UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE BIOQUÍMICA E IMUNOLOGIA

LUMMY MARIA OLIVEIRA MONTEIRO

Deciphering the architecture/function relationship in complex bacterial promoters through Synthetic Biology approaches.

RIBEIRÃO PRETO
2020

LUMMY MARIA OLIVEIRA MONTEIRO

Deciphering the architecture/function relationship in complex bacterial promoters through Synthetic Biology approaches.

Tese de Doutorado apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do Título de Doutor em Ciências – Área de Concentração: Bioquímica.

Orientador: Prof. Dr. Rafael Silva Rocha.

RIBEIRÃO PRETO

2020

Monteiro, Lummy Maria Oliveira

Desvendando as relações arquitetura/função de promotores bacterianos complexos utilizando abordagens de biologia sintética. Ribeirão Preto, 2020.

99 p.: il. ; 30 cm

Tese de Doutorado, apresentada à Faculdade de Medicina de Ribeirão Preto/USP. Área de concentração: Bioquímica.

Orientador: Silva-Rocha, Rafael.

1. Regulação gênica; 2. Promotores Complexos; 3. Fatores de Transcrição; 4. Engenharia de Proteínas; 5. Design de circuitos; 6. Machine Learning

Monteiro, L.M.O. **Deciphering the architecture/function relationship in complex bacterial promoters through Synthetic Biology approaches.** Tese apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do Título de Doutor em Ciências – Área de Concentração: Bioquímica. Ribeirão Preto, 2020.

Aprovado em:___/___/___

Banca Examinadora

Prof. Dr. _____Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____Instituição: _____

Julgamento: _____ Assinatura: _____

AGRADECIMENTOS

Na realização de um trabalho científico como este, se faz necessário a colaboração de pessoas e instituições para o bom desenvolvimento do mesmo, desde o projeto até o trabalho final. Assim como a participação ativa da família e amigos, que contribuem de forma essencial para seguir em frente e atingir com êxito aos objetivos estabelecidos. Dessa maneira, torna-se fundamental, agradecer a elas. Em especial, quero expressar meus sinceros agradecimentos:

Ao povo brasileiro, que com muita dificuldade financia a ciência em nosso país;

À minha família, Cassiana, Maria Helena, Ocacílio, Alessandro, Alexandre Filho (Fi), Felipinho e Alexandre, que mesmo de longe estão sempre ao meu lado me incentivando, acreditando no meu potencial, financiando meus sonhos e vibrando a cada conquista. Sem vocês, nada teria sido possível! O suporte de vocês foi essencial para que eu pudesse ir mais longe! Amo vocês;

Ao Jonathan, por todo amor e suporte, por apoiar todas as minhas decisões mesmo quando elas me levam para o outro lado do oceano. Obrigada por fazer toda distância parecer insignificante. “You are my lobster!”. Por ter trazido com você uma segunda família para mim. Fran, Silvino e Bruna, muito obrigada por todo apoio, carinho e incentivo. Vocês são a minha família em SP;

Ao meu orientador, Prof. Dr. Rafael Silva-Rocha, a quem eu sou muito grata por tudo o que aconteceu durante o Doutorado. Obrigada por ter acreditado em mim desde a nossa primeira conversa na FMRP, mesmo quando eu ainda não sabia quase nada de Biologia Sintética. Por todas as oportunidades e portas que você abriu para mim; se hoje defendo esse trabalho, foi porque tive seu apoio em todas as etapas;

Aos amigos do SSBLab: Léo, Greicy, Kauan, Luisa, Ananda, Juliana, Murilo, Bianca, Ninna e Felipe, agradeço por terem tornado melhor todos os dias de trabalho. Obrigada pelos ensinamentos e disponibilidade, por todas as nossas discussões produtivas, cafés e risadas! Aprendi muito com vocês! Um agradecimento bem especial à Claudinha, por todo carinho e amizade! Vou sentir saudades dos nossos cafezinhos pra começarmos bem o nosso dia de trabalho!

À Prof. Maria Eugenia Guazzaroni e aos amigos do Laboratório de Metagenômica Funcional, Luana, Guilherme, Gabriel, Tiago e Ítalo, por toda amizade e ensinamentos. Muito obrigada por toda orientação, sugestões de experimentos, colaborações em trabalhos e discussões científicas ao longo desses 4 anos.

À Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto e ao Programa de pós-graduação em Bioquímica pelo ensino público de qualidade. À Ivone pela disponibilidade e atenção com todos nós!

To my supervisor in Germany, Dr. Ulisses Nunes da Rocha, for having welcomed me into his group for one year, for guiding me in the best way possible. For having believed in our project, many times even more than myself; for helping me in every detail. I grew up a lot in your group and I am extremely grateful for all the teachings I received. I learned to believe in my potential and to see science with more “Why? How? and What?”.

Many thanks to Ulisses, João, Felipe, Rodolfo, Natascha, Junia, Havva, N’afiu, René, and Nicole for all assistance during 2019 at UFZ/UMB. I learned so much and I had a great time of my professional life with you. To all my amazing friends in Germany: Junia, Natascha, Felipe, Rodolfo, Mayara, Ruan, Juliana, Gabriela, Joyce, Laszlon, Bruna, Adriana, Canan, Peter, Amir, Krupa, Xin, Francesco, Fernando, Bijing, Swamini, Judith and Katharina. If my days in Germany were the best days of my life, it was because I had you guys participating of these incredible months;

To Helmholtz Centre for Environmental Research (UFZ), which received me as a guest scientist for one year and which during that time opened doors from all researchers and laboratories to me. It also showed me how productive it is to work at a research institute and that is possible to have a non-competitive work environment, with generosity and empathy.

Às minhas amigas Cataguasenses e aos amigos da BQI/UFV que longe ou perto estão sempre me apoiando e torcendo pelo meu sucesso.

À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (Processos: 2016/19179-9 e 2018/21133-2) pelo apoio financeiro no Brasil e no exterior. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo apoio financeiro, imprescindível para a realização deste trabalho.

A todos que me inspiraram, incentivaram e auxiliaram, o meu muito obrigada por fazerem parte dos quatro anos mais importantes e incríveis da minha vida (até agora).

E que venham os novos desafios!

“Around here, however, we don't look backwards for very long. We keep moving forward, opening up new doors and doing new things, because we're curious...and curiosity keeps leading us down new paths.”

-Walt Disney

GENERAL INDEX

RESUMO

Monteiro, L.M.O. **Desvendando as relações arquitetura/função de promotores bacterianos complexos utilizando abordagens de biologia sintética.** Tese de Doutorado. Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto, 2020.

A regulação gênica tem sido estudada extensivamente, no entanto, a complexidade dos mecanismos regulatórios ainda permanece desconhecida. Entender os mecanismos da regulação gênica é importante não apenas para desvendar a complexidade de um organismo, mas para postular novas regras, caracterizar novas partes biológicas e então permitir novos designs de circuitos biológicos, por exemplo. Uma possível estratégia para desvendar os mecanismos de ação e complexidade dos promotores bacterianos seria combinar o conhecimento da regulação gênica com o uso de abordagens da biologia sintética e da bioinformática, que, por sua vez, permitem projetar e construir novas funções em sistemas biológicos. O progresso na biologia sintética é frequentemente possibilitado por poderosas ferramentas de bioinformática que permitem a integração das fases de design, construção e teste do ciclo de engenharia biológica. Consequentemente, o desenvolvimento de novas ferramentas de bioinformática é útil e importante para os cientistas que trabalham para estender ou modificar o comportamento dos organismos e projetá-los para realizar novas tarefas. Nesse contexto, a presente tese descreveu (i) a existência de propriedades emergentes em promotores sintéticos complexos em *Escherichia coli*, que poderiam ser extrapoladas para sistemas regulatórios de ocorrência natural e impactariam significativamente a engenharia de circuitos biológicos sintéticos em bactérias. Em resumo, esses dados demonstram como pequenas mudanças na arquitetura dos promotores bacterianos podem resultar em mudanças drásticas na lógica regulatória final do sistema, com implicações importantes na compreensão de promotores complexos naturais em bactérias e sua engenharia para novas aplicações; (ii) o mecanismo de reconhecimento do indutor de dois reguladores AraC/XylS de *Pseudomonas putida* (BenR e XylS) para a criação de um novo sistema de expressão responsivo ao ácido acetil salicílico (aspirina). Usando homologia de proteínas e docking molecular com o indutor benzoato e um conjunto de análogos químicos, identificamos o sítio de ligação conservado dessas duas proteínas. Como resultado, uma coleção de fatores de transcrição (TFs) engenheirados foram gerados com respostas aprimoradas a uma molécula bem caracterizada e amplamente inócua com um potencial para induzir a expressão heteróloga de genes bacterianos em animais; (iii) a complexidade dos fatores de transcrição em comunidades microbianas ambientais. Criamos um banco de dados de fatores de transcrição bacteriano (BacTFDB) que foi usado para treinar um modelo de *Machine Learning* para prever novos TFs e suas famílias em amostras metagenômicas e metranscriptômicas (PredicTF). PredicTF fornece a primeira ferramenta para traçar o perfil de TFs em bactérias ainda a serem cultivadas e abre o potencial para avaliar redes regulatórias em comunidades microbianas complexas. PredicTF é um pipeline de código aberto flexível capaz de prever e anotar TFs em genomas e metagenomas. PredicTF está disponível em https://github.com/mdsufz/PredicTF.

Palavras chaves: Regulação gênica, Promotores Complexos, Fatores de Transcrição, Engenharia de Proteínas, Design de circuitos, *Machine Learning*.

ABSTRACT

Monteiro, L.M.O. **Deciphering the architecture/function relationship in complex bacterial promoters through Synthetic Biology approaches.** Doctoral Thesis. Ribeirão Preto Medical School, Ribeirão Preto, 2020.

Gene regulation has been studied extensively, however the complexity of the regulatory mechanisms still remains unknown. Understanding how gene regulation occurs is important not only to better understand the complexity of an organism but to postulate new rules, characterize new biological parts and then allow new design of biological circuits, for example. A possible strategy to unravel the mechanisms of action and complexity of bacterial promoters would be to combine the knowledge of gene regulation with the use of approaches from synthetic biology and bioinformatics, which, in turn, allow to design and build new functions in biological systems. Progress in synthetic biology is often made possible by powerful bioinformatics tools that allow the integration of the design, construction and testing stages of the biological engineering cycle. Consequently, the development of new bioinformatics tools is useful and important for scientists working on the design, development and testing of parts to extend or modify the behavior of organisms and design them to perform new tasks. In this context, the present thesis described (i) the existence of emergent properties in complex synthetic promoters in *Escherichia coli*, which could be extrapolated to naturally occurring regulatory systems and would significantly impact the engineering of synthetic biological circuits in bacteria. Taken together, these data demonstrate how small changes in the architecture of bacterial promoters could result in drastic changes in the final regulatory logic of the system, with important implications for the understanding of natural complex promoters in bacteria and their engineering for novel applications; (ii) the inducer recognition mechanism of two AraC/XylS regulators from *Pseudomonas putida* (BenR and XylS) for creating a novel expression system responsive to acetyl salicylate (i.e. Aspirin). Using protein homology modeling and molecular docking with the cognate inducer benzoate and a suite of chemical analogues, we identified the conserved binding pocket of these two proteins. As a result, a collection of engineered transcription factors (TFs) was generated with enhanced response to a well characterized and largely innocuous molecule with a potential for eliciting heterologous expression of bacterial genes in animal carriers; (iii) the complexity of transcription factors in environmental communities. We created one bacterial transcription factor database (BacTFDB) that was used to train a deep learning model to predict novel TFs and their families in metagenomics and metranscriptomics samples (PredicTF). PredicTF provides the first tool to profile TFs in yet-to be cultured bacteria and it opens the potential to evaluate regulatory networks in complex microbial communities. PredicTF is a flexible, open source pipeline able to predict and annotate TFs in genomes and metagenomes. PredicTF is available at https://github.com/mdsufz/PredicTF.

Key words: Gene Regulation, Complex Promoters, Transcription Factors, Protein Engineering, Design of Circuits, Machine Learning.

I. INTRODUCTION

1. Regulation of gene expression in prokaryotes

Living organisms have mechanisms associated with the process of maintaining life. Considering the objectives of such processes, these could be basically grouped into two broad categories. The first would comprise processes of a structural nature, that is, related to the conversion of energy and raw material into new components necessary for the maintenance and growth of the cell. The second category would comprise processes of a regulatory nature, that is, those dedicated to the control of processes belonging to the first category. This second group has the function of coordinating the different cellular processes depending on factors present in the external and intracellular environment, necessary for the cell to maximize the use of the limited resources available in the environment. According to our current knowledge, the main mechanism for controlling cellular processes is obtained through the regulation of gene expression. The importance of adequate control of gene expression is so great that the greater the complexity of organisms, the greater the sophistication and the abundance of molecular mechanisms used for this purpose (Cases et al., 2003; Konstantinidis and Tiedje, 2004; Koonin and Wolf, 2008).

The regulation of gene expression can be defined as the regulation of information encoded in the gene to result in a gene product or function. It includes a wide variety of mechanisms that are used by cells to increase or decrease the production of specific gene products, such as proteins or RNA. Sophisticated gene expression programs can be observed in biology, for example, to trigger developmental pathways, respond to environmental stimuli or adapt to new food sources. Virtually any step of gene expression can be modulated, from initiation of transcription, processing of RNA, to post-translational modification of a protein. Often, one regulatory gene control another, and so on, in a transcriptional network. Gene regulation is therefore essential for all organisms, as it increases the versatility and adaptability of organisms, allowing the cell to express a protein when needed (Singleton, 2009).

Prokaryotes, since they have a simpler genetic structure compared to eukaryotes (Lawrence, 1999), and as they have easy tools for efficient genetic manipulation (Sambrook et al., 1989), can be considered a good experimental model for the search and the study of mechanisms associated with the control of gene expression (Silva-Rocha and De Lorenzo, 2010; Gama-Castro et al., 2016). Bacteria use a variety of mechanisms to target RNA polymerase to specific promoters in order to activate transcription in response to growth signals or environmental stimuli. Activation can occur due to factors that somehow interact with specific promoters, increasing the transcription directed by these promoters. Alternatively, activation may be due to factors that interact with RNA polymerase, changing their preferences for target promoters (Browning and Busby, 2004).

Bacterial transcription occurs due to the action of the DNA-dependent multi-subunit RNA polymerase enzyme, which is able to synthesize RNA, but is unable to locate promoters. A sigma subunit is necessary for bacterial RNA polymerase to recognize a promoter. The promoters, therefore, control the transcription of all genes, since the initiation of transcription requires the interaction of the promoter with RNA polymerase forming an open complex. The key step in the initiation of transcription is the recognition of the promoter by RNA polymerase, and different elements of DNA sequence responsible for this have been studied (Browning and Busby, 2004).

Both the recognition of the promoter by RNA polymerase and its activity are regulated by a series of transcriptional factors (TFs) that join areas adjacent to the union region of this enzyme (Browning and Busby, 2004). These TFs can modulate the activity of target promoters in a positive (in the case of activators) or negative (repressive) ways through a variety of mechanisms (Browning and Busby, 2004). In turn, TFs modulate gene expression, through different mechanisms, such as the DNA binding affinity for TFs, which can be modulated by small ligands or by covalent modifications, and, in addition, changes in cell concentration of TFs can control promoter activity. Finally, TF can be sequestered by a regulatory protein which binds to it and consequently

modulates its activity (Browning and Busby, 2016). The main TFs in *Eschericha coli*, as well as the mechanisms of gene regulation, will be better described in the following topics.

2. Transcription Factors and Signal Integration in Bacterial Promoters

The *Escherichia coli* genome has more than 300 genes that encode proteins that are able to bind to promoters, regulating transcription in a positive or negative way. Most of these proteins are specific DNA-binding molecules and this ensures that their actions are targeted at specific promoters. Many of these proteins control a vast number of genes, while others control only few genes (Martínez-Antonio et al., 2003; Browning and Busby, 2004). It is estimated that the seven transcription factors CRP, FNR, IHF, Fis, ArcA, NarL and Lrp, are capable of controlling 50% of all *E. coli* genes, while another 60 transcription factors are capable of controlling only one sole promoter (Martínez-Antonio et al., 2003). In this sense, TFs with a larger repertoire of targets are known as global regulators and coordinate the expression of the various genes according to priority stimuli, such as carbon source, oxygen concentration, physiological state of the cell, etc (Martínez-Antonio et al., 2003; Browning and Busby, 2004).

In the context of gene expression modulation, repressors bind to DNA targets that overlap essential elements at their target promoters, thereby occluding access of RNA polymerase (Figure 1A) (Bintu et al., 2005b, 2005a). In many cases, repression is enhanced by multiple binding of repressor molecules, which at some promoters bind distally to each other and interact with each other via DNA loops (Figure 1B). At other promoters, RNA polymerase is able to engage but is blocked at the promoter by the repressor (Figure 1C). Most bacterial transcription factors that function as activators bind to DNA targets located just upstream of the essential elements at their target promoters (Lee et al., 2012). Such factors often interact with RNA polymerase, and this results in its recruitment to the target promoter, thereby increasing transcript initiation (Figure 2D) (Busby and Ebright, 1994). The activation surface on the factor, known as the Activating Region,

is usually composed of a small cluster of amino acid sidechains that make direct contact with a cognate surface somewhere on RNA polymerase, usually on Domain 4 of the RNA polymerase sigma subunit or the C-terminal domain of the RNA polymerase alpha subunit. Other activators induce conformation changes in promoter DNA that result in adjustment in the spacing between different essential elements such that they can be served by RNA polymerase (Figure 2E) (Brown et al., 2003). For the majority of activators that function at promoters served by RNA polymerase carrying the “housekeeping” sigma factor (or one related to the housekeeping sigma), transcript activation occurs without any major conformation change in the RNA polymerase. However, for RNA polymerase carrying a sigma factor related to Sigma-54, this is not the case, as major conformation changes are required for transcript initiation (Yang et al., 2015). These conformation changes are driven by activator– RNA polymerase interactions energized by ATP hydrolysis by a special class of activators known as enhancer-binding proteins (Huo et al., 2009).

Molecular analysis of the regulatory regions of many bacterial transcription units has shown that they are often not simple, with the involvement of many different transcription factors (Barnard et al., 2004). Since the activity of most bacterial transcription factors is regulated by just one signal, we can regard bacterial promoters as integration devices converting messages from the different factors into a single output. Hence, here, we consider three classes of promoters: those controlled by both an activator and a repressor, those controlled just by repressors, and those controlled by two activators.

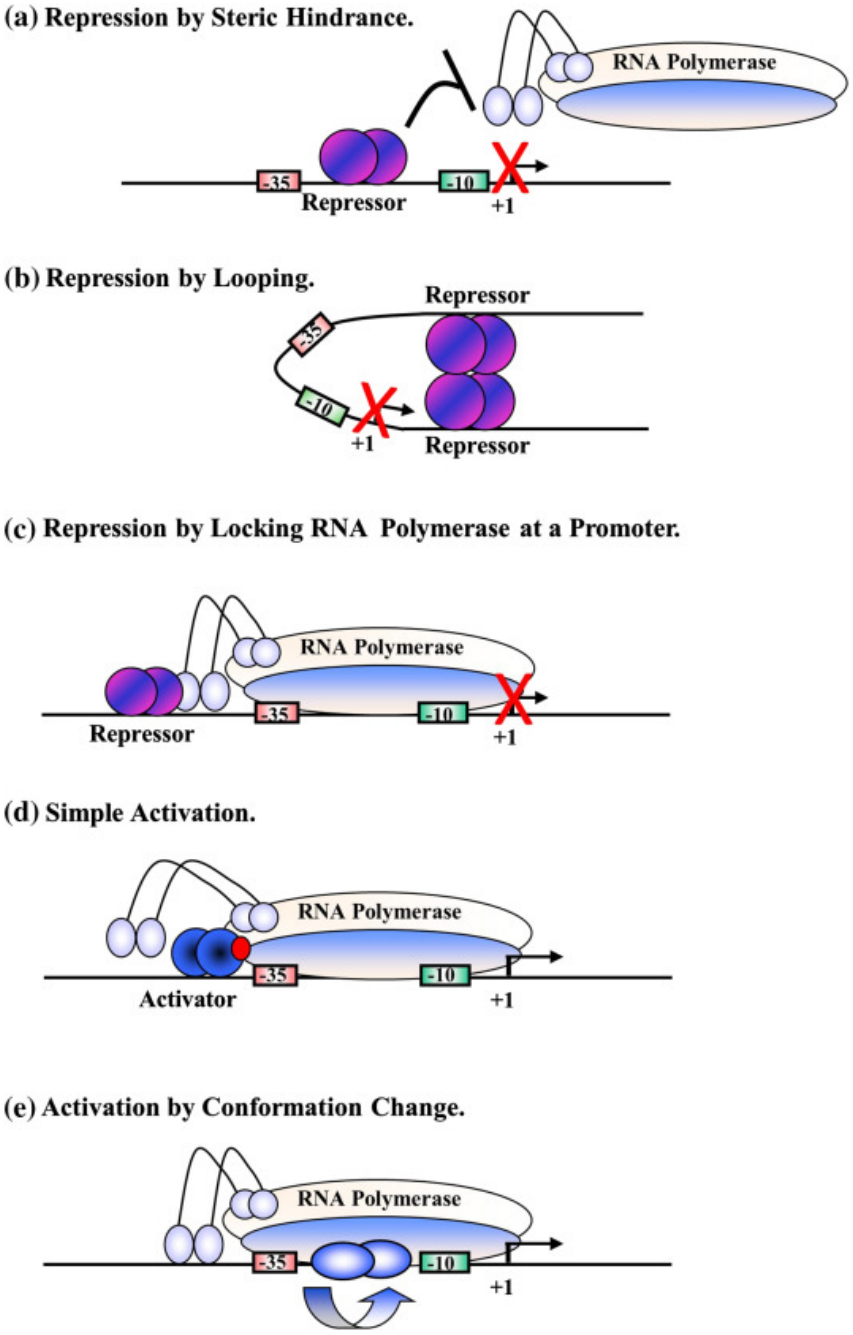


Figure 1. Mechanisms of repression and activation by transcription factors at bacterial promoters. In each panel, the target promoter region is shown as a line with different promoter elements shown by rectangles, and the transcript start, marked + 1, indicated with a bent arrow that shows the direction of transcription. Transcription factors are shown as circular or oval dimers. The multisubunit RNA polymerase is sketched, as in Fig. 1, with the two catalytic subunits, β and β', drawn as a larger oval, the sigma subunit drawn as a smaller darker-shaded oval, and each of the two α subunits drawn as a dumbbell with a curved line to illustrate the flexible linker between the N- and C-terminal domains (illustrated by the two lobes of each dumbbell). (a) A repressor binds adjacent to key promoter elements and prevents RNA polymerase engagement. (b) Repressor dimers bind at some distance from promoter elements but interact, thereby preventing RNA polymerase access to the promoter. (c) RNA polymerase binds to the promoter but is jammed by repressor binding. (d) Activator provides direct contact (small circle) with RNA polymerase, thereby recruiting RNA polymerase to the promoter and facilitating transcript initiation. (e) Activator alters the juxtaposition of essential promoter elements so as to enable RNA polymerase binding and subsequent transcript initiation. Figure taken from Busby, 2019 (Browning et al., 2019).

3. Synthetic Biology

Synthetic biology is a young discipline that seeks to design and construct new biological entities such as enzymes, genetic circuits, and cells or the redesign of existing biological systems. The goal of synthetic biology is to extend or modify the behavior of organisms and engineer them to perform new tasks. One useful analogy to conceptualize both the goal and methods of synthetic biology is the computer engineering hierarchy (Figure 2) (Andrianantoandro et al., 2006). Within the hierarchy, every constituent part is embedded in a more complex system that provides its context. Design of new behavior occurs with the top of the hierarchy in mind but is implemented bottom-up (Andrianantoandro et al., 2006). At the bottom of the hierarchy are, for example, DNA, RNA, proteins, and metabolites (including lipids and carbohydrates, amino acids, and nucleotides), analogous to the physical layer of transistors, capacitors, and resistors in computer engineering. At the next layer, are biochemical reactions that regulate the flow of information responsible for physical processes, equivalent to engineered logic gates that perform computations in a computer. At the module layer, the synthetic biologist uses a diverse library of biological devices to assemble complex pathways that function like integrated circuits. The connection of these modules to each other and their integration into host cells allows the synthetic biologist to extend or modify the behavior of cells (Andrianantoandro et al., 2006).

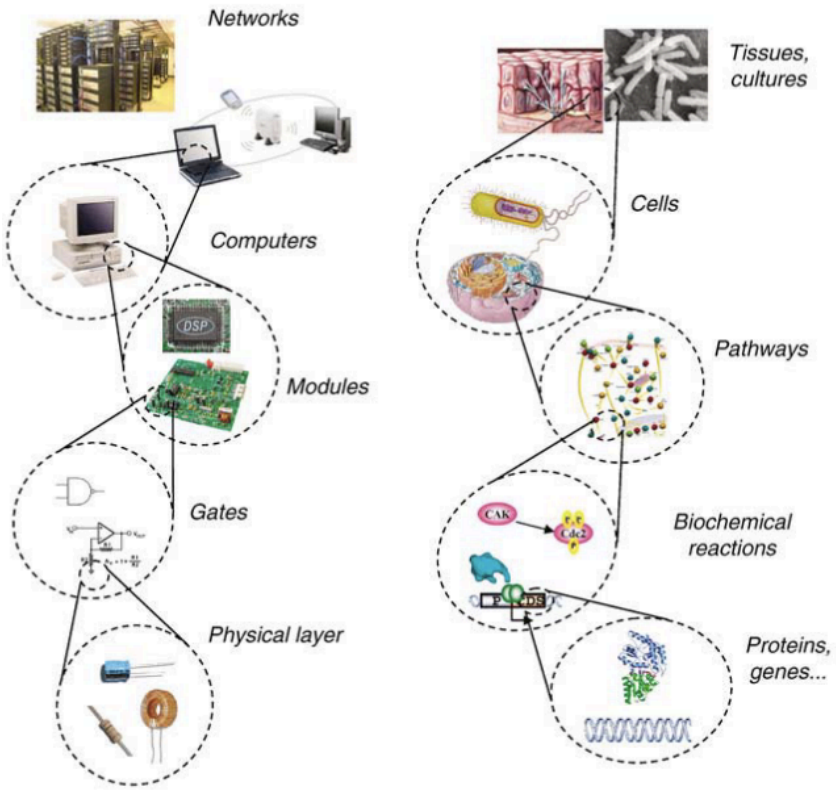


Figure 2. A possible hierarchy for synthetic biology is inspired by computer engineering. Figure taken from Andrianantoandro, 2006 (Andrianantoandro et al., 2006).

The Synthetic biology fundaments goes through the designing of a new biological entity using bioinformatics or modelling approaches, the building of an entity by DNA design and synthesis, the testing or validation of the project using, for example, automated genomes or plasmid engineering and finally, the learning by assessment and enrichment of engineering biological parts (Figure 3). Thus, synthetic biologists have several objectives to be achieved as design and build engineered biological systems. Synthetic biology includes several working to develop like: 1) Standardized biological parts - identify and catalog the standardized genomic parts that can be used (and quickly synthesized) to build new biological systems (Elowitz and Lim, 2010; Way et al., 2014; Sanches-Medeiros et al., 2018); 2) Applied protein design - redesigning the existing biological parts and expanding the set of functions of the natural protein to new processes (Hyeon et al., 2016a; Liu and Chen, 2016; Wang et al., 2018); 3) Natural product synthesis - design microbes to produce all the enzymes and biological functions needed to perform complex multi-

stage production of natural products (Hyeon et al., 2016b; Noda et al., 2016; Siu et al., 2017); and 4) Synthetic genomics - design and build a "simple" genome for a natural bacterium (Zhang and Voytas, 2018; Zhang et al., 2020).

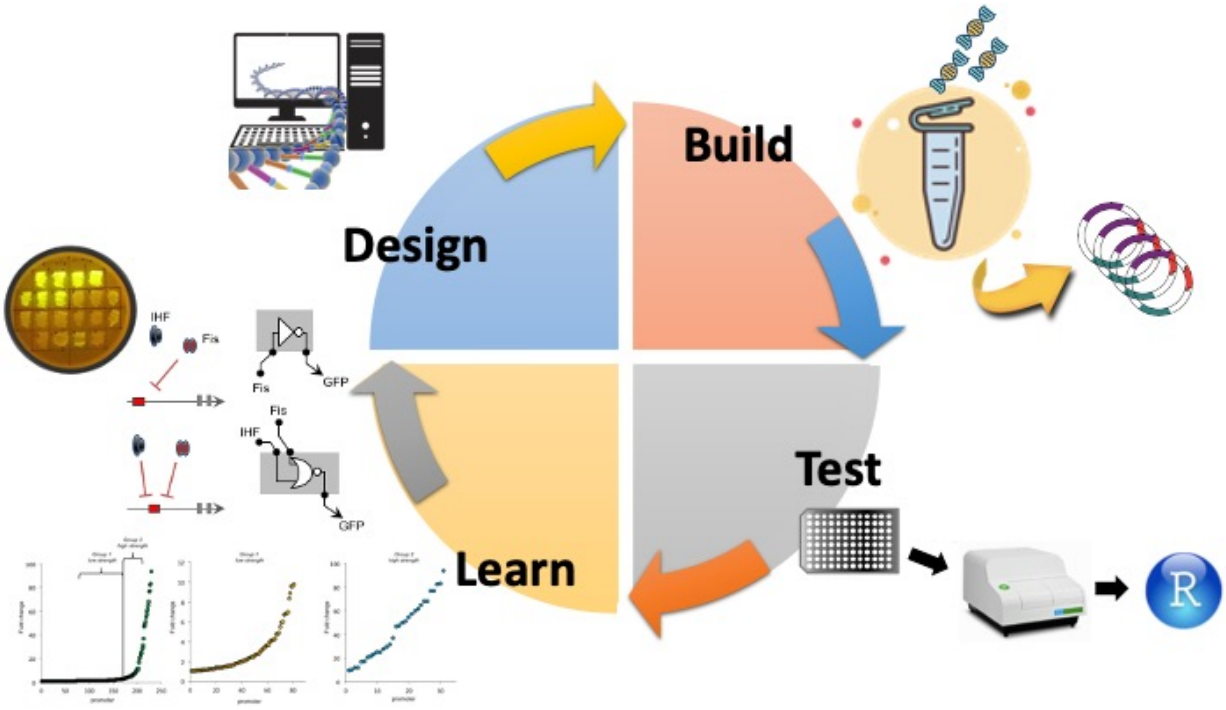


Figure 3. Synthetic biology fundaments. Desing, Build, Test, Learning.

The element that distinguishes synthetic biology from traditional molecular and cellular biology is the focus on the design and construction of core components (parts of enzymes, genetic circuits, metabolic pathways, etc.) that can be modeled, understood, and tuned to meet specific performance criteria, and the assembly of these smaller parts and devices into larger integrated systems to solve specific problems (Khalil and Collins, 2010; Deplazes-Zemp, 2012). Unlike many other areas of engineering, biology is incredibly non-linear and less predictable, and there is less knowledge of the parts and how they interact. Hence, the overwhelming physical details of natural biology (gene sequences, protein properties, biological systems) must be organized and recast via a set of design rules that hide information and manage complexity, thereby enabling the

engineering of many-component integrated biological systems. It is only when this is accomplished that designs of significant scale will be possible (Benner and Sismour, 2005; Cameron et al., 2014).

4. Synthetic Biology and the redesign of biological circuits

The use of logical models for the analysis of regulatory networks, as well as the implementation of new "logic gates" in biological networks through the engineering of regulatory elements, belong to the field of Synthetic Biology (Cameron et al., 2014). Synthetic Biology seeks to reprogram biological circuits in order to modify living systems, generating new behaviors of interest in fields such as biotechnology (Cameron et al., 2014). An advantage of this approach is that the process of reissuing the regulatory network also allows unveiling unknown properties of these systems, which could not be found through classical strategies. Thus, Synthetic Biology allows to explore these properties of the regulatory networks in addition to generating practical knowledge about the study systems, considering that the ultimate goal of this field is to generate a biological system with new properties (Andrianantoandro et al., 2006).

Two main approaches have been used for the redesign of logical behavior in biological systems. The first is to use pre-existing regulatory elements, such as TFs and promoters and reconnect them, following a specific design to obtain the desired behavior (Silva-Rocha and De Lorenzo, 2011; Tamsir et al., 2011). The second is to re-edit the regulatory elements to modify their intrinsic properties and thus generate new logical behaviors. In this sense, it has been demonstrated that, in fact, the modification of *cis*-regulatory elements at target promoters is sufficient to modify the logic of their signal integration (Hunziker et al., 2010).

As Boolean logic gates are widely used in electronic circuits to build digital devices, logic operations are encoded in gene regulatory networks that cells use to integrate multiple environmental and cellular signals to respond accordingly (Wang et al., 2011). The combination of logic gates together with new experimental designs has enabled important advances in

understanding the relationship between the architecture of regulatory elements (mainly promoters) and the final response generated as a result of the signal integration process (Tamsir et al., 2011). However, our knowledge of the relationship between the promoters' architecture and the resulting logic is still quite limited. For example, while deciphering the logic of promoters of low complexity (that is, with few operator sequences) is a relatively easy task, the same cannot be said for promoters formed by numerous regulatory sequences (Ishihama, 2010). In addition, recent progress in the massive identification of operator sequences on a genomic scale (for example, through chromatin immunoprecipitation experiments and computational tools) has shown that the integration of multiple operators into bacterial promoters is more frequent than previously imagined (Shimada et al., 2011a, 2011b; Chen et al., 2018; Barne et al., 2019).

5. Synthetic biology and the development of bioinformatics tools to explore the microbial communities' complexity

Progress in synthetic biology is enabled by powerful bioinformatics tools allowing the integration of the design, build and test stages of the biological engineering cycle. The development of new bioinformatics tools is helpful and important for scientists working on design, build and test parts to extend or modify the behavior of organisms and engineer them to perform new tasks. Bioinformatics tools for the DESIGN and BUILD stages include tools for the selection, synthesis, assembly and optimization of parts (enzymes and regulatory elements), devices (pathways) and systems (chassis). TEST tools include those for screening, identification and quantification of metabolites for rapid prototyping (Carbonell et al., 2016).

The functional potential of microbial communities can be determined by the genetic content of its constituent members. However, genetic content alone does not guarantee that a given function or enzymatic reaction will be performed as predicted (Liu et al., 2019). In this scenario, Transcription Factor proteins (TFs) play a central and critical role in gene regulation. Since TFs

may determine when and which genes are expressed, profiling TFs can help understand the regulation of gene expression and to build regulatory networks in complex microbial communities. Further, defining which factors control gene expression may offer insights into the mechanisms controlling ecosystem processes and even interactions between species of a microbial community.

One of the major goals in the manipulation of microbiomes for ecological and biotechnological applications is to control the outcome of their functions (Widder et al., 2016). As TFs are key to potentially control which genes are expressed, one of the best ways to study and understand gene regulation in a microbiome may be to profile its TFs. Unfortunately, to date, no platform supports prediction and classification of novel bacterial TF from 'omics data recovered from microbial communities.

Although the regulation of gene expression in prokaryotes has been studied extensively, there is still a void regarding the effects that different regulatory elements would have on the final system logic. Unraveling these rules would help to create a predictive system able to decipher the logic of natural transcription factors and promoters. These rules would allow a significant advance in the understanding of the combinatorial mechanisms of control of gene expression in bacteria, with the potential impact on biotechnological applications on the re-engineering of biological circuits. The results of this endeavor contribute not only to a fundamental understanding of the signal integration mechanisms for bacteria and microbial communities but also produce new rules and methods for the design and build of standardized, integrated biological systems to accomplish many tasks of biotechnological interest.

II. OBJETIVES

1) Main objective

Unraveling rules is fundamental for the understanding of the signal integration mechanisms for bacteria and microbial communities. Decipher the logic of natural transcription factors and promoters is a start point to create a predictive system with potential impact on biotechnological applications on the re-engineering of biological circuits. In this way, the general objective of this thesis was to engineer biological systems at the level of regulators and promoters through synthetic biology approaches and, in addition, to use a systemic approach to decipher the complexity of microbial communities through the mining of regulatory elements in massive data.

2) Specific objectives

- Decipher the logic of integrating signals in complex promoters, seeking to characterize biological parts and understand the relationship between the architecture of complex promoters and the logic of gene regulation dependent on global regulators in bacteria.
- Engineer a set of bacterial transcription factors from *Pseudomonas putida* with enhanced response to a well characterized and largely innocuous molecule with a potential for eliciting heterologous expression of bacterial genes in animal carriers.
- Decode the complexity of transcription factors in environmental communities by mapping the complex regulatory profile of (meta)genomes of interest by searching for know and novel transcription factors through bioinformatics tools.

III. RESULTS

CHAPTER I

Emergent Properties in Complex Synthetic Bacterial Promoters

This chapter was published as:

(Monteiro et al., 2018) MONTEIRO, Lummy Maria Oliveira; ARRUDA, Leticia Magalhaes; SILVA-ROCHA, Rafael. Emergent properties in complex synthetic bacterial promoters. **ACS synthetic biology**, v. 7, n. 2, p. 602-612, 2018.

ACS Synthetic Biology

Research Article

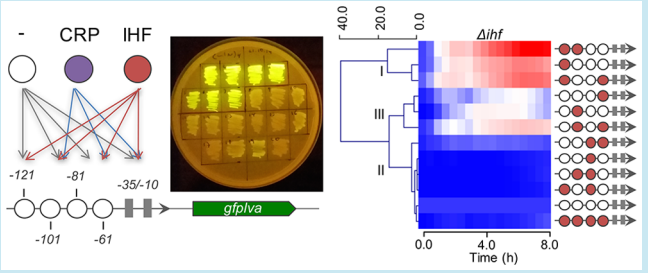
pubs.acs.org/synthbio

Emergent Properties in Complex Synthetic Bacterial Promoters

Lummy Maria Oliveira Monteiro,[†] Letícia Magalhães Arruda,[†] and Rafael Silva-Rocha*

Systems and Synthetic Biology Lab, Ribeirao Preto Medical School, University of São Paulo, Ribeirão Preto, São Paulo, Brazil

ABSTRACT: Regulation of gene expression in bacteria results from the interplay between hundreds of transcriptional factors (TFs) at target promoters. However, how the arrangement of binding sites for TFs generates the regulatory logic of promoters is not well-known. Here, we generated and fully characterized a library of synthetic complex promoters for the global regulators, CRP and IHF, in *Escherichia coli*, which are formed by a weak −35/−10 consensus sequence preceded by four combinatorial binding sites for these two TFs. Using this approach, we found that while *cis*-elements for CRP preferentially activate promoters when located immediately upstream of the promoter consensus, binding sites for IHF mainly function as "UP" elements and stimulate transcription in several different architectures in the absence of this protein. However, the combination of CRP- and IHF-binding sites resulted in emergent properties in these complex promoters, where the activity of combinatorial promoters cannot be predicted from the individual behavior of its components. Taken together, the results presented here add to the information on architecture-logic of complex promoters in bacteria.



