (light blue box) for a TFR of interest – upstream of the *lux* operon (yellow arrows). Cells containing the reporter vector produce spontaneous bioluminescence (left). However, this bioluminescence production will be repressed when an additional vector expressing the cognate TFR is introduced to the cells (right, TFR represented in green).

We have explored the utility of this biosensor mechanism in great depth focusing on two TFRs from *Streptomyces coelicolor* – SCO7222 and ActR – using TetR as a control. In Chapter 2, we have demonstrated that the target of SCO7222 is the divergently transcribed gene *SCO7223* which encodes a probable monooxygenase. Similarly, ActR regulates the divergently transcribed *actAB* operon encoding putative efflux pumps for an antibiotic actinorhodin [123,128].

We show that SCO7222 and ActR interact with the palindromic operators previously predicted (Table 2.5) in the *SCO7222*/*7223* and *actR*/*actA* intergenic sequences, respectively. The operator sequences bound by ActR and SCO7222 are related and indeed, the two proteins exhibit considerable affinities for each other's operators. Our biosensors show that while TetR responds to tetracycline and other structurally-related molecules, culture supernatant from *S. coelicolor* relieves ActR-based repression. SCO7222 does not respond to the ligands recognized by either ActR or TetR. Mutagenesis of the binding sites for SCO7222 and ActR reveals that the two proteins recognize distinct nucleotides and, most importantly, that the DNA/protein interactions observed in our biosensors and *in vitro* perfectly correlate. We suggest therefore that this biosensor mechanism is likely to be broadly applicable to other TFRs.

3.3 Results

3.3.1 Identification of the binding sites for ActR and SCO7222

Based on the previous footprint data (Figure 2.7B) and sequence analysis (Table 2.5) of the protected regions, we predicted that the likely operators of SCO7222 would be the three sequences exhibiting significant similarity to the nearly perfect palindrome with the consensus sequence C/TTGGAACGNCGTTC CAG/C (N1, N2, and N3 in Figure 3.2A). This is similar to the intergenic DNA between *actR* and *actA* that includes the putative ActR-binding sequences (Figure 2.8C and Table 2.5) P1, P2, and P3 that are weaker palindromes but contain the direct or inverted sequence of CCACCGTT (Figure 3.2A).

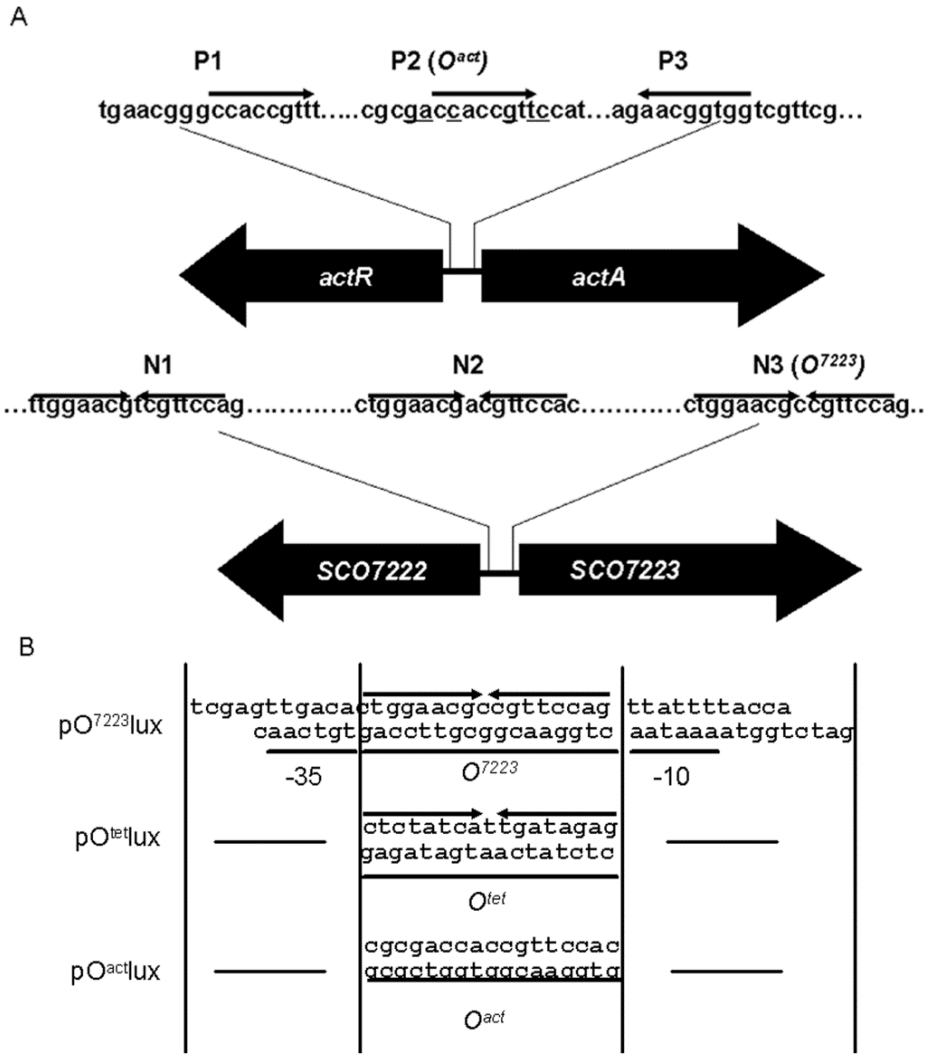

Figure 3.2. Putative binding sites for ActR and SCO7222. (A) Repeated elements in the *actR*/*actA* and *SCO7222*/*SCO7223* intergenic regions are shown. Between *actR* and *actA*, there are three weakly palindromic sequences (underlined) that exhibit a low degree of conservation. Between *SCO7222* and *SCO7223* there are three 17 bp-repeats that are perfectly palindromic and highly conserved. Arrows indicate repeated sequence (*actR/actA*) and palindromes (*SCO7222/7223*). **(B)** Sequence of the synthetic promoter used in the pO^{7223}lux, pO^{tet}lux, and pO^{act}lux. DNA fragment contained -10 and -35 regions from the Tn*10 tetA* promoter, flanking the known or putative operator sequence of the respective TFR. Arrows indicate the palindromic nucleotides in O^{7223} and O^{tet}.

3.3.2 Construction of biosensors using TetR, ActR, and SCO7222

We designed synthetic promoters in which each of N1, N2 and N3 (O^{7223}) was introduced between functional *E. coli* -10 and -35 promoter elements (one with O^{7223} shown in Figure 3.2B), and introduced them upstream of the *lux* operon in the vector pCS26-*Pac*, resulting in the plasmids pN1lux, pN2lux, and pO^{7223}lux, respectively. Similarly, the putative reporter plasmids for ActR – pP1lux, pO^{act}lux, and pP3lux – were constructed using the P1, P2 (O^{act}), and P3 sequences (one with O^{act} shown in Figure 3.2B) respectively. These plasmids were introduced into *E. coli* along with pO^{tet}lux (reporter vector for TetR containing the known TetR operator, Figure 3.2B) and all conferred significant bioluminescence (175-fold to 487-fold higher) compared to promoterless reporter vector pCS26-*Pac* (Table 3.1, note that only the values from pO^{7223}lux and pO^{act}lux are shown for the SCO7222 and ActR reporters, respectively), confirming that the synthetic promoter fragments are functional.

Table 3.1. Effects of TetR/ActR/SCO7222 on various reporter plasmids

Plasmids[a]	Bioluminescence[b]
pCS26-*Pac*	318
$pO^{tet}lux$	1.55×10^5
$pO^{tet}lux$ + pACYC184	1.33×10^5
$pO^{tet}lux$ + pTetR	368
$pO^{tet}lux$ + pActR	1.31×10^5
$pO^{tet}lux$ + pSCO7222	1.24×10^5
$pO^{7223}lux$	8.87×10^4
$pO^{7223}lux$ + pACYC184	5.70×10^4
$pO^{7223}lux$ + pSCO7222	262
$pO^{7223}lux$ + pTetR	8.22×10^4
$pO^{7223}lux$ + pActR	360
$pO^{act}lux$	5.58×10^4
$pO^{act}lux$ + pACYC184	3.46×10^4
$pO^{act}lux$ + pActR	212
$pO^{act}lux$ + pTetR	4.22×10^4
$pO^{act}lux$ + pSCO7222	1.86×10^4

[a] Analysis was carried out using *E. coli* DH5α as the host.
[b] All values are in relative light units (RLUs) and represent the average of at least three independent readings.

We constructed TetR-, ActR-, and SCO7222-producing plasmids by amplifying the respective genes with oligonucleotides that included a Shine-Dalgarno sequence upstream of the initiator codon. These fragments were ligated downstream of the *tetA* promoter in the vector pACYC184 to create the plasmids pTetR, pActR, and pSCO7222. We introduced these expression vectors and the control pACYC184 into the *E. coli* strains bearing $pO^{tet}lux$, $pO^{act}lux$ and $pO^{7223}lux$, and tested each strain for bioluminescence.

As expected, pTetR eliminated pO^{tet}lux-dependent bioluminescence, while the control plasmid pACYC184 (as well as pActR and pSCO7222) had no effect (Table 3.1). This result is consistent with the fact that TetR specifically interacts with the O^{tet} sequence, validating our biosensor system. Similarly, pSCO7222 eliminated the bioluminescence produced by pN1lux, pN2lux (data not shown), and pO^{7223}lux (Table 3.1). The function of these three sites in our biosensor system agrees with the mobility shift demonstrating three complexes between SCO7222 and the *SCO7222/SCO7223* intergenic sequence and we focused the rest of our work on N3, which we refer to O^{7223}.

On the other hand, pActR eliminated pO^{act}lux-dependent bioluminescence (Table 3.1) while it had no such effect on the bioluminescence produced by pP1lux and pP3lux (data not shown). Although further investigation is required to fully understand the exact identity of the other ActR binding site(s) in the *actR*/*actA* intergenic region (three binding sites predicted based on the electrophoretic mobility shift assays, EMSAs, shown in Figure 2.4A), we successfully identified one operator sequence for ActR, which we refer to O^{act}, to be used for further studies.

We noticed that the operators recognized by ActR (CGCGACCACCGTTC CAT) and SCO7222 (CTGGAACGACGTTCCAG) were similar, particularly in their right half-sites. To determine whether each repressor could bind the other's operator, we combined pActR with pO^{7223}lux biosensor and pSCO7222 with pO^{act}lux biosensor. We observed strong repression of pO^{7223}lux-dependent

bioluminescence by pActR (to background) and partial repression of pO^{act}lux-dependent bioluminescence by pSCO7222 (1.5-fold) (Table 3.1).

3.3.3 Effects of various ligands on TetR, ActR, and SCO7222

To determine whether tetracycline could relieve repression by TetR, ActR, and SCO7222, we introduced plasmid pUAO1, bearing the *tetO* gene from *Campylobacter jejuni* [236], to each biosensor strain to protect them against the antibiotic. The TetO gene product is a ribosomal protection protein that allows relatively high cytoplasmic concentrations of the antibiotic to be tolerated without substantial damage to the cells [40]. The addition of pUAO1 had no major impact although it slowed bacterial growth to ~1 doubling per hour (data not shown).

To determine whether the resulting strains could detect tetracycline, we cultured each of them in the presence of varying concentrations of the drug (Figure 3.3A). We observed a dose-responsive induction of bioluminescence in the TetR biosensor that peaked when 4 µg/ml of tetracycline was added. Higher drug concentrations inhibited cell growth, in spite of the presence of the *tetO* gene, but even at levels where there was significant inhibition of growth, we could still detect bioluminescence in the TetR biosensor. This observation is consistent with the previously characterized ability of tetracycline to relieve repression by TetR [54,55,56], and this supports that our biosensor system can be employed to discover activating ligands for TFRs. Importantly, tetracycline did

not relieve repression by ActR or SCO7222 (Figure 3.3A), suggesting that the induction of the TetR biosensor by tetracycline is specific.

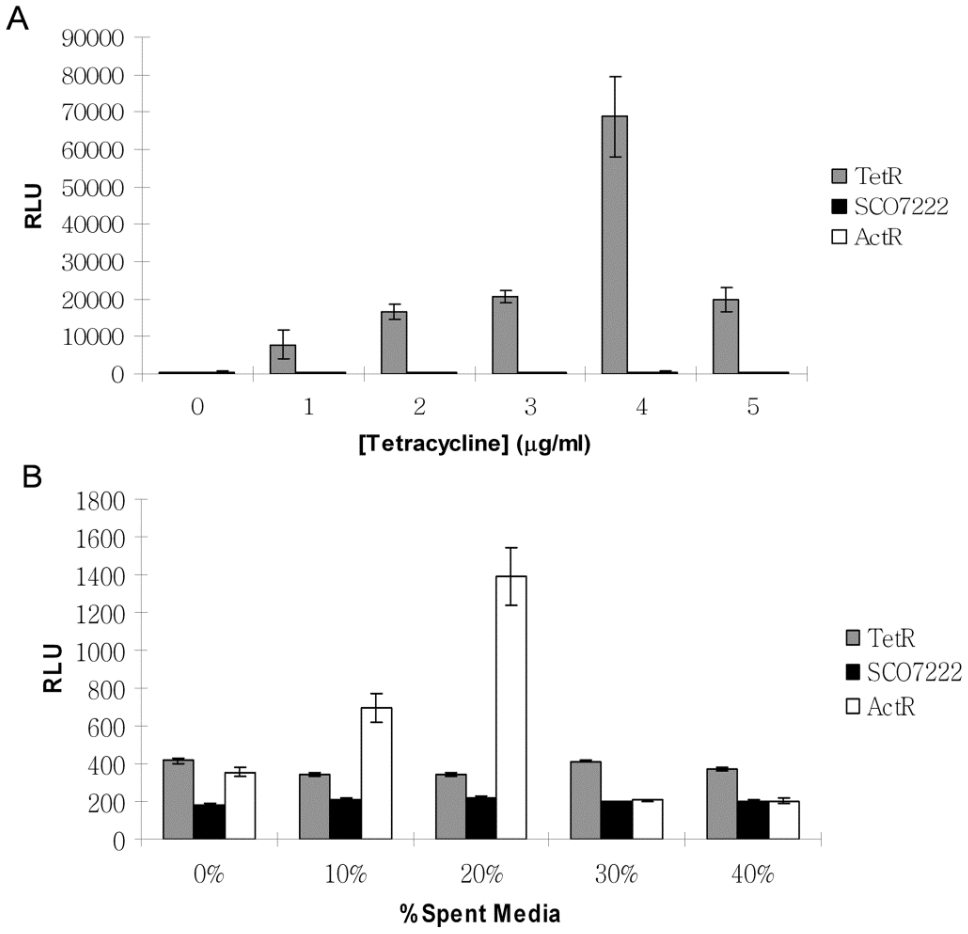

Figure 3.3. Effect of purified tetracycline or *S. coelicolor* supernatant on the TetR-, ActR-, and SCO7222-mediated repression. (A) 1-5 µg of pure tetracycline was added to 1 ml of the *E. coli* biosensors harboring pUOA1 and either pO^{tet}lux, pO^{act}lux or pO^{7223}lux along with the vector expressing the cognate repressors, respectively. **(B)** 10-40% v/v of *S. coelicolor* supernatant was added to the same biosensor strains described above. All values were measured in RLUs, and error bars indicate +/- 1 standard deviation of values obtained from three independent readings.

Similar analysis was conducted using varying concentrations of chlortetracycline, demethylchlortetracycline, doxycycline, methacycline, and oxytetracycine – all antibiotics related in structure and mechanism to tetracycline (Figure 1.4). While the concentration required for half-maximal induction varied (Table 3.2), all the tetracycline derivatives analyzed were good inducers of the TetR biosensor. Again, none of these molecules had any effect on the ActR or the SCO7222-based biosensor strains (data not shown).

Table 3.2. Induction of TetR-controlled gene expression by various tetracyclines

Antibiotic	Concentration for half maximum induction[a]	R^2
Chlortetracycline	1.0	0.98
Demethylchlortetracycline	0.18	0.63
Doxycycline	0.18	0.84
Methacycline	0.42	0.97
Oxytetracycline	0.47	0.86

[a] All concentration values are in µg/ml. Arbitrary light units were calculated from at least three independent readings.

We then applied the culture supernatant from *S. coelicolor* to all three biosensors to determine whether this streptomycete produced any metabolite that could interact with TetR, ActR, or SCO7222. As demonstrated in Figure 3.3B, the *S. coelicolor* supernatant was able to activate bioluminescence in the ActR-based biosensor, while it had no effect on the TetR- and SCO7222-based biosensors (Of note, *S. coelicolor* is not a producer of tetracycline or its derivatives. On the other hand, the TetR biosensor responded to the culture supernatant from tetracycline-producing *Streptomyces aureofaciens* [128]). Under the growth condition we have employed in this work therefore, *S.*

coelicolor does not produce a SCO7222 ligand although it is possible that there was some inducing molecule in the supernatant that could not cross the *E. coli* envelope.

3.3.4 Mutagenesis of operator sequences

Unlike pCS26-*Pac*, which is a low copy number plasmid (4 to 5 copies per cell) [237], pACYC184 is propagated at a moderately high copy number in *E. coli* (~20 copies per cell) [238]. As a result, the TFRs in our biosensors are probably in excess of their target operators, potentially impairing small-molecule induction and explaining the observed ActR/SCO7222 cross-talk. To provide a basis for comparing the interactions between the TFRs and operators in our reporter system with their biochemical affinities, we constructed mutants of O^{act} and O^{7223} using two strategies. In the first, we subjected each operator sequence to randomized mutation and screened for non-functional operators that exhibited bioluminescence in the presence of their cognate repressors (see Materials and Methods). We screened ~1600 colonies (~800 for each repressor) and isolated 16 O^{act} sequences impaired in their interaction with ActR and 18 O^{7223} sequences impaired in their interaction with SCO7222 (Table 3.3).

Table 3.3. Sequences of O^{7223} or O^{act} mutants not recognized by SCO7222 or ActR[a]

O^{7223}	C	T	G	G	A	A	C	G	C	C	G	T	T	C	C	A	G
pMTlux1	-	-	-	-	-	-	-	-	Δ[b]	N[c]	-	-	-	-	-	-	-
2	-	-	-	-	-	T	-	-	-	-	-	-	-	G	-	-	-
3	-	-	-	-	-	T	-	-	-	-	-	-	-	-	A	-	-
4	-	-	-	-	C	G	-	-	-	G	-	-	-	-	-	-	-
5	-	-	-	-	T	-	-	-	-	-	-	-	-	T	-	-	-
6	-	-	-	-	-	-	-	-	Δ[b]	N[c]	-	-	-	-	-	G	-
7	-	-	-	-	Δ[b]	N[c]	-	-	-	-	-	-	-	-	-	-	-
8	-	-	C	-	-	-	-	-	-	-	-	-	-	G	-	-	-
8	-	-	Δ[b]	N[c]	-	-	-	-	G	-	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	T	-
11	-	-	-	-	G	-	-	-	A	-	-	G	-	-	-	-	-
12	-	C	-	-	-	-	-	-	-	-	C	A	-	-	A	-	-
13	-	-	-	-	G	G	-	-	A	-	-	-	-	-	-	-	-
14	-	-	-	-	C	-	-	-	-	-	-	C	-	-	-	-	-
15	-	-	-	-	-	-	-	T	-	-	-	-	-	G	-	-	-
16	-	-	-	-	G	-	-	-	-	-	-	-	-	-	A	-	-
17	-	-	C	-	-	T	-	T	-	-	C	C	A	-	-	-	-
18	-	-	-	-	G	-	-	-	-	-	C	-	-	-	-	-	-
O^{act}	**C**	**G**	**C**	**G**	**A**	**C**	**C**	**A**	**C**	**C**	**G**	**T**	**T**	**C**	**C**	**A**	**C**
pMAlux 1	A	-	-	-	-	-	-	-	-	-	-	C	-	-	A	-	-
2	-	A	-	-	-	-	-	T	-	-	-	-	-	-	-	-	T
3	-	-	-	-	-	T	-	-	-	-	-	-	-	T	-	-	-
4	-	-	-	-	-	G	-	G	G	-	-	-	-	-	-	-	G
5	-	-	-	-	-	G	-	-	-	-	-	A	-	-	-	-	-
6	A	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-
7	-	-	A	-	-	-	-	-	-	-	C	-	-	-	-	-	-
8	-	-	-	-	G	-	-	-	G	-	C	C	-	-	-	-	-
9	-	-	-	-	-	-	-	G	-	A	-	-	-	-	-	-	-
10	-	-	-	-	-	-	-	-	-	-	C	C	-	-	-	-	-
11	-	A	A	-	-	-	-	-	-	-	C	C	-	-	-	-	-
12	-	A	-	-	-	-	-	-	-	-	-	-	G	T	-	-	-
13	-	-	-	-	-	-	-	-	-	-	-	C	A	-	-	-	-
14	-	-	-	-	-	-	-	G	-	-	C	-	-	-	-	-	-
15	-	-	G	-	-	-	-	-	-	-	-	C	-	G	-	-	-
16	-	-	A	T	-	-	-	-	-	-	-	C	-	-	-	-	-

[a] Only nucleotides found to be different from those present at the same position in the wild type operator sequences are shown.
[b] nucleotide deleted.
[c] Not applicable as either this nucleotide or previous one has been deleted in the same operator sequence analyzed.

