



Sensory Systems and Transcriptional Regulation in *Escherichia coli*

Georgette Femerling^{1†}, Socorro Gama-Castro^{1†}, Paloma Lara^{1†}, Daniela Ledezma-Tejeida², Víctor H. Tierrafría^{1,3}, Luis Muñoz-Rascado¹, César Bonavides-Martínez¹ and Julio Collado-Vides^{1,3,4*}

¹Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, México, ²Institute of Molecular Systems Biology, ETH Zürich, Zurich, Switzerland, ³Department of Biomedical Engineering, Boston University, Boston, MA, United States, ⁴Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Universitat Pompeu Fabra (UPF), Barcelona, Spain

OPEN ACCESS

Edited by:

Rodolfo García-Contreras,
National Autonomous University of
Mexico, Mexico

Reviewed by:

Fang-Fang Wang,
Institute of Microbiology (CAS), China
Eveline Peeters,
Vrije University Brussel, Belgium
Corina-Diana Ceapa,
Universidad Nacional Autónoma de
México, Mexico

*Correspondence:

Julio Collado-Vides
colladojulio@gmail.com

[†]These authors have contributed
equally to this work and share first
authorship

Specialty section:

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

Received: 26 November 2021

Accepted: 18 January 2022

Published: 14 February 2022

Citation:

Femerling G, Gama-Castro S, Lara P,
Ledezma-Tejeida D, Tierrafría VH,
Muñoz-Rascado L,
Bonavides-Martínez C and
Collado-Vides J (2022) Sensory
Systems and Transcriptional
Regulation in *Escherichia coli*.
Front. Bioeng. Biotechnol. 10:823240.
doi: 10.3389/fbioe.2022.823240

In free-living bacteria, the ability to regulate gene expression is at the core of adapting and interacting with the environment. For these systems to have a logic, a signal must trigger a genetic change that helps the cell to deal with what implies its presence in the environment; briefly, the response is expected to include a feedback to the signal. Thus, it makes sense to think of genetic sensory mechanisms of gene regulation. *Escherichia coli* K-12 is the bacterium model for which the largest number of regulatory systems and its sensing capabilities have been studied in detail at the molecular level. In this special issue focused on biomolecular sensing systems, we offer an overview of the transcriptional regulatory corpus of knowledge for *E. coli* that has been gathered in our database, RegulonDB, from the perspective of sensing regulatory systems. Thus, we start with the beginning of the information flux, which is the signal's chemical or physical elements detected by the cell as changes in the environment; these signals are internally transduced to transcription factors and alter their conformation. Signals transduced to effectors bind allosterically to transcription factors, and this defines the dominant sensing mechanism in *E. coli*. We offer an updated list of the repertoire of known allosteric effectors, as well as a list of the currently known different mechanisms of this sensing capability. Our previous definition of elementary genetic sensory-response units, SENSOR units for short, that integrate signals, transport, gene regulation, and the biochemical response of the regulated gene products of a given transcriptional factor fit perfectly with the purpose of this overview. We summarize the functional heterogeneity of their response, based on our updated collection of SENSORS, and we use them to identify the expected feedback as part of their response. Finally, we address the question of multiple sensing in the regulatory network of *E. coli*. This overview introduces the architecture of sensing and regulation of native components in *E. coli* K-12, which might be a source of inspiration to bioengineering applications.

Keywords: signal, gene regulation, sensory-response genetic units, feedback, biosensor, transcriptional regulatory network, *E. coli*, transcription factor

INTRODUCTION

The ability to adapt to changes in the environment is a fundamental property of life, which in bacterial systems has been studied for decades at the molecular level, thanks to advances in genetics and the relative simplicity of these organisms.

During more than two and a half decades our laboratory has been gathering knowledge on the regulation of transcription initiation and the organization and expression of the regulated genes in *Escherichia coli* K-12. This knowledge can be accessed in two databases, RegulonDB, and EcoCyc (Santos-Zavaleta et al., 2019; Keseler et al., 2021). This corpus shows the complex architecture of multiple sensing and regulatory systems currently known in *E. coli* and is the basis for the review presented here from the perspective of sensing systems.