KEYWORDS: *emergent properties, global regulators, promoter architecture, regulatory networks, synthetic biology*

The experience of the past decade has greatly increased our knowledge of how cells coordinate gene expression in response to changing environmental and physiological conditions. Since the seminal description of the first gene regulatory mechanism by Jacob and Monod in the 60s, thousands of molecular studies have described the different mechanisms by which transcriptional factors (TFs) coordinate gene expression in bacteria. In particular, the model organism *Escherichia coli* has been used for decades to investigate the different ways in which TFs activate or repress gene expression, and a number of mechanisms have been elucidated in this and other bacteria.[1−8] With an increase in our knowledge of these mechanisms, it was soon evident that bacterial promoters are usually regulated by several TFs that bind to specific *cis*-regulatory elements located in close proximity to the promoter site and interact with one another in different ways. In this sense, the existence of synergy or competition between TFs for binding sites in the DNA will ultimately determine the level and timing of expression for each particular gene depending on the combination of specific molecular signals available to the bacteria.[9,10] Additionally, compilation of the regulatory interactions known for *E. coli* resulted in the classification of TFs as global and local regulators, where the first group is composed of TFs capable of controlling a large number of target genes, whereas the second group has a more limited regulatory scope.[11,12] This analysis also showed that some environmental and physiological signals that control global regulators are higher in the regulatory hierarchy since their presence will lead to major regulatory effects in the organisms compared to the presence of signals for local regulators. For instance, the CRP global regulator controls the expression of a large number of genes in *E. coli* in response to changes in cAMP levels, which in turn is modulated by glucose.[13,14] Although it is well established that cAMP synthesis is modulated by glucose, it is noteworthy that several sugars independent of phosphotransferase system can also reduce cAMP levels.[15] In another case, the nucleoid associated protein IHF has an important role in DNA organization in response to bacterial growth and can modulate the expression of a number of genes.[16,17] Furthermore, many global regulators are known to co-occur frequently at target promoters,[5,11] and this co-occurrence could indicate the existence of some interaction mechanisms between these pairs of regulators.[18,19]

With the advent of synthetic biology, using the current knowledge on gene regulatory mechanisms in bacteria to reprogram these organisms for novel applications has been of special interest.[20−22] In order to accomplish this task, many studies have addressed the modification of native promoters to construct synthetic regulatory systems with enhanced and/or modified performance.[23−26] Moreover, some initial studies have focused on the shuffling of *cis*-regulatory elements to reconstruct complex promoters in bacteria;[27−30] this approach could not only provide novel regulatory systems but also reveal some of the hidden roles regarding the interaction of multiple TFs in target promoters. Though these approaches have resulted in significant progress such as the knowledge that promoter arrangement indeed determines the final regulatory logic of systems, these studies have mainly used local regulators and it is not yet known whether global regulators would follow the same rules. Moreover, the standard model for gene regulation in bacteria states that we could anticipate the

Received: September 27, 2017
Published: November 1, 2017

DOI: 10.1021/acssynbio.7b00344
ACS Synth. Biol. 2018, 7, 602−612

regulatory behavior of complex promoters by analyzing the individual contributions of each TF and its respective *cis*-elements, as evidenced by the widely used mathematical frameworks available to model gene regulation.[31,32] However, in an alternative scenario, the combination of several *cis*-regulatory elements for specific TFs (mainly for global regulators that naturally act together) could result in promoters with emergent properties, where the final response of the system would not be anticipated based on known individual contributions. This hypothesis is also motivated by the fact that many biological systems have been shown to display emergent properties.[33]

In order to get insights into the regulatory mechanisms of combinatorial bacterial promoters, we investigated here the relationship between promoter architecture and gene expression regulation. These insights are relevant since promoters in their natural form are known to co-occur frequently at target promoters.[18,19] This study is important, since many *E. coli* genes can be regulated by CRP, IHF or for both at the same time. This is the case of *gcd* gene that is regulated by one CRP (repressor) and one IHF (activator). This gene encodes for quinoprotein glucose dehydrogenase.[34] Another natural example is the *hpt* gene that is regulate by one CRP (activator) and one IHF (repressor). This gene encodes for hypoxanthine phosphoribosyltransferase. This protein is involved in step 1 of the subpathway that synthesizes IMP (Inosine monophosphate) from hypoxanthine.[35] And finally, the *acs* gene that is regulate by at least two CRP (activator) and three IHF (repressor). This gene encodes for acetyl-CoA synthetase. Acetyl-CoA synthetase is an enzyme involved in metabolism of acetate. There are many other known genes that are regulated by CRP and IHF at the same time. In addition, possibly many other genes have not yet been studied.[36] For this, we constructed and characterized a library of synthetic promoters containing an array of *cis*-elements for CRP and IHF, two global regulators of *E. coli*, using a GFP(LVA) reporter assay. Our data clearly indicated that though CRP and IHF have very different regulatory effects, many binding site combinations for both TFs resulted in novel regulatory activities that were not anticipated by the analysis of individual elements when their sites were placed in different positions and arrangements. These results demonstrate the existence of emergent properties in complex synthetic promoters in *E. coli*, which could be extrapolated to naturally occurring regulatory systems and would significantly impact the engineering of synthetic biological circuits in bacteria.

RESULTS

CRP Strongly Activates Synthetic Promoters with *cis*-Elements Immediately Upstream of a Core Promoter. In order to investigate the architecture-logic relationship in synthetic bacterial promoters, we employed a minimal design as presented in Figure 1. First, we designed a promoter composed of a weak core element (comprising the −35 and −10 boxes of *Plac*) preceded by 20 bp sequences that could be occupied by *cis*-elements for the target TFs (Figure 1A). In this design, the *cis*-elements can be centered at regions −61, −81, −101, and −121 related to the transcriptional start site (TSS) of the promoter. By fixing these positions, we would expect the effect of TF to be stimulatory at the resultant promoter, based on previous systematic inspections on the effect of *cis*-element localization on gene regulation.[5,37] For each of the four potential positions, we designed double stranded DNA

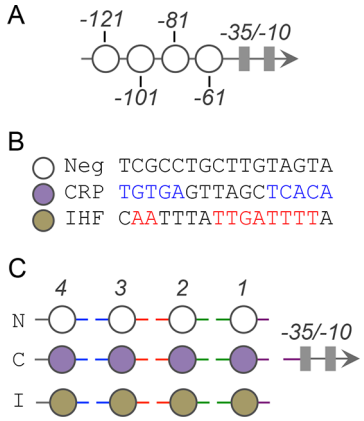


Figure 1. Construction of the complex promoter library. (A) Schematic representation of the promoter library, showing the positions −121, −101, −81, and −61 (white circles) at which *cis*-elements were inserted. The −35 and −10 boxes (gray rectangles) correspond to the core promoter. (B) Nucleotide sequences for Neutral (N), CRP (C), IHF (I) *cis*-elements. (C) Simplified scaffold scheme for the minimal synthetic promoter library. Motifs positions are identified as 4, 3, 2, and 1 respective to the core promoter, and colored lines represent the cohesive sequences for DNA ligation. (D) *E. coli* library transformants showing different promoter strengths.

oligonucleotides with a consensus sequence for CRP, IHF, and a control sequence (called "Neg", Figure 1B) that does not display any stimulatory effect *in vivo*.[18] In this sense, each double stranded DNA fragment has 3′ overhang elements with four nucleotides that specify each position where the fragment can be ligated (Figure 1C). In this manner, we used a portfolio of 12 different fragments that could be used to construct up to 81 (3^4) different combinatorial promoters. Using this setup, we constructed a library of synthetic promoters by ligating these *cis*-elements and the core promoter into a GFP(LVA) reporter vector, allowing the measurement of promoter activities *in vivo* to determine the effect of promoter architecture in the final output of promoters (exemplified by the different clones represented in Figure 1D).

In order to determine the effect of different arrangements of *cis*-elements for CRP, we analyzed the promoter activity of 10 synthetic promoters in the wild type strain growing in minimal media for a period of 8 h. As shown in Figure 2, clustering of promoter activities reveals the existence of two clear groups, one (marked as **I** in the figure) composed of six promoters with activities similar to the reference promoter (the one in the top with four Neg sites) and another group (marked as **II**) composed of four promoters with a high level of activity (about 80 times the level of the reference promoter). By analyzing the architecture of each promoter, it is easily notable that all members of the highly active group have a *cis*-element for CRP at position 1 (equivalent to the −61 relative to the TSS), which is in accordance to previous reports on this TF.[19,37,38] Moreover, the addition of another CRP *cis*-element at positions 2, 3, and 4 (boxes −81, −101, and −121) only marginally affects the activity of a promoter harboring a *cis*-element at position 1. In summary, this result demonstrates the potential of our approach to investigate the effect of *cis*-element arrangement on promoter activity, as the resulting synthetic promoters reproduced the expected behavior for CRP.

***cis*-Element for IHF Enhances Promoter Activity in the Complex Promoters in the Absence of This TF.** In order to investigate the effect of *cis*-elements for IHF in our complex

DOI: 10.1021/acssynbio.7b00344
ACS Synth. Biol. 2018, 7, 602−612

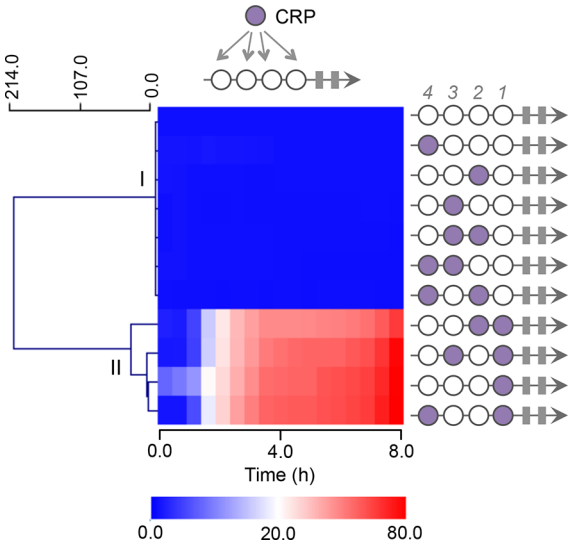


Figure 2. CRP motif at position 1 is fundamental for high promoter activity. A subset of 11 synthetic promoters containing shuffled CRP and Neutral *cis*-elements displaying two clear activity patterns (groups **I** and **II**). In group **I** are promoters that do not present promoter activity while group **II** includes promoters with high transcription rates. Circles in magenta represent the positions of CRP sites. Relative promoter activity was measured for 8 h, calculated based on the Neutral full promoter, and displayed on an intensity scale from 0.0 to 80.0. Plots were calculated based on the average of three independent experiments.

promoter design, we analyzed 11 synthetic promoters harboring combinations of one or two IHF binding site and Neg sequences in addition to the full IHF P_{III} promoter (Figure 3). The experiments were performed similarly as before but both in the wild type and Δ*ihf* mutant strains of *E. coli*. Promoter activity analysis allowed clustering of the data into three major groups as shown in Figure 3A for the wild type strain. In this sense, group **I** (high activity) was formed of two promoters with maximal activity not higher than 8 times that observed for the reference promoter, whereas group **II** (Medium activity) (5 promoters) displayed activities comparable to those of the reference, and group **III** (Low activity) (4 promoters) showed intermediate activity. When the same set of promoters was assayed in Δ*ihf* mutant strains of *E. coli*, we observed two major features (Figure 3B). First, a generalized increase in promoter activity was observed for groups **I** and **III**, where the former was still formed of promoters with stronger activity. Second, the composition of the groups was almost unchanged, with the exception of one promoter (with two IHF *cis*-elements at positions 4 and 3) that displayed no activity in the wild type but showed the highest activity of the group in the mutant, and another promoter (composed of four IHF binding sites) that did not gain activity in the mutant, were clustered into group **II** in the mutant. These expression profiles clearly indicate that the *cis*-elements for IHF could stimulate promoter activity mainly in the absence of this global regulator. In addition, groups I and III which display significant promoter activity in mutant strain, they have IHF *cis*-elements at all positions except 2 (equivalent to the −81 region), whereas promoters with position 2 occupied displayed very low activity regardless of occupancy at other sites (group **II**). These results suggest that *cis*-elements for IHF could operate as an RNAP transcriptional activity enhancer, probably as a UP element-like motif as described previously[39,40] although it have not been described UP-element at position −60 to −120 yet. Our hypothesis is that the IHF *cis*-element would be acting recruiting the RNAP to the promoter; however, experiments proving this hypothesis are still necessary. Thus, when IHF binds to its cognate *cis*-element, it blocks RNAP contact with the UP element-like sequence, thus preventing transcriptional stimulation of the promoter. However, the reason why the

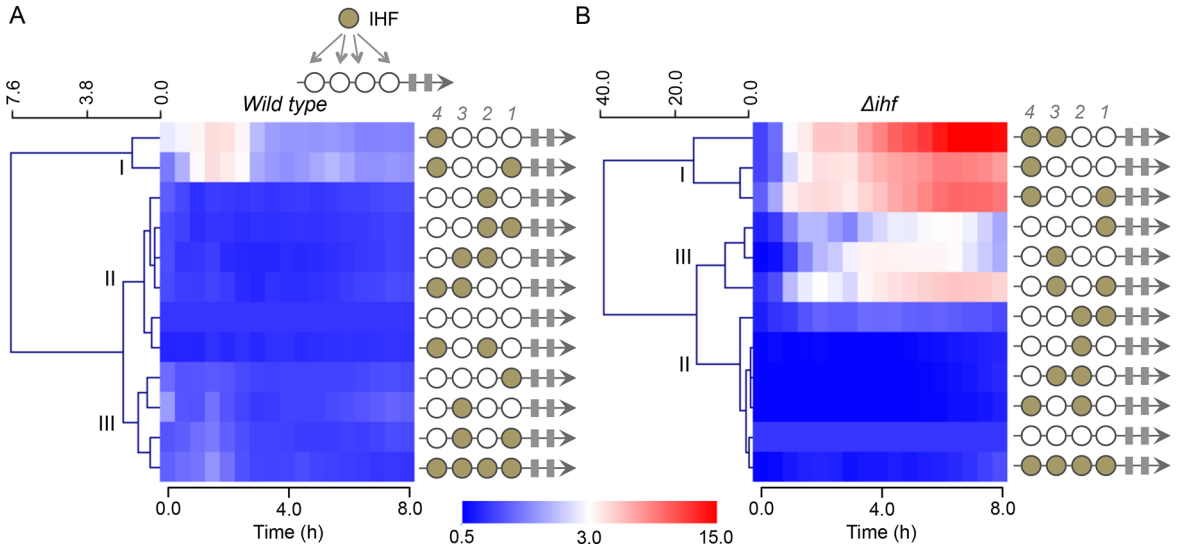


Figure 3. IHF motif enhanced promoter activity in *E. coli* Δ*ihf* strain. A subset of shuffled IHF and Neutral motif promoters were assayed in the wild type and Δ*ihf* mutant strains and grouped according to their relative activity. Circles in beige represent the positions of IHF sites. (A) IHF *vs* Neutral motifs assayed in the wild type strain. Synthetic promoters that showed higher promoter activities are clustered in group **I**, group **II** is formed of promoters with low activity, whereas group **III** is formed of promoters with intermediate promoter activity. (B) The same set of promoters were assayed in the *E. coli* Δ*ihf* mutant strain, highlighting that in the absence of IHF transcription factor, promoter activity was generally improved for the groups **I** and **III**. Relative promoter activity was measured for 8 h, calculated based on the Neutral full promoter, and displayed on an intensity scale from 0.0 to 15.0. Plots were calculated based on the average of three independent experiments.

DOI: 10.1021/acssynbio.7b00344
ACS Synth. Biol. 2018, 7, 602−612

existence of a *cis*-element for IHF at position 2 renders the promoters inactive regardless of the identity of other positions is unknown and may be related to some intrinsic property of the DNA sequence itself.

Rise of Emergent Properties in Complex Promoters with CRP and IHF *cis*-Elements. Once we determined that the CRP and IHF *cis*-elements have distinct effects on promoter activity, we wondered what would happen if the CRP and IHF binding sequences were combined, as occurs in the natural promoters of *E. coli*. In this sense, would the resulting promoter represent the sum of each contribution of the isolated *cis*-elements, or would it display a novel regulatory logic? To address these questions, we used as the start point, two architectures containing *cis*-elements for IHF that displayed activity both in the wild type and Δ*ihf* mutant strains as represented by the members of group I in Figure 3A. These two promoters possess either one position (position 4) or two (positions 4 and 1) occupied by the IHF *cis*-element. Using these two basic architectures, we introduced *cis*-elements for CRP at either position 3, 2, or both, and assayed the resulting promoter activity (Figure 4). In this data set, we did not test

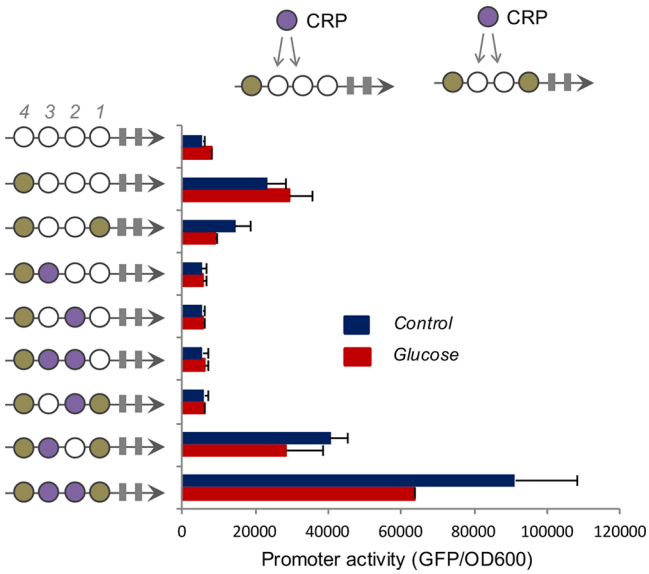


Figure 4. Emergence in combinatorial promoters based on CRP and IHF *cis*-elements. Complex promoters were constructed based on two promoters shown in Figure 3 and were assayed in the *E. coli* wild type strain, in the absence (blue) or presence (red) of 0.4% glucose. The promoters were constructed by fixing the IHF *cis*-element at position 4 and at positions 4 and 1 and by shuffling the CRP motifs at positions 2 and 3. Relative promoter activity was measured after 4 h of incubation. Plots were calculated based on the average of three independent experiments.

position 1 since CRP at this position has a strong stimulatory effect regardless of other upstream elements (as already presented in Figure 2). Notably, when positions 3 and 2 where occupied by CRP *cis*-elements (either in isolation or simultaneously), no significant promoter activity was detected (Figure 2). When we assayed combinatorial promoters based on single IHF *cis*-elements in the wild type strain, we observed that the introduction of single or double *cis*-elements for CRP resulted in complete abolishment of promoter activity. When these promoters were evaluated in the presence of 0.4% glucose (which in our tested condition, is sufficient to block CRP activity[41]), we did not recover the original promoter activity, suggesting that this effect was not dependent on CRP binding but rather on the combination of the DNA elements itself. When we performed a similar analysis on the variants of the promoters harboring the two IHF *cis*-elements (at positions 4 and 1), we observed a remarkably different behavior. When single CRP *cis*-elements were placed individually at positions 3 or 2 (P_{ICNI} and P_{INCI}), we observed the same promoter blocking effect as described previously, and this effect was not alleviated in the presence of glucose (Figure 4). However, when both positions, 3 and 2 (P_{ICCI}), where occupied by *cis*-elements for CRP resulting in a synthetic promoter with two IHF sites flanking two CRP sites, we observed a strong increase in promoter activity compared to that of the original promoter. It is striking to note that the presence of two CRP sites at positions 2 and 3 provides an expressive promoter activity and that withdrawal of only one CRP binding site at position 2 or 3 significantly impairs the promoter force. It is very clear the requirement for the double CRP site when comparing PICCI with PICNI and PINCI promoters. Interestingly, the addition of glucose resulted in complete abolishment of promoter activity, indicating that the strong enhancement of promoter activity was indeed dependent on CRP activity. These data strongly support the notion that the combination of *cis*-elements for global regulators such as CRP and IHF leads to the appearance of emergent properties, since the final regulatory behavior of the complex promoter does not represent the sum of the behavior of the original architectures (*i.e.*, the promoter harboring two IHF sites at positions 4 and 1 and the promoter harboring two CRP sites at positions 3 and 2).

In order to further evaluate the promoter architecture effect, we expanded the number of architectures assayed and performed experiments in the absence of IHF and by modulating CRP activity (*i.e.*, in the presence or absence of glucose). For a better presentation of the results, the experiment was divided into three subgroups. Figure 5A shows constructs that have one IHF *cis*-element fixed at position 4 (−121 region) and different combinations of CRP *cis*-element at positions 2 and 3 (−81 and −101 regions). As shown in the figure, a promoter containing a single IHF site at position 4 displays strong activity in the Δ*ihf* mutant strain that was insensitive to glucose presence. Moreover, addition of single or double CRP *cis*-elements at positions 3 and 2 completely abolishes promoter activity, and this could not be reverted by the addition of glucose to the media. These results agree with the previous analysis and indicate that addition of CRP *cis*-elements blocks the activity of the original promoter independently of CRP activity. Next, we investigated the effect of the presence of CRP binding sites at different positions in promoters with a single IHF *cis*-element fixed at position 1 (Figure 5B). In this condition, though the initial promoter displayed detectable activity with a fold-change about three times that of the reference promoter, addition of a single CRP *cis*-element immediately upstream of the IHF site (position 2) completely abolished the promoter activity. Interestingly, moving the CRP site far from the IHF site (for instance, from position 2 to position 3) generates a marginally detectable activity, whereas placing the site in the farthest position (*i.e.*, at position 4) generates a combinatorial promoter with activity similar to that of the original harboring a single IHF site at position 1 (Figure 5B). These results indicate a position dependent effect of CRP *cis*-elements, which was not related to

DOI: 10.1021/acssynbio.7b00344
ACS Synth. Biol. 2018, 7, 602−612

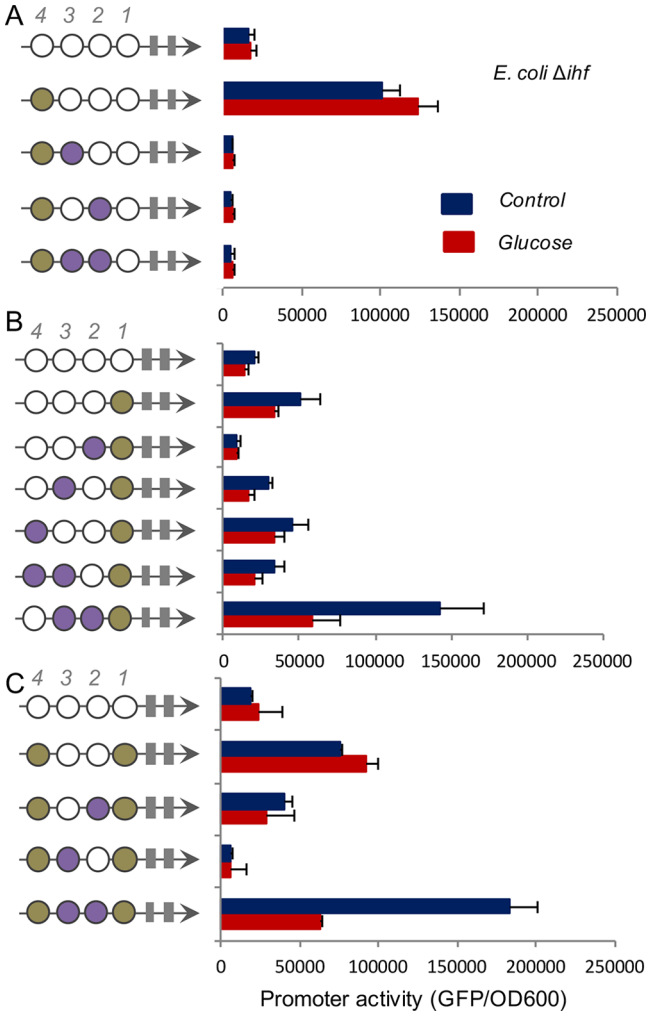


Figure 5. Systematic investigation of complex promoters for CRP and IHF in the *E. coli* Δ*ihf* strain. All experiments were performed in the absence (blue) or presence (red) of 0.4% glucose. (A) Synthetic promoters with a single IHF site fixed at position 4 and varying CRP sites at positions 3 and 2. (B) Synthetic promoters with a single IHF site fixed at position 1 and varying CRP sites at positions 4, 3, and 2. (C) Synthetic promoters with two IHF sites fixed at positions 4 and 1 and varying CRP sites at positions 3 and 2. Relative promoter activity was measured after 4 h of incubation. Plots were calculated based on the average of three independent experiments.

CRP activity since the addition of glucose to the media resulted in very similar expression profiles. This notion is also supported by the fact that addition of two CRP binding sites at positions 4 and 3 resulted in similar activity to that where only position 3 was occupied by a CRP *cis*-element. Additionally, introduction of two CRP binding sites at positions 3 and 2 generated completely different behavior, resulting in a promoter with strong activity that was completely dependent on CRP (as addition of glucose substantially decreased its activity, Figure 5B). These results indicate that the emergence of strong CRP-dependent activity requires two tandem CRP binding sites (at positions 3 and 2) followed by a single IHF *cis*-element (at position 1). Finally, when the CRP binding sites were combined with IHF *cis*-elements fixed at position 4 and 1 (Figure 5C), we observed the same behavior presented in Figure 5B, since addition of one site at position 3 or 2 strongly impairs promoter activity whereas two tandem sites generate a CRP-dependent promoter with activity stronger than that of the parental architecture.

IHF *cis*-Elements Generates Fine-Tuning for CRP Activated Promoters. In the previous sections, we demonstrated that additional CRP *cis*-elements could influence the regulatory behavior of a promoter harboring IHF sites. Since CRP binding sites could strongly influence these promoters, we investigated how the addition of IHF could modulate promoters containing CRP *cis*-elements at position 1, which were previously demonstrated to generate strong CRP dependent activation (Figure 2). For this, we sampled several combinatorial promoters where additional IHF and CRP sites were mixed upstream of a CRP *cis*-element located at position 1. As shown in Figure 6A, all promoters displayed strong activity in the wild type strain of *E. coli* and this activity was severely impaired when glucose was added to the growth media. Additionally, certain degree of heterogeneity can be observed in promoter activities indicating that the additional sites contributed to the final activity observed. However, when the same experiments were performed in a Δ*ihf* mutant strain of *E. coli*, we observed that the level of heterogeneity was strongly reduced both in the active and repressed conditions (Figure 6B). Taken together, these data strongly indicate that additional IHF sites could produce a fine-tuning effect that modulates the final activity of the constructed combinatorial promoters.

Emergent Properties Are Consequence of the *cis*-Elements. Once we affirm the existence of the emergent properties in our promoters, we proceed to study whether this property was given by the *cis*-elements located upstream to the core promoter, or if in some way this emergent promoter activity was given by the generation of a new promoter formed in our constructions. For this, we focused on two promoters (P_{ICCI} and P_{NCCI}) with emergent properties to study this property further. For this, deletions of 5 and 15 base pairs were made in the core promoter (Figure 7A). As it is possible to note that, with the 5 base pair deletion, the −10 and −35 boxes remain intact, while upon deletion of 15 base pairs the −10 box is completely removed. These deletions were also performed for P_{NNNN} as a control. The experiments were carried out in *wt* and Δ*ihf* strains and in presence or absence of glucose. After these deletions, it was possible to conclude that the 5 base pairs deletion (D5) for both P_{ICCI} and P_{NCCI} promoter the activity remained similar to the parental promoter, showing no significant difference. On the other hand, after the 15 base pairs deletion (D15) the promoter activity fell sharply to almost zero (Figure 7B and 7C). In this way, it was possible to verify that the promoter activities as well as the emergent properties found in this work are due to the different architectures promoting the *cis*-elements upstream to the core promoter.

DISCUSSION

Regulation of gene expression at the level of RNAP recruitment to target promoters is known to be a combinatorial mechanism where multiple transcriptional factors binding to target *cis*-regulatory elements and their interplay defines the timing and intensity of gene expression. This combinatorial control has been extensively described in bacteria and in single-celled and multicellular eukaryotes, and the so-called *regulatory code* is known to play a major role in the way living organisms develop and interact with the environment.[27,42,43] However, while classical approaches to understand this code are based on a case-by-case dissection of the *cis*-regulatory elements of

DOI: 10.1021/acssynbio.7b00344
ACS Synth. Biol. 2018, 7, 602−612

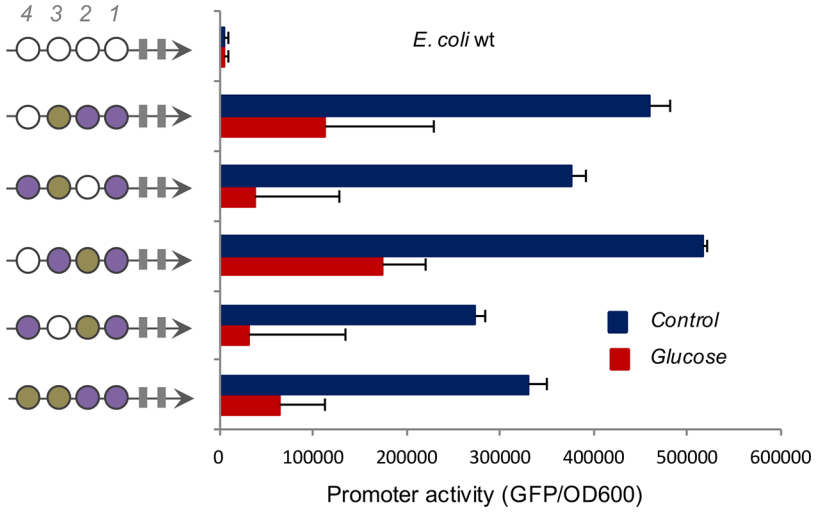


Figure 6. Fine-tuning of CRP-dependent synthetic promoters by IHF sites. Synthetic promoters with the CRP site fixed at position 1 and varying CRP and IHF sites at positions 4, 3, and 2 were assayed in the absence (left) or presence (right) of 0.4% glucose. (A) Analysis of promoter activity in the *E. coli* wild type strain. (B) Analysis of promoter activity in the *E. coli* Δ*ihf* strain. Relative promoter activity was measured after 4 h of incubation. Plots were calculated based on the average of three independent experiments.

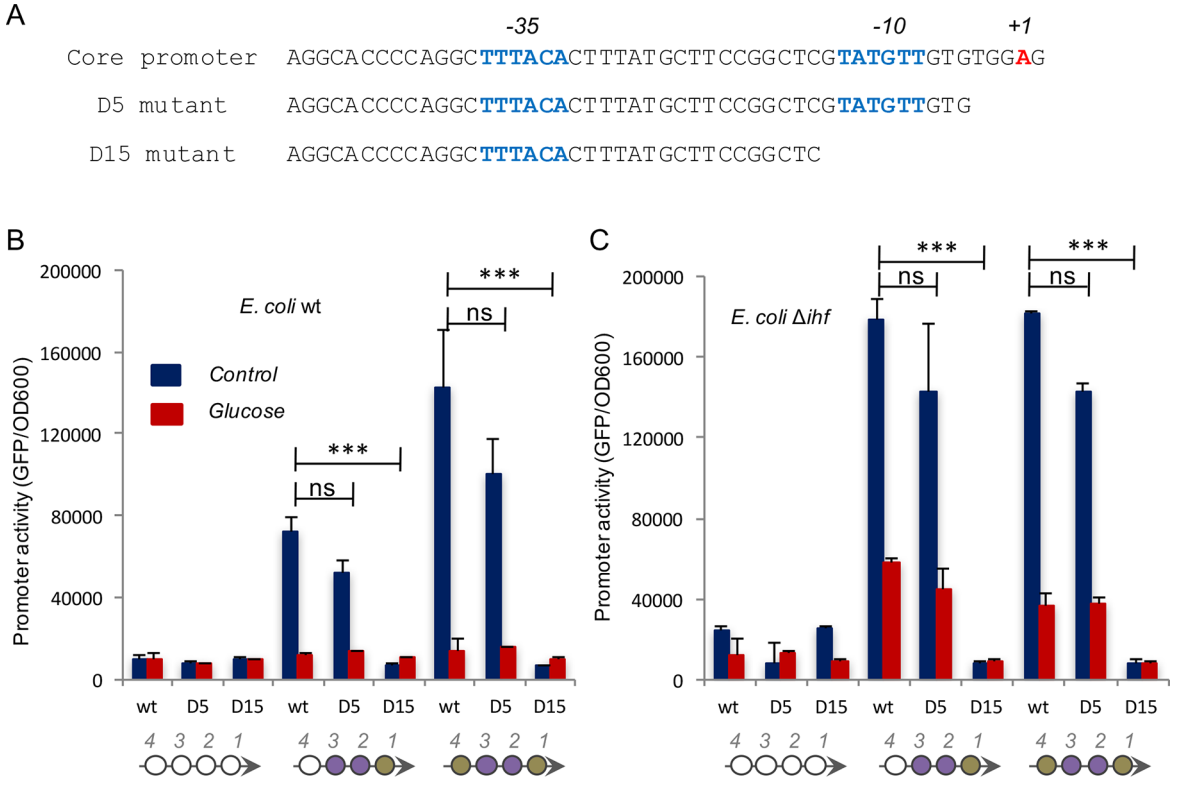


Figure 7. Emergent properties are consequence of the *cis*-elements. All experiments were performed in the absence (blue) or presence (red) of 0.4% glucose. (A) Deletions of 5 bp (D5) and 15 bp (D15) were performed on the core promoter of P_{NNNN} (control), P_{NCCI} and P_{ICCI}. (B) Analysis of promoter activity in the *E. coli* wild type strain. (C) Analysis of promoter activity in the *E. coli* Δ*ihf* strain. Relative promoter activities were measured after 4h of incubation. Plots were calculated based on the average of three independent experiments. Statistics differences are highlighted by (***) as analyzed using Student's *t* test with p-value $p < 0.0005$; ns, not significant.

particular genes, several studies have now described the systematic investigation of combinatorial promoters through the construction and evaluation of synthetic promoters built from *cis*-regulatory elements. In this sense, Cox III and colleagues constructed a library of synthetic promoters for two local activators (AraC and LuxR) and two local repressors (LacI and TetR) at three different promoter positions (upstream, downstream, or overlapping the core −35/−10 box). From this work, the authors described a number of rules for engineering combinatorial promoters for synthetic biology; for instance, activators were only efficient upstream of the core whereas efficacy of repression was higher at the core and then at the downstream region, with only minor effects at the upstream position.[30] However, this work only used local TFs,

DOI: 10.1021/acssynbio.7b00344
ACS Synth. Biol. 2018, 7, 602−612

Table 1. Strains, Plasmids, and Primers Used in This Study

strains, plasmids, and primers	description	reference
Strains		
E. coli DH10B	*F⁻ endA1 deoR⁺ recA1 galE15 galK16 nupG rpsL* Δ*(lac)X74 φ80lacZ*Δ*M15 araD139* Δ*(ara,leu)7697 mcrA* Δ*(mrr-hsdRMS-mcrBC) Str^R λ⁻*	51
E. coli BW25113	$lacI^{+}rrnB_{T14}$ $\Delta lacZ_{WJ16}$ *hsdR514* $\Delta araBAD_{AH33}$ $\Delta rhaBAD_{LD78}$ *rph-1* Δ*(araB−D)567* Δ*(rhaD−B)568* Δ*lacZ4787(::rrnB-3) hsdR514 rph-1*	52
E. coli JW1702	*E. coli* BW25113 with Δ*ihf* mutation	50
Plasmids		
pMR1	Cm^{R}; orip15a; Promoter probe vector with mCherry and GFPlva reporters	18
pMR1-P_{NNNN}	pMR1 with a reference promoter with four nonregulatory sequences	This study
pMR1-P_{INNN}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 4	This study
pMR1-P_{NINN}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 3	This study
pMR1-P_{NNIN}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 2	This study
pMR1-P_{NNNI}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 1	This study
pMR1-P_{IINN}	pMR1 with a synthetic promoter with IHF *cis*-elements at positions 4 and 3	This study
pMR1-P_{NIIN}	pMR1 with a synthetic promoter with IHF *cis*-elements at positions 3 and 2	This study
pMR1-P_{NNII}	pMR1 with a synthetic promoter with IHF *cis*-elements at positions 2 and 1	This study
pMR1-P_{ININ}	pMR1 with a synthetic promoter with IHF *cis*-elements at positions 4 and 2	This study
pMR1-P_{NINI}	pMR1 with a synthetic promoter with IHF *cis*-elements at positions 3 and 1	This study
pMR1-P_{INNI}	pMR1 with a synthetic promoter with IHF *cis*-elements at positions 4 and 1	This study
pMR1-P_{IIII}	pMR1 with a synthetic promoter with IHF *cis*-elements at positions 4, 3, 2 and 1	This study
pMR1-P_{CNNN}	pMR1 with a synthetic promoter with a CRP *cis*-element at position 4	This study
pMR1-P_{NCNN}	pMR1 with a synthetic promoter with a CRP *cis*-element at position 3	This study
pMR1-P_{NNCN}	pMR1 with a synthetic promoter with a CRP *cis*-element at position 2	This study
pMR1-P_{NNNC}	pMR1 with a synthetic promoter with a CRP *cis*-element at position 1	This study
pMR1-P_{CCNN}	pMR1 with a synthetic promoter with CRP *cis*-elements at positions 4 and 3	This study
pMR1-P_{NCCN}	pMR1 with a synthetic promoter with CRP *cis*-elements at positions 3 and 2	This study
pMR1-P_{NNCC}	pMR1 with a synthetic promoter with CRP *cis*-elements at positions 2 and 1	This study
pMR1-P_{CNCN}	pMR1 with a synthetic promoter with CRP *cis*-elements at positions 4 and 2	This study
pMR1-P_{NCNC}	pMR1 with a synthetic promoter with CRP *cis*-elements at positions 3 and 1	This study
pMR1-P_{CNNC}	pMR1 with a synthetic promoter with CRP *cis*-elements at positions 4 and 1	This study
pMR1-P_{ICNN}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 4 and a CRP *cis*-element at position 3	This study
pMR1-P_{INCN}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 4 and a CRP *cis*-element at position 2	This study
pMR1-P_{ICCN}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 4 and CRP *cis*-elements at positions 3 and 2	This study
pMR1-P_{INCI}	pMR1 with a synthetic promoter with a IHF *cis*-elements at positions 1 and 4 and CRP *cis*-element at position 2	This study
pMR1-P_{ICNI}	pMR1 with a synthetic promoter with a IHF *cis*-elements at positions 1 and 4 and CRP *cis*-element at position 3	This study
pMR1-P_{ICCI}	pMR1 with a synthetic promoter with a IHF *cis*-elements at positions 1 and 4 and CRP *cis*-elements at positions 3 and 2	This study
pMR1-P_{NNCI}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 1 and CRP *cis*-element at position 2	This study
pMR1-P_{NCNI}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 1 and CRP *cis*-element at position 3	This study
pMR1-P_{CNNI}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 1 and CRP *cis*-element at position 4	This study
pMR1-P_{CCNI}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 1 and CRP *cis*-elements at positions 2 and 3	This study
pMR1-P_{NCCI}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 1 and CRP *cis*-elements at position 3 and 2	This study
pMR1-P_{NICC}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 1 and CRP *cis*-element at position 2	This study
pMR1-P_{CINC}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 3 and CRP *cis*-elements at positions 1 and 4	This study
pMR1-P_{NCIC}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 2 and CRP *cis*-elements at positions 1 and 3	This study
pMR1-P_{CNIC}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 2 and CRP *cis*-elements at positions 1 and 2	This study
pMR1-P_{IICC}	pMR1 with a synthetic promoter with a IHF *cis*-element at position 4 and 3 and CRP *cis*-elements at positions 1 and 2	This study
pMR1-$P_{NNNN\text{-}D5}$	Version pMR1-P_{NNNN} of with a 5bp deletion at the 3′ region	This study
pMR1-$P_{NNNN\text{-}D15}$	Version pMR1-P_{NNNN} of with a 15bp deletion at the 3′ region; Removes the −10 region	This study
pMR1-$P_{NCCI\text{-}D5}$	Version pMR1-P_{NCCI} of with a 5bp deletion at the 3′ region	This study
pMR1-$P_{NCCI\text{-}D15}$	Version pMR1-P_{NCCI} of with a 15bp deletion at the 3′ region; Removes the −10 region	This study
pMR1-$P_{ICCI\text{-}D5}$	Version pMR1-P_{ICCI} of with a 5bp deletion at the 3′ region	This study
pMR1-$P_{ICCI\text{-}D15}$	Version pMR1-P_{ICCI} of with a 15bp deletion at the 3′ region; Removes the −10 region	This study
Primers[a]		
P1−C5	AATTCTGTGAGTTAGCTCACA	This study
P1−C3	CGCCTGTGAGCTAACTCACAG	This study
P1−I5	AATTCCAATTTATTGATTTTA	This study
P1−I3	CGCCTAAAATCAATAAATTGG	This study
P1−N5	AATTCTCGCCTGCTTGTAGTA	This study
P1−N3	CGCCTACTACAAGCAGGCGAG	This study
P2−C5	GGCGTGTGAGTTAGCTCACA	This study

DOI: 10.1021/acssynbio.7b00344
ACS Synth. Biol. 2018, 7, 602−612

Table 1. continued

strains, plasmids, and primers	description	reference
P2−C3	GCGGTGTGAGCTAACTCACA	This study
P2−I5	GGCGCAATTTATTGATTTTA	This study
P2−I3	GCGGTAAAATCAATAAATTG	This study
P2−N5	GGCGTCGCCTGCTTGTAGTA	This study
P2−N3	GCGGTACTACAAGCAGGCGA	This study
P3−C5	CCGCTGTGAGTTAGCTCACA	This study
P3−C3	CCAATGTGAGCTAACTCACA	This study
P3−I5	CCGCCAATTTATTGATTTTA	This study
P3−I3	CCAATAAAATCAATAAATTG	This study
P3−N5	CCGCTCGCCTGCTTGTAGTA	This study
P3−N3	CCAATACTACAAGCAGGCGA	This study
P4−C5	TTGGTGTGAGTTAGCTCACA	This study
P4−C3	CAAGTGTGAGCTAACTCACA	This study
P4−I5	TTGGCAATTTATTGATTTTA	This study
P4−I3	CAAGTAAAATCAATAAATTG	This study
P4−N5	TTGGTCGCCTGCTTGTAGTA	This study
P4−N3	CAAGTACTACAAGCAGGCGA	This study
CoreP-5	CTTGAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAG	This study
CoreP-3	GATCCTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCT	This study
pMR1-F	CTCGCCCTTGCTCACC	This study
pMR1-R	ACAAGAATTGGGACAACTCC	This study

[a]Restriction sites are underlined in the primer sequences.

which are limited to a few natural targets and, thus, are not found in naturally complex promoter architectures as global regulators are. Moreover, the work by Cox III only explored a single binding site at the upstream promoter regions, which does not allow the investigation of combinatorial effects generated by *cis*-element arrangements and identities in this region. Therefore, our work addresses a more realistic combinatorial situation by mimicking the manner in which promoters are organized naturally, and indeed, our result of *cis*-element mediated repression of gene expression has not been reported previously. The effect of promoter architecture in gene regulation has also been extensively investigated in single-celled eukaryotes such as yeast, with especial interest in the work of Sharon and co-workers.[44] In this study, the authors synthesized and analyzed using a high-throughput approach, thousands of different promoters for several TFs of *Saccharomyces cerevisiae*,[44] thus allowing them to investigate the effect of number, position, and affinity of binding sites on gene expression. However, the fundamental difference between transcription initiation in prokaryotes and eukaryotes, due to the sophisticated process of chromatin remodeling required in the latter, makes it impossible to extrapolate the conclusions drawn by Sharon *et al.* to a bacterial organism. However, the approach used in this study was analogous to the approach used by Sharon *et al.*, since we could inspect the effect of binding site multiplicity, location, and identity.