Interestingly, all but one (pMTlux10) of the repressor-resistant alleles exhibited multiple mutations, suggesting either that the high concentration of the repressor in our biosensors resulted in DNA binding that was relatively resistant to mutation, or that the affinity of each TFR for its cognate operator was simply very high due to a large number of specific protein/operator interactions. In O^{act}, many of the alleles were altered at positions 8 or 11 while in O^{7223} many alleles had mutations at positions 5, 14 or 15 (positions 14 and 15 were two nucleotides more frequently involved in O^{7223} double mutants, Table 3.3).

In our second strategy therefore, we subjected these base pairs to site-directed mutagenesis. We generated the alleles O^{act}A8G, O^{act}G11C, O^{7223}A5G, O^{7223}C14G, and O^{7223}C15A and assessed each for repression by both their cognate and heterologous repressors. Position 5 of O^{act} (O^{act}A5G) was also tested since it is one of the palindromic nucleotides in the operator. As shown in Table 3.4, while the point mutation at position 8 of O^{act} had no effect, those at 5 and 11 eliminated repression by ActR in *E. coli*. None of the point mutations in O^{7223} altered repression of bioluminescence by SCO7222.

Table 3.4. Summary of bioluminescence and *in vitro* assays with SCO7222 and ActR

Operator	Repression[b]		K_d
O^{act}-related operators	SCO7222	ActR	SCO7222
O^{act}	N	Y	30 nM
O^{act}A8G	N	Y	23 nM
O^{act}G11C	N	N	N/A[c]
O^{act}A5G	N	N	15 nM
O^{7223}-related operators	SCO7222	ActR	SCO7222
O^{7223}	Y	Y	2 nM
O^{7223}A5G	Y	N	2 nM
O^{7223}C14G	Y	Y	2 nM
O^{7223}C15A	Y	N	2 nM

[a] For point mutants, only the specific nucleotide changes made are shown.
[b] Y: there was strong repression; and N: either there was no or weak repression (less than 5-fold reduction).
[c] Not applicable as no shift was observed even at the highest concentration of protein tested.

We also assessed the effects of each point mutation on cross-repression by ActR and SCO7222 and found that the results were quite different. For example, O^{7223}A5G and O^{7223}C15A, which had no effect on repression by SCO7222, eliminated cross-repression by ActR (Table 3.4). Clearly therefore, while the operator sequences are similar, the manner in which they are recognized by the two proteins in the context of our biosensors is subtly different.

3.3.5 Correlation of *in vitro* binding strengths with repression in biosensors

To assess the significance of these results, we measured the affinities of SCO7222 and ActR for O^{7223} and O^{act} and for the relevant point mutants described above. Using purified His_6-SCO7222 and His_6-ActR we carried out gel mobility shift experiments with radioactively labeled probes corresponding to each operator sequence. As shown in Figure 3.4, strong interactions of both proteins with their cognate operators were observed. Consistent with the fact that SCO7222 caused complete inhibition of transcription on O^{7223} but not on O^{act} in *E. coli*-based assays (Table 3.1), dissociation constants (K_ds) for these interactions were ~2 nM for O^{7223} and ~30 nM for O^{act} (Figure 3.4A). Even though the affinity of SCO7222 for O^{act} was much lower compared to O^{7223}, binding at higher concentrations possibly explains the consistent weak repression this protein brought about on the O^{act}-regulated promoter in bioluminescence assays (Table 3.1).

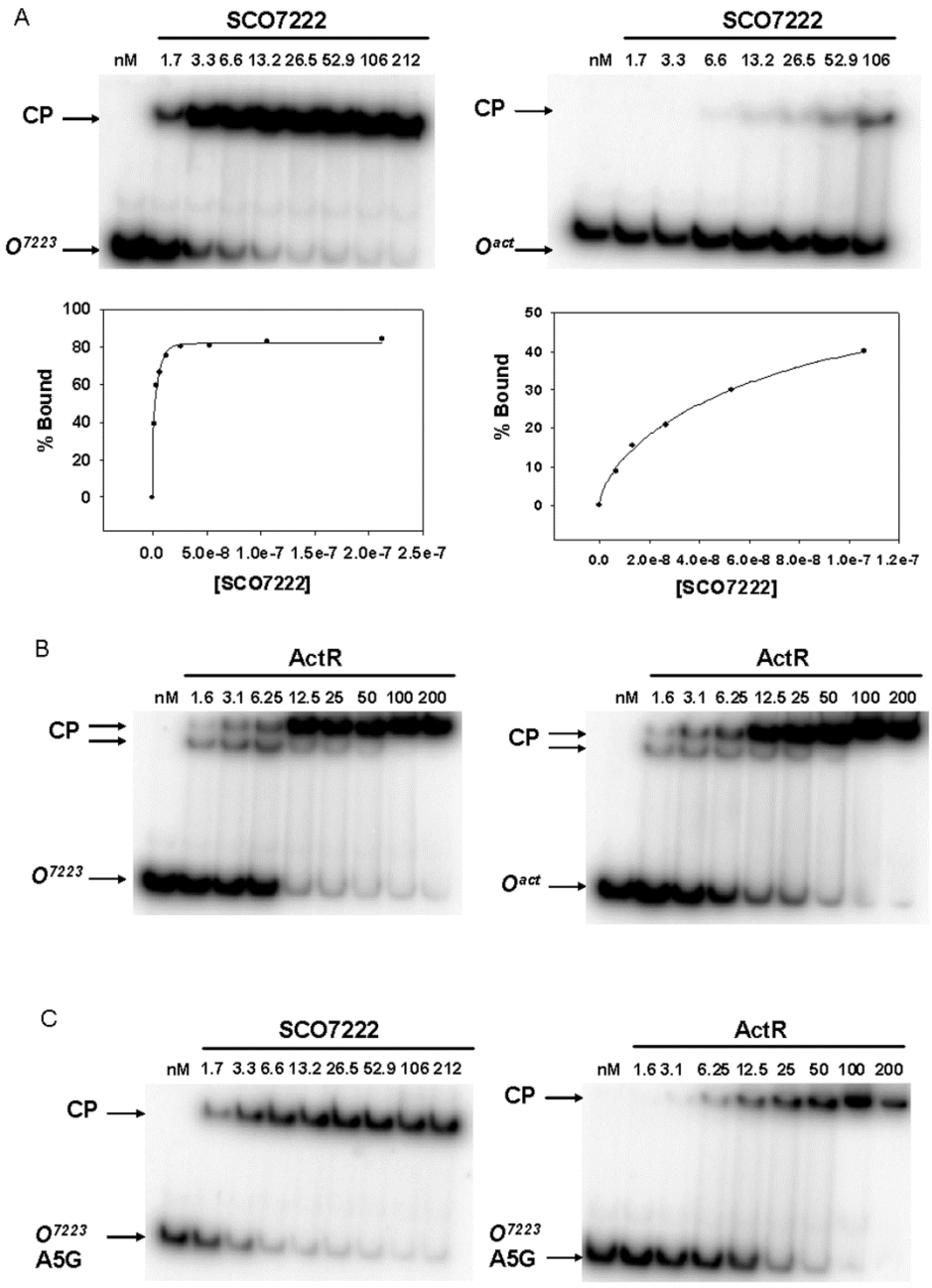

Figure 3.4. Characterization of the interactions between ActR or SCO7222 and *O*^{act} or *O*^{7223}. The ^{32}P-labelled *O*^{7223} and *O*^{act} probes were incubated with 1.5-210 nM of proteins at 30 °C for 10 minutes and then separated in 12% non-denaturing polyacrylamide gels. The interactions shown on the gels are between SCO7222 and *O*^{7223}/*O*^{act} **(A);** ActR and *O*^{7223}/*O*^{act} **(B)**; and SCO7222/ActR and *O*^{7223}A5G **(C)**. "CP" indicates the protein-DNA complex.

As shown in Figure 3.4B, ActR bound to O^{7223} as efficiently as O^{act}, in agreement with the capacity of this protein to repress both O^{act}- and O^{7223}-regulated transcription in *E. coli*-based biosensors (Table 3.1). Unlike SCO7222, ActR produced two shifted bands with both O^{act} and O^{7223}, raising the possibility that two ActR dimers interact with each site.

We went on to assess the effect of each point mutation in O^{act} and O^{7223} on the affinities of these DNAs for ActR and SCO7222 and found a nearly perfect correlation between *in vitro* binding and repression by the two TFRs in the biosensors (Figure 3.4 and Table 3.4). For example, in *E. coli*, SCO7222 repressed bioluminescence from the O^{7223}A5G operator, but ActR did not. Consistent with this, SCO7222 bound to O^{7223}A5G with a similar affinity to O^{7223} (Figure 3.4A and 3.4C) while much higher concentrations of ActR were required to form a complex with O^{7223}A5G compared to either O^{act} or O^{7223} (Figure 3.4B and 3.4C). In sum, the K_ds of SCO7222 for the three mutant operators it was able to repress in *E. coli* (O^{7223}A5G, O^{7223}C14G, and O^{7223}C15A), were all in the 2-3 nM range, whereas its K_d for those operators it could not repress in *E. coli* were at least 5-fold higher and in one case, O^{act}G11C, we observed no interaction at all (Table 3.4).

3.4 Discussion

In this study, we have developed an *E. coli*-based biosensor system that reports the interactions between TFRs of interest and their target DNAs in a whole cell-based manner. We have used this system, in combination with gel mobility shift assays, to investigate the DNA-binding properties of ActR and SCO7222 of *S. coelicolor*. We have shown that both TFRs bind to the operator sequence(s) previously predicted in Chapter 2. Importantly, our analysis of the SCO7222 operators (in addition to the known TetR operators) underscores the fact that repeated palindromic sites in the intergenic sequences can be assumed to be candidate binding sites for TFRs when a TFR-encoding gene is paired divergently with a putative target gene. Operator identification is mandatory for designing a synthetic promoter fragment in our biosensor system; therefore, it is desirable to investigate whether this finding is applicable to the majority of TFRs with a divergent neighbor and their operator sequences can be predicted through simple sequence analysis of individual intergenic regions (tested in Chapter 4).

Interestingly, our analysis has suggested that the operators of ActR and SCO7222 are related such that ActR has a similar affinity for both sequences, both in *E. coli*-based assays and *in vitro*. Similarly, SCO7222 is able to interact with the ActR operator although it binds more tightly to its own operator. These observations raise questions about possible cross-talk between ActR and SCO7222 in *S. coelicolor*. In our view, it is unlikely that the interactions of these proteins with heterologous operators are biologically meaningful as the

SCO7222/*7223* gene pair is located far outside the *act* gene cluster which contains most, if not all of the genes necessary for actinorhodin biosynthesis and self-resistance [239,240].

How might cross-talk be avoided between ActR and SCO7222 (and perhaps other members of this very large gene family) *in vivo*? One possibility is that coupled transcription and translation, which is a characteristic of all bacteria, results in the direct delivery of a TetR protein from the ribosome on which it is synthesized to its cognate operator. This scenario would be consistent with the fact that the repressor-encoding genes tend to be closely linked to their target genes. If, as is the case for TetR, these two TFRs regulate their own production in addition to expression of their target genes, they would limit their own intracellular accumulation and hence minimize or eliminate cross-talk.

In addition to the interactions of the TFRs with their target DNAs, our biosensors have also successfully demonstrated small-molecule inductions of TetR and ActR. In addition to confirming that tetracycline is the activating ligand for TetR, we have shown that repression by ActR is relieved by the culture supernatant from *S. coelicolor*. Significantly, this observation later played an integral role in discovering that actinorhodin and related molecules, including the biosynthetic pathway intermediate (*S*)-DNPA, are inducing ligands for ActR [128]. Interestingly, (*S*)-DNPA is a better ligand for ActR than actinorhodin *in vitro* and this observation led us to propose a “feed-forward” mechanism [128]. This model suggests that repression by ActR is relieved by the intermediates of actinorhodin

such that the *actAB* genes (encoding putative efflux pumps for actinorhodin) can be expressed to protect the host cell prior to complete biosynthesis and accumulation of the final antibiotic product.

More importantly for our immediate purposes, analysis of the ActR ligands further supports our current paradigm which predicts that many TFRs interact with small molecules that are related or identical to the substrates of their regulatory target gene products. The fact that our reporter strains were able to specifically detect pure molecules (i.e. tetracyclines for TetR) and molecules from a natural source (i.e. *S. coelicolor* culture supernatant for ActR) is highly promising. These observations suggest that this biosensor mechanism provides a valuable tool to investigate biological roles of the members in this large family of transcription factors.

3.5 Conclusion

We have developed a bioluminescence-based *E. coli* biosensor system that can be used to identify target operator sequences and small-molecule ligands of TFRs. Through our work with ActR and SCO7222, we have successfully revealed the high degree of correlation between *in vitro* binding and repression observed in the biosensors. Moreover, high specificity and sensitivity of the TetR- and ActR-based biosensors for their cognate ligands have suggested that our biosensor system has potential for broad application to investigate a large number of TFRs.

3.6 Materials and Methods

3.6.1 Bacterial strains and culture conditions

Bacterial strains used in this study are described in Table 3.5. *E. coli* XL1-blue was used during general cloning process, while *E. coli* DH5α was used to propagate plasmids for biosensor strain construction. *E. coli* BL21(DE3) was used for protein overexpression. All *E. coli* strains were grown using Luria Broth (LB) or LB agar medium, containing the appropriate antibiotics when required (100 μg/ml ampicillin, 50 μg/ml kanamycin, and 37 μg/ml chloramphenicol). *Streptomyces* cultures were grown at 30˚C in yeast-extract malt-extract (YEME) broth or maintained on solid mannitol soy (MS) agar medium [234].

Table 3.5. Bacterial strains used in this work

Strain	Description	Source
E. coli		
BL21(DE3)	F$^-$ *dcm ompT hsdS*(r_B^- m_B^-) *gal met* λ(DE3)	Novagen
DH5α	F'/*endA1 hsdR17* (r_K^- m_K^+) *glnV44 thi-1 recA1 gyrA*(Nal^r)*relA1 del*(*lacIZYA- argF*)*U169 deoR (80dlac (lacZ) M15)*	Stratagene
XL1-blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac*[F′ *proAB lacI*^{q} ZΔM15 Tn*10*(Tet^r)]	Stratagene
Streptomyces		
M145	*S. coelicolor* prototroph, SCP1- SCP2-	John Innes
84/25	*S. aureofaciens* tetracycline-producing strain	J. Kormanec

3.6.2 Plasmids, primers, and sequencing

The plasmids and primers used in this study are listed in Table 3.6 and Table 3.7, respectively. Oligonucleotide primers used in this study were obtained from the Institute for Molecular Biology and Biotechnology (MOBIX) facility at McMaster University or from Sigma Aldrich. Polymerase chain reactions (PCR) were carried out using Vent DNA polymerase (New England Biolabs). DNA sequencing was carried out by the MOBIX facility.

Table 3.6. Plasmids used in this study

Plasmid	Description (selection marker)[a]	Background	Reference
pCS26-*Pac*	Promoterless *luxCDABE* reporter (Kan^r)	pSC101	[241]
$pO^{tet}lux$	Reporter based on the TetRoperator (Kan^r)	pCS26-*Pac*	This work
$pO^{act}lux$	Reporter based on the ActR operator (Kan^r)	pCS26-*Pac*	This work
$pO^{7223}lux$	Reporter based on the SCO7222 operator (Kan^r)	pCS26-*Pac*	This work
pMAlux	Reporter based on the ActR operator with random mutations (Kan^r)	pCS26-*Pac*	This work
pMTlux	Reporter based on the SCO7222 operator with random mutations (Kan^r)	pCS26-*Pac*	This work
pACYC184	Plasmid required for repressor expression vector constructions. Cm^r gene from Tn*9* and Tc^r gene from pSC101 (Cm^rTc^r)	p15A	[238]
pTetR	*tetR* gene in pACYC184; TetR-expressing vector (Cm^r)	pACYC184	This work
pActR	*act*II-ORF1 (*actR*) gene in pACYC184; ActR-expressing vector (Cm^r)	pACYC184	This work
pSCO7222	*SCO7222* gene in pACYC184; SCO7222-expressing vector (Cm^r)	pACYC184	This work
pUOA1	Plasmid required to provide tetracycline resistance for DH5α-based strains. Tc^r gene from pUA466 (Tc^r Ap^r)	pUC8	[242]
pET28a-ActR	ActR-overexpressing vector for protein purification	pET28a	This work
pET28a-SCO7222	SCO7222-overexpressing vector for protein purification (Kan^r)	pET28a	This work

[a] Kan^r, kanamycin resistance; Cm^r, chloramphenicol resistance; Tc^r, tetracycline resistance and Ap^r, ampicillin resistance.

Table 3.7. Primers used in this study

Name	Sequence (5′ → 3′)[a,b]	Purpose
TT1	tcgagttgacactctatcattgatagagttattttacca	+ strand oligonucleotide for preparing the O^{tet} based promoter
TT2	gatctggtaaaataactctatcaatgatagagtgtcaac	- strand oligonucleotide for preparing the O^{tet} based promoter
TA1	tcgagttgacacgcgaccaccgttccacttatttacca	+ strand oligonucleotide for preparing the O^{act} based promoter
TA2	gatctggtaaaataagtggaacggtggtcgcgtgtcaac	- strand oligonucleotide for preparing the O^{act} based promoter
T7223-1	tcgagttgacactggaacgccgttccagttatttacca	+ strand oligonucleotide for preparing the O^{7223} based promoter
T7223-2	gatctggtaaaataactggaacggcgttccagtgtcaac	- strand oligonucleotide for preparing the O^{7223} based promoter
Tet1	taagaaggagaggaattaatgatgtc	Forward primer for amplifying *tetR* gene for pTetR construction
Tet2	aaggGGATCCtgcttttaagacccac	Reverse primer for amplifying *tetR* gene for pTetR construction
Act1	taagaaggagaggagcacatgtcgcgaagcgaggaa	Forward primer for amplifying *actR* gene for pActR construction
Act2	atgggGGATCCgaaggccgtgttcatga	Reverse primer for amplifying *actR* gene for pActR construction
7222-1	taagaaggagaggagcacatggcatccaggtcgc	Forward primer for amplifying *SCO7222* gene for pSCO7222 construction
7222-2	gggcatGGATCCagtgacccatgcaacttggg	Reverse primer for amplifying *SCO7222* gene for pSCO7222 construction
7223IR-1	cgctcatcctagaccgcttgg	Forward primer for amplifying the intergenic region between *SCO7222* and *SCO7223*
7223IR-2	ggtgcttccctcctgatggcg	Reverse primer for amplifying the intergenic region between *SCO7222* and *SCO7223*
ActIR-1	gtgctcctcatcgtatggcatgaacg	Forward primer for amplifying the intergenic region between *actR* and *actA*
ActIR-2	ggcgtcccccgggtcctc	Reverse primer for amplifying the intergenic region between *actR* and *actA*
TET-EMSA-F	cgtatcacgaggccctttcg	Forward primer for EMSA probes
TET-EMSA-R	tcatatttgccatccatttgc	Reverse primer for EMSA probes

[a] compatible cohesive ends introduced by these oligonucleotides are underlined.
[b] restriction endonuclease recognition sequences introduced by these oligonucleotides are shown in capital letters.