RegulonDB Overview

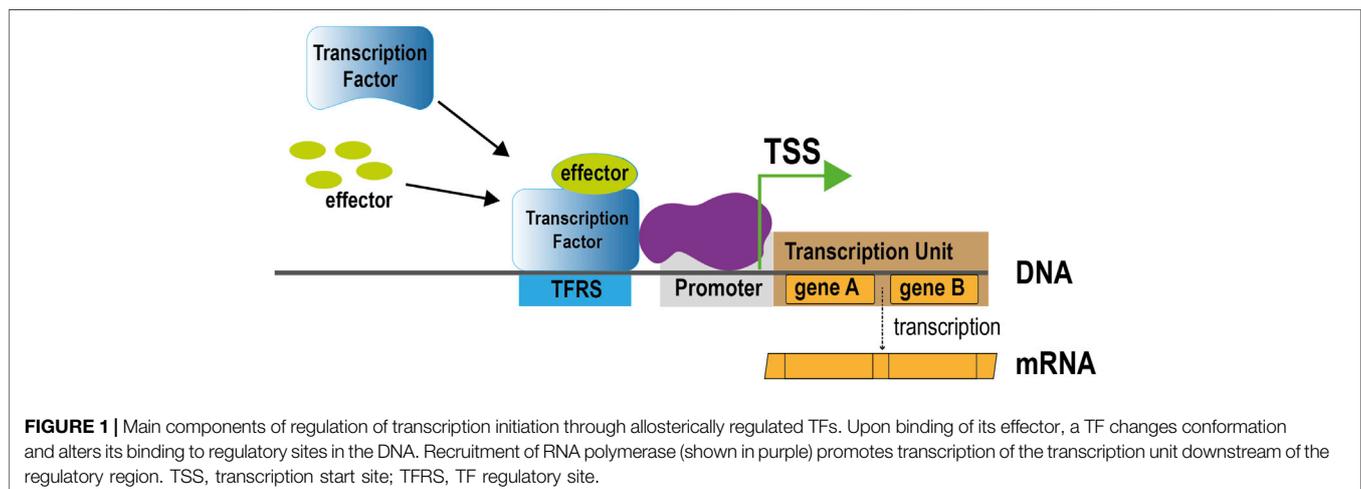
Progress through the years of our biocuration efforts on regulation of transcription initiation and operon organization in *E. coli* K-12 has been periodically published, mostly in the special issues on databases in *Nucleic Acids Research* (Huerta et al., 1998; Gama-Castro et al., 2016; Santos-Zavaleta et al., 2019). The most common way to share progress has been in terms of the number of the main players of gene regulation, such as transcription factors (TFs), the operator DNA sequences to which TFs bind, called TF binding sites (TFBS), and TF regulatory sites (TFRSs) when there is evidence of their regulatory role in addition to binding, promoters, and other regulatory elements, including transcription start sites (TSSs) and transcription units (**Figure 1**). These elements include objects based on their interactions, such as regulatory interactions, that link TFs and their activating or repressing effects on the target genes, with detailed knowledge of the promoter involved for several of them, as well as operons and simple and complex regulons defined as the collection of target genes by a single TF, or by a group of TFs, respectively, (Mejia-Almonte et al., 2020). Together they define the transcriptional regulatory network of *E. coli*.

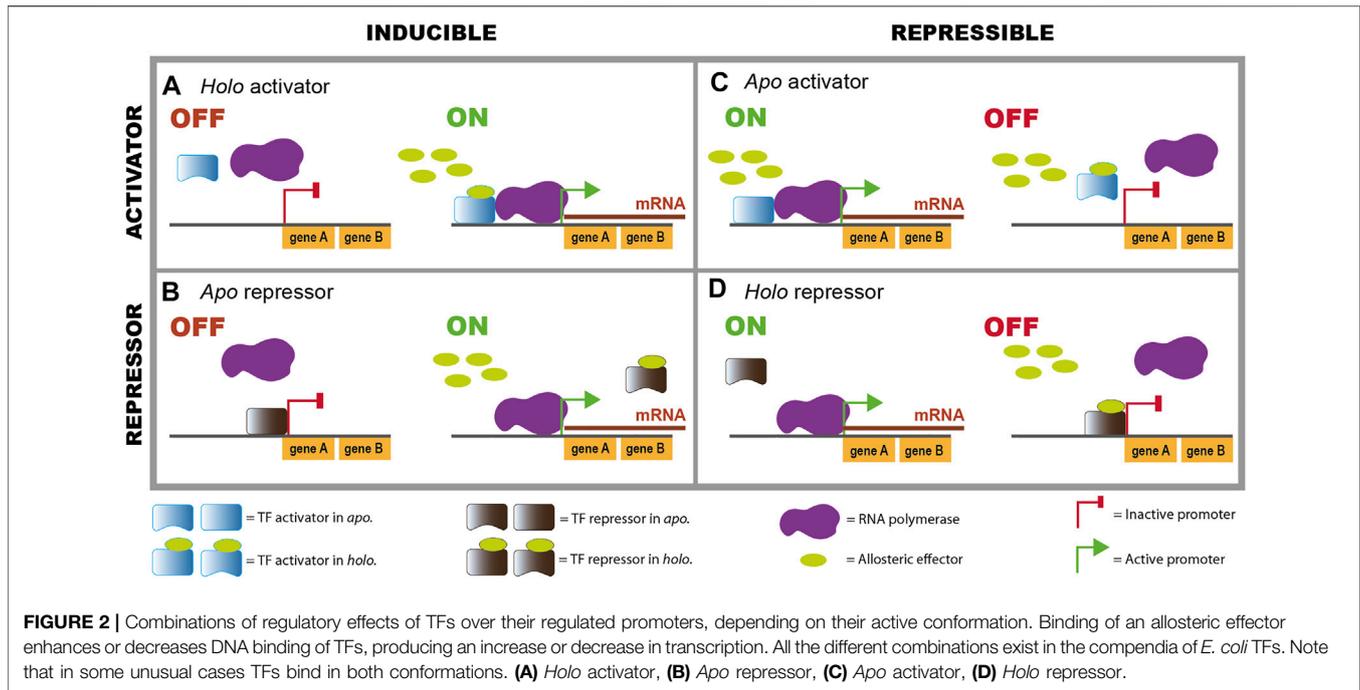
RegulonDB was originally organized within a conceptual framework that maintains a clean genomic picture following the classic definitions of the regulation of transcription initiation. However, these concepts required updating (Mejia-Almonte et al., 2020) to better represent the constant expansion of knowledge about transcriptional regulation, including the massive identification of regulatory elements made possible using high-resolution, genome-scale strategies (Seo et al., 2014; Yan et al., 2018; Ju et al., 2019).

