UNIVERSIDADE DE SÃO PAULO

FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

GERARDO RUIZ AMORES

Unveiling the Effect of Global Regulators in the Regulatory Network for Biofilm Formation in *Escherichia coli*

Ribeirão Preto, Brasil, 2017

GERARDO RUIZ AMORES

Unveiling the Effect of Global Regulators in the Regulatory Network for Biofilm Formation in *Escherichia coli*

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: Biologia Celular e Molecular

Orientador: Prof. Dr. Rafael Silva Rocha

Ribeirão Preto, Brasil, 2017

FICHA CATALOGRÁFICA

Amores, Ruiz Gerardo

Entendendo o Efeito dos Reguladores Globais na Rede Regulatória para a Formação de Biofilme em *Escherichia coli*.

Ribeirão Preto, 2017

128. p.: 30cm

Tese de Doutorado, apresentada à Faculdade de Medicina de Ribeirão Preto/USP. Área de concentração: Biologia Celular e Molecular de Bioagentes Patogênicos.

Orientador: Rafael Silva-Rocha.

1. Biofilme 2. Redes Regulatórias 3. Regulação da Transcrição

To my parents for all the love and education, Sandra Luz Amores Espinal

& Gerardo Ruiz Tellez

To my eternal life for all the unconditional love, Marcela Rodriguez Sanchez

To my grandmother for all the support, Altagracia Espinal Torres

To my sister for all the support, Jade Ruiz Amores

To my father in law for the encouragement, Ernesto Filidor Perez

"– A vida, senhor Visconde, é um pisca-pisca. A gente nasce, isto é, começa a piscar. Quem pára de piscar chegou ao fim, morreu. Piscar é abrir e fechar os olhos – viver é isso. É um dorme e acorda, dorme e acorda, até que dorme e não acorda mais [...] A vida das gentes neste mundo, senhor Sabugo, é isso. Um rosário de piscados. Cada pisco é um dia. Pisca e mama, pisca e brinca, pisca e estuda, pisca e ama, pisca e cria filhos, pisca e geme os reumatismos, e por fim pisca pela última vez e morre. – E depois que morre?, perguntou o Visconde. – Depois que morre, vira hipótese. É ou não é?"

Monteiro Lobato

But darkness descended, and mists of myth and mysticism settled in for a thousand of years. Ways of knowing dwindled to one: higher authority. Remnants to Aristotle were deemed authoritative, but how he known? Infinite regress loomed.

Our scientific theory can indeed go wrong, and precisely in the familiar way: through failure of predicted observation. But what if, we have achieved a theory that is conformable to every possible observation, past and future? In what sense could the world be said to deviate from what the theory claims?

W.V. Quine, 1981 and 1995

ACKNOWLEDGMENTS

To Dr. Rafael Silva Rocha, because he believe in me and guided me to be a better scientist and person. My entire admiration owing to your capability to analyze the biological phenomenons and find the answer, also owing the wisdom shared.

To Dr. Aitor de Las Heras, a brilliant scientific mind with PROG until the bones. My entire gratefully owing the wise advices shared in science and life. Thanks for the great times in Edinburgh!

To Dr. Alistair Elfick, owing to the affable and friendly reception during my visit to Edinburgh and for accepting me at his laboratory at the University of Edinburgh. My gratitude owing to the knowledge shared.

I am grateful to the Comissão de Aperfeiçoamento de Pessoal do Nível Superior CAPES for my PhD fellowship.

Finally I would like to thanks at all the people who helped me in some way; from medium culture preparation to scientific talks at laboratory, coffee and others. As is impossible to mention all of them. I would like to thank to all the excellent and commitment academic and technical staff at FMRP-USP, my colleagues and friends. Specially, I appreciate and will miss the great moments spent at laboratory with all my friends (http://ssblfmrp.blogspot.com.br/p/people.html).

Amores, G, R. **Entendendo o Efeito dos Reguladores Globais na Rede Regulatória para a Formação de Biofilme em *Escherichia coli*.** Tese apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para a obtenção do título de Doutor em ciências em Biologia Celular e Molecular.

Aprovado em: _____ de_____ de 2017

Banca Examinadora

Prof. Dr. _____Instituição: _____

Julgamento. _____Assinatura: _____

Prof. Dr. _____Instituição: _____

Julgamento. _____Assinatura: _____

Prof. Dr. _____Instituição: _____

Julgamento. _____Assinatura: _____

Prof. Dr. _____Instituição: _____

Julgamento. _____Assinatura: _____

Prof. Dr. _____Instituição: _____

Julgamento. _____Assinatura: _____

GENERAL INDEX

INDEX OF FIGURES

INDEX OF TABLES

ABBREVATIONS

˚C	Celsius grades
A. U.	Arbitrary Units
AC	Adenylate cyclase
APE	A Plasmid Editor
c-di-GMP	di-cyclic GMP
cAMP	cyclic AMP
CCR	Carbon catabolite repression
CR	Congo Red
CRP	cAMP receptor protein
CTD	Carboxyl terminal domain
Curli	Fimbriae protein generated by the transcription of *csgAB* operon
DGC	Diguanylate cyclase
DNA	Deoxyribonucleic acid
Dps	DNA-binding protein
Fis	Factor for inversion stimulation
GFP	Green fluorescent protein
Glu	Glucose
GR	Global regulator
H-NS	Histone-like nucleoid-structuring protein
IHF	Integration host factor

kDa	Kilo-Daltons
LB	Luria-Bertani medium
Lrp	Leucine-responsive regulatory protein
mL	milliliter
NAP	nucleoid-associated proteins
ng	Nanogram
Node	Biological part - acid nucleic, protein or metabolite
NTD	Amine terminal domain
PCR	Polymerase chain reaction
PDE	Phosphodiesterase
pmol	Picomol
PS	Polystyrene
RNA	Ribonucleic acid
StpA	suppressor of td mutant phenotype A
TF	Transcription Factor
U	Enzyme Unit
UTR	Untranslated region
WT	wild type
µl	Microliters

PREFACE

In this thesis, we show the effect of the global regulators CRP, IHF and Fis in the flagella-biofilm transcriptional regulatory network and show that, despite its important role to regulate the system, those GRs are not entirely necessary, since remotion of any GRs studied leads, despite affected, to motility and biofilm formation. Altogether, this data strongly suggest that the flagella-biofilm network present an intrinsic capacity to support nodes remotion, and supporting the robustness of the flagella-biofilm transcriptional regulatory network.

One of the synthetic biology aims is to standardize and understand the signal integration and the behavior of biological parts for its eventual circuit integration and properly function. Due this, as a first part of my PhD, I worked to generate synthetic cis-elements that will be recognized by three transcriptions factors as well as understand the regulatory mechanisms which drives this sequence in bacteria (Attached article). For this, we based our study on the well-known transcription activation and repression mechanisms of gene expression in bacteria (Browning and Busby, 2016). Specifically, we used the principle of CCR induction in bacteria, in which a specific *cis*-element recognized by CRP global regulator is sufficient to modulate cognate genes, as for wild type *lac* promoter. Therefore, in order to evaluate the hypothesis of "compress information", we wondering whether it could be possible to generate a synthetic sequence that would be recognized by three different global regulators. First, we unveil the consensus sequences that are recognized by CRP, IHF and Fis and we used an evolutionary algorithm to generate a synthetic *cis*-element that does not exist in bacteria and, that in principle, would be recognized by the three different transcription factors. Next, those sequences were cloned upstream of a GFPlva reporter system to determine the functionality of the sequences. For this, bacteria transformed with the reporter systems were surveyed *in vivo* to determine its dynamic of gene expression. The results shown that (i) the synthetic sequence, that does not exist in bacteria, is completely functional. (ii) The synthetic promoter regions generated were, in fact, modulated by three GRs, showing that it is possible to compress information. This concept was published in ACS Syntheticc Biology by the end of 2015 and the article is attached in this thesis.

ABSTRACT

Amores, G.R. **Unveiling the Effect of Global Regulators in the Regulatory Network for Biofilm Formation in *Escherichia coli*.** 2017. 128p. Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto. 2017.

In nature, biofilm is a complex structure resulted of multicellular bacterial communities that provide important nutritional functions and the acquisition of protective traits such as antibiotics resistance and horizontal gene transfer. The development from the planktonic, lonely bacteria, to the mature multilayered biofilm structure consists of three main phases: motility, attachment and biofilm maturation. At cellular level, the process is controlled by several genes such as *flhD*, *fliA, rpoS*, *csgD*, adrA, *cpxR* all acting as master regulators. Additionally, the global regulators CRP, IHF, Fis, and others in less frequency, have been related to biofilm formation, although blurry information has been provided. In this thesis we used synthetic, molecular and cellular biology approaches to understand the effect of CRP, IHF and Fis in the transcriptional regulatory network in the bacterium *Escherichia coli*. In the first chapter, we employed network analysis to reconstruct and analyze part of the entire regulatory network described to modulate the flagella-biofilm program. With this analysis we identified some critical interactions responsible for the planktonic-biofilm transition. Next, we selected the top ten effectors nodes of the network and cloned the promoter region of those genes in a reporter system. As extensively explained in chapter II, this system allowed us to validate as well as suggest new interactions in the network. Additionally, the measurement of the promoter activity during bacterial development show that CRP, IHF and Fis differentially modulate most of the surveyed genes suggesting that those Global Regulators participate to modulate gene expression in different phases of the planktonic-biofilm development. At chapter three, to get a better overview of the entire process, we performed motility, adherence/early biofilm and mature biofilm assays. We describe the intrinsic ability of *E. coli* to perform motility, adherence and mature biofilm at 37˚C. In contrast, the absence of *ihf, fis* as well as Carbon Catabolite Repression (CCR), lead to altered phenotypes at both motility and biofilm development. At the end, we discussed how the changes of promoter activity of target genes, together with our network analysis, could explain part of the altered phenotypes observed. For instance, we observed changes at the main stress responders *rpoS* and *rpoE* that, in combination with alterations at specific genes such as *fliA*, can explain the enhanced motility in the *E. coli* Δ*ihf* strain. Altogether, in this thesis, we provided evidence that CRP, IHF and Fis control the activity of the promoter regions of genes involved in the planktonic-biofilm development.

KEYWORDS: Biofilm, Regulatory Network, Transcriptional Regulation.

RESUMO

Amores, G.R. **Entendendo o Efeito dos Reguladores Globais na Rede Regulatória para a Formação de Biofilme em *Escherichia coli*.** 2017. 128p. Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto. 2017.

Na natureza, o biofilme é uma estrutura complexa resultante de comunidades bacterianas multicelulares que fornece importantes funções nutricionais e a aquisição de traços de proteção como resistência a antibióticos e transferência horizontal de genes. O desenvolvimento das bactérias planctônicas solitárias para uma estrutura de biofilme maduro consiste em três fases principais: motilidade, fixação e maturação do biofilme. Ao nível celular, o processo é controlado por vários genes tais como *flhD, fliA, rpoS, csgD, adrA, cpxR,* todos agindo como reguladores mestre. Além disso, os reguladores globais CRP, IHF, Fis e outros em menor freqüência, têm sido relacionados à formação de biofilme, embora tenham sido fornecidas informações nao conclusivas sobre esse processo. Nesta tese foram utilizadas abordagens de bioinformática, assim como de biologia molecular e celular para entender o efeito de CRP, IHF e Fis na rede reguladora da transição de motilidade para biofilme na bactéria *Escherichia coli*. No primeiro capítulo, utilizamos a análise de rede para reconstruir e analisar parte da rede regulatória descrita para modular o programa flagelo-biofilme. Com esta análise identificamos algumas interações críticas responsáveis pela transição planctônica-biofilme. Em seguida, selecionamos os dez principais nós efetores da rede e clonamos a região promotora desses genes em um sistema repórter. Conforme explicado amplamente no capítulo II, este sistema nos permitiu validar e sugerir novas interações na rede. Adicionalmente, a medição da atividade do promotor durante o desenvolvimento bacteriano mostra que a CRP, a IHF e a Fis modulam diferencialmente a maioria dos genes analisados sugerindo que estes Reguladores Globais participam para modular a expressão génica em diferentes fases do desenvolvimento de estado planctónico para biofilme. No capítulo três, para obter uma melhor visão geral de todo o processo, realizamos ensaios de motilidade, aderência / biofilme precoce e biofilmes maduros. Descrevemos a capacidade intrínseca de *E. coli* para realizar motilidade, adesão e biofilme maduro a 37 °C. Em contraste, a ausência de *ihf*, *fis*, bem como o fenômeno de Repressão de Catabolite de Carbono (CCR), levam a fenótipos alterados, tanto na motilidade como no desenvolvimento do biofilme. No final, discutimos como as mudanças da atividade do promotor de genes alvo, juntamente com a nossa análise de rede, poderia

explicar parte dos fenótipos alterados observados. Por exemplo, observamos mudanças nos principais respondedores de estresse *rpoS* e *rpoE* que, em combinação com alterações em genes específicos como *fliA*, podem explicar a motilidade aumentada na estirpe de *E. coli Δihf*. Em conjunto, nesta tese, apresentamos evidências de que CRP, IHF e Fis controlam a atividade das regiões promotoras de genes envolvidos no desenvolvimento planctônico-biofilme.

PALAVRAS CHAVES: Biofilme, Redes Regulatórias, Regulação da Transcrição.

I. INTRODUCTION

1. Regulatory gene expression in Bacteria

Bacteria sense and integrate a plethora of external signals (Input) to produce complex regulatory pathways. Those pathways allow the properly gene expression in space and time resulting in a complex network which optimizes the fitness of the bacterial cell (i.e., Output of the process) (**Figure 1**). Gene expression regulation studies had pointed at two major mechanisms of regulation. One is the physiological level which supports changes in gene expression in function of physical constrains such as growth (Klumpp et al., 2009). The other is the regulated gene expression mechanisms, which support the existence of two programs of regulation called global and local regulation, both controlled by protein interactions at an upstream region of cognate genes called promoter sequence (Gottesman, 1984).

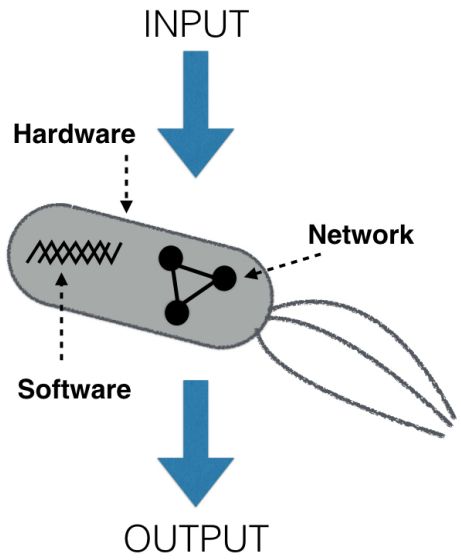


Figure 1: General Bacteria response. An environmental signal (input) is sensed by bacteria (the hardware), and modulates the expression of bacterial genome (the software) in order to activate/repress the expression of key genes which modulates a specific pathway, resulting in a complex network exerting a specific output to optimize bacterial fitness.

In bacteria, the rate of cell proliferation and the level of gene expression are intimately intertwined (Scott et al., 2010; Scott and Hwa, 2011). Schaechter et al., in 1958 demonstrated for the first time that macromolecular composition of the bacterial cell is related to its metabolic activity (Schaechter Et Al., 1958). In this sense, physiological characteristics such as cell size and macromolecular

compositions, e.g., the ratio of protein, RNA, and DNA have been related to growth conditions, the quality and availability of nutrients (Bremer and Dennis, 2008). In support, bacteria growing in different media but maintaining the same growth rate exhibited the same macromolecular composition (Klumpp et al., 2009). Additionally, changes in growth speed correlates with abundance changes in constitutive proteins as ribosomes and RNA polymerases. For instance, the total number of RNAPs per cell was determined to increase from 1,500 at slow growth (0.6 doublings per hour) to 11,400 at fast growth (2.5 doublings per hour) (Klumpp and Hwa, 2008).

Those intrinsic characteristics suggested that physiological constrains and gene expression affect one another. Quantitative predictions of those biological constraints demonstrated important effects in gene expression related to bacterial physiology (Scott, et al., 2010), pointing that microbial "growth laws" could explain part of the complexity of regulation of gene expression (Berthoumieux et al., 2013; Gerosa et al., 2013; Klumpp and Hwa, 2008, 2014; Klumpp et al., 2009; Lovén et al., 2012; Scott and Hwa, 2011).

1.1 Global transcription regulation

In general, for most of the genes in bacteria, recognition and transcription initiation of the promoter region by RNA polymerase is the pivotal step to modulate gene expression. The core RNA polymerase is an active catalytic complex of ~380-kDa composed of β and β'-subunit, two α-subunits with a flexible carboxy-terminal domain and ω-subunit. This complex can synthesize RNA and terminate transcription at appropriate sites. However, promoter DNA recognition and transcription initiation are dependent on a dissociable sixth subunit, the sigma factor (σ) (**Figure2A**) (Browning and Busby, 2016; Österberg et al., 2011).

In *Escherichia coli*, the housekeeping sigma factor *rpoD* (σ70) is responsible for the transcription of most promoters, while six other factors provide selectivity for the transcription of different sets of genes required for adaptive responses (Österberg et al., 2011). In this sense, *rpoS* (σ38) is the general stress factor associated with stationary phase and a variety of growth-impairing stresses; while *rpoH* (σ32) controls heat shock responsive promoters (Treviño-Quintanilla et al., 2013). Additionally, *rpoF* (σ28) controls flagellum functions, *rpoE* (σ24) controls responses to extreme heat

shock stress, FecI (σ19) transcribe the iron dependent promoters and *rpoN* (σ54) transcribe nitrogen assimilation dependent promoters (Österberg et al., 2011; Treviño-Quintanilla et al., 2013) (**Figure2B**). The different RNA holoenzymes in the cell are associated by competition. Thus, composition of the holoenzyme pools modulates gene transcription and provides the basis for other regulatory molecules and DNA-binding transcriptional regulators (activators and repressors) which fine-tune the final output of the promoter (For extensive review, see (Browning and Busby, 2016).

The complex that is formed by the binding of a sigma factor to the core-enzyme is named RNA polymerase holoenzyme (RNAP holoenzyme). The recognition and interaction of the double-stranded DNA by this complex is enough to unwind the DNA in the region of the transcription start site (positions +1 and +2). The unwound template single strand is located at the catalytic domain leading a transcriptionally competent open complex. At this point, nucleoside triphosphates (NTPs) are added initializing the synthesis of RNA transcript and eventually the DNA is pulled into the initiating complex (scrunching process). If scrunched complex is held at the promoter, initiation transcription initiation can be aborted and small RNA fragments are generated (Browning and Busby, 2016). When core-RNA polymerase scape at promoter, the sigma factor is released allowing the elongation of RNA transcript by the continuously addition of NTP's until RNA polymerase encounters a transcriptional terminator. Finally, all the components are released and the core-RNA polymerase is available to re-engage with a sigma factor and start the process again (Lee et al., 2012). It is worth to notice that promoter sequences share specific characteristics for the different holoenzymes. For example, the core-promoter sequence for the canonical *rpoD* holoenzyme presents specific characteristics that are sufficient to start the transcription. The sigma 70 factor recognizes the -10 (TATAAT) and -35 (TTGACA) elements, the extended –10 element (TGTG) and the discriminator element (GGG). The flexible carboxy terminal of α-subunits recognizes the UP element AAAWWtWtTTTNNNAAANN (located at –37 to –58 relative to the +1) (**Figure2C**). Additionally, promoter sequences present different discrete sequences motives called *cis-elements*. Those short DNA sequences are recognized by different transcription factors (TF), which modulates the affinity or accessibility of RNA polymerase holoenzyme at the core-promoter region in a mechanism called specific transcriptional regulation. Therefore, TFs regulates positively or negatively the gene transcription.

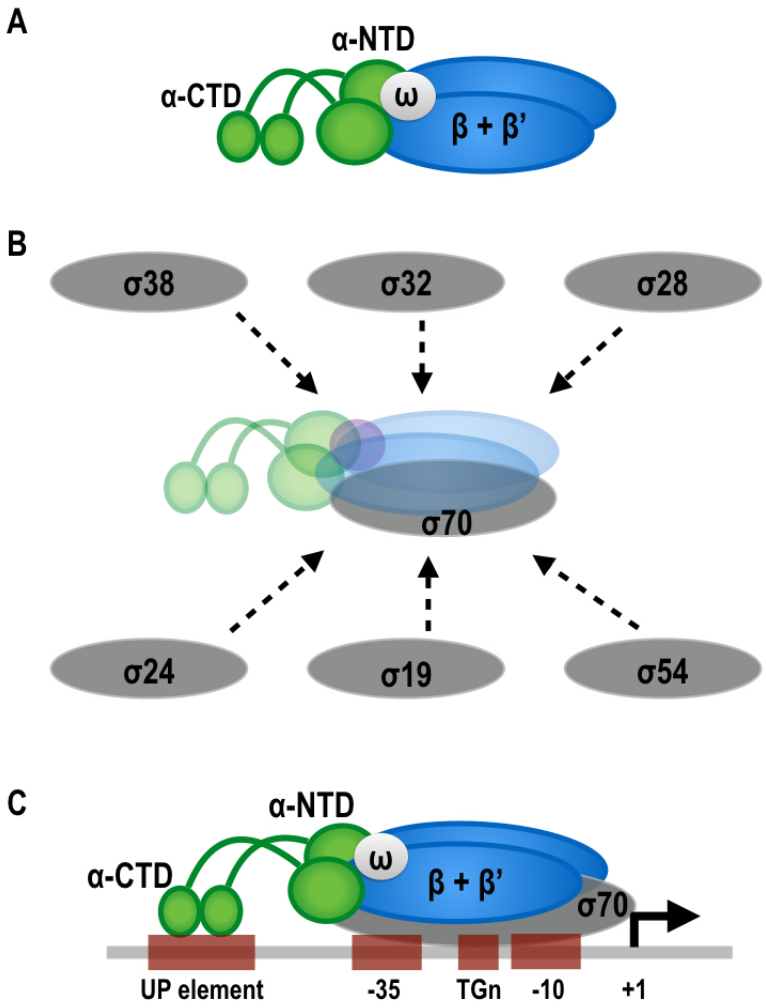


Figure 2: Holo-RNA polymerase drives global gene transcription. **(A)** In *E. coli*, the core RNA polymerase is composed of 2 β subunits, two α-subunits; NTD to N terminal domain and CTD to carboxyl terminal domain and one and ω-subunit, this complex have the capability to start transcription. **(B)** Selectivity at different promoters is given by the seven different sigma factors RpoS (σ38), RpoH (σ32), RpoF (σ28), RpoE (σ24), FecI (σ19) and RpoN (σ54) which responds to specific inputs and competes with the canonical sigma 70 (rpoD). **(C)** Specific parts of the sigma factors recognize specific places and sequences at promoter giving selectivity for specific promoters which means that each promoter presents a particular characteristic. For the canonical rpoD sigma factor the sequences are (when the transcriptional start site is denoted as +1), the –35 element (positions –35 to –30), the extended –10 element (Ext; positions –17 to –14), the –10 element (positions –12 to –7) and the discriminator element (Dis; –6 to –4). The UP element (positions –37 to –58), which leads specific mechanism of regulation. The sigma factor and of the carboxy-terminal domain of the α-subunit of RNA polymerase (αCTD) interact with those promoter elements as depicted.

1.2 Specific transcription regulation

Transcription factors activities are modulated by environmental signals. Those signals lead to changes in the abundance and availability of TF by induction of covalent modifications, regulation of their synthesis, turnover or sequestration (Seshasayee et al., 2011). Besides, proteins induced by external signals or small molecules, bind directly to transcription factors modulating their activity (Lee et al., 2012; Seshasayee et al., 2011). Altogether, homodimerization, dimerization or multimerization (among others) of TF leads to their association with *cis-elements* in the promoter region of target genes. This event induces different mechanisms of transcriptional regulation, which range from simply blocking access to the promoter, induction of supercoiled structures or direct association of TF with RNA holoenzyme to stimulate its activity (Browning and Busby, 2004; Gerosa et al., 2013; Lee et al., 2012).

In *E. coli*, there is a hierarchical organization of transcription factors. In this sense, TF that modulates many promoters are known as global regulators, those which regulates a group of promoters related to a specific function are called master regulators and in the lower level of regulation are specific transcription factors which regulates a single promoter (Martínez-Antonio and Collado-Vides, 2003). Because promoter regions presents different *cis-elements*, it is possible a simultaneous regulation by global, master and specific regulators, generating a complex regulatory network (Cosentino Lagomarsino et al., 2007). Therefore, understanding the complexity of a signaling pathway becomes in a scalar issue when a particular phenomenon is studied (Thieffry et al., 1998). Nonetheless, different mechanism for activation and repression have being described at promoter level (Browning and Busby, 2016).

The promoter activity is increased by three mechanisms. For the mechanism of promoter conformation change, the activator binds between -10 and -35 elements leading to transcription. This is possible since the space between the elements is not optimal, and thus the activator binds and flex the core-promoter element generating an optimal space allowing its recognition by the holoenzyme **(Figure 3A)**. The second mechanism is when an activator binds upstream of the core promoter element and recruits the RNAP holoenzyme to induce activation. This mechanism is called class I activation. This mechanism is exclusive when suboptimal -10 and -35 elements are present at core promoter avoiding the proper association of holo-RNA. Therefore, polymerase recruitment at the core

promoter is by interactions between an exposed domain of an activator protein and the C-terminal domain of α-subunit of the RNAP holoenzyme (protein-protein) **(Figure 3B)**. Finally, in class II activation, the *cis-element* of the activator protein overlaps the -35 element of the core-promoter. Recruitment of the activator protein allows its interaction with different subunits of the RNAP except with C-terminal domain of α-subunit, which is able to bind to another proteins. Those characteristics allows that class II and class I activation work together leading to synergic effects at target promoters (Browning and Busby, 2016; Lee et al., 2012).

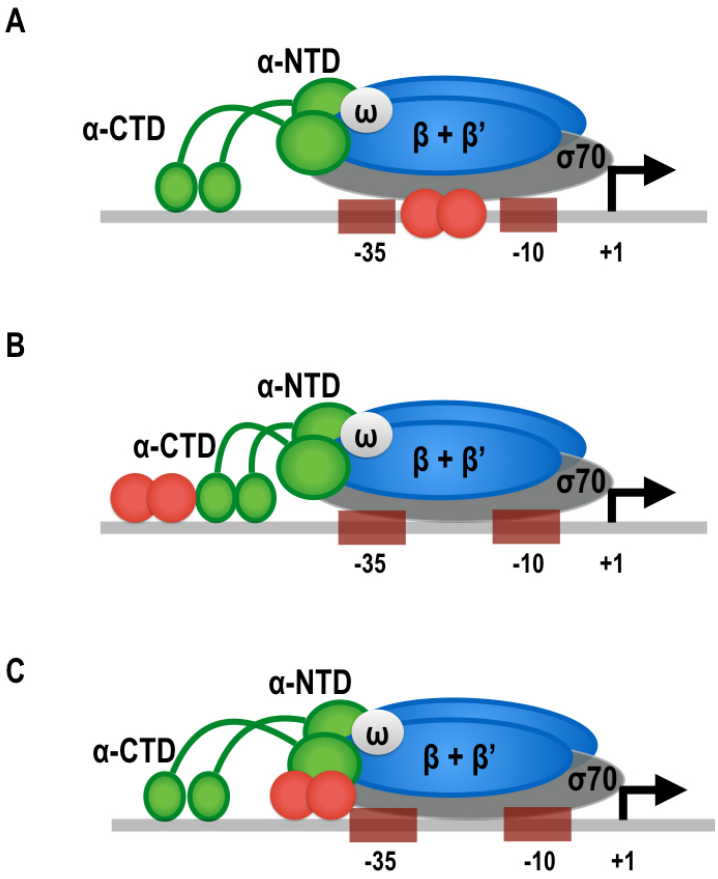


Figure 3: Specific gene transcription — activation mechanisms. **(A)** In the first mechanism, activators induce a conformational change in the promoter DNA to activate transcription. These activators bind at, or near to, the core RNA polymerase recognition elements of the promoter and often realign the –10 and –35 elements so that they can be recognized by RNA polymerase, recruiting it leading activation of transcription. **(B)** In class I activation, the activator binds to a site upstream of the promoter and recruits RNA polymerase to the promoter by contacting the carboxy-terminal domain of the α-subunit (αCTD). **(C)** In class II activation, the activator binds to a site in the promoter adjacent to (or overlapping with) the –35 element, where it recruits RNA polymerase through direct interactions with the sigma factor. Red circles represent the activators.

Three mechanism of repression by TFs have been described to the canonical σ70 promoter of bacteria, which represent most of the repression mechanisms for all promoters. The first repression mechanism is by steric hindrance. The repressor associates between -35 and -10 elements, blocking the capability of RNAP to recognize its core-promoter sequence **(Figure 4A)**. Second, in repression by looping, repressors recognize *cis-elements* upstream and downstream of the core promoter, those proteins interact between them inducing a DNA-loop and "hiding" the core promoter from RNAP recognition **(Figure 4B)**. The third mechanism occurs on promoters that are modulated by activators. In this case, the repressor acts over the activator protein and not by masking the promoter sequence, which means the repressor block the exposed site of the activator protein that interacts with RNAP avoiding its interaction and transcription stimulation **(Figure 4C)**. All those mechanism affects RNAP activity resulting in a repression of transcription. Interestingly, there are promoters with activity repressed constitutively by any of those mechanism and removal of repression activates the promoter or release the *cis-element* to be occupied by an activator leading to transcription (Browning and Busby, 2004, 2016).

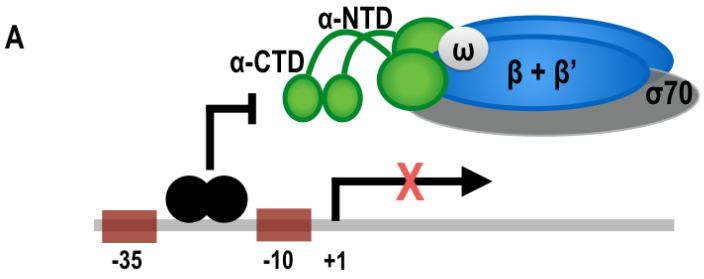


Figure 4: Specific repression mechanisms of gene transcription. **(A)** In repression by steric hindrance, the repressor binds to a site that overlaps the core elements of the promoter that is recognized by RNA polymerase blocking the transcription. **(B)** In repression by looping, protein–protein interactions are formed between repressors that bind to sites upstream and downstream of the (Konstantinidis and Tiedje, 2004)promoter, leading looping of the DNA affecting the recognition of promoter elements by RNA polymerase repressing the transcription. **(C)** Genes that require activators for transcription, repressors can modulate activators to prevent recruitment of RNA polymerase or by blocking association of the activators. Black circles are the repressors.

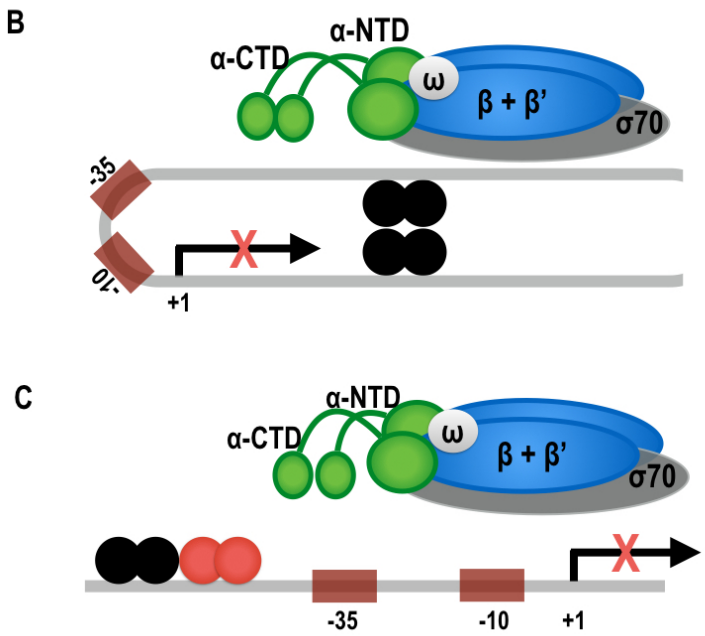


Figure 4: Continue.

Bacteria respond rapidly to environmental constraints because its capability to change, in short time, its gene expression patterns. As stated, there are different mechanisms of gene expression regulation in which specific, master and global transcription factors are involved. There are nine global regulators in bacteria, CRP, FNR, IHF, Fis, ArcA, NarL, Lrp, H-NS and Fur all of them binds to DNA and modulates the transcription of pivotal genes with important biological functions (Martínez-Antonio and Collado-Vides, 2003). We are interested in CRP, IHF and Fis because of their characteristics as describe below.

Bacteria can selectively use substrates from a mixture of different carbon sources trough a rapid reorganization of its cellular metabolism. Carbon catabolite repression (CCR) is a well known process of metabolic reorganization in which, in the presence of glucose, bacteria prevents the expression, and often also the activity, of catabolic systems that enable the use of secondary substrates. For instance, in pathogenic bacteria, CCR inhibits expression of virulence factors (for extensive review , see (Görke and Stülke, 2008). CCR mechanism is mainly controlled by cyclic AMP (cAMP) and CRP global regulator (cAMP receptor protein). On standard conditions, cAMP is produced by adenylate cyclase (AC) and eventually is recognized by the homodimer CRP receptor, the resulting complex

cAMP–CRP is able to activate the promoters of many catabolic genes. When glucose is present the amounts of cAMP decrease impairing the formation of the active complex. Therefore, the activity of CRP is diminished. This causes repression of a set of catabolite-sensitive operons such as lac, ara, mal and others (González-Gil et al., 1998; Görke and Stülke, 2008). These characteristic allows experimentally controlling CRP activity by adding glucose to media.

The genetic information of bacteria, like others domains, needs to be organized for storage in ways that are compatible with DNA replication, expression and segregation. The 1.5 mm of DNA in bacteria is condensed about 1000-fold by constrained supercoiled loops, random coiling, and binding of cations and proteins (Wang et al., 2013). The nucleoid-associated proteins (NAPs), which have not any structural similarities with histones of eukaryotes, can profoundly affect the local as well as global structure of the chromosome (Dillon and Dorman, 2010). Furthermore, NAPs have been related to specific DNA reactions such as the regulation of transcription and DNA recombination. The main NAPs are integration host factor (IHF), histone-like nucleoid-structuring protein (H-NS), suppressor of td mutant phenotype A (StpA), factor for inversion stimulation (Fis), DNA-binding protein from starved cells (Dps), and leucine-responsive regulatory protein (Lrp). From those, IHF, Fis, H-NS and Lrp are also global regulators (Browning et al., 2010; McLeod and Johnson, 2001). Interestingly, IHF and Fis are expressed differently during bacterial growth, making them an important study model to understand the effects of any phenomenon during bacteria development.

The integration host factor (IHF) is expressed in a growth rate dependent manner. The α-subunit and β-subunit generates the IHF heterodimer which binds to AT-rich domains introducing a u-turn into the DNA. This effect quenches the labile DNA strand of nearest sites but, in contrast, helps to the formation of the open transcription complex at distal promoters (McLeod and Johnson, 2001). Additionally, IHF is capable to recruit RNA polymerase at promoter regions. Altogether, IHF influences at the global and specific transcription in bacteria (Dillon and Dorman, 2010; Freundlich et al., 1992; Travers, 1997).

Fis (Factor for inversion stimulation) is maximally expressed in the early exponential phase of growth (McLeod and Johnson, 2001). Fis participates in different DNA-based cellular activities, including DNA transcription, replication and recombination. Fis homodimers recognizes an AT rich sequence at the promoter regions modulating its activity. This protein can repress promoter transcription by

blocking the core promoter sequence avoiding the recognition by RNA polymerase. Paradoxically, Fis can also interact physically with RNA polymerase, recruiting it at promoter regions activating transcription (Finkel and Johnson, 1992; Nasser et al., 2001)

2. Synthetic biology and basic network principles

Synthetic biology, a multidisciplinary discipline based on biology, chemistry, computer science, mathematics, engineering, among others, aims to simplify biological systems through abstractions, decoupling and standardizing each biological part (acid nucleic, protein or metabolite) in order to rewire, construct a new system or understand the impact of the network as a system in different cellular processes (Cameron et al., 2014; Chen et al., 2012; Lu et al., 2009).

In biological systems, proteins, DNA, RNA and small molecules present dynamic interactions that contribute to the structure and function of a living cell. Experimental data exerts evidence of how and when these molecules interact with each other giving an overview of a specific network studied. Besides, none of these networks are independent, instead they form a 'network of networks' that is responsible for the behavior of the cell. Because of this, understanding the topological and dynamic properties of different networks that control the behavior of the cell is a hard task. Interestingly, it has been observed that cellular functions with complex molecular interactions and architectures share the same principle than other complex systems as Internet, computer chips and society (Barabási and Oltvai, 2004). This make it possible to implement "network laws" to study regulatory interactions in order to understand a particular phenomenon (Barabási, 2009).

Networks theory is in continuously improvement, uncovering the organizing principles that govern the formation and evolution of technological and social networks. This is achieved by heuristic mathematical algorithms comparing the connectivity of each component of the network trough pairwise interactions. Nowadays, this field is already making an impact on cell biology (Barabási and Oltvai, 2004). Therefore, to implement network laws on biological systems it is necessary understand that cells are complex systems with an orchestrated activity of many components that interact with each other through pairwise interactions. For this, abstractions are required in order to reduce all the components of the system (proteins, DNA, RNA and small molecules) to a series of nodes which are

connected each other by links (Albert, 2005; Barabási and Oltvai, 2004). Those links represent the interaction between two components as protein-protein or acid nucleic-protein among others. Therefore, interactions or connections in a node is known as degree and will be further explained in detail. A defined number of elements and their relationships are called network or, in more formal mathematical language, a graph, and it is possible to analyze it by mathematical models or algorithms. Moreover, pointing directions into the network can be considered when a metabolic flow is being studied or when a TF activates a gene, but it is avoided in protein-protein interactions networks. When a group of nodes are physically or functionally linked to achieve a specific function it is called as motif, which is an advantage since it allows to analyze a motif separately from a network (*modularity*) (Barabási and Oltvai, 2004; Brophy and Voigt, 2014). This allows new interactions between hubs to be found and also the effect of any interaction into the network to be understood (**Figure 5**).

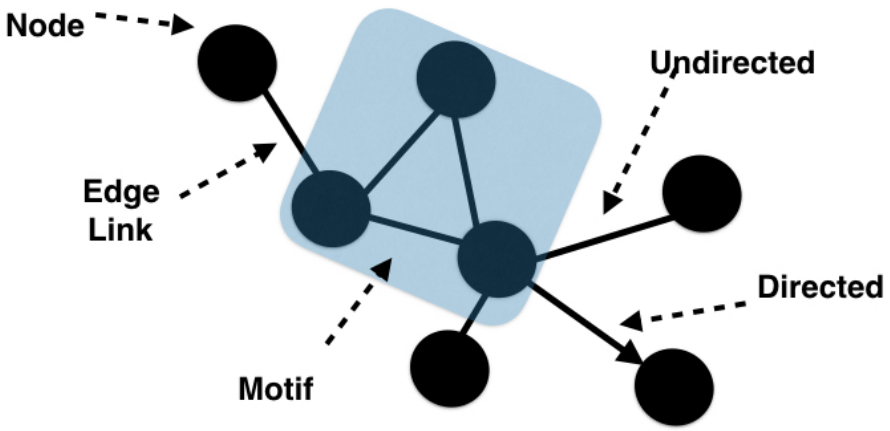


Figure 5: Basic network principles. The network is composed of nodes that can be any element of the cell (proteins, DNA, RNA and small molecules). Their biological interactions are represented as links with (black continuous arrow) or without directions (continuous black line). When a group of nodes develop a specific task, in a network, is known as motif.

2.1 Network models to biological systems

In network science, any system (social, economic, computer, etc.) with regulatory interactions between its elements can be represented by a network graph (Bollobás, 1979). Within a cell, all the elements that compound the system are in continuously regulatory interactions changing according to the presented inputs allowing the properly bacteria fitness (Albert, 2005). This has allow to implement network models in order to understand the structure of cellular interactions networks, providing better understanding of the biological studied system (Barabási and Oltvai, 2004).

Network science encompasses different theories which have allowed the understanding of different phenomenons as in society, epidemiology, etc., for biological phenomenas two main graphs families had provide significative knowledge on biological network research, mainly at metabolic networks and gene regulatory networks which are: linear graphs, and random graphs (Albert, 2005; Barabási and Oltvai, 2004; Newman, 2003). Linear graphs have been widely used as a model of an isolated signal transduction pathway. Those graphs are represented by nodes and edges connecting between them sequentially, absence of feedbacks is another characteristic. Thus, the graph increases linearly with the number of nodes **(Figure 6A)** (Albert, 2005).