3.6.3 Construction of *lux*-based reporter plasmids and expression vectors for the TetR, ActR and SCO7222 biosensors

Oligonucleotides were annealed to prepare synthetic promoters consisting of the -10 and -35 regions flanking the known (in the case of TetR) or candidate (P1, P2, or P3 for ActR and N1, N2, or N3 for SCO7222) binding sites. The fragments were introduced into the *Bam*HI-*Xho*I sites of pCS26-*Pac* to give reporter plasmids pO^{tet}lux, pO^{act}lux, and pO^{7223}lux, in which the expression of the *lux* operon was under the control of TetR, ActR, and SCO7222, respectively.

The *tetR*, *actR*, and *SCO7222* genes were amplified by PCR with the introduction of Shine-Dalgarno sequences and downstream recognition sites for *Bam*HI. After digestion with *Bam*HI, the fragments were ligated to *Bam*HI-*Eco*RV-cut pACYC184 to give pTetR, pActR, and pSCO7222, respectively.

3.6.4 Bioluminescence measurements

Isolated *E. coli* DH5α colonies containing individual reporter plasmids +/- expression vector were used to inoculate one-milliliter amounts of reporter cultures, which were grown for 16-20 hours before measuring luminescence using a Lumat 9507 luminometer (Bertholt Technologies). In some cases, *E. coli* reporter cultures were supplemented with filter-sterilized spent *Streptomyces* culture supernatant or with purified tetracyclines. The half maximal concentrations of various tetracyclines required for induction were determined using a standard sigmoidal dose response regression equation: y = bottom +

(top-bottom)/(1+10$^{(Ec50-x)*hillslope}$), where bottom (y_{min}) was set to 1 as data was normalized using signal to noise ratios.

3.6.5 Operator mutagenesis

Double stranded DNA products, obtained by annealing the imperfectly matched primers TA3 and TA2, and T7223-3 and T7223-2, were inserted into the *Bam*HI-*Eco*RV sites of pCS26-*Pac* to give pMAlux (TA3-TA2) and pMTlux (T7223-3-T7223-2), respectively. TA3 and T7223-2 are identical to TA1 and T7223-1, except that the operator sequences had been doped during synthesis as follows (conducted by Sigma Aldrich). At each operator position while the concentration of the correct nucleotide was as usual during synthesis, the other three nucleotides were also present as defined by x/3n, where x = concentration of principle nucleotide and n = length of the operator sequence. The result of this is that each mutagenic oligonucleotide would be expected to possess at least one nucleotide sequence change in the operator region, embedded in an otherwise 'wild-type' synthetic promoter. In both cases we mutagenized the top strands shown in Figure 3.2B.

The plasmids pMAlux and pMTlux were introduced into *E. coli* strains containing pActR or pSCO7222, respectively, to isolate colonies that produced luminescence, indicating that the respective repressors could not bind to the mutagenized operators. Plasmid DNA was isolated from these strains and reintroduced into *E. coli* to isolate kanamycin-resistant and chloramphenicol-

sensitive colonies, which had lost pActR or pSCO7222, but still contained pMAlux or pMTlux. The DNA sequences of the operators cloned in pMAlux and pMTlux were determined for further analysis.

3.6.6 Expression and Purification of His_6-ActR and His_6-SCO7222

Previously described pET28a-ActR and pET28a-SCO7222 were used to express ActR and SCO7222, respectively. *E. coli* BL21(DE3) cultures containing individual vectors were grown at 37°C to an OD_{600} of ~0.4-0.6 and were then induced with 1 mM isopropyl β-D-thiogalactopyranoside for 20 hours at 37°C. Cell pellets were resuspended in 25 ml lysis buffer (10 mM Tris-HCl (pH7.8), 150 mM NaCl and 1x Complete Mini EDTA-free protease inhibitor (Roche)) and lysed by 3x passage through a French press cell. 15 ml of Ni-NTA agarose resin (Qiagen) was added and the slurry was incubated for 30 minutes with gentle shaking. The resin was loaded onto a column, washed with 500 ml of 10 mM Tris-HCl (pH7.8), 150 mM NaCl, and 50 mM imidazole, and then eluted with 25 ml 10 mM Tris-HCl (pH7.8), 150 mM NaCl, and 0.5 M imidazole. His_6-ActR or His_6-SCO7222 was transferred to 10 mM Tris-HCl (pH8), 500 mM NaCl and 20% glycerol by buffer exchange using a 10 kDa cut-off Amicon Ultra Centrifugal Filter Device (Millipore) and was stored at -80°C. Protein concentrations were determined by Bradford assay (Bio-Rad).

3.6.7 EMSAs

The primers TET-EMSA-F and TET-EMSA-R along with $pO^{act}lux$, $pO^{7223}lux$, or the reporter plasmids with point mutagenized operators (as templates) were used in PCR reactions to isolate the double stranded DNA fragments containing the operator regions, which served as probes for electrophoretic mobility shift assays. The DNA fragments were end-labeled using [γ-^{32}P]ATP (Amersham) and T4 polynucleotide kinase (New England Biolabs).

Labeled probe (1 ng), varying amounts of purified protein and 90 ng of salmon sperm DNA were used in 15 µl reactions containing 1x EMSA reaction buffer (10 mM Tris-Cl (pH 7.8), 150 mM NaCl, 2 mM DTT and 10% glycerol). Reactions were incubated at 30°C for 10 minutes and were fractionated on 12% non-denaturing polyacrylamide gels containing 1.5% glycerol. The gels were exposed using a phosphor screen (Amersham) and bands were detected using a PhosphorImager (Molecular Dynamics).

3.6.8 Determination of dissociation constants (K_d)

ImageQuant software (Molecular Dynamics) was used to analyze EMSA results to determine the percentages of shifted and un-shifted probes, which represent bound and unbound substrate, respectively. Saturation curves (% probe bound against the protein concentration) were drawn with SigmaPlot 2000 to determine K_d.

CHAPTER 4: STREAMLINING TETR FAMILY TRANSCRIPTIONAL REGULATOR REPORTER SYSTEM

4.1 Abstract

The two-plasmid biosensor mechanism described in Chapter 3 is a useful tool for investigating TetR family transcriptional regulators (TFRs). In this study, we aim to further advance this mechanism by improving its sensitivity and applicability. First, we investigate the use of a one-plasmid biosensor system. In this system, a TFR-encoding gene and its target operator sequence are introduced into a single vector, and expression of a TFR on this plasmid is under the control of an inducible promoter. We show that, using TetR, this new reporter system is at least 1000x more sensitive to tetracycline than the previous two-plasmid system. In addition, we create a reporter for SCO0310, a previously uncharacterized TFR encoded in *Streptomyces coelicolor*. Unlike previously tested ActR and SCO7222, no clear operator candidate can be identified in the *SCO0310*/*SCO0311* intergenic sequence although it is very likely that SCO0310 regulates *SCO0311*. Through *in vitro* analysis, we find that SCO0310 binds a 40 bp region within the intergenic region, and we test the use of a "road-block" strategy in which the entire sequence of this binding region is placed downstream of the -10 and -35 promoter elements – not between them (as in the previous reporters) – to prepare a synthetic promoter. We observe that SCO0310 represses the activity of this promoter in the biosensor. This strategy should be generally applicable to other TFRs whose operators cannot be easily predicted by simple sequence analysis.

4.2 Introduction

In the previous chapter, we have described a two-plasmid *Escherichia coli*-based reporter mechanism that can be utilized for testing putative binding sites and discovering small-molecule ligands for TetR family transcriptional regulators (TFRs) [128,235]. This system has allowed us to develop the TetR-based biosensor that specifically detects tetracycline and its derivatives. Furthermore, the ActR-based biosensor we developed has specifically responded to the culture supernatant from *Streptomyces coelicolor*, leading to discovery of actinorhodin and its related molecules as the inducing ligands.

Despite the previous success, there are two significant areas that we believe should be addressed in order to maximize the use of this biosensor mechanism and make it more generally applicable to other TFRs. First of all, improving sensitivity of the biosensors for their cognate ligands is highly desirable. The two-plasmid system has been designed such that a TFR-expressing vector (pACYC184 backbone) is propagated at a higher copy number in *E. coli* than a reporter vector (pCS26-*Pac* backbone) containing the synthetic promoter-fused *lux* operon. As a result, the TFR in the biosensor is likely in excess of its target operator, potentially increasing the ligand concentration required to observe induction.

In this study, we have explored a newly developed one-plasmid biosensor system (Figure 4.1). Importantly, expression of a TFR in this system is driven by the *araBAD* promoter (P*araBAD*) that is under the control of the transcription factor AraC. AraC regulates expression of the *araBAD* operon that is involved in the conversion of L-arabinose to D-xylose-5-phosphate [243], and it possesses a dual repressor/activator activity depending on the intracellular concentration of L-arabinose [244]. In the absence of arabinose, AraC has a negative effect on P*araBAD* while it enhances the promoter activity when it binds the sugar. Here, we have created a TetR-based biosensor using the one-plasmid approach and find that this reporter system is at least 1000x more sensitive to tetracycline than the previous two-plasmid system.

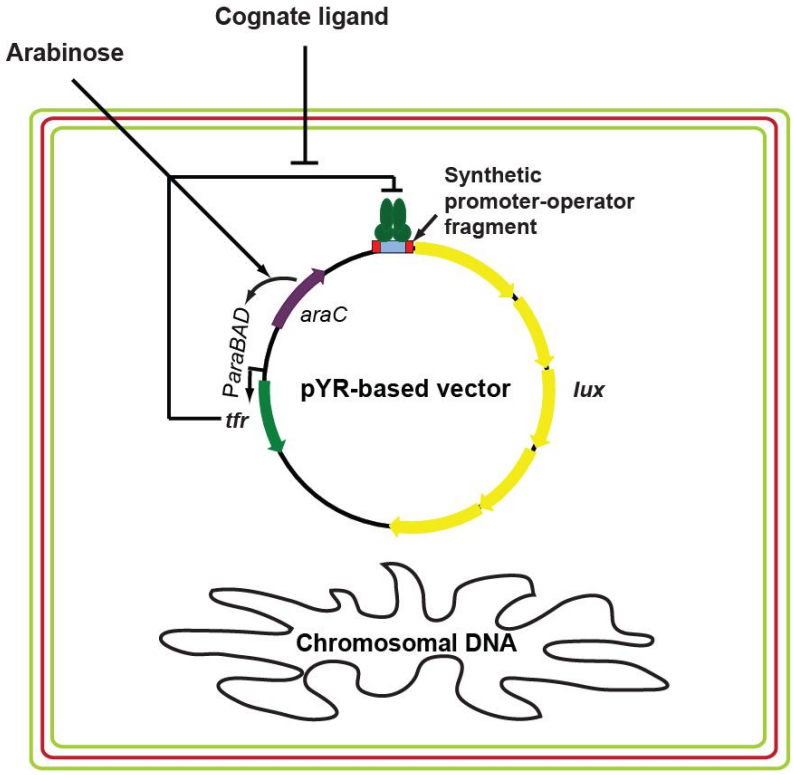

Figure 4.1. Schematic diagram of the one-plasmid TFR biosensors. A reporter vector – using the plasmid pYR as a backbone – is constructed by inserting a synthetic promoter fragment, which contains the -10 and -35 promoter elements (red boxes) surrounding a putative TFR binding site (light blue box), upstream of the *lux* operon (yellow arrows). In addition, a gene encoding the TFR (green) is cloned downstream of the *araBAD* promoter (P*araBAD*), divergently to the *araC* gene (purple). The TFR expression will be induced by the addition of arabinose, causing repression of the *lux* genes. The repression will be relieved only in the presence of a cognate ligand of the TFR.

In addition to improving sensitivity, identification of a TFR operator, required for designing a synthetic promoter, provides another vital challenge in reporter construction. Our previous analysis has suggested that TFRs - when divergently transcribed from their putative target genes – likely bind to repeated palindromic sequence elements in the intergenic sequences. This is consistent with our prediction that many of these TFRs interact with multiple operators, as

homodimers, to regulate expressions of not only their divergent target genes but also their own genes.

In this study, we have analyzed 94 TFRs of *S. coelicolor* that are separated by less than 200 bp from their divergently transcribed putative targets (i.e. ones that follow the "200 bp" rule described in Chapter 2). We find that only 33 of them (35%) have repeated palindromic operator candidates that are easily recognizable through simple sequence analysis of the individual intergenic regions, making it a greater challenge to create reporters for the remaining 61 TFRs.

To develop a generally applicable method to construct biosensors with the TFRs lacking predictable operator sequences, we have focused on SCO0310, a previously uncharacterized TFR from *S. coelicolor*, as the model protein. By conducting *in vitro* assays we find that this TFR binds a 40 bp region within the intergenic region between *SCO0310* and *SCO0311*. Unlike the previous cases where we have placed the TFR operators between the -10 and -35 promoter elements, we successfully create an SCO0310-based reporter by placing the entire 40 bp binding region downstream of the promoter elements. Discovery of this "road-block" strategy is significant as it should be applicable to TFRs even when their operators are hard to identify by simple sequence analysis or when they are too long to be inserted between the promoter elements, as we have done previously.

4.3 Results

4.3.1 Construction of the one-plasmid TetR reporter system

In collaboration with the Davidson laboratory at University of Toronto, we have generated one-plasmid reporter system for TFRs [245]. A synthetic promoter fragment – containing a known TetR operator placed between the -10 and -35 promoter elements – was prepared and cloned upstream of the *lux* operon to create pYRtetO. The *tetR* open reading frame was then amplified and placed under the control of the promoter P*araBAD* (and divergently oriented to the *araC* gene) on pYRtetO, resulting in the pYRtetOR plasmid. We introduced pYR, pYRtetO, and pYRtetOR separately into *E. coli*, and we compared the bioluminescence production from each of the resulting strains in the presence of varying concentrations of L-arabinose.

As shown in Figure 4.2, the pYRtetO and pYRtetOR strains produced 211-fold and 19-fold higher bioluminescence respectively than the empty vector strain in the absence of L-arabinose, confirming that the synthetic promoter for the *lux* operon on both plasmids is active. The reduced amount of light produced from pYRtetOR compared to that from pYRtetO (11-fold reduction) was consistent with previous results [245], and this suggested that leaky expression of the *araBAD* promoter in the absence of arabinose generated a sufficient amount of TetR to cause some repression. As expected, no significant change in bioluminescence was observed from the pYR and pYRtetO strains when L-arabinose was added. On the other hand, addition of 0.002% arabinose further

reduced the pYRtetOR-dependent bioluminescence by 9-fold and the light production was eliminated at arabinose concentrations of 0.02% or higher.

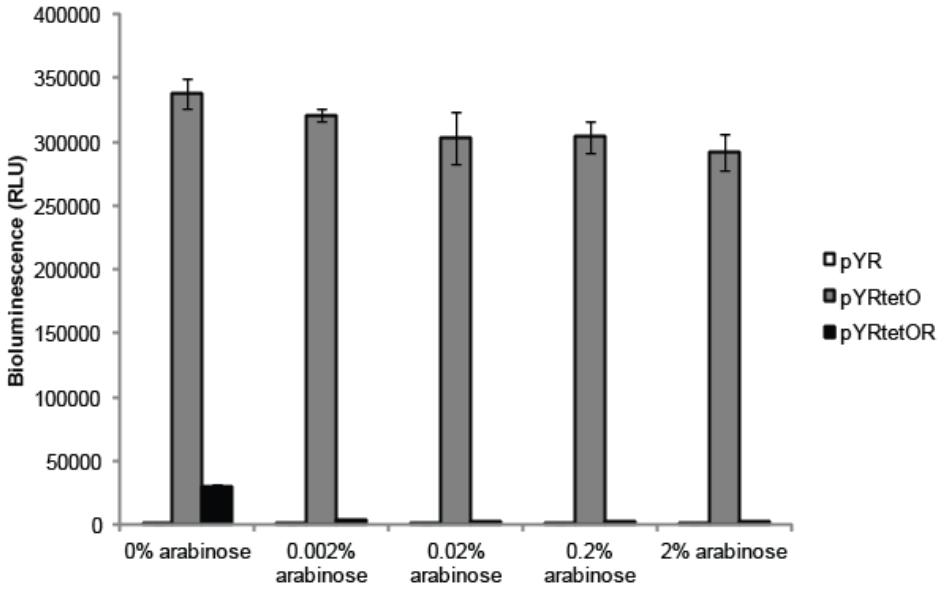

Figure 4.2. Effect of L-arabinose on the one-plasmid TetR reporter. Analysis was carried out using *E. coli* TOP10 as the host. Varying amounts of L-arabinose were added to 1 ml of the *E. coli* biosensors harboring pYR, pYRtetO, or pYRtetOR. All values were measured in relative light units (RLUs), and error bars indicate +/- 1 standard deviation of values obtained from three independent readings.

To measure sensitivity of this reporter, we cultured the pYRtetOR strain in the presence of various concentrations of tetracycline. Of note, no L-arabinose was added for more accurate evaluation of the reporter. Similar to the data obtained with the two-plasmid TetR reporter (Figure 3.3A), we observed a dose-responsive induction of bioluminescence (Figure 4.3); however, the induction peaked at the drug concentration of 10 ng/ml, which is 400-fold lower than 4

µg/ml tetracycline required previously. More importantly, the new TetR biosensor could detect the drug at a concentration as low as 1 ng/ml (6-fold induction), consistent with the observation made by our colleagues [245], suggesting that it is at least 1000x more sensitive to tetracycline than the previous biosensor (1 µg/ml required for 3-fold induction).

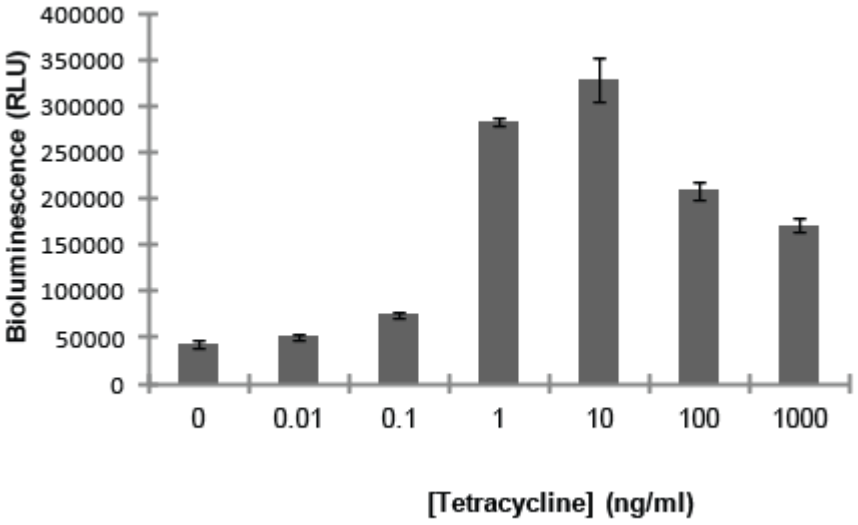

Figure 4.3. Effect of purified tetracycline on the TetR biosensor pYRtetOR. 0.01 to 1000 ng of pure tetracycline was added to 1 ml of the *E. coli* biosensors harboring pYRtetOR. All values were measured in RLUs, and error bars indicate +/- 1 standard deviation of values obtained from three independent readings.