Regulation and Sensing Systems

In some well-studied systems, the *signal*, considered herein and in RegulonDB as the metabolite or physical change (e.g., temperature, pH, osmotic pressure, etc.) that initiates a flux of information that will affect the expression of one or a group of genes, has been identified and linked to the molecular mechanisms it triggers. In the well-known case of lactose metabolism (Pardee et al., 1958), in the absence of glucose, when the signal, lactose, reaches a certain concentration outside the cell, it is transported into the cytoplasm where it is isomerized to allolactose, a different but related molecule. This metabolite, which we call an *effector*, will bind LacI, the cognate TF, which will unbind from its operator sites. This unbinding of the repressor and the binding of CRP, the TF that responds to glucose, will induce the expression of genes involved in the metabolism of lactose.

In other more complex cases, the effector is known but the signal is less clear. For example, the TFs PdhR and IclR bind pyruvate, a central metabolite that can potentially change in abundance due to many different environmental triggers (Lorca et al., 2007; Anzai et al., 2020). To go from the conditions that provoke a genetic response to deciphering the molecular mechanisms that support such a response is not always straightforward. Experimentalists have to isolate the cause of changes in gene expression from additional cellular changes. For instance, changing from nitrogen-poor to nitrogen-rich sources will affect the growth rate, which by itself provokes additional changes in gene expression, making it harder to track the precise information flux from signal to effector of nitrogen-related TFs



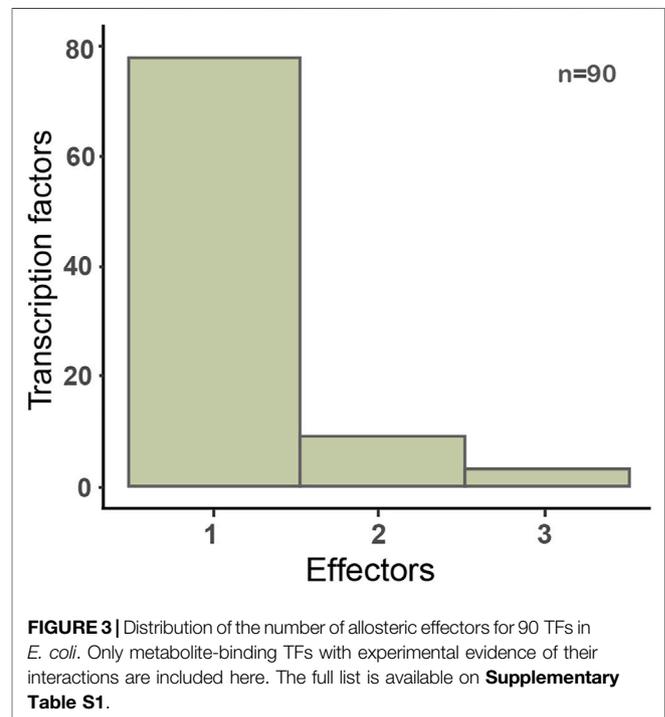


(Magasanik, 2000; Zimmer et al., 2000). Regulatory systems are also built based on responses to internal signals, and in fact, in several cases the effectors that bind to the TFs may have either an external origin as a metabolite transported into the cell or an internal origin resulting from an enzymatic activity inside the cell, or both. The distribution of internal and external signals in relation to local and global regulation has been previously analyzed (Martínez-Antonio et al., 2006).

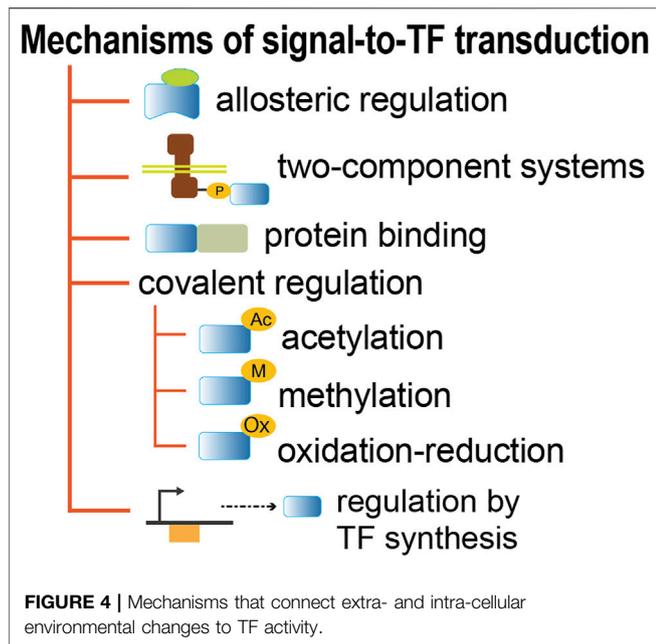
Information Transfer From Signal to TF

The information flux from the signal to the TF is achieved either by direct binding of the signal to the TF, or more frequently through a transduction process when the signal molecule is chemically transformed into the effector molecule which binds and alters the TF conformation. The current corpus of knowledge of *E. coli* sensory systems shows that by far the most frequent mechanism of this information transfer is achieved by means of allosteric interactions of specific metabolites that bind to the TF (Madan Babu and Teichmann, 2003).

For instance, for the lactose metabolism described above, LacI switches from its *apo* conformation as a free protein, which is the one that binds to TFRSs or operator sites, to its *holo* conformation as LacI-allolactose complex unbinding from its TFRSs. This explains the induction that lactose can exert on the regulated *lacXYZ* operon, via the unbinding of a repressor TF (**Figure 2B**). Another example is catabolite repression, a term used for the downregulation of several operons that encode enzymes for the utilization of alternative, nonpreferred carbon sources (Görke and Stülke, 2008). When glucose, which is largely considered the preferred carbon source in *E. coli*, is incorporated in the cell, it produces a decrease in the intracellular concentration of adenosine 3',5'-cyclic monophosphate (cyclic AMP, or cAMP), the effector of the global CRP regulator.