From the results generated in this work, the most striking was the observation that a single CRP-binding site located immediately upstream of an IHF-binding site could completely abolish transcriptional activity independently of CRP function. This result appeared in several promoter architectures tested here and would indicate that the DNA sequence itself was modulating gene expression. However, introduction of an additional CRP binding site drastically changed this process, resulting in a CRP-activated promoter. It has now been widely demonstrated that DNA can display an allosteric effect on TFs, where the binding of a protein to DNA changes the way this protein interacts with other TFs.[45−47] Moreover, another type of DNA-based allosteric event has been described where the binding of a protein to DNA can influence the binding of a second protein to an adjacent site independently of protein−protein interaction, and that this influence is transmitted through the DNA molecule.[46,47] In this sense, these processes could explain how two tandem *cis*-elements for CRP that are inactive alone (at positions 3 and 2 in Figure 2) generated a strong CRP-dependent promoter when in association with a single IHF binding site (the latter at position 1, Figure 5B). However, it certainly does not explain how a single CRP *cis*-element displays inhibitory effects in certain promoter architectures (as in many of those presented in Figure 4). Recently, an increasing number of reports have demonstrated that flanking DNA sequences can strongly affect the binding affinity of eukaryotic TFs for identical binding sites,[48,49] thus explaining why *in vitro* and *in vivo* binding assays do not always correlate. In this process, these flanking sequences generate distortions in the local DNA shape that influences the way the TFs interacts with DNA, by altering the groove width and helical parameters of DNA.[48] Though we could not find any report of this process influencing bacterial TFs, our results on synthetic complex promoters suggest that a similar process could influence the activity of bacterial promoters, thus explaining the intrinsic repressive activity of the CRP *cis*-element (independently of the presence of CRP protein) at some positions in promoters containing *cis*-elements for IHF. Our findings could thus be extended to naturally complex promoters and indicate that in those systems, not only would the nature of the TF recruited to the target promoter be imperative for gene expression, but also the *cis*-element itself could have a regulatory role in proximal sites. This evidence an unanticipated intrinsic complexity of natural bacterial promoters that should be considered both for synthetic biology projects as well as to understand the regulatory behavior of natural strains. Taken together, our results highlight the appearance of emergent properties in combinatorial control

DOI: 10.1021/acssynbio.7b00344
ACS Synth. Biol. 2018, 7, 602−612

in bacteria, thus opening new venues for understanding combinatorial regulation in bacterial genes and open new venues that could be investigated in future studies.

MATERIALS AND METHODS

Plasmids, Bacterial Strains, and Growth Conditions. The plasmids, bacterial strains, and primers used in this study are listed in Table 1. For cloning procedures, the bacterial strain used was *E. coli* DH5α. *E. coli* BW25113 was used as the wild type strain (WT) whereas *E. coli* JW1702−1 was used as the mutant for IHF transcription factor, and both were obtained from the Keio collection.[50] *E. coli* strains were grown at 37 °C in LB media with chloramphenicol at 34 μg mL^{-1} or in M9 minimal media (6.4 g L^{-1} $Na_2HPO_4{\cdot}7H_2O$, 1.5 g L^{-1} KH_2PO_4, 0.25 g L^{-1} NaCl, 0.5 g L^{-1} NH_4Cl) supplemented with chloramphenicol at 17 μg mL^{-1}, 2 mM $MgSO_4$, 0.1 mM casamino acids, and 1% glycerol as the sole carbon source. Where indicated, CRP response was depleted by using 0.4% of glucose.

Design of the Minimal Promoter Scaffold and Ligation Reactions. Promoters were constructed by ligation of 5′ end phosphorylated oligonucleotides[27,30] acquired from Sigma-Aldrich (Table 1). All single strand nucleotides were designed to carry a discrete 16 bp sequence[8] containing a CRP binding site (C), IHF binding site (I), one Neutral (N) motif with no transcription factor binding (Figure 1B), and a core promoter based on the *lac* promoter (Table 1), which is a weak promoter and therefore requires activation. All these oligonucleotides were designed to carry three base pair overhangs corresponding to their corrected insertion region on the promoter (Figure 1A). The upper and lower strand corresponding to each position were mixed at equimolar concentrations and annealed by heating at 95 °C followed by gradual cooling to room temperature. External overhangs of the fourth *cis*-element position and the core promoters reassembled on the *Eco*RI and *Bam*HI digested sites, allowing ligation to a previously digested *Eco*RI/*Bam*HI pMR1[18] plasmid. All five fragments (four *cis*-elements positions plus core promoter) were mixed at equimolar concentrations in a pool with the final concentration of 5′ phosphate termini fixed at 15 μM. For the ligase reaction, 1 μL of the pooled fragments was added to 50 ng *Eco*RI/*Bam*HI pMR1 digested plasmid in the presence of ligase buffer and ligase enzyme to a final volume of 10 μL. After 1 h at 16 °C, the ligase reaction was inactivated for 15 min at 65 °C and one aliquot of 2 μL was then electroporated into 50 μL of *E. coli* DH10B competent cells. After 1 h of regenerating in 1 mL LB media, the total volume was plated onto LB solid dishes supplemented with chloramphenicol at 34 μg mL^{-1}. Clones were confirmed by colony PCR using primers pMR1-F and pMR1-R (Table 1) using the pMR1 empty plasmid PCR reaction as a further length reference upon agarose gel electrophoresis. Clones with the potential correct length were submitted to Sanger DNA sequencing for confirming the correct promoter assembly.

GFP(LVA) Fluorescence Assay and Data Processing. To measure promoter activity, the library of 38 promoters was analyzed in different genetic backgrounds and conditions. For each experiment, a plasmid harboring the promoter of interest was used to transform *E. coli* wild type or *E. coli* Δ*ihf* mutant cells. Freshly plated single colonies were grown overnight in LB media, centrifuged, and resuspended in fresh M9 media. The culture (10 μL) was then assayed in 96-well microplates in biological triplicates with 170 μL of M9 media or M9 media supplemented with 0.4% glucose whenever required. Cell growth and GFP(LVA) fluorescence were quantified using a Victor X3 plate reader (PerkinElmer). GFP (LVA) is a GFP with a degradation tag. This degradation tag allows us to evaluate the expression over time in an efficient and robust way. Promoter response was calculated as arbitrary units by dividing the fluorescence levels by the optical density at 600 nm (reported as GFP/OD_{600}) after background correction. The same strain harboring the pMR1 empty plasmid was used as the threshold background signal during calculations. Fluorescence and absorbance measurements were taken at 30 min intervals over 8 h (However, for Figures 4, 5, 6 and 7 just the 4 h promoter activity (GFP/OD_{600}) was shown). Technical triplicates and biological triplicates were included in all experiments. Raw data were processed using *ad hoc* R script (https://www.r-project.org/) and plots were constructed using R or MeV (www.tm4.org/mev.html).

AUTHOR INFORMATION

Corresponding Author

*Tel.: +55 16 3315 3229. Fax: +55 16 3633 0728. E-mail: silvarochar@gmail.com.

ORCID

Rafael Silva-Rocha: 0000-0001-6319-631X

Author Contributions

†L.M.O.M. and L.M.A. contributed equally to this work

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors are thankful to lab members and to Dr. Maria-Eugenia Guazzaroni for critical discussion of this work. This work was supported by the Young Investigator Award of the Sao Paulo State Foundation (FAPESP, award number 2012/22921-8). LMOM was supported by a FAPESP Ph.D. fellowship (FAPESP, award number 2016/19179-9).

REFERENCES

(1) Browning, D. F., and Busby, S. J. W. (2004) The regulation of bacterial transcription initiation,. *Nat. Rev. Microbiol. 2*, 57−65.

(2) Prigent-Combaret, C., Brombacher, E., Vidal, O., Ambert, A., Lejeune, P., Landini, P., and Dorel, C. (2001) Complex regulatory network controls initial adhesion and biofilm formation in Escherichia coli via regulation of the csgD gene. *J. Bacteriol. 183*, 7213−7223.

(3) Vicente, M., Chater, K. F., and De Lorenzo, V. (1999) Bacterial transcription factors involved in global regulation. *Mol. Microbiol. 33*, 8−17.

(4) Liu, X., and Matsumura, P. (1994) The FlhD/FlhC complex, a transcriptional activator of the Escherichia coli flagellar class II operons,. *J. Bacteriol. 176*, 7345−7351.

(5) Collado-Vides, J., Magasanik, B., and Gralla, J. D. (1991) Control site location and transcriptional regulation in Escherichia coli. *Microbiol Rev. 55*, 371−394.

(6) De Lorenzo, V., Herrero, M., Giovannini, F., and Neilands, J. B. (1988) Fur (ferric uptake regulation) protein and CAP (catabolite-activator protein) modulate transcription of fur gene in Escherichia coli. *Eur. J. Biochem. 173*, 537−546.

(7) Miyada, C. G., Stoltzfus, L., and Wilcox, G. (1984) Regulation of the araC gene of Escherichia coli: catabolite repression, autoregulation, and effect on araBAD expression. *Proc. Natl. Acad. Sci. U. S. A. 81*, 4120−4124.

(8) Little, J. W., Edmiston, S. H., Pacelli, L. Z., and Mount, D. W. (1980) Cleavage of the Escherichia coli lexA protein by the recA protease. *Proc. Natl. Acad. Sci. U. S. A. 77*, 3225−3229.

DOI: 10.1021/acssynbio.7b00344
ACS Synth. Biol. 2018, 7, 602−612

(9) Hermsen, R., Ursem, B., and ten Wolde, P. R. (2010) Combinatorial gene regulation using auto-regulation. *PLoS Comput. Biol. 6*, e1000813.
(10) Buchler, N. E., Gerland, U., and Hwa, T. (2003) On schemes of combinatorial transcription logic. *Proc. Natl. Acad. Sci. U. S. A. 100*, 5136−5141.
(11) Martinez-Antonio, A., and Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol. 6*, 482−489.
(12) Pérez-Rueda, E., and Collado-Vides, J. (2000) The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12,. *Nucleic Acids Res. 28*, 1838−1847.
(13) Inada, T., Takahashi, H., Mizuno, T., and Aiba, H. (1996) Down regulation of cAMP production by cAMP receptor protein in Escherichia coli: an assessment of the contributions of transcriptional and posttranscriptional control of adenylate cyclase. *Mol. Gen. Genet. 253*, 198−204.
(14) Schmitz, A. (1981) Cyclic AMP receptor protein interacts with lactose operator DNA. *Nucleic Acids Res. 9*, 277−292.
(15) You, C., Okano, H., Hui, S., Zhang, Z., Kim, M., Gunderson, C. W., Wang, Y. P., Lenz, P., Yan, D., and Hwa, T. (2013) Coordination of bacterial proteome with metabolism by cyclic AMP signalling. *Nature 500*, 301−306.
(16) Azam, T. A., and Ishihama, A. (1999) Twelve species of the nucleoid-associated protein from Escherichia coli. Sequence recognition specificity and DNA binding affinity. *J. Biol. Chem. 274*, 33105−33113.
(17) Biek, D. P., and Cohen, S. N. (1989) Involvement of integration host factor (IHF) in maintenance of plasmid pSC101 in Escherichia coli: mutations in the topA gene allow pSC101 replication in the absence of IHF. *J. Bacteriol. 171*, 2066−2074.
(18) Guazzaroni, M. E., and Silva-Rocha, R. (2014) Expanding the logic of bacterial promoters using engineered overlapping operators for global regulators. *ACS Synth. Biol. 3*, 666−675.
(19) Lee, D. J., Minchin, S. D., and Busby, S. J. W. (2012) Activating Transcription in Bacteria. *Annu. Rev. Microbiol. 66*, 125−152.
(20) Nandagopal, N., and Elowitz, M. B. (2011) Synthetic biology: integrated gene circuits. *Science 333*, 1244−1248.
(21) Voigt, C. A. (2006) Genetic parts to program bacteria. *Curr. Opin. Biotechnol. 17*, 548−557.
(22) Benner, S. a., and Sismour, a. M. (2005) Synthetic biology. *Nat. Rev. Genet. 6*, 533−543.
(23) Brophy, J. A., and Voigt, C. A. (2014) Principles of genetic circuit design. *Nat. Methods 11*, 508−520.
(24) Rhodius, V. A., Segall-Shapiro, T. H., Sharon, B. D., Ghodasara, A., Orlova, E., Tabakh, H., Burkhardt, D. H., Clancy, K., Peterson, T. C., Gross, C. A., and Voigt, C. A. (2013) Design of orthogonal genetic switches based on a crosstalk map of sigmas, anti-sigmas, and promoters. *Mol. Syst. Biol. 9*, 702.
(25) Guet, C. C., Elowitz, M. B., Hsing, W., and Leibler, S. (2002) Combinatorial synthesis of genetic networks. *Science 296*, 1466−1470.
(26) Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000) Construction of a genetic toggle switch in Escherichia coli. *Nature 403*, 339−342.
(27) Kinkhabwala, A., and Guet, C. C. (2008) Uncovering cis regulatory codes using synthetic promoter shuffling. *PLoS One 3*, e2030.
(28) Isalan, M., Lemerle, C., Michalodimitrakis, K., Horn, C., Beltrao, P., Raineri, E., Garriga-Canut, M., and Serrano, L. (2008) Evolvability and hierarchy in rewired bacterial gene networks. *Nature 452*, 840−845.
(29) Murphy, K. F., Balázsi, G., and Collins, J. J. (2007) Combinatorial promoter design for engineering noisy gene expression. *Proc. Natl. Acad. Sci. U. S. A. 104*, 12726−12731.
(30) Cox, R. S., 3rd, Surette, M. G., and Elowitz, M. B. (2007) Programming gene expression with combinatorial promoters. *Mol. Syst. Biol. 3*, 145.
(31) Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005) Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev. 15*, 116−124.
(32) Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., Kuhlman, T., and Phillips, R. (2005) Transcriptional regulation by the numbers: applications. *Curr. Opin. Genet. Dev. 15*, 125−135.
(33) Bhalla, U. S., and Iyengar, R. (1999) Emergent properties of networks of biological signaling pathways. *Science 283*, 381−387.
(34) Yamada, M., Asaoka, S., Saier, M. H., Jr, and Yamada, Y. (1993) Characterization of the gcd gene from Escherichia coli K-12 W3110 and regulation of its expression. *J. Bacteriol. 175*, 568−571.
(35) Izu, H., Ito, S., Elias, M. D., and Yamada, M. (2002) Differential control by IHF and cAMP of two oppositely oriented genes, hpt and gcd, in Escherichia coli: significance of their partially overlapping regulatory elements. *Mol. Genet. Genomics 266*, 865−872.
(36) Kumari, S., Beatty, C. M., Browning, D. F., Busby, S. J., Simel, E. J., Hovel-Miner, G., and Wolfe, A. J. (2000) Regulation of acetyl coenzyme A synthetase in Escherichia coli. *J. Bacteriol. 182*, 4173−4179.
(37) Tebbutt, J., Rhodius, V. A., Webster, C. L., and Busby, S. J. W. (2002) Architectural requirements for optimal activation by tandem CRP molecules at a class I CRP-dependent promoter. *FEMS Microbiol. Lett. 210*, 55−60.
(38) Ushida, C., and Aiba, H. (1990) Helical phase dependent action of CRP: effect of the distance between the CRP site and the −35 region on promoter activity. *Nucleic Acids Res. 18*, 6325−6330.
(39) Giladi, H., Murakami, K., Ishihama, a., and Oppenheim, a. B. (1996) Identification of an UP element within the IHF binding site at the PL1-PL2 tandem promoter of bacteriophage lambda. *J. Mol. Biol. 260*, 484−491.
(40) Rossiter, A. E., Godfrey, R. E., Connolly, J. A., Busby, S. J. W., Henderson, I. R., and Browning, D. F. (2015) Expression of different bacterial cytotoxins is controlled by two global transcription factors, CRP and Fis, that co-operate in a shared-recruitment mechanism. *Biochem. J. 466*, 323−335.
(41) Amores, G. R., Guazzaroni, M. E., and Silva-Rocha, R. (2015) Engineering Synthetic cis-Regulatory Elements for Simultaneous Recognition of Three Transcriptional Factors in Bacteria. *ACS Synth. Biol. 4*, 1287−1294.
(42) Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muniz-Rascado, L., Garcia-Sotelo, J. S., Weiss, V., Solano-Lira, H., Martinez-Flores, I., Medina-Rivera, A., Salgado-Osorio, G., Alquicira-Hernandez, S., Alquicira-Hernandez, K., Lopez-Fuentes, A., Porron-Sotelo, L., Huerta, A. M., Bonavides-Martinez, C., Balderas-Martinez, Y. I., Pannier, L., Olvera, M., Labastida, A., Jimenez-Jacinto, V., Vega-Alvarado, L., Del Moral-Chavez, V., Hernandez-Alvarez, A., Morett, E., and Collado-Vides, J. (2013) RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res. 41*, D203−213.
(43) Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., and Segal, E. (2012) Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nat. Genet. 44*, 743−750.
(44) Sharon, E., Kalma, Y., Sharp, A., Raveh-Sadka, T., Levo, M., Zeevi, D., Keren, L., Yakhini, Z., Weinberger, A., and Segal, E. (2012) Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol. 30*, 521−530.
(45) Kim, S., Brostromer, E., Xing, D., Jin, J., Chong, S., Ge, H., Wang, S., Gu, C., Yang, L., Gao, Y. Q., Su, X. D., Sun, Y., and Xie, X. S. (2013) Probing allostery through DNA. *Science 339*, 816−819.
(46) Lefstin, J. A., and Yamamoto, K. R. (1998) Allosteric effects of DNA on transcriptional regulators. *Nature 392*, 885−888.
(47) Chaires, J. B. (2008) Allostery: DNA does it, too. *ACS Chem. Biol. 3*, 207−209.
(48) Gordan, R., Shen, N., Dror, I., Zhou, T., Horton, J., Rohs, R., and Bulyk, M. L. (2013) Genomic regions flanking E-box binding sites

DOI: 10.1021/acssynbio.7b00344
ACS Synth. Biol. 2018, 7, 602−612

influence DNA binding specificity of bHLH transcription factors through DNA shape. *Cell Rep. 3*, 1093−1104.
(49) Khoueiry, P., Rothbacher, U., Ohtsuka, Y., Daian, F., Frangulian, E., Roure, A., Dubchak, I., and Lemaire, P. (2010) A cis-regulatory signature in ascidians and flies, independent of transcription factor binding sites,. *Curr. Biol. 20*, 792−802.
(50) Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol. 2*, 2006 0008.
(51) Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York.
(52) Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A. 97*, 6640−6645.

DOI: 10.1021/acssynbio.7b00344
ACS Synth. Biol. 2018, 7, 602−612

CHAPTER II

Unraveling the Complex Interplay of Fis and IHF Through Synthetic Promoter Engineering

This chapter was published as:

(Monteiro et al., 2020) MONTEIRO, Lummy Maria Oliveira; SANCHES-MEDEIROS, Ananda; WESTMANN, Caua Antunes; & SILVA-ROCHA, Rafael. Unraveling the Complex Interplay of Fis and IHF Through Synthetic Promoter Engineering. **Frontiers in bioengineering and biotechnology**, v. 8, p. 1-13, 2020.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 18 June 2020
doi: 10.3389/fbioe.2020.00510

Unraveling the Complex Interplay of Fis and IHF Through Synthetic Promoter Engineering

Lummy Maria Oliveira Monteiro, Ananda Sanches-Medeiros, Cauã Antunes Westmann and Rafael Silva-Rocha**

Ribeirão Preto Medical School (FMRP), University of São Paulo, Ribeirão Preto, Brazil

Bacterial promoters are usually formed by multiple *cis*-regulatory elements recognized by a plethora of transcriptional factors (TFs). From those, global regulators are key elements since these TFs are responsible for the regulation of hundreds of genes in the bacterial genome. For instance, Fis and IHF are global regulators that play a major role in gene expression control in *Escherichia coli*, and usually, multiple *cis*-regulatory elements for these proteins are present at target promoters. Here, we investigated the relationship between the architecture of the *cis*-regulatory elements for Fis and IHF in *E. coli*. For this, we analyze 42 synthetic promoter variants harboring consensus *cis*-elements for Fis and IHF at different distances from the core −35/−10 region and in various numbers and combinations. We first demonstrated that although Fis preferentially recognizes its consensus *cis*-element, it can also recognize, to some extent, the consensus-binding site for IHF, and the same was true for IHF, which was also able to recognize Fis binding sites. However, changing the arrangement of the *cis*-elements (i.e., the position or number of sites) can completely abolish the non-specific binding of both TFs. More remarkably, we demonstrated that combining *cis*-elements for both TFs could result in Fis and IHF repressed or activated promoters depending on the final architecture of the promoters in an unpredictable way. Taken together, the data presented here demonstrate how small changes in the architecture of bacterial promoters could result in drastic changes in the final regulatory logic of the system, with important implications for the understanding of natural complex promoters in bacteria and their engineering for novel applications.

Keywords: regulatory network, *cis*-regulatory elements, complex promoters, global regulators, transcriptional crosstalk, fine-tuning

OPEN ACCESS

Edited by:
Tao Chen,
Tianjin University, China

Reviewed by:
Long Liu,
Jiangnan University, China
Andrew Cameron,
University of Rochester, United States
Kaneyoshi Yamamoto,
Hosei University, Japan

****Correspondence:***
Lummy Maria Oliveira Monteiro
lummymaria@gmail.com
Rafael Silva-Rocha
silvarochar@usp.br

Specialty section:
This article was submitted to Synthetic Biology, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *02 March 2020*
Accepted: *30 April 2020*
Published: *18 June 2020*

Citation:
Monteiro LMO, Sanches-Medeiros A, Westmann CA and Silva-Rocha R (2020) Unraveling the Complex Interplay of Fis and IHF Through Synthetic Promoter Engineering. Front. Bioeng. Biotechnol. 8:510. doi: 10.3389/fbioe.2020.00510

INTRODUCTION

Bacteria have evolved complex gene regulatory networks to coordinate the expression level of each gene in response to changing environmental conditions. In this aspect, a typical bacterium such as *Escherichia coli* uses around 300 different transcriptional factors (TFs) to control the expression of more than 5,000 genes, and gene regulation in bacteria has been extensively investigated in the last six decades (Lozada-Chavez, 2006). Among the known TFs from *E. coli*, global regulators are able to control the highest percentage of transcriptional units in response to significant physiological or environmental signals, such as the metabolic state of the cell, the availability of carbon sources, and the presence of oxygen (Martínez-Antonio et al., 2003; Ishihama, 2010), while local regulators are

responsible for gene regulation in response to specific signals (such as sugars and metals) (Ishihama, 2010; Browning and Busby, 2016). Most TFs control gene expression through their interaction with specific DNA sequences located near the promoter region, the *cis*-regulatory element, or transcriptional factor binding site (Browning and Busby, 2004, 2016). Over the decades, many *cis*-regulatory elements for many TFs from *E. coli* have been experimentally characterized, mapped, and compiled in databases such as RegulonDB and EcoCyc (Gama-Castro et al., 2016; Keseler et al., 2017). Analysis of these datasets demonstrates that TFs usually act in a combinatorial way to control gene expression, where multiple *cis*-regulatory elements for different TFs are located in the upstream region of the target genes (Guazzaroni and Silva-Rocha, 2014; Rydenfelt et al., 2014; Gama-Castro et al., 2016). Therefore, the arrangement of *cis*-regulatory elements at the target promoters is crucial to determine which TFs will be able to control the target gene and how these regulators interact with each other once bound to the DNA (Collado-Vides et al., 1991; Ishihama, 2010).

Several studies have explored the relationship between the architecture of *cis*-regulatory elements and the final logic of the target promoters, and initial attempts have focused on the mutation of *cis*-regulatory elements from natural promoters to investigate how these elements specify the promoter activity dynamics (Sawers, 1993; Darwin and Stewart, 1995; Izu et al., 2002; Setty et al., 2003). More recently, synthetic biology approaches have been used to construct artificial promoters through the combination of several *cis*-regulatory elements, and these have been characterized to decipher their architecture/dynamics relationship (Cox et al., 2007; Isalan et al., 2008; Kinkhabwala and Guet, 2008; Shis et al., 2014). However, while most synthetic biology approaches have focused on *cis*-elements for local regulators (which do not commonly regulate gene expression in a combinatorial manner), we recently investigated this combinatorial regulation problem with global regulators (Guazzaroni and Silva-Rocha, 2014; Amores et al., 2015; Monteiro et al., 2018). This is important because global regulators (such as IHF, Fis, and CRP) have numerous binding sites along the *E. coli* genome and frequently co-occur at target promoters (Guazzaroni and Silva-Rocha, 2014). Thus, Fis and IHF are two global regulators that play a critical role in coordinating gene expression in *E. coli* as well as in mediating DNA condensation in the cell (Azam and Ishihama, 1999; Browning and Busby, 2004; Browning et al., 2010; Ishihama, 2010). Fis, an abundant nucleoid-associated protein (NAP), is related to gene expression regulation in fast-growing cells, varying its function (as a repressor or activator transcriptional factor) according to its biding site position related to the core promoter (Hirvonen et al., 2001), while IHF is a NAP, which activity relates to changes in gene expression in cells during the transition from exponential to stationary phase (Azam and Ishihama, 1999; Azam et al., 2000; Browning et al., 2010). Moreover, IHF binds to AT-rich DNA motifs with well-defined sequence preferences, while Fis also prefers AT-rich regions with a more degenerate sequence preference (Déthiollaz et al., 1996; Ussery et al., 2001; Dorman and Deighan, 2003; Aeling et al., 2006). Additionally, cross-regulation between Fis and IHF has been demonstrated for several systems, and how specific vs. promiscuous DNA recognition can be achieved for these two global regulators is not fully understood (Browning et al., 2010; Ishihama, 2010; Rossiter et al., 2015).

We previously explored how complex synthetic promoters harboring *cis*-regulatory elements for CRP and IHF can generate diverse regulatory logic depending on the final architecture of synthetic promoters, demonstrating that it is not possible to predict the regulatory logic of complex multiple promoters from the known dynamics of their simple versions (Monteiro et al., 2018). Here, we further explore this approach to investigate the relationship between *cis*-regulatory elements for Fis and IHF. Using consensus binding sites for these 2 TFs at different promoter positions and in different numbers, we first demonstrated that while some promiscuous interactions occur between the TFs and the binding sites, some specific *cis*-regulatory architectures can completely abolish non-specific interactions. Additionally, complex promoters constructed by the combination of *cis*-elements for Fis and IHF can generate many completely different outputs, such as Fis-repressed promoters, IHF-repressed promoters, and systems where Fis and IHF act as activators. As these changes in promoter logic result from changes in promoter architecture only (and not on the affinity of the transcriptional factor to each individual *cis-element*), the data presented here reinforce the notion that complex bacterial promoters can display emergent properties, where their final behavior cannot be defined from the characterization of the individual component. Taken together, our findings present a comprehensive strategy for fine-tuning gene circuits to perform optimally in a given context (e.g., engineering of synthetic promoters) as well as provide insights for the understanding of natural complex promoters controlled by global regulators.

RESULTS AND DISCUSSION

Generation of Complex Promoters for Fis and IHF

In order to investigate the effect of promoter architecture in the regulation by Fis and IHF, we evaluated the effect of 12 complex promoters constructed in early work (Monteiro et al., 2018) and we constructed 30 new combinatorial promoters with consensus DNA sequences for Fis (Fis-BS) and IHF (IHF-BS) binding sites positioned upstream of a weak core promoter (−35/−10 region) at specific positions (1–4) centered at the −61, −81, −101, and −121 regions related to the transcriptional start site (TSS) (**Figure 1**). For that, we generated double-strand DNA sequences for Fis-BS, IHF-BS, and a neutral sequence (Neg) with no related transcriptional binding site, which were combined for the generation of a library of synthetic promoters, merging the transcriptional binding sites for Fis, IHF, and/or neutral sequence for each position (**Table 1**). The complex promoters were assembled by DNA ligation and cloned into pMR1, a mid-copy number vector harboring mCherry and GFPlva as reporter fluorescent proteins (**Figure 1**). The resulting reporter plasmids (with each promoter controlling only by GFPlva expression) were used to transform competent *E. coli* wild-type strain

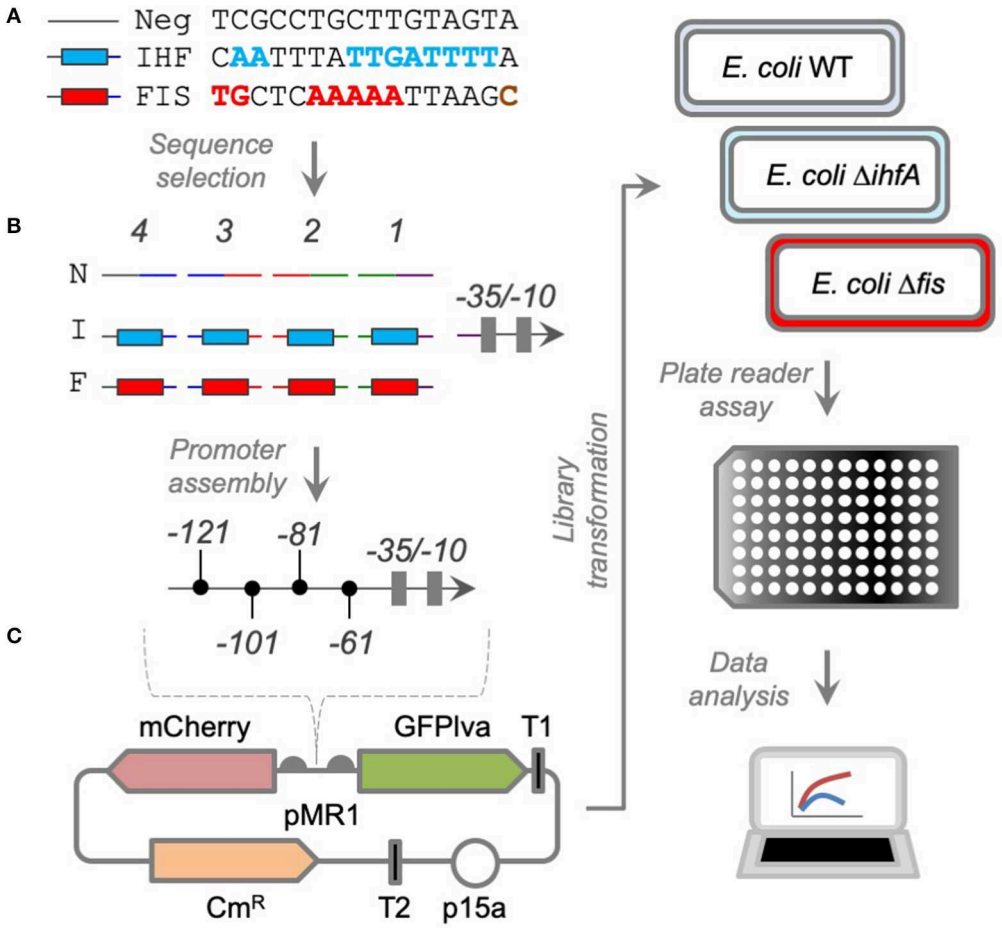


FIGURE 1 | Strategies to construct synthetic complex promoters. **(A)** DNA sequences harboring the consensus sequence for IHF or Fis binding were selected, along with a control sequence that cannot be recognized by any TF. **(B)** Double-stranded DNA fragments were produced with cohesive ends specific for each promoter position (numbered from 1 to 4) and assembled together with a weak core promoter harboring the −35/−10 boxes for RNAP recognition (Guazzaroni and Silva-Rocha, 2014). **(C)** The fragments were cloned into a promoter probe vector (pMR1) harboring resistance to chloramphenicol (CmR), a medium-copy number origin of replication (p15a), and two reporter genes (mCherry and GFPlva). The libraries were introduced into wild-type and mutant strains of E. coli from the KEIO collection (Baba et al., 2006). The resultant strains were analyzed at the population level in a plate reader and the data processed using script in R.

(BW25113—WT) and/or *E. coli* mutants for *ihfA* (Δ*ihfA*) and *fis* (Δ*fis*) (from Keio collection) (Baba et al., 2006). Using these constructs, we assayed promoter activity for 8 h in minimal media (M9 complete), measuring the relative GFP expression (GFP/OD) in all strains in the plate reader fluorimeter Victor X3 (PerkinElmer). As a negative control, we used the Neg sequence occupying the 4 possible positions before the core promoter. All data presented in this work are referred to 4 h of cell growth. In the next sections, we present the results of the promoter analysis per category to uncover the *cis*-regulatory logic for each variant.

Changing the Fis Binding Site Architecture Modulates Fis and IHF Binding Specificity

We analyzed the architecture effect for Fis *cis*-regulatory elements by evaluating the influence of position and sequence combination for Fis-BS. For that, we used promoters merging Fis-BS and Neg sequences to measure relative GFP expression (GFP/OD) levels after 4 h of cell growth in wild-type, Δ*fis*, and Δ*ihfA* *E. coli* strains, and normalized the results to our negative control (top bars in **Figure 2**). The results displayed in **Figure 2** show that most of the promoters harboring Fis-BS exhibit low activity in wild-type *E. coli*, comparable to the negative control. However, when these promoters were assayed in *E. coli* Δ*fis* strain (red bars), 4 of them displayed a significant increase in activity compared to the wild-type strain (green and gray in **Figure 2**). Particularly, in the presence of Fis protein, Fis could occupy Fis-BS and act as a repressor of promoter activity. However, not all architectures with Fis-BS at the 4th or 3rd positions display this promoter behavior. This phenomenon only occurs in two other cases with more than 1 Fis-BS combination (promoters shaded in green in **Figure 2**). This

TABLE 1 | Strains, plasmids, and primers used in this study.