4.3.2 Predicting operators by repeated palindrome search is impossible for most TFRs

The approach we have used in reporter construction for TetR, ActR, and SCO7222 depends on the identification of a cognate operator in order to design a synthetic promoter fragment. As is the case for these three TFRs, it is relatively straight-forward to construct reporter plasmids when operators of TFRs can be easily predicted by searching for palindromic sequence elements that are repeated multiple times in the regions between their genes and predicted target genes.

To determine if all TFRs possess similarly predictable operators, we analyzed 94 TFR-encoding genes that are transcribed divergently to their known or putative targets and obey the “200 bp” rule. We conducted sequence analysis of the individual intergenic regions between the TFR genes and their putative target genes. It was evident that the majority of these TFRs (61 TFRs or 65%) do not have recognizable repeated palindromes in their intergenic regions while reasonable candidates are identified for 33 TFRs (35%, Table 4.1). Of note, the *S. coelicolor* genome encodes 11 TFRs that are >200 bp away from their divergent neighboring genes. Consistent with our “200 bp” rule, which suggests local regulatory relationships are less likely for these TFRs, it is even more challenging to predict operators for these proteins (only 2 of them, or 18%, have recognizable repeated palindromes in the intergenic regions, Table 4.1).

Table 4.1. Putative operator sequences of 35 TFRs from *S. coelicolor*

TFR	Putative operators[a]	Length of intergenic region (bp)
SCO0116	TTAGATGTCGACTGGCATCTAT	116
	TTAGATTGGGGTCATAATCTAA	
SCO0241	CAGACCGGCGGTCTT	345
	CAGACCGCCGGTCTG	
SCO0250	GTGAATGGTGTTCAC	63
	ATGAACAGTGTTCAC	
SCO0253	ATCTATCGGTGAATGAT	70
	ATCTATCAGCGATAGGT	
SCO0520	AAGTGGAGGCGCCGCCACTT	136
	AAGCGGGGGCGCCTCCACTT	
SCO0646	AATGCAAGCCCACTTGCATT	111
	ATCGCAGGTCGACTTGCGTT	
SCO0772	GGAGAACGAGCGTTCTCC	96
	GGAGAACGCACGTTCTCC	
SCO0887	ATCCGGACCGGGGTCCGTTT	139
	AAACGGACCGCGGTCCAGTT	
SCO1135	ACCGGACAACTGTCCATT	188
	AACGGACAGTTGTCCGCT	
SCO1193	CTCCTTAAGATAGGGAG	101
	CTCCTCAGAGTAGGGAG	
SCO1718	AGCGCACTACGTACGGT	63
	TACGTACGGTGTACGTA	
SCO2243	CAACGCTAGTACTG	90
	CAATGCTAGGATCG	
SCO2374	GACAGCCGCTGAC	115
	GTCAGTAACTGAC	
SCO2775	GTTAATGAGCGTTAAC	97
	GTTAACGACCGCTAAC	
SCO3167	AACGAAACCGTTTCGTT	148
	TACGAAACGGTGTCGTT	
SCO3315	AGGCGAAGCCCGCATCCGCT	128
	AGTGGACGATGCCCTCCACT	
SCO3367	ACTTGACGCCCGGCTAGT	158
	ACTTGCCGGGCGGCAAGT	
SCO3979	TAACGGATAGCTTTT	80
	TAACGGTTACTATTT	
SCO4008	AACTAACTGGTTGGTT	136
	AACCAAATAGTTGGTT	
SCO4303	TCGGTACCGATCGGTACTGA	94
	TCCGTACCGATCGATACGGA	
SCO4358	TACTGAGTACACCGTA	96
	TACGCCGTACTCAGTT	
SCO4454	AGAACGTTCGGTCT	181
	AGACCGAACGTTCT	
SCO4871	CTTAACGTCGTTAAG	72
	CTTATCGCCGTTAAG	
SCO4898	ACTTGACACCGACCAGA	109
	TCTTGTCAATGACTAGA	
SCO4940	TAGCGTACGGGCGTTCGCTA	101
	TAGCGAACGCCCGTACGCTA	

SCO5068	*AA*C*GGGGT*G*C*C*C*T*C*C*G*C*T* *AAT**GGGGTG**T**C**A**C**C**C**G**TT*	155
ActR	See Figure 3.2A	110
SCO5811	*AGTGAACGC*TC*GCGTTCACT* *AGTGAACGC*AAC*CGTTCACT*	187
SCO6121	*ATCTAGCG*C*CGCTAGAT* GT*CTAGCG*GT*GCTAG*GA	92
SCO6323	*TCGCACGC*GTTG*GCGT*T*C*G*A* *AC**GCACGC**TT*GAC*CGTGC*AT*	385
SCO6694	AT*CGATACGG*C *ACCGATACGGT*	175
SCO7222	See Figure 3.2A	146
SCO7719	*TACGTACGCGCA* *TGCGTGCACGCA*	54
SCO7794	*CT*T*AAC**TAAGTT**C**AG* *CT**GAAC*GCT*GTTAAG*	67
MmfR	*TCGGTAAGCTGACCGA* C*CGGC*TG*GCTTGCCGC*	194

[a] The palindromic nucleotides are italicized, while the repeated sequences are underlined.

To develop a method for creating reporters for TFRs with unpredictable operator sequences, we chose SCO0310 as a model TFR. SCO0310 is predicted to regulate *SCO0311* and *SCO0312* (possibly *SCO0313* and *SCO0314*, too) which may constitute an operon (Figure 4.4), and the intergenic region between *SCO0310* and *SCO0311* is 108 bp long. The *SCO0311* and *SCO0312* genes encode a putative fatty acyl-CoA synthetase and a putative acyl-CoA dehydrogenase respectively, and proteins from these families are often involved in fatty acid degradation through β-oxidation [246].

Figure 4.4. SCO0310 and putative target genes. The *SCO0310* gene is divergently oriented to a putative target cluster containing four genes. SCO0311, SCO0312, and SCO0313 are putative fatty acyl-CoA synthetase, acyl-CoA dehydrogenase, and membrane transporter respectively, while SCO0314 is a protein of unassigned function (15 bp between *SCO0311* and *SCO0312*; 61 bp between *SCO0312* and *SCO0313*; *SCO0313* and *SCO0314* are overlapped).

4.3.3 Identification of the SCO0310-binding region

To test the ability of SCO0310 to bind the region between its own gene and predicted target genes, we conducted an electrophoretic mobility shift assay (EMSA) using purified SCO0310 and a radiolabeled DNA fragment corresponding to the *SCO0310*/*SCO0311* intergenic region. As shown in Figure 4.5A, two protein-DNA complexes could be observed, suggesting that there are two binding sites for SCO0310 in this region. To further localize the SCO0310 binding sites, we divided the entire intergenic region into two half-sites and performed EMSAs using these two regions as probes. Two shifts were observed when the half-site region closer to *SCO0311* was used as a probe while no shift was visible with the other probe (Figure 4.5B). These data suggest that both binding sites for SCO0310 are located relatively near *SCO0311*.

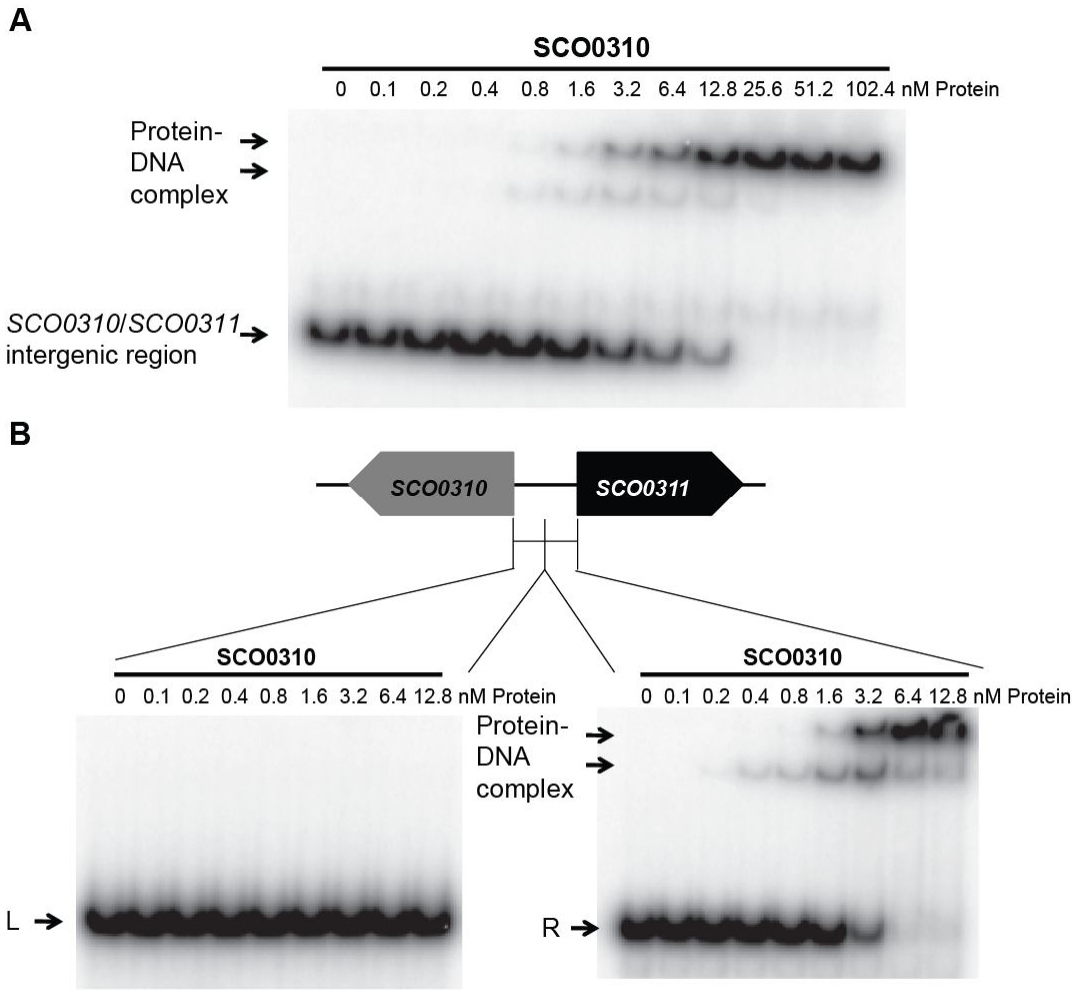

Figure 4.5. SCO0310 binds the *SCO0310*/*SCO0311* intergenic region. (A) A ^{32}P-labeled probe containing the entire *SCO0310/SCO0311* intergenic DNA sequence was incubated with the varying concentrations of SCO0310 and then separated in 12% non-denaturing polyacrylamide gels. Two distinguishable protein-DNA complexes were observed and they are indicated by arrows. **(B)** Varying concentrations of SCO0310 were incubated with two radiolabeled half-sites of the *SCO0310*/*SCO0311* intergenic region (the left half-site, L, corresponds to the region closer to *SCO0310* while the right half-site, R, is the one near *SCO0311*).

To map the locations of the SCO0310 operators more precisely, DNaseI footprinting assays were next conducted on both strands of the intergenic region. As shown in Figure 4.6A, a single protected region in the *SCO0310/SCO0311* intergenic DNA was evident in the presence of SCO0310. It is noteworthy that there was only one protected region rather than two (as suggested by the previous EMSA), possibly implying that the two binding sites of SCO0310 might be very closely located to each other. Furthermore, the length of the region protected is approximately 40 bp, which might be long enough to accept two SCO0310 dimers. Consistent with the previous EMSA results, 37 out of 40 nucleotides are located in the half-site close to *SCO0311* (Figure 4.6B); moreover, two shifts were observed when a radiolabeled probe containing this 40 bp sequence was prepared and incubated with the purified SCO0310 for a mobility shift assay (data not shown). These data suggest that SCO0310 is likely to regulate *SCO0311* (and its downstream gene(s) forming an operon) and further support our “200 bp” rule.

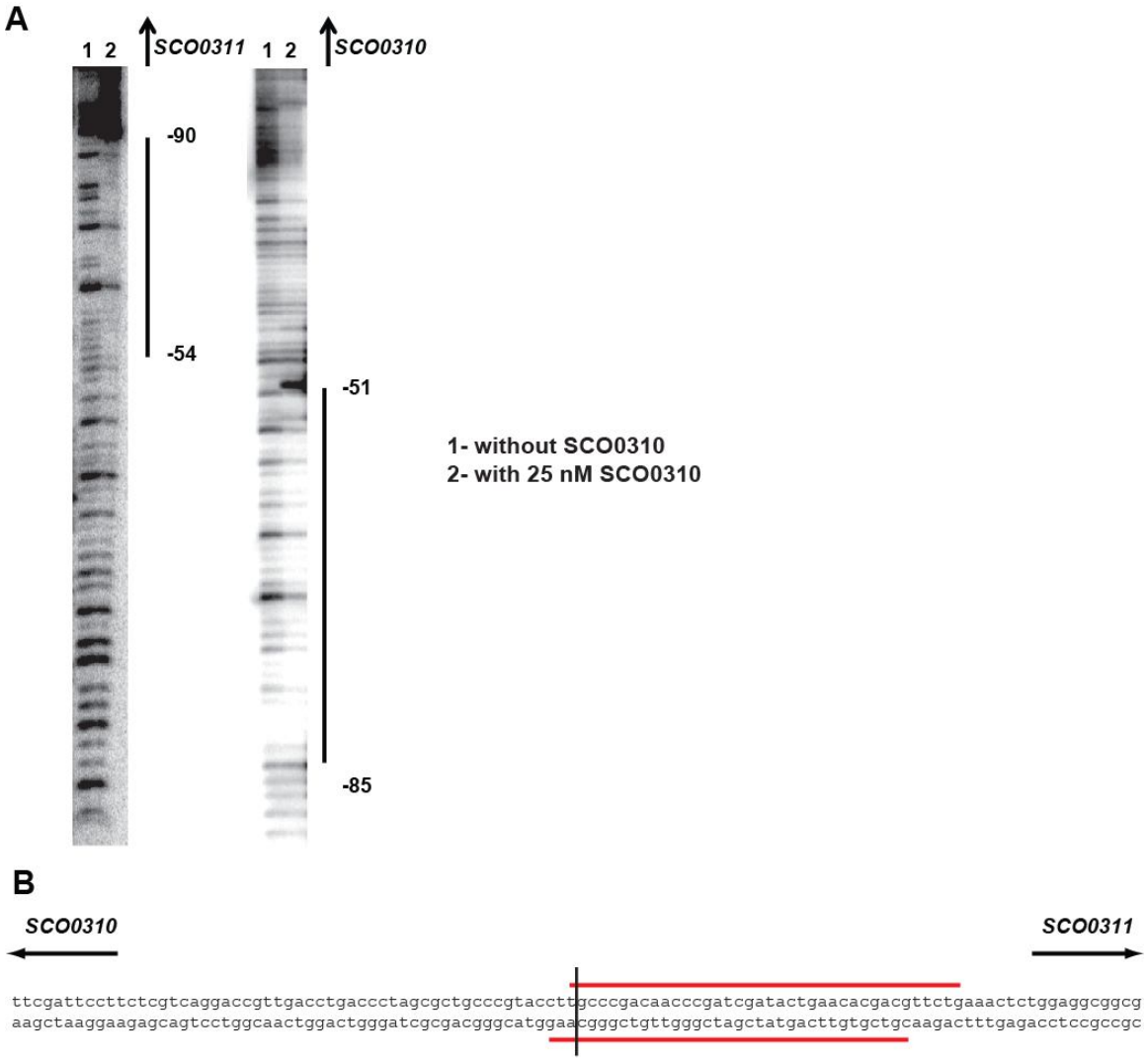

Figure 4.6. DNaseI footprinting assay with SCO0310. (A) DNaseI was added to the radiolabeled *SCO0310/SCO0311* intergenic DNA in the absence or presence of SCO0310. For the left gel, the primer that was extended toward the *SCO0311* gene (i.e. top strand of Figure 4.6B) was labeled at 5′-end to prepare the probe, while the other primer extended toward *SCO0310* was labeled to prepare the probe for the right gel. The regions protected by SCO0310 are indicated by solid vertical lines. The numbers beside each line indicate the start and end positions of the protected region relative to the translational start site of the *SCO0310* gene. **(B)** Sequence of the *SCO0310/SCO0311* intergenic DNA is shown (oriented such that the *SCO0310* gene starts after the left end). Horizontal red lines indicate the regions protected by SCO0310 on each DNA strand, while a vertical black line indicates the center of the intergenic region.

4.3.4 Construction of a one-plasmid SCO0310-based biosensor

According to the footprinting data, 40 bp in the protected region should possess the SCO0310 binding sites. If this is the case, an SCO0310-based biosensor would require the incorporation of this 40 bp sequence in the synthetic promoter fragment. Given that 17 bp (Figure 4.7A) is the optimal size of a spacer DNA between the -35 and -10 elements of a functional promoter [247] and that a change in this separation by as short as 1 bp can drastically reduce the promoter activity [248], our established strategy would not work (Figure 4.7B).

Instead, an alternative method could involve placing the 40 bp fragment downstream of the functional promoter elements (Figure 4.7C, the *tetA* promoter was used in this study) such that the interaction of SCO0310 with this fragment would block the passage of RNA polymerase even if it interacts with the promoter. We prepared a synthetic promoter fragment using this “road-block” strategy and cloned it into pYR in the absence (pYR0310O) or presence (pYR0310OR) of the *SCO0310* gene. These plasmids, along with the empty vector pYR, were separately introduced to *E. coli* and we compared the bioluminescence production from the resulting strains in the presence of varying amounts of L-arabinose.