The random graph family harbor more than 5 different models (poison, exponential, scale-free, etc), from those, the scale-free growing network model (Barabási and Oltvai, 2004; Price, 1965) presents suitable characteristics for biological networks. In this sense, random graphs were first proposed by Rapoport in 1957 to solve the net-structure (Rapoport, 1957). In 1960 Erdös–Rényi, independently, rediscover the random networks with applicabilities until nowadays (Barabási, 2003; Rényi, 1960). This model establish that all nodes in the network present the same or similar number of connections following a Poisson distribution **(Figure 6B)** (Barabási and Oltvai, 2004). Poisson graphs are used to understand social, informatics and technological but not biological networks (Barabási, 2003).

In 1965 Derek J. de Solla Price described what would now be called a scale-free network (Newman, 2003) by demonstrate that networks of scientific publications follow a power law distribution; for instance, there are some papers which present higher number of citations (Price, 1965). Interestingly, this concept was retaken by Barabási–Albert in 1999, whit extensive proves that biological systems are more complex than an equal number of interactions as described above (Albert et al., 2000; Jeong et al., 2001; Ravasz et al., 2002). In fact, biological systems follow a Power law distribution in

which higher connected nodes present higher probabilities to receive a new connections than low connected ones. This law is called “cumulative advantage” or “preferential attachment” (Barabási and Oltvai, 2004; Newman, 2003; Price, 1965) For instance, in transcriptional regulatory networks, the transcription factors activate or inhibit the transcription of specific or different genes. This implies directionality as well as different interactions with other nodes (Barabási and Oltvai, 2004; Keller, 2005).

The scale-free networks exhibit the following properties: i) contain a relatively small, but significant number of highly connected nodes, which are practically absent in homogeneous networks; ii) are self-similar i.e. any part of the network is statistically similar to the whole, iii) any two nodes can be connected via a small number of intermediate nodes called as “small-world behavior”, and iv) are highly tolerant to errors when nodes with low connections are removed but vulnerable when high connected ones are removed (Barabási, 2009; Wolf et al., 2002). Subsequently, centrality measurements and topology network analysis can be performed in order to understand the impact of any node on the behavior of all networks. Identifying the most important nodes and connections in a large complex network is of fundamental importance in a number of application areas. For instance, in genetics, neuroscience, proteomics, signal transduction studies, transportation systems, food-webs, sexual networking, epidemiology, economics, etc. (Goh et al., 2002; Hallinan et al., 2011; Strogatz, 2001).

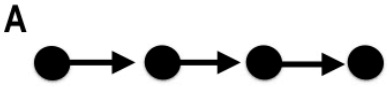


Figure 6: Linear and Random networks. **(A)** A linear pathway can be represented as a succession of directed edges connecting adjacent nodes **(B)** The Erdös–Rényi model follow Poisson distribution leading a randomly connected nodes, note that all nodes present between 1 ~ 3 links. **(C)** The Barabási–Albert model follows a power-law distribution “the richter get richter”; notice that there are major nodes which present 5 links.

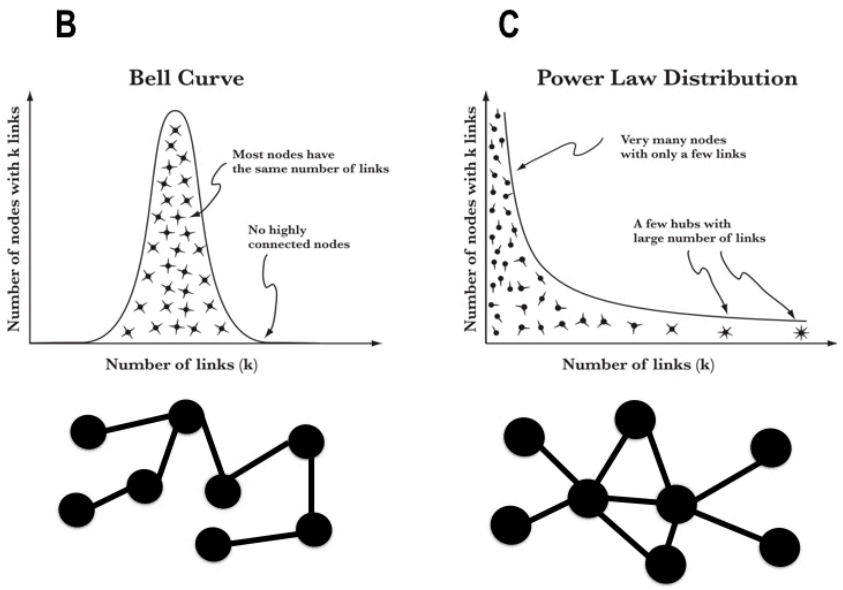


Figure 6: Continue

2.2 Analyzing the network — centrality measurements

Network studies address two mainly issues, centrality (which nodes are best connected to others or have most effect) and connectivity (whether and how nodes are connected to one another through the network). Those, depending upon the analysis, allows to identify the main effector nodes and paths, and also, how much tolerant a network could be upon edges or nodes remotion (Newman, 2003). Therefore, centrality graph measurements describes the organization of any network (Mason and Verwoerd, 2007).

The degree of a node in undirected graphs, is determined by simply quantification of the number of connections to each node (Mason and Verwoerd, 2007). In directed networks, degree centrality presents an incoming degree, which denotes the number of links that point to a node, and an outgoing degree, which denotes the number of links that start from it. (Hahn and Kern, 2005).

Betweenness centrality determine the frequency of node participation in paths connecting other components (Albert, 2005). Therefore, this node plays an important role to maintain the network performance. Remotion of higher nodes will destabilize the network. Proteins with high betweenness are called bottlenecks because they control the flow of information across a network and, thus, can

be important for minimizing response times within a cell (Jeong et al., 2001; McDermott et al., 2009; Ravasz et al., 2002).

Closeness centrality measures, indicates the average number of nodes connecting a protein to all other proteins, takes into account both direct and indirect interactions between proteins in the network. An important node is typically “close” to, and can communicate quickly with, the other nodes in the network. This measurement is good for undirected graphs and analysis for directed graphs must be interpreted carefully (Hahn and Kern, 2005).

Additional to nodes, links also exerts determinant effects into the network and disturbance at any link could affect the network performance. Therefore, an edge with a high edge betweenness centrality score represents a bridge-like connector between two parts of a network, and its removal may affect the communication between many pairs of nodes (Barabási, 2009).

2.3 Graph topology of a network

Biological networks can be construct by simple data as protein-protein interactions as well as using high-throughput experimental approaches. Whatever the case, networks may facilitating both visual browsing and computational analysis (Hallinan et al., 2011). Graphs are abstract of mathematical objects designed for describe relations between objects to visualize information (Kaufmann and Wagner, 2001). As established, centrality graph measurements identifies the main effectors nodes and paths in the network, these data, must be graphed in order to it properly visualization (Barabási, 2009). Therefore, nodes may be drawn as dots, circles, boxes, or by their name labels. Edges may be drawn as straight lines, orthogonal polygonal paths, arbitrary polygonal paths, or arbitrary curves. Another important characteristics to identify the components of the network such as add name to nodes and edges, different colors, thickness of lines, size of boxes, etc., are important to the properly graph visualization (Goh et al., 2002; Kaufmann and Wagner, 2001).

The hierarchical layout algorithm is used for representing main direction or “flow” within a network. In this algorithm, a first layer of main nodes with higher connections are plotted; subsequently, different layers of lower connected nodes are added avoiding edge crosses (Clauset et al., 2008). For instance, in a bacterial gene transcription network, a hierarchical graph display the major nodes at

the top of the graph (global regulators) spreading regulatory signals to lower connected nodes (specific transcription factors and genes) (Görg C., 2005; Hallinan et al., 2011). This representation when combined with other statistical tools as maximum-likelihood approach and Monte Carlo algorithm allows accurate predictions of interactions into the network **(Figure 7A)** (Clauset et al., 2008).

Methods based on physical analogies consider repulsive forces between all nodes, but also attractive forces between nodes which are adjacent (Kobourov, 2012). This model analyzes the network based in two physical concepts; "The Springs" and "Force-Directed Placement" it also consider the lengths of shortest paths between nodes. Improvements to the heuristics models have allow ameliorate the representation of those networks (Görg C., 2005). Basic force-directed approaches are suitable for to small graphs, but not to bigger ones (Kobourov, 2012). Because of this, we used organic graph algorithm (cytoscape 3.0) which combine the spring-embedded algorithm with elements of other algorithms to show the clustered structure of a graph (Kobourov, 2012) **(Figure 7B)**.

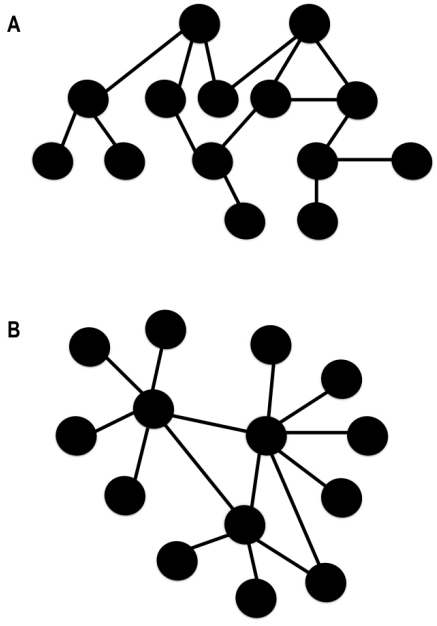


Figure 7: Hierarchical and Organic Networks. **(A)** Hierarchical network represent the flow of the information with hierarchical position of the must important nodes. **(B)** The Organic algorithm present an equilibrated network, this allow generation of cluster of related nodes. Notice that both network present the same number of nodes and links.

2.4 Bacterial promoters to understand transcriptional regulatory networks

Synthetic biology aims a predictive understanding of the biological systems. With the availability of high throughput transcriptional regulatory databases as RegulonDB and EcoCyc-DB, it has been possible to describe the functional connections in the complex regulatory transcriptional network of *E. coli*. Those studies proved that the transcriptional regulatory network structure of *E. coli* has a scale-free power-law degree distribution with a hierarchical arrangement in which the upper nine GR controls the most of the lower-tier nodes (more than 500 genes), which are more sparsely connected in the network (Isalan et al., 2008; Martínez-Antonio and Collado-Vides, 2003; Martínez-Antonio et al., 2008). That control is trough direct or indirect interactions exerting a fine-tuning modulation of different biological functions.

There is an observed variation in the activity depending on variation of other related genes, and this variation can be measured by RNA abundance using transcriptome data. There are some constraints to be considered in this technique such as the inaccurate correlation between protein and mRNA dynamics concentrations over time and different half-life time of the molecules. First attempts to identify gene activity and interactions under a particular input were developed using lac-Z gene fusion techniques; however, the principal limitation is the requirement of cell lysis, making it laborious to follow the dynamic of activity (Gottesman, 1984; Kalir et al., 2001; Stefan et al., 2015). The first system for real-time monitoring of the transcriptional activation over the time course of living cells was employed to analyze a specific structure network and it was developed by Kalir et al. in 2001. The system consisted of a low-copy plasmid with a fluorescent reporter gene and the promoter regions of flagella network of *E. coli* were cloned upstream of it. Kalir et al. in 2001 proved that this system was accurate to classify the precise order of transcription of the various flagella operons, which correlates with the observed order of the spatial position of gene products during flagellar motor assembly. Later, Ronen, et al., in 2002, used the same system coupled with mathematical models to evaluate the regulatory network of the SOS DNA repair system. They showed that relative protein levels might be determined from purely transcriptional data, suggesting that this technique will provide the dynamics flow of gene expression, assigning kinetic parameters to transcriptional networks on a genomic scale (Ronen et al., 2002). Recently, Stefan et al., in 2015, used transcriptional fusions (promoter + reporter gene) coupled with mathematical models in which they considered a higher number of biological constraints such as concentrations of RNA polymerase and ribosome, gene

copy number, amino acid sequence, protein half-life, nucleotide pools, among others, to prove that integration of all this parameters allows determine the differences between mRNA and protein concentrations as well as global physiological effects provide a reliable reconstruction of transcriptional regulatory networks in bacteria.

Networks laws support that complex systems display a tolerance degree against errors (Albert et al., 2000). Meaning that removing of low connected nodes will have less impact in the network contrary to removing of highly connected nodes. Therefore, perturbation at any level, which exerts expression changes in highly connected genes, would impact the network drastically. Seeking to understand that effect, most of the studies has depleted the highly connected genes and evaluated the changes in RNA expression profiles suggesting activation or inhibition effect over any gene based on RNA change expression levels. In addition to the achievement of the great understanding about the metabolic fluxes networks in bacteria (Chua et al., 2006), there is still the question whether those network descriptions reflects the predictive power over the effect of perturbations at one or more nodes of the network. Elegantly, Baumstark, et al., 2015, used 85 combinatorial promoter-TF constructions to rewire 85 different networks in *E. coli*, found that just around 20 underlying potential states of the system account for most of the observed changes. This was one of the first studies using combinatorial promoter regions driving different TF over expression with transcriptional analysis to understand the tolerance to rewire networks in bacteria. This study demonstrated the impact at topology network by perturbing a specific node, providing analytical information for studying *E. coli* gene expression in response to rewiring (Baumstark et al., 2015).

The use of transcriptional fusions allows the understanding of dynamic gene expression for a specific promoter (which reflects the effect over that specific gene) in the presence of any constrain as input changes or perturbation at any node of the network. This makes transcriptional fusions a suitable technique to understand modulations or changes into a specific network.

3. Biofilm formation in Bacteria

The world we live in is only possible as a result of the evolution and innovation of metabolic products generated by both lonely and sociable microorganisms. These products resulted from different

processes like carbon, sulfur, light and nitrogen cycles, which are carried out by different microorganisms as Archaea and bacteria (Shih, 2015). Therefore, without these processes, life as we know would not exist. Bacteria, unicellular planktonic cells with binary fission and ability to colonize different environments, have been considered and studied at laboratories as lonely microorganisms by many years. Nevertheless, in nature, bacteria perform cell-cell contact, aggregate and produce different signaling molecules facilitating communication between individual cells, thus leads to complex structures of multicellular communities called biofilm. This is found in medical, industrial and natural environments, and also can be engineered in vitro for a broadly use in biotechnology. (All reviewed in (Laverty et al., 2014; Serra et al., 2013; Shapiro, 1988, 1998; Srivastava and Bhargava, 2016; Williams et al., 2007; Zargar et al., 2015)

This bacterial higher order of organization is encased in a self-produced polymeric matrix which is constituted by polysaccharides, amyloid fibrils, proteins, nucleic acids, phospholipids and humic substances providing different advantageous characteristics for its community (Hufnagel et al., 2015; Shapiro, 1998). Physiologically, biofilms present a heterogeneous complex structure modulated in space and time, permitting communication between cells and presenting continuously sharing of molecules allowing to control the quorum of a bacteria population (known as a quorum sensing QS). Besides, channels are generated allowing a continuous flux of water, oxygen and nutrients and harmful substances. Altogether, biofilms provide important nutrient functions and acquisition of traits functions as horizontal gene transfer (Claessen et al., 2014; Petrova and Sauer, 2016; Samanta et al., 2013). As a result, the protective function in the biofilm is improved providing resistance against antibiotics, pH shifts, environmental stress, oxygen radicals, UV radiation, metal toxicity, phagocytosis, desiccation resistance, acid exposure, detergents, protection against predators, and host immune defense response substances (DePas et al., 2014; Monds and O'Toole, 2009).

Those characteristics make biofilms attractive for research and of great economic importance mainly due to antibiotic resistance. In this sense, the Center for Disease Control and Prevention (CDC) estimates occurrence of 2 million of bacterial infections in the United States and 48,000 deaths attributable to antibiotic-resistant infections each year in the United States and Europe reaching $20 billion and €1.5 billion of costs, respectively (Center for Disease Dynamics, 2015; U. S. Department of Health and Human Services. Centers for Disease, 2013). At laboratories, additionally to the importance established above (understanding the mechanisms of antibiotic resistance), biofilm also

allows synchronizing particular behaviors on a population at wide-scale, considered interesting to study as a multicellular-like organism, which is comparable to Eukaryotic life style (Claessen et al., 2014; Hufnagel et al., 2015; Monds and O'Toole, 2009; Srivastava and Bhargava, 2016).

3.1 Escherichia coli: stages of biofilm structure formation and signaling pathways

It is clear that bacteria modulate its population density, morphology and gene expression patterns in function of different environmental signals as temperature, pH, osmolarity, oxidative stress, starvation, and quorum sensing signals (Petrova and Sauer, 2016; Srivastava and Bhargava, 2016). These signals allow bacteria to shift between planktonic and biofilm states. Specifically, these transitions have been described in more detail in *E. coli* and involve, (i) planktonic state, (ii) reversible followed by an irreversible attachment phase at live or nonlive surfaces, (iii) aggregation or microcolony formation, (iv) maturation phase and (v) escape of biofilm (Beloin et al., 2008; Laverty et al., 2014; Solano et al., 2014) (**Figure 8**).

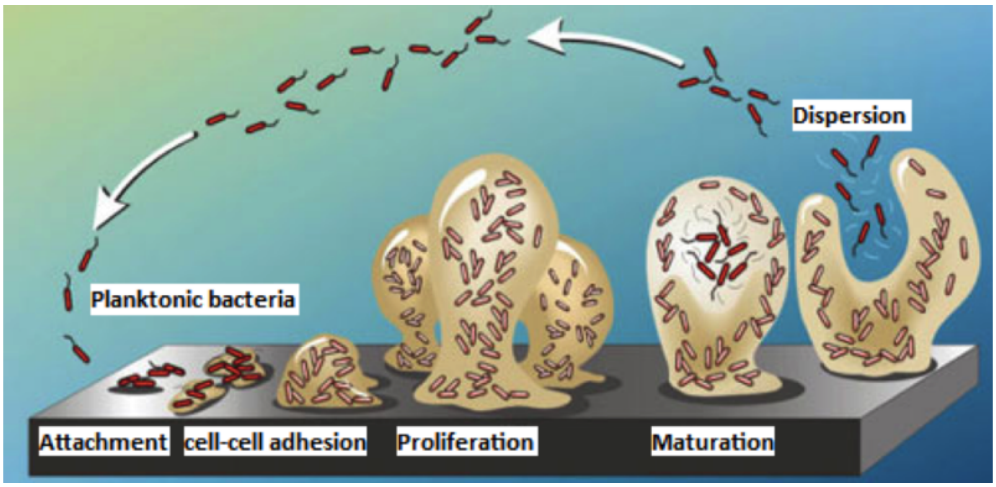


Figure 8: Phases of biofilm development. The biofilm development is stratified as (i) planktonic state which are highly motile cells, (ii) reversible followed by an irreversible attachment phase at live or nonlive surfaces which is present as attachment phasethe, (iii) aggregation or microcolony formation in which cell-cell adhesion take place followed of a proliferation state, (iv) maturation phase represented as maturation and (v) Biofilm scape or Dispersion. Each phase is related with specific gene expression and morphological changes as described in this chapter. (Adapted from (Stoodley et al., 2002)

All those phases of differentiation present changes at both growth phases inside of biofilm as well as changes on gene expression patterns that allows expression of different proteins and signaling molecules. The interaction between those features modulates the shift from planktonic to biofilm formation (Hosseinkhan et al., 2015; Mauter et al., 2013). Many different signaling pathways and molecules that control those states have already been described, like the master regulators *rpoS*, *csgD, rpoE* and *flhD* (Dudin et al., 2014; Liu et al., 2015; Ogasawara et al., 2010; Ogasawara et al., 2011; Serra et al., 2013). It is known that these regulators are considered indispensable to modulate and control this shift, and they will be described below. However, how signals are integrated at the promoter level of these regulators during the shift from planktonic to biofilm formation has not been fully elucidated yet (Claessen et al., 2014; Guttenplan and Kearns, 2013; Laverty et al., 2014; Monds and O'Toole, 2009; Petrova and Sauer, 2016; Srivastava and Bhargava, 2016) (**Figure 9**).

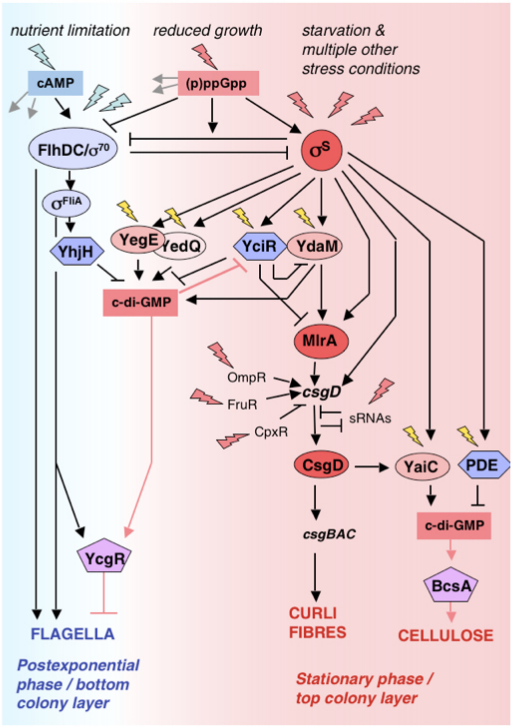


Figure 9: Signaling pathway of the flagella-biofilm program. cAMP and ppGpp drives the regulation of two intricately signaling pathways which modulate the flagella and biofilm program. Those small molecules activates the master regulators *rpoS* and *flhDC* that eventually activates downstream effectors. Flagella genes are related with motility, adhesion and cell-cell contact. At stationary phase, *csgD* is active by *rpoS* and other effectors. This allow the activation of the biofilm regulatory pathway (Adapted from (Serra et al., 2013).

3.1.1 Planktonic stage

Of the five phases mentioned above, the best documented are planktonic and attachment stages. Planktonic phase is characterized by synthesis of flagella and motility, giving advantages to overcome repulsive electrostatic forces allowing cells to swim in liquid or to swarm over solid surface (Laverty et al., 2014). This allows some bacteria to colonize different environments and, depending of the species, will lead to an infection, illness or medical difficulties. Flagellum is a complex protein system assembled from over 30 proteins and is regulated at both gene expression and chemotaxis level (Guttenplan and Kearns, 2013). Also, it is considered the most powerful motor in the earth generating between 400-3800 pN-nm of torque depending on the species. Thus, its generation and function requires a high-energy investment (Guttenplan and Kearns, 2013). Flagella genes are organized into 3 classes: first class genes, which are positively controlled by the global transcription regulator CRP, encode the master transcriptional regulator FlhDC, which controls the number of flagella and activate the transcription of the second and third classes of genes; second class genes that are involved in hook-basal body complexes in the morphogenic pathway of flagellar structure (*flgAB, flhB, fliAEFL*) and; the third class operons that are responsible for formation of the filament, the final step of flagellar assembly (the *flgK, fliD,* and *fliC* operons) or for the function of the complete flagellum (the *motA* and *tar* operons) (All reviewed in (Guttenplan and Kearns, 2013; Kutsukake et al., 1990; Laverty et al., 2014; Yamaguchi et al., 1986).

3.1.2 Reversible and irreversible attachment phase

Flagellum is “rapidly” inhibited at the functional level and “slowly” inhibited at gene expression level (Guttenplan and Kearns, 2013). Phosphodiesterases (PDE) and diguanylate cyclases (DGC) are continuously and broadly expressed in bacteria; those enzymes induce degradation and production of the small cytoplasmic signaling molecule cyclic-di-GMP (c-di-GMP). Changes in this molecule are highly associated with the transition oh both planktonic and biofilm phases (Sommerfeldt et al., 2009; Valentini and Filloux, 2016). In this sense, there are 29 genes coding different PDE and DGC in *E. coli* that controls the levels of c-di-GMP (Sommerfeldt et al., 2009); However, the inputs that controls those levels are poorly understood. Additionally, c-di-GMP interacts directly with YcgR protein, which

attaches to the flagella motor and inhibits its function (fast inhibition), thus stimulating the shift from planktonic to biofilm state (Gao et al., 2016; Guttenplan and Kearns, 2013; Sommerfeldt et al., 2009).

Once flagella function has been rapidly disrupted, expression of *fim* and *mat* operons leads to the production of type I, type IV and P pili, and these structures allow the reversible attachment of *E. coli* at different surfaces. This reversible attachment is characterized by association at the cell surface through a single pole making possible to return to planktonic phase. For example: type I pili has a mannosespecific FimH receptor at the tip, which is responsible for adhesion to host tissue surfaces as the bladder epithelium, resulting in cystitis (for an extensive review read (Laverty et al., 2014; Martínez-Santos et al., 2012).

By another hand, the "slowly" flagella inhibition mechanism (gene expression regulation level) is modulated by the two-component systems EnvZ/OmpR for osmoregulation and RcsCDB system for colonic acid production (Laverty et al., 2014). OmpR positively regulates curli (extracellular proteinaceous structures extending from the cell surface for attachment during biofilm development) and Rcs positively regulate type I fimbriae and colonic acid capsule. Interestingly, those genes are differentially expressed during developmental phases contributing differentially in space and time to modulate the sessile but also the release from biofilm phases (Petrova and Sauer, 2016; Samanta et al., 2013). Thus, as the levels of c-di-GMP and OmpR levels increase, they associate at the promoter region of CsgD leading its activation (Laverty et al., 2014). CsgD is the master regulator of curli fibers. These proteinaceous adhesive filaments are responsible for attachment at different surfaces and to cell-cell attachment being the first level of irreversible attachment phase. Bacteria in this phase are deprived of flagella and present rood-shaped format (Guttenplan and Kearns, 2013; Hufnagel et al., 2014, 2015; Saldaña et al., 2009).

3.1.3 Aggregation or micro/macro-colony formation

At aggregation/microcoloni phase, the CsgD regulator controls cellulose production through *adrA* gene activation, leading the formation of an extracellular matrix (Laverty et al., 2014). The extracellular matrix as well as colonic acid produced by Rcs abets the establishment of a more complex multicellular structure characterized by the presence of differentiated pillar-like structures

with fluid-filled channels, known as microcolonies (Davey et al., 2003). Additionally, Ma, et al., in 2009 have shown an alternative signaling pathway that is necessary to allow *E. coli* K-12 to efficiently attach on polystyrene (PS) surfaces (Ma and Wood, 2009). This system inhibits the "classic" cellulose production (CsgD-AdrA) and activates biofilm formation pathway to attach at hydrophobic surfaces (PS, PVC and PP). This pathway is modulated by the outer membrane OmpA that activate the CpxRA two component system. CpxRA two-component when activated inhibits *csgD* and *adrA* activation (Ma and Wood, 2009). However, OmpR activates cellulose production trough activate the GGDEF protein YedQ resulting in an alternative pathway that modulates biofilm formation at different surfaces (Da Re and Ghigo, 2006). Therefore, bacteria have the ability to perform biofilm formation at different surfaces by modulate its gene expression patter.

3.1.4 Maturation phase

The maturation phase is the most complex state because requires temporal and spatial modulation of bacteria community morphology in function of the demand of biofilm growth, as well as by external stimulus as nutrients availability and stress signals. Some of these signals are produced by the accumulation of waste products and toxins generated by the biofilm itself, thus multiple regulatory networks are involve in this phase (Monds and O'Toole, 2009; Petrova and Sauer, 2016; Stoodley et al., 2002). Therefore, as the biofilm grows, the environment changes and stress signals are generated, triggering the activation of *rpoS* sigma stress factor (Hengge, 2009; Liu et al., 2015; Suh et al., 1999). This leads to the activation of the two-component system Cpx/CpxR. Consequently, Rcs system is activated and *rcsABCF*, *wca* and *ompC* operons are transcribed, allowing the higher production of colonic acid leading to biofilm maturation (Gualdi et al., 2007; Guttenplan and Kearns, 2013; Laverty et al., 2014). At the final step of maturation, quorum sensing systems are activated in order to the properly biofilm structure maintenance (Laverty et al., 2014; Solano et al., 2014). In *E. coli* there are four different quorum sensing systems. (i) The *sdiA* system (homologous to LuxR) senses and binds to N-acyllhomoserine allowing the activation of *ftsQAZ* operon that upregulate cell division in the biofilm (Shimada et al., 2013; Shimada et al., 2014). (ii) The *luxS*/autoinducer-II system that is modulated by *lsrRK* operon, its activation allows the autoinducer-II molecule production and eventually the blockage of adhesion and flagella related genes as *csgAEFG, htrE,* and *flgN* (Laverty

et al., 2014; Pereira et al., 2013; Torres-Escobar et al., 2014). (iii) An autoinducerIII/epinephrine/norepinephrine system that is more related to expression of virulence genes (LEE operon). Activation of LEE operon is controlled by epinephrine through activation of QsecCBEFAD proteins (Laverty et al., 2014; Sharma and Bearson, 2013). (iv) An indole mediated system that is generated by tryptophanase (TnaA) and its expression requires cyclic AMP. Indole is exported and imported by AcrEF and Mtr transporters respectively. It enhances expression of diverse xenobiotic exporter genes (*mdtAE, cusB, emrK, and yceL*) via two-component signal transduction systems (BaeSR and CpxAR), increasing antibiotic resistance and also modulating its own production (Kim and Park, 2015). Despite the "simplistic" way to describe the biofilm maturation, there are processes, mechanisms and signals that cause activation and repression that have not been fully elucidated.

3.1.5 *Escape from the biofilm*

Finally, detachment phase is necessary to bacteria survival. As biofilm grows, harm substances are present in the environment, making the survival of these trapped bacteria challenging. Thus, biofilm located cells developed mechanisms to leave the biofilm allowing both self and species preservation. The signals that modulate this natural process involve matrix degrading enzymes, surfactants, chelating agents, signaling molecules, and environmental cues such as nitric oxide, oxygen levels and variation in carbon and energy source availability (Petrova and Sauer, 2016). The mechanisms described are desorption, detachment and dispersion. RpoE sigma factor is activated under environmental stresses therefore it competes with RpoS. Furthermore, RpoE sigma factor is required to maintain cell envelope integrity and in *E. coli* has been shown to promote flagellar gene expression (Xie et al., 2016). Additionaly, CpxA/cpxR system is controlled by RpoS and negatively controls csgD therefore flagella-related genes are transcriptionally activated again (Laverty et al., 2014; Ogasawara et al., 2010; Ogasawara et al., 2011). The last reported regulation is given by c-di-GMP levels, and contrary to the mechanism describe above to biofilm formation, c-di-GMP decreases release the inhibition of *flhD* promoter, thus flagella production is activated (for extended review read (Petrova and Sauer, 2016).

4. The issue at stake

Biofilm structure is focus of attention due to its involvement in the production of different compounds and also due its protective effects in bacteria. At molecular level, it is known that global regulators are expressed differentially during bacterial growth, and even inside biofilm structure there are cells in different growth phase. Also, global regulators modulate genes of both flagella and biofilm network (Eisenstein et al., 1987; Martínez-Antonio and Collado-Vides, 2003; Venturi, 2006). This suggests that global regulators will present an effect inside the complex topology of the planktonic-biofilm network. In this context, we were interested in discovering how would the dynamics of genes of the planktonic-biofilm network behave in the presence or absence of one or more global regulators. This with main focus to understand the effect of CRP, IHF and Fis in the transcription regulatory network of ten main nodes that drives the planktonic-biofilm transition. Subsequently, we wondered what would be the effect of that network response to the biological (physiological) characteristics of bacteria. This aim disclosed the biological effects of the global regulators in the motility and biofilm formation program. Therefore; to achieve both objectives, we used a multidisciplinary approach, combining computational tools and gene reporter assays as well as morphological analyzes in order to solve those and other biological questions of biofilm development in bacteria. To the first aim, by using computational tools and a promoter-GFP reporter system, we disclosed the effect of the global regulators over the gene expression (activation or repression) of the main effector nodes as well as the behavior during the growth curve in a batch system which mimics the biofilm development. The network graph obtained was analyzed and the main effectors nodes and paths were obtained. This valuable information, helped to understand the biological differences observed at flagella and biofilm program obtained by our microbiological assays on mutant strains of IHF and Fis as well as in CCR induction. Altogether, we provided a better understanding of the effect of the global regulators CRP, IHF and Fis during the transition between planktonic to biofilm formation. And we prove that the flagella-biofilm transcriptional regulatory network tolerates remotion of IHF and Fis as well as down-activation of CRP; leading to, despite the affections, motility and biofilm formation.

II. OBJECTIVES

General objective

The general objective of this Thesis was to understand the effect of CRP, IHF and Fis global regulators in flagella and biofilm formation in the model bacteria *Escherichia coli*.

Specific objectives

1. Understand the transcriptional regulatory network of the flagella-biofilm program by using network analysis.
2. Disclose the effect of global regulators CRP, IHF and Fis on genes that modulate the transition between flagella and biofilm formation.
3. Integrate the newly discovered interactions into the existing flagella-biofilm network to gain a better understanding of the program to biofilm formation.
4. Investigate the outcome of motility and biofilm formation in bacteria in the absence of CRP, IHF and Fis.

III. MATERIAL AND METHODS

1. Bacterial strains

The list of bacterial strains is presented in Table1. *E. coli* wild type (BW25113) or strains genetically depleted of IHF or Fis global regulators (Baba et al., 2006)were used as the host of all plasmids. *E. coli* DH5α or DH10b were used as cloning strains. DNA of *E. coli* BW25113 and MG1655 were used as template for PCR amplification of DNa fragments of interest.

2. Media and growth conditions

E. coli cells were grown in Luria Broth (LB) or M9 minimal media (6.4g/L Na_2HPO_4•$7H_2O$, 1.5g/L KH_2PO_4, 0.25g/L NaCl, 0.5g/L NH_4Cl) supplemented with 2mM $MgSO_4$, 0.1mM $CaCl_2$, 0.1mM casamino acids, and 1% glycerol or 0.4% glucose as the carbon source. When required, chloramphenicol (Cm, 35 µg/ml) was added to ensure plasmid maintenance. Cells were growth at 37°C with constant shaking at 220 rpm overnight.

3. Plasmid construction

Plasmids and primers used in this work are listed on table 2 and 3 respectively. As a control of all experiments we transform *E. coli* BW25113 wild-type and mutant strains with pMR1 empty vector, pMR1-Placwt and pMR1-*Plac*Δ*crp* (Guazzaroni and Silva-Rocha, 2014). To analyze the signal integration of global regulators at the studied genes, we amplified the promoter or promoter plus the 5'UTR regions by PCR and cloned into pMR1 vector by standard protocols (Sambrook, 1989). DNA from *E. coli* BW25113 wild type was used as a template to all constructions except for *flhD* promoter, for which DNA from *E. coli* MG1655 was used. Polymerase Chain Reaction (PCR) was performed by use 50 ng of DNA template with 50pmol of each primer (purchased from Sigma-Aldrich and Extendd) and 1U Phusion® High-Fidelity DNA Polymerase (New England Biolabs) in 50 µl reaction volume. Initial denaturation was at 98˚C by five minutes; subsequently, amplification was done by 30 cycles at 98˚C, 60˚C and 72˚C of 30 seconds each one. Ten minutes at 72˚C was used as final extension step. All amplifications were cloned into pMR1 vector. Briefly, PCR fragments and plasmid DNA were purified and digested with EcoRI-BamHI (transcriptional fusions) or EcoRI-SphI (translational fusions), and posterior cloning procedures were done by standard protocols (Sambrook, 1989). We

were not able to clone *matA*. All cloned fragments were check by DNA sequencing using dideoxy terminal method.

4. Promoter activity assay

For promoter activity measurements, single colonies of wild type and mutant *E. coli* strains harboring the different plasmids were picked from fresh plates. Each strain was inoculated and grown overnight in 2 mL of M9 medium with glycerol or glucose containing chloramphenicol and incubated at 37 °C with shaking at 225 rpm and aeration. Stationary phase cultures were diluted to a final optical density (OD) of 0.05 in 200 µL of M9 medium containing either glycerol or glucose as required, and supplemented with chloramphenicol. Cell growth and GFP fluorescence was quantified each 20 minutes during 8 h while incubated at 37 °C using Victor X3 plate reader (PerkinElmer, Waltham, Massachusetts, USA). Data were analyzed from at least three biological samples, arbitrary units were calculated by normalized fluorescence values over optical density upon background subtraction. For data representation, average and standard deviations for each promoter were calculated and graphed using Microsoft excel and R software.

5. Bacteria motility assay

The effect of global regulators on motility was evaluated as follows. *E. coli* wild type or mutant strains harboring pMR1 plasmid were growth overnight at 37˚C in LB plates. Inoculation of single colony into motility plates (tryptone 1%, NaCl 0.25%, agar 0.3%) was done by using toothpick (Sperandio et al., 2002). When indicated glucose was added to the motility media. The motility halos were measured at 8, 12, 18 and 24 h. Each strain was evaluated by three independent experiments.

6. Biofilm formation in liquid media

Biofilm formation was assayed by Microtiter Dish Biofilm Formation Assay (O'Toole, 2011). Single colonies of *E. coli* wild type or mutant strains harboring all the constructions were growth overnight at 37˚C in M9 with glycerol or glucose. The cultures were diluted 1:100 in 200µl of fresh M9 media with either glycerol or glucose. 96 well plates were incubated at 37˚C by 72 h. Next, cells were gently washed four times with distilled water and left at room temperature by 15 minutes with 200µL of crystal violet solution at 0.1%. Subsequently, vigorously washed of the wells was performed with distilled water (four times) and left dry by 24 h at room temperature. Finally, 200 µL acetic acid 30%

was added and after 15 minutes, the solution was transferred at new plate and OD-550 was measured in VictorX3 plate reader (PerkinElmer, Waltham, Massachusetts, USA). Data from biological triplicates were analyzed by Microsoft excel software. Standard deviation was determined.

7. Biofilm formation in solid media

Biofilm morphology was evaluated by congo red assay according to (Zhou et al., 2013). Chloramphenicol was added during all steps to ensure plasmid maintenance. Wild type and mutant *E. coli* strains with pMR1 plasmid were growth overnight at 37˚C on LB agar plates. Single colonies were picked and growth 14 h. at 37˚C on 1ml LB at 220rpm. The cultures were washed twice with MgSO4 and resuspended in M9 with either glycerol or glucose to an OD600nm of 0.5. Five micro-liters drop of each culture was added into YESCA-CR plates (with few modifications (Zhou et al., 2013). YESCA media (1 g/L yeast extract, 20 g/L agar) was complemented with congo red at 50 μg/ml diluted in KPi buffer (50 mM potassium phosphate buffer, pH 7.2; 28.9 mM KH2 PO4, 21.1 mM K2 HPO4 in water). When indicated, glucose was added at 0.4% final concentration. YESCA-CR seeded plates were allowed to dry in sterile environment and allowed to grow at 37˚C for several days. Biofilm development was followed (documented) by 6 days using a Fluorescent Stereoscope Leica MZ16. Biological triplicated images were analyzed by ImageJ and prisma software was used to determine the statistical significance of the experiments by using one-way ANOVA .

8. Network construction

The interactions (activation or repression) of proteins and genes that modulate the transition between planktonic to biofilm formation were collected manually from scientific research publications. Therefore, is possible the human error to identify all the genes in the pathway. Next, the connections were validated with KEGG pathway DB. All gathered information was represented in the website (analysis don't shown). Subsequently we converted as activation all reports of molecular or biochemical evidence of over expression; when contrary, the effect was consider as repression. Cytoscape 3.4.1 was used to generate the transcriptional regulatory network with slight modifications (Cline et al., 2007). Briefly, with our gathered data of activation and repression. We construct a Random graph by use the orthogonal algorithm for properly cluster visualization. Next, the generated networks were analyzed by centrality measurements, p. ie. degree, betweenness centrality, Partner of multi Edge Node Pairs and topological coefficient. Finally, we transform our GFP dynamics data

obtained from our promoter activity assay into a functional interaction (inhibits or activates) and those, were integrated into the network. Same centrality measurements were performed.

9. *In-silico* analysis of promoter regions

In-silico cloning of the promoter regions (and promoter regions plus 5'UTR) of the genes which drives the flagella-biofilm differentiation were done by select the respective promoter region of *rpoS, rpoE, csgD, cpxR, flhD, fliA, ompR, matA, yeaJ,* and *adrA* from KeggDB. The individual sequences were in-silico cloned upstream of a GFFP-lva sequence. RegulonDB databases use high throughput datasets and computational predictions to determine the *cis-elements* of any promoter of *E. coli*. Therefore, we obtain the *cis-elements* from each promoter at regulonDB and located them within our *in-silico* constructions. All the analysis were performed using A Plasmid Editor (APE) software.