Strains, plasmids, and primers	Description	References
Strains		
E. coli DH10B	*F⁻ endA1 deoR⁺ recA1 galE15 galK16 nupG rpsL Δ(lac)X74 ϕ80lacZΔM15 araD139 Δ(ara,leu)7697 mcrA Δ(mrr-hsdRMS-mcrBC) Str^R λ⁻*	Sambrook et al., 1989
E. coli BW25113	*lacI+rrnBT14 ΔlacZWJ16 hsdR514 ΔaraBADAH33 ΔrhaBADLD78 rph-1 Δ(araB–D)567 Δ(rhaD–B)568 ΔlacZ4787(::rrnB-3) hsdR514 rph-1*	Datsenko and Wanner, 2000
E. coli JW1702	*E. coli* BW25113 with Δ*ihfA* mutation	Baba et al., 2006
E. coli JW3229	*E. coli* BW25113 with Δ*fis* mutation	Baba et al., 2006
Plasmids		
pMR1	Cm^R; orip15a; Promoter probe vector with mCherry and GFPlva reporters	Guazzaroni and Silva-Rocha, 2014
pMR1-NNNN	pMR1 with a reference promoter with 4 non-regulatory sequences	Monteiro et al., 2018
pMR1-FNNN	pMR1 with a synthetic promoter with a Fis *cis*-elements at position 4	This study
pMR1-NFNN	pMR1 with a synthetic promoter with a Fis *cis*-elements at position 3	This study
pMR1-NNFN	pMR1 with a synthetic promoter with a Fis *cis*-elements at position 2	This study
pMR1-NNNF	pMR1 with a synthetic promoter with a Fis *cis*-elements at position 1	This study
pMR1-NNFF	pMR1 with a synthetic promoter with Fis *cis*-elements at positions 2 and 1	This study
pMR1-FNNF	pMR1 with a synthetic promoter with Fis *cis*-elements at positions 4 and 1	This study
pMR1-FFNN	pMR1 with a synthetic promoter with Fis *cis*-elements at positions 4 and 3	This study
pMR1-NFFN	pMR1 with a synthetic promoter with Fis *cis*-elements at positions 3 and 2	This study
pMR1-NFNF	pMR1 with a synthetic promoter with Fis *cis*-elements at positions 3 and 1	This study
pMR1-FNFN	pMR1 with a synthetic promoter with Fis *cis*-elements at positions 4 and 2	This study
pMR1-FFNF	pMR1 with a synthetic promoter with Fis *cis*-elements at positions 4, 3 and 1	This study
pMR1-FNFF	pMR1 with a synthetic promoter with Fis *cis*-elements at positions 4, 2 and 1	This study
pMR1-NFFF	pMR1 with a synthetic promoter with Fis *cis*-elements at positions 3, 2 and 1	This study
pMR1-FFFN	pMR1 with a synthetic promoter with Fis *cis*-elements at positions 4, 3 and 2	This study
pMR1-FFFF	pMR1 with a synthetic promoter with Fis *cis*-elements at positions 4, 3, 2 and 1	This study
pMR1-INNN	pMR1 with a synthetic promoter with a IHF cis-element at position 4	Monteiro et al., 2018
pMR1-NINN	pMR1 with a synthetic promoter with a IHF cis-element at position 3	Monteiro et al., 2018
pMR1-NNIN	pMR1 with a synthetic promoter with a IHF cis-element at position 2	Monteiro et al., 2018
pMR1-NNNI	pMR1 with a synthetic promoter with a IHF cis-element at position 1	Monteiro et al., 2018
pMR1-IINN	pMR1 with a synthetic promoter with IHF cis-elements at positions 4 and 3	Monteiro et al., 2018
pMR1-NIIN	pMR1 with a synthetic promoter with IHF cis-elements at positions 3 and 2	Monteiro et al., 2018
pMR1-NNII	pMR1 with a synthetic promoter with IHF cis-elements at positions 2 and 1	Monteiro et al., 2018
pMR1-ININ	pMR1 with a synthetic promoter with IHF cis-elements at positions 4 and 2	Monteiro et al., 2018
pMR1-NINI	pMR1 with a synthetic promoter with IHF cis-elements at positions 3 and 1	Monteiro et al., 2018
pMR1-INNI	pMR1 with a synthetic promoter with IHF cis-elements at positions 4 and 1	Monteiro et al., 2018
pMR1-IIII	pMR1 with a synthetic promoter with IHF cis-elements at positions 4, 3, 2 and 1	Monteiro et al., 2018
pMR1-FNNI	pMR1 with a synthetic promoter with a IHF *cis*-element at position 1 and Fis *cis*- element at position 4	This study
pMR1-NFNI	pMR1 with a synthetic promoter with a IHF *cis*-element at position 1 and Fis *cis*- element at position 3	This study
pMR1-NNFI	pMR1 with a synthetic promoter with a IHF *cis*-element at position 1 and Fis *cis*- element at position 2	This study
pMR1-NFFI	pMR1 with a synthetic promoter with a IHF *cis*-element at position 1 and Fis *cis*- elements at positions 3 and 2	This study
pMR1-FFNI	pMR1 with a synthetic promoter with a IHF *cis*-element at position 1 and Fis *cis*- elements at positions 4 and 3	This study
pMR1-IFNN	pMR1 with a synthetic promoter with a IHF *cis*-element at position 4 and Fis *cis*- element at position 3	This study
pMR1-INFN	pMR1 with a synthetic promoter with a IHF *cis*-element at position 4 and Fis *cis*- element at position 2	This study
pMR1-INNF	pMR1 with a synthetic promoter with a IHF *cis*-element at position 4 and Fis *cis*- element at position 1	This study
pMR1-IFFN	pMR1 with a synthetic promoter with a IHF *cis*-element at position 4 and Fis *cis*- elements at positions 3 and 2	This study
pMR1-IFNF	pMR1 with a synthetic promoter with a IHF *cis*-element at position 4 and Fis *cis*- elements at positions 3 and 1	This study
pMR1-INFF	pMR1 with a synthetic promoter with a IHF *cis*-element at position 4 and Fis *cis*- elements at positions 2 and 1	This study
pMR1-IFFF	pMR1 with a synthetic promoter with a IHF *cis*-element at position 4 and Fis *cis*- elements at positions 3, 2 and 1	This study
pMR1-IFFI	pMR1 with a synthetic promoter with a IHF *cis*-elements at positions 4 and 1. Fis *cis*- elements at positions 3 and 2	This study

(Continued)

TABLE 1 | Continued

Strains, plasmids, and primers	Description	References
pMR1-IFNI	pMR1 with a synthetic promoter with a IHF *cis*-elements at positions 4 and 1. Fis *cis*- element at position 3	This study
pMR1-INFI	pMR1 with a synthetic promoter with a IHF *cis*-elements at positions 4 and 1. Fis *cis*- element at position 2	This study
Primers		
P1-N5	AATTCTCGCCTGCTTGTAGTA*	Monteiro et al., 2018
P1-N3	CGCCTACTACAAGCAGGCGAG	Monteiro et al., 2018
P2-N5	GGCGTCGCCTGCTTGTAGTA	Monteiro et al., 2018
P2-N3	GCGGTACTACAAGCAGGCGA	Monteiro et al., 2018
P3-N5	CCGCTCGCCTGCTTGTAGTA	Monteiro et al., 2018
P3-N3	CCAATACTACAAGCAGGCGA	Monteiro et al., 2018
P4-N5	TTGGTCGCCTGCTTGTAGTA	Monteiro et al., 2018
P4-N3	CAAGTACTACAAGCAGGCGA	Monteiro et al., 2018
P1-I5	AATTCCAATTTATTGATTTTA*	Monteiro et al., 2018
P1-I3	CGCCTAAAATCAATAAATTGG	Monteiro et al., 2018
P4-I5	TTGGCAATTTATTGATTTTA	Monteiro et al., 2018
P4-I3	CAAGTAAAATCAATAAATTG	Monteiro et al., 2018
P1-F5	AATTCTGCTCAAAAATTAAGC*	This study
P1-F3	CGCCGCTTAATTTTTGAGCAG	This study
P2-F5	GGCGTGCTCAAAAATTAAGC	This study
P2-F3	GCGGGCTTAATTTTTGAGCA	This study
P3-F5	CCGCTGCTCAAAAATTAAGC	This study
P3-F3	CCAAGCTTAATTTTTGAGCA	This study
P4-F5	TTGGTGCTCAAAAATTAAGC	This study
P4-F3	CAAGGCTTAATTTTTGAGCA	This study
CoreP-5	CTTGAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAG	Monteiro et al., 2018
CoreP-3	GATCCTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCT*	Monteiro et al., 2018
pMR1-F	CTCGCCCTTGCTCACC	Monteiro et al., 2018
pMR1-R	ACAAGAATTGGGACAACTCC	Monteiro et al., 2018

**Restriction sites are underlined in the primer sequences.*

reveals a complex association between promoter architecture and expression profile, which seems to be dependent on the Fis-BS position and arrangement.

We also assayed Fis-BS promoters in the *E. coli* Δ*ihfA* strain (blue bars) to evaluate the specificity of Fis for Fis-BS. Strikingly, despite most promoters display similar activity levels in the Δ*ihfA* strain as in the wild-type, 1 single promoter variant harboring Fis-BS at the 3rd position (−101 relative to the TSS) displayed a substantial increase in activity in the *ihfA* mutant relative to the wild-type strain (promoter shaded in gray in **Figure 2**). This result indicates that IHF also acts as a repressor of this promoter variant. Although it was restricted to a single promoter variant, these results suggest that non-specific IHF binding to the Fis-BS exists, suggesting that promiscuous regulatory interaction could occur and seems to be dependent on promoter architecture, since this phenomenon is detected only for Fis-BS at the 3rd position. Altogether, these results suggest a complex interplay between the position and combination of Fis-BS and the regulation of gene expression.

IHF Binding Sites Can Be Recognized by the Fis Regulator in an Architecture-Dependent Manner

Using the same strategy as in **Figure 2**, we investigated the regulatory logic of promoters harboring multiple *cis*-regulatory elements for IHF, merging IHF-BS, and Neg sequences. **Figure 3** shows that most promoters displayed low activity in the wild type strain of *E. coli* and higher activity in *E. coli* Δ*ihfA* (blue bars), in agreement with previous data on complex IHF promoters (green and gray shaded) (Monteiro et al., 2018). However, when these promoters were assayed in *E. coli* Δ*fis* strain (red bars), we observed that 4 promoter architectures also displayed higher activity in this mutant (promoters shaded in green in the figure), indicating that Fis was also able to repress these promoter variants, highlighting a possible crosstalk (Cepeda-Humerez et al., 2015; Friedlander et al., 2016) between these 2 TFs, which should be further investigated in the future. However, it is worth noticing that the promoter variants harboring *cis*-regulatory elements for IHF at 4th or 3rd and 4th positions

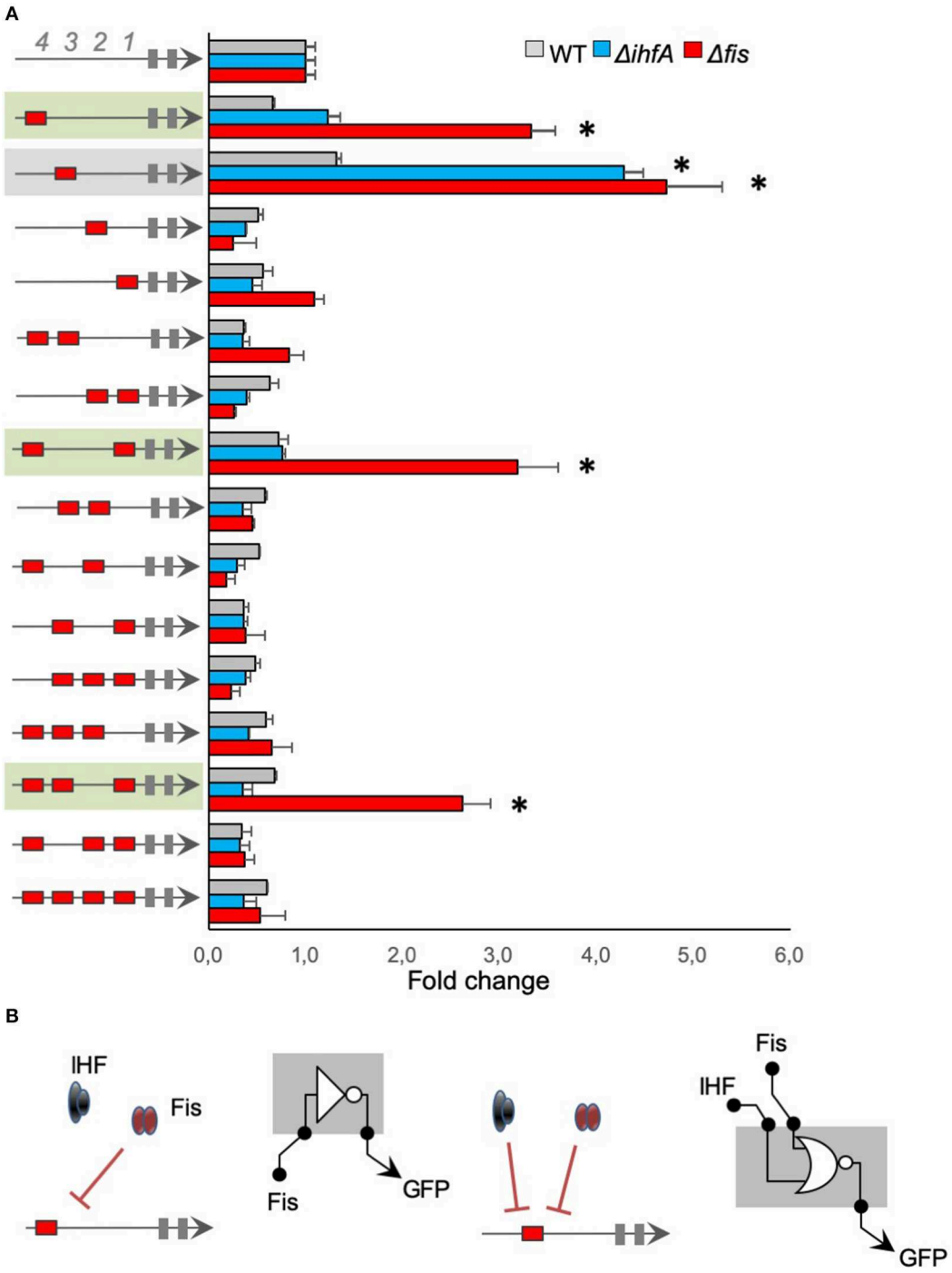


FIGURE 2 | The activity of promoters harboring Fis-binding sites **(A)** The architecture of the synthetic promoter is shown on the left (red boxes represent Fis-BS). Promoter activities are shown in bars and normalized based on the activity of the reference promoter (i.e., a promoter with 4 neutral sequences). Promoter analyses were performed for 4 h of growth, three genetic backgrounds of *E. coli* (wild type—gray bars, Δ*fis*—red bars, and Δ*ihfA*—blue bars). Promoters that displayed a significant increase in activity compared to the wild-type strain were shaded in green or gray for easy viewing. Statistical differences between synthetic promoters and their control (wild type condition) are highlighted by (*) as analyzed using Student's *t*-test with $p < 0.05$. **(B)** Summary of most significant changes in promoter architecture leading to changes in promoter logic.

(−101 and −121 relative to the TSS) displayed both a strong repression by IHF but no modulation by Fis (promoter shaded in gray in **Figure 3**). Again, these results reinforce that the gene expression pattern and the promiscuous or specific binding to transcriptional factors allows for the fine-tuning of promoter activities based on their architectures.

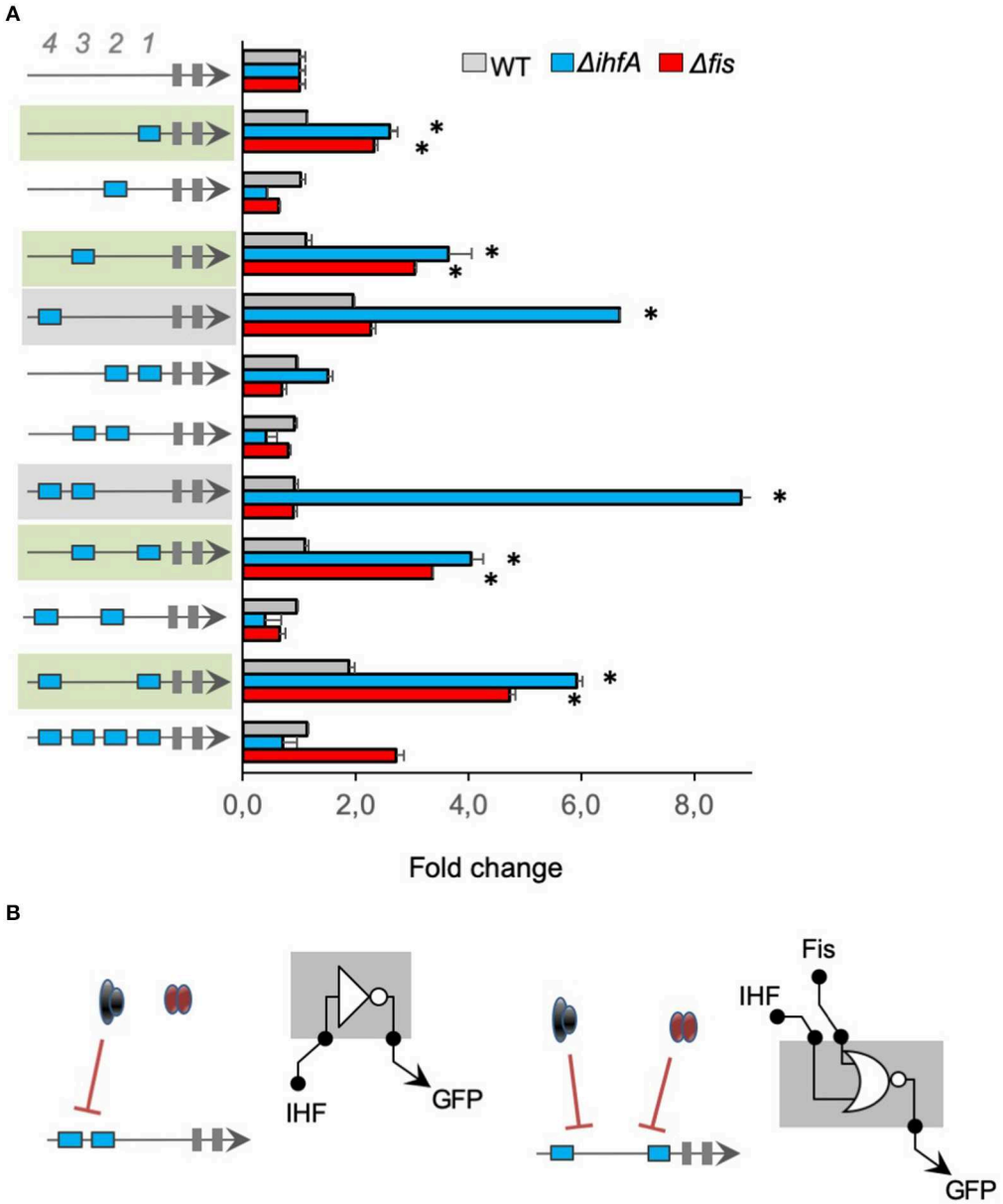


FIGURE 3 | The activity of promoters harboring IHF-binding sites. **(A)** The architecture of the synthetic promoter is shown on the left (blue boxes represent IHF-BS). Promoter activities are shown in bars and normalized based on the activity of the reference promoter (i.e., a promoter with 4 neutral sequences). Eleven promoter variants previously described (Monteiro et al., 2018) were analyzed in wild type (gray bars) Δ*fis* (red bars), and Δ*ihfA* (blue bars) mutant strains of *E. coli*. Promoter architectures that displayed higher activity in *E. coli* Δ*fis* and Δ*ihfA* are shaded in green, indicating that Fis was also able to repress these promoter variants, highlighting a possible crosstalk. Promoter that displayed a strong repression by IHF but no modulation by Fis are shaded in gray, reinforce that the gene expression pattern and the promiscuous or specific binding to transcriptional factors allows for the fine-tuning of promoter activities based on their architectures. Statistical differences between synthetic promoters and their control (wild type condition) are highlighted by (*) as analyzed using Student's *t*-test with $p < 0.05$. **(B)** Summary of most significant changes in promoter architecture leading to changes in promoter logic. Statistical differences between synthetic promoters and their control are highlighted by (*) as analyzed using Student's *t*-test with $p < 0.05$.

Merging IHF-BS and Fis-BS Leads to an Unpredictable Expression Pattern

After we investigated the regulatory interactions for promoters harboring *cis*-regulatory elements for a single transcriptional factor (IHF or Fis), we constructed promoters combining binding sites for both TFs and Neg sequences. In order to systematically investigate the effect of combined transcriptional factor-binding sites on promoter logic, we first fixed 1 IHF-BS at the 1st position (−61) and varied Fis-BS for the 2nd, 3rd, and 4th positions. As shown in **Figure 4A**, 1 promoter harboring 1 single IHF-BS at

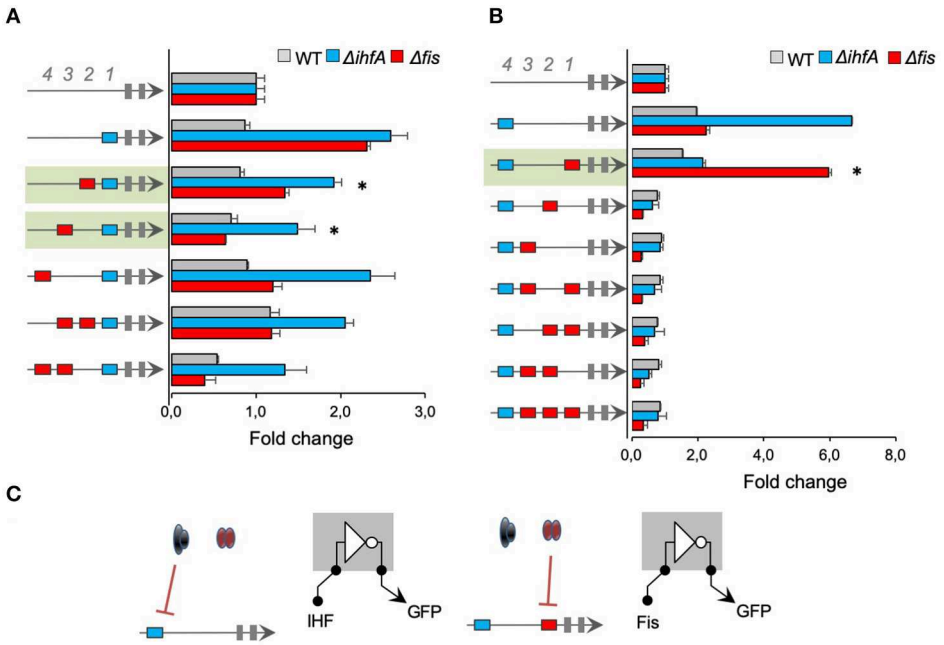


FIGURE 4 | The activity of promoters with combined IHF- and Fis-binding sites. The architecture of the synthetic promoter is shown on the left (blue boxes represent IHF-BS and red boxes represent Fis-BS). Promoter activities are shown in bars and normalized based on the activity of the reference promoter (i.e., a promoter with 4 neutral sequences). Promoter variants were analyzed in wild type (gray bars) Δ*fis* (red bars), and Δ*ihfA* (blue bars) mutant strains of *E. coli*. **(A)** Characterization of promoters with a single IHF-BS fixed at position 1 (−61) and varying Fis-BS. Promoters shaded in green present reduced activity in the *ihfA* and *fis* mutant strains, compared to IHF-BS at the 1st position. Statistical differences between synthetic promoters and their control (PNNNI) are highlighted by (*) as analyzed using Student's *t*-test with $p < 0.05$. **(B)** Characterization of promoters with a single IHF-BS fixed position 4 (−121) and varying Fis-BS. Promoter shaded in green displayed increased activity in the *E. coli* Δ*fis* strain, while it showed no activity in the wild type and Δ*ihfA* strains. Statistical differences between this promoter in Δ*fis* condition and wild type and Δ*ihfA* are highlighted by (*) as analyzed using Student's *t*-test with $p < 0.05$. **(C)** Summary of most significant changes in promoter architecture leading to changes in promoter logic. Statistical differences between synthetic promoters and their control are highlighted by (*) as analyzed using Student's *t*-test with $p < 0.05$.

the 1st position showed no activity in the wild-type *E. coli* strain but increased activity in the *fis* and *ihfA* mutant strains. However, adding Fis-BS at the 2nd or 3rd position resulted in promoters with reduced activity in the *ihfA* and *fis* mutant strains, compared to IHF-BS at the 1st position (promoters shaded in green in **Figure 4A**). Comparison of these green shaded promoters to promoters with 1 single Fis-BS at the 2nd or 3rd positions in **Figure 2**, we cannot observe any patterns between the merging of binding sites for these transcriptional factors, that is, the activity of promoters consisting of both Fis-BS and IHF-BS is not the sum of behaviors from Fis-BS and IHF-BS individually. When 1 single IHF-BS was fixed at the 4th position (−121), the resulting promoter displayed strong activity in Δ*ihfA* strains (**Figure 4B**). However, when 1 single Fis-BS was added at the 1st position (−61), the resulting promoter displayed increased activity in the *E. coli* Δ*fis* strain, while it showed no activity in the wild type and Δ*ihfA* strains. Therefore, this promoter architecture may be being repressed, especially by Fis regulator (shaded in green). However, for promoters with Fis-BS fixed at the 1st position (**Figure 2**), we observed a reduction in the promoter activity in the Δ*fis* strain, demonstrating that the presence of IHF in this specific position may influence a positive expression in the absence of Fis. Finally, the addition of 1 single or multiple Fis-BS at different positions completely blocked promoter activity, and this was not relieved in either Δ*fis* or Δ*ihfA* strains, showing that transcriptional factors and binding site sequences of IHF and Fis contribute to promoter complexity. A mutant for *ihfA* and *fis* should be a compelling model to completely understand this promoter logic, but a mutant for both TFs has proven to be difficult to construct. It is important to note that IHF and Fis, which are transcriptional factors, are also NAPs, so the gene expression identified here could be related to possible changes in the DNA geometry (Déthiollaz et al., 1996). Taken together, these results also suggest that Fis and IHF proteins and their binding sites exert complex regulatory patterns, hampering promoter behavior predictions.

Combination of Fis and IHF Binding Sites Generates Strong Fis and IHF Activated Promoters

In all promoters presented until this point, while the combination of different *cis*-regulatory genes was able to determine the regulatory logic displayed by IHF and Fis, the 2 TFs acted as repressors of promoter activity (**Figures 2–4**). However, this behavior shifted when we constructed promoter versions harboring IHF-BS at the 1st and 4th positions and varying sites for Fis-BS (**Figure 5**). As shown in this figure, when 1

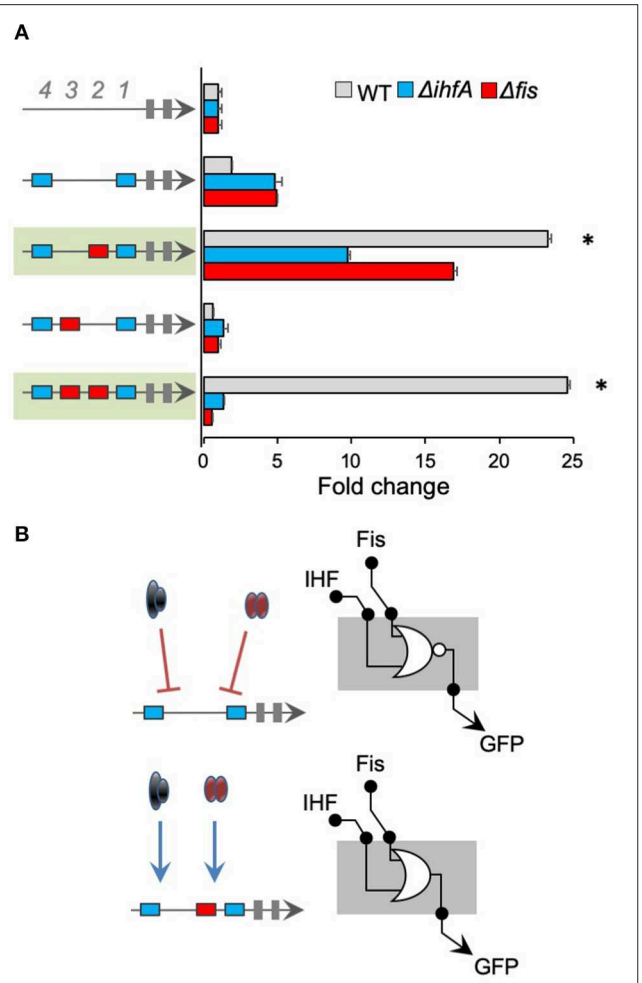


FIGURE 5 | Analysis of promoters with 2 fixed IHF-binding sites. **(A)** The architecture of the synthetic promoter is shown on the left (blue boxes represent IHF-BS and red boxes represent Fis-BS). Promoter activities are shown in bars and normalized based on the activity of the reference promoter (i.e., a promoter with 4 neutral sequences). Promoter variants were analyzed in wild type (gray bars) Δ*fis* (red bars), and Δ*ihfA* (blue bars) mutant strains of *E. coli*. For this analysis, IHF *cis*-regulatory elements were placed at positions 1 and 4, and additional Fis-BS were introduced into the promoters. Promoter that displayed a strong activity in the wild-type when compared to the Δ*fis* or Δ*ihfA* mutant of *E. coli* are shaded in green. Statistical differences between synthetic promoters in wild type condition and Δ*fis* and Δ*ihfA* condition are highlighted by (*) as analyzed using Student's *t*-test with $p < 0.05$. **(B)** Summary of most significant changes in promoter architecture leading to changes in promoter logic. Statistical differences between synthetic promoters and their control are highlighted by (*) as analyzed using Student's *t*-test with $p < 0.05$.

single Fis-BS was added at the 2nd position (−81), the resulting promoter displayed a strong activity in the wild-type strain of *E. coli*, when compared to the version lacking this element (promoter in the green shaded region in **Figure 5**). Furthermore, when these promoters were assayed in *E. coli* Δ*fis* or Δ*ihfA*, we observed a substantial reduction in their activity, indicating that both TFs acted as activators of the combinatorial promoter. The same behavior was also observed for a promoter harboring 2 IHF-BS (at the 1st and 4th positions) and 2 Fis-BS (2nd and 3rd positions), where reduction in the gene expression was even more evident. The same does not occur for a promoter harboring the 2 sites of IHF-BS and Fis-BS at the 3rd position, indicating the dependence and complexity of the relationship between promoter architecture and gene expression. These results highlight the rise of emergent properties in complex promoters for global regulators (Monteiro et al., 2018), as increasing the number of *cis*-regulatory elements can drastically shift the final regulatory logic of the system.

Conclusions

Bacteria are naturally endowed with complex promoters harboring multiple binding sites for several TFs. While several works based on mathematical modeling have argued that combinatorial regulation can be predicted from the characterization of individual promoter elements (Yuh, 1998; Bintu et al., 2005; Hermsen et al., 2006; Zong et al., 2018), along with the previous report (Monteiro et al., 2018) and here we provide growing evidence that small changes in the architecture of *cis*-regulatory elements can drastically change the final response of the system (Kreamer et al., 2016). The unpredictable behaviors observed in these studies might also depict a deeper evolutionary trend in gene regulation that has selected molecular systems/mechanisms capable of promoting both evolvability and robustness of gene expression levels through non-linear gene regulation (Steinacher et al., 2016). Thus, understanding the way the architecture of *cis*-regulatory elements determines gene expression behavior is pivotal not only to understand natural bacterial systems but also to provide novel conceptual frameworks for the construction of synthetic promoters for biotechnological applications (Monteiro et al., 2019b). Frequently, in genetic bioengineering applications, it is also necessary to fine-tune and balance specific gene expression due to the complexity of regulatory networks (Boyle and Silver, 2012; Scalcinati et al., 2012; Steinacher et al., 2016). Several recent studies have focused on the improvement of this strategy for diverse purposes (Egbert and Klavins, 2012; Siegl et al., 2013; Hwang et al., 2018). The present adjusting approach could be used as a strategy for the fine-tuning of genetic circuits to perform optimally in a given context. Our approach provided a library (from this study and from our previous work (Monteiro et al., 2018) of 74 promoter architectures characterized in different strains and conditions for in total of 230 outputs (different promoters in different strains and growth conditions) (**Figure 6** and **Table S1**). Promoters from our synthetic promoter library with small adaptations could be used for diverse purposes in the biotechnological and bacterial network gene regulation fields.

Abstracting all the gene regulations investigated in this work, we are able to provide a visual summary of the findings reported here from a Boolean logic perspective (**Figure 7**). As shown in **Figure 7A**, changing a perfect Fis binding in 20 bp (from position −121 to −101) can turn a specific Fis-repressed promoter into a system repressed by both Fis and IHF. Using a more formal logic gate definition (Amores et al., 2015), this modification can turn a promoter with a NOT logic into one with an NOR logic. However, a promoter harboring 2 IHF-binding sites at

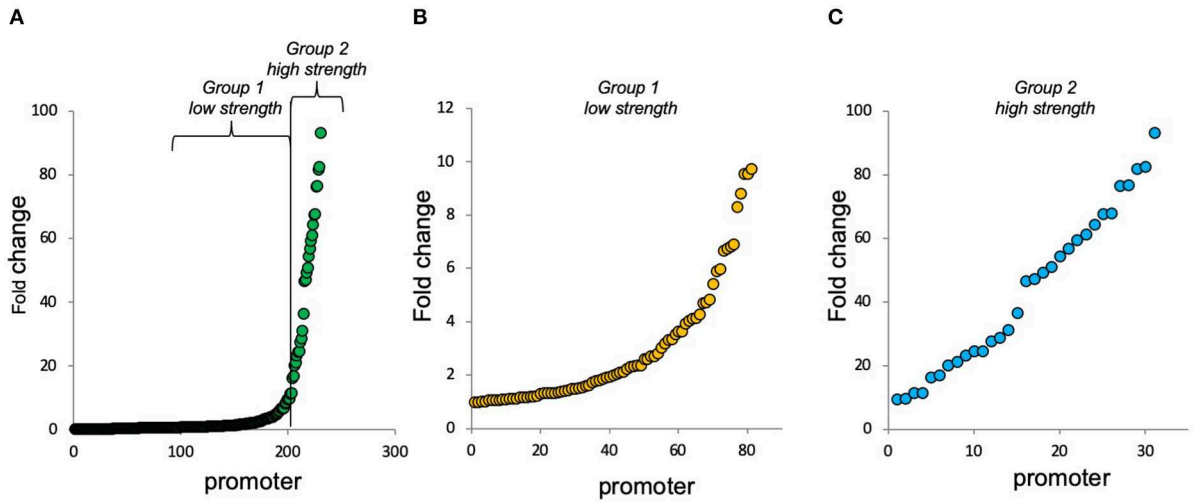


FIGURE 6 | Fold change of complex synthetic promoters at 4 h of growth. Complex promoters' approach could be used as a strategy for the fine-tuning of genetic circuits to perform optimally in a given context. A total of 230 promoter characterizations (74 architectures characterized in different strains and conditions). All promoter details are shown in **Table S1**. **(A)** All 230 promoter characterizations (from the lowest to the highest fold change) that were divided into 2 groups of interests. **(B)** Promoters with fold change between 1 and 10 (group 1: low strength). **(C)** Promoters with fold change up to 10 (group 2—high strength).

positions –121 and –101 displayed specific IHF-repression, while changing the second binding site to position –61 resulted in a promoter repressed by both IHF and Fis (**Figure 7B**). In terms of promoter logic, this change in cis-element architecture also turns a promoter with NOT logic into one with an NOR logic. When a single IHF-binding site was presented at position –121, the final promoter was only repressed by IHF (**Figure 7C**). Yet, introducing an additional Fis-binding site at position –61 of this promoter turned it into a system exclusively repressed by Fis. This change maintained the NOT logic of the promoter but changed the TF able to repress the activity. Finally, and more remarkably, while a promoter with 2 IHF-BS (at positions –121 and –61) was repressed by both Fis and IHF, adding a third binding site for Fis at position –81 resulted in a promoter strongly activated by both TFs (**Figure 7D**). Therefore, this single-change cis-element architecture turned a promoter with NOR logic into an entirely OR promoter responsive to the same TFs. This remarkable regulatory versatility and unpredictability unveiled by synthetic combinatorial promoter shows that we only start to understand the complexity of gene regulation in bacteria. While the work presented here covers two of the main global regulators of *E. coli*, further studies are still necessary to uncover the hidden complexity of combinatorial gene regulation in this bacterium.

MATERIALS AND METHODS

Plasmids, Bacterial Strains, and Growth Conditions

E. coli DH10B was used for cloning procedures, while *E. coli* BW25113 was used as the wild-type strain (WT); *E. coli* JW1702-1 was used as a mutant for the IHF transcription factor (TF), and *E. coli* JW3229 was used as a mutant for the Fis TF. All strains were obtained from the Keio collection. For the procedures and analyses, *E. coli* strains were grown in M9 minimal media (6.4 g/L, $Na_2HPO_4 \bullet 7H_2O$, 1.5 g/L KH_2PO_4, 0.25 g/L NaCl, 0.5 g/L NH_4Cl) supplemented with chloramphenicol at 34 μg/mL, 2 mM $MgSO_4$, 0.1 mM casamino acids, and 1% glycerol as the sole carbon source (Complete M9) at 37°C. Plasmids, bacterial strains, and primers used in this study are listed in **Table 1**.