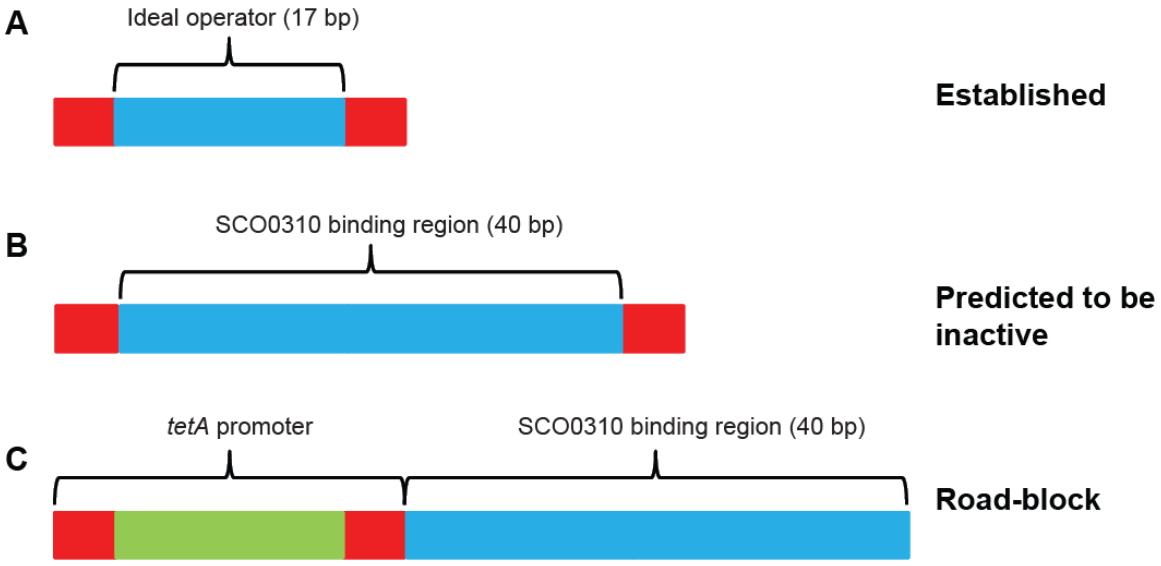

Figure 4.7. Construction of a promoter fragment for the SCO0310 reporter using the 40 bp protected region identified by DNaseI footprinting. (A) Promoter fragments have been designed to contain 17 bp operator sequences between the promoter regions (red boxes) **(B)** If the entire 40 bp fragment was inserted between the promoter elements, the spacing would be too long for the promoter to be functional. **(C)** One possible solution is to put the fragment downstream of a known functional promoter (the *tetA* promoter was used in this study).

As shown in Figure 4.8, the pYR0310O strain conferred significant bioluminescence compared to the pYR control strain at all of the L-arabinose concentrations tested, suggesting that the presence of the 40 bp fragment did not prevent *tetA* promoter activity. Unlike what was observed from the TetR-based reporter (Figure 4.2), no partial repression by SCO0310 was observed from pYR0310OR in the absence of L-arabinose. However, the bioluminescence production from the pYR0310OR strain was reduced as arabinose was added (4-fold reduction with 0.002% arabinose and complete elimination with 0.02% and 0.2% arabinose). These data indicate that SCO0310 is repressing *lux* expression

by interacting with the operator-containing 40 bp fragment and thus that this reporter is functional.

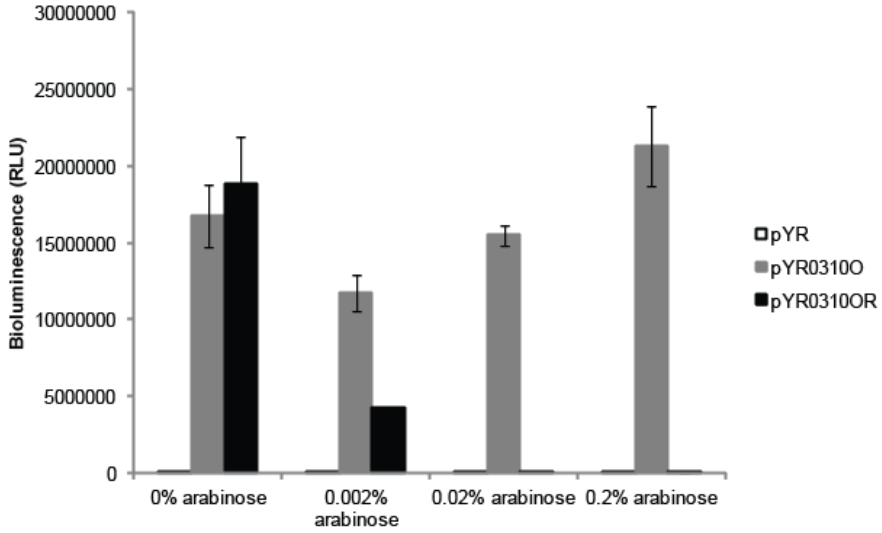

Figure 4.8. The SCO0310-based biosensor using the "road-block" strategy is functional. Varying amounts of L-arabinose were added to 1 ml of the *E. coli* biosensors harboring pYR, pYR0310O, or pYR0310OR. All values were measured in RLUs, and error bars indicate +/- 1 standard deviation of values obtained from three independent readings.

4.4 Discussion

In this study, we have described two strategies to improve our biosensor mechanism in terms of its sensitivity and ease of construction. First, the utilization of pYR allows us to introduce a TFR-encoding gene and its cognate operator sequence on the same vector and, more importantly, the TFR expression level on this plasmid can be fine-tuned by the arabinose-inducible promoter to prevent potential over-repression. As demonstrated with the increased sensitivity of the TetR-based biosensor for tetracycline, this one-plasmid reporter system should enable our biosensors to detect their cognate ligands at lower concentrations.

Higher sensitivity for small molecules is especially useful when we screen biosensors against antibiotics, as it is highly important to observe induction before the drugs reach minimum inhibitory concentrations. In addition, this improved sensitivity makes our biosensor more amenable to high-throughput screening. Promisingly, the one-plasmid TetR-based biosensor has been successfully used to specifically detect tetracycline and its derivatives when it was screened against a compound library of 3,000 bioactive compounds (Cuthbertson, unpublished data).

In addition to testing the one-plasmid reporter system, we have also developed the “road-block” strategy that can be applied, in combination with DNase I footprinting assays, to construct reporters for the TFRs whose exact operator sequences cannot be predicted through simple sequence analysis. This

discovery is crucial as our study of the *S. coelicolor* TFRs suggests that operator prediction is challenging for most of these proteins, and this trend might be observed from TFRs encoded by the majority of other bacterial genomes in the public databases. Furthermore, the fact that a TFR binding site is not used as a spacer DNA between the -10 and -35 elements in this method implies that length of operator is no longer a limiting factor when designing a synthetic promoter. This is particularly beneficial for the TFRs that interact with long operators, and previous studies have indicated that some TFRs such as QacR (28 bp) and EthR (55 bp) recognize the DNA sequences that are much longer than 17 bp (Previously described in Chapter 1, section 4.3).

One possible concern with the “road-block” strategy is the ability of RNA polymerase to overcome the protein obstacles and continue the transcription process. Bacterial RNA polymerase pauses and backtracks for several nucleotides when it encounters a protein road-block [249,250], but a number of transcription elongation factors involved in rescuing the backtracked RNA polymerase to resume transcription and read through the blocked site have been identified [251,252,253,254]. In addition, Epshtein and colleagues have demonstrated another anti-road-block strategy that involves multiple RNA polymerase molecules initiating transcription from the same promoter [255]. This model is based on the “pushing” effect exerted by the trailing RNA polymerase molecules on the leading one – which has been interfered by a protein obstacle – to translocate it forward.

It is noteworthy that the bioluminescence production from pYR0310OR (the SCO0310 reporter constructed using the “road-block” strategy) was repressed only in the presence of arabinose (Figure 4.8), whereas strong partial repression was observed from pYRtetOR (the TetR reporter constructed without using the “road-block” strategy) even in the absence of arabinose (Figure 4.2). It is possible that TetR has a higher affinity for its operator sequence than SCO0310 does for its own. However, if these observations are attributed to the anti-road-block strategies possessed by bacteria, it clearly opposes our effort to minimize the TFR concentration required to observe repression from the reporters for optimal ligand screening.

Importantly, some of the elongation factors involved in reactivating the paused RNA polymerase are dispensable for cell viability under laboratory conditions [252]. Therefore, it would be worth testing the reporters (e.g. pYR0310OR) using the *E. coli* host in which the genes encoding these elongation factors are knocked out. While the “pushing” mechanism is proposed to be a more efficient and generally used strategy in bacteria [255], sufficient distance between a promoter and a blocked site is required for this to take place (as multiple RNA polymerases have to be allowed to initiate transcription and apply the “pushing” effect). This emphasizes the importance of conducting DNaseI footprinting assays for designing a synthetic promoter in our “road-block” strategy such that a TFR binding site can be located as closely as possible to the promoter elements.

The improvements we have made (and will make) on our *E. coli*-based biosensor mechanism should allow us, in principle, to construct a reporter for any TFR and identify its ligand(s) more effectively. Therefore, we anticipate that these increased applicability and sensitivity make our *E. coli*-based reporter system a more powerful means of characterizing bacterial small-molecule responses mediated by TFRs.

4.5 Conclusion

We have tested the use of the one-plasmid reporter system and the "road-block" strategy. Our analysis indicates that the one-plasmid system should allow our biosensors to detect their cognate ligands at lower concentrations than they do with the two-plasmid system. Discovery of the "road-block" strategy is significant as it allows us to prepare a synthetic promoter with an operator with any size and to construct reporters for TFRs whose exact operator sequences cannot be predicted through sequence analysis. Therefore, the combined use of the two strategies would make our biosensor mechanism more effective and applicable to investigate a larger number of TFRs and identify their small-molecule ligands.

4.6 Materials and Methods

4.6.1 Bacterial strains and culture conditions

Bacterial strains used in this study are described in Table 4.2. *E. coli* XL1-blue was used during general cloning process. In order to propagate pYR-based plasmids for biosensor construction, *E. coli* TOP10 was used as a host. This strain is capable of transporting L-arabinose but not capable of metabolizing it. This is an important feature as TFR expression on the pYR-based reporters is under the control of an arabinose-inducible promoter. All *E. coli* strains were grown using Luria Broth (LB) or LB agar medium, containing the appropriate antibiotics when required (100 μg/ml ampicillin, 50 μg/ml kanamycin, and 37 μg/ml chloramphenicol).

Table 4.2. Bacterial strains used in this work

Strain/Plasmid	Description	Source
E. coli		
BL21(DE3)	F^- *dcm ompT hsdS*($r_B^- m_B^-$) *gal met* λ(DE3)	Novagen
TOP10	F- *mcrA* Δ(*mrr-hsdRMS-mcrBC*) ψ80*lacZ*ΔM15 Δ(*ara-leu*)*7697 galE15 galK16 rpsL*(Str^R) *endA1* λ^-	Invitrogen
XL1-blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac*[F′ *proAB lacI*^{q} ZΔM15 Tn*10*(Tet^r)]	Stratagene

4.6.2 Plasmids, primers, and sequencing

The plasmids and primers used in this study are listed in Table 4.3. Oligonucleotide primers used in this study were obtained from the Institute for Molecular Biology and Biotechnology (MOBIX) facility at McMaster University. Polymerase chain reactions (PCR) were carried out using Vent DNA polymerase (New England Biolabs). DNA sequencing was carried out by the MOBIX facility.

Table 4.3. Plasmids and primers used in this study

Plasmid	Description (selection marker)[a]	Background	Reference
pYR	Promoterless *luxCDABE* reporter (Kan^r)	pSC101	[245]
pYRtetO	TetR reporter based on the TetR operator (Kan^r)	pYR	[245]
pYRtetOR	TetR reporter based on the TetR operator and the *tetR* gene(Kan^r)	pYR	[245]
pYR0310O	SCO0310 reporter based on the SCO0310 binding region (Kan^r)	pYR	This work
pYR0310OR	SCO0310 reporter based on the SCO0310 binding region (Kan^r)	pYR	This work
pET28a-SCO0310	SCO0310-overexpressing vector for protein purification (Kan^r)	pET28a	This work
Primer	**Sequence (5′ → 3′)[b, c]**	**Purpose**	
SCO0310IR-F	ttcgattccttctcgtcaggacc	Forward primer for amplifying the intergenic region between *SCO0310* and *SCO0311*	
SCO0310IR-R	cgccgcctccagagtttc	Reverse primer for amplifying the intergenic region between *SCO0310* and *SCO0311*	
SCO0310IRL-1	ttcgattccttctcgtcaggaccgttgacctgaccctagcgctgcccgtacctt	+ strand oligonucleotide for preparing the half intergenic region closer to *SCO0310* (the “L” probe in the text)	
SCO0310IRL-2	aaggtacgggcagcgctagggtcaggtcaacggtcctgacgagaaggaatcgaa	- strand oligonucleotide for preparing the half intergenic region closer to *SCO0310* (the “L” probe in the text)	
SCO0310IRR-1	gcccgacaacccgatcgatactgaacacgacgttctgaaactctggaggcggcg	+ strand oligonucleotide for preparing the half intergenic region closer to *SCO0311* (the “R” probe in the text)	

SCO0310IRR-2	cgccgcctccagagtttcagaacgtcgtgttca gtatcgatcgggttgtcgggc	- strand oligonucleotide for preparing the half intergenic region closer to *SCO0311* (the "R" probe in the text)
SCO0310OE-F	ggggggCATATGgacatgaagcccgaag	Forward primer for SCO0310 over-expression
SCO0310OE-R	gggggGAATTCtacacgggttcggcgcc	Reverse primer for SCO0310 over-expression
0310-1	cttgacactctatcattgatagagttattttac ccttgcccgacaacccgatcgatactgaacacg acgttctgggcc	+ strand oligonucleotide for preparing the synthetic promoter based on SCO0310-binding region
0310-2	ccagaacgtcgtgttcagtatcgatcgggttgt cgggcaagggtaaaataactctatcaatgatag agtgtcaaggtac	- strand oligonucleotide for preparing the synthetic promoter based on SCO0310-binding region
SCO0310-F	gGAATTCgacatgaagcccgaagaccgccg	Forward primer for amplifying the *SCO0310* gene for pYR0310OR
SCO0310-R	gCTCGAGctgtcgctccgctcgttc	Reverse primer for amplifying the *SCO0310* gene for pYR0310OR

[a] Kan^r, kanamycin resistance
[b] compatible cohesive ends introduced by these oligonucleotides are underlined.
[c] restriction endonuclease recognition sequences introduced by these oligonucleotides are shown in capital letters.

4.6.3 Expression and Purification of His_6-SCO0310

A fragment encoding SCO0310 was PCR-amplified, digested with *Nde*I and *Eco*RI, and then ligated to pET28a, giving pET28a-SCO0310. An *E. coli* BL21(DE3) culture containing pET28a was grown at 37°C to an optical density (OD_{600}) ~0.4-0.6 and were then induced with 1 mM isopropyl β-D-thiogalactopyranoside for 20 hours at 37°C. The cell pellet was resuspended in 25 ml lysis buffer (10 mM Tris-HCl (pH7.8), 150 mM NaCl and 1x Complete Mini EDTA-free protease inhibitor (Roche)) and lysed by 3x passage through a French press cell. 15 ml of Ni-NTA agarose resin (Qiagen) was added and the slurry was incubated for 30 minutes with gentle shaking. The resin was loaded

onto a column, washed with 500 ml of 10 mM Tris-HCl (pH7.8), 150 mM NaCl, and 50 mM imidazole, and then eluted with 25 ml 10 mM Tris-HCl (pH7.8), 150 mM NaCl, and 0.5 M imidazole. His_6-SCO0310 was transferred to 10 mM Tris-HCl (pH8), 500 mM NaCl and 20% glycerol by buffer exchange using a 10 kDa cut-off Amicon Ultra Centrifugal Filter Device (Millipore) and was stored at -80°C. The protein concentration was determined by Bradford assay (Bio-Rad).

4.6.4 EMSAs

The primers SCO0310IR-F and SCO0310IR-R along with the *S. coelicolor* M145 chromosomal DNA (as template) were used in PCR reactions to isolate the double stranded DNA fragment containing the entire intergenic region between *SCO0310* and *SCO0311*. In addition the pairs of SCO0310IRL-1/2 and SCO0310IRR-1/2 were annealed together to serve as probes for mobility shift assays. The DNA fragments were end-labeled using [γ-^{32}P]ATP (Amersham) and T4 polynucleotide kinase (New England Biolabs).

Labeled probe (12.7 fmole), varying amounts of purified protein and 90 ng of salmon sperm DNA were used in 15 µl reactions containing 1x EMSA reaction buffer (10 mM Tris-Cl (pH 7.8), 150 mM NaCl, 2 mM DTT and 10% glycerol). Reactions were incubated at 30°C for 10 minutes and were fractionated on 12% non-denaturing polyacrylamide gels containing 1.5% glycerol. The gels were exposed using a phosphor screen (Amersham) and bands were detected using a PhosphorImager (Molecular Dynamics).

4.6.5 DNase I footprinting assays

The same pair of primers to amplify the entire intergenic sequence in the previous EMSA were used for DNase I footprinting. The probes in the assays were prepared by PCR using one unlabeled primer and one 5′-end labeled primer (using [γ-^{32}P] ATP and T4 polynucleotide kinase). 150,000 cpm of a labeled DNA probe, 25 nM of purified SCO0310, and 90 ng of salmon sperm DNA (Sigma-Aldrich) were used in 40 µl reactions containing 1x EMSA reaction buffer. After the reactions were incubated at 30°C for 10 minutes, 10 µl DNase I solution (0.8 U in 10 mM $CaCl_2$) was added. The incubation was continued for 60 seconds at room temperature and reactions were stopped by adding 140 µl DNase I stop solution (200 mM NaCl, 30 mM EDTA, and 1% SDS). The digested samples were then precipitated with ethanol and resuspended in 5 µl Stop Solution (from Thermosequenase Cycle Sequencing Kit (USB): 95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol). Samples were heated at 80°C for 3 minutes, cooled on ice, and separated on 8% polyacrylamide/7 M urea sequencing gels. Dried gels were exposed using a phosphor screen (Bio-Rad) and bands were detected using a PhosphorImager (Molecular Dynamics). Sequencing ladders were prepared using Thermosequenase Cycle Sequencing Kit (USB).

4.6.6 Construction of pYR-based reporter plasmids for SCO0310

The TetR reporters, pYRtetO and pYRtetOR, were previously prepared [245]. The 0310-1 and 0310-2 oligonucleotides were annealed to prepare a synthetic promoter consisting of the 40 bp SCO0310 binding region located downstream of the *tetA* promoter. This fragment was introduced into the *Kpn*I-*Apa*I sites of pYR to give a reporter plasmid pYR0310O. The *SCO0310* gene was amplified by PCR using the primers SCO0310-F and SCO0310-R. After digestion with *Eco*RI and *Xho*I, the fragment was ligated to pYR0310O to give pYR0310OR.

4.6.7 Bioluminescence measurements

Isolated *E. coli* TOP10 colonies containing individual reporter plasmids were used to inoculate one-milliliter amounts of reporter cultures in the presence of varying amounts of L-arabinose. The cultures were grown for 16-20 hours were subcultured in fresh LB media containing kanamycin and varying amounts of L-arabinose. These subcultures were incubated for additional 4-5 hours before measuring luminescence using a Lumat 9507 luminometer (Bertholt Technologies). In some cases, *E. coli* reporter cultures were supplemented with purified tetracycline.

SUMMARY AND CONCLUDING REMARKS

Investigating transcriptional regulators is crucial to our understanding of cellular physiology, and characterization of TetR family transcriptional regulators (TFRs) is highly desirable due to their abundance and ubiquity in prokaryotes. A central theme of TFRs involves a functional relationship that seems to present between their small-molecule ligands and regulatory target genes, and this paradigm has been the driving force of my graduate studies. In the work presented here, I have therefore developed generally applicable tools for characterizing these two basic elements (i.e. ligands and targets) of the biological roles of TFRs.

In the first part of my thesis, I described the use of genome context as a predictive tool to identify regulatory targets of TFRs. The wealth of bacterial genome sequences available in the public databases was instrumental in this approach. I have shown that the majority of the genes encoding TFRs are divergently transcribed from their immediate neighbors and most of these TFRs are likely to regulate their neighboring genes. The “200 bp” rule presented in this study will serve as an important starting point for predicting TFR target genes through simple genome analysis. Importantly, it has to be emphasized that this was the first study to directly address what many research groups have speculated and have been misinformed about TFRs: while the majority of TFR-encoding genes exhibit the *tetR*/*tetA* genetic configuration with their adjacent genes and are predicted to be repressors of those neighbors (like TetR), efflux-

mediated antibiotic resistance is not the most common cellular function regulated by TFRs. Instead, there is a much greater diversity of biochemical functions under the control of these transcription factors and therefore, my work suggests that investigation of TFRs would improve our knowledge in microbiology more extensively than previously anticipated.