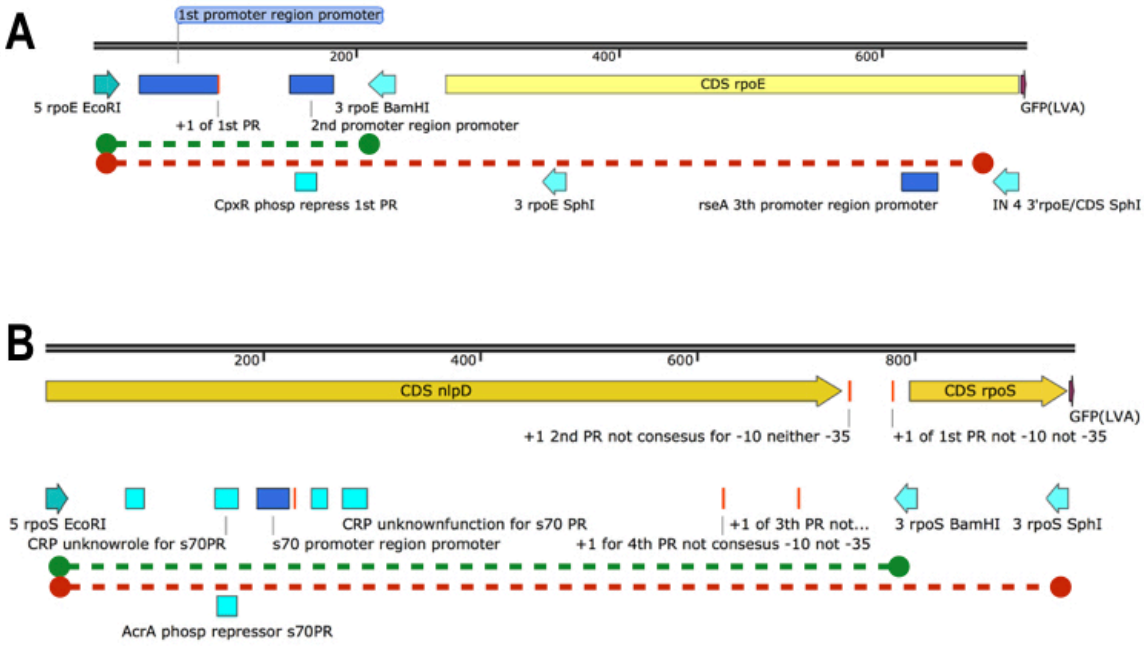


Figure 10: *cis-element* promoter architecture of main effectors of the flagella-biofilm network. The cloning procedure was performed in vitro and the sites at RegulonDB were loaded into the promoter region. Double black line (DNA), yellow boxes/arrows are the open reading frame (ORF) next to the core-promote in some cases a part of promoter region was embedded in the ORF. Blue boxes are the *cis-elements*. Green segmented line represents the transcriptional fusion and red-segmented line indicates translational fusions.

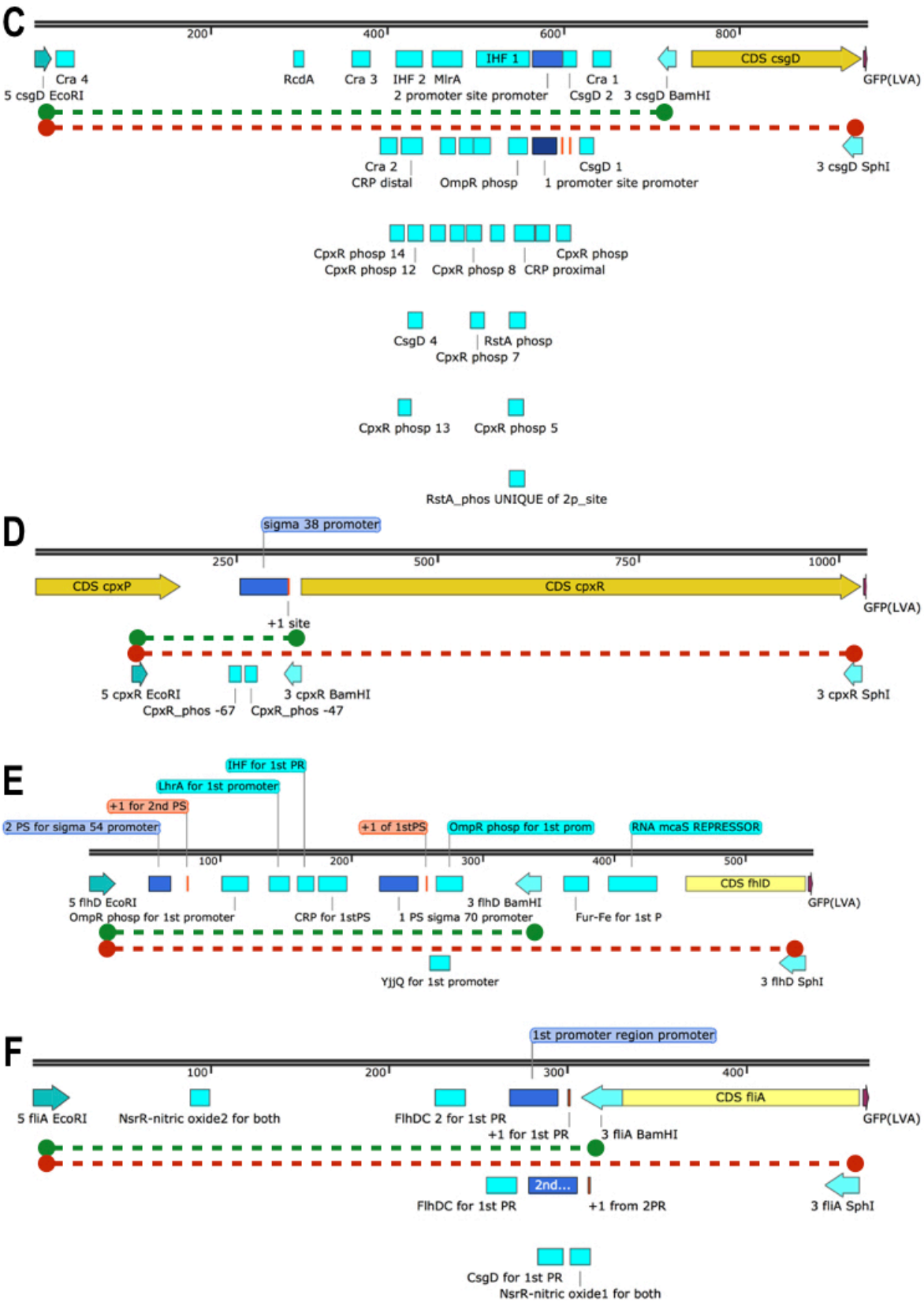


Figure 10: ***Continue***

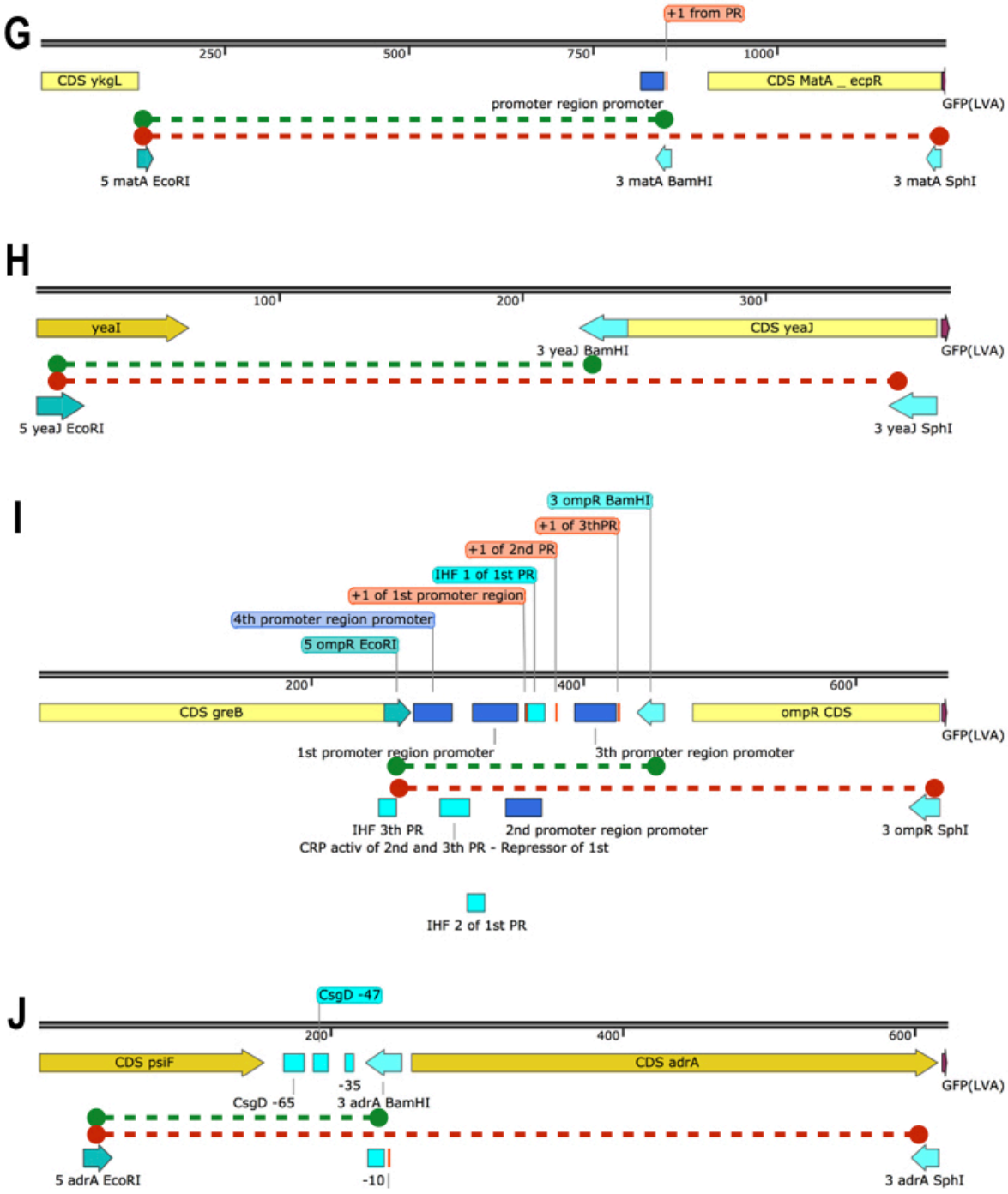


Figure 10: ***Continue***

Table 1. Strains Used In This Study

strain	description	reference
E. coli DH5α	*F− endA1 gln V44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(rK− mK+), λ−*	Grant, S. G., et al., 1990
E. coli BW25113	*Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ−, rph-1, Δ(rhaD-rhaB)568, hsdR514. Wild type strain*	datsenko 2006
E. coli JW1702-1	*Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ−, ΔihfA786::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514. ΔihfA mutant strain*	datsenko 2006
E. coli JW3229-1	*Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ−, Δf is-779::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514.Δfis mutant strain.*	datsenko 2006
E. coli MG1655	*F− λ−ilvG- rfb-50 rph-1*	Bachmann, 1972

Table 2. Plasmids Used In This Study

plasmids	description	reference
pUC19	ApR, oripColE1, lacZα; standard cloning vector	Yanisch-Perron, C, et al., 1985
pMR1	CmR, orip15a; GFPlva promoter probe vector	Guazzaroni & Silva-Rocha 2014
pMR1-PrpoS	CmR, orip15a; pMR1-rpoS-GFPlva transcriptional fusion	This study
pMR1-PrpoE	CmR, orip15a; pMR1-rpoE-GFPlva transcriptional fusion	This study
pMR1-PcsgD	CmR, orip15a; pMR1-csgD-GFPlva transcriptional fusion	This study
pMR1-PflhD	CmR, orip15a; pMR1-flhD-GFPlva transcriptional fusion	This study
pMR1-PfliA	CmR, orip15a; pMR1-fliA-GFPlva transcriptional fusion	This study
pMR1-PyeaJ	CmR, orip15a; pMR1-yeaJ-GFPlva transcriptional fusion	This study
pMR1-PadrA	CmR, orip15a; pMR1-adrA-GFPlva transcriptional fusion	This study
pMR1-PacrA	CmR, orip15a; pMR1-acrA-GFPlva transcriptional fusion	This study
pMR1-PcpxR	CmR, orip15a; pMR1-cpxR-GFPlva transcriptional fusion	This study
pMR1-PmatA	CmR, orip15a; pMR1-matA-GFPlva transcriptional fusion	This study
pMR1-PompR	CmR, orip15a; pMR1-ompR-GFPlva transcriptional fusion	This study
pMR1-PfliA/ IHFmut	CmR, orip15a; pMR1-fliA/IHFmut-GFPlva transcriptional fusion	This study

plasmids	description	reference
pMR1-PyeaJ/ Fismut	CmR, orip15a; pMR1-yeaJ/Fismut-GFPlva transcriptional fusion	This study
pMR1-PrpoS-5UTR	CmR, orip15a; pMR1-rpoS-5UTR-GFPlva transcriptional fusion	This study
pMR1-PrpoE-5UTR	CmR, orip15a; pMR1-rpoE-5UTR-GFPlva translational fusion	This study
pMR1-PcsgD-5UTR	CmR, orip15a; pMR1-csgD-5UTR-GFPlva translational fusion	This study
pMR1-PflhD-5UTR	CmR, orip15a; pMR1-flhD-5UTR-GFPlva translational fusion	This study
pMR1-PfliA-5UTR	CmR, orip15a; pMR1-fliA-5UTR-GFPlva translational fusion	This study
pMR1-PyeaJ-5UTR	CmR, orip15a; pMR1-yeaJ-5UTR-GFPlva translational fusion	This study
pMR1-PadrA-5UTR	CmR, orip15a; pMR1-adrA-5UTR-GFPlva translational fusion	This study
pMR1-PacrA-5UTR	CmR, orip15a; pMR1-acrA-5UTR-GFPlva translational fusion	This study
pMR1-PcpxR-5UTR	CmR, orip15a; pMR1-cpxR-5UTR-GFPlva translational fusion	This study
pMR1-PmatA-5UTR	CmR, orip15a; pMR1-matA-5UTR-GFPlva translational fusion	This study
pMR1-PompR-5UTR	CmR, orip15a; pMR1-ompR-5UTR-GFPlva translational fusion	This study

Table 3. Primers Used In This Study

primers	description	reference
5 rpoS EcoRI	5’-GCGCGAATTCGATTACTGGCAACGATTTCCG-3’	This study
3 rpoS BamHI	5’-GCGCGGATCCACTCATAAGGTGGCTCCTAC-3’	This study
3 rpoS SphI	5’-GCGCGCATGCCCTGCGATAACAGTTCCTCT-3’	This study
5 rpoE EcoRI	5’-GCGCGAATTCGTGGAGGATTAGGTGGTGAA-3’	This study
3 rpoE BamHI	5’-GCGCGGATCCTCGAAACGCCACTCCATTAG-3’	This study
3 rpoE SphI	5’-GCGCGCATGCGCTGATAGCGCACTACCA-3’	This study
4 rpoE/ORF SphI	5’-GCGCGCATGCTTGCCATGCGTAAATCTTCC-3’	This study

primers	description	reference
5 csgD EcoRI	5’-GCGCGAATTCCTGGACCTGGTCGTACATT-3’	This study
3 csgD BamHI	5’-GCGCGGATCCTTATTGATCGCACACCTGACAG-3’	This study
3 csgD SphI	5’-GCGCGCATGCTATCCGCTTCCATCATATCCAG-3’	This study
5 flhD EcoRI	5’-GCGCGAATTCGAATTGAGCGGCATAACCTG-3’	This study
3 flhD BamHI	5’-GCGCGGATCCCTGCAATAAGCAGAACCACC-3’	This study
3 flhD SphI	5’-GCGCGCATGCACGCTTTGTCCTGAACAATC-3’	This study
5 fliA EcoRI	5’-GCGCGAATTCGCAGGGTTAATCGTTGTAACC-3’	This study
3 fliA BamHI	5’-GCGCGGATCCGATAAACAGCCCTGCGTTATATG-3’	This study
3 fliA SphI	5’-GCGCGCATGCGCAGATCGTCAAGTTCCAC-3’	This study
5 yeaJ EcoRI	5’-GCGCGAATTCGATAAAGCCTTACGCGAAGC-3’	This study
3 yeaJ BamHI	5’-GCGCGGATCCGATAAATGGCACTCCTTCGG-3’	This study
3 yeaJ SphI	5’-GCGCGCATGCCGATATAGCGCAGATAGGCA-3’	This study
5 adrA EcoRI	5’-GCGCGAATTCCGAAAAAAGTTTGACGCCAC-3’	This study
3 adrA BamHI	5’-GCGCGGATCCCAATTTTCCCAAATTATAGAGACGG-3’	This study
3 adrA SphI	5’-GCGCGCATGCGCACGTTTACGCCCATTAC-3’	This study
5 cpxR EcoRI	5’-GCGCGAATTCGCGTGGCTTAATGAACTGAC-3’	This study
3 cpxR BamHI	5’-GCGCGGATCCTGTTTAAATACCTCCGAGGCA-3’	This study
3 cpxR SphI	5’-GCGCGCATGCATGAAGCAGAAACCATCAGATAG-3’	This study
5 matA EcoRI	5’-GCGCGAATTCTTTTCACTCAAACTGTTAAGATG-3’	This study
3 matA BamHI	5’-GCGCGGATCCCCGGAAGTAAATAAGATACGT-3’	This study
3 matA SphI	5’-GCGCGCATGCACCAATAATTTGCTAAGGCC-3’	This study
5 ompR EcoRI	5’-GCGCGAATTCCTCGTTGATTCCCTTTGTCT-3’	This study
3 ompR BamHI	5’-GCGCGGATCCGCAACAATTTGTAAGCGTGT-3’	This study
3 ompR SphI	5’-GCGCGCATGCCACCAGGTAACATTAAATCCAG-3’	This study
5 acrA EcoRI	5’-GCGCGAATTCTGCCATATGTTCGTGAATTTACAG-3’	This study
3 acrA BamHI	5’-GCGCGGATCCTCATATGTAAACCTCGAGTGTCC-3’	This study
3 acrA SphI	5’-GCGCGCATGCCAGGATCAATCTGATAGAGAGAGAC-3’	This study
5 fliA - IHFmut overlap	5’-CCCGGGTTGCACATTCCCGGGGGCCGGATAAGGCGT-3’	This study

primers	description	reference
3 fliA - IHFmut overlap	5’-CCCCGGGAATGTGCAACCCGGGTAAATTGCAATTCAACTTGTAGGC-3’	This study
5 yeaJ -Fismut EcoRI	5’-GCGCGAATTCGAAGCGAAAAGCGAGGG-3’	This study

IV. RESULTS

Chapter I

Analysis of the regulatory network controlling planktonic to biofilm transition in *E. coli*

1. Basis

Bacteria shift its free life style to a community structure because of the beneficial characteristics that biofilm structure provides. Differentiation from planktonic to biofilm formation is characterized by specific stages. Each functional differentiation stage is accompanied by changes at the expression of specific genes in order to the properly development of the flagella-biofilm network. The complex transcriptional regulatory network of flagella function and curli production, the principal biofilm structure indicator, has been investigated by different reports **(Figure 9)** (Ogasawara et al., 2011; Sanchez-Torres et al., 2011; Serra et al., 2013).

The process is controlled by the master regulators *rpoS* sigma factor and *flhDC* regulator. These master regulators receive major regulatory inputs from c-di-GMP, cAMP and ppGpp, which are modulated by DGCs and PDEs (Pesavento et al., 2008; Sommerfeldt et al., 2009; Valentini and Filloux, 2016). The *flhDC* is expressed at post exponential phase and controls more than 60 genes involved in flagella synthesis and related functions such as chemotaxis (Chevance and Hughes, 2008). At stationary phase, the general stress response master regulator *rpoS* is expressed controlling over 500 genes in a highly complex network (Serra et al., 2013). Therefore, the interplay between those master regulators and downstream-activated genes modulates the complex transition between planktonic and biofilm stages. Flagella maintenance is due to FlhDC activation of effectors molecules such as FliA (RpoF), leading to the activation of the flagella expression program. Others molecules as *cpxR* and *clpX* play a complex dual role during bacterium development to inhibit or activate both programs. Moreover, at stationary phase, biofilm formation initiates by increasing levels in the second messenger c-di-GMP as well as expression and activation of *rpoS,* leading to the activation of *csgD* (Mika and Hengge, 2014; Ogasawara et al., 2011). While, others DGCs are activated amplifying the response. Additionally, there are alternative mechanisms involving different key genes of the biofilm pathway, which leads to *csgD* activation and properly curli fibers production (Hengge, 2011; Pesavento et al., 2008; Valentini and Filloux, 2016). Undoubtedly, the interplay between those molecules modulate the properly development of the flagella-biofilm program. Yet, the full dynamics of the complex regulatory network controlling planktonic to biofilm transition is not completely understood. In this sense, *s*cale free networks describe biological phenomena by abstraction of the cell's components and compare their interactions through pairwise interactions. In network laws, connectivity describes the behavior among members of genetic networks, metabolic

pathways, protein interaction and any other biological network. This leads to the uncovering of important effectors in the pathway.

Analysis of transcriptomic data shown that global regulators as CRP and IHF initiate the circuits fluxes that activate specialized TF, culminating in fine-tuning of both pathways. Interestingly, a proportion of those TFs, are auto-activated themselves (Martínez-Antonio et al., 2008). This suggest that there is a hierarchical order at flagella-biofilm transcriptional regulatory network. Therefore, our first aim was to understand the effect of the different TF and GR already described in the regulatory transcriptional pathway of flagella and curli fimbriae production using network analysis, which allows us to find important effectors and paths in the network. Specifically, information of activation and repression between 35 nodes that participates in the motility and biofilm program were collected from original scientific publications studying their gene expression levels using several methods. The gathered information was analyzed by generate flagella, biofilm and the integrated network. The centrality measurements obtained allowed us to understand the essential genes and their principal interactions in the network.

2. Planktonic and biofilm regulatory network of E. coli

In order to accurately re-construct the flagella-biofilm regulatory network, the gathered information from the scientific articles was contrasted with RegulonDB (Huerta et al., 1998) and KEGG DB (Wixon and Kell, 2000), which provides information of the transcriptional regulation of *E. coli*. Interestingly, some contrasting features were found. First, from collected data from scientific literature, some TF were reported with dual effect (activation and repression) for the same target gene. For instance, *rpoS* expression was reported to be active but also inactive by the CRP global regulator (Hengge, 2009; Jofré et al., 2014; Mika and Hengge, 2005). Second, most of genes that were reported as being regulated by different TF, in the web-databases the same gene did not presented *cis-element* to any TF **(Figure 10)**. To overcome those constrains we consider for the network that any transcription factor, master regulator, global regulator, small molecule and RNAs could be a node. This last, despite gathered, were not included because RNA-node analysis of regulatory gene expression increases the complexity of the network (Mika and Hengge, 2013, 2014) and because of the complex requirements of molecular tools to evaluate this topic. For edges

direction analysis, we consider the activation or inhibition most reported or both when divergent reports were found. The data were plotted in a power-law graph using an organic algorithm to proper cluster visualization generating the transcriptional regulatory network of both planktonic and biofilm structure. Next, we performed degree analysis in order to present a general view of the most connected nodes, following by edge and node betweenness analysis in order to disclose the main effectors and the logic pathway that could describe the network. Finally, we also performed the out degree analysis at integrated network to identify the potential talkative nodes **(Figure 11)**.

The topology of flagella network shows that *flhD* is the most connected node followed by *rpoS,* while all other nodes are connected in a minor degree **(Figure 11A)**. Moreover, *rpoS* is the main effector in the network and it is modulated by different inputs. From those, c-di-GMP presents high betweenness and edge connection values **(Figure 11B),** which means that this small molecule exerts important functions to induce biofilm formation (Valentini and Filloux, 2016). The flagella network is maintained by the highly connected node *flhD* **(Figure 11A)**. From its different connections, the outstanding connection is with the sigma factor *fliA* (RpoF) which activates different genes of the flagella programs **(Figure 11B)**. Additionally, the node and edge betweenness analysis presented three paths that are important to maintain the balance in the network. The first one is the connection between *rpoS* and *rpoE,* which activates both *fliA* and *matA* for motility and common pilus to adherence. Second, *rpoS* with the TF and kinase protein *cpxR*, which not only modulates *rpoE*, but has also an important connection with the TF *nsrR,* which repress *fliA* expression. Third, the activation of *clpX* ATPase by *rpoS* exerts *flhD* inhibition at posttranscriptional and/or posttranslational levels (Tomoyasu et al., 2002). Altogether, this analysis shows that connections between positive and negative modulators are important to maintain the functionality of flagella network. That is, the connections between *rpoS-rpoE-cpxR-fliA-matA* and *csgD* in a minor degree seems to play a major role in processing of information for the operation of the flagella regulatory network.

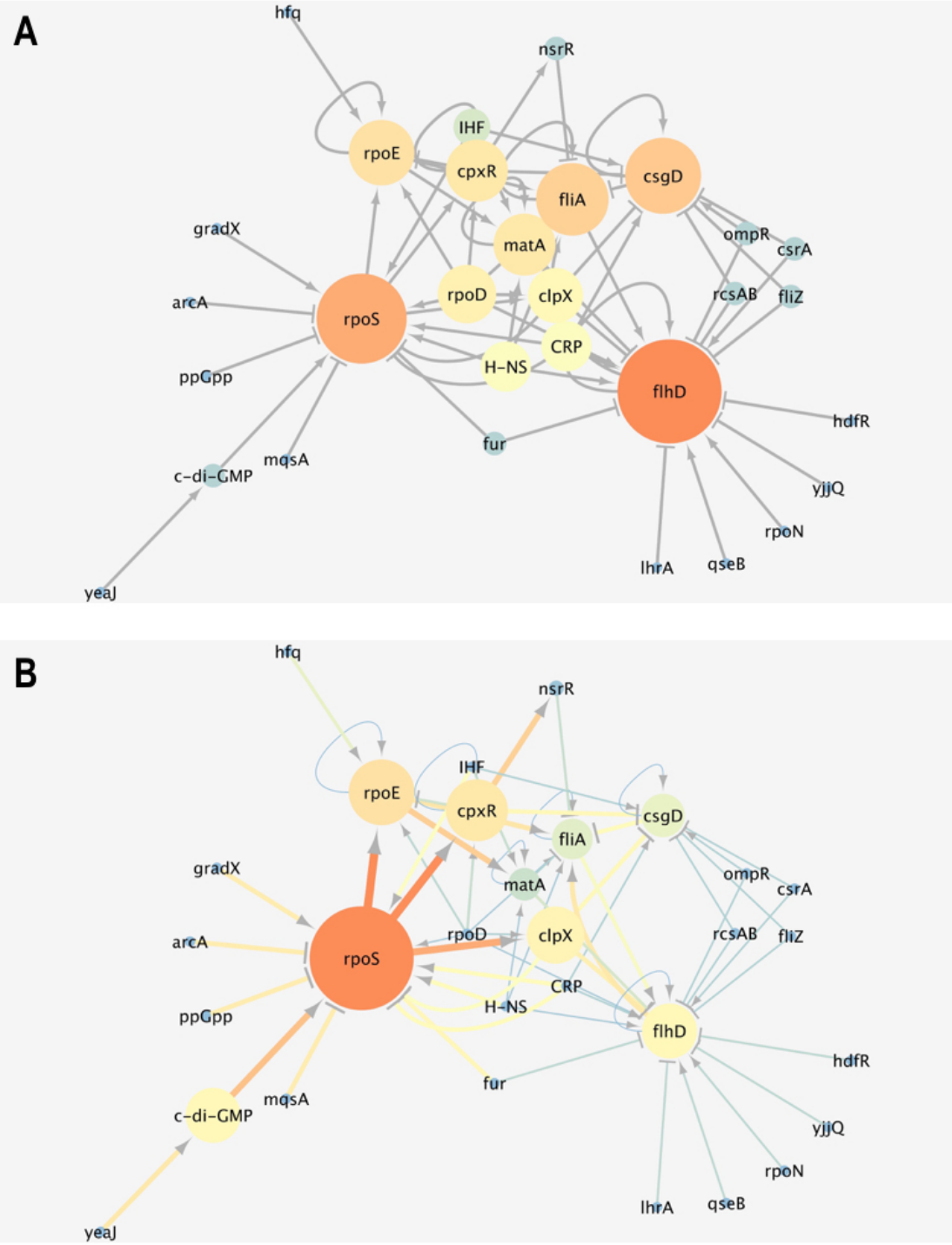


Figure 11: Flagella transcriptional regulatory network. Biochemical and molecular data of protein and gene expression of the different molecules that modulates the flagella-biofilm program were collected from original research articles and the data were transformed into nodes and edges. Those data were loaded into Cityscape 3.4 and an organic algorithm was used to the properly cluster visualization. The network was analyzed by **(A)** degree and **(B)** betweenness with edge betweenness centrality measurements.

In the case of the biofilm network, the analysis shows that *rpoS* and *csgD* are the highly connected nodes **(Figure 12A)**. Entrance to stationary phase leads to increase of c-di-GMP, the principal *rpoS* input activator. Both edge and betweenness analyses support this connection **(Figure 12B)**. Once activated, *rpoS* controls and spreads the information to another nodes becoming the most important effector in the network. *csgD*, another major effector, receives different input signals to triggers curli program. The most outstanding connections are *rpoS - cpxR* and *rpoS - clpX, since* both interact with same edge value to modulate *csgD* effect. This makes sense since *cpxR* activates the phase of scape of biofilm and the *clpX* down-regulates surface proteins to remove cell-cell contacts, impairing biofilm structure. Interestingly, despite node analysis provided the lowest value to *adrA*, the principal cellulose activator; edge analysis shows that *adrA* is an important element to maintain the program activated because receives information of the two major effectors *rpoS* and *csgD*. Therefore, the biofilm network seems to be in fine balance to maintain the properly development of all the phases of biofilm development.

Flagella and biofilm networks present intrinsic properties and topologies that are important for each biological function. In nature, both programs are modulated differentially in function of the different inputs sensed by the colony. Therefore, in order to understand the logic interplay of an integrated topology, we merged both regulatory transcriptional networks. The degree analysis shows that *flhD* is the more connected node of the network, followed by *csgD*, *rpoS* and *fliA*, while all other nodes have less connections **(Figure 13A)**. Additionally, we wondered which nodes can exerts more signals to the entire network. In this sense, the out degree analysis shows that *rpoD* and CRP distributes the higher number of connections to other nodes in the network **(Figure 13B),** as proposed by Martines (Martínez-Antonio et al., 2008).

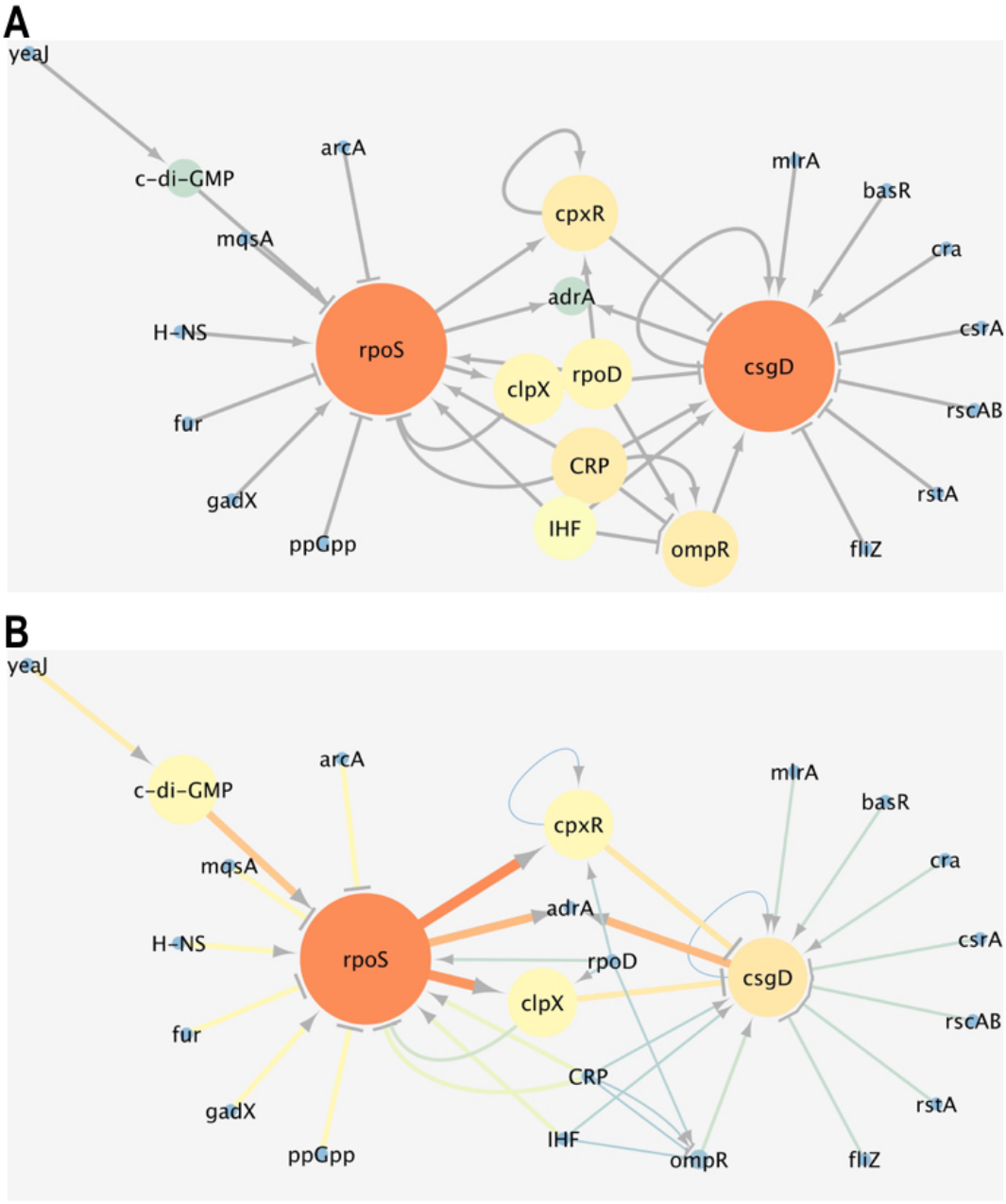


Figure 12: Biofilm transcriptional regulatory network. Protein and gene expression of the different effectors in the flagella-biofilm program were collected from original research articles. This information was transform into nodes and edges. Those data were loaded into Cityscape 3.4 and an organic algorithm was used to the properly cluster visualization. The network was analyzed by (A) degree and (B) betweenness centrality with edge betweenness centrality measurements.

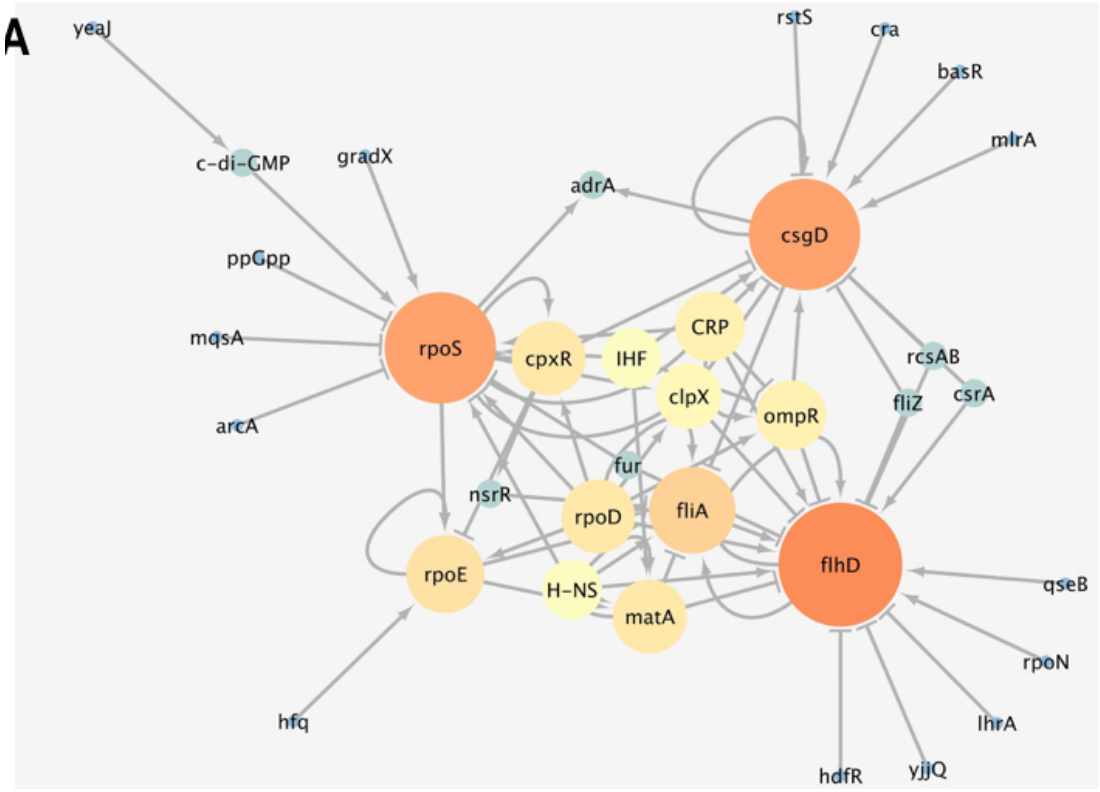


Figure 13: Flagella-biofilm transcriptional regulatory network. The flagella and biofilm network were merged to a fully understanding of the effectors in an integrated view. Information of both networks were loaded into Cityscape 3.4 and an organic algorithm was used to the properly cluster visualization. The complex network was analyzed by use the centrality measurements algorithms of **(A)** degree, **(B)** out-degree with edge betweenness centrality and **(C)** betweenness centrality with edge betweenness centrality measurements.

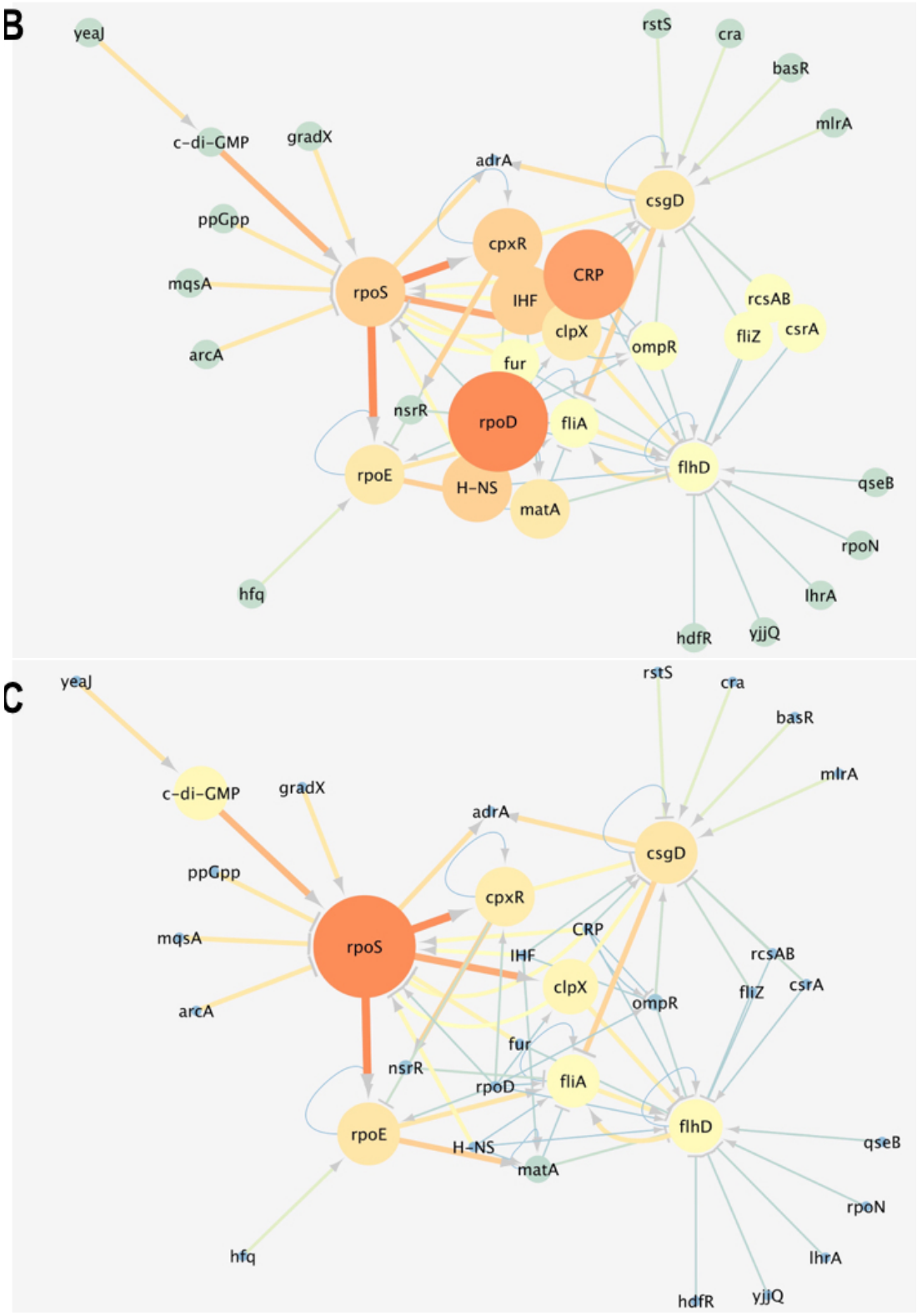


Figure 13: Continue

The topology representing the betweenness and edge betweenness centrality values clearly show an integrated and balanced distribution of the nodes in the network. Thus, *rpoS* is the main effector in the network activated by c-di-GMP. Other six nodes (*csgD, rpoE, cpxR, clpX, flhD* and *fliA*) present a balanced betweenness value, suggesting an orchestrated interplay throughout the growth to maintain the proper function of the network **(Figure 13C)**. In planktonic stage, the *rpoS-flhD-fliA* is the main program activated, together with the dual functional *cpxR* and *clpX*. Subsequently, later during bacteria growth, the different biofilm phases are modulated by a balance between *rpoS* and *csgD* and the later activators *rpoE-cpxR-clpX*.