CRP is a transcriptional regulator of several operons involved in metabolism of different carbon sources; it binds to its TFRSs in a holo conformation as the complex CRP-cAMP. Thus, some operons are turned off in the presence of glucose (which leads to low cAMP levels), due to the unbinding of an activator that changes from its holo to its apo conformation (**Figure 2A**). Other TFs use mechanisms depicted in **Figures 2B-D**.



In summary, there are four possible combinations of activator or repressor and *apo* or *holo* conformations that bind to the operator sites (Figure 2). This is the general behavior although in some exceptional cases a TF can bind both in *apo* and in *holo*, and even upstream of the same promoter as illustrated by AraC (Schleif, 2010).

The CRP regulon is an example of regulation by an allosteric metabolite, which is the most frequent mechanism of information flux from the signal to the TF. The first comprehensive compilation of the literature on the different conformations and metabolites that bind to TFs was reported by Balderas et al. (Balderas-Martinez et al., 2013). Since then, we have continued gathering this knowledge as it becomes available, so that currently of the 221 TFs with evidence of at least one regulatory interaction experimentally identified, we know the effector for 90 of them. Although by far current knowledge tells us that TFs respond to only one specific allosteric metabolite, in some cases TFs may sense more than two or even more different metabolites (Figure 3, Supplementary Table S1).

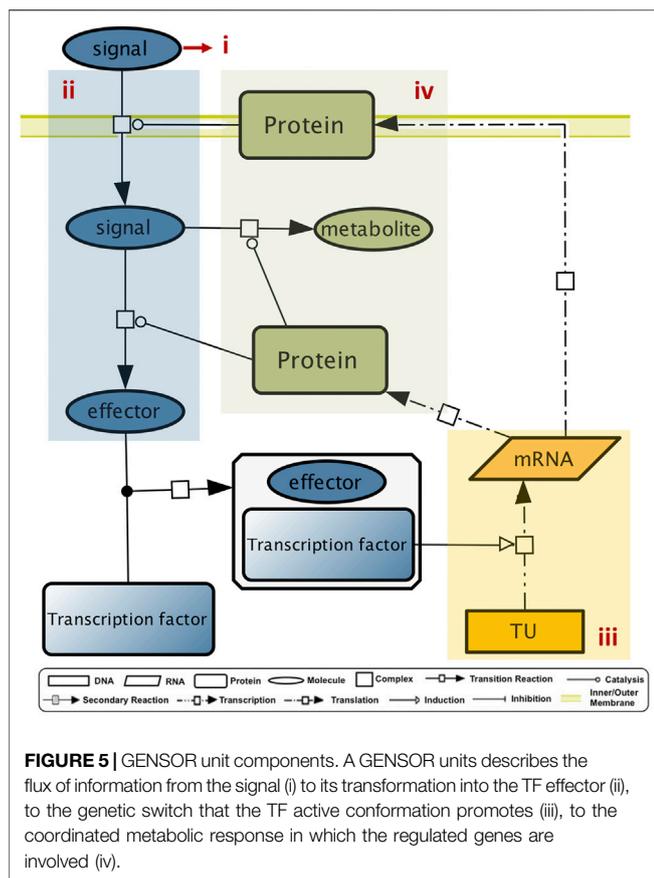
Some of those multiple effectors for a single TF, such as the metal-binding proteins ZntR and NikR, have been identified *in vitro*, but it is not fully certain that they have a physiological significance (Changela et al., 2003; Leitch et al., 2007). For instance, if the kinetics of the metal-TF complex only allows an interaction when metal concentrations are lethal for the bacterium, then the interaction is unlikely to have physiological significance. Another interesting case is the Lrp regulator whose main effector is leucine, but it has been shown that other amino acids, such as methionine and alanine can also bind allosterically to Lrp, but their physiological significance has not been well established (Hart and Blumenthal, 2011). It is important to mention that our TF-effector statistics reflect only those effectors which support regulatory interactions in *E. coli* K-12.

Less clear at this time is whether different conformations for the same TF can recognize different subsets of TFRSs, thus regulating different subsets of target genes. For instance, it was recently shown that, depending on the composition of the culture medium, different sets of genes are subject to regulation by the leucine-responsive regulator (Lrp) (Kroner et al., 2019). This is supported by other studies showing that binding of leucine to Lrp favors one specific Lrp conformation over others that coexist *in vivo*, increasing the DNA-binding affinity of Lrp to sites where it would not bind if leucine were not present (Chen et al., 2001; Chen and Calvo, 2002).

The signal-to-TF information flux can also be executed by mechanisms different from allosteric binding to TFs (Figure 4, Supplementary Table S1). The two-component systems involve the covalent phosphorylation-dephosphorylation of the TF by the cognate histidine kinase sensor. In some cases, it has been shown that the kinase activity depends on the allosteric binding of an effector, as happens with ArcA (Jeon et al., 2001). Other TFs in bacteria rely on protein-protein interactions to modify the conformation capable of altering their DNA binding ability, such as toxin-antitoxin systems, as well as MalT (Schreiber et al., 2000; Joly et al., 2002; Mandrich et al., 2002; Schlegel et al., 2002), Mlc (Lee et al., 2000; Nam et al., 2001), and TorR, which are involved in responses to alkaline/acid stress in which the binding of TorI can affect recruitment of the RNA polymerase (Ansaldi et al., 2004). Other TFs, such as Ada, RcsB, and FNR, can be modulated by covalent modifications that include methylation, acetylation, and oxidation-reduction (Takahashi et al., 1988; Khoroshilova et al., 1997; Hu et al., 2013). The regulatory function of TFs can also be triggered by their own synthesis, without the presence of any effector metabolite. These usually have a slower response given the time needed to synthesize the full active protein in adequate concentrations. This is the case for some TFs like IHF, HNS, and other so-called nucleoid-associated proteins, whose changes in concentration are associated with changes of growth phase (Ali Azam et al., 1999; Azam and Ishihama, 1999).