Analysis of TFRs having alternative genomic orientations with respect to their neighbors (especially the ones with ambiguous orientation, Figure 2.1C) will further improve our ability to predict their regulatory target genes. Finding broadly applicable rules for these transcription factors is expected to be more challenging due to the fact that examination of intergenic separations between the TFR-coding genes and their neighboring genes might not be sufficient for reliable predictions (e.g. it is hard to predict if two genes are co-transcribed unless these genes are overlapped or separated by a noticeably short length of DNA). However, I anticipate that genome context of these TFRs and their relative neighbors can still provide a powerful indication of local regulation, as genetic arrangements of the TFR- and target-encoding genes in one organism – when clustered together – tend to be conserved in other organisms which possess the orthologs of both.

For example, I have previously described AcnR (Chapter 1, section 4.3) whose gene is co-transcribed with its target gene *acn* in *Corynebacterium glutamicum* (Figure 1.9A) [114]. In my view, genome analysis is not sufficient to predict that these two genes are co-transcribed as their annotated open reading

frames are separated by >100 bp. Therefore, AcnR would have been considered as a TFR with ambiguous orientation according to my classification method. However, Krug and colleagues correctly anticipated the local regulation of *acn* by AcnR based on the fact that the *acn-acnR* configuration is conserved by the orthologs of the two genes in all corynebacteria [114].

In addition, our lab is currently investigating a previously uncharacterized TFR from *Streptomyces coelicolor*, SCO1699. Similar to the genetic arrangement of the *dhaQ-dhaS* pair represented in Figure 1.10, the translational stop sites of *SCO1699* and its neighboring gene *SCO1698* face each other (making SCO1699 another example of TFRs with ambiguous orientation). SCO1699 has been shown to bind upstream of *SCO1698 in vitro* (conducted by a project student Jana Cmorejova) and thus has been proposed to regulate this gene. Probable orthologs of *SCO1698* and *SCO1699* have been identified in all of the sequenced *Streptomyces* genomes (Cuthbertson, unpublished data). Consistent with the situation exhibited by AcnR, the relative organization of these genes is identical to the *SCO1698-SCO1699* pair. Importantly therefore, in addition to analyzing a single genome to search for putative targets (as described in this thesis), I suggest that careful examination of multiple genomes to look for possible conservation in the relative TFR-neighbor arrangements can offer another practical method for target prediction.

In the second part of the thesis, I explored the use of whole-cell based reporter systems to identify small-molecule ligands of TFRs. Before I began this work, we lacked proper reporter systems that are generally applicable to investigate TFRs encoded by various bacterial genomes. Now, we can create standardized reporters for almost all TFRs in streptomycetes and closely related actinomycetes using natural promoters in *Streptomyces* (Gram-positive, high GC) or through the design of synthetic promoters in *Escherichia coli* (Gram-negative, lower GC)

Each of these two reporter systems has advantages over the other but they also complement with each other very well. For example, the use of the *E. coli*-based reporter system to study TFRs from *Streptomyces* is definitely attractive due to the faster-growing trait of the host organism. In addition, heterologous investigation also allows us to observe more direct relationships between TFRs and their cognate operators/ligands. On the other hand, the *Streptomyces*-based reporter system leads us to more accurately evaluate regulatory activity of TFRs on their natural target promoters, and it has a major advantage concerning the membrane permeability that allows easier access of screened molecules into the cells. To further improve the reporter system based on a Gram-positive organism, the use of faster-growing streptomycetes such as *Streptomyces venezuelae* and/or related actinomycetes such as *Rhodococcus* as general reporter hosts could be investigated.

The fact that actinomycetes are rich sources of TFRs makes development of the *Streptomyces*-based reporter system highly significant. In addition, while the work presented in this thesis has primarily focused on the TFRs from streptomycetes, my experience with the *E. coli*-based biosensor mechanism (i.e. TetR from *E. coli*; and ActR, SCO7222, and SCO0310 from *S. coelicolor*) offers promise that this system is capable of accommodating TFRs from diverse organisms that grow very differently from each other. Therefore, I believe that the combined use of all the predictive and screening tools that I have created will pave the way for us to characterize the physiological roles of the majority of TFRs that have been and will be identified.

In my view, one of the top priorities for our lab should be to biologically investigate the hundreds of TFRs whose ligands and regulatory target genes can be predicted by our phylogenetic framework (described in Chapter 1, section 4.4) and the "200 bp" rule, respectively. Once perfect correlation between the predictions and reporter data is obtained for a TFR of interest (regarding its ligand and target), this information can be used to elucidate a biochemical function of the target gene product (e.g. if the ligand is an antibiotic, it can be tested to see if introduction of its target gene into a susceptible organism has an effect on bacterial survival in the presence of the drug). Importantly, it should be noted that the new knowledge obtained from these TFRs and their target genes can be integrated into our phylogenetic and genomic methodologies to further improve their reliability and applicability for the remaining TFRs.

The ultimate goal of the lab is to investigate all TFRs from organisms of our particular interest. For exploring the extent of biological functions under the regulation of TFRs in greater depth, organisms with large numbers of these transcription factors such as members in the genus *Streptomyces* (most of the sequenced genomes possess >100 TFRs) can be good candidates to consider. On the other hand, while genomes of pathogens such as *Acinetobacter baumanii* (20 TFRs), *Mycobacterium tuberculosis* (49 TFRs), *Pseudomonas aeruginosa* (40 TFRs), *Salmonella enteric* serovar Typhimurium (13 TFRs), and *Staphylococcus aureus* (5 TFRs) encode fewer TFRs, these proteins should be highly attractive to characterize due to their potential clinical impact.

I believe that the work conducted as part of my doctoral thesis provides a solid foundation to achieve this goal and fully characterize one of the most important families of prokaryotic transcription factors. Given the overwhelming number of bacterial genome sequences available, our approaches to examine TFRs in a global scale should be highly beneficial; moreover, the advantage of these strategies is the fact that they can be further improved upon over time. I suggest that research in other one-component transcriptional regulator families (see Table 1.1) would potentially benefit from similar approaches as well, and this will help us gain better insights on how bacteria sense challenges from their surroundings and respond to them for growth and survival. Importantly, knowledge of the regulatory mechanisms governing biological processes such as (primary/secondary) metabolism and antibiotic resistance in bacteria could lead

to new ideas for human benefits, including enhanced productions of useful metabolites (e.g. amino acids and antibiotics) and development of better drugs.

REFERENCES

1. Dorman CJ (2004) H-NS: a universal regulator for a dynamic genome. Nat Rev Microbiol 2: 391-400.
2. Wosten MM (1998) Eubacterial sigma-factors. FEMS Microbiol Rev 22: 127-150.
3. Huffman JL, Brennan RG (2002) Prokaryotic transcription regulators: more than just the helix-turn-helix motif. Curr Opin Struct Biol 12: 98-106.
4. Stock AM, Robinson VL, Goudreau PN (2000) Two-component signal transduction. Annu Rev Biochem 69: 183-215.
5. Merino E, Yanofsky C (2005) Transcription attenuation: a highly conserved regulatory strategy used by bacteria. Trends Genet 21: 260-264.
6. Wassarman KM, Zhang A, Storz G (1999) Small RNAs in *Escherichia coli*. Trends Microbiol 7: 37-45.
7. Bastet L, Dube A, Masse E, Lafontaine DA (2011) New insights into riboswitch regulation mechanisms. Mol Microbiol 80: 1148-1154.
8. Caron MP, Lafontaine DA, Masse E (2010) Small RNA-mediated regulation at the level of transcript stability. RNA Biol 7: 140-144.
9. Kobir A, Shi L, Boskovic A, Grangeasse C, Franjevic D, et al. (2011) Protein phosphorylation in bacterial signal transduction. Biochim Biophys Acta 1810: 989-994.
10. Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. Nature 475: 324-332.
11. Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. J Mol Biol 3: 318-356.
12. Arsene F, Tomoyasu T, Bukau B (2000) The heat shock response of *Escherichia coli*. Int J Food Microbiol 55: 3-9.
13. Shimotsu H, Kuroda MI, Yanofsky C, Henner DJ (1986) Novel form of transcription attenuation regulates expression the *Bacillus subtilis* tryptophan operon. J Bacteriol 166: 461-471.
14. Reidl J, Klose KE (2002) *Vibrio cholerae* and cholera: out of the water and into the host. FEMS Microbiol Rev 26: 125-139.
15. Browning DF, Busby SJ (2004) The regulation of bacterial transcription initiation. Nat Rev Microbiol 2: 57-65.
16. Parkinson JS, Kofoid EC (1992) Communication modules in bacterial signaling proteins. Annu Rev Genet 26: 71-112.
17. Hoch JA (2000) Two-component and phosphorelay signal transduction. Curr Opin Microbiol 3: 165-170.
18. Dutta R, Qin L, Inouye M (1999) Histidine kinases: diversity of domain organization. Mol Microbiol 34: 633-640.
19. Marina A, Waldburger CD, Hendrickson WA (2005) Structure of the entire cytoplasmic portion of a sensor histidine-kinase protein. EMBO J 24: 4247-4259.

20. Casino P, Rubio V, Marina A (2009) Structural insight into partner specificity and phosphoryl transfer in two-component signal transduction. Cell 139: 325-336.
21. Ulrich LE, Koonin EV, Zhulin IB (2005) One-component systems dominate signal transduction in prokaryotes. Trends Microbiol 13: 52-56.
22. Harrison SC, Aggarwal AK (1990) DNA recognition by proteins with the helix-turn-helix motif. Annu Rev Biochem 59: 933-969.
23. Weiss V, Kramer G, Dunnebier T, Flotho A (2002) Mechanism of regulation of the bifunctional histidine kinase NtrB in *Escherichia coli*. J Mol Microbiol Biotechnol 4: 229-233.
24. Calogero S, Gardan R, Glaser P, Schweizer J, Rapoport G, et al. (1994) RocR, a novel regulatory protein controlling arginine utilization in *Bacillus subtilis*, belongs to the NtrC/NifA family of transcriptional activators. J Bacteriol 176: 1234-1241.
25. Wright GD (2007) The antibiotic resistome: the nexus of chemical and genetic diversity. Nat Rev Microbiol 5: 175-186.
26. Giedraitiene A, Vitkauskiene A, Naginiene R, Pavilonis A (2011) Antibiotic resistance mechanisms of clinically important bacteria. Medicina (Kaunas) 47: 137-146.
27. Wilson M, DeRisi J, Kristensen HH, Imboden P, Rane S, et al. (1999) Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. Proc Natl Acad Sci U S A 96: 12833-12838.
28. Alland D, Kramnik I, Weisbrod TR, Otsubo L, Cerny R, et al. (1998) Identification of differentially expressed mRNA in prokaryotic organisms by customized amplification libraries (DECAL): the effect of isoniazid on gene expression in *Mycobacterium tuberculosis*. Proc Natl Acad Sci U S A 95: 13227-13232.
29. Chopra I, Roberts M (2001) Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiol Mol Biol Rev 65: 232-260 ; second page, table of contents.
30. Chopra I, Hawkey PM, Hinton M (1992) Tetracyclines, molecular and clinical aspects. J Antimicrob Chemother 29: 245-277.
31. Livermore DM (2005) Tigecycline: what is it, and where should it be used? J Antimicrob Chemother 56: 611-614.
32. Argast M, Beck CF (1985) Tetracycline uptake by susceptible *Escherichia coli* cells. Arch Microbiol 141: 260-265.
33. McMurry LM, Cullinane JC, Petrucci RE, Jr., Levy SB (1981) Active uptake of tetracycline by membrane vesicles from susceptible *Escherichia coli*. Antimicrob Agents Chemother 20: 307-313.
34. Schnappinger D, Hillen W (1996) Tetracyclines: antibiotic action, uptake, and resistance mechanisms. Arch Microbiol 165: 359-369.
35. Levy SB (1984) Resistance to the tetracyclines. Orlando, Fla: Academic Press.

36. Hughes VM, Datta N (1983) Conjugative plasmids in bacteria of the 'pre-antibiotic' era. Nature 302: 725-726.
37. Mendez B, Tachibana C, Levy SB (1980) Heterogeneity of tetracycline resistance determinants. Plasmid 3: 99-108.
38. McMurry L, Petrucci RE, Jr., Levy SB (1980) Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in *Escherichia coli*. Proc Natl Acad Sci U S A 77: 3974-3977.
39. Yamaguchi A, Udagawa T, Sawai T (1990) Transport of divalent cations with tetracycline as mediated by the transposon Tn*10*-encoded tetracycline resistance protein. J Biol Chem 265: 4809-4813.
40. Connell SR, Trieber CA, Dinos GP, Einfeldt E, Taylor DE, et al. (2003) Mechanism of Tet(O)-mediated tetracycline resistance. EMBO J 22: 945-953.
41. LeBlanc DJ, Lee LN, Titmas BM, Smith CJ, Tenover FC (1988) Nucleotide sequence analysis of tetracycline resistance gene *tetO* from *Streptococcus mutans* DL5. J Bacteriol 170: 3618-3626.
42. Speer BS, Bedzyk L, Salyers AA (1991) Evidence that a novel tetracycline resistance gene found on two *Bacteroides* transposons encodes an NADP-requiring oxidoreductase. J Bacteriol 173: 176-183.
43. Yang W, Moore IF, Koteva KP, Bareich DC, Hughes DW, et al. (2004) TetX is a flavin-dependent monooxygenase conferring resistance to tetracycline antibiotics. J Biol Chem 279: 52346-52352.
44. Tamura N, Konishi S, Iwaki S, Kimura-Someya T, Nada S, et al. (2001) Complete cysteine-scanning mutagenesis and site-directed chemical modification of the Tn*10*-encoded metal-tetracycline/H^+ antiporter. J Biol Chem 276: 20330-20339.
45. Paulsen IT, Brown MH, Skurray RA (1996) Proton-dependent multidrug efflux systems. Microbiol Rev 60: 575-608.
46. Marger MD, Saier MH, Jr. (1993) A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. Trends Biochem Sci 18: 13-20.
47. Yamaguchi A, Adachi K, Akasaka T, Ono N, Sawai T (1991) Metal-tetracycline/H^+ antiporter of *Escherichia coli* encoded by a transposon Tn*10*. Histidine 257 plays an essential role in H^+ translocation. J Biol Chem 266: 6045-6051.
48. Levy SB (1989) Evolution and spread of tetracycline resistance determinants. J Antimicrob Chemother 24: 1-3.
49. Alekshun MN, Levy SB (2007) Molecular mechanisms of antibacterial multidrug resistance. Cell 128: 1037-1050.
50. Hillen W, Berens C (1994) Mechanisms underlying expression of Tn*10* encoded tetracycline resistance. Annu Rev Microbiol 48: 345-369.
51. Bertrand KP, Postle K, Wray LV, Jr., Reznikoff WS (1983) Overlapping divergent promoters control expression of Tn*10* tetracycline resistance. Gene 23: 149-156.

52. Bertrand KP, Postle K, Wray LV, Jr., Reznikoff WS (1984) Construction of a single-copy promoter vector and its use in analysis of regulation of the transposon Tn*10* tetracycline resistance determinant. J Bacteriol 158: 910-919.
53. Meier I, Wray LV, Hillen W (1988) Differential regulation of the Tn*10*-encoded tetracycline resistance genes *tetA* and *tetR* by the tandem *tet* operators O1 and O2. EMBO J 7: 567-572.
54. Takahashi M, Altschmied L, Hillen W (1986) Kinetic and equilibrium characterization of the Tet repressor-tetracycline complex by fluorescence measurements. Evidence for divalent metal ion requirement and energy transfer. J Mol Biol 187: 341-348.
55. Takahashi M, Degenkolb J, Hillen W (1991) Determination of the equilibrium association constant between Tet repressor and tetracycline at limiting Mg^{2+} concentrations: a generally applicable method for effector-dependent high-affinity complexes. Anal Biochem 199: 197-202.
56. Orth P, Saenger W, Hinrichs W (1999) Tetracycline-chelated Mg^{2+} ion initiates helix unwinding in Tet repressor induction. Biochemistry 38: 191-198.
57. Kisker C, Hinrichs W, Tovar K, Hillen W, Saenger W (1995) The complex formed between Tet repressor and tetracycline-Mg^{2+} reveals mechanism of antibiotic resistance. J Mol Biol 247: 260-280.
58. Epe B, Woolley P (1984) The binding of 6-demethylchlortetracycline to 70S, 50S and 30S ribosomal particles: a quantitative study by fluorescence anisotropy. EMBO J 3: 121-126.
59. Hinrichs W, Kisker C, Duvel M, Muller A, Tovar K, et al. (1994) Structure of the Tet repressor-tetracycline complex and regulation of antibiotic resistance. Science 264: 418-420.
60. Orth P, Schnappinger D, Hillen W, Saenger W, Hinrichs W (2000) Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. Nat Struct Biol 7: 215-219.
61. Reichheld SE, Yu Z, Davidson AR (2009) The induction of folding cooperativity by ligand binding drives the allosteric response of tetracycline repressor. Proc Natl Acad Sci U S A 106: 22263-22268.
62. Ramos JL, Martinez-Bueno M, Molina-Henares AJ, Teran W, Watanabe K, et al. (2005) The TetR family of transcriptional repressors. Microbiol Mol Biol Rev 69: 326-356.
63. Yu Z, Reichheld SE, Savchenko A, Parkinson J, Davidson AR (2010) A comprehensive analysis of structural and sequence conservation in the TetR family transcriptional regulators. J Mol Biol 400: 847-864.
64. Pareja E, Pareja-Tobes P, Manrique M, Pareja-Tobes E, Bonal J, et al. (2006) ExtraTrain: a database of Extragenic regions and Transcriptional information in prokaryotic organisms. BMC Microbiol 6: 29.

65. Sahota G, Stormo GD (2010) Novel sequence-based method for identifying transcription factor binding sites in prokaryotic genomes. Bioinformatics 26: 2672-2677.
66. Schell MA (1993) Molecular biology of the LysR family of transcriptional regulators. Annu Rev Microbiol 47: 597-626.
67. Kustu S, North AK, Weiss DS (1991) Prokaryotic transcriptional enhancers and enhancer-binding proteins. Trends Biochem Sci 16: 397-402.
68. Gallegos MT, Michan C, Ramos JL (1993) The XylS/AraC family of regulators. Nucleic Acids Res 21: 807-810.
69. Martinez-Hackert E, Stock AM (1997) Structural relationships in the OmpR family of winged-helix transcription factors. J Mol Biol 269: 301-312.
70. Fuqua WC, Winans SC, Greenberg EP (1994) Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. J Bacteriol 176: 269-275.
71. Haydon DJ, Guest JR (1991) A new family of bacterial regulatory proteins. FEMS Microbiol Lett 63: 291-295.
72. Sulavik MC, Gambino LF, Miller PF (1995) The MarR repressor of the multiple antibiotic resistance (*mar*) operon in *Escherichia coli*: prototypic member of a family of bacterial regulatory proteins involved in sensing phenolic compounds. Mol Med 1: 436-446.
73. Summers AO (1992) Untwist and shout: a heavy metal-responsive transcriptional regulator. J Bacteriol 174: 3097-3101.
74. Weickert MJ, Adhya S (1992) A family of bacterial regulators homologous to Gal and Lac repressors. J Biol Chem 267: 15869-15874.
75. Busenlehner LS, Pennella MA, Giedroc DP (2003) The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. FEMS Microbiol Rev 27: 131-143.
76. Friedberg D, Midkiff M, Calvo JM (2001) Global versus local regulatory roles for Lrp-related proteins: *Haemophilus influenzae* as a case study. J Bacteriol 183: 4004-4011.
77. Ramos JL, Duque E, Gallegos MT, Godoy P, Ramos-Gonzalez MI, et al. (2002) Mechanisms of solvent tolerance in gram-negative bacteria. Annu Rev Microbiol 56: 743-768.
78. Brandi A, Pon CL, Gualerzi CO (1994) Interaction of the main cold shock protein CS7.4 (CspA) of *Escherichia coli* with the promoter region of *hns*. Biochimie 76: 1090-1098.
79. Sunnarborg A, Klumpp D, Chung T, LaPorte DC (1990) Regulation of the glyoxylate bypass operon: cloning and characterization of *iclR*. J Bacteriol 172: 2642-2649.
80. Chandler MS (1992) The gene encoding cAMP receptor protein is required for competence development in *Haemophilus influenzae* Rd. Proc Natl Acad Sci U S A 89: 1626-1630.
81. Cases I, de Lorenzo V, Ouzounis CA (2003) Transcription regulation and environmental adaptation in bacteria. Trends Microbiol 11: 248-253.