Design of Synthetic Promoter Scaffolds and Ligation Reactions

The construction of synthetic promoters was performed by the ligation reaction of 5′–end phosphorylated oligonucleotides acquired from Sigma Aldrich (**Table 1**). The design of all single strands was projected to carry a 16 bp sequence containing the Fis binding site (F), IHF binding site (I), or a neutral motif (N), which is a sequence where any TF is able to bind (**Figure 1A**). These locations were identified as positions 1, 2, 3, and 4, respectively (**Figure 1B**) and to be located at −61, −81, −101, or −121 bp upstream of the core promoter (**Figure 1C**). In addition to the 16 bp oligonucleotides, all single strands were designed to contain 3 base pairs overhang for its corrected insertion on the promoter (**Figure 1C**). Additionally, a core promoter based on the lac promoter, which is a weak promoter and therefore requires activation. The design of the synthetic promoters and the positions of the cis-elements were made based on strategies already performed by our group (Monteiro et al., 2018), aiming to arrange the cis-elements aligned to the transcription initiation site, considering the DNA curvature. To assemble the synthetic promoters, the 5′ and 3′ strands corresponding to each position were mixed at equimolar concentrations and annealed by heating at 95°C for 5 min, followed by gradual cooling to room temperature (25°C) for 5 min, and finally maintained at 0°C for 5 min. The external overhangs of the *cis*-element at position 4 and the core promoter were designed to carry EcoRI and BamHI

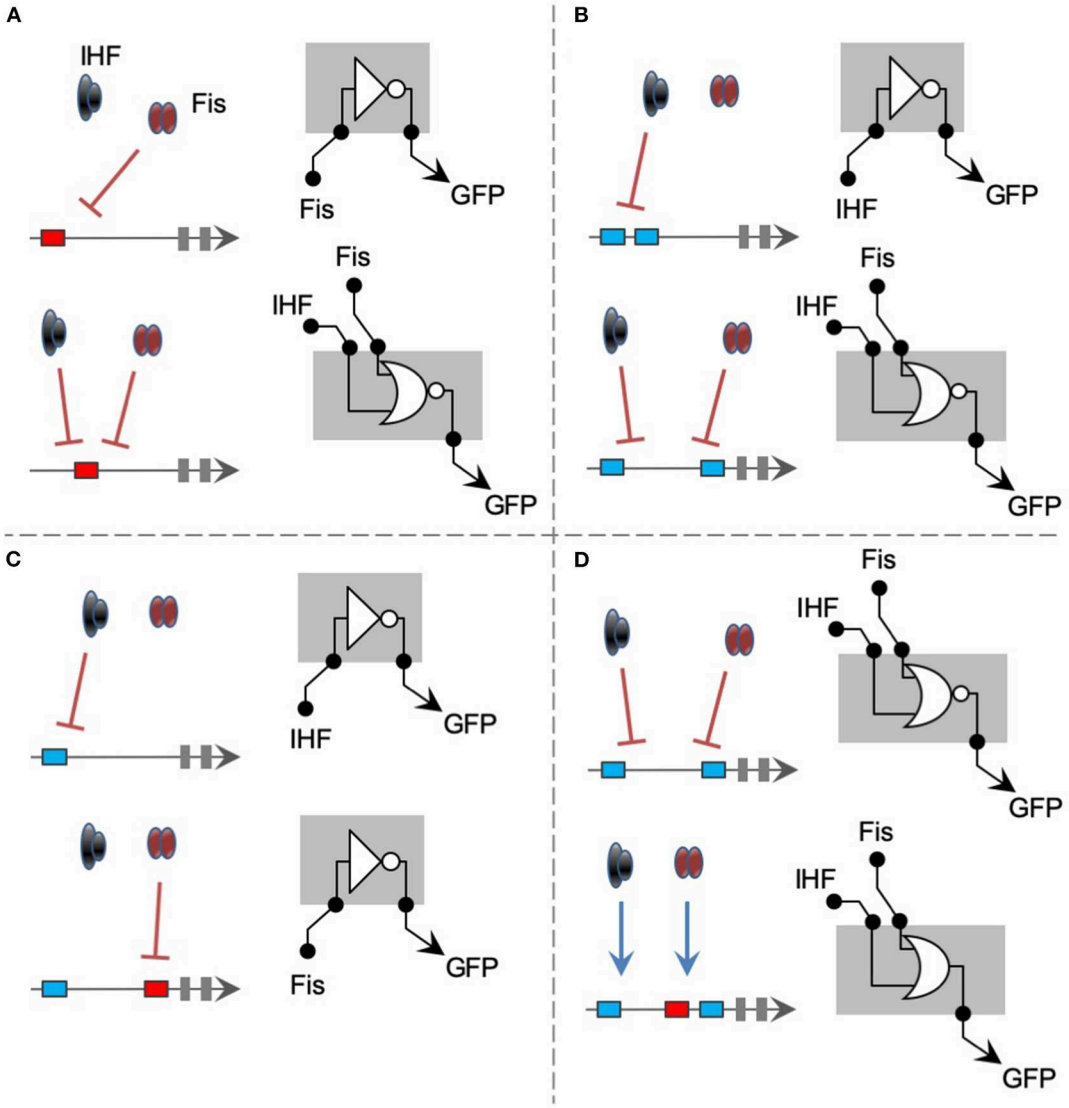


FIGURE 7 | Summary of most significant changes in promoter architecture leading to changes in promoter logic. The figures are represented using logic gate representation for gene regulation, even. **(A)** The change of a single Fis-BS from position −121 to −101 turns a NOT gate for Fis into an NOR gate for Fis and IHF. **(B)** Two IHF-BSs (at positions −121 and −101) worked as a NOT gate for IHF, while changing 1 site to position −61 generates an NOR gate for both IHF and Fis. **(C)** While a single IHF-binding site at position −121 results into a NOT gate for IHF, adding a Fis-binding site to the position −61 creates a NOT gate exclusively dependent on Fis. **(D)** Finally, adding a Fis-BS (position −81) to the NOR gate promoter presented in B drastically changes its logic to an OR gate, where the 2 TFs act as activators.

digested sites. In this way, it was allowed to ligate to a previously digested EcoRI/BamHI pMR1 plasmid. All five fragments (4 *cis*-elements positions plus core promoter) were mixed equally in a pool with the final concentration of 5′ phosphate termini fixed at 15 μM. For the ligase reaction, 1 μL of the fragment pool was added to 50 ng EcoRI/BamHI pMR1 digested plasmid in the presence of ligase buffer and ligase enzyme to a final volume of 10 μL. Ligation was performed for 1 h at 16°C, after which the ligase reaction was inactivated for 15 min at 65°C. Two μL of the ligation was used to electroporate 50 μL of *E. coli* DH10B competent cells. After 1-h regenerating in 1 mL LB media, the total volume was plated in LB solid dishes supplemented with chloramphenicol at 34 μg/mL. Clones were confirmed by colony PCR with primers pMR1-F and pMR1-R (**Table 1**) using pMR1 empty plasmid PCR reaction as further length reference on electrophorese agarose gel. Clones with a potential correct length were submitted to Sanger DNA sequencing to confirm correct promoter assembly.

Promoter Activity Analysis and Data Processing

Promoter activity was measured for all 42 promoters at different genetic backgrounds and conditions. For each experiment,

the plasmid containing the promoter of interest was used to transform *E. coli* wild type, *E. coli* Δ*ihfA* mutant, or *E. coli* Δ*fis* mutant, as indicated. Freshly plated single colonies were selected with sterile loops and then inoculated in 1 mL of M9 media. After 16 h 10 μL of this culture was assayed in 96 wells microplates in biological triplicate with 190 μL of M9 media. Cell growth and GFP fluorescence were quantified using a Victor X3 plate reader (PerkinElmer) that was measured for 8 h at intervals of 30 min. All graphics were constructed based on 4 h of cell growth since under our experimental setup and previous work (Monteiro et al., 2018), most promoters reach maximal activity at 4 h of growth. Therefore, this is the best time point to compare maximal promoter activity. Promoter activities were calculated as arbitrary units dividing the GFP fluorescence levels by the optical density at 600 nm (reported as GFP/OD_{600}) after background correction. Technical triplicates and biological triplicates were performed in all experiments. Raw data were processed using *ad hoc* R script (https://www.r-project.org/), and plots were constructed using R (version R-3.6.3). For all analyses, we calculated fold-change expression using pMR1-NNNN as the promoter reference.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

RS-R and LM designed the experimental strategy. LM, AS-M, and CW performed the experiments. LM, AS-M, CW, and RS-R analyzed and interpreted the data. LM and RS-R wrote the manuscript. All authors have read and approved the final version of the manuscript.

FUNDING

This work was supported by the São Paulo Research Foundation (FAPESP, award # 2012/22921-8 and 2017/50116-6). LM, AS-M, and CW were supported by FAPESP PhD and Master Fellowships (award # 2016/19179-9, 2018/04810-0, and 2016/05472-6).

ACKNOWLEDGMENTS

The authors are thankful to their lab colleagues for insightful comments on this work. This manuscript has been released as a pre-print at bioRxiv (Monteiro et al., 2019a).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2020.00510/full#supplementary-material

REFERENCES

Aeling, K. A., Opel, M. L., Steffen, N. R., Tretyachenko-Ladokhina, V., Hatfield, G. W., Lathrop, R. H., et al. (2006). Indirect recognition in sequence-specific DNA binding by *Escherichia coli* integration host factor. *J. Biol. Chem.* 281, 39236–39248. doi: 10.1074/jbc.M606363200

Amores, G. R., Guazzaroni, M. E., and Silva-Rocha, R. (2015). Engineering synthetic cis-regulatory elements for simultaneous recognition of three transcriptional factors in bacteria. *ACS Synth. Biol.* 4, 1287–1294. doi: 10.1021/acssynbio.5b00098

Azam, T. A., Hiraga, S., and Ishihama, A. (2000). Two types of localization of the DNA-binding proteins within the *Escherichia coli* nucleoid. *Genes Cells* 5, 613–626. doi: 10.1046/j.1365-2443.2000.00350.x

Azam, T. A., and Ishihama, A. (1999). Twelve species of the nucleoid-associated protein from *Escherichia coli*. *J. Biol. Chem.* 274, 33105–33113. doi: 10.1074/jbc.274.46.33105

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2, 1–11. doi: 10.1038/msb4100050

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., et al. (2005). Transcriptional regulation by the numbers: models. *Curr. Opin. Genet. Dev.* 15, 116–124. doi: 10.1016/j.gde.2005.02.007

Boyle, P. M., and Silver, P. A. (2012). Parts plus pipes: synthetic biology approaches to metabolic engineering. *Metab. Eng.* 14, 223–232. doi: 10.1016/j.ymben.2011.10.003

Browning, D. F., and Busby, S. J. (2004). The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.* 2, 57–65. doi: 10.1038/nrmicro787

Browning, D. F., and Busby, S. J. W. (2016). Local and global regulation of transcription initiation in bacteria. *Nat. Rev. Microbiol.* 14, 638–650. doi: 10.1038/nrmicro.2016.103

Browning, D. F., Grainger, D. C., and Busby, S. J. (2010). Effects of nucleoid-associated proteins on bacterial chromosome structure and gene expression. *Curr. Opin. Microbiol.* 13, 773–780. doi: 10.1016/j.mib.2010.09.013

Cepeda-Humerez, S. A., Rieckh, G., and Tkačik, G. (2015). Stochastic proofreading mechanism alleviates crosstalk in transcriptional regulation. *Phys. Rev. Lett.* 115:248101. doi: 10.1103/PhysRevLett.115.248101

Collado-Vides, J., Magasanik, B., and Gralla, J. D. (1991). Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol. Rev.* 55, 371–394.

Cox, R. S., Surette, M. G., and Elowitz, M. B. (2007). Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.* 3:145. doi: 10.1038/msb4100187

Darwin, A. J., and Stewart, V. (1995). Nitrate and nitrite regulation of the fnr-dependentaeg-46.5 promoter of *Escherichia coli* K-12 is mediated by competition between homologous response regulators (NarL and NarP) for a common DNA-binding site. *J. Mol. Biol.* 251, 15–29. doi: 10.1006/jmbi.1995.0412

Datsenko, K. A., and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* 97, 6640–6645. doi: 10.1073/pnas.120163297

Déthiollaz, S., Eichenberger, P., and Geiselmann, J. (1996). Influence of DNA geometry on transcriptional activation in *Escherichia coli*. *EMBO J.* 15, 5449–5458. doi: 10.1002/j.1460-2075.1996.tb00928.x

Dorman, C. J., and Deighan, P. (2003). Regulation of gene expression by histone-like proteins in bacteria. *Curr. Opin. Genet. Dev.* 13, 179–184. doi: 10.1016/S0959-437X(03)00025-X

Egbert, R. G., and Klavins, E. (2012). Fine-tuning gene networks using simple sequence repeats. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16817–16822. doi: 10.1073/pnas.1205693109

Friedlander, T., Prizak, R., Guet, C. C., Barton, N. H., and Tkačik, G. (2016). Intrinsic limits to gene regulation by global crosstalk. *Nat. Commun.* 7:12307. doi: 10.1038/ncomms12307

Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., Ledezma-Tejeida, D., Muñiz-Rascado, L., García-Sotelo, J. S., et al. (2016). RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucl. Acids Res.* 44, D133–D143. doi: 10.1093/nar/gkv1156

Guazzaroni, M.-E., and Silva-Rocha, R. (2014). Expanding the logic of bacterial promoters using engineered overlapping operators for global regulators. *ACS Synth. Biol.* 3, 666–675. doi: 10.1021/sb500084f

Hermsen, R., Tans, S., and ten Wolde, P. R. (2006). Transcriptional regulation by competing transcription factor modules. *PLoS Comput. Biol.* 2:e164. doi: 10.1371/journal.pcbi.0020164

Hirvonen, C. A., Ross, W., Wozniak, C. E., Marasco, E., Anthony, J. R., Aiyar, S. E., et al. (2001). Contributions of UP elements and the transcription factor FIS to expression from the seven rrn P1 promoters in *Escherichia coli*. *J. Bacteriol.* 183, 6305–6314. doi: 10.1128/JB.183.21.6305-6314.2001

Hwang, H. J., Lee, S. Y., and Lee, P. C. (2018). Engineering and application of synthetic nar promoter for fine-tuning the expression of metabolic pathway genes in *Escherichia coli*. *Biotechnol. Biofuels* 11:103. doi: 10.1186/s13068-018-1104-1

Isalan, M., Lemerle, C., Michalodimitrakis, K., Horn, C., Beltrao, P., Raineri, E., et al. (2008). Evolvability and hierarchy in rewired bacterial gene networks. *Nature* 452, 840–845. doi: 10.1038/nature06847

Ishihama, A. (2010). Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiol. Rev.* 34, 628–645. doi: 10.1111/j.1574-6976.2010.00227.x

Izu, H., Ito, S., Eilas, M. D., Yamada, M. (2002). Differential control by IHF and cAMP of two oppositely oriented genes, hpt and gcd, in *Escherichia coli* : significance of their partially overlapping regulatory elements. *Mol. Genet. Genomics* 266, 865–872. doi: 10.1007/s00438-001-0608-7

Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., et al. (2017). The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucl. Acids Res.* 45, D543–D550. doi: 10.1093/nar/gkw1003

Kinkhabwala, A., and Guet, C. C. (2008). Uncovering cis regulatory codes using synthetic promoter shuffling. *PLoS ONE* 3:e2030. doi: 10.1371/journal.pone.0002030

Kreamer, N. N., Phillips, R., Newman, D. K., and Boedicker, J. Q. (2016). Predicting the impact of promoter variability on regulatory outputs. *Sci. Rep.* 5:18238. doi: 10.1038/srep18238

Lozada-Chavez, I. (2006). Bacterial regulatory networks are extremely flexible in evolution. *Nucl. Acids Res.* 34, 3434–3445. doi: 10.1093/nar/gkl423

Martínez-Antonio, A., Collado-Vides, J., Balleza, E., López-Bojorquez, L. N., Martínez-Antonio, A., Resendis-Antonio, O., et al. (2003). Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol.* 33, 482–489. doi: 10.1016/j.mib.2003.09.002

Monteiro, L. M. O., Arruda, L. M., Sanches-Medeiros, A., Martins-Santana, L., Alves, L. D. F., Defelipe, L., et al. (2019b). Reverse engineering of an aspirin-responsive transcriptional regulator in *Escherichia coli*. *ACS Synth. Biol.* 8, 1890–1900. doi: 10.1021/acssynbio.9b00191

Monteiro, L. M. O., Arruda, L. M., and Silva-Rocha, R. (2018). Emergent properties in complex synthetic bacterial promoters. *ACS Synth. Biol.* 7, 602–612. doi: 10.1021/acssynbio.7b00344

Monteiro, L. M. O., Sanches-Medeiros, A., Westmann, C. A., and Silva-Rocha, R. (2019a). Modulating Fis and IHF binding specificity, crosstalk and regulatory logic through the engineering of complex promoters. *bioRxiv*. 1–11. doi: 10.1101/614396

Rossiter, A. E., Godfrey, R. E., Connolly, J. A., Busby, S. J. W., Henderson, I. R., and Browning, D. F. (2015). Expression of different bacterial cytotoxins is controlled by two global transcription factors, CRP and Fis, that co-operate in a shared-recruitment mechanism. *Biochem. J.* 466, 323–335. doi: 10.1042/BJ20141315

Rydenfelt, M., Garcia, H. G., Cox, R. S., and Phillips, R. (2014). The influence of promoter architectures and regulatory motifs on gene expression in *Escherichia coli*. *PLoS ONE* 9:e114347. doi: 10.1371/journal.pone.0114347

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Available online at: https://www.cabdirect.org/cabdirect/abstract/19901616061 (accessed April 2, 2020).

Sawers, G. (1993). Specific transcriptional requirements for positive regulation of the anaerobically inducible pfl operon by ArcA and FNR. *Mol. Microbiol.* 10, 737–747. doi: 10.1111/j.1365-2958.1993.tb00944.x

Scalcinati, G., Knuf, C., Partow, S., Chen, Y., Maury, J., Schalk, M., et al. (2012). Dynamic control of gene expression in Saccharomyces cerevisiae engineered for the production of plant sesquitepene α-santalene in a fed-batch mode. *Metab. Eng.* 14, 91–103. doi: 10.1016/j.ymben.2012.01.007

Setty, Y., Mayo, A. E., Surette, M. G., and Alon, U. (2003). Detailed map of a cis-regulatory input function. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7702–7707. doi: 10.1073/pnas.1230759100

Shis, D. L., Hussain, F., Meinhardt, S., Swint-Kruse, L., and Bennett, M. R. (2014). Modular, multi-input transcriptional logic gating with orthogonal LacI/GalR family chimeras. *ACS Synth. Biol.* 3, 645–651. doi: 10.1021/sb500262f

Siegl, T., Tokovenko, B., Myronovskyi, M., and Luzhetskyy, A. (2013). Design, construction and characterisation of a synthetic promoter library for fine-tuned gene expression in actinomycetes. *Metab. Eng.* 19, 98–106. doi: 10.1016/j.ymben.2013.07.006

Steinacher, A., Bates, D. G., Akman, O. E., and Soyer, O. S. (2016). Nonlinear dynamics in gene regulation promote robustness and evolvability of gene expression levels. *PLoS ONE* 11:e0153295. doi: 10.1371/journal.pone.0153295

Ussery, D., Schou Larsen, T., Trevor Wilkes, K., Friis, C., Worning, P., Krogh, A., et al. (2001). Genome organisation and chromatin structure in *Escherichia coli*. *Biochimie* 83, 201–212. doi: 10.1016/S0300-9084(00)01225-6

Yuh, C. (1998). Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science* 279, 1896–1902. doi: 10.1126/science.279.5358.1896

Zong, D. M., Cinar, S., Shis, D. L., Josić, K., Ott, W., and Bennett, M. R. (2018). Predicting transcriptional output of synthetic multi-input promoters. *ACS Synth. Biol.* 7, 1834–1843. doi: 10.1021/acssynbio.8b00165

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

CHAPTER III

Reverse Engineering of an Aspirin-Responsive Transcriptional Regulator in *Escherichia coli*

This chapter was published as:

(Monteiro et al., 2019) MONTEIRO, Lummy Maria Oliveira; ARRUDA, Letícia Magalhaes; SANCHES-MEDEIROS, Ananda; MARTINS-SANTANA, Leonardo; ALVES, Luana de Fátima; DEFELIPE, Lucas; TURJANSKI, Adrian Gustavo; GUAZZARONI, Marıa-Eugenia & SILVA-ROCHA, Rafael. Reverse engineering of an aspirin-responsive transcriptional regulator in *Escherichia coli*. **ACS synthetic biology**, v. 8, n. 8, p. 1890-1900, 2019.

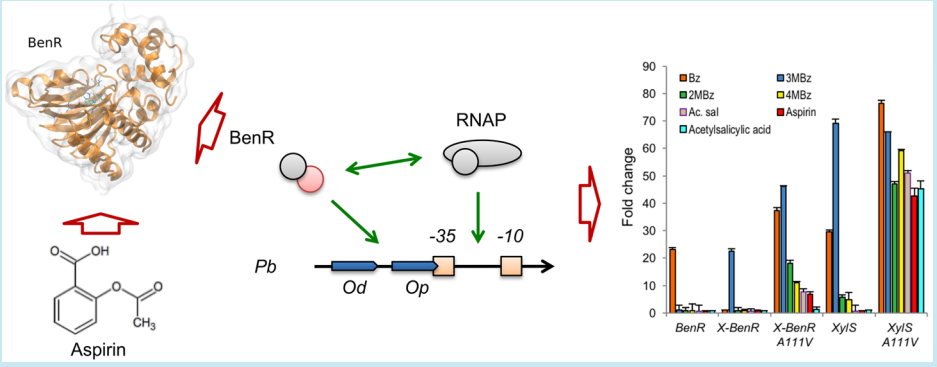
ACS SyntheticBiology
Cite This: ACS Synth. Biol. 2019, 8, 1890−1900
Research Article
pubs.acs.org/synthbio

Reverse Engineering of an Aspirin-Responsive Transcriptional Regulator in *Escherichia coli*

Lummy Maria Oliveira Monteiro,[†] Letícia Magalhães Arruda,[†] Ananda Sanches-Medeiros,[†] Leonardo Martins-Santana,[†] Luana de Fátima Alves,[‡] Lucas Defelipe,[§,∥] Adrian Gustavo Turjanski,[§,∥] María-Eugenia Guazzaroni,[‡] Víctor de Lorenzo,[⊥] and Rafael Silva-Rocha*[,†]

[†]Cell and Molecular Biology Department, FMRP − University of São Paulo, Ribeirão Preto, São Paulo 14049-900, Brazil
[‡]Biology Department, FFCLRP − University of São Paulo, Ribeirão Preto, São Paulo 14040-901, Brazil
[§]Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires 1428, Argentina
[∥]IQUIBICEN/UBA-CONICET, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires 1428, Argentina
[⊥]Systems Biology Program, National Center of Biotechnology − CSIC, Madrid 28049, Spain

Supporting Information



ABSTRACT: Bacterial transcription factors (TFs) are key devices for the engineering of complex circuits in many biotechnological applications, yet there are few well-characterized inducer-responsive TFs that could be used in the context of an animal or human host. We have deciphered the inducer recognition mechanism of two AraC/XylS regulators from *Pseudomonas putida* (BenR and XylS) for creating a novel expression system responsive to acetyl salicylate (i.e., aspirin). Using protein homology modeling and molecular docking with the cognate inducer benzoate and a suite of chemical analogues, we identified the conserved binding pocket of BenR and XylS. By means of site-directed mutagenesis, we identified a single amino acid position required for efficient inducer recognition and transcriptional activation. Whereas this modification in BenR abolishes protein activity, in XylS, it increases the response to several inducers, including acetyl salicylic acid, to levels close to those achieved by the canonical inducer. Moreover, by constructing chimeric proteins with swapped N-terminal domains, we created novel regulators with mixed promoter and inducer recognition profiles. As a result, a collection of engineered TFs was generated with an enhanced response to benzoate, 3-methylbenzoate, 2-methylbenzoate, 4-methylbenzoate, salicylic acid, aspirin, and acetylsalicylic acid molecules for eliciting gene expression in *E. coli*.

KEYWORDS: *transcriptional regulation, protein engineering, homology modeling, gene regulatory network, reverse engineering, bacterial expression system*

Transcriptional regulation plays a central role in the adaptation of cells to changing environmental conditions. In bacteria, this step is mainly regulated by the interaction of RNA polymerase (RNAP) with the promoter region through the use of many sigma factors and by a large number of transcription factors (TFs) that can promote or block RNAP binding or further steps in transcription.[1] With the growing interest in the engineering of living cells for novel biotechnological and biomedical applications, a special focus has emerged in understanding gene regulation at the molecular level.[1−4] In this context, many different classes of TFs have been extensively characterized in the molecular detail from bacteria to mammals, and this knowledge has allowed a number of engineering projects, where natural systems can be

Received: April 29, 2019
Published: July 30, 2019

DOI: 10.1021/acssynbio.9b00191
ACS Synth. Biol. 2019, 8, 1890−1900

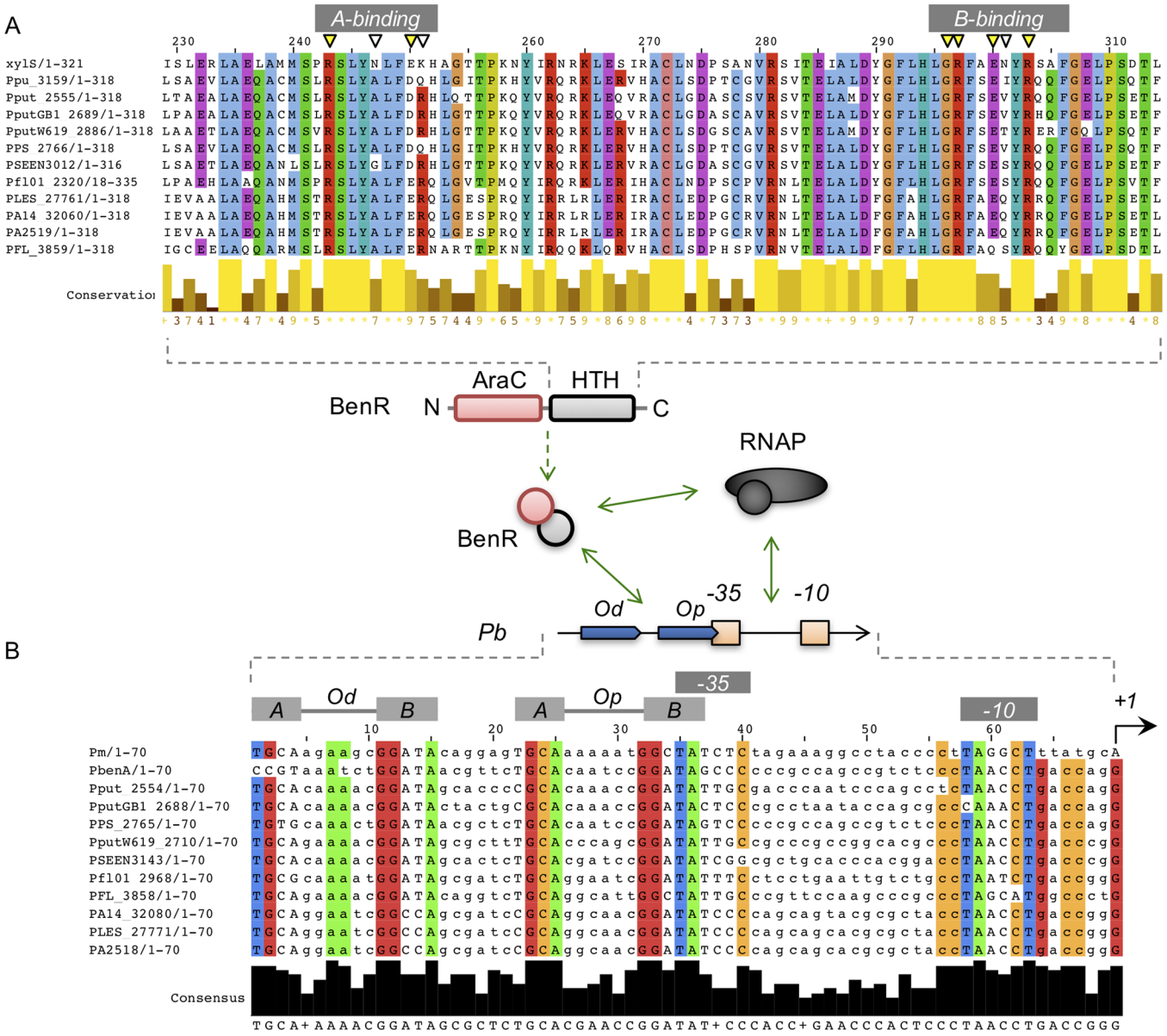


Figure 1. Analysis of BenR homologues and target promoters in *Pseudomonas*. (A) Analysis of protein conservation between XylS from *Pseudomonas putida* mt-2 and BenR homologues from strains of *P. putida* (Ppu_3159, Pput_2555, PputGB1_2689, PputW619_2886, PPS_2765), *P. entomophila* (PSEEN3143), *P. fluorescens* (Pfl01_2968, PFL_3858), and *P. aeruginosa* (PA14_32080, PLES_27771, PA2518). *PbanA* in the second line indicates the promoter from *P. putida* KT2440 (Ppu_3159). Only the region relative to the HTH domain is shown. Critical aa's for DNA recognition (labeled as A-binding and B-binding) are marked with inverted triangles, with conserved regions colored in yellow. In the middle, the schematic representation of the BenR interaction with the RNAP and the σ factor (necessary for the correct initiation of transcription) and the target promoter shows the two binding sites (*Od* and *Op*) and the −35/−10 boxes at *Pb* and *Pm*. (B) Promoter alignment for *Pm* from *P. putida* mt-2 and for *Pb* from several *Pseudomonas* strains. The two conserved boxes (A and B) from *Od* and *Op* binding sites are highlighted.

repurposed to display novel behaviors.[1,5−9] Attempts in this direction are very diverse, and examples include the construction of mutated variants of natural TFs with enhanced or modified performance,[10−15] the recombination of protein domains to create TFs with completely altered specificity or dynamical behavior,[16,17] and the mining of novel regulators from genomes or metagenomes.[18,19] Additionally, the revolution provided by the CRISPR/Cas9 system has also impacted the field of gene regulation, as this system has been repurposed to construct fully synthetic expression systems based on RNA/DNA interaction.[20−22]

Despite the progress in the engineering of novel expression systems, a critical bottleneck relies on the selection of suitable signal-recognition modules related to the application of interest. In other words, whereas many different TFs are well-characterized as responsive to small molecules (sugars, ions, aromatics, etc.), many times, the application at stake requires systems responsive to unusual compounds.[23] Therefore, the construction of TF variants with enhanced responsiveness to non-natural ligands has become more appealing. Approaches to accomplish this task range from the use of laborious random mutagenesis followed by selection[11,12] to the use of computational analysis to guide rational design.[24] Here we focused on the engineering of novel expression systems responsive to commercially available drugs suitable to *in vivo* administration to a mammalian animal. Our

DOI: 10.1021/acssynbio.9b00191
ACS Synth. Biol. 2019, 8, 1890−1900

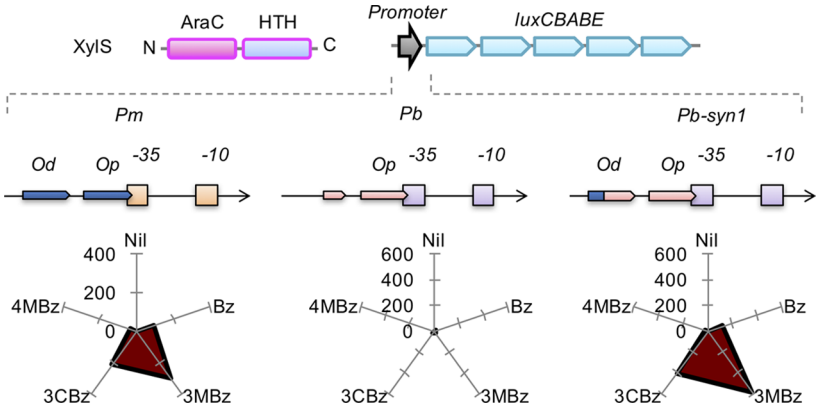


Figure 2. Recognition of *Pm* and *Pb* by XylS. For the analysis, *E. coli* DH5α strain was transformed with pSEVA438 (harboring a functional XylS expressed with its native promoter[33]) and pSEVA226 (a reporter vector with the *luxCDABE* operon[33]) harboring *Pm*, *Pb*, or *Pb-syn1*, a variant of *Pb* endowed with the A box of *Od* from *Pm*.[27] All strains were grown on minimal media to the midexponential phase and then exposed for 3 h to 100 μM benzoate (Bz), 3-methylbenzoate (3MBz), 3-chlorobenzoate (3CBz), or 4-methylbenzoate (4MBz). The graphs represent the fold change of promoter activity relative to the noninduced condition.

particular interest was focused on acetylsalicylic acid (ASA or aspirin), a longstanding, safe and widely used drug. This compound has been applied in synthetic regulatory circuits to deliver lytic proteins to tumors *in vivo*, representing a promising field for the development of tumor-targeting circuits for clinical applications.[25] As a starting point, we sought to investigate the molecular mechanisms of signal recognition by two homologous regulators of *Pseudomonas putida*, namely, BenR and XylS. These two TFs are members of the AraC/XylS family of transcriptional activators[26−28] that recognize different aromatic compounds with structural similarities to ASA. Yet, whereas both regulators share ∼60% amino acid (aa) identity, BenR is responsive to only benzoate,[27] whereas XylS can recognize a large number of substituted variants.[29] Additionally, some crosstalk between these two regulators and their target promoters has been characterized, as XylS can recognize only its target promoter *Pm*, whereas BenR can efficiently activate its natural target *Pb* as well as *Pm*.[27]

In this study, we have investigated the molecular mechanisms of signal recognition by these two regulators using computational tools and in vivo validation. By constructing a model for the ligand-binding domain of BenR and performing molecular docking with benzoate and a collection of analogues, we identified a potential binding pocket strongly conserved between these two TFs. Thereby, we used site-directed mutagenesis and the construction of chimeric proteins to validate the identified binding pocket of the protein. Finally, we demonstrated how a single aa position plays a critical and opposite role in the activity of both proteins. Some changes in this position in BenR resulted in the complete loss of protein activity, whereas the same in XylS triggered an enhanced response to benzoate analogues, including ASA. The results presented here thus provide insights into not only the mechanism of signal recognition by members of the AraC/XylS family but also the engineering of a regulatory device responsive to aspirin.

RESULTS

Analysis of Conserved Elements in BenR and XylS Close Homologues and Target Promoters. To investigate the molecular mechanisms accounting for the functional differences between BenR and XylS, we analyzed the close homologues of these proteins present in the genomes of some species of *Pseudomonas*. As represented schematically in Figure 1, these proteins are TFs formed by two domains, the N-terminal (AraC domain), which is required for ligand recognition,[30] and the C-terminal domain composed of two helix-turn-helix (HTH) regions required for the recognition of the distal and proximal operators (*Od* and *Op*) upstream of the target promoters.[27,31] It is proposed that two monomers of XylS are required for the activation of the *Pm* promoter, each binding to an operator region and contacting an A and B box conserved within this region. A previous study[32] used alanine scanning mutagenesis to identify four residues in XylS required for the recognition of the A boxes (Arg242, Asn246, Glu249, and Lys250) and five required for the interaction with boxes B (G295, Arg296, Asp299, Asn300 and Arg302). As can be observed in the protein alignment between BenR and XylS homologues, most of these positions are well conserved in the proteins analyzed, with the exception of residues Asn246, Lys250, and Asn300 (Figure 1A). It is surprising to notice that whereas the change of Asn to Ala at position 246 in XylS reduced the capability of this protein to activate *Pm* by half, Ala is found to be well-conserved in most BenR proteins analyzed. This indicates that BenR homologues might be less stringent in the interaction at the A box of the target promoter.[27] In the same direction of this hypothesis, the analysis of *Pm* and *Pb* promoters from several *Pseudomonas* strains reveals that most features (A and B boxes, −35/10 region) are well conserved, except for the A box of *Pb* (the target of the BenR studied here) from *Pseudomonas putida* KT2440 (Figure 1B). In this sense, to check the effect of the A box in the interaction between XylS and *Pm* and *Pb* promoters, we assayed the promoter activity using a *lux* reporter system. We used a wild-type *xylS* gene expressed from a pSEVA vector[33] and *Pm*, *Pb*, and *Pb-syn1* (a *Pb* variant with the recomposed A box of the *Od* region[27]). Using this system, we observed that whereas XylS could recognize *Pm* very efficiently, it was not able to induce *Pb* activity in response to the inducers tested (Figure 2). However, when a version of *Pb* with the reconstituted A box (*Pb-syn1*) was used, a strong induction of promoter activity was observed in response to the inducers used (Figure 2).[27] Taken together, these results reinforce the notion that whereas XylS has a critical requirement for complete A and B boxes at the *Od* and *Op* regions, BenR is less stringent for promoter recognition.

DOI: 10.1021/acssynbio.9b00191
ACS Synth. Biol. 2019, 8, 1890−1900

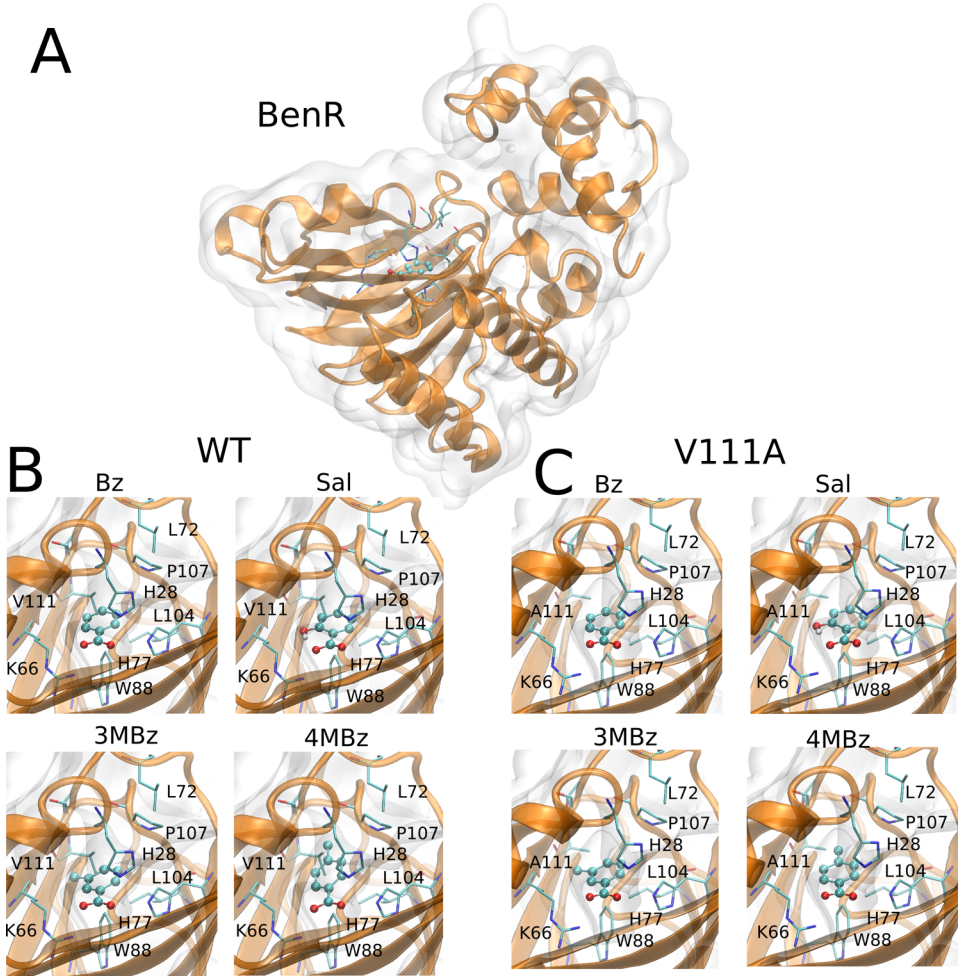


Figure 3. Mapping ligand binding sites in BenR by protein structure prediction and molecular docking. (A) Homology model of wild-type BenR showing the putative binding site of the ligands. (B) Docking of benzoate (Bz), 3-methylbenzoate (3MBz), 4-methylbenzoate (4MBz), and salicylate (Sal) at the wild-type BenR protein, highlighting the critical aa participating in the putative binding site. (C) Docking of BenR-V111A variant showing the putative binding site, as in panel B.

Single Amino Acid Position Is Critical for Aromatic Recognition in BenR and XylS. After tracing critical differences in the DNA recognition requirements between BenR and XylS, we decided to investigate aa differences that could explain the divergence in the ligand selectivity between these two TFs. As presented before, BenR has a very narrow inducer selectivity because this TF can only respond to benzoate as an inducer under natural conditions. However, XylS can be activated by a diverse collection of aromatic inducers, such as benzoate and methylated or chlorinated benzoate analogues.[29] To gain insight into the molecular mechanisms responsible for these differences, we constructed a 3D protein model for the N-terminal region of BenR using homology modeling. The resulting model was subjected to molecular docking using benzoate, 3-methylbenzoate (3MBz), 4-methylbenzoate (4MBz), and salicylate (Sal). Using this approach, we obtained a protein structural model (Figure 3A) and identified a potential cavity on the protein surface that accommodates a benzoate molecule (Figure 3B). Additionally, our results suggest that the identified binding pocket binds with less affinity for the benzoate analogues, indicating that the inducer selectivity could be based on size exclusion. By analyzing the model and the predicted binding pocket, we could identify eight aa's (His 28, Arg 66, Leu 72, His 77, Trp 88, Leu 104, Pro 107, and Val 111) that contributed to the surface of the cavity (Figure 3B,C). We then compared these aa's between BenR and XylS, hypothesizing that a change in some of these aa's could explain the difference in ligand specificity between these two regulators. To our surprise, six out eight aa's from the predicted binding pocket were conserved between the two proteins. The differences were at positions 28 and 111, representing histidine (28) and valine (111) in BenR and tyrosine (28) and alanine (111) in XylS. Because alanine has a shorter side chain, we hypothesized that this could lead to a bigger binding pocket in XylS that could better accommodate the methylated or chlorinated benzoate analogues (Figure 3C and Table S3).