Although it is not well understood why a particular mechanism is used in each signal-to-TF transduction process, the demand theory of gene regulation suggests a framework where the optimal solution is selected in evolution (Wall et al., 2004). Much remains unknown when searching for rational explanations, given the difficulty to test them. At the other extreme, TF orthologs may control different biological processes, as illustrated by CRP. Whereas this TF functions as the master regulator of catabolite repression in *E. coli*, its ortholog in *Pseudomonas* species regulates membrane-related functions (Milanesio et al., 2011).

Irrespective of the mechanisms for sensing, the signal induces a change in the conformation of the TF, and this change will either increase or decrease the concentration of active TF. This, in turn, will promote either binding or unbinding to its DNA target sites, exerting their regulatory effect on the corresponding regulated promoters. The sensing aspect ends by conveying the information originated by the signal to the genome. Here begins the genetic response encoded in the target genes of the regulated promoters. As a first approximation, the whole process



can be described as a genetic sensory-response unit, or GENSOR unit for short.

Genetic Sensory-Response Regulated Units

An elementary GENSOR unit includes the events from signal detection to the outset of a functional response, mediated by an individual TF. This process is summarized in four components: (Santos-Zavaleta et al., 2019) the signal, (Keseler et al., 2021) the conversion of signal to the effector, (Huerta et al., 1998) the set of genetic switches, and (Gama-Castro et al., 2016) the response (Figure 5, (Ledezma-Tejeida et al., 2017)).

Methodologically, the assembly of elementary GENSOR units centers around an individual TF and includes all the genes whose promoters are regulated by the TF, their mRNAs, their gene products, and the protein complexes they belong to. If the gene product is an enzyme, the reaction catalyzed, its substrates, and products are also included. From the information organized in RegulonDB and EcoCyc, we previously built GENSOR units for 189 TFs (these can be found in RegulonDB v10.9, grouped in different classes; http://regulondb.ccg.unam.mx/central_panel_menu/integrated_views_and_tools/gensor_unit_groups). The TF collection was updated for this work, (See Data Availability), and we now have 204 GENSOR units, for which an effector is described for 87 TFs. Each GENSOR unit is a multilevel network linking

transcriptional regulation to metabolism by including the cellular components that are directly affected by the regulatory activity of a TF. They were used to predict metabolites allosterically regulating a TF, after the observation that 83% of TFs for which a binding molecule was known had it in their GENSOR unit (Ledezma-Tejeida et al., 2017). Further analyses focused on quantifying the functional homogeneity of the metabolic response mediated by a single TF. We have previously shown through several metrics that only ~25% of TFs regulate one biological process, as opposed to two or more, and in 16% of them the gene products encode reactions that cannot be linked to classical pathways (Ledezma-Tejeida et al., 2019).

These are elementary GENSOR units, as they are limited to one TF. As discussed below under “Genomic processing of multiple signals in *E. coli*,” this is still a simplification of the complex architecture of sensing and genetic regulation exerted by TFs in orchestrating changes in gene expression. Nonetheless, they have been useful as a first multilevel integration, as discussed elsewhere (Ledezma-Tejeida et al., 2019). Furthermore, and as illustrated below, they are amenable to computational processing to identify if their response includes feedback as expected. GENSOR units must be periodically updated as new regulatory interactions are identified, particularly with the combination of high-throughput technologies for the identification of target sites for TFs in the genome and genome-wide expression profiles (Aquino et al., 2017; Santos-Zavaleta et al., 2019).

Response to Signal Feedback

Signals trigger a change in the expression of one or more genes, and this response helps the cell to deal with what is implied by its presence in the environment. It is thus expected that the response includes feedback to the signal, for instance, the presence of carbon sources induces the expression of gene products used in their utilization, the presence of amino acids can repress their synthesis and save the energy needed in their synthesis. Furthermore, feedback to the signal should help TFs return to their previous state after transient changes. In order to explore this question, we reassembled GENSOR units by combining data from EcoCyc version 25.1 (Keseler et al., 2017) and RegulonDB version 10.9 (Santos-Zavaleta et al., 2019).