82. Hodgson DA (2000) Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. Adv Microb Physiol 42: 47-238.
83. Flardh K, Buttner MJ (2009) *Streptomyces* morphogenetics: dissecting differentiation in a filamentous bacterium. Nat Rev Microbiol 7: 36-49.
84. Ma D, Alberti M, Lynch C, Nikaido H, Hearst JE (1996) The local repressor AcrR plays a modulating role in the regulation of *acrAB* genes of *Escherichia coli* by global stress signals. Mol Microbiol 19: 101-112.
85. Le TB, Fiedler HP, den Hengst CD, Ahn SK, Maxwell A, et al. (2009) Coupling of the biosynthesis and export of the DNA gyrase inhibitor simocyclinone in *Streptomyces antibioticus*. Mol Microbiol 72: 1462-1474.
86. Poole K, Gotoh N, Tsujimoto H, Zhao Q, Wada A, et al. (1996) Overexpression of the *mexC-mexD-oprJ* efflux operon in *nfxB*-type multidrug-resistant strains of *Pseudomonas aeruginosa*. Mol Microbiol 21: 713-724.
87. Grkovic S, Brown MH, Roberts NJ, Paulsen IT, Skurray RA (1998) QacR is a repressor protein that regulates expression of the *Staphylococcus aureus* multidrug efflux pump QacA. J Biol Chem 273: 18665-18673.
88. Sanchez P, Alonso A, Martinez JL (2002) Cloning and characterization of SmeT, a repressor of the *Stenotrophomonas maltophilia* multidrug efflux pump SmeDEF. Antimicrob Agents Chemother 46: 3386-3393.
89. Folcher M, Morris RP, Dale G, Salah-Bey-Hocini K, Viollier PH, et al. (2001) A transcriptional regulator of a pristinamycin resistance gene in *Streptomyces coelicolor*. J Biol Chem 276: 1479-1485.
90. Noguchi N, Takada K, Katayama J, Emura A, Sasatsu M (2000) Regulation of transcription of the *mph*(A) gene for macrolide 2'-phosphotransferase I in *Escherichia coli*: characterization of the regulatory gene *mphR*(A). J Bacteriol 182: 5052-5058.
91. Feng Y, Cronan JE (2011) Complex binding of the FabR repressor of bacterial unsaturated fatty acid biosynthesis to its cognate promoters. Mol Microbiol 80: 195-218.
92. Chen CH, Cheng JC, Cho YC, Hsu WH (2005) A gene cluster for the fatty acid catabolism from *Pseudonocardia autotrophica* BCRC12444. Biochem Biophys Res Commun 329: 863-868.
93. Uguru GC, Stephens KE, Stead JA, Towle JE, Baumberg S, et al. (2005) Transcriptional activation of the pathway-specific regulator of the actinorhodin biosynthetic genes in *Streptomyces coelicolor*. Mol Microbiol 58: 131-150.
94. Miyamoto CM, Chatterjee J, Swartzman E, Szittner R, Meighen EA (1996) The role of *lux* autoinducer in regulating luminescence in *Vibrio harveyi*; control of *luxR* expression. Mol Microbiol 19: 767-775.
95. Engohang-Ndong J, Baillat D, Aumercier M, Bellefontaine F, Besra GS, et al. (2004) EthR, a repressor of the TetR/CamR family implicated in ethionamide resistance in mycobacteria, octamerizes cooperatively on its operator. Mol Microbiol 51: 175-188.

96. Croxatto A, Chalker VJ, Lauritz J, Jass J, Hardman A, et al. (2002) VanT, a homologue of *Vibrio harveyi* LuxR, regulates serine, metalloprotease, pigment, and biofilm production in *Vibrio anguillarum*. J Bacteriol 184: 1617-1629.
97. MacEachran DP, Stanton BA, O'Toole GA (2008) Cif is negatively regulated by the TetR family repressor CifR. Infect Immun 76: 3197-3206.
98. Pompeani AJ, Irgon JJ, Berger MF, Bulyk ML, Wingreen NS, et al. (2008) The *Vibrio harveyi* master quorum-sensing regulator, LuxR, a TetR-type protein is both an activator and a repressor: DNA recognition and binding specificity at target promoters. Mol Microbiol 70: 76-88.
99. Ostash I, Ostash B, Luzhetskyy A, Bechthold A, Walker S, et al. (2008) Coordination of export and glycosylation of landomycins in *Streptomyces cyanogenus* S136. FEMS Microbiol Lett 285: 195-202.
100. Rey DA, Nentwich SS, Koch DJ, Ruckert C, Puhler A, et al. (2005) The McbR repressor modulated by the effector substance *S*-adenosylhomocysteine controls directly the transcription of a regulon involved in sulphur metabolism of *Corynebacterium glutamicum* ATCC 13032. Mol Microbiol 56: 871-887.
101. Nakada Y, Jiang Y, Nishijyo T, Itoh Y, Lu CD (2001) Molecular characterization and regulation of the *aguBA* operon, responsible for agmatine utilization in *Pseudomonas aeruginosa* PAO1. J Bacteriol 183: 6517-6524.
102. Mermod M, Magnani D, Solioz M, Stoyanov JV (2012) The copper-inducible ComR (YcfQ) repressor regulates expression of ComC (YcfR), which affects copper permeability of the outer membrane of *Escherichia coli*. Biometals 25: 33-43.
103. Kloosterman TG, van der Kooi-Pol MM, Bijlsma JJ, Kuipers OP (2007) The novel transcriptional regulator SczA mediates protection against Zn^{2+} stress by activation of the Zn^{2+}-resistance gene czcD in Streptococcus pneumoniae. Mol Microbiol 65: 1049-1063.
104. Su CC, Rutherford DJ, Yu EW (2007) Characterization of the multidrug efflux regulator AcrR from *Escherichia coli*. Biochem Biophys Res Commun 361: 85-90.
105. Itou H, Watanabe N, Yao M, Shirakihara Y, Tanaka I (2010) Crystal structures of the multidrug binding repressor *Corynebacterium glutamicum* CgmR in complex with inducers and with an operator. J Mol Biol 403: 174-184.
106. Grkovic S, Hardie KM, Brown MH, Skurray RA (2003) Interactions of the QacR multidrug-binding protein with structurally diverse ligands: implications for the evolution of the binding pocket. Biochemistry 42: 15226-15236.
107. Le TB, Stevenson CE, Fiedler HP, Maxwell A, Lawson DM, et al. (2011) Structures of the TetR-like simocyclinone efflux pump repressor, SimR,

and the mechanism of ligand-mediated derepression. J Mol Biol 408: 40-56.
108. Le TB, Schumacher MA, Lawson DM, Brennan RG, Buttner MJ (2011) The crystal structure of the TetR family transcriptional repressor SimR bound to DNA and the role of a flexible N-terminal extension in minor groove binding. Nucleic Acids Res 39: 9433-9447.
109. Grkovic S, Brown MH, Schumacher MA, Brennan RG, Skurray RA (2001) The staphylococcal QacR multidrug regulator binds a correctly spaced operator as a pair of dimers. J Bacteriol 183: 7102-7109.
110. Schumacher MA, Miller MC, Grkovic S, Brown MH, Skurray RA, et al. (2002) Structural basis for cooperative DNA binding by two dimers of the multidrug-binding protein QacR. EMBO J 21: 1210-1218.
111. Rouch DA, Cram DS, DiBerardino D, Littlejohn TG, Skurray RA (1990) Efflux-mediated antiseptic resistance gene *qacA* from *Staphylococcus aureus*: common ancestry with tetracycline- and sugar-transport proteins. Mol Microbiol 4: 2051-2062.
112. Dover LG, Corsino PE, Daniels IR, Cocklin SL, Tatituri V, et al. (2004) Crystal structure of the TetR/CamR family repressor *Mycobacterium tuberculosis* EthR implicated in ethionamide resistance. J Mol Biol 340: 1095-1105.
113. Frenois F, Engohang-Ndong J, Locht C, Baulard AR, Villeret V (2004) Structure of EthR in a ligand bound conformation reveals therapeutic perspectives against tuberculosis. Mol Cell 16: 301-307.
114. Krug A, Wendisch VF, Bott M (2005) Identification of AcnR, a TetR-type repressor of the aconitase gene *acn* in *Corynebacterium glutamicum*. J Biol Chem 280: 585-595.
115. Lin J, Akiba M, Sahin O, Zhang Q (2005) CmeR functions as a transcriptional repressor for the multidrug efflux pump CmeABC in *Campylobacter jejuni*. Antimicrob Agents Chemother 49: 1067-1075.
116. Lin J, Michel LO, Zhang Q (2002) CmeABC functions as a multidrug efflux system in *Campylobacter jejuni*. Antimicrob Agents Chemother 46: 2124-2131.
117. Lei HT, Shen Z, Surana P, Routh MD, Su CC, et al. (2011) Crystal structures of CmeR-bile acid complexes from *Campylobacter jejuni*. Protein Sci 20: 712-723.
118. Lin J, Sahin O, Michel LO, Zhang Q (2003) Critical role of multidrug efflux pump CmeABC in bile resistance and *in vivo* colonization of *Campylobacter jejuni*. Infect Immun 71: 4250-4259.
119. Christen S, Srinivas A, Bahler P, Zeller A, Pridmore D, et al. (2006) Regulation of the Dha operon of *Lactococcus lactis*: a deviation from the rule followed by the Tetr family of transcription regulators. J Biol Chem 281: 23129-23137.

120. Park H, Ro YT, Kim YM (2011) MdoR is a novel positive transcriptional regulator for the oxidation of methanol in *Mycobacterium* sp. strain JC1. J Bacteriol 193: 6288-6294.
121. Chattoraj P, Mohapatra SS, Rao JL, Biswas I (2011) Regulation of transcription by SMU.1349, a TetR family regulator, in *Streptococcus mutans*. J Bacteriol 193: 6605-6613.
122. Jakoby M, Nolden L, Meier-Wagner J, Kramer R, Burkovski A (2000) AmtR, a global repressor in the nitrogen regulation system of *Corynebacterium glutamicum*. Mol Microbiol 37: 964-977.
123. Caballero JL, Malpartida F, Hopwood DA (1991) Transcriptional organization and regulation of an antibiotic export complex in the producing *Streptomyces* culture. Mol Gen Genet 228: 372-380.
124. Nothaft H, Rigali S, Boomsma B, Swiatek M, McDowall KJ, et al. (2010) The permease gene *nagE2* is the key to *N*-acetylglucosamine sensing and utilization in *Streptomyces coelicolor* and is subject to multi-level control. Mol Microbiol 75: 1133-1144.
125. Rkenes TP, Lamark T, Strom AR (1996) DNA-binding properties of the BetI repressor protein of *Escherichia coli*: the inducer choline stimulates BetI-DNA complex formation. J Bacteriol 178: 1663-1670.
126. Matsuoka H, Hirooka K, Fujita Y (2007) Organization and function of the YsiA regulon of *Bacillus subtilis* involved in fatty acid degradation. J Biol Chem 282: 5180-5194.
127. Zhang H, White-Phillip JA, Melancon CE, 3rd, Kwon HJ, Yu WL, et al. (2007) Elucidation of the kijanimicin gene cluster: insights into the biosynthesis of spirotetronate antibiotics and nitrosugars. J Am Chem Soc 129: 14670-14683.
128. Tahlan K, Ahn SK, Sing A, Bodnaruk TD, Willems AR, et al. (2007) Initiation of actinorhodin export in *Streptomyces coelicolor*. Mol Microbiol 63: 951-961.
129. Peng WT, Nester EW (2001) Characterization of a putative RND-type efflux system in *Agrobacterium tumefaciens*. Gene 270: 245-252.
130. Kieboom J, de Bont J (2001) Identification and molecular characterization of an efflux system involved in *Pseudomonas putida* S12 multidrug resistance. Microbiology 147: 43-51.
131. Martin FA, Posadas DM, Carrica MC, Cravero SL, O'Callaghan D, et al. (2009) Interplay between two RND systems mediating antimicrobial resistance in *Brucella suis*. J Bacteriol 191: 2530-2540.
132. Chan YY, Tan TM, Ong YM, Chua KL (2004) BpeAB-OprB, a multidrug efflux pump in *Burkholderia pseudomallei*. Antimicrob Agents Chemother 48: 1128-1135.
133. Cerda-Maira FA, Ringelberg CS, Taylor RK (2008) The bile response repressor BreR regulates expression of the *Vibrio cholerae breAB* efflux system operon. J Bacteriol 190: 7441-7452.

134. Quillin SJ, Schwartz KT, Leber JH (2011) The novel *Listeria monocytogenes* bile sensor BrtA controls expression of the cholic acid efflux pump MdrT. Mol Microbiol 81: 129-142.
135. Lee LF, Huang YJ, Chen CW (2003) Repressed multidrug resistance genes in *Streptomyces lividans*. Arch Microbiol 180: 176-184.
136. Lee LF, Chen YJ, Kirby R, Chen C, Chen CW (2007) A multidrug efflux system is involved in colony growth in *Streptomyces lividans*. Microbiology 153: 924-934.
137. Hirakawa H, Takumi-Kobayashi A, Theisen U, Hirata T, Nishino K, et al. (2008) AcrS/EnvR represses expression of the *acrAB* multidrug efflux genes in *Escherichia coli*. J Bacteriol 190: 6276-6279.
138. Rodriguez-Garcia A, Santamarta I, Perez-Redondo R, Martin JF, Liras P (2006) Characterization of a two-gene operon *epeRA* involved in multidrug resistance in *Streptomyces clavuligerus*. Res Microbiol 157: 559-568.
139. Wenzel M, Lang K, Gunther T, Bhandari A, Weiss A, et al. (2012) Characterization of the flavonoid-responsive regulator FrrA and its binding sites. J Bacteriol 194: 2363-2370.
140. Lechardeur D, Cesselin B, Liebl U, Vos MH, Fernandez A, et al. (2012) Discovery of intracellular heme-binding protein HrtR, which controls heme efflux by the conserved HrtB-HrtA transporter in *Lactococcus lactis*. J Biol Chem 287: 4752-4758.
141. Buroni S, Manina G, Guglierame P, Pasca MR, Riccardi G, et al. (2006) LfrR is a repressor that regulates expression of the efflux pump LfrA in Mycobacterium smegmatis. Antimicrob Agents Chemother 50: 4044-4052.
142. Hirooka K, Kunikane S, Matsuoka H, Yoshida K, Kumamoto K, et al. (2007) Dual regulation of the *Bacillus subtilis* regulon comprising the *lmrAB* and *yxaGH* operons and *yxaF* gene by two transcriptional repressors, LmrA and YxaF, in response to flavonoids. J Bacteriol 189: 5170-5182.
143. Fukumori F, Hirayama H, Takami H, Inoue A, Horikoshi K (1998) Isolation and transposon mutagenesis of a *Pseudomonas putida* KT2442 toluene-resistant variant: involvement of an efflux system in solvent resistance. Extremophiles 2: 395-400.
144. Chuanchuen R, Narasaki CT, Schweizer HP (2002) The MexJK efflux pump of *Pseudomonas aeruginosa* requires OprM for antibiotic efflux but not for efflux of triclosan. J Bacteriol 184: 5036-5044.
145. Matsuo Y, Eda S, Gotoh N, Yoshihara E, Nakae T (2004) MexZ-mediated regulation of *mexXY* multidrug efflux pump expression in Pseudomonas aeruginosa by binding on the *mexZ-mexX* intergenic DNA. FEMS Microbiol Lett 238: 23-28.
146. Hagman KE, Shafer WM (1995) Transcriptional control of the *mtr* efflux system of *Neisseria gonorrhoeae*. J Bacteriol 177: 4162-4165.
147. Sobel ML, Hocquet D, Cao L, Plesiat P, Poole K (2005) Mutations in PA3574 (*nalD*) lead to increased MexAB-OprM expression and multidrug

resistance in laboratory and clinical isolates of *Pseudomonas aeruginosa*. Antimicrob Agents Chemother 49: 1782-1786.
148. Gristwood T, Fineran PC, Everson L, Salmond GP (2008) PigZ, a TetR/AcrR family repressor, modulates secondary metabolism via the expression of a putative four-component resistance-nodulation-cell-division efflux pump, ZrpADBC, in *Serratia* sp. ATCC 39006. Mol Microbiol 69: 418-435.
149. Cho YH, Kim EJ, Chung HJ, Choi JH, Chater KF, et al. (2003) The *pqrAB* operon is responsible for paraquat resistance in *Streptomyces coelicolor*. J Bacteriol 185: 6756-6763.
150. Gonzalez-Pasayo R, Martinez-Romero E (2000) Multiresistance genes of *Rhizobium etli* CFN42. Mol Plant Microbe Interact 13: 572-577.
151. Sun X, Zahir Z, Lynch KH, Dennis JJ (2011) An antirepressor, SrpR, is involved in transcriptional regulation of the SrpABC solvent tolerance efflux pump of *Pseudomonas putida* S12. J Bacteriol 193: 2717-2725.
152. Teran W, Felipe A, Segura A, Rojas A, Ramos JL, et al. (2003) Antibiotic-dependent induction of *Pseudomonas putida* DOT-T1E TtgABC efflux pump is mediated by the drug binding repressor TtgR. Antimicrob Agents Chemother 47: 3067-3072.
153. Alatoom AA, Aburto R, Hamood AN, Colmer-Hamood JA (2007) VceR negatively regulates the *vceCAB* MDR efflux operon and positively regulates its own synthesis in *Vibrio cholerae* 569B. Can J Microbiol 53: 888-900.
154. Bentley SD, Chater KF, Cerdeno-Tarraga AM, Challis GL, Thomson NR, et al. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417: 141-147.
155. Marchler-Bauer A, Bryant SH (2004) CD-Search: protein domain annotations on the fly. Nucleic Acids Res 32: W327-331.
156. Webb EC (1992) Enzyme nomenclature 1992. San Diego, California: Academic Press.
157. Dym O, Eisenberg D (2001) Sequence-structure analysis of FAD-containing proteins. Protein Sci 10: 1712-1728.
158. Kleiger G, Eisenberg D (2002) GXXXG and GXXXA motifs stabilize FAD and NAD(P)-binding Rossmann folds through C(alpha)-H... O hydrogen bonds and van der waals interactions. J Mol Biol 323: 69-76.
159. Rossmann MG, Argos P (1976) Exploring structural homology of proteins. J Mol Biol 105: 75-95.
160. Spector MP, DiRusso CC, Pallen MJ, Garcia del Portillo F, Dougan G, et al. (1999) The medium-/long-chain fatty acyl-CoA dehydrogenase (*fadF*) gene of *Salmonella typhimurium* is a phase 1 starvation-stress response (SSR) locus. Microbiology 145 (Pt 1): 15-31.
161. Heinzelmann E, Berger S, Muller C, Hartner T, Poralla K, et al. (2005) An acyl-CoA dehydrogenase is involved in the formation of the Delta *cis*3