In summary, the flagella-biofilm transcriptional regulatory network reconstruction consisted of 35 nodes representing the connections between different TF, GR and small molecules. From those, 13 nodes were highly connected (**Figure 13A**) and the betweenness analysis (**Figure 13C)** suggested that the properly gene expression of the network is regulated by 8 principal effectors (*rpoS, csgD, rpoE, c-di-GMP, cpxR, clpX, flhD, fliA*). Moreover, nodes that presented important biological functions as *matA* (related to fimbriae production), *adrA* (activator of cellulose production), *ompR* (flhD TF inhibitor) and *yeaJ* (DGC-related to modulate motility at 37˚C) were also important in our different centrality analysis and some nodes have been less explored experimentally. Interestingly, CRP and IHF presented a good value of out degree measurement (**Figure 13B**), suggesting that global regulators play an important role in the network, although these regulators have been less analyzed experimentally.

Interestingly, when we gathered the information, we found different scientific reports, mainly of *Enteroaggregative E. coli* (EAEC), which relates Fis regulator to the modulation of biofilm formation. Interestingly, any of those reports present evidence that Fis is regulating the genes chosen for our study (Gualdi et al., 2008; Jakovleva et al., 2012; Moor et al., 2014; Prigent-Combaret et al., 2012; Saldaña et al., 2009; Sheikh et al., 2001; Vijay et al., 2015). Therefore, Fis node was not considered in the reconstructed network. Clearly, the balanced betweenness values of the main nodes suggest that this network could present tolerance to nodes remotion. In order to disclose more connections in the network to understand the role of global regulators we perform in-vivo promoter activity assay. This approach will show a better understanding of the effect of GRs in the flagellala-biofilm transcriptional regulatory network.

Chapter II

Regulatory transcriptional effect of CRP, IHF and Fis in key nodes of the flagella-biofilm network

1. Basis

Centrality measurements applied to our reconstructed network allowed us to identify important effectors as well as cross-talk paths between flagella-biofilm program development (Chapter I). Additionally, comparison of connections also allowed us to infer the effect of the GR in the network. Clearly, network understanding is based in connectivity and two important laws support this fact. The Preferential attachment states that the probability of an old node to receive a new link from a new node is proportional to the number of pre-existent connections in this node, in other words, “The rich get richer”. Also, a network with highly connected nodes can support removal of less connected nodes. This leads to the small world effect, which means that no matter how large the network could be, any two nodes are connected via a relatively small chain of links. Therefore, disclosing the largest number of connections in a network becomes essential to obtain an accurately description of any biological phenomenon studied (To extensive go to Introduction 2).

A “typical” bacterial genome has around 5 million bp encoding about 5,000 proteins (Land et al., 2015). Nine proteins known as global regulators modulate the proper transcription of the bacterial genome during development trough specific transcriptional mechanisms related to recognition of *cis-elements* in the promoter region of cognate genes. Global regulators are expressed differentially during growth or under specific input signals. Specifically, CRP is related to the control of metabolic processes in bacteria and it is differentially active under different substrates, while IHF (growth rate dependent) and Fis (expressed at early exponential phase) are dual functional proteins, working both as activators and repressors (Dillon and Dorman, 2010; McLeod and Johnson, 2001). During the transition from motile cells to biofilm structure, a mixture of different developmental bacterial stages occurs. Therefore, small molecules and global regulators direct the activation of signaling proteins that are sensed by neighborhood cells in order to construct its community complex structure. Altogether, the study of connections of global regulators and the dynamics of key genes of flagella-biofilm network will provide a better understanding of this developmental-like program.

In recent years, cloning of promoter sequence upstream of a gene reporter is a well-accepted methodology to understand gene expression dynamics of a specific gene as well as protein interactions at the promoter sequence. Amores et al., in 2015, used bioinformatics algorithms to engineering a synthetic promoter sequence capable to be recognized by CRP, IHF and Fis global

regulators, and its dynamics of protein interactions were inferred by promoter reporter assay (Attached article (Amores et al., 2015) (More examples at introduction chapter: promoters to understand transcriptional regulatory networks). Therefore, from the network analysis, we chose ten principal nodes in order to investigate the effect of the global regulators CRP, IHF and Fis during the planktonic to biofilm transition. While CRP and IHF have been implicated in this process, the role of Fis has been suggested but not demonstrated.

2. Workflow to determine the activity and dynamics of the main network effectors

Individual and merged networks were analyzed and the nodes were chosen in function of the different analysis performed (as described in the previous chapter); *rpoS* > *rpoE/csgD* > *cpxR/clpX* > *flhD/fliA/matA* > *ompR/adrA/yeaJ*. Nodes *clpX* and *cpxR* belong to the same group with same biological effect, thus, we decide replace *clpX* by *yeaJ.* This because YeaJ is a GDEF protein in the involvement of c-di-GMP production (Hengge, 2016; Pesavento et al., 2008). Upon selection, we analyzed the architecture of the promoter region of the chosen genes. For this, data from regulonDB, EcoCycDB and KEGG DB were collected and the *cis-element* information of the promoter regions of interest were compared with the interactions represented in our network. We expected that, the interactions between nodes in our network should coincide with all the *cis-element* of the promoter regions studied. In fact, all the promoter regions presented different topologies of *cis-elements* recognized by different TF and GR. Despite the majority of connections from EcoCycDB and KEGGDB website confirm the gathered interactions as displayed each promoter at website, none of those sites suggest the *cis-element* at promoter regions. Thus, the confirmation of predicted *cis-elements* from transcriptomic data for each promoter from RegulonDB show just a few connections matching with our gathered data **(Figure 10)**. Altogether, the connections of the flagella-biofilm transcriptional regulatory network of our gathered data correlates with information at EcoCyc and KEGG, although the promoter sequence analysis show a few connections.

To address the role of the global regulators at promoter regions, we amplified the promoter region sequence (transcriptional fusion) and the promoter region plus 5'UTR (translational fusion) of each gene and were cloned upstream of a GFPlva reporter system (Guazzaroni and Silva-Rocha, 2014). As a positive and negative control, we used BioBrick constitutive promoter PJ100 and pMR1 empty

vector, respectively. The twenty-two constructions generated were used to transform *E. coli* BW25113 wild-type (wt) and the dynamics of gene expression was evaluated by reporter expression assay. Therefore, when the promoter presented activity, we considered this as the intrinsic capability of the promoter to respond in the presence of the three global regulators and its dynamics was measured trough the growth (blue line in all graphs of this chapter). None translational fusions showed activity, we strongly believe because of the design of our fusions. For instance, for *rpoE* translational fusion we considered another promoter region who is suggested at 200bp inside its ORF (Figure 10). While, cloning of 10 codons of any ORF has been proved to be enough to induce the *lacZ*-reporter systems. (Sommerfeldt et al., 2009)

Next, *E. coli* strains genetically depleted for IHF (Δ*ihf*) and Fis (Δ*fis*) strains used to uncover their individual effect over the generated constructions (in all graphs Δ*ihf* strains are shown in red lines while Δ*fis* strains are represented by green lines). Finally, we took advantage of the CCR mechanism to down modulate CRP and to evaluate its effect over the studied promoters. For this, *E. coli* BW25113 wild-type (wt) was exposed to glucose, generating a conditional Δ*crp* strain. Interestingly, this condition can be applied to all the mutant strains, therefore is possible to generate a double mutant-like phenotypes. Promoter activity was followed during eight hours in static conditions, which is the main concept for liquid biofilm formation system. Finally, the experimental data were transformed into connections between the global regulators and the promoter sequences studied and the data were loaded into integrated flagella-biofilm transcriptional network in order to get a better understanding of the role of CRP, IHF and Fis in the network. The flowchart of the experimental design is presented in **Figure14**.

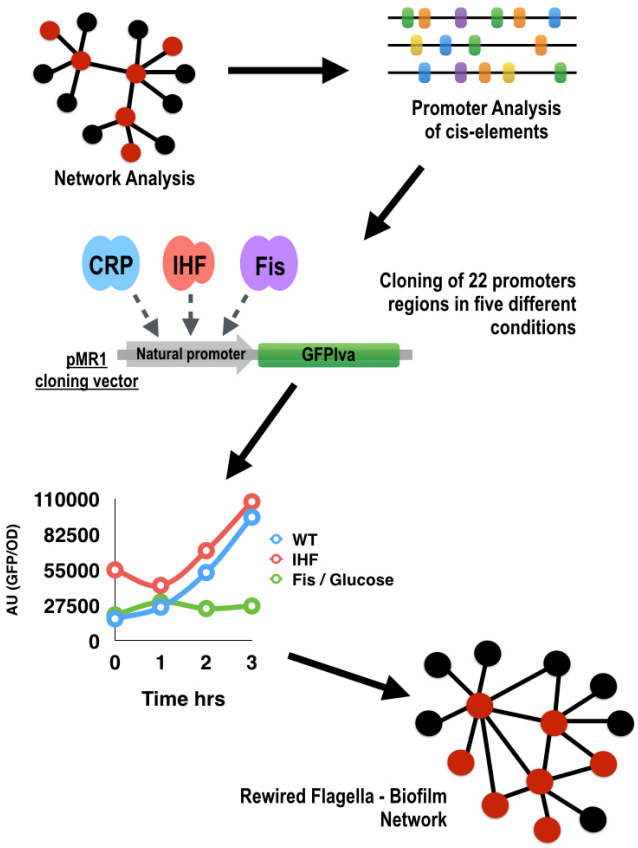


Figure 14: Design of the promoter node reporter assay. Data of the flagella-biofilm program were transformed into nodes and edges and analyzed by Cityscape software. The architecture of the promoter regions of the principal nodes effectors of the network was analyzed to confirm the interactions reported. Subsequently, cloning of the natural promoter regions in a reporter system was developed. Promoter activity was determined for each promoter in different conditions as established in material and methods. Finally, GFP activity was transform into connectivity data and loaded into flagella-biofilm network to gain a better understanding of this program.

3. *CRP, IHF and Fis modulate the major effector of the flagella-biofilm network - rpoS*

rpoS has been proposed as the main effector of flagella-biofilm network (Pesavento et al., 2008; Serra et al., 2013) and we also shows that in our network analysis, we were interested in understanding the effect of the three global regulators on the promoter region of the *rpoS* gene. For this, overnight cultures of *E. coli* BW25113 (wild-type) carrying the transcriptional *PrpoS*-GFPlva fusion were diluted 1:100 in fresh M9 minimal medium supplemented with glycerol as sole carbon source, and GFP fluorescence was measured at 20 minutes time intervals at 37˚C during 8 hours. As

can be seen in blue line F**igure 15A** left graph, the analyzed promoter presented a growth-phase-dependent activity with a slightly increase at the end of the growth, ranging from ~4e^4 AU at lag phase to ~5e^4 AU at stationary phase. Therefore, in the presence of CRP, IHF and Fis, *rpoS* promoter sequence is constitutively active. Also, it has its higher activity at stationary phase, as has been previously described (Hengge-Aronis, 2002; Klauck et al., 2007; Liu et al., 2015).

Next, to uncover the role of IHF and Fis global regulators on the promoter region of *rpoS, o*vernight culture of *E. coli* mutant strains for those global regulators harboring the pMR1-*rpoS* system were evaluated by GFP promoter activity assay. The lag phase of the static culture of *E. coli* Δ*ihf* strain with *PrpoS*-GFPlva reporter system starts with similar activity than *wt* strain ~4.5e^4 AU. Interestingly, the GFP activity quickly increases and sustains at ~8e^4 AU during the growth (red line vs. blue line **Figure15A** left panel). The P*rpoS*-GFPlva into *E. coli* Δ*fis* strain culture present the higher levels of GFP activity in a growth rate dependent manner from ~5e^4 to 2e^5 AU (green line **Figure15A** left panel). This data strongly suggest that IHF and Fis global regulators are playing a repressive effect over the *rpoS* promoter since, in their absence the GFP activity increases over those of the wild type strain (**Figure15A** left panel).

To understand the role of CRP over the *rpoS* promoter region, we employed the glucose inducible CCR system to down-modulate the activity of CRP GR. For this, overnight cultures of *E. coli* BW25113 *wt,* Δ*ihf* and Δ*fis* carrying *PrpoS*-GFPlva system were diluted 1:100 in M9 medium supplemented with glycerol and glucose, and the experiments were performed as before. The *E. coli* wild type strain in the presence of glucose (blue line right panel) show lower levels of AU activity than when its absence (left panel). Nonetheless, this activity is higher than pMR1 empty vector (**Figure 15C**) suggesting that, while CRP participates as an activator, there are other TFs that maintain the *PrpoS*-GFPlva activity. The genetically depleted strains Δ*ihf* and Δ*fis* induced at CCR condition (red and green lines, right graph) presented similar AU activity during the first 100 minutes of growth that the strains without glucose (red and green lines, left panel). Interestingly, strong decrease of GFP activity after 100 minutes is observed in the presence of glucose (right panel). Those effects are, in fact, by CRP at the promoter sequence, since PJ100 constitutive promoter was only slight affected over all experimental conditions (**Figure 15B**). Altogether, these data suggest that CRP is an activator of *rpoS* while IHF and Fis act as repressors of *rpoS* gene expression.

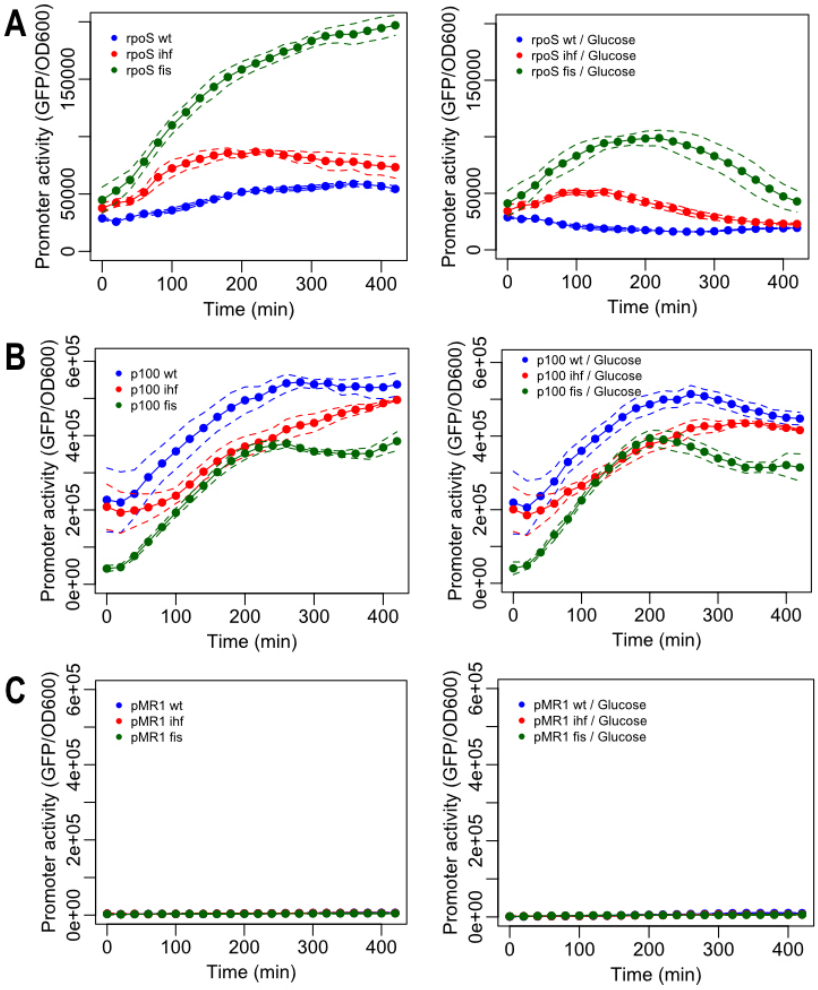


Figure 15: Promoter activity of *rpoS* by CRP, IHF and Fis global regulators. Each *E. coli* strain was grown overnight in M9 media supplemented with 0.2% CAA and 1% glycerol. After pregrowth, cells were diluted in fresh M9 CAA/glycerol media either without (strain name) or with (promotername-strain / glucose) 0.4% glucose. GFP fluorescence was measured each 20 minutes at 37˚C during 8 hours in static conditions. **(A)** *E. coli* BW25113 (pMR1-*PrpoS*). **(B)** *E. coli* BW25113 (pMR1-p100). (**C)** *E. coli* BW25113 (pMR1-empty vector). Vertical bars represent the standard error of three independent experiments.

4. *Global regulators CRP, IHF and Fis differentially modulates rpoE and csgD network effectors*

The RpoE sigma factor is essential for growth and cell wall maintenance in *E. coli* at all temperatures (Noor, 2015). In mature biofilm, is *rpoE* is expressed and controls the expression of different flagella and biofilm genes (Beloin et al., 2008; Du et al., 2011b; Gibson et al., 2010). The network betweenness analysis shows that *rpoE* is an important effector with significant connections between the main effector *rpoS* and flagella genes (as represented in Chapter I). In this sense, we surveyed the effect of the GRs CRP, IHF and Fis on the promoter region of RpoE sigma factor. For this, overnight cultures of *E. coli* BW25113 wild type, Δ*ihf* and Δ*fis* transformed with the *PrpoE*-GFPlva system were diluted 1:100 in fresh M9 medium supplemented with glycerol and glucose as indicated and GFP activity was measured each 20 minutes at 37˚C by 8 hours. As can be seen at blue line in **figure 16A,** left panel, the *rpoE* promoter in the wt strain display a constitutive activity increasing from $5e^4$ to $1e^5$. Interestingly, in the absence of IHF or Fis protein (red and green lines **Figure 16A,** left panel) the promoter activity increases in the first 100 minutes of culture to ~$2e^5$ and is sustained in Δ*ihf* strain and down-regulated in Δ*fis* strain at the end of the growth curve. This suggests that both IHF and Fis act as repressors of the *rpoE* promoter. Additionally, in the presence of glucose (**Figure 16A,** right panel**)** the *rpoE* promoter activity was maintained unchanged when compared to wild type strain without glucose. The *E. coli* mutants strains of IHF and Fis in the presence of glucose present the same behavior with increased levels in GFP activity than Δ*ihf* and Δ*fis* strains without of glucose. These data suggest that, IHF and Fis are exerting a repressive effect over the *rpoE* promoter, even though the data obtained for CRP is not conclusive.

CsgD is the central protein related in biofilm formation, since *csgDEFG* and *csgBA* operon transcription leads to curli fimbriae production, which is the main indicator of biofilm formation (Chirwa and Herrington, 2003; Mika and Hengge, 2014). Analysis of *csgD* promoter activity in *E. coli* BW25113 *wt* **(**blue line **Figure 16B** left panel**)** shows that in the presence of CRP IHF and Fis the promoter is active with a dynamics of low but significant GFP values at lag phase, subsequently an increase activity after 100 minutes of culture with a maximum peak of activity at 200 min (~$5.5e^4$ AU), posterior sustained activation ($4e^4$AU) during all growth curve is observed. These data show that this promoter has a late respond during growth, in accordance with the participation of CsgD in activation of biofilm formation.

Assessment of *csgD* promoter behavior on Δ*ihf* and Δ*fis* strains **(**red and green lines **Figure 16B** left panel**)** shows a slight decrease in GFP activity when compared with Wt strain. This data suggest that IHF and Fis are acting as activators of *csgD* promoter. CRP effect was evaluated by CCR in all strains, as can be seen in the right panel of **Figure 16B**, all promoters presented slightly decreases in GFP activity. The wild type strain present a sustained activity of ~4e^{4}AU, while mutant strains decreased the signal activation between ~1e^4 to 2e^4 AU. Altogether, these results suggest that there is interplay between GR (CRP, IHF and Fis) to maintain the *csgD* promoter activated during the growth curve.

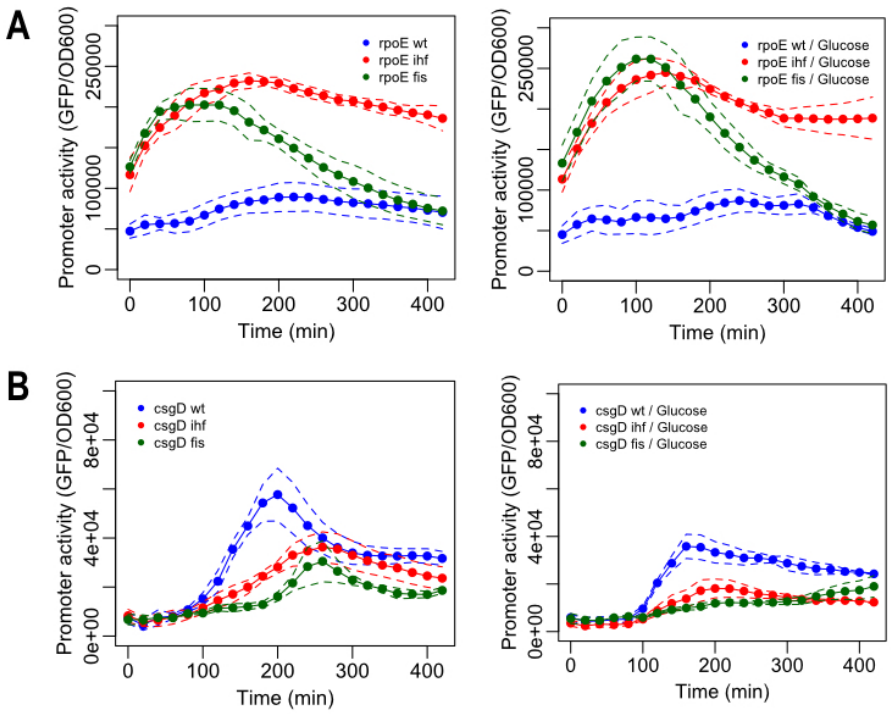


Figure 16: Dynamics of GFP expression of *rpoE* and *csgD* promoters by CRP, IHF and Fis. Each *E. coli* strain was grown overnight in M9 media supplemented with 0.2% CAA and 1% glycerol. After pregrowth, cells were diluted in fresh M9 CAA/glycerol media either without (strain name) or with (promotername-strain / glucose) 0.4% glucose. GFP fluorescence was measured each 20 minutes at 37˚C during 8 hours in static conditions. **(A)** *E. coli* BW25113 (pMR1-*PrpoE*). **(B)** *E. coli* BW25113 (pMR1-*PcsgD*). Vertical bars represent the standard error of three independent experiments.

5. IHF but not CRP neither Fis modulate the dual effector cpxR

As mentioned previously, CpxR perform a boundary role between both networks. Activation of CpxR in *E. coli* has been associated with inhibition of flagella genes as *fliA* and also inhibitor of *csgD* (Dudin et al., 2014; Ogasawara et al., 2010). Following the blue line in the right panel **Figure 17A** we see that the promoter activity of *cpxR* surveyed in *E. coli* BW25113 wild type oscillates between 4,000 AU to ~7,000AU. Slightly decreases can be observed in the absence of Fis global regulator (green line left panel **Figure 17A**). Depletion of IHF protein turns the promoter activity on with an oscillating activity of ~10,000 AU (red line left panel **Figure 17A**). The effect of CRP protein was evaluated by CCR on all the *E. coli* strains **(Figure 17A** right panel**)**. As can be seen in the figures, the *cpxR* promoter activity on the wild type strain plus glucose presented an oscillating behavior between 5,000 to 10,000 AU, which is slightly higher than wild type strain without glucose. Furthermore, promoter activity in *E. coli* Δ*ihf* strain plus mimic Δ*crp* conditions presented a diminished activity fluorescence when compared with Δ*ihf* strain. Evaluation of *cpxR* promoter activity in *E. coli* Δ*fis* strain plus glucose was unaffected when compared with Δ*fis*. These results cannot be conclusive about a specific role of CRP and Fis GR since the GFP values are similar to the empty vector pMR1 (**Figure 17B**). However, this results suggest that IHF global regulator is playing a repressive role at *cpxR* promoter.

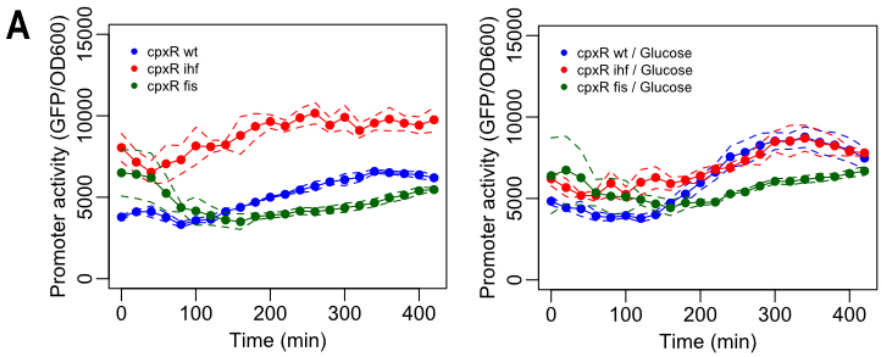


Figure 17: Effect of CRP, IHF and Fis on the promoter activity of *cpxR*. Each *E. coli* strain was grown overnight in M9 media supplemented with 0.2% CAA and 1% glycerol. After pregrowth, cells were diluted in fresh M9 CAA/glycerol media either without (strain name) or with (promotername-strain / glucose) 0.4% glucose. GFP fluorescence was measured each 20 minutes at 37˚C during 8 hours in static conditions. **(A)** *E. coli* BW25113 (pMR1-PcpxR). **(B)** *E. coli* BW25113 (pMR1-empty vector). Vertical bars represent the standard error of three independent experiments.

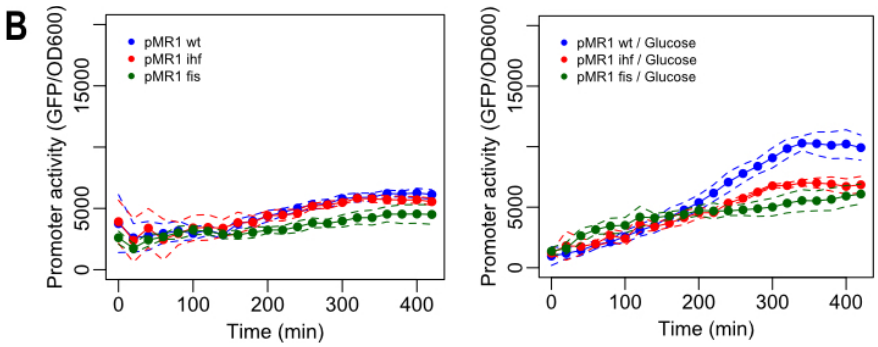


Figure 17: Continue

6. Influence of CRP, IHF and Fis on the promoter activity of flagella nodes — flhD and fliA

The *flhD* gene is the master regulator of flagella functions mainly because activation of the specific flagella sigma factor *fliA* which drives the flagella functional responses (Liu and Matsumura, 1994). In our experimental conditions, *flhD* was the promoter that presented the highest levels of GFP (up to $6e^5$ AU) in the presence of CRP, IHF and Fis (blue line left panel **Figure 18A**). Dynamics of *flhD* promoter activity in the *E. coli* Δ*ihf* and Δ*fis* strains were unchanged (red and green line left panel **Figure 18A**), suggesting that neither IHF nor Fis global regulators are necessary to modulate *flhD* output. Notably, when CCR was induced in all the *E. coli* strains wild type, Δ*ihf* and Δ*fis*, which mimics the conditions Δ*crp* to wt and "double mutant" to the genetically depleted strains, the promoter activity of *flhD* was affected **(Figure 18A** right panel**)**. These results suggest that CRP but not IHF neither Fis participates to modulate *flhD* transcriptional response.

Dynamics of *fliA* promoter present a constitutive fluorescence value of ~10,000 AU in the presence of the three global regulators (blue line left panel **Figure 18B**). Activity of *fliA* in the *E. coli* Δ*fis* strain was unchanged, suggesting that Fis protein does not participate to modulate *fliA* activity (green line left panel **Figure 18B**). Divergently, when *fliA* promoter was evaluated in *E. coli* Δ*ihf* strain (red line left panel **Figure 18B**), the GFP levels reach over 300,000 AU, strongly suggesting that IHF participates as repressor over the *fliA* promoter. The performance of *fliA* promoter activity of *E. coli wild type and* Δ*fis* in CCR induction (blue and green line right panel **Figure 18B**) presents insignificant changes. Besides, the Δ*ihf* plus glucose condition, it present a decrease in GFP output from $3e^5$ to $2e^5$. These results strongly suggest that IHF is a repressor for the *fliA* promoter and no

directly effect can be assigned to CRP and Fis global regulators. However, an additive effect is observed in Δ*ihf* strain in the presence of glucose condition. Altogether, we can conclude that IHF will be controlling the flagella program through the regulation of *fliA* expression.

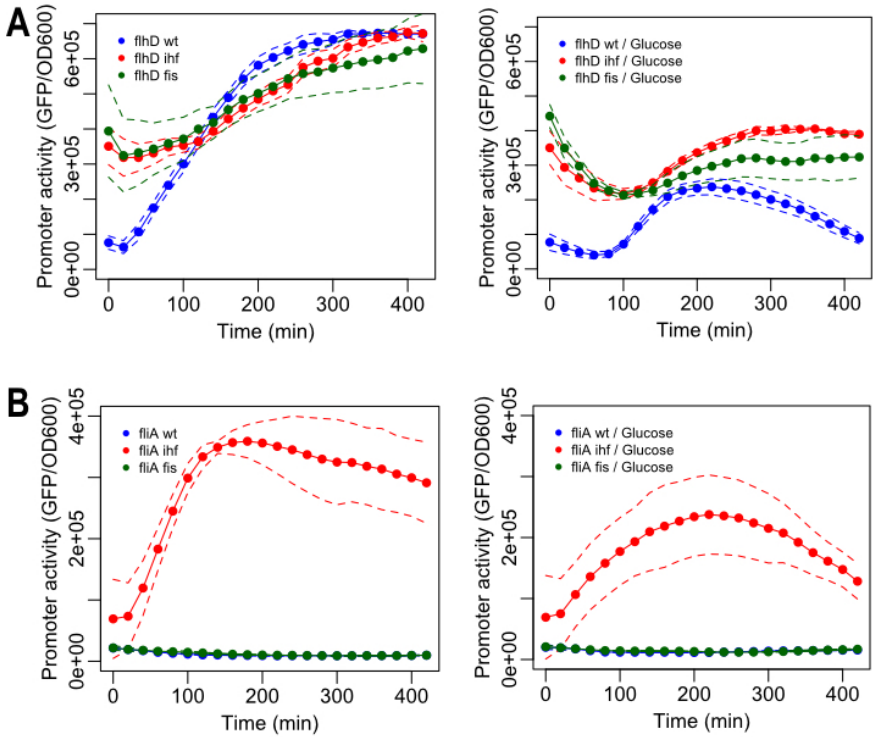


Figure 18: CRP, IHF and Fis effects on the promoter activity of *flhD* and *fliA* nodes. Each *E. coli* strain was grown overnight in M9 media supplemented with 0.2% CAA and 1% glycerol. After pregrowth, cells were diluted in fresh M9 CAA/glycerol media either without (strain name) or with (promotername-strain / glucose) 0.4% glucose. GFP fluorescence was measured each 20 minutes at 37˚C during 8 hours in static conditions. **(A)** *E. coli* BW25113 (pMR1-*PflhD*). **(B)** *E. coli* BW25113 (pMR1-*PfliA*). Vertical bars represent the standard error of three independent experiments.

7. *Impact of CRP, IHF and Fis in the promoter activity of the minor effector nodes — yeaJ, ompR and adrA*

Diguanylate cyclases (DGC) modulate the levels of c-di-GMP. Changes at concentrations of this molecule lead to different transcriptional responses (Hengge, 2016). The knockdown of the GGDEF protein YeaJ shows less motility at 37˚C (Pesavento et al., 2008). The effect of the global regulators CRP, IHF and Fis were surveyed on the promoter region of the DGC *yeaJ*. The dynamics of gene

expression in *E. coli* wild type shows that *yeaJ* is active at lag phase to ~50,000AU increasing its activity throughout the time, with a peak of activity of ~150,000 at 200 min. Subsequently, promoter activity decreases and this is sustained by the rest of the bacteria growth with ~100,000 AU, as can be seen at left panel blue lines **Figure 19A**. The promoter activity of *yeaJ* promoter in the Δ*ihf* strain presents similar behavior than wild type strain, suggesting that IHF is not participating to modulate the promoter region of *yeaJ* (**Figure 19A** left panel red line). Interestingly, the promoter activity in the Δ*fis* strains show a significant reduction. These results clearly show that, Fis global regulator exerts an activation effect on the *yeaJ* promoter region; however, as the GFP activation never drops (sustained activation ~50,000AU), also suggest that other molecules could be regulating the *yeaJ* promoter (**Figure 19A** left panel green line). The equal *E. coli* strains carrying the *yeaJ* promoter in the CCR condition, which mimics a Δ*crp* condition, show similar behavior but with increased peak of fluorescence from ~$1.5E^5$ to ~$2.5E^5$ AU when compares to wild type without glucose **(Figure 19A)**, suggesting that CRP is repressing the *yeaJ* promoter activity. Slightly increases at GFP activity are observed in Δ*fis* strain in the presence of glucose, however not significative. Altogether, we can conclude that CRP represses and Fis activates the *yeaJ* promoter while IHF seems to not participate.

Finally, we evaluated the promoter region of the *OmpR* protein, which inhibits *flhD* leading to biofilm formation (Samanta et al., 2013). We also assessed the promoter region of the *adrA* gene which is related to cellulose production and provides strength to biofilm structure (Römling, 2005). Promoter activity of *ompR* was maintained constantly with a value of 5,000AU in the *E. coli* BW225113 wild type strain **(Figure 19B** left panel blue line**)**. Absence of Fis global regulator does not change the behavior of this promoter during the growth curve **(Figure 19B** left panel green line**)**. An increase of ~3,000 AU was observed on the *ompR* promoter activity when surveyed in *E. coli* Δ*ihf* strain **(Figure 19B** left panel red line**)**. The absence of CRP in all *E. coli* strains did not affect the GFP levels **(Figure 19B** right panel**)**. The same behavior was observed for the *adrA* promoter system (**Figure 19C**). Therefore, despite the changes in the dynamics of the promoter activity of *ompR* and *adrA* systems **(Figure 19B and C)**, they were highly similar to our negative control pMR1 **(Figure 17B)**. Thus, we were unavailable to determine any effect of the global regulators CRP, IHF or Fis on the *ompR* and *acrA* promoter activity.

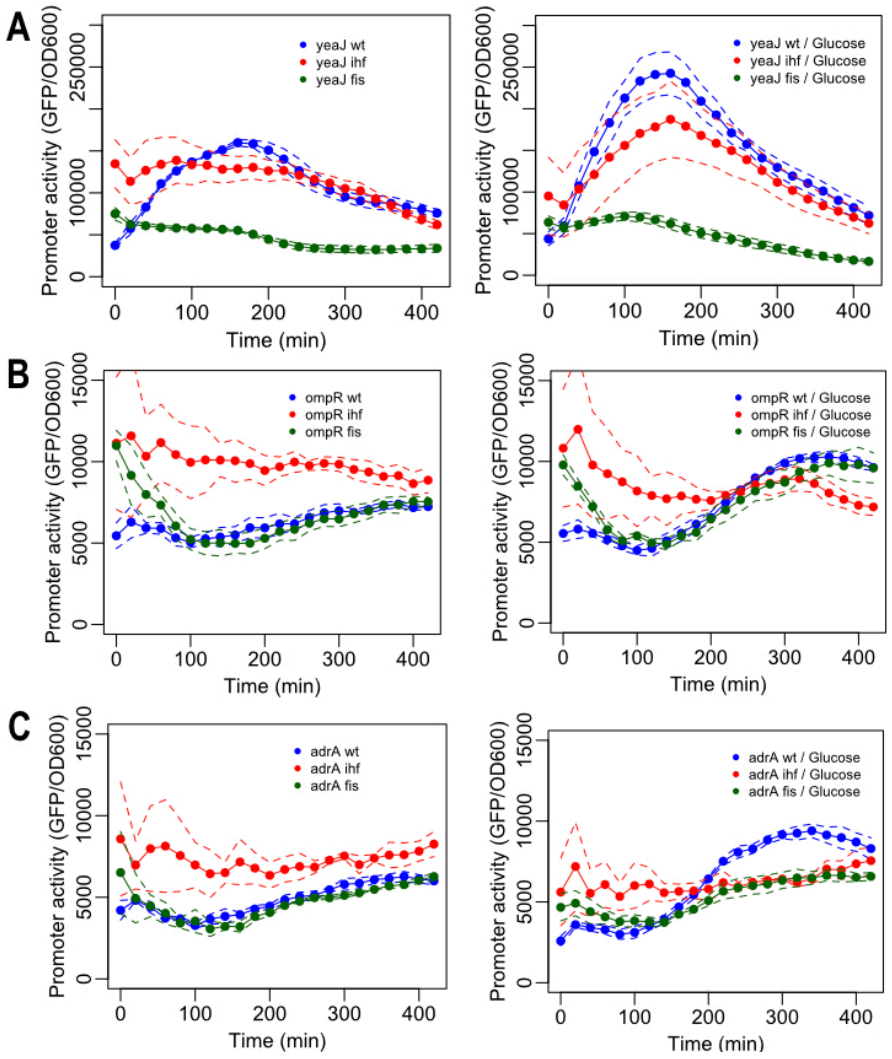


Figure 19: Promoter activity of *yeaJ*, *ompR* and *adrA* by signals of CRP, IHF and Fis. Each *E. coli* strain was grown overnight in M9 media supplemented with 0.2% CAA and 1% glycerol. After pregrowth, cells were diluted in fresh M9 CAA/glycerol media either without (strain name) or with (promotername-strain / glucose) 0.4% glucose. GFP fluorescence was measured each 20 minutes at 37˚C during 8 hours in static conditions. **(A)** *E. coli* BW25113 (pMR1-*PyeaJ*). **(B)** *E. coli* BW25113 (pMR1-*PompR*). (C) *E. coli* BW25113 (pMR1-*PadrA*). Vertical bars represent the standard error of three independent experiments.

8. Integration of experimental data into the flagella-biofilm transcriptional regulatory network

The effect of CRP, IHF and Fis over the promoter regions obtained by our promoter activity assay was transformed into activation or inhibition links. From the 27 interactions tested, we suggested 7 new interactions, while 5 more interactions and one of the two effects of CRP over *rpoS* were confirmed (**Table 4**). Subsequently, those interactions were loaded into the flagella-biofilm transcriptional regulatory network and centrality measurements were performed. In fact, integration of the experimental data slightly changed the topology of the network. The new network consist of 36 nodes; from those, 13 nodes were the most connected **(Figure 20A)**. Besides, the out degree analysis show that RpoD, CRP, IHF and Fis, this last in minor degree, exerts important out-connections to modulate the flagella-biofilm network (**Figure 20B**). This analysis represents the already described effect (**Figure 13B**) of the global regulators CRP and IHF and also shows the new node Fis (**Figure 20B**) as inducers of important signals to other nodes. The topology representing the betweenness and edge betweenness centrality values clearly show an integrated and balanced distribution of the nodes in the network (**Figure 20C**), *rpoS*, the main effector in the network, is activated by c-di-GMP which is modulated by *yeaJ*. Interestingly, we proved that *yeaJ* now is under the control of CRP and Fis. Other six nodes (*csgD, rpoE, cpxR, clpX, flhD and fliA*) present a balanced betweenness value suggesting an orchestrated interplay through the growth to maintain the proper function of the network. Interestingly, and despite edge analysis did not change over the main nodes, the addition of new edges and Fis node leads to a new path with important edge value. Thus, Fis is exerting now important connections with the pivotal effectors *rpoS* and *rpoE* nodes that modulate the entire network.