First, we focused on the simplest type of feedback: a metabolite signaling the TF is produced, transformed, or transported by an enzyme whose gene is directly regulated by the TF. We implemented a computational pipeline to automatically retrieve feedback from a GENSOR unit by identifying whether the effector molecule takes part in a reaction that is part of the GENSOR unit, excluding of course the effector-TF binding reaction (See the Material and Methods section, below). Of 90 TFs with information on their effectors, 87 had a GENSOR unit assembled. We found direct feedback in 71 GENSOR units, retrieving feedback computationally in 82% of the cases (for detailed results see **Supplementary Table S2**). Because we expected every sensory-response unit to have feedback, we analyzed in detail the remaining 16 GENSOR units where the feedback was not detected (**Supplementary Table S3**). We found four major reasons that explained why the feedback was not found:

- 1) The allosteric effector of the TF is not present in a metabolic reaction in the GENSOR unit but a compound either in the subclass or superclass (within the EcoCyc classification) of the effector is part of the GENSOR unit. For example, the effector of GalR D-galactose is not part of a metabolic reaction in the GENSOR unit, but a subclass of D-galactose, α -D-galactopyranose, is found in a reaction that is part of the GENSOR unit. Similar cases were found for AraC, ExuR, FadR, GalR, GalS, RbsR, and TreR. In **Supplementary Table S3** we show the allosteric effectors and the metabolites occurring in other reactions in the GENSOR unit. For these cases, we consider the feedback to be supported by evidence even if it was not found computationally.
- 2) The second case is that of TFs that must be analyzed in connection with additional TFs in which the feedback is not automatically detected unless reactions from a second regulator or regulated TF are considered. This is the case for AllS, Cbl, and RhaR. Allantoin is the effector of both AllS and AllR, and AllR regulates *allS* as well as *allB*. Since the enzyme encoded by *allB* catalyzes the first step in the assimilation of allantoin, the missing feedback of AllS is recovered when considering that it is regulated by AllR. The allosteric effector of Cbl is adenosine 5'-phosphosulfate, which is synthesized by the enzyme sulfate adenylyltransferase, encoded by *cysD* and *cysN*; both genes are regulated by CysB, which in turn is regulated by Cbl. Thus, these CysB-regulated reactions have to be considered to identify the feedback of Cbl. For RhaR, the allosteric effector is α -L-rhamnopyranose, a subclass of L-rhamnose, whose transport is mediated by RhaT, transcriptionally regulated by RhaS, which in turn is regulated by RhaR. Thus, the biological feedback is clear but is mediated by an additional TF.
- 3) We found two GENSOR units in which the feedback was not automatically identified for TFs that have an enzymatic function, BirA and DnaA. Up to now these additional activities of TFs have not been included in the GENSOR units. BirA is a DNA-binding transcriptional repressor and a biotin-[acetyl-CoA-carboxylase] ligase, and its allosteric effector is biotinyl-5'-adenylate. BirA negatively regulates BioB, which synthesizes biotin, and biotin is used by BirA to produce biotinyl-5'-adenylate. When biotin is in excess, it is transformed to biotinyl-5'-adenylate by BirA, and BirA-biotinyl-5'-adenylate negatively regulates *bioB* involved in synthesis of biotin (Sirithanakorn and Cronan, 2021). Then the feedback is evidenced when this additional activity of BirA is taken into consideration. On the other hand, DnaA is a transcriptional dual regulator and a chromosomal replication initiator protein with ATPase activity. The allosteric effector of DnaA is ATP. When DnaA interacts with two homodimers of Had, the DnaA-bound ATP is hydrolyzed (Su'etsugu et al., 2005). Hence, once the hydrolysis reaction is included, the feedback becomes evident.
- 4) Finally, in other cases identifying the feedback is hard because of the limited knowledge for the components of the GENSOR units, and no clear conclusion can be made. We consider this to be the case for MarR, PurR, PyrR, and ComR. In the case of

the MarR GENSOR unit, no enzyme has been reported in the genome to produce, consume, or transport salicylate, which is the allosteric effector. PurR regulates enzymes whose reactions are involved in IMP, a precursor of hypoxanthine, its allosteric effector. ComR regulates BhsA, a multiple stress resistance outer membrane protein that appears to reduce the outer membrane permeability to copper (Mermod et al., 2012), its allosteric effector. Once this process is characterized and a reaction of inhibition of copper uptake is annotated, it will be possible to identify automatically the feedback for ComR. Finally, the only target gene identified to date for PyrR (whose effector is pyruvate) is YhjX, an ABC transporter of unknown function. It has been reported that YhjX forms hetero-oligomers with YjiY, a pyruvate transporter; however, it is not yet known if YhjX affects pyruvate transport by YjiY (Behr et al., 2014).

These detailed analyses show that all GENSOR units have evidence of feedback, except for four cases which lack sufficient data (**Supplementary Table S2**). The computational program used to identify feedback is publicly available (<https://github.com/PGC-CCG/Feedback-in-GUs>) and can be used to search for a metabolite in any collection of reactions. The detailed curation work that expanded the feedback discovery illustrates both the complexities in the biology of signaling processing and the concomitant difficulties in the adequate representation of this knowledge in databases. Our analysis here was limited to TFs with allosteric effectors, and a similar analysis should be done with all other mechanisms depicted in **Figure 4**.

Biosensors Based on TFs

E. coli, in addition to being a model for the study of microbial physiology, has been widely used as a “chassis” for synthesis of valuable compounds through metabolic engineering. The integration of genetic regulation knowledge during metabolic engineering design is mandatory to improve the titer, yield, and productivity to afford the processes that are economically feasible. In particular, TFs modulated by signal metabolites are considered valuable tools for metabolic pathway engineering, since they can be used for monitoring or to control metabolic fluxes in the production of biofuels, organic acids, polymer precursors, and drugs, among others (Li et al., 2020).