The difference in binding energy suggests that mutation of Val111 to Ala increases the affinity of 3MBz (Table S3). To investigate this possibility, we constructed point mutations in *benR* and *xylS* at aa positions 111 and 110, as we noticed that this last position, while not involved in the binding pocket, was also not conserved between the two regulators (Figure 4A). In addition to the point mutations, we constructed a chimaera between the N-terminal part of XylS and the C-terminal part of BenR and also subjected this TF to mutagenesis. All assays were performed using the cognate promoter for each TF (i.e., *Pb* for BenR and *Pm* for XylS) controlling a green fluorescent

DOI: 10.1021/acssynbio.9b00191
ACS Synth. Biol. 2019, 8, 1890−1900

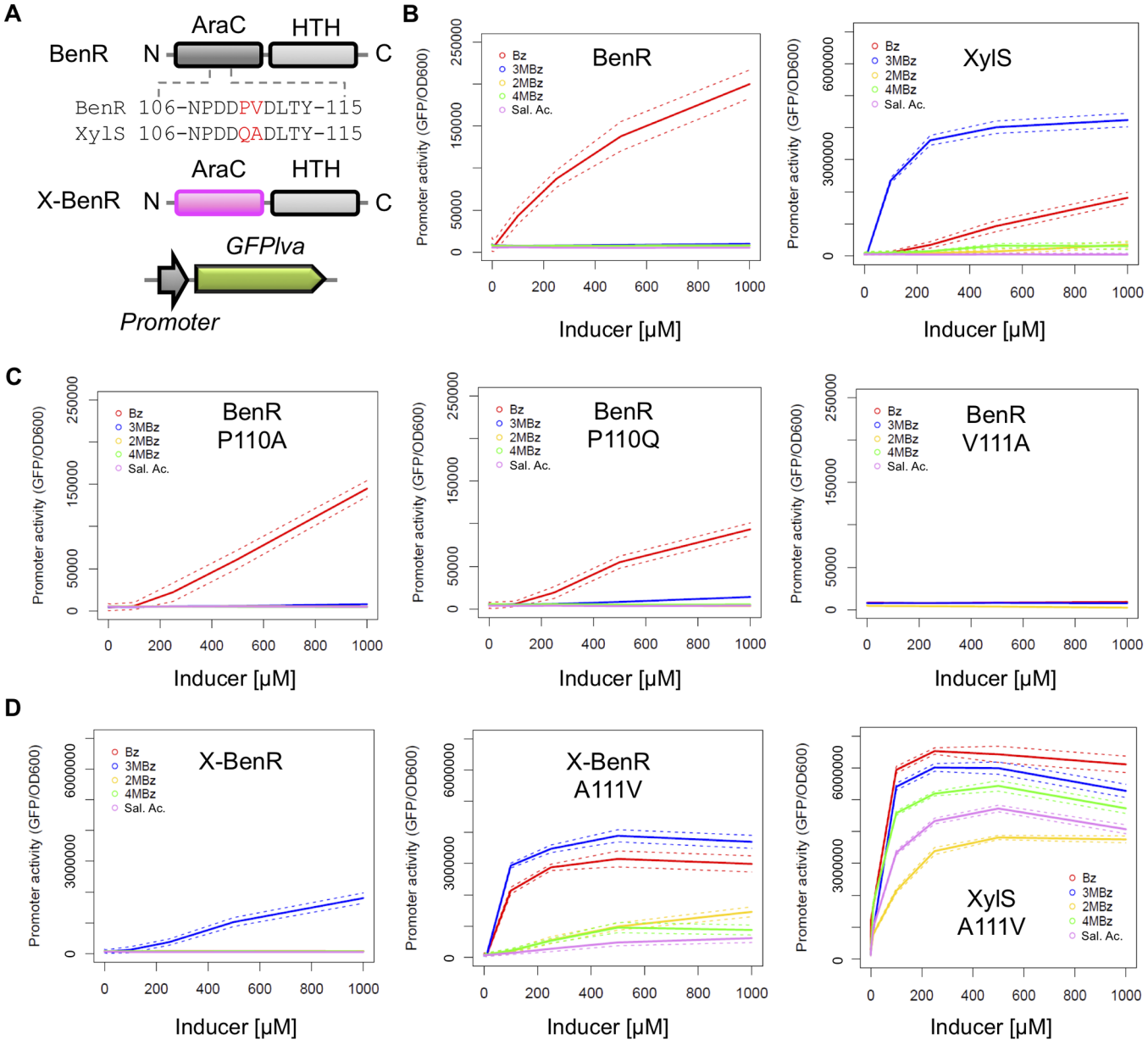


Figure 4. Experimental identification of critical aa's for inducer recognition in BenR and XylS. (A) Schematic representation of BenR and the chimeric protein X-BenR (a fusion between the N-terminal domain of XylS with the C-terminal domain of BenR), showing the nonconserved positions 110 and 111 tested here. All protein variants were tested with a GFPlva reporter under the control of the cognate promoter (i.e., *Pb* for BenR and X-BenR and *Pm* for XylS). (B) Analysis of the promoter activity of BenR-*Pb* and XylS-*Pm* in response to different concentrations of benzoate (Bz), 3-methylbenzoate (3MBz), 2-methylbenzoate (2MBz), 4-methylbenzoate (4MBz), and salicylic acid (Ac Sal). The solid line indicates the average from three independent experiments, whereas the dashed lines represent the lower and higher limits of the standard deviation. (C) Analysis of the promoter activity for BenR mutants at positions 110 (BenR-P110A and BenR-P110Q) and 111 (BenR-V111A). (D) Analysis of the promoter activity for chimeric X-BenR protein and its variant with a mutation in position 111 (X-BenR-A111V) as well as for the mutated version of XylS (XylS-A111V). The promoter activities (panels B–D) reported here were calculated after 5 h of exposure to the different inducers.

protein (GFP) reporter gene (Figure 4A) to allow investigation at the single-cell level[34,35] (Figures S3–S5). The promoter activities were calculated after 5 h of exposure to the different inducers. As can be observed in Figure 4B, wild-type BenR was specific to benzoate at all concentrations tested, whereas wild-type XylS displayed a preferential response to 3MBz, an intermediated response to benzoate, and a lower response to 2MBz and 4MBz. When we mutated position 110 of BenR from Pro to Ala or Gln (the aa found at this position in XylS), we could observe that the mutants presented the same expression profile as the wild type but with reduced efficacy (Figure 4C). However, when position 111 was changed from Val to Ala in BenR, the resulting protein variant did not display any response to the inducers tested. Contrary to the expected, this change (Val 111 to Ala) did not widen the inducer specificity of BenR. By the same token, the construction of BenR mutants with changes at both positions 110 and 111 also resulted in nonfunctional proteins (Figure S1), potentially due to the role of Val111 in signal recognition. Therefore, it is worth noticing that the exchange of a shorter side-chain aa (Ala) for a major side-chain aa (Val) does not necessarily result in a spatial decrease in the aa binding pocket.

DOI: 10.1021/acssynbio.9b00191
ACS Synth. Biol. 2019, 8, 1890–1900

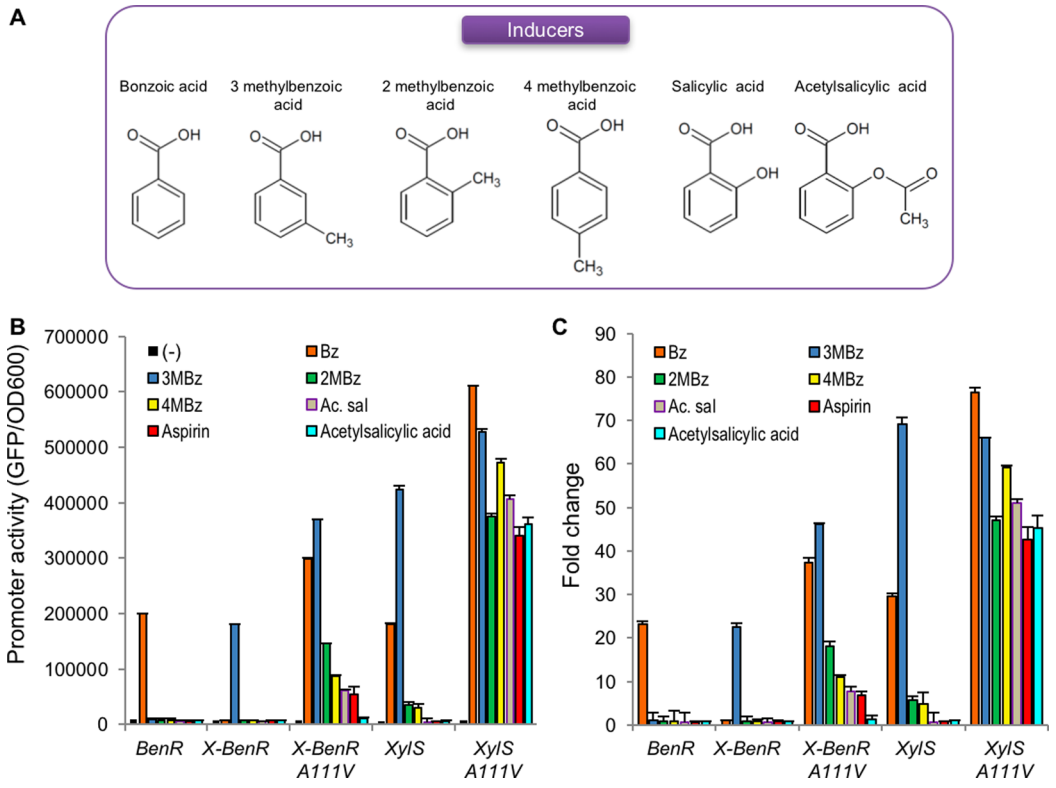


Figure 5. Analysis of transcriptional response to aromatic compounds and aspirin. (A) Schematic representation of the chemical structure of the different inducers tested. (B) Promoter activity of the different proteins reported here after 3 h of exposure to the different inducers. (All compounds were tested at 1 mM.) Aspirin refers to the commercial formulation obtained from a conversional drug store (with concentration adjusted to 100 μM), whereas acetylsalicylic acid specifies the pure compound obtained from Sigma-Aldrich. (C) Fold change calculated for each expression system in response to the different inducers used. Error bars represent the standard deviations from three independent experiments.

Taken together, these results indicate that Val111 is key for inducer recognition by BenR, and it is possible that it is related to the strengthening of its association with the effector binding pocket identified by homology modeling and molecular docking.

The construction of the chimeric protein harboring the N-terminal of XylS (the region responsible for ligand recognition) and the C-terminal of BenR (which recognizes DNA) resulted in the new protein X-BenR that displayed an induction profile similar to that of XylS for the 3MBz induction but with a reduced efficacy (Figure 4D). However, X-BenR increased its inducer response to 3MBz at the same level as that which BenR responds to Bz (Figure 4D and Figure S2), perhaps because *Pm* has a higher promoter activity than *Pb*. To check the role of position 111 in the inducer selectivity of XylS, we constructed point mutations of XylS and X-BenR by changing the alanine at this position for a valine. Unexpectedly, the resulting mutant proteins displayed an enhanced response to the optimal inducer (3MBz) as well as to the suboptimal inducers benzoate, 2MBz, 4MBz, and salicylic acid (Figure 4D). From these results, it appears that the C-terminal of BenR does not allow as much promiscuity as that found in XylS. These results strengthen the notion that the required elements for inducer selectivity are placed in the first 196 aa's of these proteins,[36] although the effector response is not completely independent of, but is also related to, the C-terminal DNA-binding domain. Finally, attempts to construct a chimeric protein harboring the N-terminal of BenR and the C-terminal of XylS (named B-XylS) resulted in nonfunctional products with no detectable response to benzoate or 3MBz (Figure S2).

Yet the single-cell analysis of promoter induction by flow cytometry showed that whereas wild-type BenR- and XylS-based expression systems presented clear unimodal patterns (i.e., with a single population of fluorescent cells), the X-BenR chimaera and mutated versions of XylS and X-BenR displayed a wider population distribution that could indicate stable subpopulations (Figures S3−S5). Taken together, these results evidenced a remarkably different role of position 111 between BenR and XylS, which led to the identification of two TF variants with an enhanced response to a wide range of benzoate derivatives.

Transcriptional Factors Responsive to a New Set of Aromatic Compounds and Aspirin. After the inducer recognition profiles of BenR and XylS are characterized, we assay the TF variants for a new set of inductors. Figure 5 represents the overall performances of the TFs in response to a new set of benzoate derivatives (Figure 5A). As shown in Figure 5B, BenR (*Pb*-BenR) and X-BenR (X-BenR/*Pb*) produced the lowest promoter outputs and were exclusively responsive to benzoate (BenR) or 3MBz (X-BenR). Because these proteins have a BenR C-terminal domain, they can recognize both *Pm* and *Pb*. On the contrary, XylS has an intermediate level of promoter output and a preference for 3MBz, followed by benzoate, and only a minor response to 2MBz and 4MBz. Yet the mutated version of X-BenR (X-BenR-A111V/*Pb*) promoted an overall increased response to the suboptimal inducers of XylS (XylS/*Pm*) and also a response to aspirin and ASA. Finally, the mutated version of XylS (XylS-A111V) displays a remarkable gain of response to all benzoate derivatives tested, including aspirin and ASA, with

DOI: 10.1021/acssynbio.9b00191
ACS Synth. Biol. 2019, 8, 1890−1900

promoter outputs similar to those of wild-type XylS induced with its optimal effector 3MBz (Figure 5B). When fold-change is calculated relative to noninduced conditions, it can be noticed that the maximal induction of the XylS-A111V (XylS-A111V/*Pm*) system (~76-fold) exceeds that of the wild-type XylS (~69-fold, Figure 5C). In this new system, the response to ASA reaches 42- and 45-fold, respectively, yet because this enhanced response could be the result of the construction of a TF with a promiscuous effector specificity, we tested the response of XylS-A111V to toluene, xylene, phenol , and a number of nonaromatic inducers (L-arabinose, fructose, glucose, isopropyl β-D-1-thiogalactopyranoside (IPTG), and arsenite; Figure S6A). As can be observed in Figure S6B, the system did not respond to any of these compounds, whereas further tests with vanillin, 4-hydroxybenzoic acid, and benzilic alcohol also show only a minor effect of this last compound on the activity of X-BenR-A111V and XylS-A111V (Figure S6C). Taken together, these data show that benzoate analogues are the preferential inducers of the newly generated systems described here.

Promoter Engineering Further Enhances X-BenR-A111V Response to Nonoptimal Inducers. The data presented above demonstrate how changes in the aa sequence of BenR and XylS could drastically change the inducer specificity of these two regulators, yet, in the case of BenR, previous studies by our group have demonstrated that changes in the operator sequence of its target promoter (*Pb*) could also impact the way this regulator responds to suboptimal inducers. More specifically, because *Pb* is formed by one complete operator sequence (harboring boxes A and B, Figure 1B) plus another incomplete operator (where the A box is missing), the addition of the A box to restore the second binding site makes the BenR/Pb system more responsive to the suboptimal inducer 3MBz.[27] In this sense, because we noticed a gain of function for the X-BenRA111V chimeric protein constructed here, we decided to investigate if this regulator would have an improved response to the new inducers when activating a mutated version of *Pb* harboring two functional operators (*Pb-syn1*, Figure 6A). As shown in Figure 6B, the newly created X-BenR-A111V regulators displayed an increased response to most inducers tested, with a more than two-fold change in promoter activity when induced with Bz, 3MBz, 2MBz, and 4MBz. These results confirm the previous notion that promoter architecture is an important element controlling the inducer response to suboptimal inducers in BenR regulators and also adds a new modulation level (the promoter itself) for the optimization of biosensors from the XylS family.

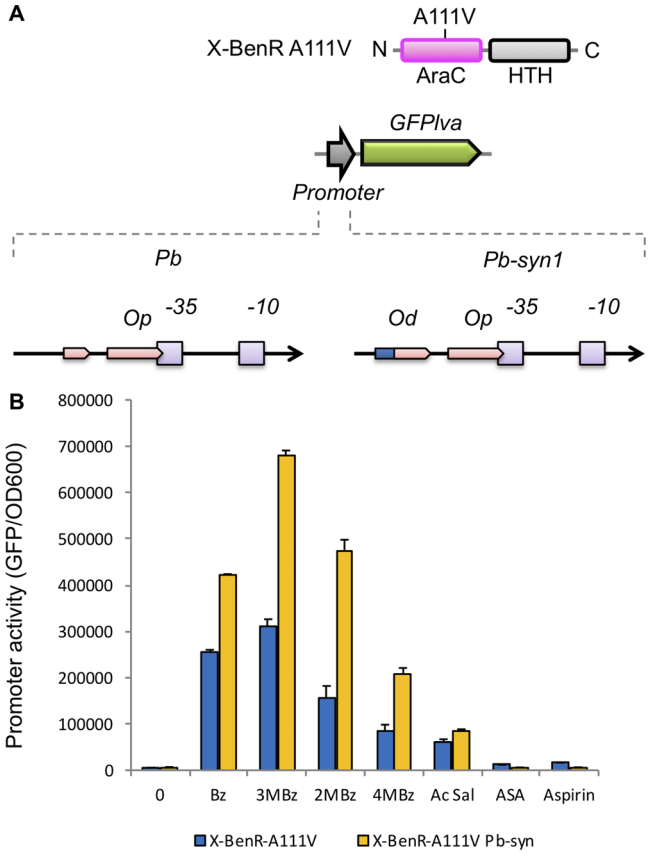


Figure 6. Effect of promoter architecture on the sensitivity of the X-BenR A111V regulator. (A) Schematic representation of the assayed system, where a wild-type or a synthetic *Pb* promoter[27] harboring a completed *Op* binding site has been used to control GFP expression. (B) Analysis of promoter activity for the X-BenR A111V mutant in response to benzoate (Bz), 3-methylbenzoate (3MBz), 2-methylbenzoate (2MBz), 4-methylbenzoate (4MBz), salicylic acid (Ac Sal), acetyl salicylic acid (ASA), and aspirin. All inducers were used at 100 μM. Vertical bars are the standard deviations from three independent experiments.

DISCUSSION

The results presented here shed some light on the molecular mechanisms for ligand and promoter recognition of BenR and XylS from *P. putida*. Of particular interest, XylS has been extensively investigated both in the context of the natural regulation of the meta pathways in *P. putida* mt-2[37−39] as well as for its applicability as a universal expression system for Gram-negative bacteria.[12,40,41] Previous attempts have investigated the critical aa for inducer recognition and promoter activation,[12,32,42] but these studies have not provided a clear molecular proposition on how this protein interacts with its ligands or the promoter. On the contrary, fewer studies have investigated BenR at the molecular level.[26,27,43] As for the findings presented here, we initially expect that changes in the aa of the identified binding pocket of BenR could adjust the ligand selectivity of the protein, as reported for the TtgV regulator that can discriminate between molecules with one or two aromatic rings.[44,45] Yet, contrary to the initial size exclusion model proposed, our experimental validation of the binding site prediction supports a role for position 111 in both BenR and XylS as a key element connecting the binding of the ligand to the changes in domain arrangement of the protein, similarly to the mechanism of activation of AraC of *E. coli*.[46] In this sense, the modeling of BenR with both N- and C-terminal domains suggests that residue 111 is close to the interface between the ligand binding and DNA binding domains (Figures S7, S8, and S10). In this scenario, valine could make critical interactions in BenR that are disrupted when this aa is changed for alanine. In the same way, changing alanine in XylS (or in X-BenR) for valine would allow the establishment of novel interactions that could enhance the performance of the protein, allowing the recognition of novel inducers such as salicylate or ASA. Previous work aiming at the construction of a XylS-*Pm* expression cassette with an increased response to 3MBz has reported several mutations in both the N- and C-terminal regions of XylS. Among these mutants, an A111V mutation has been reported before, also based on the 3D modeling of XylS, suggesting that this A111 residue interacts

DOI: 10.1021/acssynbio.9b00191
ACS Synth. Biol. 2019, 8, 1890−1900

Table 1. Strains and Plasmids Used in This Work

strains and plasmids	description	reference
Strains		
P. putida KT2440	*P. putida* mt-2 derivative	59
E. coli DH5α	F- φ80 ΔlacZ ΔM15 Δ(lacZYAA-argF) U169 recA1 endA1 hsdR17 R-M+ supE4 thi gyrA relA	50
Plasmids		
pSEVA438	Sm/Sp^R, ori pBBR1; Expression vector harboring the *xylS-Pm* expression system	33
pSEVA226	Km^R, ori RK2; reporter vector harboring the *luxCDABE* operon	33
pSEVA226-*Pb*	Km^R, ori RK2; pSEVA226 with the *Pb* promoter cloned as a *Eco*RI/*Bam*HI fragment	27
pSEVA226-*Pm*	Km^R, ori RK2; pSEVA226 with the *Pm* promoter cloned as a *Eco*RI/*Bam*HI fragment	27
pSEVA226-*Pbsyn1*	Km^R, ori RK2; pSEVA226 with the *Pbsyn1* promoter cloned as a *Eco*RI/*Bam*HI fragment	27
pMR1	Cm^R, ori p15a; dual mCherry/GFPlva promoter probe vector	51
pMR1-BenR-Pb	Cm^R, ori p15a; pMR1 variant with *benR-Pb* expression system closed as a *Bgl*II/*Eco*RI fragment	this study
pMR1-BenR-V111A	Cm^R, ori p15a; pMR1-BenR-Pb with *benR* gene mutated at position V111A	this study
pMR1-BenR-P110A	Cm^R, ori p15a; pMR1-BenR-Pb with *benR* gene mutated at position P110A	this study
pMR1-BenR-P110Q	Cm^R, ori p15a; pMR1-BenR-Pb with *benR* gene mutated at position P110Q	this study
pMR1-BenR-A110A111	Cm^R, ori p15a; pMR1-BenR-Pb with *benR* gene mutated at positions P110A and V111A	this study
pMR1-BenR-Q110A111	Cm^R, ori p15a; pMR1-BenR-Pb with *benR* gene mutated at positions P110Q and V111A	this study
pMR1-X-BenR	Cm^R, ori p15a; pMR1-BenR-Pb with *benR* with the N-terminal region replaced by that of *xylS*	this study
pMR1- X-BenR-A111V	Cm^R, ori p15a; pMR1-X-BenR with chimeric X-*benR* gene mutated at position A111V	this study
pMR1-XylS-*Pm*	Cm^R, ori p15a; pMR1 variant with *xylS-Pm* expression system closed as a *Bgl*II/*Eco*RI fragment	this study
pMR1-XylS-*Pm*-A111V	Cm^R, ori p15a; pMR1-XylS-*Pm* with *xylS* gene mutated at position A111V	this study
pMR1-B-XylS	Cm^R, ori p15a; pMR1-XylS-*Pm* with *xylS* with the N-terminal region replaced by that of *benR*	this study

with the effector binding pocket. However, in this work, authors performed the A111V mutation together with one or two additional aa positions.[12] Therefore, the role of position 111 in XylS has never been investigated in isolation. It was interesting to notice that changing the promoter architecture of *Pb* allowed the increased responsivity of the chimeric X-BenR-A111V to benzoate analogues, yet this strategy was not able to improve the responsivity to ASA or aspirin. This could indicate that the responsivity to these two compounds would require some additional interactions with the TF that are possible only in XylS, and this should be further investigated in the future. Additionally, most of the previously reported mutations affecting the XylS inducer response specificity are located close to the predicted binding pocket identified in this work (Figure S9), increasing the confidence of the computational approach used here.

It is interesting to notice that the mutations affecting the signal specificity of BenR and XylS either completely abolish the protein activity or generate regulators with an enhanced response to a series of ligands. In other words, it was not possible to switch the specificity of the ligand-binding domains from one compound to another. This notion resembles the stem protein model investigated for XylR (another aromatic responsive regulator from *P. putida* mt-2), where the selection of mutant proteins responsive to new ligands resulted in variants promiscuous to several aromatic compounds.[47] It is also important to notice that the computational approach used here predicted the ligand binding pocket site together with the conservation analysis of phylogenetically related protein homologues, which represents a powerful tool to guide the rational design of TF variants. Similar approaches could be applied to other members of the AraC/XylS family of TFs as well as to regulators from different families, aiming at the generation of novel expression systems for inducers of interest. Additionally, recent approaches based on the construction and high-throughput characterization of chimeric proteins have been used to create new benzoate responsive regulators. Yet this approach has generated very modest induction levels for the final engineered proteins (with approximately three-fold changes), whereas the approach used here yielded an induction of about 40 times for the mutated version of XylS.[48] Whereas NahR, a LysR-type transcriptional regulator, is able to induce gene expression in response to salicylate,[10] the TFs engineered here represent a new set of tools for the expression of genes of interest in response to salicylate and ASA. Additionally, the expression system based on XylS-A111V displays a ∼10-fold change in response to 10 μM of ASA, which is in the same range of the observed sensitivity for the natural ASA-responsive regulator NahR (which reaches a 20-fold change in response to a similar concentration of the compound[10]). These concentrations are in the range of the physiological concentrations of ASA in blood, as this molecules can reach levels as high as ∼30 μM after 20 min of administration of the drug.[49] Yet, whereas the new aspirin-responsive regulator presented here is not specific to this compound, it is reasonable to think that during real applications (i.e., *in vivo* in a mammalian cell model), systems would not be exposed to benzoate or any of its analogues. Therefore, the lack of exclusive responsivity to aspirin would not be an issue under real case applications. Taken together, these results demonstrate the expansion of the genetic toolbox for the engineering of synthetic circuits inducible to safe drugs.

MATERIALS AND METHODS

Bacterial Strains and Growth Conditions. The plasmids and bacterial strains used in this study are listed in Table 1. Cloning and assay procedures were performed in *E. coli* DH5α. All DNA manipulations, including cloning, polymerase chain reaction (PCR), and transformations of *E. coli*, were performed according Sambrook et al.[50] Bacterial strains were routinely grown in LB media supplemented with 36 μg mL^{-1} chloramphenicol or, when necessary, in M9 minimal media (6.4 g L^{-1} $Na_2HPO_4{\cdot}7H_2O$, 1.5 g L^{-1} KH_2PO_4, 0.25 g L^{-1} NaCl, 0.5 g L^{-1} NH_4Cl, 2 mM $MgSO_4$, 0.1 mM casamino acids, 1% glycerol) supplemented with chloramphenicol at 36 μg mL^{-1}. Liquid cultures were shaken at 180 rpm at 37 °C for

DOI: 10.1021/acssynbio.9b00191
ACS Synth. Biol. 2019, 8, 1890−1900

~16 h. The aromatic compounds used as inducers were all purchased from Sigma-Aldrich. The inducers and their catalogue numbers are benzoic acid (242381), 3-methylbenzoic acid (3117714), 2-methylbenzoic acid (169978), 4-methylbenzoic acid (117390), salicylic acid (S5922), ASA (A5376), Bayer aspirin (1000 μM) and isopropyl β-D-1-tiogalactopiranosida (IPTG) (I5502), sodium arsenite (S7400), toluene (244511), xylene (214736), phenol (P1037), L-arabinose (A3256), D-(−)-fructose (F0127), and D-(+)-glucose (G8270).

Plasmid Construction. The *benR* gene and *PbenR* and *Pb* promoters were amplified by PCR using specific primers (Table S1) and *P. putida* KT2440 genomic DNA as the template. The PCR products were digested with specific restriction enzymes (see the underlined sequences in Table S1) and cloned into the pMR1 vector,[51] yielding the pMR1-BenR-Pb (BenR) construct. BenR mutants, pMR1-BenR-P110A (BenR-P110A), pMR1-BenR-P110Q (BenR-P110Q), pMR1-BenR-V111A (BenR-V111A), pMR1-BenR-A110A111 (BenR-A110A111), and pMR1-BenR-Q110A111 (BenR-Q110A111), were generated by circular polymerase extension cloning (CPEC) site-directed mutagenesis methodology[52] using the pMR1-BenR-Pb construct as the template and the primers listed in Table S1. (Mutated base pairs are highlighted in bold and underlined.) The *xylS* gene and *Ps* and *Pm* promoters were amplified by PCR using pSEVA438 vector as the template,[33] yielding the pMR1-XylS-*Pm* (XylS) construct. XylS mutant pMR1-XylS-A111V (XylS-A111V) was constructed by CPEC site-directed mutagenesis using the pMR1-XylS-*Pm* construct as the template and the primers listed in Table S1. (Mutated base pairs are highlighted in bold and underlined.) Two chimeric transcription factors were constructed. The first construct was generated by directly linking the N-terminal domain of XylS and the C-terminal domain of BenR using, respectively, *P. putida* mt-2 and KT2440 strains as templates. The second construct was generated by linking the N-terminal domain of BenR and the C-terminal domain of XylS using the pMR1-BenR-*Pb* and the pMR1-XylS-*Pm* as templates. All fragments were amplified by PCR. In the first construct, the *PbenR* promoter was cloned upstream the chimaera, and the *Pb* promoter was cloned upstream the GFPlva reporter gene, yielding the pMR1-X-BenR (X-BenR). The chimaera mutant pMR1-X-BenR-A111V (X-BenR-A111V) was constructed using the vector pMR1-X-BenR as the template. In the second construct, the *Ps* and *Pm* promoters were cloned upstream the chimaera and the GFPlva reporter gene, respectively, generating the pMR1-B-XylS (B-XylS). All PCR amplifications were performed using Phusion high-fidelity DNA polymerase (Thermo Fisher Scientific). All resulting constructs were sequenced using dideosyterminal methods to confirm the correct assembly prior to the fluorescence assays. The aa sequences of the final construct generated here are represented in Table S2.

GFP Fluorescence Assay and Data Processing. To measure the activity of all constructions performed in this work, plasmids were transformed into *E. coli* DH5α. Freshly plated single colonies were grown in M9 minimal media supplemented with suitable antibiotics. The cultures (10 μL) were then assayed in a 96-well microplates with 170 μL of M9 minimal media and 20 μL of the different compounds tested. When required, benzoic acid (Bz), 3-methylbenzoic acid (3MBz), 2-methylbenzoic acid (2MBz), 4-methylbenzoic acid (4MBz), salicylic acid (0−1000 μM), ASA, Bayer aspirin (1000 μM) and IPTG, sodium arsenite, toluene, xylene, phenol, L-arabinose, fructose, and glucose (100 μM) were used. Cell growth and GFP fluorescence were quantified using a Victor X3 plate reader (PerkinElmer). The responsiveness of regulators was calculated as arbitrary units using the ratio between fluorescence levels and the optical density at 600 nm (reported as GFP/OD_{600}) or the luminescence by optical density at 600 nm after background correction. As a control, all assays were performed without the addition of compounds (inducers) as the threshold background signal during calculations. Fluorescence and absorbance measurements were taken at 30 min intervals up to 8 h at 37 °C. All experiments were performed in technical and biological triplicates. Raw data were processed using *ad hoc* R script (https://www.r-project.org/).

Flow Cytometry Analysis. High-throughput single-cell analysis of the bacteria carrying the BenR, XylS, XBenR, XylS mutant, or XBenR mutant systems was run as follows. First, we selected single colonies of the transformed strain (*E. coli* DH5α) and cultivated it overnight in M9 minimal medium (containing 6.4 g/L $Na_2HPO_4{\cdot}7H_2O$, 1.5 g/L KH_2PO_4, 0.25 g/L NaCl, and 0.5 g/L NH4Cl) supplemented with 2 mM $MgSO_4$, 0.1 mM CaCl2, 0.1 mM casamino acids, chloramphenicol (36 μg/mL), and 1% glycerol as the sole carbon source (supplemented M9) at 37 °C and 180 rpm. Next, overnight grown cells were diluted to 1:10 in fresh supplemented M9 and were grown for 3 h at 37 °C and 180 rpm. At this point, the cultures were induced with different concentrations (0, 10, 100, 250, 500, and 1000 μM) of the inducers. The BenR system was induced with benzoate, XylS was induced with benzoate and 3-metilbenzoate, X-BenR was induced with 3-metilbenzoate, and XylS mutant and X-BenR mutant were induced with benzoate, 3-metil-benzoate, salicylic acid, and acetyl-salicylic acid. After 3 h of induction, the cultures were stored in ice and immediately analyzed for GFP fluorescence using the Millipore Guava EasyCyte mini flow cytometer (Millipore). The results were analyzed by R scripts using the flowCore and flowViz packages available on Bioconductor (https://bioconductor.org/).

3D Structure Model Construction and Docking Analysis. The 3D models presented here were generated by SWISS-MODEL server (https://swissmodel.expasy.org/) using the best homologue for each protein. For the visualization of the models, PyMol (https://pymol.org/) and Chimaera (https://www.cgl.ucsf.edu/chimera/) were used. To predict the potential binding site for the aromatic ligands, Swiss-Docking (https://www.swissdock.ch/) and Docking Server (https://www.dockingserver.com/web) were used. Additionally, Jalview (http://www.jalview.org/) was used for the visualization of protein and DNA alignments generated by T-coffee (http://tcoffee.crg.cat/apps/tcoffee/all.html). From the generated 3D models using SwissModel,[53] models were first inspected visually for histidine protonation and Asn and Gln positioning. Models were further refined by minimizing their structure with the AMBER FF14SB[54] force field using AMBER.[55]

Molecular Docking. The energy maps were computed inside the grid, with a grid size of 26 × 30 × 40. Dockings were carried out using the LGA/LS algorithm implemented on Autodock 4[56] (version 4.2.6, with a maximum of 27 000 generations or 2 500 000 energy evaluations). Fifty independent runs were performed, and the resulting poses were clustered according to the ligand heavy-atom rmsd using a

DOI: 10.1021/acssynbio.9b00191
ACS Synth. Biol. 2019, 8, 1890−1900

cutoff of 2 Å, thus defining a result. 3D structures of the ligands were generated using OpenBabel[57] and manually inspected. Images were created with VMD.[58]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.9b00191.

Protein sequences reported here and Figures S1−S10 and Tables S1−S3 (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel: +55 16 3602 3107. Fax: +55 16 3633 6840. E-mail: silvarochar@usp.br.

ORCID

Lucas Defelipe: 0000-0001-7859-7300
Víctor de Lorenzo: 0000-0002-6041-2731
Rafael Silva-Rocha: 0000-0001-6319-631X

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the lab members for insightful discussion on this work. We thank Gabriel Lencione Lovate for help with acquiring some chemicals used in this work. R.S.-R. and M.-E.G. were supported by Young Research Awards, grant nos. 2012/22921-8 and 2015/04309-1, São Paulo Research Foundation (FAPESP). L.M.O.M., A.S.-M., L.M.-S., and L.F.A. were supported by FAPESP Ph.D. Fellowships (grant nos. 2016/19179-9, 2016/06323-4, 2017/17924-1, and 2018/04810-0).

■ REFERENCES

(1) Browning, D. F., and Busby, S. J. W. (2016) Local and Global Regulation of Transcription Initiation in Bacteria. *Nat. Rev. Microbiol. 14*, 638−650.

(2) Galán-Vásquez, E., Sánchez-Osorio, I., and Martínez-Antonio, A. (2016) Transcription Factors Exhibit Differential Conservation in Bacteria with Reduced Genomes. *PLoS One 11* (1), e0146901.

(3) Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., Ledezma-Tejeida, D., Muñiz-Rascado, L., García-Sotelo, J. S., Alquicira-Hernández, K., Martínez-Flores, I., Pannier, L., Castro-Mondragón, J. A., Medina-Rivera, A., Solano-Lira, H., Bonavides-Martínez, C., Pérez-Rueda, E., Alquicira-Hernández, S., Porrón-Sotelo, L., López-Fuentes, A., Hernández-Koutoucheva, A., Del Moral-Chávez, V., Rinaldi, F., and Collado-Vides, J. (2016) RegulonDB Version 9.0: High-Level Integration of Gene Regulation, Coexpression, Motif Clustering and Beyond. *Nucleic Acids Res. 44*, D133−D143.

(4) Steinacher, A., Bates, D. G., Akman, O. E., and Soyer, O. S. (2016) Nonlinear Dynamics in Gene Regulation Promote Robustness and Evolvability of Gene Expression Levels. *PLoS One 11* (4), e0153295.

(5) Baumstark, R., Hänzelmann, S., Tsuru, S., Schaerli, Y., Francesconi, M., Mancuso, F. M., Castelo, R., and Isalan, M. (2015) The Propagation of Perturbations in Rewired Bacterial Gene Networks. *Nat. Commun. 6*, 10105.

(6) Mannan, A. A., Liu, D., Zhang, F., and Oyarzun, D. A. (2017) Fundamental Design Principles for Transcription-Factor-Based Metabolite Biosensors. *ACS Synth. Biol. 6* (10), 1851−1859.

(7) Lee, J. W., Gyorgy, A., Cameron, D. E., Pyenson, N., Choi, K. R., Way, J. C., Silver, P. A., Del Vecchio, D., and Collins, J. J. (2016) Creating Single-Copy Genetic Circuits. *Mol. Cell 63* (2), 329−336.

(8) Nielsen, A. a K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. a, Ross, D., Densmore, D., and Voigt, C. a. (2016) Genetic Circuit Design Automation. *Science 352* (6281), aac7341.

(9) Unkles, S. E., Valiante, V., Mattern, D. J., and Brakhage, A. A. (2014) Synthetic Biology Tools for Bioprospecting of Natural Products in Eukaryotes. *Chem. Biol. 21* (4), 502−508.

(10) Shin, H. J. (2010) Development of Highly-Sensitive Microbial Biosensors by Mutation of the NahR Regulatory Gene. *J. Biotechnol. 150* (2), 246−250.

(11) Garmendia, J., de las Heras, A., Galvao, T. C., and de Lorenzo, V. (2008) Tracing Explosives in Soil with Transcriptional Regulators of Pseudomonas Putida Evolved for Responding to Nitrotoluenes. *Microb. Biotechnol. 1* (3), 236−246.

(12) Vee Aune, T. E., Bakke, I., Drablos, F., Lale, R., Brautaset, T., and Valla, S. (2010) Directed Evolution of the Transcription Factor XylS for Development of Improved Expression Systems. *Microb. Biotechnol. 3* (1), 38−47.

(13) Bates, D. M., Popescu, C. V., Khoroshilova, N., Vogt, K., Beinert, H., Munck, E., and Kiley, P. J. (2000) Substitution of Leucine 28 with Histidine in the Escherichia Coli Transcription Factor FNR Results in Increased Stability of the [4Fe-4S](2+) Cluster to Oxygen. *J. Biol. Chem. 275* (9), 6234−6240.

(14) Chen, J. X., Steel, H., Wu, Y. H., Wang, Y., Xu, J., Rampley, C. P. N., Thompson, I. P., Papachristodoulou, A., and Huang, W. E. (2019) Development of Aspirininducible Biosensors in Escherichia Coli and SimCells. *Appl. Environ. Microbiol. 85*, No. e02959-18.

(15) Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J., and Voigt, C. A. (2019) Escherichia Coli "Marionette" Strains with 12 Highly Optimized Small-Molecule Sensors. *Nat. Chem. Biol. 15*, 196.