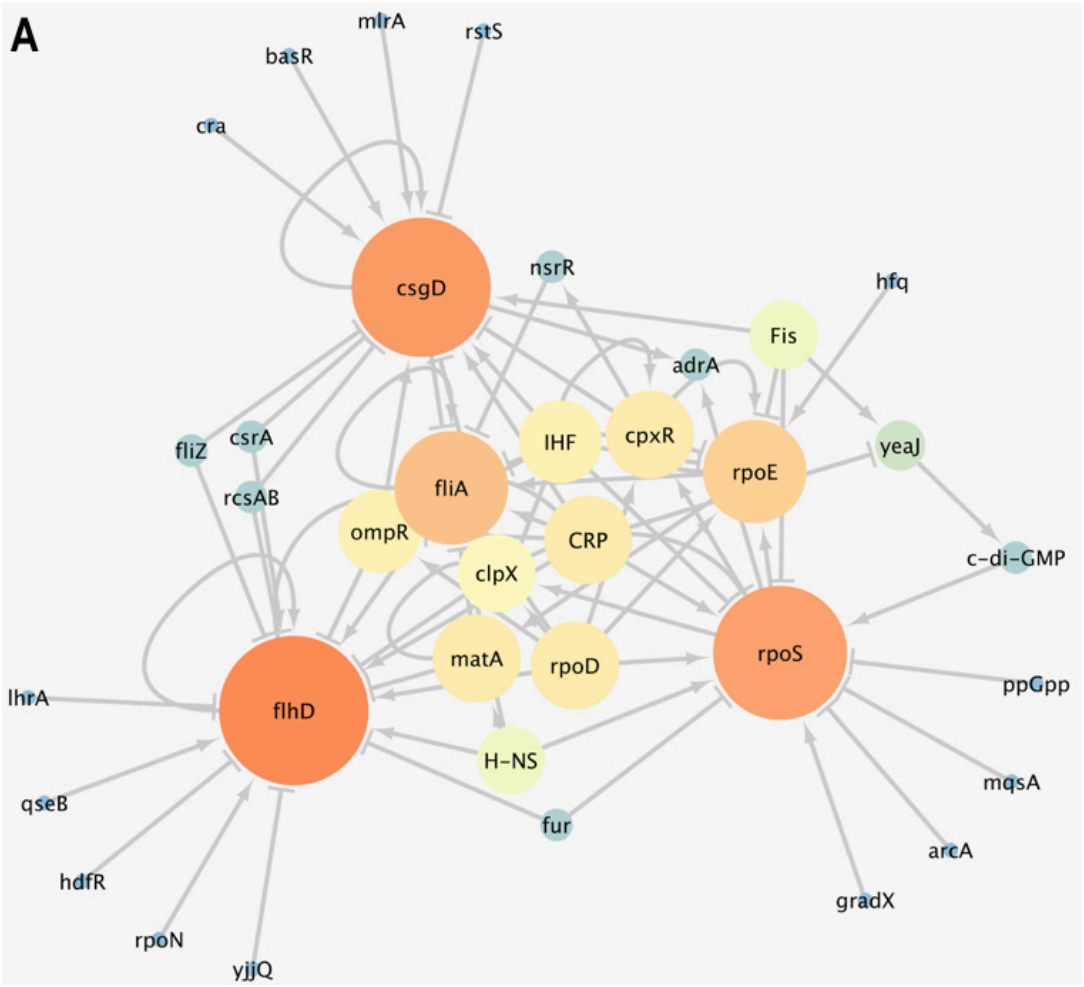


Figure 20: Integration of experimental data into the flagella-biofilm transcriptional regulatory network. The GFP values were transformed into nodes and edges and were incorporated into the flagella and biofilm network. The transformed data were loaded into Cityscape 3.4 and an organic algorithm was used to the properly cluster visualization. The re wired network was analyzed by use the centrality measurements algorithms of (A) Edge acoount, (B) out-degree with edge betweenness centrality and (C) betweenness centrality with edge betweenness centrality measurements.

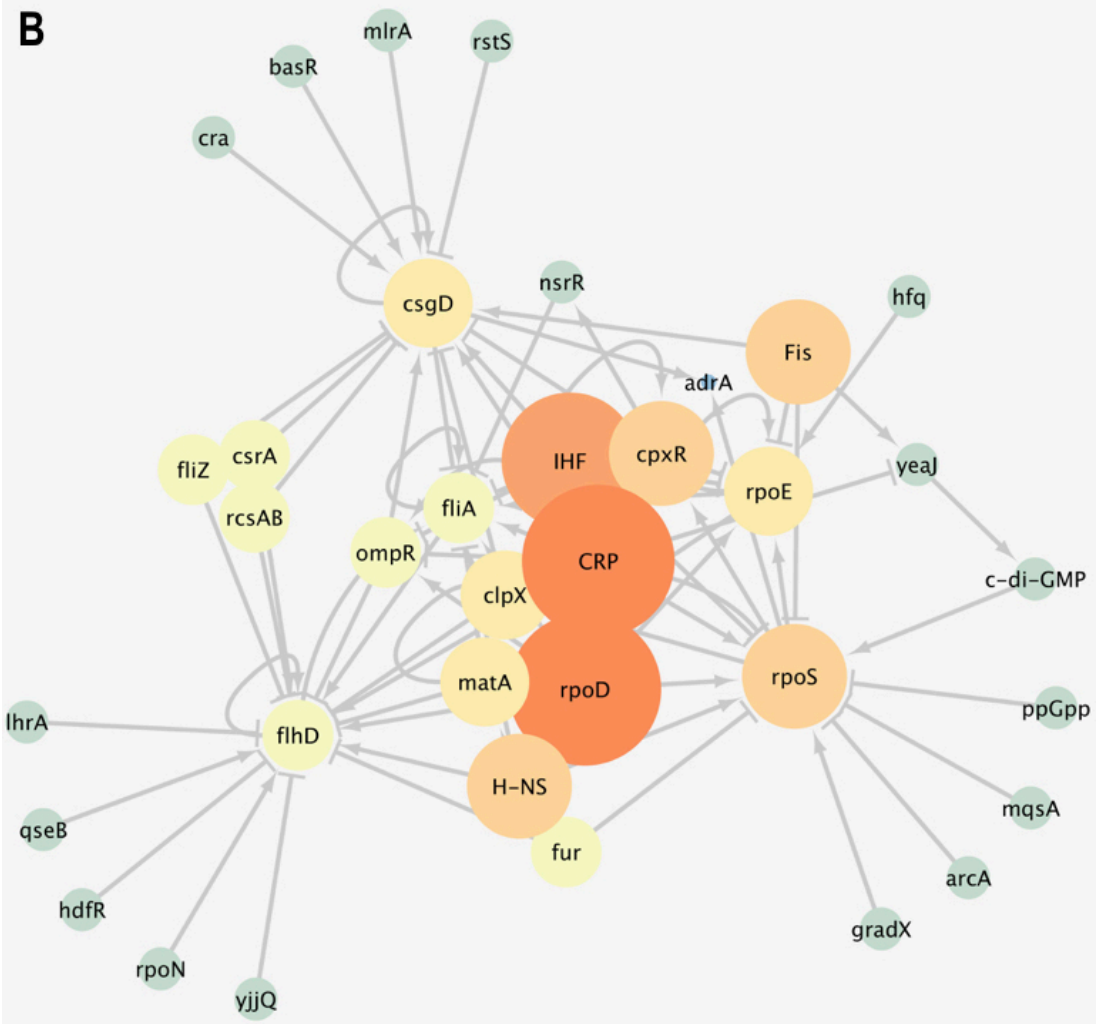


Figure 20: Continue

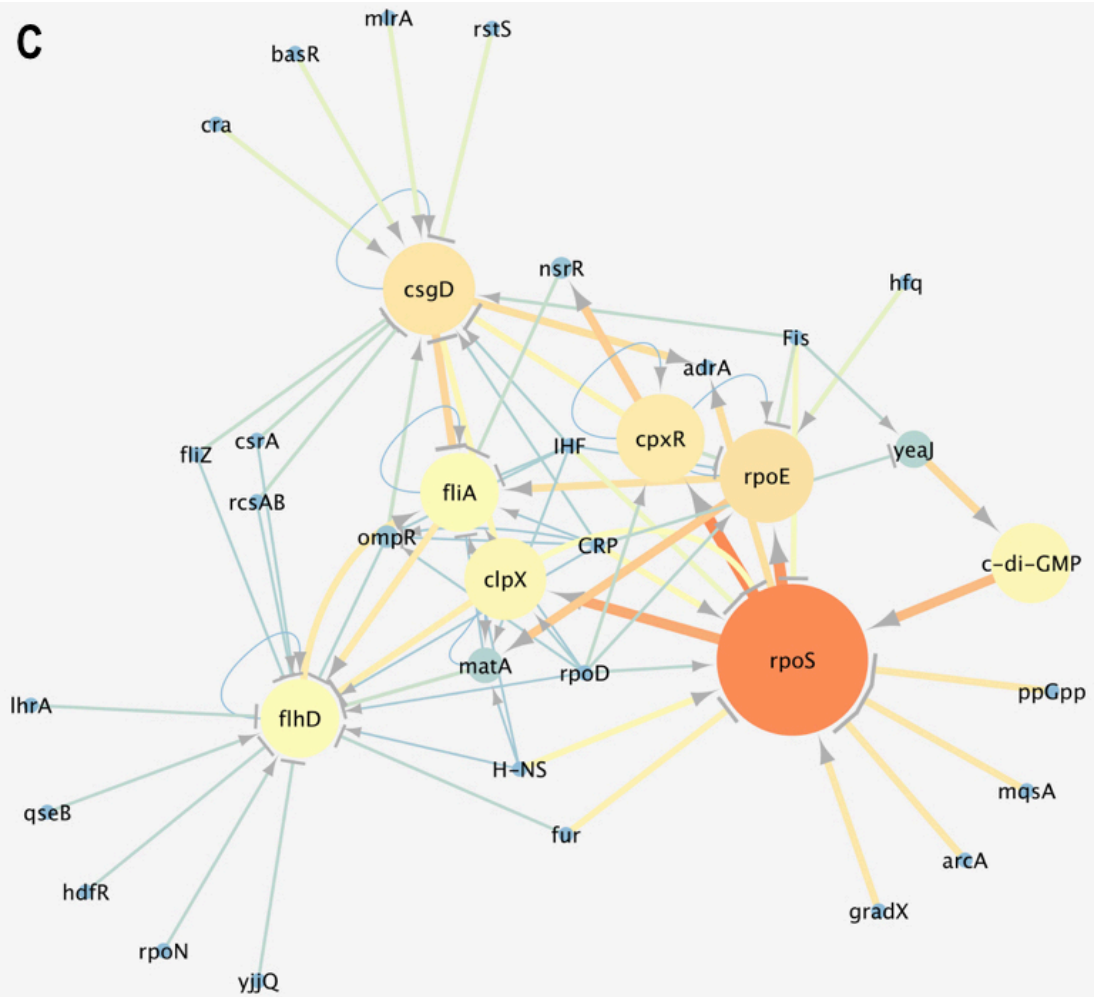


Figure 20: Continue

Altogether, in this chapter we use promoter activity assay to understand how GFP dynamics levels of promoter genes can be used to understand the connections between the GRs CRP, IHF and Fis with key nodes of the flagella-biofilm network. This technique, in addition to uncover the connections in a network, also allowed us to determine the behavior of the promoter trough the growth curve. This could be considered as the gene expression behavior in the presence or absence of any GR through all the physiological differentiation phases.

Table 4: New connections into the flagella-biofilm network

Global Regulator	Connection	Node	Contribution	Bibliography
CRP	activates	*rpoS*	confirmed	Hengge-Aronis, 2002
IHF	inhibits	*rpoS*	confirmed	Mangan, et al., 2006
Fis	inhibits	*rpoS*	New	This work
CRP	Not conclusive	*rpoE*	-	-
IHF	inhibits	*rpoE*	New	This work
Fis	inhibits	*rpoE*	New	This work
CRP	activates	*csgD*	confirmed	Hufnagel, et al., 2016
IHF	activates	*csgD*	confirmed	Gerstel, et al., 2003
Fis	activates	*csgD*	New	This work
CRP	Not conclusive	*cpxR*	-	-
IHF	Not conclusive	*cpxR*	-	-
Fis	Not conclusive	*cpxR*	-	-
CRP	activates	*flhD*	confirmed	Soutourina, et al., 1999
IHF	Not conclusive	*flhD*	-	-
Fis	Not conclusive	*flhD*	-	-
CRP	activates	*fliA*	New	This work
IHF	Not conclusive	*fliA*	-	-
Fis	Not conclusive	*fliA*	-	-
CRP	inhibits	*yeaJ*	New	This work
IHF	Not conclusive	*yeaJ*	-	-
Fis	activates	*yeaJ*	New	This work
CRP	Not conclusive	*ompR*	-	
IHF	Not conclusive	*ompR*	-	-
Fis	Not conclusive	*ompR*	-	-
CRP	Not conclusive	*adrA*	-	-
IHF	Not conclusive	*adrA*	-	-
Fis	Not conclusive	*adrA*	-	-

Chapter III

Effect of CRP, IHF And Fis in motility and biofilm formation in *E. coli* BW25113

1. Basis

Biofilm structure protects bacteria from different inputs as antibiotics and disinfectants leading to antibiotic resistance. This effect causes infections and deaths throughout the world. The generalist and commensal *bacteria* dwell, besides hospital and industrial supplies, the human intestine and evolve into pathogenic bacteria leading to urinary tract infections, hemorrhagic diarrhea, among others. Bacteria have the capability to generate biofilm inside human body (Hall-Stoodley et al., 2004). Experimentally, the transition between flagella and biofilm formation is evaluated by motility assays in which cells are allowed to swim in a low agar concentration media to determine the flagella effects. Biofilm formation is studied by two manners. The first is called “batch cells” that consist of a microliter dish in which cells are allowed to growth and attach at chemical inert surface on static conditions generating biofilm in an “aquatic environment”. The other is called “macro-colonies” in which a bacterial population grows over extended periods of time on agar plates leading to strikingly morphological patterns. This system resembles biofilms growing in organic material (To extensive go to Introduction 3).

O. Serra, et al., 2013 used the “visual hallmark” approach to determine the architecture and the cellular physiology of biofilm structure. Since liquid batch cultures are genetically and physiologically related to different growth phases. O. Serra et al. in 2013 employed stereomocroscopy, cryosection label assay and scanning electron microscopy on *E. coli* K12 wild type and mutant strains in order to uncover the physiological phases inside of biofilm structure and to correlate the morphology of the macro-colony with the regulatory network of the flagella-biofilm program. Interestingly, they show that exponential shaped cells are found at the edge and at bottom of the macro-colony, stationary cells are found at the surface of macro-colony, while a mixture of exponential and stationary shaped cells are found at the middle of the macro-colony. Additionally, from the edge to the center of the macro-colony, three zones have been postulated. At the edge, zone I is colonized by exponentially growing cells and is related with macro-colony expansion. At the middle - zone II - the exponential and stationary cells drive the expression of curli fibers (*csgAB*) generating wrinkles and concentric rings. At the center - zone III - the upper stationary cells produce high curli concentrations, while the bottom cells do not produce curli. Also, important correlations between genes of the regulatory network with the biofilm morphology were done. Briefly, the exponential cells are capable to produce flagella genes that are related with macro-colony expansion and with cell-cell contact, allowing the eventual

generation of a strong biofilm structure. Stationary cells curli and/or cellulose production are activated generating the extracellular matrix that allows hold the cells together. Importantly, those experiments were performed below 30˚C.

Our study design to evaluate the impact of GR in the dynamics of gene expression from genes of flagella and biofilm regulatory network by measurement of the promoter activity at 37˚C (human body temperature) demonstrated some of the interactions that have already been described (as explained in the previous chapters). Also, it allowed us to redirect those interactions that presented reports of both functions (activation and inhibition) at literature. Relevantly, our analysis suggests new regulatory interactions at the flagella-biofilm network by CRP, IHF and Fis. Subsequently, in order to fully characterize the effect of CRP, IHF and Fis on flagella and biofilm phases we performed motility and biofilm assays.

2. Motility program is actively regulated by CRP, IHF and Fis

Bacterial flagella are filamentous organelles that drive cell locomotion and adhesion. This organelle plays an important function during the different developmental phases of biofilm formation such as attachment at a surface as well as holding bacterial cells together inside of biofilm. Therefore, we were interested in understanding the effect of the global regulators CRP, IHF and Fis on the flagella function. For this, we growth *E. coli* BW25113 wild type and mutant strains for IHF and Fis harboring the empty vector pMR1 in fresh LB plates. A single colony was seeded (toothpick) in motility media plates 0.3% agar (Sperandio et al., 2002)complemented or not with glucose to induce down-activation of CRP (as indicated). Plates were incubated at 37˚C and motility was assessed during 8, 12, 18 and 24 hours.

Under our experimental conditions we were unable to detect motility at 8 neither at 12 h. As can be seen in **Figures 21** and **22** upper left panel, at 18 h and 24 h, *E. coli* BW25113 wild type strain shows its intrinsic capability to perform motility. This function represent the normal motility function in the presence of the tree global regulators CRP, IHF and Fis making it possible to compare it with all other conditions. CCR condition on *E. coli* BW25113 wild type by add glucose (mimicking Δ*crp*), negatively influenced motility at 18 and 24 hours **(Figure 21-22** upper right panel**)**. For the motility assay in *E.*

coli Δ*ihf* strain in the absence of glucose at 18 h induce a high motility phenotype, the effect was more notorious after 24 h, as full spread over the plate can be seen **(Figure 21-22** middle left panel**)**. The experimental condition of *E. coli* Δ*ihf* in the presence of glucose to down-activate CRP, shows reduced motility **(Figure 21-22** middle right panel**)**, remarkably, this condition still presented similar motility than wild type strain. Finally, *E. coli* Δ*fis* strain **(Figure 21-22** lower left panel**)** present an affected motility phenotype which is more evident at 24 h. Additionally, motility in *E. coli* Δ*fis* strain surveyed in motility media with glucose show the most detrimental effect **(Figure 21-22** lower right panel**)**. Altogether, our results show that (i) there is an intrinsic capability of *E. coli* wild type to perform motility in our experimental conditions; (ii) the absence of IHF improves motility, suggesting its pivotal role to modulate this program; (iii) The *E. coli* mutant strain for Fis presented reduced motility suggesting its role to activate motility and is strongly required to appropriate motility function. Therefore, we could conclude that CRP and Fis modulate positively the motility program while IHF repress the program since depletion of any of those regulators leads to strong phenotypic effects. Furthermore, all *E. coli* strains displayed different motility phenotypes in the presence of glucose, suggesting the involvement of a synergic interplay between those GR as well as the evidently interplay of other factors to regulate the motility network.

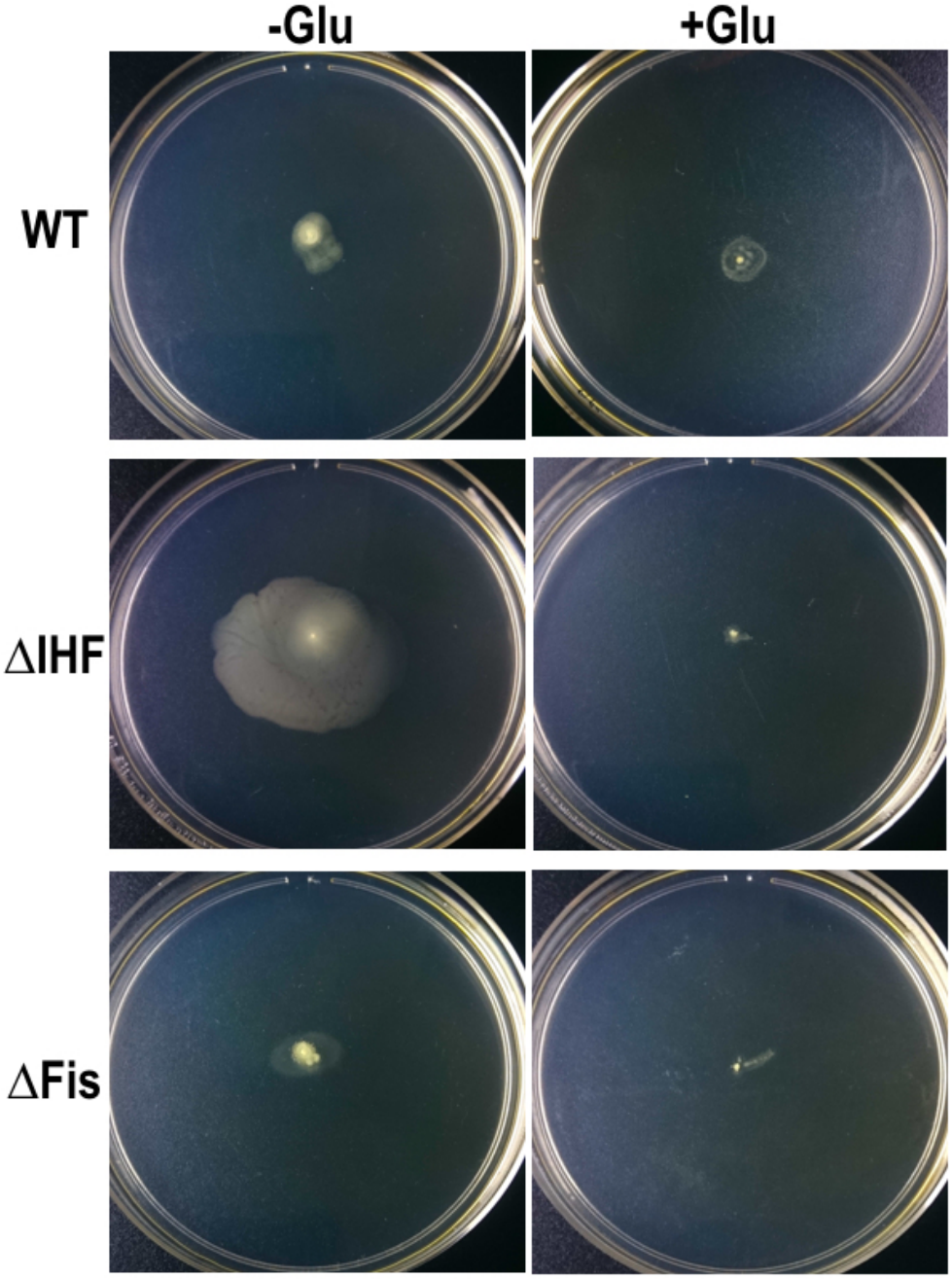


Figure 21: Effect of global regulators in the motility program at 18 h. Motility phenotype of *E. coli* BW25113 wild type and mutant strains were evaluated by cell motility assay at 18 h. Single colony of BW25113 wt, Δ*ihf* and Δ*fis* carrying empty vector pMR1 from fresh overnight LB plates, were seeded into motility plates in the presence or absence of glucose as depicted and incubated by 18 h. The results are representative of 3 independent experiments

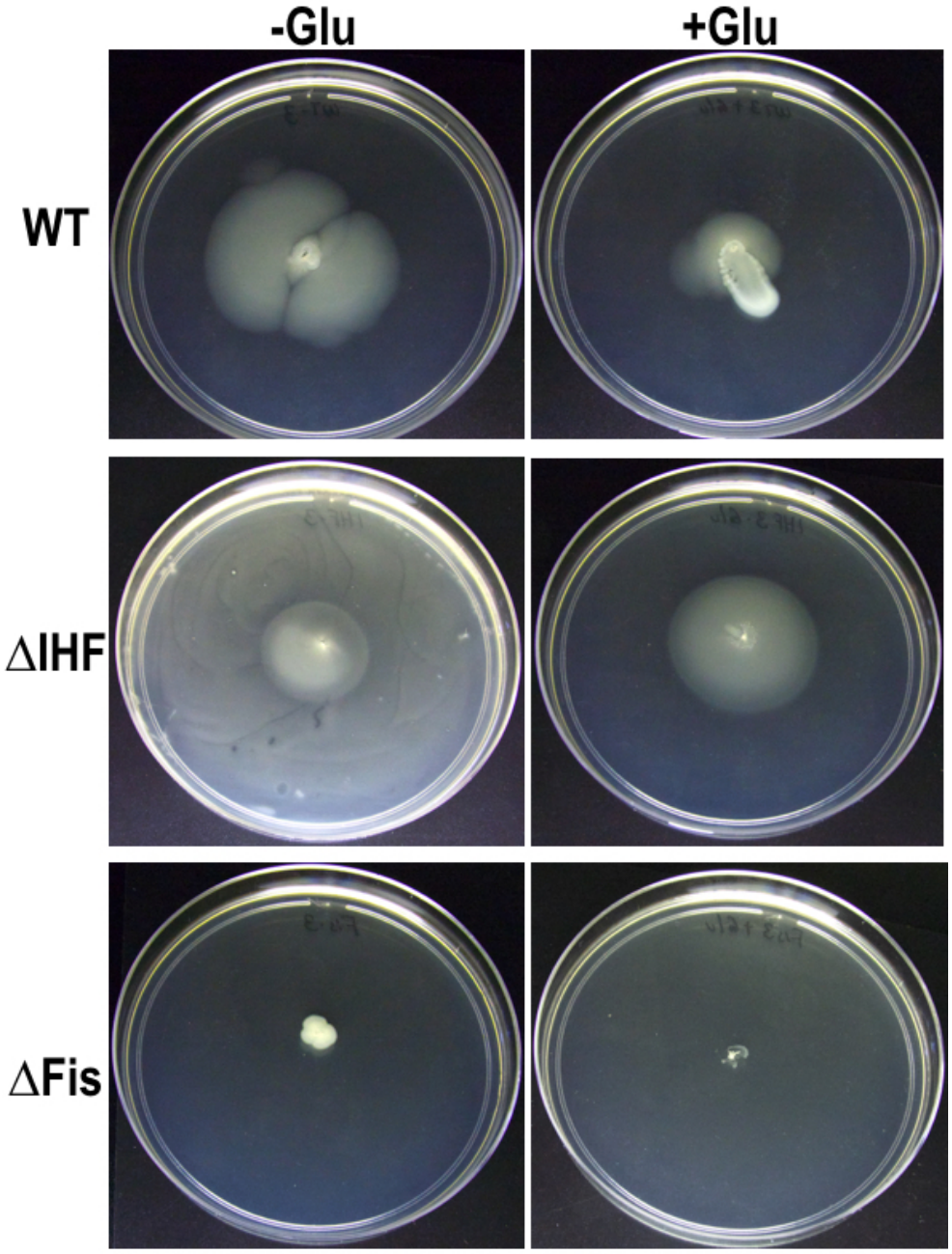


Figure 22: Global regulators modulate the motility program at 24 h. Motility phenotype of *E. coli* BW25113 *wt* and mutant strains were evaluated by cell motility assay at 24 h. Single colony of BW25113 *wt*, Δ*ihf* and Δfis carrying empty vector pMR1 from fresh overnight LB plates, were seeded into motility plates in the presence or absence of glucose as depicted and incubated by 24 h. The results are representative of 3 independent experiments.

3. Gene expression and microbiology experimental integration to understand the flagella-b*iofilm network*

The capability of *E. coli* to form biofilm was evaluated by liquid (crystal violet assay) and solid (congo red assay) media. Microtiter dish assay was used to measure the amount of biofilm generated due to the capacity of adherence of *E. coli* cells at the wall of a 96-well plate whereas Congo red was used to understand the intricate morphology of biofilm generated by *E. coli* community. Additionally, the morphological reports associating phenotype with specific modulators of the flagella-biofilm regulatory network have being established at 30˚C or below (Römling, 2005; Serra et al., 2013; Weber et al., 2006), although at 37˚C the same species can also perform biofilm in liquid media or express biofilm genes (Gualdi et al., 2008; Römling, 2005). For this reason, we speculate that specific characteristics of the biofilm must be shared in a temperature independent manner. Using this consideration, our results of promoter-gene expression profile explain some of the observed phenotypes at 37˚C. Therefore, we evaluated biofilm in liquid and solid media at both temperatures.

4. CRP, IHF And Fis modulate the adherence program in a temperature dependent manner

Early stages of biofilm formation are studied using microtiter dish assay (adherence phase). This because static batch-conditions does not allow the formation of mature biofilms. Interestingly, the assay has been effective to identify many factors required for initiation of biofilm formation (i.e, flagella, pili, adhesins, enzymes involved in cyclic-di-GMP binding and metabolism) as well as genes involved in extracellular polysaccharide production (O'Toole, 2011). Therefore, to start the biofilm characterization, *E. coli* BW25133 *wt* and mutant strains carrying the different plasmids were grown in LB. Overnight cultures were washed and resuspended 1:100 in fresh M9 medium in the presence or absence of glucose. They were added to 96-well plates and were incubated at 30˚C or 37˚C by five days. Finally, the capability of *E. coli* strains to adhere was measured at 600nm by crystal violet assay as described in methodology.

In general, *E. coli* cells adhered to the wall of the 96-well plates at all conditions. At 37˚C, two or three fold increases in biofilm yields were observed when compared with *E. coli* cells cultured at 30˚C **(FIGURE 23)**. This suggests a higher ability of *E. coli* to adhere to the tested surface at 37˚C.

Specifically, at 30˚C, *E. coli* Δ*fis* strain showed equal yields of early biofilm formation than the wild type strain, while IHF show a several decreases in this value (blue bars). In the presence of glucose (red bars), mimicking a Δ*crp* condition, the yields of early biofilm formation were affected. Exceptionally, *E. coli* Δ*ihf* in the presence of glucose, there were still low values of biofilm formation when compared with wild type strain, but they were higher than the *E. coli* Δ*ihf* strain without glucose. This data suggests that, at 30˚C, Fis protein seems to not to play a major role in the modulation of the early biofilm program and strongly support that mainly IHF but also CRP are necessary to activate the biofilm gene expression program.

At 37˚C, *E. coli* wild type strain showed the intrinsic capability to perform biofilm in liquid medium, and absence of IHF leads to a decrease in the biofilm yields. Interestingly, absence of Fis protein leads to a strong increase in adherence (blue bars). All those levels are affected in the presence of glucose (red bars). This data suggests that, at 37˚C, CRP and IHF are important effectors to activate the early biofilm program. It also suggests that Fis is acting as a repressor or down effector of the early biofilm program, since in its absence, at 37˚C, the motility is increased. Altogether, early biofilm yields are better induced at 37˚C than at 30˚C, and the three global regulators CRP, IHF and Fis participate differentially to regulate the adherence of *E. coli*.

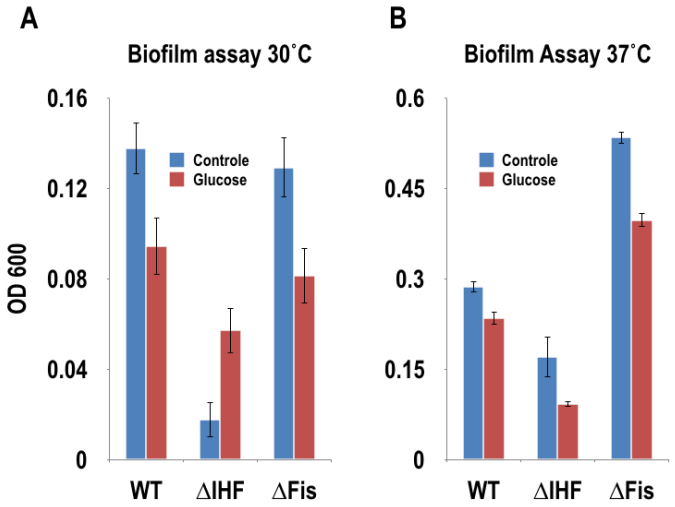


Figure 23: Temperature determine the adherence program output. Overnight cultures of *E. coli* BW25113 wild type, Δ*ihf* and Δ*fis* carrying all the constructions were diluted in fresh M9 and incubated at 30˚ and 37˚C with and without glucose as indicated. Capability to generate liquid biofilm was quantify by crystal violet assay at 600nm. Vertical bars are standard deviations called from three independent experiments.

5. CRP, IHF Aand Fis modulate the morphology of E. coli mature biofilm

Morphological characteristics are related to different physiological phases of bacteria inside of biofilm structure. Also, we can correlate mutant phenotypes with biofilm morphology to suggest the effect of the deleted gene into the regulatory network (Anwar et al., 2014; Serra et al., 2013). In this sense, we now know that specific characteristics such as, wrinkles and a central ring-like structure are generated by high levels of curli (*csgD* and *csgAB* operon). The flagella also participate at the beginning of wrinkles development. Those rules were determined at 28˚C, nevertheless we observe that bacteria induce better yields of liquid biofilm formation at 37˚C and the gene expression profile of the flagella-biofilm nodes was performed at 37˚C. Therefore, we seek to understand both the capability of *E. coli* to perform mature biofilm structure at 37˚C and the morphological characteristics of the macro-colony as consequence of the absence of any of the GR studied.

First, in order to evaluate the capability of bacteria to perform biofilm on solid media *E. coli* BW25133 wild type and mutant strains carrying pMR1 empty vector were grown in LB. Overnight cultures were washed and resuspended at same OD (0.5) in fresh M9 in the presence or absence of glucose. Five micro-liters drop of each culture were added into YESCA-CR plates supplemented or not with glucose (as indicated) and biofilm formation was allowed by incubate at 30 and 37˚C during six days. As depicted in **Figure 24**, all the strains and conditions are capable to develop complex biofilm structures at 30 and 37˚C in the presence or absence of glucose. In general, a central ring and wrinkles structures in all conditions is evident giving a volcano like-morphology. *E coli* cells grown at 30˚C show a soft surface structure while those grown at 37˚C conditions show an aggressive structure. The red-pale color of *E. coli* Δ*ihf* strain-structure at 30˚C suggest low curli production, while all structures strongly stained with Congo red.

Specifically, colonies grown at 30˚C **(Figure 25)** presented highly similar biofilm characteristics from those depicted to 28˚C (Serra et al., 2013). *E. coli* wild type strain in absence of glucose show a concentric ring delimitating the center of the colony (zone III), and discrete wrinkles structures are homogeneously distributed starting from the central ring to the edge of the coloni (zone II). A small expansion zone is observed. The *E. coli* Δ*ihf* strain presented an evident small size with a discrete red-pale central ring; red-less color coloni and absence of wrinkles highly suggest defections in curli production. Biofilm structure from Δ*fis* strain present similar red color, discrete wrinkles structures, a

concentric ring and slightly bigger size than wild type biofilm structure. *E. coli* cells at 30˚C in the presence of glucose, show similar characteristics than those without of glucose. Altogether, these data suggest that IHF is very important to biofilm formation at 30˚C, while CRP and Fis seems to slightly modulate the mature biofilm program.

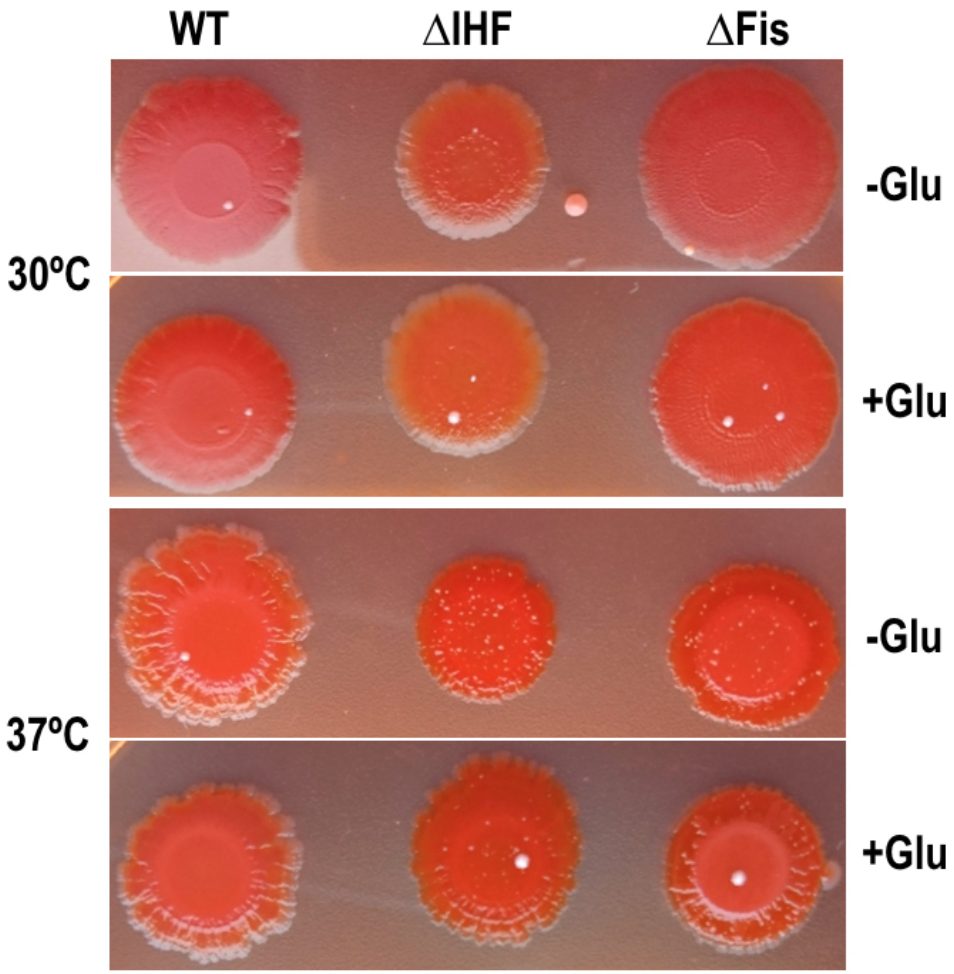


Figure 24: Capability of *E. coli* and mutant strains to develop mature biofilm. Overnight cultures of *E. coli* BW25113 wt, Δ*ihf* and Δ*fis* strains carrying pMR1 empty vector were washed and resuspended in M9 at 0.5 OD. Five micro-liters drop was seeded into YESCA-CR plates in the presence or absence of glucose as indicate. Capability to generate biofilm was documented after six days of incubation at 30˚C and 37˚C. The results are representative of 3 independent experiments.

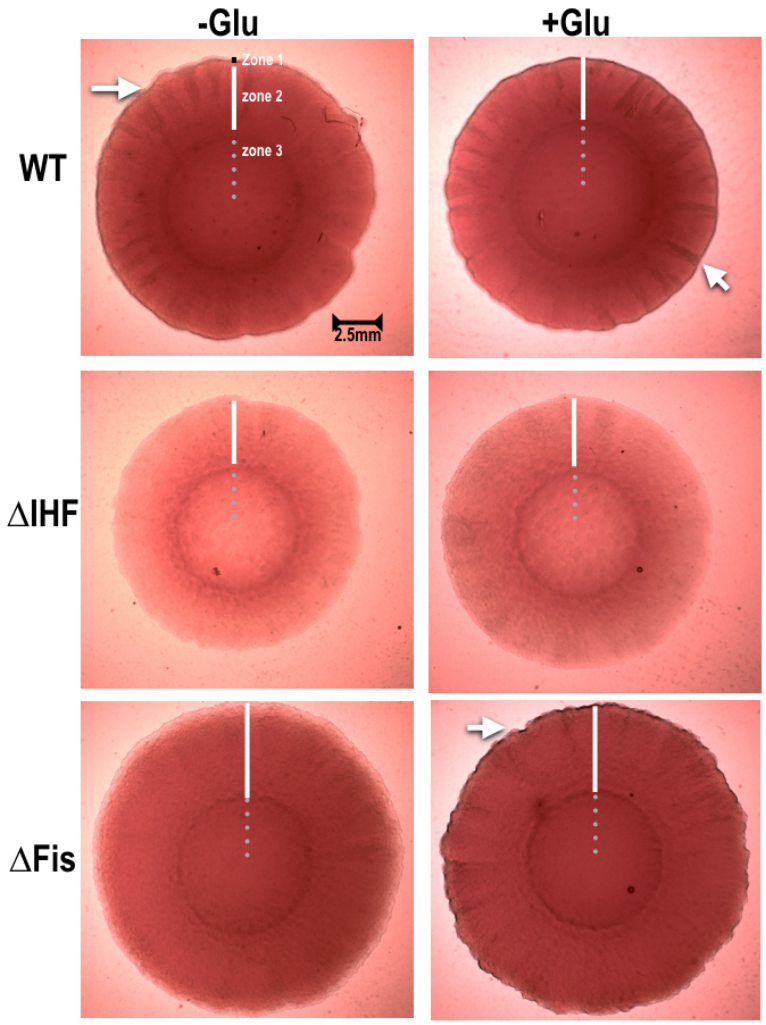


Figure 25: Biofilm morphology of *E. coli* BW25113 and mutant strains at 30˚C. Congo red assay was used to evaluate the biofilm morphology on *E. coli* BW25113 wild type and mutant strains (zoom 0.8X is shown). Overnight cultures of *E. coli* BW25113 wild type, Δ*ihf* and Δ*fis* strains carrying pMR1 empty vector were washed and resuspended in M9 at 0.5 OD. Five micro-liters drop was seeded into YESCA-CR plates in the presence or absence of glucose as indicate. Capability to generate biofilm was documented during six days of incubation at 30˚C. Three biological experiments were developed. Dot line, Zone III; white line, zone II; black line, zone I; narrows represent wrinkles structures.