Whole-cell TF-based biosensors have been constructed with a TF sensing a chemical compound modulating a reporter gene (Fernandez-López et al., 2015). These systems have been employed for high-throughput library screening to select new bacterial strains with desired metabolic characteristics (Kaczmarek and Prather, 2021). Similarly, TF-based biosensors can be used to monitor organic and inorganic pollutants in industrial sewage or in the environment. The high sensitivity of TFs enables the detection of pollutants in concentrations below the limit of some analytic techniques, quickly and cheaply. Examples include cadmium and mercury sensing by CadR (Tao et al., 2013). Manipulation of transcriptional regulation to improve the yield of an engineered system can be achieved in several ways, for example, by constructing hybrid promoters to reduce the concentration of toxic intermediaries in synthetic

pathways (Zhang et al., 2012) or by increasing the flexibility of the engineered process. A successful case of this strategy was reported by Zhang et al. (Zhang et al., 2012), in which the yield and titer of biodiesel production by a bacterial strain were improved. The strain was previously engineered to overexpress ethanol (precursor), fatty acids (precursors) and fatty acid ethyl esters (biodiesel) biosynthesis pathways. In that report, a synthetic promoter that includes a 17-bp FadR-binding DNA sequence was cloned upstream of *adhB* and *pdc* (encoding the enzymes for the ethanol biosynthesis), resulting in their repression by FadR. As a consequence, only when acyl-CoAs are present, the ethanol biosynthetic pathway is transcriptionally derepressed and biodiesel synthesis is carried out, thus avoiding high concentrations of ethanol in the cell, improving the stability of the strain (Zhang et al., 2012).

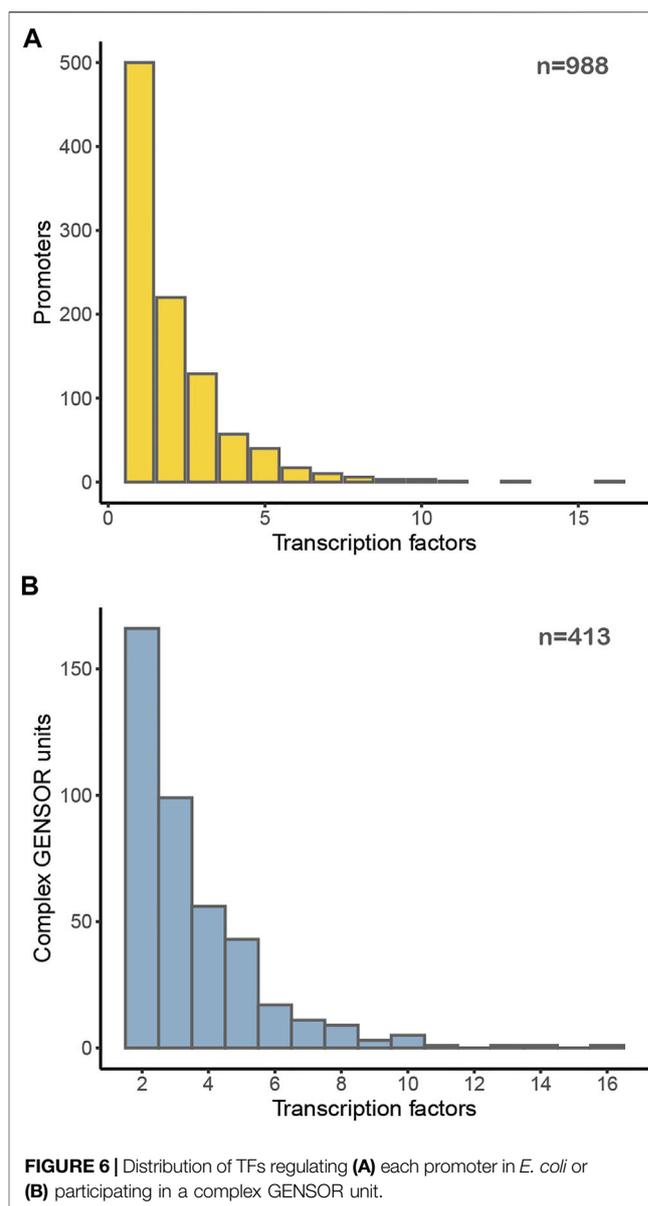
Supplementary Table S1 in the supplementary material contains the list of effectors for the 90 TFs with currently characterized noncovalent effectors. We consider this repertoire of TFs involved in naturally sensing different metabolites as a valuable resource for bioengineering purposes. Moreover, GENSOR units could be a reference point from which to search key compounds of biosynthetic pathways. If the signal, effector, or an intermediate of a GENSOR unit is present in the process of interest, the transcriptional regulator can be integrated into the model for monitoring or directing metabolic fluxes.

The implementation of TFs as biosensors has shown to be useful in stabilizing the yield for the production of several compounds, however, to date this biotechnological application is limited to well characterized TF-effector regulatory interactions. Furthermore, the knowledge of transcriptional regulatory elements in *E. coli* and other bacteria is rapidly expanding with technologies to identify the binding of TFs in the whole genome (Santos-Zavaleta et al., 2018), increasing for instance, the number of TFs with known binding sites in *E. coli* to 189 TFs (Gao et al., 2021). Future progress should expand to eventual completion of the TF-gene regulatory repertoire of interactions, enabling the definition of complete GENSOR units; this information shall improve the use of *E. coli* as a chassis in synthetic biology. Also, insights into molecular mechanisms governing allosteric interaction and regulatory effects could be employed to engineer allosteric TFs to sense new compounds, which could allow the use of TF-based biosensors in a wider range of biosynthetic processes (Tao et al., 2013; Taylor et al., 2016).

Most applications have exploited and implemented one signal-sensing system. As mentioned before, the GENSOR units we have discussed are constructed around a single TF. We know, however, that in addition to the predominantly one-to-one relationship between signals and TFs (Figure 2), transcriptional regulation offers an additional layer of integration at the level of promoters.