double bond in the acyl residue of the lipopeptide antibiotic friulimicin in *Actinoplanes friuliensis*. Microbiology 151: 1963-1974.
162. Kunau WH, Dommes V, Schulz H (1995) beta-oxidation of fatty acids in mitochondria, peroxisomes, and bacteria: a century of continued progress. Prog Lipid Res 34: 267-342.
163. Kim JJ, Battaile KP (2002) Burning fat: the structural basis of fatty acid beta-oxidation. Curr Opin Struct Biol 12: 721-728.
164. Sanli G, Dudley JI, Blaber M (2003) Structural biology of the aldo-keto reductase family of enzymes: catalysis and cofactor binding. Cell Biochem Biophys 38: 79-101.
165. Nelson DR, Kamataki T, Waxman DJ, Guengerich FP, Estabrook RW, et al. (1993) The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. DNA Cell Biol 12: 1-51.
166. Loferer H, Wunderlich M, Hennecke H, Glockshuber R (1995) A bacterial thioredoxin-like protein that is exposed to the periplasm has redox properties comparable with those of cytoplasmic thioredoxins. J Biol Chem 270: 26178-26183.
167. Johansson E, Steffens JJ, Lindqvist Y, Schneider G (2000) Crystal structure of saccharopine reductase from *Magnaporthe grisea*, an enzyme of the alpha-aminoadipate pathway of lysine biosynthesis. Structure 8: 1037-1047.
168. Mande SS, Sarfaty S, Allen MD, Perham RN, Hol WG (1996) Protein-protein interactions in the pyruvate dehydrogenase multienzyme complex: dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase. Structure 4: 277-286.
169. Romao MJ, Archer M, Moura I, Moura JJ, LeGall J, et al. (1995) Crystal structure of the xanthine oxidase-related aldehyde oxido-reductase from *D. gigas*. Science 270: 1170-1176.
170. van Der Ploeg JR, Iwanicka-Nowicka R, Bykowski T, Hryniewicz MM, Leisinger T (1999) The *Escherichia coli ssuEADCB* gene cluster is required for the utilization of sulfur from aliphatic sulfonates and is regulated by the transcriptional activator Cbl. J Biol Chem 274: 29358-29365.
171. Canaan S, Sulzenbacher G, Roig-Zamboni V, Scappuccini-Calvo L, Frassinetti F, et al. (2005) Crystal structure of the conserved hypothetical protein Rv1155 from *Mycobacterium tuberculosis*. FEBS Lett 579: 215-221.
172. Selengut JD, Haft DH (2010) Unexpected abundance of coenzyme F(420)-dependent enzymes in *Mycobacterium tuberculosis* and other actinobacteria. J Bacteriol 192: 5788-5798.
173. Kirch HH, Bartels D, Wei Y, Schnable PS, Wood AJ (2004) The ALDH gene superfamily of *Arabidopsis*. Trends Plant Sci 9: 371-377.

174. Titus GP, Mueller HA, Burgner J, Rodriguez De Cordoba S, Penalva MA, et al. (2000) Crystal structure of human homogentisate dioxygenase. Nat Struct Biol 7: 542-546.
175. Reed DW, Hartzell PL (1999) The Archaeoglobus fulgidus D-lactate dehydrogenase is a Zn(2+) flavoprotein. J Bacteriol 181: 7580-7587.
176. Que L, Jr., Widom J, Crawford RL (1981) 3,4-Dihydroxyphenylacetate 2,3-dioxygenase. A manganese(II) dioxygenase from *Bacillus brevis*. J Biol Chem 256: 10941-10944.
177. Holland PJ, Hollis T (2010) Structural and mutational analysis of *Escherichia coli* AlkB provides insight into substrate specificity and DNA damage searching. PLoS One 5: e8680.
178. Tsai MH, Saier MH, Jr. (1995) Phylogenetic characterization of the ubiquitous electron transfer flavoprotein families ETF-alpha and ETF-beta. Res Microbiol 146: 397-404.
179. Breton C, Snajdrova L, Jeanneau C, Koca J, Imberty A (2006) Structures and mechanisms of glycosyltransferases. Glycobiology 16: 29R-37R.
180. Lee PT, Hsu AY, Ha HT, Clarke CF (1997) A C-methyltransferase involved in both ubiquinone and menaquinone biosynthesis: isolation and identification of the *Escherichia coli ubiE* gene. J Bacteriol 179: 1748-1754.
181. Hanks SK, Quinn AM (1991) Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. Methods Enzymol 200: 38-62.
182. Gruez A, Roig-Zamboni V, Valencia C, Campanacci V, Cambillau C (2003) The crystal structure of the *Escherichia coli* YfdW gene product reveals a new fold of two interlaced rings identifying a wide family of CoA transferases. J Biol Chem 278: 34582-34586.
183. Parisi G, Echave J (2004) The structurally constrained protein evolution model accounts for sequence patterns of the LbetaH superfamily. BMC Evol Biol 4: 41.
184. Beaman TW, Sugantino M, Roderick SL (1998) Structure of the hexapeptide xenobiotic acetyltransferase from *Pseudomonas aeruginosa*. Biochemistry 37: 6689-6696.
185. Isono K, Isono S (1980) Ribosomal protein modification in *Escherichia coli*. II. Studies of a mutant lacking the N-terminal acetylation of protein S18. Mol Gen Genet 177: 645-651.
186. Humm A, Fritsche E, Steinbacher S, Huber R (1997) Crystal structure and mechanism of human L-arginine:glycine amidinotransferase: a mitochondrial enzyme involved in creatine biosynthesis. EMBO J 16: 3373-3385.
187. Davies DR, Mahnke Braam L, Reznikoff WS, Rayment I (1999) The three-dimensional structure of a Tn*5* transposase-related protein determined to 2.9-A resolution. J Biol Chem 274: 11904-11913.

188. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, et al. (1994) Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. Science 266: 1981-1986.
189. Ollis DL, Cheah E, Cygler M, Dijkstra B, Frolow F, et al. (1992) The alpha/beta hydrolase fold. Protein Eng 5: 197-211.
190. Pallen MJ, Lam AC, Loman N (2001) Tricorn-like proteases in bacteria. Trends Microbiol 9: 518-521.
191. Lovgren A, Zhang M, Engstrom A, Dalhammar G, Landen R (1990) Molecular characterization of immune inhibitor A, a secreted virulence protease from *Bacillus thuringiensis*. Mol Microbiol 4: 2137-2146.
192. Ullah JH, Walsh TR, Taylor IA, Emery DC, Verma CS, et al. (1998) The crystal structure of the L1 metallo-beta-lactamase from *Stenotrophomonas maltophilia* at 1.7 A resolution. J Mol Biol 284: 125-136.
193. Chen L, Roberts MF (2000) Overexpression, purification, and analysis of complementation behavior of *E. coli* SuhB protein: comparison with bacterial and archaeal inositol monophosphatases. Biochemistry 39: 4145-4153.
194. Stukey J, Carman GM (1997) Identification of a novel phosphatase sequence motif. Protein Sci 6: 469-472.
195. Korolev S, Hsieh J, Gauss GH, Lohman TM, Waksman G (1997) Major domain swiveling revealed by the crystal structures of complexes of *E. coli* Rep helicase bound to single-stranded DNA and ADP. Cell 90: 635-647.
196. Hisano T, Hata Y, Fujii T, Liu JQ, Kurihara T, et al. (1996) Crystal structure of L-2-haloacid dehalogenase from *Pseudomonas* sp. YL. An alpha/beta hydrolase structure that is different from the alpha/beta hydrolase fold. J Biol Chem 271: 20322-20330.
197. Old LA, Lowes S, Russell RR (2006) Genomic variation in *Streptococcus mutans*: deletions affecting the multiple pathways of beta-glucoside metabolism. Oral Microbiol Immunol 21: 21-27.
198. Ounissi H, Courvalin P (1985) Nucleotide sequence of the gene *ereA* encoding the erythromycin esterase in *Escherichia coli*. Gene 35: 271-278.
199. Arthur M, Autissier D, Courvalin P (1986) Analysis of the nucleotide sequence of the *ereB* gene encoding the erythromycin esterase type II. Nucleic Acids Res 14: 4987-4999.
200. Varghese JN, Hrmova M, Fincher GB (1999) Three-dimensional structure of a barley beta-D-glucan exohydrolase, a family 3 glycosyl hydrolase. Structure 7: 179-190.
201. Aravind L, Koonin EV (1998) Phosphoesterase domains associated with DNA polymerases of diverse origins. Nucleic Acids Res 26: 3746-3752.
202. Robertus J (1991) The structure and action of ricin, a cytotoxic N-glycosidase. Semin Cell Biol 2: 23-30.
203. Chang TH, Abelson J (1990) Identification of a putative amidase gene in yeast *Saccharomyces cerevisiae*. Nucleic Acids Res 18: 7180.

204. Shirai H, Blundell TL, Mizuguchi K (2001) A novel superfamily of enzymes that catalyze the modification of guanidino groups. Trends Biochem Sci 26: 465-468.
205. Hemila H, Koivula TT, Palva I (1994) Hormone-sensitive lipase is closely related to several bacterial proteins, and distantly related to acetylcholinesterase and lipoprotein lipase: identification of a superfamily of esterases and lipases. Biochim Biophys Acta 1210: 249-253.
206. Barak I, Behari J, Olmedo G, Guzman P, Brown DP, et al. (1996) Structure and function of the *Bacillus* SpoIIE protein and its localization to sites of sporulation septum assembly. Mol Microbiol 19: 1047-1060.
207. Schneider A, Marahiel MA (1998) Genetic evidence for a role of thioesterase domains, integrated in or associated with peptide synthetases, in non-ribosomal peptide biosynthesis in *Bacillus subtilis*. Arch Microbiol 169: 404-410.
208. Derewenda ZS, Derewenda U (1998) The structure and function of platelet-activating factor acetylhydrolases. Cell Mol Life Sci 54: 446-455.
209. Aravind L, Koonin EV (1998) The HD domain defines a new superfamily of metal-dependent phosphohydrolases. Trends Biochem Sci 23: 469-472.
210. Thornalley PJ (2003) Glyoxalase I--structure, function and a critical role in the enzymatic defence against glycation. Biochem Soc Trans 31: 1343-1348.
211. Takano E, Chakraburtty R, Nihira T, Yamada Y, Bibb MJ (2001) A complex role for the gamma-butyrolactone SCB1 in regulating antibiotic production in *Streptomyces coelicolor* A3(2). Mol Microbiol 41: 1015-1028.
212. Yeats C, Bentley S, Bateman A (2003) New knowledge from old: *in silico* discovery of novel protein domains in *Streptomyces coelicolor*. BMC Microbiol 3: 3.
213. Horinouchi S, Kumada Y, Beppu T (1984) Unstable genetic determinant of A-factor biosynthesis in streptomycin-producing organisms: cloning and characterization. J Bacteriol 158: 481-487.
214. Gerischer U, Durre P (1990) Cloning, sequencing, and molecular analysis of the acetoacetate decarboxylase gene region from *Clostridium acetobutylicum*. J Bacteriol 172: 6907-6918.
215. Iyer LM, Koonin EV, Aravind L (2001) Adaptations of the helix-grip fold for ligand binding and catalysis in the START domain superfamily. Proteins 43: 134-144.
216. Mukhtar TA, Koteva KP, Hughes DW, Wright GD (2001) Vgb from *Staphylococcus aureus* inactivates streptogramin B antibiotics by an elimination mechanism not hydrolysis. Biochemistry 40: 8877-8886.
217. Roper DI, Subramanya HS, Shingler V, Wigley DB (1994) Preliminary crystallographic analysis of 4-oxalocrotonate tautomerase reveals the oligomeric structure of the enzyme. J Mol Biol 243: 799-801.
218. Ha NC, Choi G, Choi KY, Oh BH (2001) Structure and enzymology of Delta5-3-ketosteroid isomerase. Curr Opin Struct Biol 11: 674-678.

219. Marini P, Li SJ, Gardiol D, Cronan JE, Jr., de Mendoza D (1995) The genes encoding the biotin carboxyl carrier protein and biotin carboxylase subunits of *Bacillus subtilis* acetyl coenzyme A carboxylase, the first enzyme of fatty acid synthesis. J Bacteriol 177: 7003-7006.
220. Conti E, Franks NP, Brick P (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. Structure 4: 287-298.
221. Saidijam M, Benedetti G, Ren Q, Xu Z, Hoyle CJ, et al. (2006) Microbial drug efflux proteins of the major facilitator superfamily. Curr Drug Targets 7: 793-811.
222. Tomii K, Kanehisa M (1998) A comparative analysis of ABC transporters in complete microbial genomes. Genome Res 8: 1048-1059.
223. Tseng TT, Gratwick KS, Kollman J, Park D, Nies DH, et al. (1999) The RND permease superfamily: an ancient, ubiquitous and diverse family that includes human disease and development proteins. J Mol Microbiol Biotechnol 1: 107-125.
224. Tate CG, Muiry JA, Henderson PJ (1992) Mapping, cloning, expression, and sequencing of the *rhaT* gene, which encodes a novel L-rhamnose-H^+ transport protein in *Salmonella typhimurium* and *Escherichia coli*. J Biol Chem 267: 6923-6932.
225. Yerushalmi H, Lebendiker M, Schuldiner S (1995) EmrE, an *Escherichia coli* 12-kDa multidrug transporter, exchanges toxic cations and H^+ and is soluble in organic solvents. J Biol Chem 270: 6856-6863.
226. Ravcheev DA, Gel'fand MS, Mironov AA, Rakhmaninova AB (2002) [Purine regulon of gamma-proteobacteria: a detailed description]. Genetika 38: 1203-1214.
227. Weinitschke S, Denger K, Cook AM, Smits TH (2007) The DUF81 protein TauE in *Cupriavidus necator* H16, a sulfite exporter in the metabolism of C2 sulfonates. Microbiology 153: 3055-3060.
228. Craney A, Hohenauer T, Xu Y, Navani NK, Li Y, et al. (2007) A synthetic *luxCDABE* gene cluster optimized for expression in high-GC bacteria. Nucleic Acids Res 35: e46.
229. Hillerich B, Westpheling J (2008) A new TetR family transcriptional regulator required for morphogenesis in *Streptomyces coelicolor*. J Bacteriol 190: 61-67.
230. Zhang YM, Zhu K, Frank MW, Rock CO (2007) A *Pseudomonas aeruginosa* transcription factor that senses fatty acid structure. Mol Microbiol 66: 622-632.
231. Forster-Fromme K, Jendrossek D (2010) AtuR is a repressor of acyclic terpene utilization (Atu) gene cluster expression and specifically binds to two 13 bp inverted repeat sequences of the *atuA-atuR* intergenic region. FEMS Microbiol Lett 308: 166-174.

232. Chan YY, Chua KL (2005) The *Burkholderia pseudomallei* BpeAB-OprB efflux pump: expression and impact on quorum sensing and virulence. J Bacteriol 187: 4707-4719.
233. Santangelo MP, Blanco FC, Bianco MV, Klepp LI, Zabal O, et al. (2008) Study of the role of Mce3R on the transcription of *mce* genes of *Mycobacterium tuberculosis*. BMC Microbiol 8: 38.
234. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA (2000) Practical *Streptomyces* Genetics. Norwich, England: The John Innes Foundation.
235. Ahn SK, Tahlan K, Yu Z, Nodwell J (2007) Investigation of transcription repression and small-molecule responsiveness by TetR-like transcription factors using a heterologous *Escherichia coli*-based assay. J Bacteriol 189: 6655-6664.
236. Taylor DE, Hiratsuka K, Ray H, Manavathu EK (1987) Characterization and expression of a cloned tetracycline resistance determinant from *Campylobacter jejuni* plasmid pUA466. J Bacteriol 169: 2984-2989.
237. Bjarnason J, Southward CM, Surette MG (2003) Genomic profiling of iron-responsive genes in *Salmonella enterica* serovar typhimurium by high-throughput screening of a random promoter library. J Bacteriol 185: 4973-4982.
238. Chang AC, Cohen SN (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J Bacteriol 134: 1141-1156.
239. Fernandez-Moreno MA, Caballero JL, Hopwood DA, Malpartida F (1991) The *act* cluster contains regulatory and antibiotic export genes, direct targets for translational control by the *bldA* tRNA gene of *Streptomyces*. Cell 66: 769-780.
240. Bartel PL, Zhu CB, Lampel JS, Dosch DC, Connors NC, et al. (1990) Biosynthesis of anthraquinones by interspecies cloning of actinorhodin biosynthesis genes in streptomycetes: clarification of actinorhodin gene functions. J Bacteriol 172: 4816-4826.
241. Bjarnason J, Southward CM, Surette MG (2003) Genomic profiling of iron-responsive genes in *Salmonella enterica* serovar typhimurium by high-throughput screening of a random promoter library. J Bacteriol 185: 4973-4982.
242. Taylor DE, Hiratsuka K, Ray H, Manavathu EK (1987) Characterization and expression of a cloned tetracycline resistance determinant from *Campylobacter jejuni* plasmid pUA466. J Bacteriol 169: 2984-2989.
243. Englesberg E, Wilcox G (1974) Regulation: positive control. Annu Rev Genet 8: 219-242.
244. Schleif R (2010) AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. FEMS Microbiol Rev 34: 779-796.

245. Yu Z, Reichheld SE, Cuthbertson L, Nodwell JR, Davidson AR (2010) Characterization of tetracycline modifying enzymes using a sensitive *in vivo* reporter system. BMC Biochem 11: 34.
246. Banchio C, Gramajo HC (1997) Medium- and long-chain fatty acid uptake and utilization by *Streptomyces coelicolor* A3(2): first characterization of a gram-positive bacterial system. Microbiology 143 (Pt 7): 2439-2447.
247. deHaseth PL, Helmann JD (1995) Open complex formation by *Escherichia coli* RNA polymerase: the mechanism of polymerase-induced strand separation of double helical DNA. Mol Microbiol 16: 817-824.
248. Jensen PR, Hammer K (1998) The sequence of spacers between the consensus sequences modulates the strength of prokaryotic promoters. Appl Environ Microbiol 64: 82-87.
249. Toulme F, Mosrin-Huaman C, Artsimovitch I, Rahmouni AR (2005) Transcriptional pausing *in vivo*: a nascent RNA hairpin restricts lateral movements of RNA polymerase in both forward and reverse directions. J Mol Biol 351: 39-51.
250. Komissarova N, Kashlev M (1997) Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. Proc Natl Acad Sci U S A 94: 1755-1760.
251. Borukhov S, Sagitov V, Goldfarb A (1993) Transcript cleavage factors from E. coli. Cell 72: 459-466.
252. Orlova M, Newlands J, Das A, Goldfarb A, Borukhov S (1995) Intrinsic transcript cleavage activity of RNA polymerase. Proc Natl Acad Sci U S A 92: 4596-4600.
253. Komissarova N, Kashlev M (1997) RNA polymerase switches between inactivated and activated states By translocating back and forth along the DNA and the RNA. J Biol Chem 272: 15329-15338.
254. Toulme F, Mosrin-Huaman C, Sparkowski J, Das A, Leng M, et al. (2000) GreA and GreB proteins revive backtracked RNA polymerase in vivo by promoting transcript trimming. EMBO J 19: 6853-6859.
255. Epshtein V, Toulme F, Rahmouni AR, Borukhov S, Nudler E (2003) Transcription through the roadblocks: the role of RNA polymerase cooperation. EMBO J 22: 4719-4727.