E. coli BW25113 wild type in absence of glucose at 37˚C **(Figure 26A)** display the three zones already described to *E. coli* biofilm structure at 28˚C (Serra et al., 2013). A concentric ring delimitates the zone III, which presents an intense red color suggesting curli formation by the different layers of stationary phase cells. Wrinkles structures were homogeneously distributed on the three zones as can be corroborated in a closer view (A lower panel). Zone II represents the intrinsic capability of wild

type community to produce wrinkles and curli. Finally, zone I presents weak red color suggesting that, bacteria at exponential phase are related with colony expansion program more than curli production. Altogether, we presented the intrinsic morphology of a mature biofilm in the *E. coli* BW25113 wild type strain at 37˚C. In the presence of glucose **(Figure 26B)**, the behavior of biofilm formation is a quite similar. As depicted, *E. coli* BW25113-wt in the presence of glucose (Δ*crp* mimic condition) presents a slight small size than wild type strain, however not significative **(Figure 29)**. Expansion zone show a red-pale color showing the involvement of flagellated cells. At zone II, curli fimbriae were considered as positive since the red color was very similar to the one from wild type strain. Wrinkles structures were present at the three zones suggesting a coordinately interplay between post exponential and stationary cells. Altogether, this data suggest that CRP protein discreetly modulate the mature biofilm program in and apparently CCR dependent mechanism as reported to *E. coli* and other species (Hufnagel et al., 2016; Willias et al., 2015).

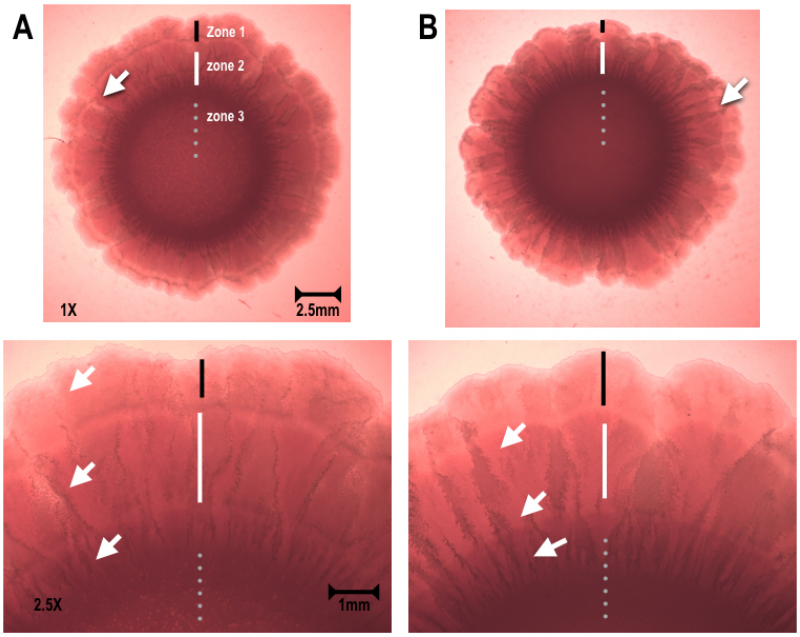


Figure 26: Biofilm morphology of *E. coli* BW25113 wild type strain and +glucose condition at 37˚C. Congo red assay was used to evaluate the biofilm morphology on *E. coli* BW25113 wt strain, zoom 0.8X (upper panel) and 2.5X (down panel) are shown. Overnight cultures of *E. coli* BW25113 wt strain carrying pMR1 empty vector were washed and resuspended in M9 at 0.5 OD. Five micro-liters drop was seeded into YESCA-CR plates in the presence or absence of glucose as indicate. Capability to generate biofilm was documented during six days of incubation at 37˚C. Three biological experiments were developed. Dot line, Zone III; white line, zone II; black line, zone I; narrows show wrinkles structures.

The *E. coli* Δ*ihf* community **(Figure 27A)** presented the most detrimental size when compared with all conditions of biofilm formation **(Figure 29)**. The expansion zone is evidently reduced. The zone of curli fimbriae formation display a red-pale color and size reduction, suggesting less curli formation. Soft wrinkles structures are present at zone II and I but not in aggregation zone (B low panel). Interestingly, this mutant presents a particular structure that we call as "cleft" (white narrows); different from wrinkles structures, clefts present a slight dark-red color, clearly suggesting bacterial cells in other physiological state than their neighborhood cells. Strikingly, biofilm structure of Δ*ihf* strain in the presence of glucose **(Figure 27B)** presents evidently increased expansions at zone I when compared with Δ*ihf* biofilm. This effect, leads to similar biofilm size than wild type strain **(Figure 29)**. Wrinkles and clefts-like structures are observed only at expansion and curli fimbriae zone. Red-pale color suggests less curli fimbriae formation in this condition. All this data suggest that IHF is a mayor modulator of the mature biofilm program, and that the interplay between CRP and IHF is necessary to the properly biofilm development.

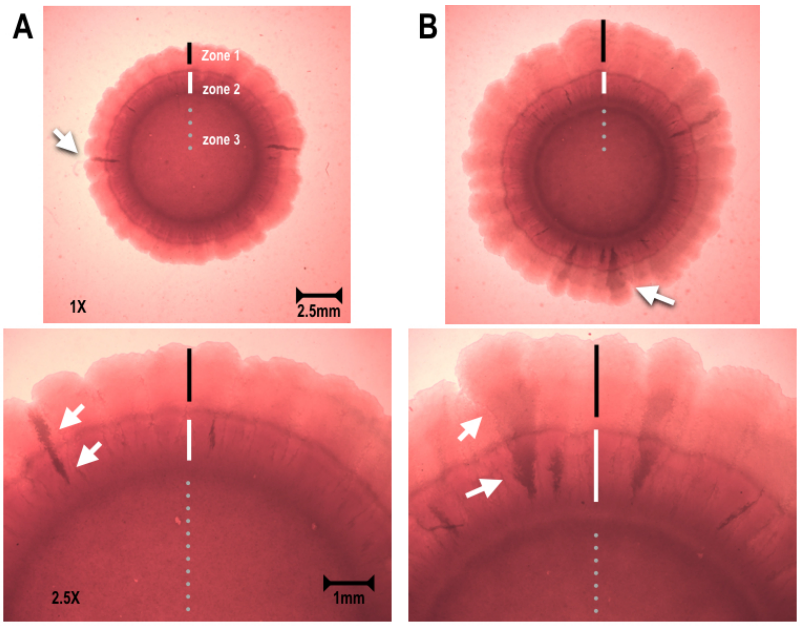


Figure 27: Biofilm morphology of *E. coli* BW25113 Δ*ihf* strain and Δ*ihf* / +glucose condition at 37˚C. Congo red assay was used to evaluate the biofilm morphology on *E. coli* BW25113 Δ*ihf* strain, zoom 0.8X (uper panel) and 2.5X (down panel) are shown. Overnight cultures of *E. coli* BW25113 *wt* strain carrying pMR1 empty vector were washed and resuspended in M9 at 0.5 OD. Five micro-liters drop was seeded into YESCA-CR plates in the presence or absence of glucose as indicate. Capability to generate biofilm was documented during six days of incubation at 37˚C. Three biological experiments were developed.

Depletion of Fis protein in *E. coli* leads to an affected biofilm structure **(Figure 28A)** but with the same size than the wild type strain in the same conditions **(Figure 29)**. Aggregation and curli zone present a dark red color, suggesting curli fimbriae production. Cleft structures are visible but not wrinkle structures, suggesting that this structure could, in fact, be different to wrinkles structures, possible a channel with a specific function for the community. Interestingly, at the bottom of this stricture (low panel) is observed a bacterial community with red-less color was observed, suggesting exponential cells possible related to colony expansion. Finally, *E. coli* Δ*fis* in the CCR condition (**Figure 28B**) showed size reduction **(Figure 29)**. Equal to the Δ*fis*, aggregation and curli zone presented a dark red color, suggesting an increase in the levels of curli fimbriae production even when compared to wild type biofilm. Cleft structures were observed only at zone 2. Interestingly, in a closer view at the edge (low panel), is possible to appreciate discrete regions with a slight diminishment in red color, suggesting the presence of an expansion zone.

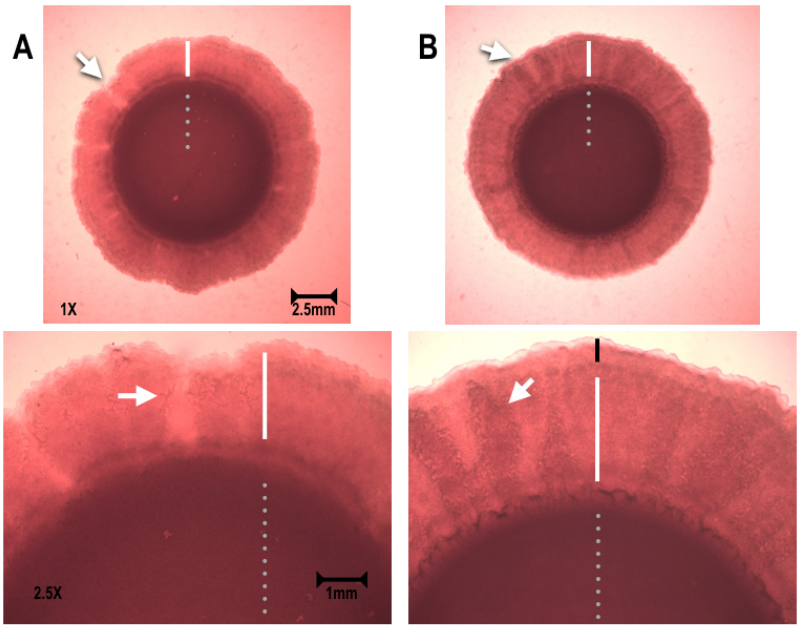


Figure 28: Biofilm morphology of *E. coli* BW25113 Δ*fis* strain and Δ*fis* / +glucose condition at 37˚C. Congo red assay was used to evaluate the biofilm morphology on E. coli BW25113 Δfis strain, zoom 0.8X (uper panel) and 2.5X (down panel) are shown. Overnight cultures of E. coli BW25113 wt strain carrying pMR1 empty vector were washed and resuspended in M9 at 0.5 OD. Five micro-liters drop was seeded into YESCA-CR plates in the presence or absence of glucose as indicate. Capability to generate biofilm was documented during six days of incubation at 37˚C. Three biological experiments were developed.

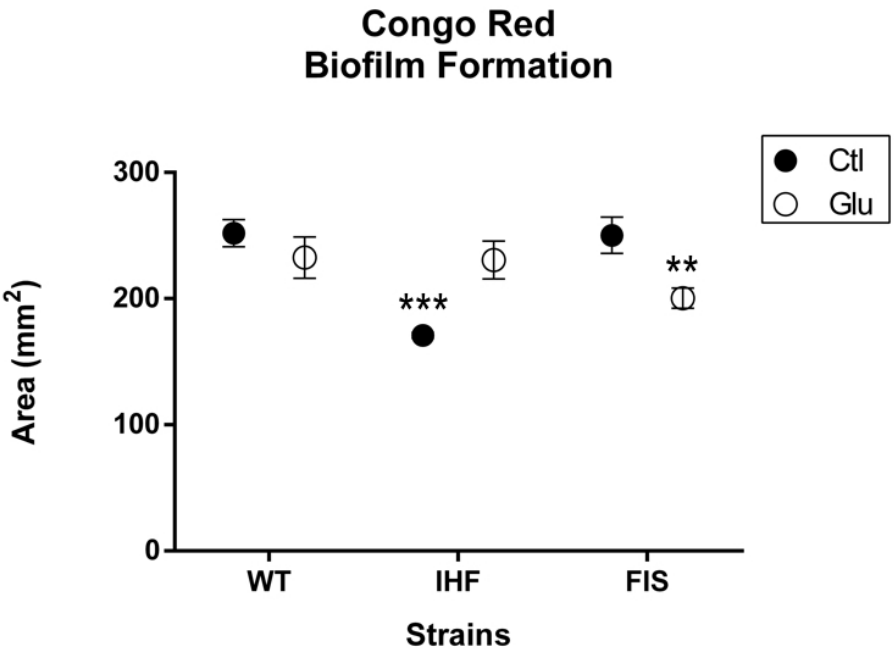


Figure 29: Effect of CRP, IHF and Fis in the biofilm size. Congo red assay was used to evaluate the biofilm morphology on *E. coli* BW25113 wild type and mutant strains. Overnight cultures of *E. coli* BW25113 wild type, Δ*ihf* and Δ*fis* strains carrying pMR1 empty vector were washed and resuspended in M9 at 0.5 OD. Five micro-liters drop was seeded into YESCA-CR plates in the presence or absence of glucose as indicate. Capability to generate biofilm was documented during six days of incubation at 37˚C. Images of biofilm structure were loaded in imageJ software. Biofilm size was compare between strains by anova test. Three biological experiments were developed.

Altogether, morphology assays have shown that IHF and Fis global regulators are important contributors to properly biofilm development, as suggested by computational analyses (Martínez-Antonio et al., 2008) and proved here experimentally. By another hand, the role of CRP protein during mature biofilm formation seems to be ambiguous in function of the community necessities. Therefore, we can conclude that CRP, IHF and Fis proteins modulate the behavior of the community at different phases by modulating the gene expression profile in a temperature dependent manner and support the hypotheses that common characteristics or “biofilm rules” are shared between different experimental biofilm structures.

V. DISCUSSION

CRP, IHF and Fis at the flagella-biofilm transcriptional regulatory network: an overview

Network principles applied to High-throughput data show that bacterial systems present a hierarchical transcriptional regulatory in which nine global regulators modulate most of the bacteria genome (Martínez-Antonio and Collado-Vides, 2003), from those, CRP and IHF global regulators were described as important modulators of the flagella-biofilm program (Martínez-Antonio et al., 2008). Also, *E. coli* mutant strains has been examined genetically and morphologically to understand the effect of that protein in the flagella-biofilm network (Pesavento et al., 2008; Serra et al., 2013). In this thesis, we used bioinformatics tools and molecular biology together at microbiology approaches to understand the effect of the global regulators CRP, IHF and Fis in the flagella-biofilm network (Gualdi et al., 2008; Moor et al., 2014; Prigent-Combaret et al., 2012; Saldaña et al., 2009; Sheikh et al., 2001; Vijay et al., 2015) **(Chapter I)**. We proved that IHF and Fis global regulators are important modulators of the flagella as well as early and mature biofilm program, while CRP seems to exert different effects **(Chapter II)**. Nevertheless, in the conditions mimicking the Δ*crp* condition by add glucose, down-activation of CRP seems to potentiate the mature biofilm program developed by the genetically depleted strain surveyed in absence of glucose, supporting the pleiotropic effect of CRP and this will be discussed in all subsections. Additionally, we found that early and mature biofilm programs are temperature dependent, due to the differences at the yields and morphological characteristics in biofilm formation surveyed at 30 and 37˚C. Interestingly, temperature approach in morphological assays allowed us to suggest the existence of “a common mature biofilm rules” **(Chapter III)**. In the next subsections we interpret and discuss our results at 37˚C in order to understand the role of each global regulator in the complex flagella-biofilm transcriptional regulatory network. By the end we discuss the bacterial offset to the loss of any global regulator for the proper biofilm development, that we call "compensatory mechanism". Because of this, we also discuss the necessity to improve the connectivity of this network seeking to a better understanding.

The pleiotropic effect of Cyclic-AMP Receptor Protein in the flagella-biofilm network

The Cyclic AMP Receptor Protein (CRP) is a global transcription factor which have the capability to sense external stimulus by the association with the small molecule cAMP. It modulates the CRP activity over the promoter regions of cognate genes changing the gene expression in order to perform

an appropriate metabolic output. Interestingly, the presence of different carbon sources as glycerol, glucose, fructose, xylose, and arabinose, differentially modulates the CRP activity so different metabolic networks are activated determining organism adaptability (Shimizu, 2016). In general, it is well known that CRP modulates that mechanism in a glucose concentration manner playing an important role in modulating bacterial metabolism.

One of the most important microbial survival strategy is biofilm formation. In this structure, bacteria in different physiological phases are responsible to the properly interplay between motility, adhesion, cell-cell contact, extra polysaccharide matrix formation, among others programs. CRP global regulator is differentially expressed during the growth curve. Out degree analysis show that CRP protein is highly connected with different effectors of the flagella-biofilm network evidencing its involvement in the network **(Figure 13B)**. We evaluated the role of CRP in the flagella-biofilm network by adding 0.4% of glucose to the media of *E. coli* wild type and mutant strains, this concentration is adequate to inhibit transcription of genes CRP-dependent, as we corroborate to the Plac promoter (Amores et al., 2015) - Attached. In fact, almost all promoters surveyed in *E. coli* BW25113 with glucose presented different activity when compared with *E. coli* grown in glycerol as solely carbono source suggesting that CRP modulates the gene expression of some analyzed genes as described in results and this, is also true for the mutant strains **(Chapter II)**. This allowed us to validate and to add new connections in our network **(Figure 20C and Table 4)**. It is clear that those global regulators modulate the flagella-biofilm transcriptional regulatory network with evident effects at motility, early and mature biofilm as is shown in this thesis **(Chapter III)**. Therefore, we focus in understand the morphological evidences with the genetical program developed in the surveyed strains. Beloin, et al., in 2004 surveyed the transcriptome of planktonic cells against batch cultures and biofilm structure. They identified 39 exclusive biofilm genes and from those, 25 were expressed in batch culture as *rpoE* and *cpxP*, showing that some characteristics are shared between static conditions and biofilm structure and establishing that stress and biofilm related genes are preferentially expressed than those of central metabolism (Beloin et al., 2004).

In wild type strain surveyed in the presence of glucose, mimicking a Δ*crp*, the main effectors of the network *rpoS, csgD* and *flhD* presented markedly changes in promoter activity **(Figure 15, 16, 18)**, while all other promoters showed slight changes. Thus, the control of the network seems to depend on a fine interplay between those regulators. This basic interplay correlates with the morphological

differences. For instance, low *flhD* expression in the absence of CRP **(Figure 18)** leads to defects in motility **(Figure 21, 22)**, while differences in *csgD* **(Figure 16)** leads to slight differences in early and mature wild type biofilm program **(Figure 23, 24)**. The scenario is similar in genetically depleted strains Δ*ihf* or Δ*fis* in the presence of glucose: motility is affected and early biofilm is affected (explained perhaps by *flhD* and *rpoS* decreases) but the absence of CRP seems to enhance the morphological characteristics of mature biofilm than same strain without glucose. This effect will be explained in detail below.

Specifically, for the morphology of wild type strain in the presence of glucose, the complex interplay between *flhD* and *csgD* is modulated by the main effector *rpoS.* This last node presented higher GFP levels than the negative control, suggesting that other proteins can activate the *rpoS* transcription increasing the complexity of the network **(Figure 15)**. This is in according with the well-known effect of the master regulators *rpoS, csgD* and *flhD* in the flagella-biofilm network, in which those nodes are the main responsible of global changes in gene expression, cellular physiology and thus, morphological changes. However, *rpoS* or *csgD* genetically depleted strains, although affected, are capable to generate biofilm (Serra et al., 2013). Additionally, You et al., in 2013 showed that, besides the classical knowledge of the central role of cAMP in CCR glucose mediated mechanism, cAMP also leads to a compensatory mechanism in which its signals allows allocation of the proteomic resources with different metabolic demands in different nutrient environments (You et al., 2013). Compensatory mechanisms have also been described at DNA for mutations, and at population level for antibiotic resistance in bacteria (Levin et al., 2000; MacLean and Vogwill, 2014). Altogether, the biofilm structure is an advantageous community behavior, thus a compensatory mechanism must exist to the properly network performance and biofilm development. Finally, as the CCR effect is complex, evaluating our GFP systems in Δ*crp* strain with its proper morphological description will leads a better understanding of CRP in the network supporting the flagella-biofilm compensatory mechanism.

Integration Host Factor is a key effector in the flagella-biofilm network

IHF global regulator is a modulator of genes involved in the flagella-biofilm program as *rpoS, matA* and *csgD* (Hengge-Aronis, 2002; Mangan et al., 2006; Martínez-Santos et al., 2012; Ogasawara et al., 2010). Recently, it has been proved that specific antibodies against IHF lead to bacterial biofilm

disruption. This strategy mainly disrupts the extracellular DNA matrix in which IHF is associated making part of the biofilm structure (Brandstetter et al., 2013; Brockson et al., 2014). The molecular mechanism involved in the different phases of biofilm formation by IHF remains blurry. However, in this work, we presented evidence that, IHF is a key element to biofilm formation by modulating gene expression level of the flagella-biofilm program, pointing that IHF could be a good candidate to disrupt bacterial biofilm in a molecular way more than affect the structure.

In Δ*ihf* scenario, we observe an important GFP increase in the main effector of the network, suggesting that IHF modulates the transcription of RpoS by blocking the *rpoS* promoter region **(Figure 15)**. Mangan et al., 2006, used a *Salmonella enterica serovar Typhimurium* strain depleted of IHF to understand its effect at gene expression. They suggested that IHF also modulates genes at stationary phase. However, as no changes at *rpoS* transcripts were found, they proposed that both proteins work in a synergic way to modulate the stationary cognate genes. Interestingly, this same report showed increases at RpoS in Δ*ihf* strain (Mangan et al., 2006). We proved in batch cultured *E. coli* cells that Δ*ihf* increases the *PrpoS*-GFP levels, suggesting that IHF directly represses the *rpoS* transcription **(Figure 15)**. We also observed different dynamic activities in all promoters surveyed through the bacterial development as described in results **(red lines chapter II)** with important new connections **(Table 4)**. Therefore, we discuss how those changes at gene expression by absence of IHF leads the effects observed in our morphological surveys.

The transcriptional regulatory flagella-biofilm network for *E. coli* established that, at stationary phase the highly expressed *rpoS* inhibit flagella genes in a process GGDEF/EAL proteins dependent, triggering the biofilm program (Dudin et al., 2014; Pesavento et al., 2008; Sanchez-Torres et al., 2011; Sommerfeldt et al., 2009). This is related to the heterogeneous *rpoS* expression trough the biofilm structure (Ito et al., 2009). Thus, since Δ*ihf* shows a high expression of *rpoS* **(Figure 15)**, it could be possible that this strain present affected motility. Interestingly, motility assay in Δ*ihf* strain showed a highly increased motility phenotype **(Figure 21, 22)**. GFP activity, mapping the dynamics of the main nodes in this network, also showed an increased expression of *rpoE* and *fliA* sigma factors but not *flhD* **(Figure 16, 18)**. In *Salmonella enterica serovar Typhi*, a gram-negative bacterium under hyper osmotic stress, *rpoE* activates *fliA* and not *flhD* suggesting a balance between *rpoS* and *rpoE* sigma factors to modulate gene expression (Du et al., 2011a; Du et al., 2011b). Altogether, the high

and sustained *rpoE* expression could lead *fliA* activation explaining a coherent path to the highly motility phenotype supporting the edge betweenness path **(Figure 20C)**.

Regarding the evident defective capability of Δ*ihf* to perform the early and mature biofilm programs **(Chapter III)**, we wondered how this strain, despite the high levels in *fliA* and low levels in *csgD* **(Figure 16, 18)**, which show low attachment and low biofilm performance **(Figure 23, 24)**, is still capable to perform both programs similar to the wt strain? In this sense, our results suggested that constitutive *yeaJ* expression at same levels than wt **(Figure 19)** will modulate the c-diGMP levels (or any other GGDEF/EAL protein) leading to *rpoS* activation **(Figure 15).** This mechanism eventually activates the curli-related (*csgD*) gene being possible the development of early and mature biofilm formation such as our results of liquid and solid biofilm formation. Following this reasoning, Isalan, et al., 2008 presented evidence that changes at *rpoS* and *ompR* node leads to changes in 39 genes after heat shock exposure, *csgD* being one of those, and suggest that *cpxA* network is a modulator of both pathways to activate *csgD (Isalan et al., 2008)*. Our results present at slight increase in *ompR* and *cpxR* promoters **(Figure 17, 19)**. This entire picture clearly explains the effects observed in our motility **(Figure 21, 22)**, early **(Figure 23)** and mature biofilm **(Figure 24, 27)** assays for Δ*ihf* strain. Therefore, we suggest that the compensatory mechanism activated by Δ*ihf* to modulate the properly flagella-biofilm program is through *yeaJ-rpoS-rpoE-ompR-cpxR* interplay, leading to *csgD* and *fliA* activation. Finally, in a more complex view, our results also suggest a direct modulation of *rpoS, rpoE, fliA, yeaJ* and *csgD* by IHF, since our expression assays have shown different GFP levels of those nodes in the mutant strain **(red lines Chapter II)**. This mechanism of direct regulation was demonstrated for IHF over *csgD* by Ogasawara, et al., 2010.

The *E. coli* Δ*ihf* plus glucose (mimicking a Δ*crp*), scenario is more complex. The dynamics of GFP between grow conditions (with or without glucose) presented similar dynamics with changes at expression levels i.e. *rpoS* yields drops at similar levels than *wt* strain while *rpoE* and *yeaJ* increase and are constant over the growth curve, plus the effect of glucose, leads to low *fliA* and *csgD* GFP yields than *E. coli* Δ*ihf* without glucose **(red lines Chapter II)**. As we can see, the characteristics are similar to the compensatory program described above. Except by the decreases in *flhD* promoter activity **(Figure 18)**, this possibly caused by an additional mechanism in which OmpR inhibits the *flhD* transcription (Samanta et al., 2013; Shin and Park, 1995). Altogether, this can explain the decreases at motility **(Figure 21, 22)** and in part the effect at the early biofilm program **(Figure 23)**. And we

suggest “in part” because we did not surveyed any node related to adherence that would allow us to identify the cause which, despite affected, still perform liquid biofilm **(Figure 23)**. Δ*rpoE* strain shows defects to develop liquid biofilm (Huang, et al., 2015). Thus, *rpoE* expression together *yeaJ* levels **(Figure 16)** in our data could explain the capability to generate liquid biofilm, still remaining uncover the specific node effector for the attachment program. Explaining the morphology seems as even harder task because of the low levels of *fliA* and even lower of *csgD* **(Figure 16, 18)**. Interestingly, the biofilm program is not affected, in fact, it presents a diminished zone II and pronounced zone I **(Figure 27B)**. Probably the cell resources are allocated mainly to perform motility than biofilm, explaining the GFP levels of *fliA* and *csgD* **(Figure 16, 18)**, thus suggesting that those values are enough to trigger the mature biofilm program. All this also suggests a delicate interplay between input intensity and protein production; we tried to address that issue, however, our translational fusions did not show activity.

Altogether, we proved that IHF modulates flagella-biofilm gene expression. Additionally, the studied strain seems to present robustness at its transcriptional regulatory network, since in absence of other effectors as CRP, the network is still capable to process the information for the adequate motility and biofilm structure development. Therefore, the bacteria activate a compensatory network to induce a “properly” flagella-biofilm program. In this sense, Δ*ihf* in P. putida had its transcriptome affected, genes at membrane composition and biofilm formation were affected suggesting that, besides its transcriptional regulator function, IHF can control several functions that are required at post-exponential stage (Silva-Rocha et al., 2013).

Factor for inversion stimulation exerts pivotal effects in the flagella-biofilm network

The role of Fis in biofilm formation is quite intricate. Both Δ*fis* and its over-expression strain presented higher levels of biofilm formation with production of curli and (or not) cellulose in different bacterial species (Jakovleva et al., 2012; Moor et al., 2014; Prigent-Combaret et al., 2012; Saldaña et al., 2009; Sheikh et al., 2001; Vijay et al., 2015). In a more complex interplay at 37˚C and in a *csgD* independent process, Δ*fis* strain produced high levels of curli and presented high levels of adherence due to activation of *csgBA* operon (curli proteins) and (or not) *adrA* gene the main cellulose producer (Gualdi et al., 2008; Saldaña et al., 2009).

In this background, we surveyed the role of *E. coli* Δ*fis* BW25113 on the main nodes of the flagella-biofilm program **(green lines chapter II)** plus its morphological description **(Δ*fis* strain chapter III)**. We found differences at dynamics and intensities of GFP levels at same nodes as follow. In our results, Δ*fis* strain is working as a repressor because it unblocks the transcription of *rpoS, rpoE* and *yeaJ* but not *fliA* neither *flhD*, while is acting as an activator of *csgD*. Therefore, the characteristics in this strain are: a highly expressed *rpoS* and *yeaJ* **(Figure 15, 19)**, an *rpoE* entirely Fis dependent **(Figure 16, 19)**, with *flhD* and *fliA* levels similar to the *wt* strain but with decreases at *csgD* expression **(Figure 16, 18)**. Clearly, the edge betweenness measurements describe the same program as described to Δ*ihf* and *wt* strain **(Figure 20C)**. At glance, *rpoS* hyper activation implies this may be the main effector of the "hyper phenotypes" observed in Δ*fis*.

In bacteria, at 37˚C *rpoS* differentially modulates genes of adherence and biofilm structure (Collet et al., 2008). Shifts from 37˚C to 23˚C lead to expression of biofilm, cold shock and *rpoS* dependent genes (White-Ziegler et al., 2008). Interestingly, RpoS (active <37–41°C) at stationary phase when reactive oxygen species (ROS) increase, the viable but nonculturable (VBNC) program is active in some cells and, at this growth conditions; also R*poE* is active, leading to lysis of an amount of cells in that population. In heat-shock a peak of R*poE* activation with R*poS* activation leads to a survival program (Noor, 2015). Finally, r*poE* is transcribed in biofilm structure (Beloin et al., 2004) and depletion of this gene leads impaired biofilm in *Salmonella pullorum* (Huang et al., 2015). Evidently the peak of RpoE presented in the Δ*fis* strain with the highly active RpoS at 37˚C are activating different signals which can be differentially sensed by bacteria activating different mechanisms as survival or stress, modifying the properly flagella-biofilm program. We will discuss in detail those possible mechanisms further.

The detrimental motility in Δ*fis* strain **(Figure 21, 22)** is clearly explained by the low activity of *fliA* and the detrimental levels in *yeaJ* and the unchanged *flhD* levels with respect wild type strain **(Figure 18)**. This corroborates with the affected motility observed in *E. coli* depleted of *yeaJ* at 37˚C, besides this report suggest the involvement of *yeaJ* to change the levels of c-di-GMP modulating the biofilm program (Pesavento et al., 2008). Here we suggest that Fis is activating the *yeaJ* promoter, thus the motility program is affected as reported (Pesavento, et al., 2008). While increases in early biofilm program **(Figure 23)** and morphology evidences **(Figure 24, 28)** may be explained, in part, by the changes at *rpoS*, *rpoE* and *csgD* **(Figure 15, 16)**. The peak of *rpoE* signal, perhaps, is sensed by

bacteria as a stress program rather than biofilm program. The weak, but still *yeaJ* activation, would induce a wave of the important biofilm signaling molecule c-di-GMP. Which together with RpoS hyperactive, is leading to a bulk of complex signals. Therefore, increases in early biofilm formation **(Figure 23)** could be explained by *rpoE* signals which together at low c-di-GMP are inhibiting flagella program and activating bacterial adherence, equal than Δ*ihf*, still remaining uncover in our model the effector node for this program.

Morphology of Δ*fis* strain, presented the most complex structure **(Figure 24, 28)**.The biofilm structure of *E. coli* Δ*fis* at 37 ˚C looks like 30˚C biofilm **(Figure 25)** but with a stressed, aggressive or old structure. Our results suggest that hyper activated *rpoS* with an early but short peak of *rpoE* activation plus a possible weak c-di-GMP signals induced by YeaJ, induce biofilm program as described to 23˚C (White-Ziegler et al., 2008). At the same time, the peak of *rpoE* activation together with *rpoS* constitutive signal will lead to activation of stress and survival mechanisms (Noor, 2015). Finally, the low amounts of *csgD* (caused by low *yeaJ*) or other signals seems to be enough to triggers the curli production but not cellulose as evidenced by the high red color in congo red assay **(Figure 28)** and low levels of adrA in GFP expression assay **(Figure 19C)** and showed by Saldaña, et al., 2009.

Finally, in the Δ*fis* strain with glucose (Δ*crp*-like), as explained to Δ*fis* strain, changes in GFP intensity at same nodes *yeaJ-rpoS-rpoE-csgD* and low activity of *flhD* can explain the changes observed **(green lines / glucose chapter II)** as the lowest motility and diminished quantities of early biofilm formation, still remaining the effector node responsible to attachment **(Figure 21, 22, 23)**. Therefore, we placed special interest in the morphological experiment **(Figure 28)**. This biofilm present small size **(Figure 29)** and a more aggressive structure than Δ*fis* strain **(Figure 28B)**. The still higher levels of *rpoS* with increases in *rpoE* and weak *yeaJ* with low *flhD* **(Figure 15, 16, 18, 19)**, can be playing the same behavior as describe above. Altogether, we suggest than in the absence of Fis, bacterial community is highly stressed and at same time is producing biofilm. Thus; in these conditions, besides an apparently increase in curli production, lysis program would be active changing the biofilm environment increasing ROS and inducing, perhaps, VBNC cells (Noor, 2015). This remain to be investigated.

The complexity of the flagella-biofilm network interplay — final remarks

The evidences in this thesis prove that, CRP, IHF and Fis are important effectors that modulate the flagella-biofilm network. Interestingly, none condition in this study show absence of motility or biofilm formation **(Chapter III)**, suggesting that, none of those GRs are essential genes to suppress completely the flagella-biofilm program. Additionally, and as stated, bacteria without the main effectors *rpoS* or *csgD* (also *rpoE, csgB, flhDC or fliA*) are still capable to perform biofilm (Huang et al., 2015; Serra et al., 2013). Altogether, the high plasticity observed in the flagella-biofilm program clearly show that the network present a natural behavior to compensate the node-less perturbation, showing the importance of this program to bacterial survival. In agreement with the robustness proposed to the flagella-biofilm transcriptional regulatory network, we observed that *E. coli* wild type or Δ*ihf* or Δ*fis* as well as CCR induction in all strains presented changes mainly at the major nodes *rpoS, rpoE, flhD, fliA, csgD* explaining the logic path that compensates the system in the absence of those GRs. Finally, we suggest that, those master regulators as well as the global regulators surveyed process the information for the adequate motility and biofilm structure development in the presence or absence of any of them or other nodes.

This small view of the complex flagella-biofilm transcriptional regulatory network explains the behavior of the network in the absence of *rpoS* or any other node. Clearly, the system finds an equilibrium to develop this program. Based on our data, we propose a compensatory mechanism of bacteria to develop biofilm; it has been proved that bacteria develops a strategy of metabolic coordination in CCR induction in order to the properly bacterial fitness (You et al., 2013) as well as compensatory mechanisms for antibiotic resistance and mutations (Levin et al., 2000; MacLean and Vogwill, 2014). Therefore, we suggest that the global regulators coordinate a compensatory mechanism mainly by changes at the connectivity and at transcriptional expression over the nodes of the flagella-biofilm network rather than alter the metabolism. Finally, find as much connections as possible for a specific network will allow predicting the behavior of the biological system in the absence of different TFs. Also, it will open the possibility to use genetic engineering to achieve a balanced connectivity between synthetic circuits and the entire bacterial network to perform a specific task. In other words, rational interactive synthetic circuits in accordance to the natural network of a specific organism will allow a higher performance of the desired biological task.

VI. CONCLUSIONS

From the results obtained in this thesis, we can draw the following conclusions:

1. Using network approach to construct and analyze the flagella-biofilm transcriptional regulatory network of bacteria, we were available to identify the main effectors of the network as well as the possible path that modulate the network.

2. The promoter dynamics assays using transcriptional fusions to GFP reporter allowed us to show than CRP, IHF and Fis coordinate the flagella-biofilm transcriptional regulatory network through the control of key elements identified using the network analysis developed.

3. The capability of BW25113 to perform motility and biofilm prove that, while CRP, IHF and Fis are important modulators of the network, these regulators are dispensable to the program development. Altogether, we prove that IHF and Fis are important modulators of motility, adherence and mature biofilm programs while CRP exert a differential effect in bacteria.

4. *E. coli* displayed better performance in early and mature biofilm programs at 37˚C and suggest a special attention in this topic to further research. In spite of this, as similarities were present at mature biofilm between 30 and 37˚C we suggest that a “common biofilm characteristics” must be shared in mature biofilms as a consequence of a conserved talkative population bacterial program.

5. Removal of nodes as CRP, IHF, Fis, and even combinatory depletions are tolerated by the flagella-biofilm transcriptional regulatory network, demonstrating the intrinsic robustness of this system. This demonstrate the necessity to perform a deeper studies of the network players to better understanding of this phenomenon.

VII. REFERENCES

Albert, R. (2005). Scale-free networks in cell biology. J Cell Sci *118*, 4947-4957.

Albert, R., Jeong, H., and Barabasi, A.L. (2000). Error and attack tolerance of complex networks. Nature *406*, 378-382.

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.

Anwar, N., Rouf, S.F., Römling, U., and Rhen, M. (2014). Modulation of biofilm-formation in Salmonella enterica serovar Typhimurium by the periplasmic DsbA/DsbB oxidoreductase system requires the GGDEF-EAL domain protein STM3615. PLoS One *9*, e106095.

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.

Barabási, A.-L. (2003). *Linked: The New Science of Networks*. In (American Journal of Physics 7**1**, 409).

Barabási, A.L. (2009). Scale-free networks: a decade and beyond. Science *325*, 412-413.

Barabási, A.L., and Oltvai, Z.N. (2004). Network biology: understanding the cell's functional organization. Nat Rev Genet *5*, 101-113.

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.

Beloin, C., Roux, A., and Ghigo, J.M. (2008). Escherichia coli biofilms. Curr Top Microbiol Immunol *322*, 249-289.

Beloin, C., Valle, J., Latour-Lambert, P., Faure, P., Kzreminski, M., Balestrino, D., Haagensen, J.A., Molin, S., Prensier, G., Arbeille, B.*, et al.* (2004). Global impact of mature biofilm lifestyle on Escherichia coli K-12 gene expression. Mol Microbiol *51*, 659-674.

Berthoumieux, S., de Jong, H., Baptist, G., Pinel, C., Ranquet, C., Ropers, D., and Geiselmann, J. (2013). Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. Mol Syst Biol *9*, 634.

Bollobás, B. (1979). Graph Theory An Introductory Course, Vol 63 (Springer-Verlag New York Inc.: Springer-Verlag New York).

Brandstetter, K.A., Jurcisek, J.A., Goodman, S.D., Bakaletz, L.O., and Das, S. (2013). Antibodies directed against integration host factor mediate biofilm clearance from Nasopore. Laryngoscope *123*, 2626-2632.

Bremer, H., and Dennis, P.P. (2008). Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. EcoSal Plus *3*.

Brockson, M.E., Novotny, L.A., Mokrzan, E.M., Malhotra, S., Jurcisek, J.A., Akbar, R., Devaraj, A., Goodman, S.D., and Bakaletz, L.O. (2014). Evaluation of the kinetics and mechanism of action of anti-integration host factor-mediated disruption of bacterial biofilms. Mol Microbiol *93*, 1246-1258.

Brophy, J.A., and Voigt, C.A. (2014). Principles of genetic circuit design. Nat Methods *11*, 508-520.

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

Browning, D.F., and Busby, S.J. (2016). Local and global regulation of transcription initiation in bacteria. Nat Rev Microbiol *14*, 638-650.

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.