(16) Shis, D. L., Hussain, F., Meinhardt, S., Swint-Kruse, L., and Bennett, M. R. (2014) Modular, Multi-Input Transcriptional Logic Gating with Orthogonal LacI/GalR Family Chimeras. *ACS Synth. Biol. 3* (9), 645−651.

(17) Younger, A. K., Dalvie, N. C., Rottinghaus, A. G., and Leonard, J. N. (2017) Engineering Modular Biosensors to Confer Metabolite-Responsive Regulation of Transcription. *ACS Synth. Biol. 6* (2), 311−325.

(18) Stanton, B. C., Nielsen, A. A., Tamsir, A., Clancy, K., Peterson, T., and Voigt, C. A. (2014) Genomic Mining of Prokaryotic Repressors for Orthogonal Logic Gates. *Nat. Chem. Biol. 10* (2), 99−105.

(19) Libis, V., Delepine, B., and Faulon, J. L. (2016) Sensing New Chemicals with Bacterial Transcription Factors. *Curr. Opin. Microbiol. 33*, 105−112.

(20) Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L. A. (2013) Programmable Repression and Activation of Bacterial Gene Expression Using an Engineered CRISPR-Cas System. *Nucleic Acids Res. 41* (15), 7429−7437.

(21) Bikard, D., and Marraffini, L. A. (2013) Control of Gene Expression by CRISPR-Cas Systems. *F1000Prime Rep. 5*, 47.

(22) Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013) Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell 152* (5), 1173−1183.

(23) Libis, V., Delepine, B., and Faulon, J. L. (2016) Expanding Biosensing Abilities through Computer-Aided Design of Metabolic Pathways. *ACS Synth. Biol. 5* (10), 1076−1085.

(24) de los Santos, E. L., Meyerowitz, J. T., Mayo, S. L., and Murray, R. M. (2016) Engineering Transcriptional Regulator Effector Specificity Using Computational Design and In Vitro Rapid Prototyping: Developing a Vanillin Sensor. *ACS Synth. Biol. 5* (4), 287−295.

(25) Royo, J. L., Becker, P. D., Camacho, E. M., Cebolla, A., Link, C., Santero, E., and Guzman, C. A. (2007) In Vivo Gene Regulation in Salmonella Spp. by a Salicylate-Dependent Control Circuit. *Nat. Methods 4* (11), 937−942.

(26) Perez-Pantoja, D., Kim, J., Silva-Rocha, R., and de Lorenzo, V. (2015) The Differential Response of the Pben Promoter of

DOI: 10.1021/acssynbio.9b00191
ACS Synth. Biol. 2019, 8, 1890−1900

Pseudomonas Putida Mt-2 to BenR and XylS Prevents Metabolic Conflicts in m-Xylene Biodegradation. *Environ. Microbiol. 17*, 64.
(27) Silva-Rocha, R., and de Lorenzo, V. (2012) Broadening the Signal Specificity of Prokaryotic Promoters by Modifying Cis-Regulatory Elements Associated with a Single Transcription Factor. *Mol. BioSyst. 8* (7), 1950−1957.
(28) Cowles, C. E., Nichols, N. N., and Harwood, C. S. (2000) BenR, a XylS Homologue, Regulates Three Different Pathways of Aromatic Acid Degradation in Pseudomonas Putida. *J. Bacteriol. 182* (22), 6339−6346.
(29) Xue, H., Shi, H., Yu, Z., He, S., Liu, S., Hou, Y., Pan, X., Wang, H., Zheng, P., Cui, C., Viets, H., Liang, J., Zhang, Y., Chen, S., Zhang, H. M., and Ouyang, Q. (2014) Design, Construction, and Characterization of a Set of Biosensors for Aromatic Compounds. *ACS Synth. Biol. 3* (12), 1011−1014.
(30) Tobes, R., and Ramos, J. L. (2002) AraC-XylS Database: A Family of Positive Transcriptional Regulators in Bacteria. *Nucleic Acids Res. 30* (1), 318−321.
(31) Kessler, B., de Lorenzo, V., and Timmis, K. N. (1993) Identification of a Cis-Acting Sequence within the Pm Promoter of the TOL Plasmid Which Confers XylS-Mediated Responsiveness to Substituted Benzoates. *J. Mol. Biol. 230* (3), 699−703.
(32) Domínguez-Cuevas, P., Marín, P., Marqués, S., and Ramos, J. L. (2008) XylS-Pm Promoter Interactions through Two Helix-Turn-Helix Motifs: Identifying XylS Residues Important for DNA Binding and Activation. *J. Mol. Biol. 375* (1), 59−69.
(33) Silva-Rocha, R., Martínez-García, E., Calles, B., Chavarría, M., Arce-Rodríguez, A., de las Heras, A., Páez-Espino, A. D., Durante-Rodríguez, G., Kim, J., Nikel, P. I., Platero, R., and de Lorenzo, V. (2013) The Standard European Vector Architecture (SEVA): A Coherent Platform for the Analysis and Deployment of Complex Prokaryotic Phenotypes. *Nucleic Acids Res. 41* (D1), D666−D675.
(34) Silva-Rocha, R., and de Lorenzo, V. (2012) Stochasticity of TOL Plasmid Catabolic Promoters Sets a Bimodal Expression Regime in Pseudomonas Putida Mt-2 Exposed to m-Xylene. *Mol. Microbiol. 86*, 199.
(35) Silva-Rocha, R., and de Lorenzo, V. (2012) A GFP-LacZ Bicistronic Reporter System for Promoter Analysis in Environmental Gram-Negative Bacteria. *PLoS One 7* (4), e34675.
(36) Michan, C., Zhou, L., Gallegos, M. T., Timmis, K. N., and Ramos, J. L. (1992) Identification of Critical Amino-Terminal Regions of Xy1S. The Positive Regulator Encoded by the TOL Plasmid. *J. Biol. Chem. 267* (32), 22897−22901.
(37) Moreno, R., Fonseca, P., and Rojo, F. (2010) The Crc Global Regulator Inhibits the Pseudomonas Putida PWW0 Toluene/Xylene Assimilation Pathway by Repressing the Translation of Regulatory and Structural Genes. *J. Biol. Chem. 285* (32), 24412−24419.
(38) Gonzalez-Perez, M. M., Ramos, J. L., and Marques, S. (2004) Cellular XylS Levels Are a Function of Transcription of XylS from Two Independent Promoters and the Differential Efficiency of Translation of the Two MRNAs. *J. Bacteriol. 186* (6), 1898−1901.
(39) Marques, S., Manzanera, M., Gonzalez-Perez, M. M., Gallegos, M. T., and Ramos, J. L. (1999) The XylS-Dependent Pm Promoter Is Transcribed in Vivo by RNA Polymerase with Sigma 32 or Sigma 38 Depending on the Growth Phase. *Mol. Microbiol. 31* (4), 1105−1113.
(40) Blatny, J. M., Brautaset, T., Winther-Larsen, H. C., Karunakaran, P., and Valla, S. (1997) Improved Broad-Host-Range RK2 Vectors Useful for High and Low Regulated Gene Expression Levels in Gram-Negative Bacteria. *Plasmid 38* (1), 35−51.
(41) Blatny, J. M., Brautaset, T., Winther-Larsen, H. C., Haugan, K., and Valla, S. (1997) Construction and Use of a Versatile Set of Broad-Host-Range Cloning and Expression Vectors Based on the RK2 Replicon. *Appl. Environ. Microbiol. 63* (2), 370−379.
(42) Dominguez-Cuevas, P., Marin, P., Busby, S., Ramos, J. L., and Marques, S. (2008) Roles of Effectors in XylS-Dependent Transcription Activation: Intramolecular Domain Derepression and DNA Binding. *J. Bacteriol. 190* (9), 3118−3128.
(43) Moreno, R., and Rojo, F. (2008) The Target for the Pseudomonas Putida Crc Global Regulator in the Benzoate Degradation Pathway Is the BenR Transcriptional Regulator. *J. Bacteriol. 190* (5), 1539−1545.
(44) Guazzaroni, M. E., Gallegos, M. T., Ramos, J. L., and Krell, T. (2007) Different Modes of Binding of Mono- and Biaromatic Effectors to the Transcriptional Regulator TTGV: Role in Differential Derepression from Its Cognate Operator. *J. Biol. Chem. 282* (22), 16308−16316.
(45) Guazzaroni, M. E., Krell, T., Felipe, A., Ruiz, R., Meng, C., Zhang, X., Gallegos, M. T., and Ramos, J. L. (2005) The Multidrug Efflux Regulator TtgV Recognizes a Wide Range of Structurally Different Effectors in Solution and Complexed with Target DNA: Evidence from Isothermal Titration Calorimetry. *J. Biol. Chem. 280* (21), 20887−20893.
(46) 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* (5), 779−796.
(47) Galvao, T. C., Mencia, M., and de Lorenzo, V. (2007) Emergence of Novel Functions in Transcriptional Regulators by Regression to Stem Protein Types. *Mol. Microbiol. 65* (4), 907−919.
(48) Juárez, J. F., Lecube-Azpeitia, B., Brown, S. L., Johnston, C. D., and Church, G. M. (2018) Biosensor Libraries Harness Large Classes of Binding Domains for Construction of Allosteric Transcriptional Regulators. *Nat. Commun. 9*, 3101.
(49) Cotty, V. F., Sterbenz, F. J., Mueller, F., Melman, K., Ederma, H., Skerpac, J., Hunter, D., and Lehr, M. (1977) Augmentation of Human Blood Acetylsalicylate Concentrations by the Simultaneous Administration of Acetaminophen with Aspirin. *Toxicol. Appl. Pharmacol. 41* (1), 7−13.
(50) Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York.
(51) Guazzaroni, M.-E., and Silva-Rocha, R. (2014) Expanding the Logic of Bacterial Promoters Using Engineered Overlapping Operators for Global Regulators. *ACS Synth. Biol. 3*, 666−675.
(52) Quan, J., and Tian, J. (2011) Circular Polymerase Extension Cloning for High-Throughput Cloning of Complex and Combinatorial DNA Libraries. *Nat. Protoc. 6* (2), 242−251.
(53) Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T. G., Bertoni, M., Bordoli, L., and Schwede, T. (2014) SWISS-MODEL: Modelling Protein Tertiary and Quaternary Structure Using Evolutionary Information. *Nucleic Acids Res. 42*, W252.
(54) Maier, J. A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K. E., and Simmerling, C. (2015) Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput. 11*, 3696.
(55) Case, D. A., Betz, R. M., Cerutti, D. S., Cheatham, T. E., Darden, T. A., Duke, R. E., Giese, T. J., Gohlke, H., Goetz, A. W., Homeyer, N., Izadi, S., Janowski, P., Kaus, J., Kovalenko, A., Lee, T. S., LeGrand, S., Li, P., Lin, C., Luchko, T., Luo, R., Madej, B., Mermelstein, D., Merz, K. M., Monard, G., Nguyen, H., Nguyen, H. T., Omelyan, I., Onufriev, A., Roe, D. R., Roitberg, A., Sagui, C., Simmerling, C. L., Botello-Smith, W. M., Swails, J., Walker, R. C., Wang, J., Wolf, R. M., Wu, X., Xiao, L., and Kollman, P. A. (2018) *AMBER 2018*; University of California, San Francisco, CA.
(56) Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., and Olson, A. J. (2009) Software News and Updates AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem. 30*, 2785.
(57) O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., and Hutchison, G. R. (2011) Open Babel: An Open Chemical Toolbox. *J. Cheminf. 3*, 33.
(58) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual Molecular Dynamics. *J. Mol. Graphics 14*, 33.
(59) Bagdasarian, M., Lurz, R., Ruckert, B., Franklin, F. C., Bagdasarian, M. M., Frey, J., and Timmis, K. N. (1981) Specific-Purpose Plasmid Cloning Vectors. II. Broad Host Range, High Copy Number, RSF1010-Derived Vectors, and a Host-Vector System for Gene Cloning in Pseudomonas. *Gene 16* (1−3), 237−247.

DOI: 10.1021/acssynbio.9b00191
ACS Synth. Biol. 2019, 8, 1890−1900

CHAPTER IV

PredicTF: a tool to predict bacterial transcription factors in complex microbial communities

This chapter was submmited as:

MONTEIRO, Lummy Maria Oliveira; SARAIVA, João; TOSCAN, Rodolfo Brizola; STADLER, Peter; SILVA-ROCHA, Rafael; da ROCHA, Ulisses Nunes. PredicTF: a tool to predict bacterial transcription factors in complex microbial communities.

PredicTF: a tool to predict bacterial transcription factors in complex microbial communities.

Lummy Maria Oliveira Monteiro[1,2,3], Joao Saraiva[1], Rodolfo Brizola Toscan[1], Peter F Stadler[2], Rafael Silva-Rocha[3], Ulisses Nunes da Rocha[1]*

[1] Helmholtz Center for Environmental Research (UFZ), Leipzig, Germany

[2] Universität Leipzig (UL), Leipzig, Germany

[3] Ribeirão Preto Medical School (FMRP), University of São Paulo (USP), Ribeirão Preto, Brazil

*Correspondence: Ulisses Nunes da Rocha, ulisses.rocha@ufz.de

Authors' email addresses:

Lummy Monteiro – lummymaria@gmail.com

João Saraiva - joao.saraiva@ufz.de

Rodolfo Brizola Toscan - rodolfo.toscan@ufz.de

Peter F. Stadler - studla@bioinf.uni-leipzig.de

Rafael Silva-Rocha - silvarochar@usp.br

Ulisses Nunes da Rocha – ulisses.rocha@ufz.de

Abstract

Background: Transcription Factors (TFs) are proteins controlling the rate of genetic information and regulating cellular gene expression. A better understanding of TFs in a bacterial community context may open novel revenues for exploring gene regulation in ecosystems where bacteria play a key role. Here we describe PredicTF, the first platform supporting the prediction and classification of novel bacterial TF not only in single species but also in complex microbial communities.

Results: We created our TF database (BacTFDB) by merging data from CollectTF and UniProt. The TF sequences from these two databases were merged and manually curated generating a robust bacterial TF database with 11.961 TFs sequences distributed in 99 TF families. BacTFDB was used to train a deep learning model to predict novel TFs and their families. Five model organisms were used to test the performance and the accuracy of PredicTF. PredicTF was able to identify 27 to 60% of the known TFs with an accuracy of 73 to 91% in our five model organisms. We further evaluated PredicTF using a two-step approach. First, we tested PredictTF's ability to predict TFs for the genome of an environmental isolate of *P. aeruginosa* PAO1 followed by the mapping of these TFs in the transcriptomic data for three clinical isolates. We were able to map 69 of the 199 predicted TFs to the transcriptomes. This result demonstrates the potential of PredicTF to map and compare TFs profiles under different environmental conditions. In the second evaluation step, PredicTF was used to predict TFs in a metagenome recovered from a community performing anaerobic ammonium oxidation (anammox) in a bioreactor. A total of 792 TFs were predicted in this community. These 792 TFs were mapped in 11 metatranscriptomes of reference from the same bioreactor where the metagenome was collected allowing the comparison of profiles of regulators expressed in different environmental situations.

Conclusion: PredicTF provides the first tool to profile TFs in yet-to be cultured bacteria and it opens the potential to evaluate regulatory networks in complex microbial communities. PredicTF is a flexible, open source pipeline able to predict and annotate TFs in genomes and metagenomes. PredicTF is available at https://github.com/mdsufz/PredicTF.

Keywords: Gene regulation, Transcription factors, Deep Learning, Transcription factor database, Microbial Communities

Background

The functional potential of microbial communities can be determined by the genetic content of its constituent members. However, genetic content alone does not guarantee that a given function or enzymatic reaction will be performed [1]. In this scenario, Transcription Factor proteins (TFs) play a central and critical role in gene regulation. These proteins are responsible for optimizing proteins and structural RNAs and the subsequent levels of metabolites and other properties, ensuring the survival and adaptation of organisms to the most diverse types of stress and environmental changes [2]. The activity of bacterial TFs is modulated by environmental signals (e.g. changes in the oxygen condition, temperature, pH or the lack of a specific substrate) [3]. Additionally, for many promoters, combinations of transcription factors work together to integrate different signals [2,4]. TFs can also work with other DNA-binding proteins whose primary role is to sculpt the bacterial folded chromosome [2,5]. Knowledge of the TFs profile expressed by an organism is the first step to better understand the regulatory network that controls protein expression in an organism or a community.

Since TFs may determine when and which genes are expressed, profiling TFs can help understand the regulation of gene expression and to build regulatory networks in complex microbial communities. Further, defining which factors control gene expression may offer insights into the mechanisms controlling ecosystem processes and even interactions between species of a microbial community. However, current TF databases are focused on single or small groups of genomes. These databases are largely manually curated based on literature evidence and pairwise sequence comparison of genomes from model organisms. Examples of these databases include RegulonDB for *Escherichia coli* K-12 [6], DBTBS for *Bacillus subtilis* [7], FlyBase for *Drosophila* [8], and FTFD for fungal species [9]. DBD [10], is a database generated from the prediction of TFs from 150 sequenced genomes from across the tree of life. Unfortunately, DBD has has not been updated for more than 9 years.

One of the major goals in the manipulation of microbiomes for ecological and biotechnological applications is to control the outcome of their functions [11]. As TFs are key to potentially control which genes are expressed, one of the best ways to study and understand gene regulation in a microbiome may be to profile its TFs. To date, no platform supports prediction and classification of novel bacterial TF from ‘omics data recovered from microbial communities.

Deep Learning approaches have been used to predict DNA sequence affinities [12] and to identify TF-binding sites in humans [13]. Although deep learning has been used in gene regulation, it has never been used to predict bacterial TFs. Further, the need for a user-friendly tool for prediction of TFs that could assist in gene regulation analysis motivated the development of PredicTF. PredicTF is a deep learning tool used to predict and identify TFs from full protein-length sequences. Further, we constructed a robust

database for bacterial transcriptional factors (BacTFDB) that was used to train our Deep Learning model.

Implementation

PredictTF is a command line software for prediction of novel transcription factors from genomic and metagenomic data. We created a bacterial transcription factor database (BacTFDB) by merging and manually curating TFs present in CollectTF [14] and the Universal Protein Resource (UniProt) [15]. CollectTF provides well described and characterized, *in vivo* validated, TFs while UniProt is a comprehensive resource for protein sequence and annotation data. We used BacTFDB to train a deep learning model to predict new TFs and their families in genomes and metagenomes. Five model organisms (*Escherichia coli, Bacillus subtillis, Pseudomonas fluorescens, Azotobacter vinelandii and Caulobacter crescentus*) were used to test the performance and accuracy of PredicTF. We used the same approach to predict TFs from a clinical isolate (*P. aeruginosa* PAO1) and a metagenome sample isolated from an anaerobic ammonium oxidation community. We also determined if the predicted TFs were expressed in transcriptomes (isolate) and metatranscriptomes (microbial community), respectively (Fig. 1).

BacTFDB - Bacterial Transcription Factor Data Base

To create a novel Bacterial Transcription Factor Data Base (BacTFDB), we collected data from two publicly available databases. Initially, we chose to collect data from CollecTF [14], a well described and characterized database. Since CollecTF does not provide an application programming interface (API) for bulk download, we developed a Python code (version 2.7) using the Beautiful Soup 4.4.0 library to recover the data

from CollecTF. With this strategy we listed 390 TF experimentally validated amino acid sequences distributed over 44 TF families. The script can be found at https://github.com/mdsufz/PredicTF.

Additionally, we retrieved TF amino acid sequences from UniProt using UniProt's API. We downloaded sequences of interest by adding a filter with the key words (Transcription factor, transcriptional factor, regulator, transcriptional repressor, transcriptional activator, transcriptional regulator). After, we filtered for Reviewed (Swiss-Prot) - Manually annotated sequences that belonged to the bacteria taxonomy. The UniProt API was accessed on 8th September-2019 and a total of 21.581 TF amino acid sequences, with applied filters, were collected. We merged the data collected from CollecTF and UniProt databases which resulted in a total of 21.971 TFs. Next, we removed redundant TF entries and TF sequences lacking a TF family since PredicTF was designed to also assign TF family. Finally, a manual inspection was performed to remove case sensitive and presence of characters associated to the database header. The first version of BacTFDB contains a total of 11.691 unique TF sequences. A summary of the information contained in BacTFDB can be found in the supplementary data (Additional file 1: Fig. S1). To evaluate PredicTF in model organisms we created 5 subsets of BacTFDB. The description of these subsets can be found in the supplementary data (Additional file 2: Table S1).

Mapping Transcription Factors using PredictTF

We used a deep learning approach similar to that found in DeepARG [16]. Supervised machine learning models are usually divided into characterization, training, and prediction units. Briefly, our approach uses the concept of dissimilarity-based classification [17] where sequences are represented and featured by their identity

distances to known genes. In PredicTF, sequences were represented and featured by their identity distances to known TFs families. The BacTFDB was used to train and test the deep learning model (https://github.com/mdsufz/PredicTF) and the latter validated in model organisms. Next, PredicTF was used to predict novel TFs from full protein-length sequences in genomes and in one metagenome. After prediction, the data was mapped in transcriptomes and metatranscriptomes from samples where the genetic potential was determined.

Using PredicTF, we trained five different models – one for each model organism (Additional file 3: Table S2). For each model, the TFs affiliated with the respective model organism were removed prior to training to avoid overfitting. PredicTF-no-coli was trained to predict TFs in *E. coli,* PredicTF-no-subtilis was trained to predict TFs in *B. subtilis,* PredicTF-no-crescentus was trained to predict TFs in *C. crescentus*, PredicTF-no-fluorescens was trained to predict TFs in *P. fluorescens* and PredicTF-no-vinelandii was trained to predict TFs in *A. vinelandii*.

Performance and accuracy calculation

We evaluated PredicTF by calculating accuracy and performance. Performance can be deemed to be the fulfillment of a task. In PredicTF case, performance is how good TF predictions are. Using model organisms (see later in the session *Prediction of Transcription Factors in model organisms*), performance was calculated by quantifying the number of TFs that PredicTF was able to predict divided by number of TFs already described and annotated for our model organisms (Equation 1). Accuracy indicates how correct the predictions performed by PredicTF are. Also using data of model organism, accuracy was determined by calculating the number of TFs correctly predicted divided by the total number of TFs predicted by PredicTF. We divided accuracy in two categories.

In the first accuracy category, we determined accuracy against experimentally validated TFs (Equation 2). In the second accuracy category, we determined accuracy against TFs without experimental validation (Equation 3); i.e., putative TFs. The performance, accuracy, and accuracy for putative TFs were calculated as follows:

Equation 1

$$Performance(\%) = \frac{Predicted\ TFs * 100}{Annotated\ TFs}$$

where, *Performance (%)* is calculated by the ratio of the total number of TFs predicted by PredicTF (*Predicted TFs*) to the total number of proteins annotated as TFs in NCBI (*Annotated TFs*) multiplied by 100.

Equation 2

$$Accuracy(\%) = \frac{TFs\ predicted\ correctly * 100}{TFs\ predicted}$$

where, *Accuracy (%)* is determined by the ratio of the total number of TFs predicted by PredicTF in agreement with NCBI annotation (*TFs predicted correctly*) to the total number of TFs predicted by PredicTF (*TFs predicted*) multiplied by 100.

Equation 3

$$Accuracy\ for\ putative\ TFs(\%) = \frac{putative\ TFs\ predicted\ correctly * 100}{putative\ TFs\ predicted}$$

where, *Accuracy for putative TFs (%)* is determined by the total number of putative TFs predicted correctly divided by putative TFs predicted multiplied by 100; *Putative TFs predicted correctly* is the total number of putative TFs predicted correctly by PredicTF in agreement with NCBI annotation; and, *Putative TFs predicted* is the total number of putative TFs predicted by PredicTF.

Prediction of Transcription Factors in model organisms

We selected bacterial species that have been widely studied as model organisms. Some bacterial species became model organisms for TF studies because they are easy to maintain and grow in a laboratory setting and to manipulate in pure culture experiments. Five complete genomes from model organisms (*E. coli, B. subtillis, P. fluorescens, A. vinelandii and C. crescentus*) were downloaded directly from NCBI. The strains details and accession number (RefSeq) for all selected organisms are listed in the supplementary data (Additional file 3: Table S2). By evaluating PredicTF using model organisms (Additional file 2: Table S1) we extrapolated performance and accuracy of our deep learn model. Since known TFs for each organism were removed from each the training dataset, we eliminate the possibility of mapping TFs already known and annotated for each of the different species. Performance, accuracy and accuracy for putative TFs of PredicTF for these five model organisms were calculated using Equations 1, 2 and 3.

Prediction of Transcription Factors in a clinical isolate

We demonstrated the use of PredicTF in a previously sequenced *P. aeruginosa* (PAO1) genome, a clinical isolate publicly available in NCBI (accession number **NC_002516.2**). *P. aeruginosa* PAO1 was selected because its genome has been sequenced and because of the availability of transcriptomes from three clinical mutants

of PAO1 (Y71, Y82, and Y89) grown in the presence and absence of an antibiotic cocktail. The transcriptomes of *P. aeruginosa* PAO1 mutants Y71, Y82, and Y89 are available in NCBI (Bioproject identifier **PRJNA479711**) [18]. These clinical *P. aeruginosa* PAO1 mutants were isolated from the sputa of three different pneumonia patients. Transcriptomes of *P. aeruginosa* PAO1 wild type and its mutants cultured in two different conditions (LB medium and LB medium in presence of antibiotic cocktail) have been previously described [18]. We used this data to determine the TF profile in these *P. aeruginosa* PAO1 mutants grown in two different conditions.

PredicTF was first used to predict TFs in the *P. aeruginosa* PAO1 genome. Next, the predicted TFs were mapped to the transcriptomes of the *P. aeruginosa* PAO1 mutants Y71, Y82 and Y89 (see later). Further description of the mapping of the transcriptomes to the genomes is available at https://github.com/mdsufz/PredicTF. The PredicTF model used in this step was trained with the full database BacTFDB. The performance and accuracy were calculated using Equations 1 and 2, respectively. All accession numbers used in this work are listed in the supplementary data (Additional file 3: Table S2).

Prediction of Transcription Factors in Complex Microbial Communities

To test PredicTF in a complex microbial community, we used an anaerobic ammonium oxidizing (anammox) microbial community from an anammox membrane bioreactor metagenome (LAC_MetaG_1) (data publicly available at NCBI bioproject via accession number **PRJNA511011**) [19]. We removed short and low-quality reads using Trim Galore - v0.0.4 dev according developer's instructions [20]. Over 50 million reads survived this step and were assembled using the *de novo* assembler metaSPADES - v3.12.0 [21]. The assembly was translated from nucleotide to amino acid sequences,

246 considering all possible translation frames, using emboss transeq [22]. The translated
247 assembly was then used as input for the prediction of transcription factors using PredicTF.
248 The region from each predicted TF was extracted. These putative TFs were later used in
249 the mapping TFs to metatranscriptomes.

250 We checked if the putative TFs predicted in the metagenomes were transcribed by
251 checking if the metatranscriptomic libraries were mapping to those regions. The
252 metatranscriptomic and metagenomic libraries used in this step belonged to the same
253 bioreactor. These metatranscriptomes are publicly available at the European Nucleotide
254 Archive under the accession numbers SRR7091385, SRR7523233, SRR7523244,
255 SRR7523245, SRR7091400, SRR7091401, SRR7091381, SRR7091402, SRR7091406,
256 SRR7523243, SRR7523246. These 11 metatranscriptomes were used to demonstrate the
257 effectiveness of the pipeline and to indicate the potential of PredicTF to profile
258 transcription factors in complex microbial communities. All accession numbers used in
259 this work are listed in the supplementary data (Additional file 3: Table S2).
260
261 **Mapping transcription factors to transcriptomes and metatranscriptomes**
262 Each transcriptomic and metatranscriptomic library was quality controlled by
263 removing short and low-quality reads using Trim Galore - v0.0.4 dev [20]. The 7
264 transcriptomic libraries for the *P. aeruginosa* PAO1 wild type and mutants showed at
265 least 26 million paired-end reads after quality checking. The 11 metatranscriptomic
266 libraries yielded over 50 million reads per library after quality check. After, the remaining
267 transcriptomic and metatranscriptomic reads were mapped to their respective assembled
268 genome or metagenome using Bowtie2 - v2.3.0 [23]. The number of reads mapped, and
269 the regions covered was extracted using SAMTools - v1.9 [24] and python 2.7. The
270 regions of the genome or metagenome assembly covered by transcriptomic or

metatranscriptomic reads were then crossed-referenced with the regions of their respective assembly which PredicTF assigned as putative TFs creating a TF profile for each transcript and metatranscriptome. A detailed description on the mapping of RNA-seq data to their respective genome or metagenome assembly can be found at the PredicTF github (https://github.com/mdsufz/PredicTF).

Results and Discussion

Database

BacTFDB is a robust and versatile bacterial TF database, it contains 11.691 TFs amino acid sequences spanning 1049 TF families and 720 different bacterial species. Fig. 2 shows the database distribution based on TF families and regulatory elements (Fig. 2A) and the distribution based on bacterial species (Fig. 2B). Although BacTFDB is composed by 11.961 TFs elements from 1049 different families and 720 organism's species, Fig. 2 shows TFs families and organisms' species that accumulate at least more than 50 sequences. We will update BacTFDB annually by adding novel entries deposited in UniProt and CollecTF. BacTFDB was used in PredicTF's deep learning model training. This model was later used to predict new TFs and their families in genomes and metagenomes.

Performance and Accuracy

The performance and accuracy of PredicTF were evaluated through the prediction of TFs in five model organisms (*E. coli, B. subtillis, P. fluorescens, A. vinelandii and C. crescentus*). For each model organism a different PredicTF model was trained to predict TFs from full protein-length sequences (described in the implementation section). After

training the five models, the performance and accuracy of each PredicTF model was calculated for each of the model organisms selected using *Equation 1, 2* and *3*.

The performance of PredicTF to identify TFs in the different model organisms ranged from 27.23% to 60.53% of the proteins described as TFs in the genomes of model organisms and the accuracy for experimentally validated TFs ranged from 73.91% and 91.43% (Table 1). Further, PredicTF was able to identify putative annotated TFs in the genomes of *E. coli* and *B. subtillis* with accuracies 85.71% and 100%, respectively (Table 1). No novel TF was predicted in the genome of *C. crescentus*, *P. fluorescens* and *A.vinelandii* (Table 1). TFs predicted by PredicTF for each organism, sorted by TF family, are shown in Fig. 3. For all organisms tested the most predicted TF family was LysR followed by OmpR/PhoB. The degree of accuracy obtained by PredicTF suggests that the deep learning strategy used is promising for the prediction of TFs in genomic or metagenomic data of bacterial species. PredicTF performance and accuracy can be further improved by expanding the number and diversity of sequences present in BacTFDB. As BacTFDB will be update yearly, we expect an improvement in TF identification of with every update.

Table 1. PredicTF performance, accuracy for experimentally validated Transcription Factors (Accuracy EV) and accuracy for putative Transcription Factors (Accuracy PU) in genomes of model organisms.

Organism	Performance[a]	Accuracy EV[b]	Accuracy PU[c]
E. coli k12	35.40%	88.51%	85.71%
B. subtillis	27.23%	73.91%	100%
C. crescentus	38.04%	83.93%	-[d]
P. fluorescens	51.19%	91.43%	-
A.vinelandii	60.53%	90.40%	-

[a] Performance was calculated by the ratio of the total number of TFs predicted by PredicTF (*Predicted TFs*) to the total number of proteins annotated as TFs in NCBI (*Annotated TFs*) multiplied by 100;
[b] Accuracy EV was determined by the ratio of the total number of TFs predicted by PredicTF in agreement with NCBI annotation (*TFs predicted correctly*) to the total number of TFs predicted by PredicTF (*TFs predicted*) multiplied by 100;
[c] Accuracy TU was determined by the total number of putative TFs predicted correctly divided by putative TFs predicted multiplied by 100; *Putative TFs predicted correctly is the total number of putative TFs predicted correctly by PredicTF in agreement with NCBI annotation*; and, *Putative TFs predicted* is the total number of putative TFs predicted by PredicTF;
[d] Currently there are no putative annotated TFs described in the genome of *C. crescentus, P.* fluorescens and *A.vinelandii*

Mining and Predicting TFs in Genomes and Transcriptomes from a bacterial isolate using PredicTF

PredicTF was used to predict TFs on the genome of *P. aeruginosa* PAO1 and these TFs were mapped in transcriptomes from the same isolate [18]. PredicTF predicted a total of 199 TFs in the *P. aeruginosa* PAO1 genome shown in Additional file 4: Fig. S2A by a family's distribution graphic. These 199 TFs were mapped in the transcriptomic data of a reference of *P. aeruginosa* PAO1. Initially, the mapping was done in the transcriptome of *P. aeruginosa* PAO1 cultured in LB media. Using this strategy, we were able to map 69 of the 199 predicted TFs to the transcriptomes under the experimental conditions carried out by Hwang & Yoon, 2019 (Additional file 4: Fig. S2B) [18]. Next, the mappings were done for another three clinical mutants of *P. aeruginosa* PAO1 (Y82, Y71, Y89) cultured in LB media (absence of an antibiotic cocktail) (Additional file 5: Fig. S3A, S3C and S3F). The TFs family's distribution for each *P. aeruginosa* PAO1 mutant cultured in presence of antibiotic cocktail is shown in the supplementary data (Additional file 5: Fig. S3B, S3D and S3F). These results demonstrate the potential of PredicTF in mapping regulatory elements in bacterial genomes and the use of this tool to map and compare TFs profiles after under different environmental conditions.

Mining and Predicting TFs in a Metagenome and Metatranscriptome using PredicTF

PredicTF was used to profile putative TFs in one metagenome recovered from an anaerobic ammonium oxidation community [19] followed by the mapping of the predicted TFs in metatranscriptomes recovered from the same community (metatranscriptomes accession numbers can be found in Additional file 3: Table S2). A total of 792 TFs (Fig. 4A) were predicted in the LAC_MetaG_1, an anaerobic ammonium

oxidizing microbial community from an anammox membrane bioreactor [19]. These 792 TFs are distributed across 27 TF families (Fig. 4A) and are related to the regulation of functions such as the oxygen limitation response and late symbiotic functions (NarL/FixJ), phosphate regulon response (OmpR/PhoB), transcriptional activator for nitrogen-regulated promoters (NtrC/DctD) and ferric uptake regulation (Fur). Next, the 792 TFs were mapped in 11 metatranscriptomes collected in different dates from the same bioreactor where the metagenome was recovered (Additional file 6: Table S3, Fig. 4B). Clustering analysis demonstrated the presence of five different groups of TFs families based on the number of transcription factor families expressed in each library (Fig. 4B). It is interesting to note that the two most abundant clusters in the heatmap are directly related to the oxygen limitation caused by the anaerobic ammonium oxidizing cultivation. In a bioreactor where oxygen is limited, an increase in the amount of nitrogen and phosphate is expected. The presence of these elements (Nitrogen and Phosphate) diverts the metabolism of the microbial community towards the production of regulators (TFs) that will help to maintain community stability. Clustering analyzes can be helpful to demonstrate the similarity between metatranscriptomic libraries based on the occurrence of TFs. This strategy can be useful to compare the profiles of TFs expressed in different environmental situations (comparing libraries with different metadata) creating patterns of TFs expression. Exploration of TF profiling in microbial communities (metagenomes or metatranscriptomes) will allow the exploration of regulation within complex microbial communities. Further, The recovery of metagenome assembled genomes is becoming standard in metagenomics studies [25–27]. The use of PredicTF together with the recovery of metagenome assembled genomes will allow the exploration of species-specific molecular mechanisms involved in the regulation of different ecosystem processes.

Conclusions

A better understanding of TFs in a bacterial community context open revenues for the exploration of gene regulation in ecosystems where bacteria play a key role. Our deep learning strategy was based on a novel and robust TF bacterial database (BacTFDB) with over 11 thousand TFs and their respective families. BacTFDB is a unique resource for the exploration of TFs and it provided the data to train a model within PredicTF capable of predicting novel TFs from genomes and metagenomes. PredicTF is the first pipeline designed to predict and annotate TFs in complex microbial communities. The prediction of TFs can provide information for those aiming to study and understand bacterial communities within a context of gene regulation. We also demonstrated that PredicTF can be used to predict novel TFs in metagenomes and metatrascriptomes creating the potential profile for regulatory elements in complex microbial communities.

PredicTF is a flexible open source pipeline able to predict and annotate TFs in genomes and metagenomes and can be found at https://github.com/mdsufz/PredicTF.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Acknowledgements

Funding

LMOM were supported by FAPESP PhD (award # 2016/19179-9) and FAPESP Research Internship Scholarship Abroad (award # 2018/21133-2). RSR was supported by FAPESP (award # 2019/15675-0). JS and UNR were supported by the Helmholtz Young Investigator grant VH-NG-1248 Micro 'Big Data'.

Authors' contributions

LMOM, JS, PFS, RSR, and UNR developed the concept of PredicTF. LMOM, JS, UNR developed the PredicTF workflow. LMOM, JS, and UNR performed the benchmarks. LMOM provided information and data for the creation BacTFDB dataset. RBT and UNR perfformed the metagenome and metatranscriptome analysis. LMOM and UNR wrote the manuscript. All authors read and approved the manuscript.

Availability of data and materials

Genomes of the model organisms used in the Tool Validation step are available at the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/) under the accession numbers NC_000913.3, NC_000964.3, NC_011916.1, NC_021149.1, and NC_016830. The datasets supporting the Prediction of Transcription Factors in a clinical isolate of this article are available at National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/) under the accession number NC_002516.2 (genome) and study accession PRJNA479711 (transcriptomes). The datasets used for the Prediction of Transcription Factors in Complex Microbial Communities of this study are available at National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/) under the study accession PRJNA511011. The respective data sets of metatranscriptomes used are available at National Center for Biotechnology Information

(https://www.ncbi.nlm.nih.gov/) under the SRA numbers SRR7091385, SRR7523233, SRR7523244, SRR7523245, SRR7091400, SRR7091401, SRR7091381, SRR7091402, SRR7091406, SRR7523243, SRR7523246 and the Joint Genome Institute (https://jgi.doe.gov/) under the Gold Analysis Project identifiers Gp0267156, Gp0267150, Gp0267154, Gp0267155, Gp0267157, Gp0267158, Gp026715, Gp0267159, Gp0267152, Gp0267153, Gp0267160. All analysis, results and scripts used to generate figures are available at https://github.com/mdsufz/PredicTF.

Competing of interests

Not applicable

Availability and requirements

Project name: PredicTF

Project home page: https://github.com/mdsufz/PredicTF

Operating system: Linux64

Programming languages: Python 2.7

Other requirements: DIAMOND [28]; Nolearn Lasagne deep learning library [29]; Sklearn machine learning routines (https://scikit-learn.org/stable/) [30]; Theano (http://deeplearning.net/software/theano/) [31]. Trim Galore - v0.0.4 dev (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) [20]. MetaSPADES - v3.12.0 (https://github.com/ablab/spades#meta) [21]. Emboss transeq (http://www.bioinformatics.nl/cgi-bin/emboss/transeq) [22]. Bowtie2 - v2.3.0 (https://sourceforge.net/projects/bowtie-bio/) [23]. SAMTools - v1.9 (http://github.com/samtools/) [24].