Cameron, D.E., Bashor, C.J., and Collins, J.J. (2014). A brief history of synthetic biology. Nat Rev Microbiol *12*, 381-390.

Center for Disease Dynamics, E.P. (2015). State of the World's Antibiotics, 2015. CDDEP. CENTER FOR DISEASE DYNAMICS, ECONOMICS & POLICY : Washington, DC.

Chen, Y.Y., Galloway, K.E., and Smolke, C.D. (2012). Synthetic biology: advancing biological frontiers by building synthetic systems. Genome Biol *13*, 240.

Chevance, F.F., and Hughes, K.T. (2008). Coordinating assembly of a bacterial macromolecular machine. Nat Rev Microbiol *6*, 455-465.

Chirwa, N.T., and Herrington, M.B. (2003). CsgD, a regulator of curli and cellulose synthesis, also regulates serine hydroxymethyltransferase synthesis in Escherichia coli K-12. Microbiology *149*, 525-535.

Chua, G., Morris, Q.D., Sopko, R., Robinson, M.D., Ryan, O., Chan, E.T., Frey, B.J., Andrews, B.J., Boone, C., and Hughes, T.R. (2006). Identifying transcription factor functions and targets by phenotypic activation. Proc Natl Acad Sci U S A *103*, 12045-12050.

Claessen, D., Rozen, D.E., Kuipers, O.P., Søgaard-Andersen, L., and van Wezel, G.P. (2014). Bacterial solutions to multicellularity: a tale of biofilms, filaments and fruiting bodies. Nat Rev Microbiol *12*, 115-124.

Clauset, A., Moore, C., and Newman, M.E. (2008). Hierarchical structure and the prediction of missing links in networks. Nature *453*, 98-101.

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B.*, et al.* (2007). Integration of biological networks and gene expression data using Cytoscape. Nat Protoc *2*, 2366-2382.

Collet, A., Cosette, P., Beloin, C., Ghigo, J.M., Rihouey, C., Lerouge, P., Junter, G.A., and Jouenne, T. (2008). Impact of rpoS deletion on the proteome of Escherichia coli grown planktonically and as biofilm. J Proteome Res *7*, 4659-4669.

Cosentino Lagomarsino, M., Jona, P., Bassetti, B., and Isambert, H. (2007). Hierarchy and feedback in the evolution of the Escherichia coli transcription network. Proc Natl Acad Sci U S A *104*, 5516-5520.

Da Re, S., and Ghigo, J.M. (2006). A CsgD-independent pathway for cellulose production and biofilm formation in Escherichia coli. J Bacteriol *188*, 3073-3087.

Davey, M.E., Caiazza, N.C., and O'Toole, G.A. (2003). Rhamnolipid surfactant production affects biofilm architecture in Pseudomonas aeruginosa PAO1. J Bacteriol *185*, 1027-1036.

DePas, W.H., Syed, A.K., Sifuentes, M., Lee, J.S., Warshaw, D., Saggar, V., Csankovszki, G., Boles, B.R., and Chapman, M.R. (2014). Biofilm formation protects Escherichia coli against killing by Caenorhabditis elegans and Myxococcus xanthus. Appl Environ Microbiol *80*, 7079-7087.

Dillon, S.C., and Dorman, C.J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. Nat Rev Microbiol *8*, 185-195.

Du, H., Sheng, X., Zhang, H., Zou, X., Ni, B., Xu, S., Zhu, X., Xu, H., and Huang, X. (2011a). RpoE may promote flagellar gene expression in Salmonella enterica serovar typhi under hyperosmotic stress. Curr Microbiol *62*, 492-500.

Du, H., Wang, M., Luo, Z., Ni, B., Wang, F., Meng, Y., Xu, S., and Huang, X. (2011b). Coregulation of gene expression by sigma factors RpoE and RpoS in Salmonella enterica serovar Typhi during hyperosmotic stress. Curr Microbiol *62*, 1483-1489.

Dudin, O., Geiselmann, J., Ogasawara, H., Ishihama, A., and Lacour, S. (2014). Repression of flagellar genes in exponential phase by CsgD and CpxR, two crucial modulators of Escherichia coli biofilm formation. J Bacteriol *196*, 707-715.

Eisenstein, B.I., Sweet, D.S., Vaughn, V., and Friedman, D.I. (1987). Integration host factor is required for the DNA inversion that controls phase variation in Escherichia coli. Proc Natl Acad Sci U S A *84*, 6506-6510.

Finkel, S.E., and Johnson, R.C. (1992). The Fis protein: it's not just for DNA inversion anymore. Mol Microbiol *6*, 3257-3265.

Freundlich, M., Ramani, N., Mathew, E., Sirko, A., and Tsui, P. (1992). The role of integration host factor in gene expression in Escherichia coli. Mol Microbiol *6*, 2557-2563.

Gao, J., Tao, J., Liang, W., and Jiang, Z. (2016). Cyclic (di)nucleotides: the common language shared by microbe and host. Curr Opin Microbiol *30*, 79-87.

Gerosa, L., Kochanowski, K., Heinemann, M., and Sauer, U. (2013). Dissecting specific and global transcriptional regulation of bacterial gene expression. Mol Syst Biol *9*, 658.

Gibson, J.L., Lombardo, M.J., Thornton, P.C., Hu, K.H., Galhardo, R.S., Beadle, B., Habib, A., Magner, D.B., Frost, L.S., Herman, C.*, et al.* (2010). The sigma(E) stress response is required for stress-induced mutation and amplification in Escherichia coli. Mol Microbiol *77*, 415-430.

Goh, K.I., Oh, E., Jeong, H., Kahng, B., and Kim, D. (2002). Classification of scale-free networks. Proc Natl Acad Sci U S A *99*, 12583-12588.

González-Gil, G., Kahmann, R., and Muskhelishvili, G. (1998). Regulation of crp transcription by oscillation between distinct nucleoprotein complexes. EMBO J *17*, 2877-2885.

Görg C. (2005). Dynamic Graph Drawing of Sequences of Orthogonal and Hierarchical Graphs, P.M. Birke P., Diehl S., ed. (Pach J. (eds) Graph Drawing. GD 2004. Lecture Notes in Computer Science, vol 3383. Springer, Berlin, Heidelberg).

Görke, B., and Stülke, J. (2008). Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. Nat Rev Microbiol *6*, 613-624.

Gottesman, S. (1984). Bacterial regulation: global regulatory networks. Annu Rev Genet *18*, 415-441.

Gualdi, L., Tagliabue, L., Bertagnoli, S., Ieranò, T., De Castro, C., and Landini, P. (2008). Cellulose modulates biofilm formation by counteracting curli-mediated colonization of solid surfaces in Escherichia coli. Microbiology *154*, 2017-2024.

Gualdi, L., Tagliabue, L., and Landini, P. (2007). Biofilm formation-gene expression relay system in Escherichia coli: modulation of sigmaS-dependent gene expression by the CsgD regulatory protein via sigmaS protein stabilization. J Bacteriol *189*, 8034-8043.

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.

Guttenplan, S.B., and Kearns, D.B. (2013). Regulation of flagellar motility during biofilm formation. FEMS Microbiol Rev *37*, 849-871.

Hahn, M.W., and Kern, A.D. (2005). Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol *22*, 803-806.

Hall-Stoodley, L., Costerton, J.W., and Stoodley, P. (2004). Bacterial biofilms: from the natural environment to infectious diseases. Nat Rev Microbiol *2*, 95-108.

Hallinan, J.S., James, K., and Wipat, A. (2011). Network approaches to the functional analysis of microbial proteins. Adv Microb Physiol *59*, 101-133.

Hengge, R. (2009). Proteolysis of sigmaS (RpoS) and the general stress response in Escherichia coli. Res Microbiol *160*, 667-676.

Hengge, R. (2011). Stationary-Phase Gene Regulation in Escherichia coli §. EcoSal Plus *4*.

Hengge, R. (2016). Trigger phosphodiesterases as a novel class of c-di-GMP effector proteins. Philos Trans R Soc Lond B Biol Sci *371*.

Hengge-Aronis, R. (2002). Recent insights into the general stress response regulatory network in Escherichia coli. J Mol Microbiol Biotechnol *4*, 341-346.

Hosseinkhan, N., Zarrineh, P., Rokni-Zadeh, H., Ashouri, M.R., and Masoudi-Nejad, A. (2015). Co-expressional conservation in virulence and stress related genes of three Gammaproteobacterial species: Escherichia coli, Salmonella enterica and Pseudomonas aeruginosa. Mol Biosyst *11*, 3137-3148.

Huang, J., Chen, S., Huang, K., Yang, L., Wu, B., and Peng, D. (2015). [Identification of rpoE gene associated with biofilm formation of Salmonella pullorum]. Wei Sheng Wu Xue Bao *55*, 156-163.

Huerta, A.M., Salgado, H., Thieffry, D., and Collado-Vides, J. (1998). RegulonDB: a database on transcriptional regulation in Escherichia coli. Nucleic Acids Res *26*, 55-59.

Hufnagel, D.A., DePas, W.H., and Chapman, M.R. (2014). The disulfide bonding system suppresses CsgD-independent cellulose production in Escherichia coli. J Bacteriol *196*, 3690-3699.

Hufnagel, D.A., Depas, W.H., and Chapman, M.R. (2015). The Biology of the Escherichia coli Extracellular Matrix. Microbiol Spectr *3*.

Hufnagel, D.A., Evans, M.L., Greene, S.E., Pinkner, J.S., Hultgren, S.J., and Chapman, M.R. (2016). The Catabolite Repressor Protein-Cyclic AMP Complex Regulates csgD and Biofilm Formation in Uropathogenic Escherichia coli. J Bacteriol *198*, 3329-3334.

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.

Ito, A., May, T., Taniuchi, A., Kawata, K., and Okabe, S. (2009). Localized expression profiles of rpoS in Escherichia coli biofilms. Biotechnol Bioeng *103*, 975-983.

Jakovleva, J., Teppo, A., Velts, A., Saumaa, S., Moor, H., Kivisaar, M., and Teras, R. (2012). Fis regulates the competitiveness of Pseudomonas putida on barley roots by inducing biofilm formation. Microbiology *158*, 708-720.

Jeong, H., Mason, S.P., Barabási, A.L., and Oltvai, Z.N. (2001). Lethality and centrality in protein networks. Nature *411*, 41-42.

Jofré, M.R., Rodríguez, L.M., Villagra, N.A., Hidalgo, A.A., Mora, G.C., and Fuentes, J.A. (2014). RpoS integrates CRP, Fis, and PhoP signaling pathways to control Salmonella Typhi hlyE expression. BMC Microbiol *14*, 139.

Kalir, S., McClure, J., Pabbaraju, K., Southward, C., Ronen, M., Leibler, S., Surette, M.G., and Alon, U. (2001). Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. Science *292*, 2080-2083.

Kaufmann, M., and **Wagner,** D. (2001). Drawing Graphs (Lecture Notes in Computer Science).

Keller, E.F. (2005). Revisiting "scale-free" networks. Bioessays *27*, 1060-1068.

Kim, J., and Park, W. (2015). Indole: a signaling molecule or a mere metabolic byproduct that alters bacterial physiology at a high concentration? J Microbiol *53*, 421-428.

Klauck, E., Typas, A., and Hengge, R. (2007). The sigmaS subunit of RNA polymerase as a signal integrator and network master regulator in the general stress response in Escherichia coli. Sci Prog *90*, 103-127.

Klumpp, S., and Hwa, T. (2008). Growth-rate-dependent partitioning of RNA polymerases in bacteria. Proc Natl Acad Sci U S A *105*, 20245-20250.

Klumpp, S., and Hwa, T. (2014). Bacterial growth: global effects on gene expression, growth feedback and proteome partition. Curr Opin Biotechnol *28*, 96-102.

Klumpp, S., Zhang, Z., and Hwa, T. (2009). Growth rate-dependent global effects on gene expression in bacteria. Cell *139*, 1366-1375.

Kobourov, S.G. (2012). Spring Embedders and Force Directed Graph Drawing Algorithms (University of Arizona).

Konstantinidis, K.T., and Tiedje, J.M. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. Proc Natl Acad Sci U S A *101*, 3160-3165.

Kutsukake, K., Ohya, Y., and Iino, T. (1990). Transcriptional analysis of the flagellar regulon of Salmonella typhimurium. J Bacteriol *172*, 741-747.

Land, M., Hauser, L., Jun, S.R., Nookaew, I., Leuze, M.R., Ahn, T.H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T.*, et al.* (2015). Insights from 20 years of bacterial genome sequencing. Funct Integr Genomics *15*, 141-161.

Laverty, G., Gorman, S.P., and Gilmore, B.F. (2014). Biomolecular Mechanisms of Pseudomonas aeruginosa and Escherichia coli Biofilm Formation. Pathogens *3*, 596-632.

Lee, D.J., Minchin, S.D., and Busby, S.J. (2012). Activating transcription in bacteria. Annu Rev Microbiol *66*, 125-152.

Levin, B.R., Perrot, V., and Walker, N. (2000). Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. Genetics *154*, 985-997.

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.

Liu, X., Wu, Y., Chen, Y., Xu, F., Halliday, N., Gao, K., Chan, K.G., and Cámara, M. (2015). RpoS differentially affects the general stress response and biofilm formation in the endophytic Serratia plymuthica G3. Res Microbiol.

Lovén, J., Orlando, D.A., Sigova, A.A., Lin, C.Y., Rahl, P.B., Burge, C.B., Levens, D.L., Lee, T.I., and Young, R.A. (2012). Revisiting global gene expression analysis. Cell *151*, 476-482.

Lu, T.K., Khalil, A.S., and Collins, J.J. (2009). Next-generation synthetic gene networks. Nat Biotechnol *27*, 1139-1150.

Ma, Q., and Wood, T.K. (2009). OmpA influences Escherichia coli biofilm formation by repressing cellulose production through the CpxRA two-component system. Environ Microbiol *11*, 2735-2746.

MacLean, R.C., and Vogwill, T. (2014). Limits to compensatory adaptation and the persistence of antibiotic resistance in pathogenic bacteria. Evol Med Public Health *2015*, 4-12.

Mangan, M.W., Lucchini, S., Danino, V., Cróinín, T.O., Hinton, J.C., and Dorman, C.J. (2006). The integration host factor (IHF) integrates stationary-phase and virulence gene expression in Salmonella enterica serovar Typhimurium. Mol Microbiol *59*, 1831-1847.

Martínez-Antonio, A., and Collado-Vides, J. (2003). Identifying global regulators in transcriptional regulatory networks in bacteria. Curr Opin Microbiol *6*, 482-489.

Martínez-Antonio, A., Janga, S.C., and Thieffry, D. (2008). Functional organisation of Escherichia coli transcriptional regulatory network. J Mol Biol *381*, 238-247.

Martínez-Santos, V.I., Medrano-López, A., Saldaña, Z., Girón, J.A., and Puente, J.L. (2012). Transcriptional regulation of the ecp operon by EcpR, IHF, and H-NS in attaching and effacing Escherichia coli. J Bacteriol *194*, 5020-5033.

Mason, O., and Verwoerd, M. (2007). Graph theory and networks in Biology. IET Syst Biol *1*, 89-119.

Mauter, M.S., Mauter, M., Fait, A., Elimelech, M., and Herzberg, M. (2013). Surface cell density effects on Escherichia coli gene expression during cell attachment. Environ Sci Technol *47*, 6223-6230.

McDermott, J.E., Taylor, R.C., Yoon, H., and Heffron, F. (2009). Bottlenecks and hubs in inferred networks are important for virulence in Salmonella typhimurium. J Comput Biol *16*, 169-180.

McLeod, S.M., and Johnson, R.C. (2001). Control of transcription by nucleoid proteins. Curr Opin Microbiol *4*, 152-159.

Mika, F., and Hengge, R. (2005). A two-component phosphotransfer network involving ArcB, ArcA, and RssB coordinates synthesis and proteolysis of sigmaS (RpoS) in E. coli. Genes Dev *19*, 2770-2781.

Mika, F., and Hengge, R. (2013). Small Regulatory RNAs in the Control of Motility and Biofilm Formation in E. coli and Salmonella. Int J Mol Sci *14*, 4560-4579.

Mika, F., and Hengge, R. (2014). Small RNAs in the control of RpoS, CsgD, and biofilm architecture of Escherichia coli. RNA Biol *11*, 494-507.

Monds, R.D., and O'Toole, G.A. (2009). The developmental model of microbial biofilms: ten years of a paradigm up for review. Trends Microbiol *17*, 73-87.

Moor, H., Teppo, A., Lahesaare, A., Kivisaar, M., and Teras, R. (2014). Fis overexpression enhances Pseudomonas putida biofilm formation by regulating the ratio of LapA and LapF. Microbiology *160*, 2681-2693.

Nasser, W., Schneider, R., Travers, A., and Muskhelishvili, G. (2001). CRP modulates fis transcription by alternate formation of activating and repressing nucleoprotein complexes. J Biol Chem *276*, 17878-17886.

Newman, M.E.J. (2003). The structure and function of complex networks, Vol 45 (2), (Society for Industrial and Applied Mathematics).

Noor, R. (2015). Mechanism to control the cell lysis and the cell survival strategy in stationary phase under heat stress. Springerplus *4*, 599.

O'Toole, G.A. (2011). Microtiter dish biofilm formation assay. J Vis Exp.

Ogasawara, H., Yamada, K., Kori, A., Yamamoto, K., and Ishihama, A. (2010). Regulation of the Escherichia coli csgD promoter: interplay between five transcription factors. Microbiology *156*, 2470-2483.

Ogasawara, H., Yamamoto, K., and Ishihama, A. (2011). Role of the biofilm master regulator CsgD in cross-regulation between biofilm formation and flagellar synthesis. J Bacteriol *193*, 2587-2597.

Österberg, S., del Peso-Santos, T., and Shingler, V. (2011). Regulation of alternative sigma factor use. Annu Rev Microbiol *65*, 37-55.

Pereira, C.S., Thompson, J.A., and Xavier, K.B. (2013). AI-2-mediated signalling in bacteria. FEMS Microbiol Rev *37*, 156-181.

Pesavento, C., Becker, G., Sommerfeldt, N., Possling, A., Tschowri, N., Mehlis, A., and Hengge, R. (2008). Inverse regulatory coordination of motility and curli-mediated adhesion in Escherichia coli. Genes Dev *22*, 2434-2446.

Petrova, O.E., and Sauer, K. (2016). Escaping the biofilm in more than one way: desorption, detachment or dispersion. Curr Opin Microbiol *30*, 67-78.

Price, D.J.d.S. (1965). Networks of Scientific Papers. The pattern of bibliographic references indicates the nature of the scientific research front. (Science), pp. 510-515.

Prigent-Combaret, C., Zghidi-Abouzid, O., Effantin, G., Lejeune, P., Reverchon, S., and Nasser, W. (2012). The nucleoid-associated protein Fis directly modulates the synthesis of cellulose, an essential component of pellicle-biofilms in the phytopathogenic bacterium Dickeya dadantii. Mol Microbiol *86*, 172-186.

Rapoport, A. (1957). Contribution to the theory of random and biased nets (The bulletin of mathematical biophysics: The bulletin of mathematical biophysics), pp. 257–277.

Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., and Barabási, A.L. (2002). Hierarchical organization of modularity in metabolic networks. Science *297*, 1551-1555.

Rényi, P.E.a.A. (1960). ON THE EVOLUTION OF RANDOM GRAPHS (Hungarian Academy of Sciences: Publications of the Mathematical Institute of the Hungarian Academy of Sciences,).

Römling, U. (2005). Characterization of the rdar morphotype, a multicellular behaviour in Enterobacteriaceae. Cell Mol Life Sci *62*, 1234-1246.

Ronen, M., Rosenberg, R., Shraiman, B.I., and Alon, U. (2002). Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. Proc Natl Acad Sci U S A *99*, 10555-10560.

Saldaña, Z., Xicohtencatl-Cortes, J., Avelino, F., Phillips, A.D., Kaper, J.B., Puente, J.L., and Girón, J.A. (2009). Synergistic role of curli and cellulose in cell adherence and biofilm formation of attaching and effacing Escherichia coli and identification of Fis as a negative regulator of curli. Environ Microbiol *11*, 992-1006.

Samanta, P., Clark, E.R., Knutson, K., Horne, S.M., and Prüß, B.M. (2013). OmpR and RcsB abolish temporal and spatial changes in expression of flhD in Escherichia coli biofilm. BMC Microbiol *13*, 182.

Sambrook, J. (1989). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory).

Sanchez-Torres, V., Hu, H., and Wood, T.K. (2011). GGDEF proteins YeaI, YedQ, and YfiN reduce early biofilm formation and swimming motility in Escherichia coli. Appl Microbiol Biotechnol *90*, 651-658.

SCHAECHTER, M., MAALOE, O., and KJELDGAARD, N.O. (1958). Dependency on medium and temperature of cell size and chemical composition during balanced grown of Salmonella typhimurium. J Gen Microbiol *19*, 592-606.

Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and gene expression: origins and consequences. Science *330*, 1099-1102.

Scott, M., and Hwa, T. (2011). Bacterial growth laws and their applications. Curr Opin Biotechnol *22*, 559-565.

Serra, D.O., Richter, A.M., Klauck, G., Mika, F., and Hengge, R. (2013). Microanatomy at cellular resolution and spatial order of physiological differentiation in a bacterial biofilm. MBio *4*, e00103-00113.

Seshasayee, A.S., Sivaraman, K., and Luscombe, N.M. (2011). An overview of prokaryotic transcription factors : a summary of function and occurrence in bacterial genomes. Subcell Biochem *52*, 7-23.

Shapiro, J.A. (1988). Bacteria as Multicellular Organisms Scientific American Journal, 82-89.

Shapiro, J.A. (1998). Thinking about bacterial populations as multicellular organisms. Annu Rev Microbiol *52*, 81-104.

Sharma, V.K., and Bearson, B.L. (2013). Hha controls Escherichia coli O157:H7 biofilm formation by differential regulation of global transcriptional regulators FlhDC and CsgD. Appl Environ Microbiol *79*, 2384-2396.

Sheikh, J., Hicks, S., Dall'Agnol, M., Phillips, A.D., and Nataro, J.P. (2001). Roles for Fis and YafK in biofilm formation by enteroaggregative Escherichia coli. Mol Microbiol *41*, 983-997.

Shih, P.M. (2015). Photosynthesis and early Earth. Curr Biol *25*, R855-859.

Shimada, K., Ogasawara, H., Yamada, K., Shimura, M., Kori, A., Shimada, T., Yamanaka, Y., Yamamoto, K., and Ishihama, A. (2013). Screening of promoter-specific transcription factors: multiple regulators for the sdiA gene involved in cell division control and quorum sensing. Microbiology *159*, 2501-2512.

Shimada, T., Shimada, K., Matsui, M., Kitai, Y., Igarashi, J., Suga, H., and Ishihama, A. (2014). Roles of cell division control factor SdiA: recognition of quorum sensing signals and modulation of transcription regulation targets. Genes Cells *19*, 405-418.

Shimizu, K. (2016). Metabolic Regulation and Coordination of the Metabolism in Bacteria in Response to a Variety of Growth Conditions. Adv Biochem Eng Biotechnol *155*, 1-54.

Shin, S., and Park, C. (1995). Modulation of flagellar expression in Escherichia coli by acetyl phosphate and the osmoregulator OmpR. J Bacteriol *177*, 4696-4702.

Silva-Rocha, R., Chavarría, M., Kleijn, R.J., Sauer, U., and de Lorenzo, V. (2013). The IHF regulon of exponentially growing Pseudomonas putida cells. Environ Microbiol *15*, 49-63.

Solano, C., Echeverz, M., and Lasa, I. (2014). Biofilm dispersion and quorum sensing. Curr Opin Microbiol *18*, 96-104.

Sommerfeldt, N., Possling, A., Becker, G., Pesavento, C., Tschowri, N., and Hengge, R. (2009). Gene expression patterns and differential input into curli fimbriae regulation of all GGDEF/EAL domain proteins in Escherichia coli. Microbiology *155*, 1318-1331.

Sperandio, V., Torres, A.G., and Kaper, J.B. (2002). Quorum sensing Escherichia coli regulators B and C (QseBC): a novel two-component regulatory system involved in the regulation of flagella and motility by quorum sensing in E. coli. Mol Microbiol *43*, 809-821.

Srivastava, S., and Bhargava, A. (2016). Biofilms and human health. Biotechnol Lett *38*, 1-22.

Stefan, D., Pinel, C., Pinhal, S., Cinquemani, E., Geiselmann, J., and de Jong, H. (2015). Inference of quantitative models of bacterial promoters from time-series reporter gene data. PLoS Comput Biol *11*, e1004028.

Stoodley, P., Sauer, K., Davies, D.G., and Costerton, J.W. (2002). Biofilms as complex differentiated communities. Annu Rev Microbiol *56*, 187-209.

Strogatz, S.H. (2001). Exploring complex networks. Nature *410*, 268-276.

Suh, S.J., Silo-Suh, L., Woods, D.E., Hassett, D.J., West, S.E., and Ohman, D.E. (1999). Effect of rpoS mutation on the stress response and expression of virulence factors in Pseudomonas aeruginosa. J Bacteriol *181*, 3890-3897.

Thieffry, D., Huerta, A.M., Pérez-Rueda, E., and Collado-Vides, J. (1998). From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in Escherichia coli. Bioessays *20*, 433-440.

Tomoyasu, T., Ohkishi, T., Ukyo, Y., Tokumitsu, A., Takaya, A., Suzuki, M., Sekiya, K., Matsui, H., Kutsukake, K., and Yamamoto, T. (2002). The ClpXP ATP-dependent protease regulates flagellum synthesis in Salmonella enterica serovar typhimurium. J Bacteriol *184*, 645-653.

Torres-Escobar, A., Juárez-Rodríguez, M.D., and Demuth, D.R. (2014). Differential transcriptional regulation of Aggregatibacter actinomycetemcomitans lsrACDBFG and lsrRK operons by integration host factor protein. J Bacteriol *196*, 1597-1607.

Travers, A. (1997). DNA-protein interactions: IHF--the master bender. Curr Biol *7*, R252-254.

Treviño-Quintanilla, L.G., Freyre-González, J.A., and Martínez-Flores, I. (2013). Anti-Sigma Factors in E. coli: Common Regulatory Mechanisms Controlling Sigma Factors Availability. Curr Genomics *14*, 378-387.

U. S. Department of Health and Human Services. Centers for Disease, C.a.P. (2013). Antibiotic Resistance Threats in the United States, 2013

Valentini, M., and Filloux, A. (2016). Biofilms and Cyclic di-GMP (c-di-GMP) Signaling: Lessons from Pseudomonas aeruginosa and Other Bacteria. J Biol Chem *291*, 12547-12555.

Venturi, V. (2006). Regulation of quorum sensing in Pseudomonas. FEMS Microbiol Rev *30*, 274-291.

Vijay, D., Dhaka, P., Vergis, J., Negi, M., Mohan, V., Kumar, M., Doijad, S., Poharkar, K., Kumar, A., Malik, S.S.*, et al.* (2015). Characterization and biofilm forming ability of diarrhoeagenic enteroaggregative Escherichia coli isolates recovered from human infants and young animals. Comp Immunol Microbiol Infect Dis *38*, 21-31.

Wang, X., Montero Llopis, P., and Rudner, D.Z. (2013). Organization and segregation of bacterial chromosomes. Nat Rev Genet *14*, 191-203.

Weber, H., Pesavento, C., Possling, A., Tischendorf, G., and Hengge, R. (2006). Cyclic-di-GMP-mediated signalling within the sigma network of Escherichia coli. Mol Microbiol *62*, 1014-1034.

White-Ziegler, C.A., Um, S., Pérez, N.M., Berns, A.L., Malhowski, A.J., and Young, S. (2008). Low temperature (23 degrees C) increases expression of biofilm-, cold-shock- and RpoS-dependent genes in Escherichia coli K-12. Microbiology *154*, 148-166.

Williams, P., Winzer, K., Chan, W.C., and Cámara, M. (2007). Look who's talking: communication and quorum sensing in the bacterial world. Philos Trans R Soc Lond B Biol Sci *362*, 1119-1134.

Willias, S.P., Chauhan, S., Lo, C.C., Chain, P.S., and Motin, V.L. (2015). CRP-Mediated Carbon Catabolite Regulation of Yersinia pestis Biofilm Formation Is Enhanced by the Carbon Storage Regulator Protein, CsrA. PLoS One *10*, e0135481.

Wixon, J., and Kell, D. (2000). The Kyoto encyclopedia of genes and genomes--KEGG. Yeast *17*, 48-55.

Wolf, Y.I., Karev, G., and Koonin, E.V. (2002). Scale-free networks in biology: new insights into the fundamentals of evolution? Bioessays *24*, 105-109.

Xie, X., Zhang, H., Zheng, Y., Li, A., Wang, M., Zhou, H., Zhu, X., Schneider, Z., Chen, L., Kreiswirth, B.N.*, et al.* (2016). RpoE is a Putative Antibiotic Resistance Regulator of Salmonella enteric Serovar Typhi. Curr Microbiol *72*, 457-464.

Yamaguchi, S., Fujita, H., Ishihara, A., Aizawa, S., and Macnab, R.M. (1986). Subdivision of flagellar genes of Salmonella typhimurium into regions responsible for assembly, rotation, and switching. J Bacteriol *166*, 187-193.

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.

Zargar, A., Quan, D.N., Emamian, M., Tsao, C.Y., Wu, H.C., Virgile, C.R., and Bentley, W.E. (2015). Rational design of 'controller cells' to manipulate protein and phenotype expression. Metab Eng *30*, 61-68.

Zhou, Y., Smith, D.R., Hufnagel, D.A., and Chapman, M.R. (2013). Experimental manipulation of the microbial functional amyloid called curli. Methods Mol Biol *966*, 53-75.

VIII. ANEXES

ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

Engineering Synthetic *cis*-Regulatory Elements for Simultaneous Recognition of Three Transcriptional Factors in Bacteria

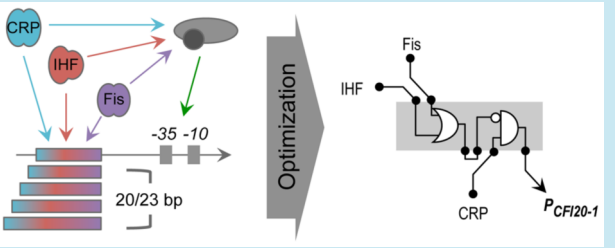
Gerardo Ruiz Amores,[†] María-Eugenia Guazzaroni,[‡] and Rafael Silva-Rocha*[,†]

[†]FMRP, University of São Paulo, Ribeirão Preto, São Paulo 05508-020, Brazil

[‡]FFCLRP, University of São Paulo, Ribeirão Preto, São Paulo 05508-020, Brazil

Supporting Information

ABSTRACT: Recognition of *cis*-regulatory elements by transcription factors (TF) at target promoters is crucial to gene regulation in bacteria. In this process, binding of TFs to their cognate sequences depends on a set of physical interactions between these proteins and specific nucleotides in the operator region. Previously, we showed that *in silico* optimization algorithms are able to generate short sequences that are recognized by two different TFs of *Escherichia coli*, namely, CRP and IHF, thus generating an AND logic gate. Here, we expanded this approach in order to engineer DNA sequences that can be simultaneously recognized by three unrelated TFs (CRP, IHF, and Fis). Using *in silico* optimization and experimental validation strategies, we were able to obtain a candidate promoter (*Plac-CFI1*) regulated by only two TFs with an AND logic, thus demonstrating a limitation in the design. Subsequently, we modified the algorithm to allow the optimization of extended sequences, and were able to design two synthetic promoters ($P_{CFI20\text{-}1}$ and $P_{CFI22\text{-}5}$) that were functional *in vivo*. Expression assays in *E. coli* mutant strains for each TF revealed that while CRP positively regulates the promoter activities, IHF and Fis are strong repressors of both the promoter variants. Taken together, our results demonstrate the potential of *in silico* strategies in bacterial synthetic promoter engineering. Furthermore, the study also shows how small modifications in *cis*-regulatory elements can drastically affect the final logic of the resulting promoter.



KEYWORDS: *synthetic promoters, cis-regulatory elements, synthetic biology*

Cells are capable of sensing environmental signals to modulate, in space and time, their gene expressions.[1] In bacteria, expression of many genes (even across different growth conditions[2]) have been quantitatively correlated with its physiological state.[3,4] These studies show that growth rate parameters [*e.g.*, RNA polymerase (RNAP), ribosomes, and sigma (σ) factor concentrations] can modulate gene expression independent of the promoter identity[3−7] in a process called "global transcriptional regulation".[6] Additionally, the transcriptional response[6] of a specific promoter depends on the precise integration of external and internal signals by TFs, leading to a fine balance between fundamental biological functions such as metabolism, development, and differentiation.[6,8,9] Concordantly, the key step in specific transcription initiation is the recognition of DNA sequences upstream to the cognate genes.[10] The fundamental structure of sigma70-dependent promoters is composed of mainly three functional elements: the upstream (UP) element, −35 box, and −10 box (the latter two being recognized by the holo form of RNAP).[11] However, the core RNAP is unable to locate promoter sequences by itself, making its interaction with the σ subunit, which confers selectivity to promoter recognition,[12,13] necessary. In *Escherichia coli*, there are seven different σ factors that represent the first level of transcriptional regulation.[12]

The second level of regulation is mediated by TFs, which bind to *cis*-regulatory elements upstream to the core promoter.[11] TFs are termed as local regulators when controlling either one or few gene targets, and as global regulators (GRs) when controlling several targets in response to specific environmental and physiological signals.[14] Since promoters of a specific gene contain regulatory elements for multiple TFs, the logic of the regulation depends on whether the TF acts as an activator or a repressor, and usually involves more than one signal.[10,15] Therefore, the intricate regulatory links operating to modulate gene expression in response to available signals give rise to very complex networks, capable of faithfully coordinating the cell's activities.[14,16,17] With increasing knowledge in the molecular mechanisms underlying the regulation of promoter activity, synthetic biology has emerged as a new discipline demanding the engineering of novel regulatory circuits in living cells for a number of biotechnological applications.[18−22]

The classical strategy to engineer new cellular circuits chiefly relies on the mixing of local TFs (*e.g.*, LacI, TetT, and AraC) with a number of different promoters (such as *Plac*, *Ptet*, and *ParaB*) in order to construct synthetic circuits with different behaviors.[23−27] The main problem associated with this strategy

Received: February 2, 2015

DOI: 10.1021/acssynbio.5b00098
ACS Synth. Biol. XXXX, XXX, XXX−XXX

is that increasing the number of regulatory interactions in a circuit leads to high levels of noise propagation in the network, which might ultimately impair its performance.[28−30] In order to circumvent this limitation, a more sophisticated strategy would be to implement higher-order regulatory logic directly into the promoter region. The designed element would integrate the presence/absence of TFs to generate the final output of the circuit.[31−36] This type of "engineered promoter logic" approach tries to mimic naturally occurring systems, since bacterial promoters are usually formed by arrays of *cis*-regulatory elements proximal to the core RNAP-binding region.[31,34,37−39] Several attempts have been made to construct synthetic promoters by reshuffling *cis*-regulatory elements with core promoters, usually through the construction and characterization of random promoter libraries *in vivo*.[32,35,40] Recently, we proposed a novel promoter engineering strategy based on *in silico* optimization algorithms to construct unique *cis*-regulatory elements capable of being recognized by two independent TFs (*e.g.*, CRP and IHF).[31] The algorithm used takes position weight matrixes (PWMs) for the TFs of interest and tries to optimize short DNA sequences through a series of mutation/selection steps. The major goal of such an approach is to "compress" regulatory information into short functional DNA sequences that can then be used in arrays to construct synthetic promoters endowed with more sophisticated logic behaviors. Here, we used this approach to engineer and test synthetic *cis*-regulatory elements that can be recognized by three different TFs, namely CRP, IHF, and Fis, which act as activators. At first, simultaneous optimization of three TF-binding sites imposed a remarkable challenge since the synthetic promoter obtained was not efficiently regulated by all the TFs selected. In order to overcome this limitation, we allowed the algorithm to optimize for increasing sequence lengths (from 20 to 23 bp) to alleviate the constraints imposed when using short (19 bp) sequences. This new approach allowed the identification of two active synthetic promoters that were regulated by the three TFs, of which two (IHF and Fis) were found to act as repressors. In summary, the results presented here highlight both the potential and limitations of *in silico* engineering of synthetic promoters, focusing on the high relevance to engineered complex regulatory logic in living cells.

RESULTS AND DISCUSSION

Engineering Triple Operators Using *In Silico* Optimization. We previously demonstrated that *in silico* optimization algorithms can be used to generate functional dual *cis*-regulatory elements using only well-characterized PWMs as inputs for the TFs of interest. Here, we aimed to construct new-to-nature *cis*-regulatory elements that, in principle, can be recognized by three different TFs (Figure 1A). Toward this, we focused on the global regulators CRP, IHF, and Fis (that have well-characterized binding site motifs[31]) as a start point to construct PWMs (Figure 1B). Using the resulting PWMs, we performed a thousand rounds of optimization and generated a list of candidate sequences for validation. From this list, we selected a DNA sequence (named CFI1) with maximal average score for the three TFs. This candidate sequence was then used to replace the wild type CRP operator of *Plac*, thus generating the *Plac-CFI1* promoter (Figure 1C). Next, this resulting promoter was cloned upstream of the *GFPlva* gene placed in the pMR1 vector.[31] As controls, we used mutated (*Plac-ΔCRP*) and wild type (*Plac-wt*) *Plac* variants along with two synthetic promoters (*Plac-CI1* and *Plac-CI2*) with dual operators (Figure 1C). In order to perform the *in vivo* promoter assay, the resulting plasmids were introduced into the wild type *E. coli* strain BW25113. To assay the promoter activities, overnight growth strains were diluted in fresh LB media and allowed to grow for 4 h (approximately midexponential phase). At this point, samples were collected and analyzed by flow cytometry to quantify the GFP levels in each strain. As shown in Figure 1D, *Plac-CFI1* was functional *in vivo* under these assay conditions, presenting a promoter activity ∼6 times higher than that of *Plac-ΔCRP*, in which the CRP-binding site was eliminated.[31] However, it is notable that the *Plac-CFI1* activity was reduced when compared to those of the other three functional promoters assayed. This activity was one and 2 orders of magnitude lower than *Plac-CI1* and *Plac-CI2*, and the fully functional wild type *Plac* activities, respectively (Figure 1D). Therefore, this result highlights a trade-off between recognition by multiple TFs and the maximal activity reached by the synthetic promoter, where gain in the former causes a reduction in the latter. Taken together, these results show that the *in silico* engineered CFI1 sequence was functional *in vivo* and thus, could be recognized by multiple TFs in *E. coli*.

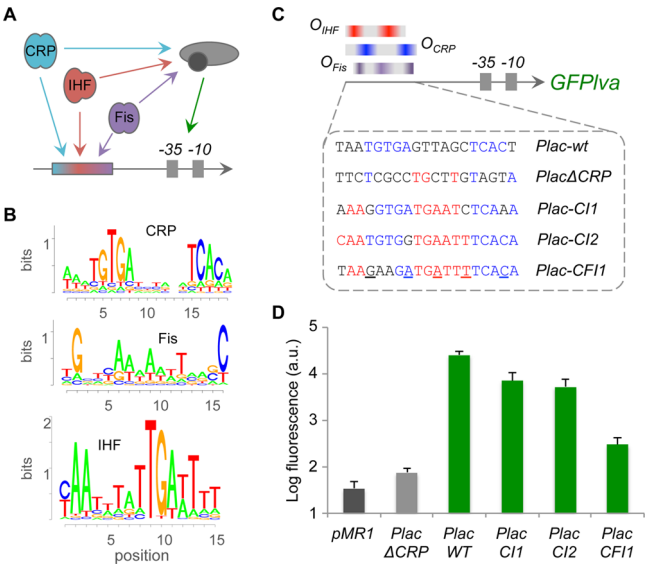


Figure 1. Construction of synthetic promoters. (A) Synthetic promoters were designed to allow the interaction of multiple TFs at the same *cis*-regulatory element that facilitate the promoter expression by recruiting RNAP to the target promoter. (B) Sequence logo representation of weight matrices for CRP, IHF, and Fis binding sequences. (C) Wild type and synthetic *lac* sequences assayed in this study. The *cis*-regulatory elements designed were constructed by optimizing the DNA sequence for the binding sequences for CRP, IHF, and Fis (represented as O_{CRP}, O_{IHF}, and O_{Fis}) located upstream to a sigma70-dependent core promoter. The boxes highlight the sequences of the *cis*-regulatory elements dual CRP/IHF promoters (*Plac-CI1* and *Plac-CI2*) and putative triple CRP/IHF/Fis promoter (*Plac-CFI1*), as well as wild type (*Plac-wt*) and unregulated (*Plac-ΔCRP*) variants. For each sequence, nucleotides highlighted in blue matched the CRP-binding consensus, those in red are close to the IHF consensus while those underlined are related to the binding consensus for Fis. (D) Steady state activity of all promoters used in this study. The empty plasmid pMR1 was used as a reference. Vertical bars represent the standard deviation of three independent experiments.

Validation of *Plac-CFI1* Activity in Mutant Strains of *E. coli*. Upon observing that incorporation of the CFI1 sequence resulted in a functional promoter, we next investigated the roles of the three regulatory proteins in its activity. Toward this, we

DOI: 10.1021/acssynbio.5b00098
ACS Synth. Biol. XXXX, XXX, XXX−XXX

first assayed promoter activities in cells growing in M9 media supplemented with glycerol (control) *versus* those exposed to both glycerol and glucose. For the latter, *E. coli* cells were subjected to *carbon catabolite repression* (CCR) condition, in which the presence of glucose resulted in a drastic reduction of the cAMP levels, which in turn inactivates CRP.[41] Figure 2

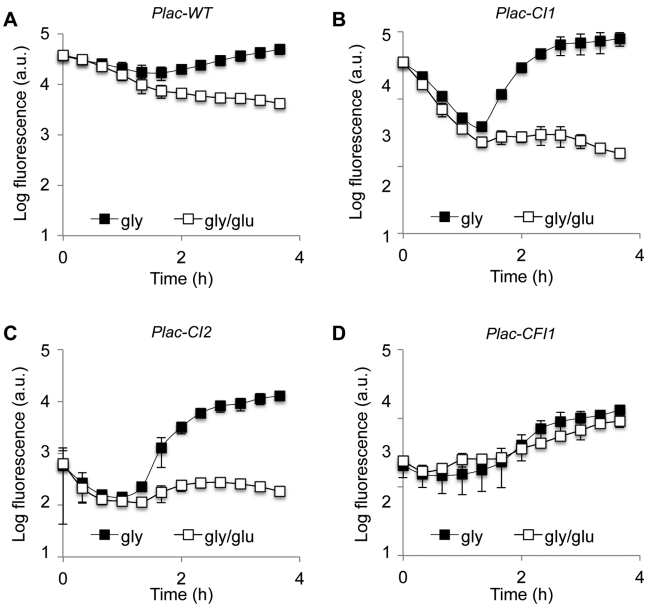


Figure 2. Effect of CCR on the dynamics of target promoter activities. Each strain was grown overnight in M9 media supplemented with 0.2% CAA and 1% glycerol. After pregrowth, cells were diluted in fresh M9 CAA/glycerol media either with (gly/glucose condition) or without (gly condition) 0.4% glucose. At fixed time intervals, samples were collected and analyzed by flow cytometry. (A) *E. coli* BW25113 (pMR1-*Plac-WT*). (B) *E. coli* BW25113 (pMR1-*Plac-CI1*). (C) *E. coli* BW25113 (pMR1-*Plac-CI2*). (D) *E. coli* BW25113 (pMR1-*Plac-CFI1*). Vertical bars represent the standard deviation of three independent experiments.

illustrates that while the activities of *Plac-WT*, *Plac-CI1*, and *Plac-CI2* were dramatically reduced during CCR (Figure 2A–C), that of *Plac-CFI1* was only marginally affected by this condition, indicating that CRP does not play a role in its regulation (Figure 2D). In order to confirm this as well as to investigate the roles of IHF and Fis in this promoter, we assayed the performance of this system in *E. coli* strains each lacking one of the TF of interest (*i.e.*, CRP, IHF, or Fis). For this, the reporter plasmid harboring this promoter was used to transform the three mutant strains of *E. coli* lacking the *crp*, *ihfA*, or *fis* gene.[42] This strategy was used to quantify the specific contribution of each TF to the final promoter activity observed, assuming the absence of any cooperativity (negative or positive) between the TFs. First, the recombinant mutant strains harboring the reporter plasmid were grown overnight in LB media. Next, the cells were diluted in fresh media and samples harvested at time intervals of 20 min to allow quantification of the promoter dynamics over time. As illustrated in Figure 3, *Plac-CFI1* assayed in the wild type *E. coli* strain displayed low activity at the beginning of the growth curve while higher levels of expression were reached at the final time points. When the same promoter was assayed in a strain devoid of the functional *crp* gene, the expression profile was almost invariant with only minor differences at the end of the growth curve. This result

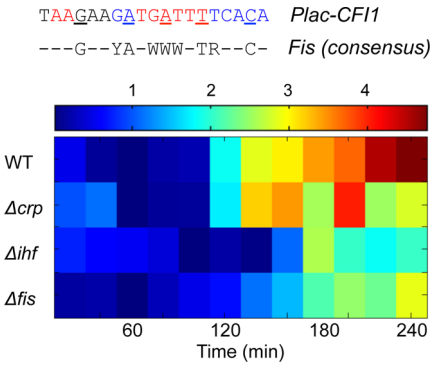


Figure 3. Regulatory effects of CRP, IHF, and Fis on *PlacCFI1* activity. Top, sequence of the *cis*-regulatory element present in the *Plac-CFI1* promoter variant, showing the binding consensus for Fis.[50] Bottom, heat map representing the expression profile of *Plac-CFI1* promoter in wild type (WT) CRP, IHF, and Fis mutant strains. Promoter activities are represented as log2 fold change relative to the minimal levels under the growth curve. Expression levels of the strains were quantified by flow cytometry.

indicates that, at least under the conditions assayed, CRP plays only a minor role, if any, in the final *Plac-CFI1* activity. However, when *Plac-CFI1* was analyzed in strains lacking either the IHF or Fis functional regulator, the promoter activity was strongly impaired during the growth curve (Figure 3). These results demonstrate that only IHF and Fis are efficient regulators of the *Plac-CFI1* activity, while CRP plays only a slight regulatory role. More importantly, these results also show that the compression of three binding sites into a single short DNA sequence imposes a challenge for the *in silico* optimization algorithm, giving rise to only partially functional *cis*-regulatory elements *in vivo*.

Implementation of a Modified Optimization Algorithm to Allow *cis*-Regulatory Engineering. Since the previously presented strategy was unable to generate functional *cis*-elements recognized by three TFs, we decided to implement a variant of the algorithm in order to overcome this limitation. The initial strategy used a 19 bp long DNA sequence for the mutation and selection steps based on the score of the PWM of interest.[31] While this short size ensures that all *cis*-elements engineered are placed in an activation-permissive position relative to the −35/−10 boxes of the promoter, it also imposes a limitation to obtaining a single sequence that has simultaneous high scores for the three PWMs. Therefore, increasing the sequence length would facilitate the generation of sequences that satisfy the PWM requirements. We implemented modifications in the original algorithm to allow optimization of DNA sequences 20, 21, 22, or 23 bp in length (Figure 4A). From each simulation, we selected the top five candidate sequences (20 in total, Table S1) to clone into the pMR1 promoter-probe vector. Subsequently, we were able to clone and test 11 different promoter variants (Table S1). From this assay, we found that only two displayed detectable activity in wild type *E. coli* growing in M9 minimal media. These promoters were named $P_{CFI20\text{-}1}$ and $P_{CFI22\text{-}5}$ since they were 20 and 22 bp in length, respectively. Figure 4B shows the sequences of the synthetic DNA elements, highlighting the conserved positions relative to the consensus sequences for the three TFs of interest.

Accessing the *cis*-Regulatory Logic of $P_{CFI20\text{-}1}$ and $P_{CFI22\text{-}5}$ Promoters *In Vivo*. In order to verify whether the two candidate promoters were indeed regulated by the TFs, we analyzed their activities in mutant strains of *E. coli* growing in M9 minimal media either in the absence or presence of glucose

DOI: 10.1021/acssynbio.5b00098
ACS Synth. Biol. XXXX, XXX, XXX−XXX

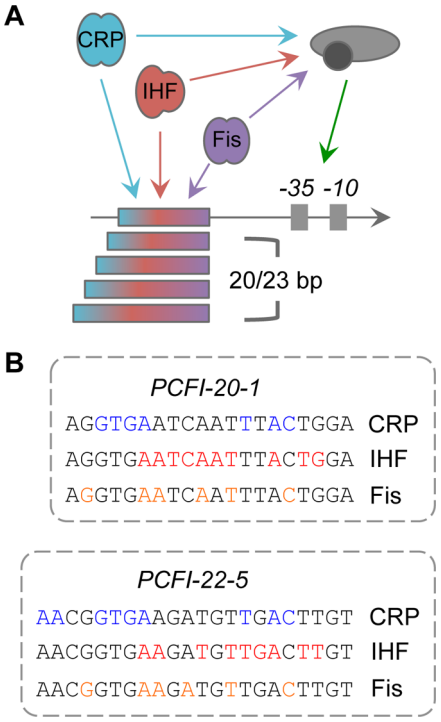


Figure 4. Expanding the optimization algorithm to construct synthetic *cis*-elements. (A) To construct functional *cis*-elements recognized by the three TFs, we implemented modifications in the algorithm to allow the optimization of sequences with 20, 21, 22, or 23 bp. (B) Two functional elements that displayed detectable activity *in vivo* are presented. The nucleotides displaying similarities to the binding consensus of CRP, IHF, and Fis are highlighted in blue, red, and orange, respectively.

(*i.e.*, under CCR). As shown in Figure 5, $P_{CFI20\text{-}1}$ displayed a constant expression in the wild type strain, being partially reduced in the absence of CRP when glucose was present in the media (Figure 5A). Likewise, promoter activity in a strain lacking IHF displayed a similar profile with lower levels, and was also affected by the addition of glucose (Figure 5B). This low promoter activity found in the *ihf*-deficient strain (about 23%) was reproducible and observed only in minimal media (see below). This indicates that the slight reduction observed in promoter activity might be due to the inherent physiology of the mutant strain growing in a poor media. In case of $P_{CFI22\text{-}5}$, promoter activity in the wild type strain was very low at the beginning of the growth curve and increased in the first 2−3 h, after which a decrement was observed (Figure 5C). Further, the promoter activity decayed rapidly after 3 h when cells were exposed to glucose, demonstrating a dependence on CRP at the final stages of growth. Remarkably, when $P_{CFI22\text{-}5}$ was assayed in a strain lacking *ihf*, its activity remained high after 3 h, indicating that IHF plays a repressive role in this promoter. However, in the presence of glucose, GFP levels dropped after 3 h of growth, reinforcing the notion that CRP plays a crucial role in late-promoter activity (Figure 5D). Analysis of promoter activities in a *fis* mutant revealed an aberrant profile (not shown) that might be due to severe impairment of growth in minimal media (Figure S1 in Supporting Information).

In order to examine unequivocally the *cis*-regulatory logic of the $P_{CFI20\text{-}1}$ and $P_{CFI22\text{-}5}$ promoters *in vivo*, wild type and mutant (*ihf* and *fis*) *E. coli* cells harboring the reporter plasmids were assayed in LB media. In order to construct functional double mutants, glucose was added to the media to impair CRP activity. Under cultivation in LB media, the mutant strains displayed more comparable growth profiles (Figure S2 in Supporting Information), allowing us to study the *in vivo* roles of CRP, IHF, and Fis in the two promoters cited. The results of GFP analysis using this setup are presented in Figure 6. In case of $P_{CFI20\text{-}1}$, promoter activity in the wild type strain was repressed after 2 h of cultivation in the presence of glucose, confirming the positive role of CRP in late activity (Figure 6A). Quite remarkably, this promoter activity was significantly enhanced in the strain mutant for *ihf*, indicating a repressive role for this TF in $P_{CFI20\text{-}1}$. Finally, an even more remarkable increase in activity was observed when $P_{CFI20\text{-}1}$ was assayed in a *fis*-minus strain, demonstrating that this TF also functions to repress this system (Figure 6A). In case of $P_{CFI22\text{-}5}$, the same regulatory profile was observed, since removal of either *ihf* or *fis* resulted in a strong increment in its activity (Figure 6B). Similar to $P_{CFI20\text{-}1}$, maximal $P_{CFI22\text{-}5}$ activity was observed in the *fis* mutant in the absence of glucose, where CRP was able to function as an activator. Notably, the two synthetic promoters displayed very different absolute activities, $P_{CFI20\text{-}1}$ being 5−10 times stronger than that of $P_{CFI22\text{-}5}$. This difference in basal promoter activity indicates that the synthetic sequences constructed have different abilities to recruit RNAP to the core promoter, which were not anticipated by the *in silico* approach used. Similarly, effects of the three TFs on the synthetic promoters were also remarkably different. For instance, while removal of IHF increased the activities of both promoters in the later stages of growth, $P_{CFI20\text{-}1}$ activity in the *fis* mutant was significantly high from the beginning of the growth curve (Figure 6). These results demonstrate that despite the same *cis*-regulatory logic being obtained with both promoters, their overall kinetics is highly dependent on the position of the *cis*-regulatory elements used. Altogether, the modifications implemented in the optimization algorithm allowed the construction of functional sequences simultaneously regulated by the three TFs (CRP, IHF, and Fis). However, the final regulatory logic observed was different from that initially anticipated, since only one TF (CRP) acted as activator while the other two (IHF and Fis) were repressors.

CONCLUSIONS

Existing regulatory networks found in living organisms have evolved for millions of years to meet the requirements imposed by varying selective pressures. Rather than being chaotic, assembling of such complex networks must obey some mechanistic rules inherent to the physical constraints of the interacting components, such as proteins and DNA molecules. Deciphering these rules might allow us to mimic those evolutionary processes in the laboratory and construct new-to-nature, application-driven regulatory circuits.[19,20,22] In this study, despite the use of very well characterized TF-binding sites for the optimization of engineered *cis*-regulatory elements, the algorithm formerly used in this work failed to generate a unique binding site (19 bp long) for the three different regulators. This indicates that the system reached a limit for logic compression of *cis*-regulatory sequences into a single DNA segment. This is not surprising since constructing such functional elements requires the simultaneous optimization of three very strict PWMs, and a change in a single nucleotide position could have a tremendous impact on the *in vivo* function of the system.[43−45] Indeed, using a more relaxed version of the algorithm that optimized longer DNA sequences did provide functional promoters that were regulated by the three TFs.

DOI: 10.1021/acssynbio.5b00098
ACS Synth. Biol. XXXX, XXX, XXX−XXX

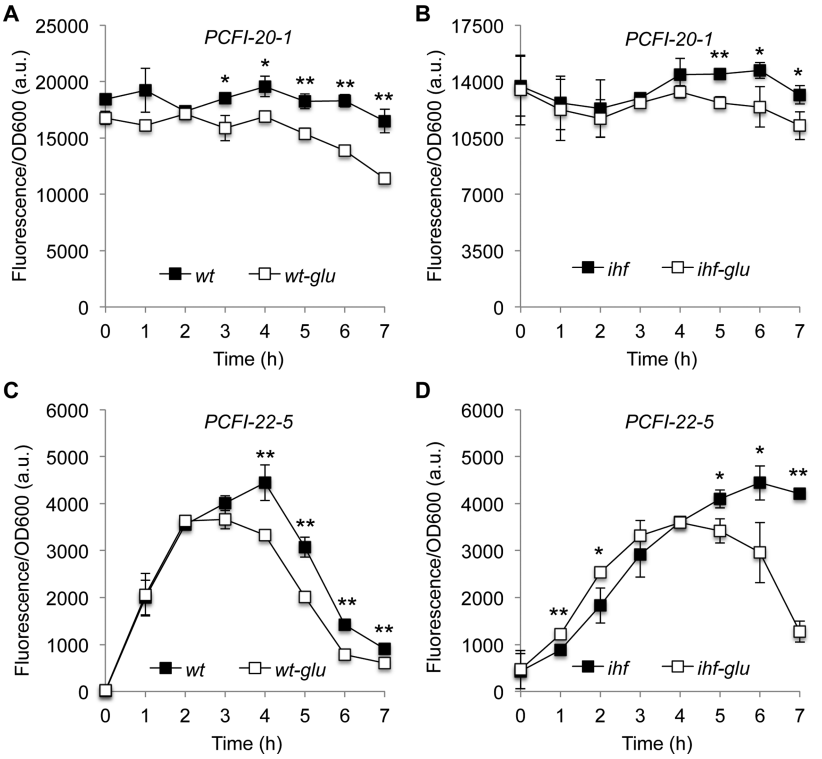


Figure 5. Analysis of $P_{CFI20\text{-}1}$ and $P_{CFI22\text{-}5}$ promoters in wild type and *ihf*-minus mutant of *E. coli* in minimal media. Cells grown overnight were diluted in fresh M9 media either with (empty symbols) or without (full symbols) 0.4% glucose. For each condition, fluorescence levels and OD_{600} were quantified in a Victor X3 plate reader. Promoter activities are expressed as the fluorescence levels of the cultures normalized by the OD_{600} (fluorescence/OD_{600}). (A) *E. coli* BW25113 (pMR1-$P_{CFI20\text{-}1}$). (B) *E. coli ihf* mutant (pMR1-$P_{CFI20\text{-}1}$). (C) *E. coli* BW25113 (pMR1-$P_{CFI22\text{-}5}$). (D) *E. coli ihf* mutant (pMR1-$P_{CFI22\text{-}5}$). Vertical bars represent the standard deviation of three independent experiments (*p < 0.05, **p < 0.01, one-way ANOVA).

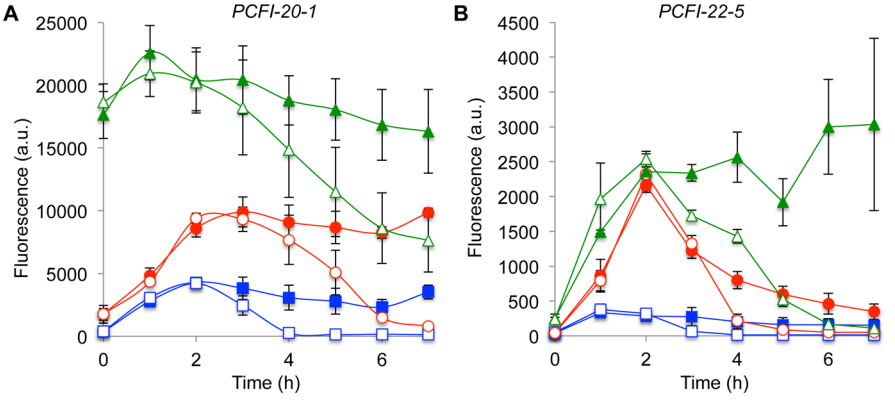


Figure 6. Analysis of $P_{CFI20\text{-}1}$ and $P_{CFI22\text{-}5}$ promoters in wild type, *ihf*, and *fis* mutants of *E. coli* in LB media. Cells grown overnight were diluted in fresh LB media either with (empty symbols) or without (full symbols) 0.4% glucose. Wild type *E. coli* is represented by blue squares, *ihf* mutant by red circles, and *fis* mutant by green triangles. (A) Promoter activity of strains carrying pMR1-$P_{CFI20\text{-}1}$ plasmid. (B) Promoter activity of strains carrying pMR1-$P_{CFI22\text{-}5}$ plasmid. Expression levels of the strains were quantified by flow cytometry. Vertical bars represent the standard deviation of three independent experiments.

However, the final logics obtained were not AND gates as initially desired, but were more sophisticated since two TFs repressed the system (Figure 7). The data presented here clearly demonstrate a trade-off between the ability of multiple TFs to bind to the same *cis*-element and the final logic presented by the system. Increasing the number of TFs regulating the same DNA element requires larger sequences that efficiently accommodate the nucleotide requirements of each TF. However, this relaxation of binding site position relative to the −35/−10 boxes might impair activation, promoting repression of the resulting promoter. While our results have only started to explore the possibilities of *in silico cis*-regulatory engineering, attempts that are more systematic should be made in the future to further understand the effect of *cis*-element composition in the final promoter logic.

METHODS

Bacterial Strains and Growth Conditions. The list of bacterial strains, plasmids, and primers used in this work are listed in Table 1. *E. coli* wild type (BW25113) and derivative

DOI: 10.1021/acssynbio.5b00098
ACS Synth. Biol. XXXX, XXX, XXX−XXX

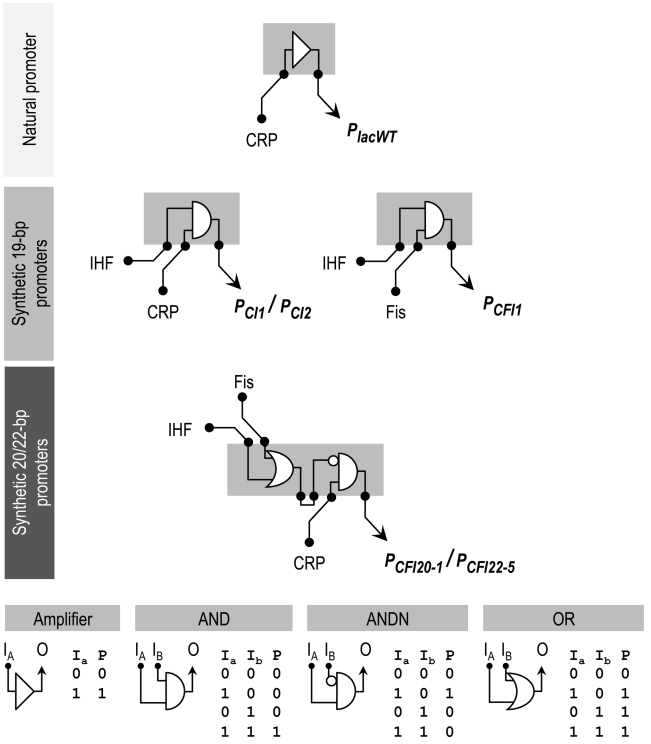


Figure 7. Overall approach to construct synthetic *cis*-regulatory elements and the resulting logic of the systems. Starting from a natural system controlled by a single input (such as the *PlacWT*), the optimization algorithm was able to generate only short (19 bp) sequences efficiently recognized by two TFs (either CRP and IHF, or Fis and IHF). This ensures an AND gate behavior since the location of the binding sites relative to the −35/−10 boxes is preserved. In order to generate functional promoters regulated by three TFs, the algorithm needs to consider longer sequences (in this case, 20 and 22 bp) to efficiently accommodate the three binding sites. However, in this case, this seems to generate binding sites placed distantly from optimal activation positions, generating repressive interactions at the resulting promoter. The logic gates used to construct the representations along with their respective truth tables are presented at the bottom of the figure.

strains harboring deletions of CRP, IHF, or Fis global regulators[42] were used as hosts of all plasmids. *E. coli* DH5α was used as the cloning strain. *E. coli* strains were grown in either Luria Broth (LB) or M9 minimal media (6.4 g/L $Na_2HPO_4 \cdot 7H_2O$, 1.5 g/L KH_2PO_4, 0.25 g/L NaCl, 0.5 g/L NH_4Cl) supplemented with 2 mM $MgSO_4$, 0.1 mM $CaCl_2$, 0.1 mM casamino acids, and either 1% glycerol or 0.4% glucose as the carbon source. Chloramphenicol was added (*Cm*, 30 μg/mL) to ensure plasmid maintenance. Isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.1 mM) was used as inductor of the repressor constructs. Cells were grown at 37 °C with constant shaking at 225 rpm.

Plasmid Construction. To assay synthetic combinatorial promoters with dual binding sites (CRP and IHF), pMR1-Plac-CI1 and pMR1-Plac-CI2 were used.[31] pMR1 (empty vector), pMR1-Plac-wt, and pMR1-Plac-ΔCRP[31] were used as controls. All vectors carry the chloramphenicol resistance gene. The synthetic combinatorial promoter with overlapping sites for the three global regulators CRP, IHF, and Fis was developed *in silico* as described previously.[31] Briefly, TF binding sites for CRP, IHF, and Fis were retrieved from the RegulonDB database;[46] only those sites with experimental support were used. Position weight matrices (PWMs) were used to generate randomly assembled DNA sequences, 19 nt in length, for the three global regulators. The sequence scores were used to determine the relative binding affinities of CRP, IHF, and Fis. Finally, an *in silico* optimization algorithm implemented in *Perl* was used to select short sequences with optimized binding sites for the three global regulators. Results were processed using scripts in R (http://www.r-project.org/). For the construction of the triple promoter, pMR1 vector[31] carrying a short-lived GFP-lva variant[47] and chloramphenicol resistance gene was used to clone an optimized sequence named CFI1 upstream to the core *Plac* sequence, thus generating the *Plac-CFI1* promoter (Figure 1A and C). First, to generate the *Plac-CFI1* promoter, pUC19 vector[48] was used as template DNA. *Plac-CFI1* was amplified using the primers 5-PlacCFI1 and M13/F-20 containing *Eco*RI/*Bam*HI restriction sites and the CFI1 19 nt sequence. The 170 bp PCR product was then cloned into the pMR1 vector digested with the same enzymes. The plasmid generated, pMR1-Plac-CFI1, was next sequenced to verify the construct. The plasmid was used to transform competent *E. coli* wild type, and mutant cells depleted of CRP, IHF, or Fis. In order to implement variants of the optimization algorithm, we generated *Perl* codes in which size of the DNA sequences was set from 20 to 23 bp, and performed the simulations. For each simulation, we selected the top five sequences based on the average score for the three regulators (Table S1 in Supporting Information). Of the 20 candidates, 11 were successfully cloned in the pMR1 vector as described above (Table 1). These new candidates were then introduced into the wild type *E. coli* strain BW25113 and assayed for promoter activity (see below). Two positive promoters were named $P_{CFI20-1}$ and $P_{CFI22-5}$, introduced into the mutant *E. coli* strains, and used for further analysis.

Promoter Activity Assay. For promoter activity measurements, single colonies of wild type and mutant *E. coli* strains harboring the different plasmids were picked from fresh plates. Each strain was inoculated and grown overnight in 2 mL of either LB or M9 medium containing chloramphenicol at 37 °C with shaking at 225 rpm. Stationary phase cultures were diluted to a final optical density (OD) of 0.05 in 20 mL of either LB or M9 medium containing either glycerol or glucose as required, and supplemented with chloramphenicol and IPTG. Cultures were then incubated at 37 °C with shaking at 225 rpm for 4 h and samples were taken each 20 min. Samples from LB cultures were washed twice with 10 mM $MgSO_4$ buffer, resuspended in 1 mL of the same buffer, and kept on ice (when using M9 minimal medium, samples were kept on ice until use). Fluorescence was quantified either in a Victor X3 plate reader (PerkinElmer) or by flow cytometry analyses as described below. Statistical significance of the experiments were analyzed using one-way ANOVA.

Flow Cytometry. Cells were analyzed by flow cytometry, using a BD FACSCanto II flow cytometer configuration 4−2−2 (BD Biosciences). Blue laser was used to induce an excitation wave of 488 nm for forward scatter (FSC), side scatter (SSC), and fluorescence. SSC detection was at 488 ± 10 nm. Fluorescence was detected at 530 ± 30 nm. The threshold parameter was 200 for FSC. The PMT voltage settings were 260 (FSC), 400 (SSC), and 380 (FITC) to decrease the electronic noise. The flow rate was set to "low." For each sample, dot-plot graph (FSC, SSC) was generated and set to collect 50,000 events inside the gate. The data was exported using the BD FACSDiva software v6.1.3.

DOI: 10.1021/acssynbio.5b00098
ACS Synth. Biol. XXXX, XXX, XXX−XXX

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

strain	description	reference
E. coli DH5α	*F⁻ endA1 gln V44 thi-1 recA1 relA1 gyrA96 deoR nupG Φ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17(r_K^- m_K^+), λ⁻*	49
E. coli BW25113	*Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻, rph-1, Δ(rhaD-rhaB)568, hsdR514.* Wild type strain	42
E. coli JW5702-4	*Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻, Δcrp-765::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514. Δcrp* mutant strain	42
E. coli JW1702-1	*Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻, ΔihfA786::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514. ΔihfA* mutant strain	42
E. coli JW3229-1	*Δ(araD-araB)567, ΔlacZ4787(::rrnB-3), λ⁻, Δfis-779::kan, rph-1, Δ(rhaD-rhaB)568, hsdR514. Δfis* mutant strain	42
Plasmids		
pUC19	Ap^R, *oripColE1, lacZα*; standard cloning vector	48
pMR1	Cm^R, *orip15a*; *GFPlva* promoter probe vector	31
pMR1-*Plac-wt*	Cm^R, *orip15a*; *Plac-wt-GFPlva* transcriptional fusion	31
pMR1-*Plac-Δcrp*	Cm^R, *orip15a*; *Plac-Δcrp-GFPlva* transcriptional fusion	31
pMR1-*Plac-CI1*	Cm^R, *orip15a*; *Plac-CI1-GFPlva* transcriptional fusion	31
pMR1-*Plac-CI2*	Cm^R, *orip15a*; *Plac-CI2-GFPlva* transcriptional fusion	31
pMR1-*Plac-CFI1*	Cm^R, *orip15a*; *Plac-CFI1-GFPlva* transcriptional fusion	This study
pMR1-$P_{CFI20\text{-}1}$	Cm^R, *orip15a*; $P_{CFI20\text{-}1}$-*GFPlva* transcriptional fusion	This study
pMR1-$P_{CFI20\text{-}2}$	Cm^R, *orip15a*; $P_{CFI20\text{-}2}$-*GFPlva* transcriptional fusion	This study
pMR1-$P_{CFI20\text{-}3}$	Cm^R, *orip15a*; $P_{CFI20\text{-}3}$-*GFPlva* transcriptional fusion	This study
pMR1-$P_{CFI20\text{-}4}$	Cm^R, *orip15a*; $P_{CFI20\text{-}4}$-*GFPlva* transcriptional fusion	This study
pMR1-$P_{CFI21\text{-}1}$	Cm^R, *orip15a*; $P_{CFI21\text{-}1}$-*GFPlva* transcriptional fusion	This study
pMR1-$P_{CFI21\text{-}2}$	Cm^R, *orip15a*; $P_{CFI21\text{-}2}$-*GFPlva* transcriptional fusion	This study
pMR1-$P_{CFI21\text{-}3}$	Cm^R, *orip15a*; $P_{CFI21\text{-}3}$-*GFPlva* transcriptional fusion	This study
pMR1-$P_{CFI22\text{-}1}$	Cm^R, *orip15a*; $P_{CFI22\text{-}1}$-*GFPlva* transcriptional fusion	This study
pMR1-$P_{CFI22\text{-}5}$	Cm^R, *orip15a*; $P_{CFI22\text{-}5}$-*GFPlva* transcriptional fusion	This study
pMR1-$P_{CFI23\text{-}1}$	Cm^R, *orip15a*; $P_{CFI23\text{-}1}$-*GFPlva* transcriptional fusion	This study
pMR1-$P_{CFI23\text{-}4}$	Cm^R, *orip15a*; $P_{CFI23\text{-}4}$-*GFPlva* transcriptional fusion	This study
Primers		
5-PlacCFI1	5′-GCGCGAATTCTAAGAAGATGATTTTCACACATTAGGCACCCCAGG-3′	This study
M13/F-20	5′-GTAAAACGACGGCCAGT-3′	48

Data Analysis. Data from a binary flow cytometry standard (FCS) format was converted into an ASCII text format using FCSExtract Utility (http://research.stowers-institute.org/mcm/efg/ScientificSoftware/Utility/FCSExtract/index.htm) for subsequent analysis. Fluorescence geometric mean of the gated population was calculated. The arithmetic estimations and graphs were created using scripts in R and MatLab software (MathWorks).

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acssynbio.5b00098.

Table S1. List of designed sequences used in this study. Figure S1. Growth profiles of wild type and *ihf*- and *fis* mutant strains of *E. coli* in minimal media. Figure S2. Growth profiles of wild type and *ihf*- and *fis* mutant strain of *E. coli* in LB media. (PDF)

■ AUTHOR INFORMATION

Corresponding Author

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

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This study was supported by the National Council of Technological and Scientific Development of Brazil (project CNPq-Universal 14/2013, ref. number 472893/2013-0 and 14/2014 ref. number 441833/2014-4), and by the Sao Paulo State Foundation (FAPESP, grant number 2012/22921-8). M.-E.G. benefits from a Young Talent Fellowship of the National Council of Technological and Scientific Development of Brazil (CNPq, ref. number 370630/2013-0). G.R.A. was supported by a CAPES Doctoral Fellowship.

■ REFERENCES

(1) Ulrich, L. E., Koonin, E. V., and Zhulin, I. B. (2005) One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol. 13*, 52−56.

(2) Keren, L., Zackay, O., Lotan-Pompan, M., Barenholz, U., Dekel, E., Sasson, V., Aidelberg, G., Bren, A., Zeevi, D., Weinberger, A., Alon, U., Milo, R., and Segal, E. (2013) Promoters maintain their relative activity levels under different growth conditions. *Mol. Syst. Biol. 9*, 701.

(3) Scott, M., and Hwa, T. (2011) Bacterial growth laws and their applications. *Curr. Opin. Biotechnol. 22*, 559−565.

(4) Klumpp, S., and Hwa, T. (2014) Bacterial growth: global effects on gene expression, growth feedback and proteome partition. *Curr. Opin. Biotechnol. 28*, 96−102.

(5) Lovén, J., Orlando, D. A., Sigova, A. A., Lin, C. Y., Rahl, P. B., Burge, C. B., Levens, D. L., Lee, T. I., and Young, R. A. (2012) Revisiting global gene expression analysis. *Cell 151*, 476−482.

(6) Gerosa, L., Kochanowski, K., Heinemann, M., and Sauer, U. (2013) Dissecting specific and global transcriptional regulation of bacterial gene expression. *Mol. Syst. Biol. 9*, 658.

(7) Rothschild, D., Dekel, E., Hausser, J., Bren, A., Aidelberg, G., Szekely, P., and Alon, U. (2014) Linear superposition and prediction of bacterial promoter activity dynamics in complex conditions. *PLoS Comput. Biol. 10*, e1003602.

(8) Kalir, S., McClure, J., Pabbaraju, K., Southward, C., Ronen, M., Leibler, S., Surette, M. G., and Alon, U. (2001) Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria. *Science 292*, 2080−2083.

DOI: 10.1021/acssynbio.5b00098
ACS Synth. Biol. XXXX, XXX, XXX−XXX

(9) Zaslaver, A., Mayo, A. E., Rosenberg, R., Bashkin, P., Sberro, H., Tsalyuk, M., Surette, M. G., and Alon, U. (2004) Just-in-time transcription program in metabolic pathways. *Nat. Genet. 36*, 486−491.
(10) Browning, D. F., and Busby, S. J. (2004) The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol. 2*, 57−65.
(11) Lee, D. J., Minchin, S. D., and Busby, S. J. (2012) Activating transcription in bacteria. *Annu. Rev. Microbiol. 66*, 125−152.
(12) Osterberg, S., del Peso-Santos, T., and Shingler, V. (2011) Regulation of alternative sigma factor use. *Annu. Rev. Microbiol. 65*, 37−55.
(13) Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., Jimenez-Jacinto, V., Salgado, H., Juarez, K., Contreras-Moreira, B., Huerta, A. M., Collado-Vides, J., and Morett, E. (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in E. coli. *PLoS One 4*, e7526.
(14) Martínez-Antonio, A., and Collado-Vides, J. (2003) Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr. Opin. Microbiol. 6*, 482−489.
(15) Ishihama, A. (2010) Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks. *FEMS Microbiol Rev. 34*, 628−645.
(16) Silva-Rocha, R., and de Lorenzo, V. (2010) Noise and robustness in prokaryotic regulatory networks. *Annu. Rev. Microbiol. 64*, 257−275.
(17) Janga, S. C., Salgado, H., and Martinez-Antonio, A. (2009) Transcriptional regulation shapes the organization of genes on bacterial chromosomes. *Nucleic Acids Res. 37*, 3680−3688.
(18) 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*, 502−508.
(19) Brophy, J. A., and Voigt, C. A. (2014) Principles of genetic circuit design. *Nat. Methods 11*, 508−520.
(20) Church, G. M., Elowitz, M. B., Smolke, C. D., Voigt, C. A., and Weiss, R. (2014) Realizing the potential of synthetic biology. *Nat. Rev. Mol. Cell Biol. 15*, 289−294.
(21) Guazzaroni, M. E., Silva-Rocha, R., and Ward, R. J. (2015) Synthetic biology approaches to improve biocatalyst identification in metagenomic library screening. *Microb. Biotechnol. 8*, 52−64.
(22) Singh, V. (2014) Recent advancements in synthetic biology: current status and challenges. *Gene 535*, 1−11.
(23) Siuti, P., Yazbek, J., and Lu, T. K. (2013) Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol. 31*, 448−452.
(24) 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.
(25) Zhan, J., Ding, B., Ma, R., Ma, X., Su, X., Zhao, Y., Liu, Z., Wu, J., and Liu, H. (2010) Develop reusable and combinable designs for transcriptional logic gates. *Mol. Syst. Biol. 6*, 388.
(26) Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S., and Elowitz, M. B. (2007) Accurate prediction of gene feedback circuit behavior from component properties. *Mol. Syst. Biol. 3*, 143.
(27) Voigt, C. A. (2006) Genetic parts to program bacteria. *Curr. Opin. Biotechnol. 17*, 548−557.
(28) Murphy, K. F., Adams, R. M., Wang, X., Balázsi, G., and Collins, J. J. (2010) Tuning and controlling gene expression noise in synthetic gene networks. *Nucleic Acids Res. 38*, 2712−2726.
(29) Hooshangi, S., Thiberge, S., and Weiss, R. (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc. Natl. Acad. Sci. U. S. A. 102*, 3581−3586.
(30) Pedraza, J. M., and van Oudenaarden, A. (2005) Noise propagation in gene networks. *Science 307*, 1965−1969.
(31) 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.
(32) Hunziker, A., Tuboly, C., Horvath, P., Krishna, S., and Semsey, S. (2010) Genetic flexibility of regulatory networks. *Proc. Natl. Acad. Sci. U. S. A. 107*, 12998−13003.
(33) Gertz, J., and Cohen, B. A. (2009) Environment-specific combinatorial cis-regulation in synthetic promoters. *Mol. Syst. Biol. 5*, 244.
(34) Gertz, J., Siggia, E. D., and Cohen, B. A. (2009) Analysis of combinatorial cis-regulation in synthetic and genomic promoters. *Nature 457*, 215−218.
(35) Kinkhabwala, A., and Guet, C. C. (2008) Uncovering cis regulatory codes using synthetic promoter shuffling. *PLoS One 3*, e2030.
(36) Alper, H., Fischer, C., Nevoigt, E., and Stephanopoulos, G. (2005) Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci. U. S. A. 102*, 12678−12683.
(37) Hermsen, R., Ursem, B., and ten Wolde, P. R. (2010) Combinatorial gene regulation using auto-regulation. *PLoS Comput. Biol. 6*, e1000813.
(38) Mayo, A. E., Setty, Y., Shavit, S., Zaslaver, A., and Alon, U. (2006) Plasticity of the cis-regulatory input function of a gene. *PLoS Biol. 4*, e45.
(39) Hermsen, R., Tans, S., and ten Wolde, P. R. (2006) Transcriptional regulation by competing transcription factor modules. *PLoS Comput. Biol. 2*, e164.
(40) Cox, R. S., 3rd, Surette, M. G., and Elowitz, M. B. (2007) Programming gene expression with combinatorial promoters. *Mol. Syst. Biol. 3*, 145.
(41) Bruckner, R., and Titgemeyer, F. (2002) Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization. *FEMS Microbiol. Lett. 209*, 141−148.
(42) 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.
(43) 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*, 1950−1957.
(44) Berg, J., Willmann, S., and Lassig, M. (2004) Adaptive evolution of transcription factor binding sites. *BMC Evol. Biol. 4*, 42.
(45) Sadler, J. R., Sasmor, H., and Betz, J. L. (1983) A perfectly symmetric lac operator binds the lac repressor very tightly. *Proc. Natl. Acad. Sci. U. S. A. 80*, 6785−6789.
(46) 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.
(47) Andersen, J. B., Sternberg, C., Poulsen, L. K., Bjorn, S. P., Givskov, M., and Molin, S. (1998) New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl. Environ. Microbiol. 64*, 2240−2246.
(48) Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene 33*, 103−119.
(49) Grant, S. G., Jessee, J., Bloom, F. R., and Hanahan, D. (1990) Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proc. Natl. Acad. Sci. U. S. A. 87*, 4645−4649.
(50) Ussery, D., Larsen, T. S., Wilkes, K. T., Friis, C., Worning, P., Krogh, A., and Brunak, S. (2001) Genome organisation and chromatin structure in Escherichia coli. *Biochimie 83*, 201−212.

DOI: 10.1021/acssynbio.5b00098
ACS Synth. Biol. XXXX, XXX, XXX−XXX