List of Abbreviations

Transcription Factors (TFs)

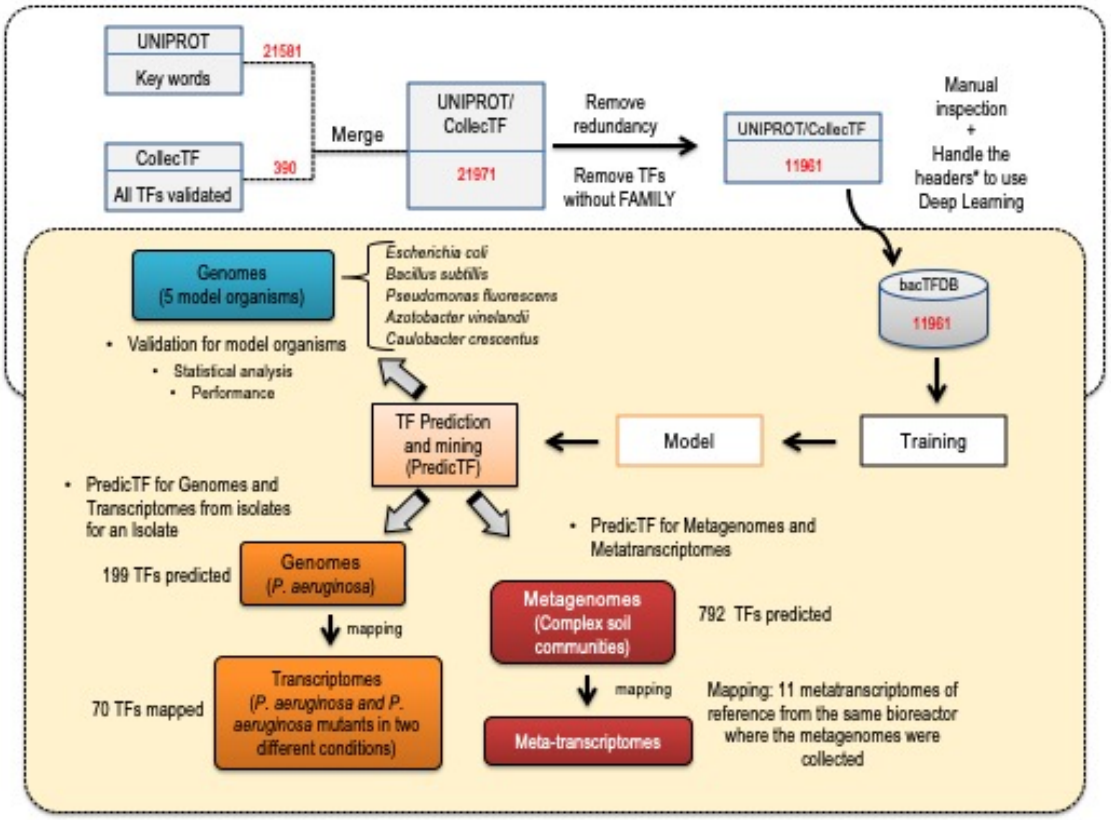
Bacterial Transcription Factor Data Base (BacTFDB)

Transcription factor binding sites (TFBSs)

anaerobic ammonium oxidizing (anammox)

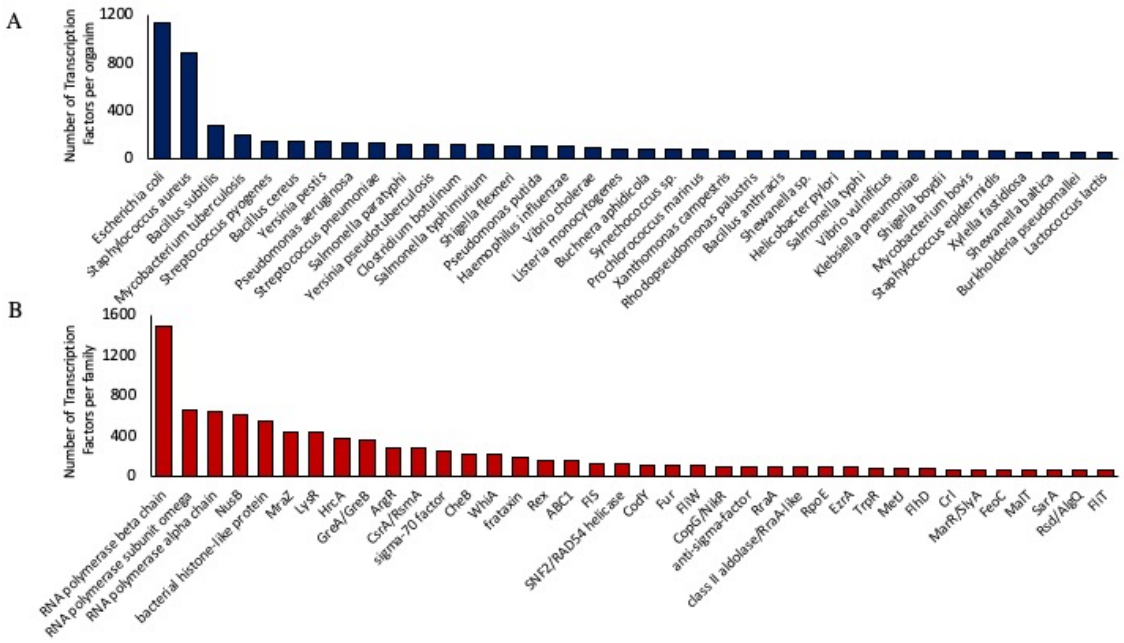
Figure legends:

Fig. 1



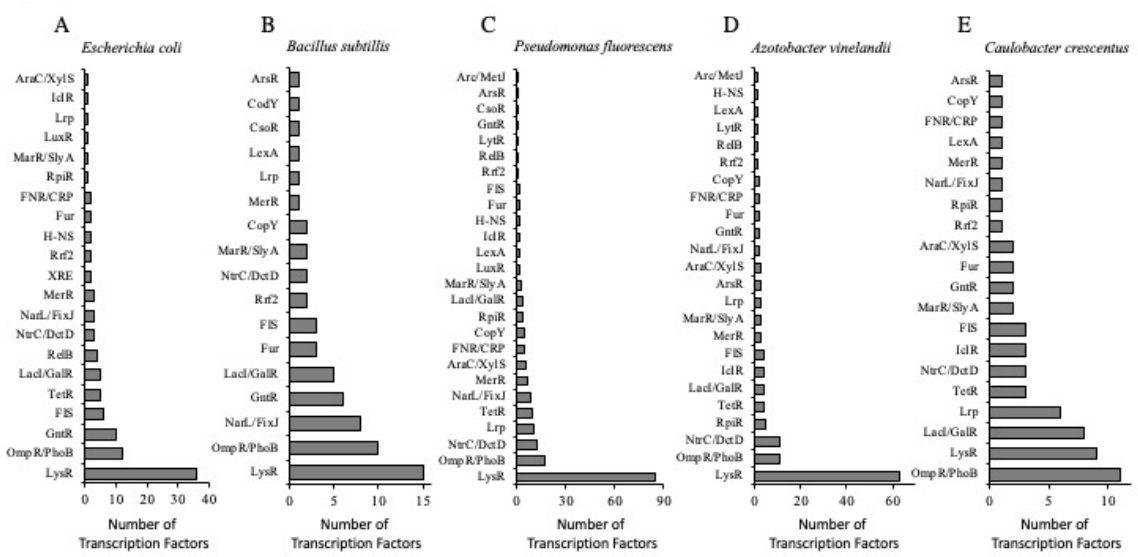
PredicTF workflow and testing. We collected publicly available data on TFs from two different databases: CollecTF and UNIPROT. After removing redundancies and filtering TFs well characterized, this data (BacTFDB) was used to train a deep learning model to predict new TFs and their families. Five model organisms (*Escherichia coli, Bacillus subtillis, Pseudomonas fluorescens, Azotobacter vinelandii and Caulobacter crescentus*) were used to test the accuracy of PredicTF. Later, we used the same approach to predict TFs from an insolate (*P. aeruginosa*) and mapped TFs predicted in transcriptomics data (*P. aeruginosa* and mutants in two experimental conditions). Finally, we used our tool to predict TF for complex communities (metagenome) and mapped these TFs in their respective meta-transcriptomes.

Fig. 2



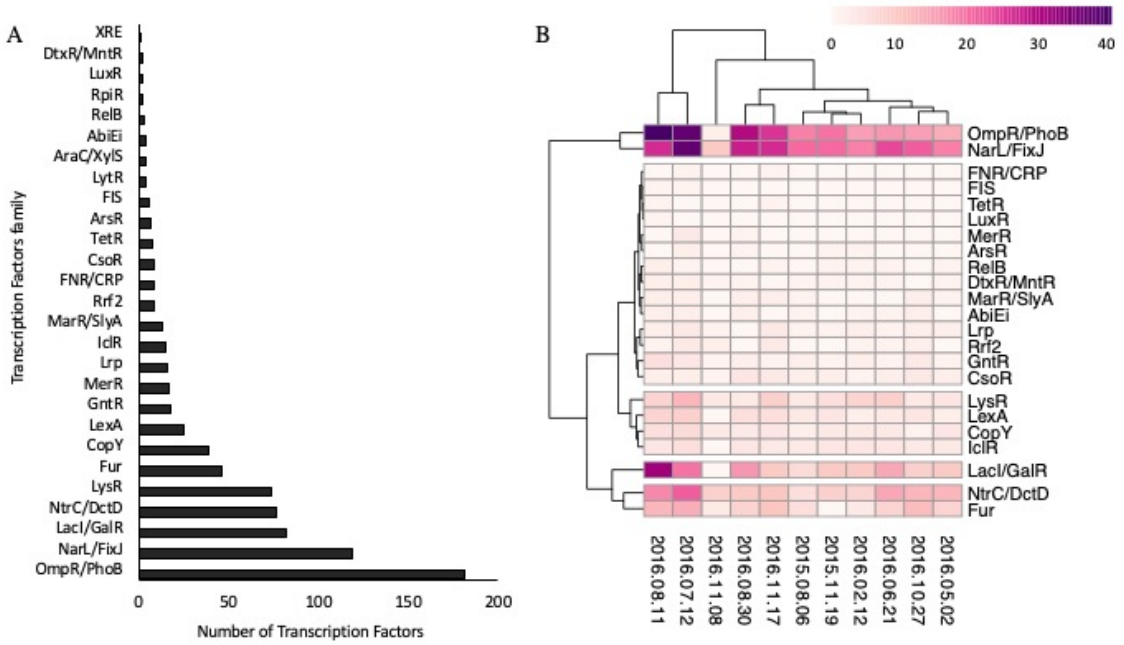
Database composition: Transcription Factor Database (BacTFDB) distribution. A) Database distribution based on the TFs and **B)** Regulatory Elements families and Organisms species. In these graphics only families with up to 50 sequences and only organisms that contributed with more than 50 sequences are shown.

Fig. 3



Prediction of TFs by PredicTF for genomes of model organisms. Prediction of TFs or 5 model organisms sorted by family. **A)** *Escherichia coli* **B)** *Bacillus subtillis* **C)** *Caulobacter crescentus* **D)** *Pseudomonas fluorescens* **E)** *Azotobacter vinelandii*

Fig. 4



Recovery of novel Transcription Factors in one metagenome and eleven metatranscriptomes. A) PredicTF predicted 792 TFs were predicted in one anaerobic ammonium oxidizing microbial communities from anammox membrane bioreactor (LAC_MetaG_1) and were grouped by family. **B)** Using 792 TFs predicted in one metagenome, we mapped these TFs for 11 metatranscriptomes of reference from the same bioreactor where the metagenome was recovered.

Additional files

Additional file 1: Fig. S1

Bacterial Transcription Factor Data Base (BacTFDB) were created from from two publicly available databases. We collect 390 TFs from CollecTF and 21.581 from UniProt (accessed 8-Sep-2019) accumulating 21.581 Transcription Factor (TF) amino acid sequences. We merged the data from CollecTF and UniProt databases resulting in a total of 21.971 TFs amino acid. We removed redundant TF entries and since PredicTF was designed to also assign TF family, TF sequences lacking a TF family were removed. Finally, a manual inspection was performed to remove misleading of spelling, case sensitive and presence of characters associate to the database header. The final database (BacTFDB) contains a total of 11.691 TF unique sequences.

Additional file 2: Table S1

Description of the bacterial transcriptional factors database (BacTFDB) subsets used to train models to predict Trancription Factors in model organisms

Additional file 3: Table S2

Accession number for 5 model organisms, *Pseudomonas aeruginosa* PAO1 genome and transcriptomes and Complex Microbial Communities used to validate and test PredicTF.

Additional file 4: Fig. S2

Transcription factor (TF) families predicted for *Pseudomonas aeruginosa* PAO1 genome (accession number NC_002516.2) [18] using PredicTF and their mapping to *P. aeruginosa* PAO1 growing in LB medium. A) A total of 199 TFs distributed in 25 TF

564 families were predicted in the *P. aeruginosa* PAO1 genome. B) These 199 TFs were
565 mapped in the transcriptomic data of a reference of *P. aeruginosa* PAO1 (Bioproject
566 identifier PRJNA479711) [18]. Initially, the mapping was done in the transcriptome of *P.*
567 *aeruginosa* PAO1 cultured in LB media. Using this strategy, we were able to map 69 of
568 the 199 predicted TFs to the transcriptome.
569
570 **Additional file 5: Fig. S3**
571 Transcription Factor (TF) family profiles in three *Pseudomonas aeruginosa* PAO1
572 mutants. After the prediction of Transcription Factors (TFs) using *P. aeruginosa* PAO1
573 genome, we mapped transcriptomes from three *P. aeruginosa* PAO1 mutants (Y82, Y71,
574 Y89) cultured in LB media (A, C and F). After, the mapping was done for each *P.*
575 *aeruginosa* PAO1 mutant cultured in presence of antibiotic cocktail (B, D and E). *P.*
576 *aeruginosa* PAO1 mutant Y82 (A, B); P. aeruginosa PAO1 mutant Y71 (C, D); *P.*
577 *aeruginosa* PAO1 mutant Y89 (E, F).
578
579 **Additional file 6: Table S3**
580 Number of Transcription Factors (TFs) per TF family mapped to each of the 11
581 metatranscriptomes of reference from the same bioreactor where the metagenome
582 (accession number PRJNA511011, NCBI) used to predict the putative TFs was collected.
583 The different metatranscriptomes are represented by their European Nucleotide Archive
584 accession numbers.
585
586
587
588

589

590 **References**

591 1. Liu J, Meng Z, Liu · Xiaoyue, Zhang X-H. Microbial assembly, interaction,
592 functioning, activity and diversification: a review derived from community compositional
593 data. Mar Life Sci Technol [Internet]. 2019 [cited 2020 Jul 20];1:112–28. Available from:
594 https://doi.org/10.1007/s42995-019-00004-3

595 2. Browning DF, Busby SJW. The regulation of bacterial transcription initiation. Nat Rev
596 Microbiol [Internet]. 2004;2:57–65. Available from:
597 http://www.ncbi.nlm.nih.gov/pubmed/15035009

598 3. Browning DF, Butala M, Busby SJW. Bacterial Transcription Factors: Regulation by
599 Pick "N" Mix. J. Mol. Biol. Academic Press; 2019. p. 4067–77.

600 4. Browning DF, Busby SJW. Local and global regulation of transcription initiation in
601 bacteria. Nat Rev Microbiol [Internet]. Nature Publishing Group; 2016;14:638–50.
602 Available from: http://www.nature.com/doifinder/10.1038/nrmicro.2016.103

603 5. Browning DF, Grainger DC, Busby SJ. Effects of nucleoid-associated proteins on
604 bacterial chromosome structure and gene expression. Curr Opin Microbiol. 2010;13:773–
605 80.

606 6. Gama-Castro S, Salgado H, Santos-Zavaleta A, Ledezma-Tejeida D, Muñiz-Rascado
607 L, García-Sotelo JS, et al. RegulonDB version 9.0: high-level integration of gene
608 regulation, coexpression, motif clustering and beyond. Nucleic Acids Res.
609 2016;44:D133–43.

610 7. Sierro N, Makita Y, De hoon M, Nakai K. DBTBS: A database of transcriptional
611 regulation in Bacillus subtilis containing upstream intergenic conservation information.
612 Nucleic Acids Res [Internet]. 2008 [cited 2020 Jun 15];36. Available from:
613 https://academic.oup.com/nar/article-abstract/36/suppl_1/D93/2507686

614 8. Consortium F. FlyBase: a Drosophila database. Flybase Consortium. Nucleic Acids
615 Res [Internet]. 1998 [cited 2020 Jun 15];26:85–8. Available from: http://astorg.u-
616 strasbg.fr:7081/

617 9. Park J, Park J, Jang S, Kim S, Kong S, Choi J, et al. FTFD: An informatics pipeline
618 supporting phylogenomic analysis of fungal transcription factors. Bioinformatics.
619 2008;24:1024–5.

620 10. Kummerfeld SK. DBD: a transcription factor prediction database. Nucleic Acids Res
621 [Internet]. 2006 [cited 2020 Jun 15];34:D74–81. Available from:
622 https://academic.oup.com/nar/article-abstract/34/suppl_1/D74/1133788

623 11. Widder S, Allen RJ, Pfeiffer T, Curtis TP, Wiuf C, Sloan WT, et al. Challenges in
624 microbial ecology: Building predictive understanding of community function and
625 dynamics [Internet]. ISME J. Springer Nature; 2016 [cited 2020 Jul 20]. p. 2557–68.
626 Available from: www.nature.com/ismej

12. Alipanahi B, Delong A, Weirauch MT, Frey BJ. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol [Internet]. 2015 [cited 2020 Jun 15];33:831–8. Available from: http://tools.genes.toronto.edu/

13. Pan X, Shen H Bin. RNA-protein binding motifs mining with a new hybrid deep learning based cross-domain knowledge integration approach. BMC Bioinformatics. BioMed Central Ltd.; 2017;18.

14. Kiliç S, White ER, Sagitova DM, Cornish JP, Erill I. CollecTF: A database of experimentally validated transcription factor-binding sites in Bacteria. Nucleic Acids Res [Internet]. 2014 [cited 2020 Jun 15];42. Available from: https://academic.oup.com/nar/article-abstract/42/D1/D156/1051934

15. UniProt Consortium T. UniProt: the universal protein knowledgebase. Nucleic Acids Res [Internet]. 2018 [cited 2020 Jul 20];46:2699–2699. Available from: https://academic.oup.com/nar/article-abstract/32/suppl_1/D115/2505378

16. Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L. DeepARG: A deep learning approach for predicting antibiotic resistance genes from metagenomic data. Microbiome. BioMed Central Ltd.; 2018;6.

17. Sørensen L, Loog M, Lo P, Ashraf H, Dirksen A, Duin RPW, et al. Image dissimilarity-based quantification of lung disease from CT. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2010. p. 37–44.

18. Hwang W, Yoon SS. Virulence Characteristics and an Action Mode of Antibiotic Resistance in Multidrug-Resistant Pseudomonas aeruginosa. Sci Rep [Internet]. 2019 [cited 2020 Jun 15];9. Available from: https://www.nature.com/articles/s41598-018-37422-9

19. Keren R, Lawrence JE, Zhuang W, Jenkins D, Banfield JF, Alvarez-Cohen L, et al. Increased replication of dissimilatory nitrate-reducing bacteria leads to decreased anammox bioreactor performance. Microbiome. BioMed Central Ltd.; 2020;8.

20. Krueger F. Babraham Bioinformatics - Trim Galore! [Internet]. Version 0.5.0. 2018 [cited 2020 Jul 29]. Available from: https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

21. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. MetaSPAdes: A new versatile metagenomic assembler. Genome Res [Internet]. 2017 [cited 2020 Jul 29];27:824–34. Available from: http://www.genome.org/cgi/doi/10.1101/gr.213959.116.

22. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, et al. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res [Internet]. 2019 [cited 2020 Jul 29];47:W636–41. Available from: https://academic.oup.com/nar/article-abstract/47/W1/W636/5446251

23. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods [Internet]. 2012 [cited 2020 Jul 29];9:357–9. Available from: http://bowtie-bio.sourceforge.net/bowtie2/index.shtml.

24. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics [Internet]. 2009 [cited 2020 Jul 29];25:2078–9. Available from: https://academic.oup.com/bioinformatics/article-abstract/25/16/2078/204688

25. Parks DH, Rinke C, Chuvochina M, Chaumeil PA, Woodcroft BJ, Evans PN, et al. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol [Internet]. 2017 [cited 2020 Jul 29];2:1533–42. Available from: https://www.nature.com/articles/s41564-017-0012-7/

26. Pasolli E, Asnicar F, Manara S, Zolfo M, Karcher N, Armanini F, et al. Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell [Internet]. 2019 [cited 2020 Jul 29];176:649-662.e20. Available from: https://www.sciencedirect.com/science/article/pii/S0092867419300017

27. Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data [Internet]. 2018 [cited 2020 Jul 29];5. Available from: https://www.nature.com/articles/sdata2017203

28. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat. Methods. Nature Publishing Group; 2014. p. 59–60.

29. van Merriënboer B, Bahdanau D, Dumoulin V, Serdyuk D, Warde-Farley D, Chorowski J, et al. Blocks and Fuel: Frameworks for deep learning. arxiv.org [Internet]. 2015 [cited 2020 Jun 15]; Available from: https://arxiv.org/abs/1506.00619

30. Pedregosa FABIANPEDREGOSA F, Michel V, Grisel OLIVIERGRISEL O, Blondel M, Prettenhofer P, Weiss R, et al. Scikit-learn: Machine Learning in Python Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX, GRAMFORT ET AL. Matthieu Perrot [Internet]. J. Mach. Learn. Res. 2011. Available from: http://scikit-learn.sourceforge.net.

31. The Theano Development Team, Al-Rfou R, Alain G, Almahairi A, Angermueller C, Bahdanau D, et al. Theano: A Python framework for fast computation of mathematical expressions. 2016 [cited 2020 Jun 15]; Available from: https://groups.google.com/group/theano-dev/

IV. GENERAL CONCLUSIONS

- In charpters I and II we demonstrate the existence of emergent properties in complex synthetic promoters in *E. coli*, which could be extrapolated to naturally occurring regulatory systems and would significantly impact the engineering of synthetic biological circuits in bacteria. Additionally, our results demonstrated in a systematic way that the arrangement and number of these *cis*-regulatory elements are crucial for the final expression dynamics of the target promoters. Taken together, these data presented here demonstrate how small changes in the architecture of bacterial promoters could result in drastic changes in the final regulatory logic of the system, with important implications for the understanding of natural complex promoters in bacteria and their engineering for novel applications. Our findings also present a comprehensive strategy for fine-tuning gene circuits to perform optimally in a given context (e.g., engineering of synthetic promoters) as well as provide insights for the understanding of natural complex promoters controlled by global regulators.

- In charpter III a collection of engineered transcription factors was generated with enhanced response to a well characterized and largely innocuous molecule with a potential for eliciting heterologous expression of bacterial genes in animal carriers. These results demonstrate the expansion of the genetic toolbox for the engineering of synthetic circuits inducible by safe chemical compounds.

- In charpter IV we provide a deep learning strategy based on a novel and robust TF bacterial database (BacTFDB). BacTFDB is a unique resource for the exploration of TFs and it provides the data to train a model within PredicTF capable of predicting novel TFs from genomes and metagenomes. PredicTF is the first pipeline designed to predict and annotate TFs in complex microbial communities. We also demonstrated that PredicTF can be used to

predict novel TFs in metagenomes and metatrascriptomes creating the potential profile for regulatory elements in complex microbial communities.

V. ADDITIONAL INFORMATION

1. Articles published in journals (related to this thisis)

- **MONTEIRO, Lummy Maria Oliveira;** ARRUDA, Leticia Magalhaes; SILVA-ROCHA, Rafael. Emergent properties in complex synthetic bacterial promoters. ACS synthetic biology, v. 7, n. 2, p. 602-612, 2018. https://doi.org/10.1021/acssynbio.7b00344

- **MONTEIRO, Lummy Maria Oliveira**; SANCHES-MEDEIROS, Ananda; WESTMANN, Caua Antunes; & SILVA-ROCHA, Rafael. Modulating Fis and IHF binding specificity, crosstalk and regulatory logic through the engineering of complex promoters. bioRxiv, p. 614396, 2019. https://doi.org/10.1101/614396

- **MONTEIRO, Lummy Maria Oliveira**; ARRUDA, Letícia Magalhaes; SANCHES-MEDEIROS, Ananda; MARTINS-SANTANA, Leonardo; ALVES, Luana de Fátima; DEFELIPE, Lucas; TURJANSKI, Adrian Gustavo; GUAZZARONI, María-Eugenia & SILVA-ROCHA, Rafael. Reverse engineering of an aspirin-responsive transcriptional regulator in *Escherichia coli*. ACS synthetic biology, v. 8, n. 8, p. 1890-1900, 2019. https://doi.org/10.1021/acssynbio.9b00191

Submitted

- **MONTEIRO, Lummy Maria Oliveira;** SARAIVA, João; TOSCAN, Rodolfo Brizola; STADLER, Peter; SILVA-ROCHA, Rafael; da ROCHA, Ulisses Nunes. PredicTF: a tool to predict bacterial transcription factors in complex microbial communities. (Submitted)

2. Articles published in journals (Not related to this thesis)

- SANCHES-MEDEIROS, Ananda; **MONTEIRO, Lummy Maria Oliveira**; SILVA-ROCHA, Rafael. Calibrating transcriptional activity using constitutive synthetic promoters in mutants for global regulators in *Escherichia coli*. International journal of genomics, v. 2018, 2018. https://doi.org/10.1155/2018/9235605

- ARRUDA, Letícia Magalhaes; **MONTEIRO, Lummy Maria Oliveira**; SILVA-ROCHA, Rafael. The Chromobacterium violaceum ArsR arsenite repressor exerts tighter control on its cognate promoter than the *Escherichia coli* system. Frontiers in microbiology, v. 7, p. 1851, 2016. https://doi.org/10.3389/fmicb.2016.01851

- NORA, Luísa Czamanski, WESTMANN, Caua Antunes, MARTINS-SANTANA, Leonardo; ALVES, Luana de Fátima; **MONTEIRO, Lummy Maria Oliveira**; GUAZZARONI, Maria Eugenia; SILVA-ROCHA, Rafael. The art of vector engineering: towards the construction of next-generation genetic tools. Microbial biotechnology, v. 12, n. 1, p. 125-147, 2019. https://doi.org/10.1111/1751-7915.13318

Submitted

- TLÁSKAL, Vojtěch; BRABCOVÁ, Vendula; VĚTROVSKÝ, Tomáš; JOMURA, Mayuko; LÓPEZ-MONDÉJAR, Rubén; **MONTEIRO, Lummy Maria Oliveira**; SARAIVA, João Pedro; HUMAN, Zander Rainier; CAJTHAML, Tomáš; DA ROCHA, Ulisses Nunes; BALDRIAN, Petr. Cross-domain cooperation between bacteria and fungi mediates wood decomposition. (Submitted)

VI. ATTACHMENTS

Emergent Properties in Complex Synthetic Bacterial Promoters

Author:
Lummy Maria Oliveira Monteiro, Letícia Magalhães Arruda, Rafael Silva-Rocha

Publication: ACS Synthetic Biology
Publisher: American Chemical Society
Date: Feb 1, 2018

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK CLOSE WINDOW

Reverse Engineering of an Aspirin-Responsive Transcriptional Regulator in Escherichia coli

Author:
Lummy Maria Oliveira Monteiro, Letícia Magalhães Arruda, Ananda Sanches-Medeiros, et al

Publication: ACS Synthetic Biology

Publisher: American Chemical Society

Date: Aug 1, 2019

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats, and translations.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.

BACK CLOSE WINDOW

VII. REFERENCES

Andrianantoandro, E., Basu, S., Karig, D. K., and Weiss, R. (2006). Synthetic biology: New engineering rules for an emerging discipline. Mol. Syst. Biol. 2. doi:10.1038/msb4100073.

Barnard, A., Wolfe, A., and Busby, S. (2004). Regulation at complex bacterial promoters: How bacteria use different promoter organizations to produce different regulatory outcomes. Curr. Opin. Microbiol. 7, 102–108. doi:10.1016/j.mib.2004.02.011.

Barne, S. L., Belliveau, N. M., Ireland, W. T., Kinney, J. B., and Phillips, R. (2019). Mapping DNA sequence to transcription factor binding energy in vivo. PLoS Comput. Biol. 15. doi:10.1371/journal.pcbi.1006226.

Benner, S. A., and Sismour, A. M. (2005). Synthetic biology. Nat. Rev. Genet. 6, 533–543. doi:10.1038/nrg1637.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., et al. (2005a). Transcriptional regulation by the numbers: Applications. Curr. Opin. Genet. Dev. 15, 125–135. doi:10.1016/j.gde.2005.02.006.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., et al. (2005b). Transcriptional regulation by the numbers: models. Curr. Opin. Genet. Dev. 15, 116–124. doi:10.1016/j.gde.2005.02.007.

Brown, N. L., Stoyanov, J. V., Kidd, S. P., and Hobman, J. L. (2003). The MerR family of transcriptional regulators. FEMS Microbiol. Rev. 27, 145–163. doi:10.1016/S0168-6445(03)00051-2.

Browning, D. F., and Busby, S. J. W. (2004). The regulation of bacterial transcription initiation. Nat. Rev. Microbiol. 2, 57–65. doi:10.1038/nrmicro787.

Browning, D. F., and Busby, S. J. W. (2016). Local and global regulation of transcription initiation in bacteria. Nat. Rev. Microbiol. 14, 638–650. doi:10.1038/nrmicro.2016.103.

Browning, D. F., Butala, M., and Busby, S. J. W. (2019). Bacterial Transcription Factors: Regulation by Pick “N” Mix. J. Mol. Biol. 431, 4067–4077. doi:10.1016/j.jmb.2019.04.011.

Busby, S., and Ebright, R. H. (1994). Promoter structure, promoter recognition, and transcription activation in prokaryotes. Cell 79, 743–746. doi:10.1016/0092-8674(94)90063-9.

Cameron, D. E., Bashor, C. J., and Collins, J. J. (2014). A brief history of synthetic biology. Nat. Rev. Microbiol. 12, 381–390. doi:10.1038/nrmicro3239.

Carbonell, P., Currin, A., Jervis, A. J., Rattray, N. J. W., Swainston, N., Yan, C., et al. (2016). Bioinformatics for the synthetic biology of natural products: Integrating across the Design-Build-Test cycle. Nat. Prod. Rep. 33, 925–932. doi:10.1039/c6np00018e.

Cases, I., De Lorenzo, V., and Ouzounis, C. A. (2003). Transcription regulation and environmental adaptation in bacteria. Trends Microbiol. 11, 248–253. doi:10.1016/S0966-842X(03)00103-3.

Chen, Y., Ho, J. M. L., Shis, D. L., Gupta, C., Long, J., Wagner, D. S., et al. (2018). Tuning the dynamic range of bacterial promoters regulated by ligand-inducible transcription factors. Nat. Commun. 9, 1–8. doi:10.1038/s41467-017-02473-5.

Deplazes-Zemp, A. (2012). The Conception of Life in Synthetic Biology. Sci. Eng. Ethics 18, 757–774. doi:10.1007/s11948-011-9269-z.

Elowitz, M., and Lim, W. A. (2010). Build life to understand it. Nature 468, 889–890. doi:10.1038/468889a.

Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., Ledezma-Tejeida, D., Muñiz-Rascado, L.,

García-Sotelo, J. S., et al. (2016). RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. Nucleic Acids Res. 44, D133–D143. doi:10.1093/nar/gkv1156.

Hunziker, A., Tuboly, C., Horváth, P., Krishna, S., and Semsey, S. (2010). Genetic flexibility of regulatory networks. Proc. Natl. Acad. Sci. U. S. A. 107, 12998–3003. doi:10.1073/pnas.0915003107.

Huo, Y. X., Zhang, Y. T., Xiao, Y., Zhang, X., Buck, M., Kolb, A., et al. (2009). IHF-binding sites inhibit DNA loop formation and transcription initiation. Nucleic Acids Res. 37, 3878–3886. doi:10.1093/nar/gkp258.

Hyeon, J. E., Shin, S. K., and Han, S. O. (2016a). Design of nanoscale enzyme complexes based on various scaffolding materials for biomass conversion and immobilization. Biotechnol. J. 11, 1386–1396. doi:10.1002/biot.201600039.

Hyeon, J. E., Shin, S. K., and Han, S. O. (2016b). Design of nanoscale enzyme complexes based on various scaffolding materials for biomass conversion and immobilization. Biotechnol. J. 11, 1386–1396. doi:10.1002/biot.201600039.

Ishihama, A. (2010). Prokaryotic genome regulation: Multifactor promoters, multitarget regulators and hierarchic networks. FEMS Microbiol. Rev. 34, 628–645. doi:10.1111/j.1574-6976.2010.00227.x.

Khalil, A. S., and Collins, J. J. (2010). Synthetic biology: Applications come of age. Nat. Rev. Genet. 11, 367–379. doi:10.1038/nrg2775.

Konstantinidis, K. T., and Tiedje, J. M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. Available at: www.pnas.orgcgidoi10.1073pnas.0308653100 [Accessed August 18, 2020].

Koonin, E. V., and Wolf, Y. I. (2008). Genomics of bacteria and archaea: The emerging dynamic view of the prokaryotic world. Nucleic Acids Res. 36, 6688–6719. doi:10.1093/nar/gkn668.

Lawrence, J. (1999). Selfish operons: The evolutionary impact of gene clustering in prokaryotes and eukaryotes. Curr. Opin. Genet. Dev. 9, 642–648. doi:10.1016/S0959-437X(99)00025-8.

Lee, D. J., Minchin, S. D., and Busby, S. J. W. (2012). Activating Transcription in Bacteria. Annu. Rev. Microbiol. 66, 125–152. doi:10.1146/annurev-micro-092611-150012.

Liu, H., and Chen, Q. (2016). Computational protein design for given backbone: Recent progresses in general method-related aspects. Curr. Opin. Struct. Biol. 39, 89–95. doi:10.1016/j.sbi.2016.06.013.

Liu, J., Meng, Z., Liu, · Xiaoyue, and Zhang, X.-H. (2019). Microbial assembly, interaction, functioning, activity and diversification: a review derived from community compositional data. Mar. Life Sci. Technol. 1, 112–128. doi:10.1007/s42995-019-00004-3.

Martínez-Antonio, A., Collado-Vides, J., Balleza, E., López-Bojorquez, L. N., Martínez-Antonio, A., Resendis-Antonio, O., et al. (2003). Identifying global regulators in transcriptional regulatory networks in bacteria. Curr. Opin. Microbiol. 33, 482–489. doi:10.1016/j.mib.2003.09.002.

Monteiro, L. M. O., Arruda, L. M., Sanches-Medeiros, A., Martins-Santana, L., Alves, L. D. F., Defelipe, L., et al. (2019). Reverse Engineering of an Aspirin-Responsive Transcriptional Regulator in Escherichia coli. ACS Synth. Biol. doi:10.1021/acssynbio.9b00191.

Monteiro, L. M. O., Arruda, L. M., and Silva-Rocha, R. (2018). Emergent Properties in Complex Synthetic Bacterial Promoters. ACS Synth. Biol. 7, 602–612.

doi:10.1021/acssynbio.7b00344.

Monteiro, L. M. O., Sanches-Medeiros, A., Westmann, C. A., and Silva-Rocha, R. (2020). Unraveling the Complex Interplay of Fis and IHF Through Synthetic Promoter Engineering. Front. Bioeng. Biotechnol. 8. doi:10.3389/fbioe.2020.00510.

Noda, S., Shirai, T., Oyama, S., and Kondo, A. (2016). Metabolic design of a platform Escherichia coli strain producing various chorismate derivatives. Metab. Eng. 33, 119–129. doi:10.1016/j.ymben.2015.11.007.

Sambrook, J., Fritsch, E., and Maniatis, T. (1989). Molecular cloning: a laboratory manual. Available at: https://www.cabdirect.org/cabdirect/abstract/19901616061 [Accessed April 2, 2020].

Sanches-Medeiros, A., Monteiro, L. M. O., and Silva-Rocha, R. (2018). Calibrating transcriptional activity using constitutive synthetic promoters in mutants for global regulators in Escherichia coli. Int. J. Genomics 2018. doi:10.1155/2018/9235605.

Shimada, T., Fujita, N., Yamamoto, K., and Ishihama, A. (2011a). Novel roles of camp receptor protein (CRP) in regulation of transport and metabolism of carbon sources. PLoS One 6. doi:10.1371/journal.pone.0020081.

Shimada, T., Yamamoto, K., and Ishihama, A. (2011b). Novel Members of the Cra Regulon Involved in Carbon Metabolism in Escherichia coli †. J. Bacteriol. 193, 649–659. doi:10.1128/JB.01214-10.

Silva-Rocha, R., and De Lorenzo, V. (2010). Noise and robustness in prokaryotic regulatory networks. Annu. Rev. Microbiol. 64, 257–275. doi:10.1146/annurev.micro.091208.073229.

Silva-Rocha, R., and De Lorenzo, V. (2011). Implementing an OR-NOT (ORN) logic gate with components of the SOS regulatory network of Escherichia coli. Mol. Biosyst. 7, 2389–2396. doi:10.1039/c1mb05094j.

Singleton, P. (2009). Dictionary of DNA and Genome Technology. John Wiley & Sons, Ltd doi:10.1002/9780470689127.

Siu, Y., Fenno, J., Lindle, J. M., and Dunlop, M. J. (2017). Design and Selection of a Synthetic Feedback Loop for Optimizing Biofuel Tolerance. ACS Publ. 7, 16–23. doi:10.1021/acssynbio.7b00260.

Tamsir, A., Tabor, J. J., and Voigt, C. A. (2011). Robust multicellular computing using genetically encoded NOR gates and chemical “wires”. Nature 469, 212–5. doi:10.1038/nature09565.

Wang, B., Kitney, R. I., Joly, N., and Buck, M. (2011). Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. Nat. Commun. 2, 1–9. doi:10.1038/ncomms1516.

Wang, J., Cao, H., Zhang, J. Z. H., and Qi, Y. (2018). Computational Protein Design with Deep Learning Neural Networks. Sci. Rep. 8. doi:10.1038/s41598-018-24760-x.

Way, J. C., Collins, J. J., Keasling, J. D., and Silver, P. A. (2014). Integrating biological redesign: Where synthetic biology came from and where it needs to go. Cell 157, 151–161. doi:10.1016/j.cell.2014.02.039.

Widder, S., Allen, R. J., Pfeiffer, T., Curtis, T. P., Wiuf, C., Sloan, W. T., et al. (2016). Challenges in microbial ecology: Building predictive understanding of community function and dynamics. ISME J. 10, 2557–2568. doi:10.1038/ismej.2016.45.

Yang, Y., Darbari, V. C., Zhang, N., Lu, D., Glyde, R., Wang, Y. P., et al. (2015). Structures of the RNA polymerase-σ54 reveal new and conserved regulatory strategies. Science (80-.). 349,

882–885. doi:10.1126/science.aab1478.

Zhang, F., and Voytas, D. F. (2018). Synthetic genomes engineered by SCRaMbLEing. Sci. China Life Sci. 61, 975–977. doi:10.1007/s11427-018-9325-1.

Zhang, W., Mitchell, L. A., Bader, J. S., and Boeke, J. D. (2020). Synthetic Genomes. Annu. Rev. Biochem. 89, 77–101. doi:10.1146/annurev-biochem-013118-110704.