Genomic Processing of Multiple Signals in *E. coli*

Transcriptional regulators bind around one or multiple promoter regions to control the expression of the downstream gene or collection of genes organized in polycistronic units in bacteria. The GENSOR units we have discussed were built using the group



of genes that have a binding site for a particular TF in their upstream promoter region, that is, a simple regulon. However, the architecture of transcription can be rather complex with multiple nearby promoters upstream of genes and multiple binding sites for different TFs (Mejía-Almonte et al., 2020). Given the finite number of TFs and promoters in a genome, different productive approaches have modeled gene regulation from a combinatorial perspective, either with grammatical models to describe the multiple combinations of sites for different TFs in a promoter region (Collado-Vides, 1992), in quantitative thermodynamic models (Bintu et al., 2005a; Bintu et al., 2005b), or in combinatorial logic modules of gates of regulation (Buchler et al., 2003).

We can think of this genomic organization as supporting the collection of decisions genetically encoded in the genome. An

interesting question is what is the number of signals that the genetically encoded collection of decisions integrates. In the first layer, that of the number of effectors binding to TFs (Figure 3), it is hard to say it conveys integration, since in the reduced number of cases of multiple effectors, it is not clear if multiple binding happens under the same growing conditions. Thus, we consider that, most likely, there is no major integration of signals in decision processes at this level. Figure 6A, however, shows that approximately 50% of promoters are subject to regulation by multiple TFs, with a distribution of up to ~10 different TFs affecting the same promoter. Although we cannot say that in all these cases the different TFs simultaneously affect promoters or if they bind separately under different conditions, in a good number of well-studied cases it has been shown that multiple TFs work together in affecting promoter activity. Projects requiring quantitative approaches could benefit from the combinatorial repertoires in the approaches we have mentioned before, those of the Terry Hwa (Buchler et al., 2003) and Rob Phillips (Bintu et al., 2005a; Bintu et al., 2005b) laboratories. Figure 6B shows the number of complex GENSOR units, defined based on complex regulons and their jointly regulated target genes. Genetic decisions at the level of transcription initiation can integrate the simultaneous occurrence of multiple signals, as illustrated by carbon source decisions involving the absence of glucose mediated by the catabolite repressor CRP and the presence of another sugar (e.g., lactose, maltose, or arabinose with LacI, MalT, or AraC, respectively), to mention a few well-known cases.

We consider complex GENSOR units to better reflect the activity of decision-making in *E. coli* than elementary GENSOR units, as they are derived from the organization of TFs regulating promoters in the genome. This is further supported by the increased correlation with co-expression of complex GENSOR units at a level similar to that of transcription units (see Figure 4 in (Ledezma-Tejeida et al., 2019)).

Despite these warnings, Figure 6B shows our best approximation to the distribution of the number of different signals that regulatory TF-based decisions are integrated by these sensing processes.

DISCUSSION

The transcriptional network of the *E. coli* genome is estimated to involve ~300 TFs (Pérez-Rueda and Collado-Vides, 2000; Madan Babu and Teichmann, 2003), of which we have experimental evidence for 221 TFs and evidence of allosteric interactions for 90 of them (40%). Recalling that two-component systems also have an allosteric component, this number could increase to 118 (53%). However, sequence-based computational predictions estimate that 75% of the complete set of TFs bind small molecules (Madan Babu and Teichmann, 2003), a higher fraction than what is currently experimentally supported.

It is remarkable to note the few number of allosteric effectors that bind overall to TFs, as shown in Figure 4, with 1.16 effectors per TF on average. It is possible that this number is greatly underestimating the TF-metabolite interactions that happen *in*

vivo, since, as already mentioned, these numbers are limited to the effectors for which there is evidence of a regulatory interaction. A recent *in vivo* systematic analysis combining transcriptomics and metabolomics identified new effectors for 30 TFs, in many cases increasing the number of known effectors for a given TF (Lempp et al., 2019). Although high-throughput analysis inherently identify false positives and the authors of the study acknowledge that a disadvantage of their method is the false identification of metabolites with similar dynamics to the true effectors, studies in enzyme-metabolite interactions have also identified a large number of potential ligands *in vivo* (Piazza et al., 2018). The compendia of TF-metabolite interactions presented here is backed up by *in vitro* experimental evidence, it is likely that the list will grow as more systematic TF-metabolite identification studies appear, and the challenge will shift to identifying those that are functionally relevant. Alternatively, TFs might have been selected to have a reduced number of allosteric interactions favoring the one-to-one mediated information flux from signals to promoters as opposed to enzymes (Huang et al., 2011).

As we have discussed, this implies that the integration of multiple signals happens in bacteria, predominantly, via the binding of multiple TFs governing the level of individual promoters. Transcriptional regulation is also governed by TFs that change their conformation via different mechanisms, as schematized in Figure 4. The mechanisms can be rather complex, like the multiple cascades of enzymatic transformation in two-component systems (van Heeswijk et al., 2013; Groisman, 2016). Also as mentioned, other TFs seem to have no multiplicity of conformations but are regulated directly by their synthesis, like IHF. We are aware that this variability supports a diverse dynamic in the signaling processes used in *E. coli*.

We have argued that sensing is an essential aspect of the larger context of gene regulation when considering those cases that evolution has generated to enable the cell to address changes in the environment. The genetic sensory response units, or GENSOR units, we previously assembled fit perfectly with this perspective. GENSOR units are multilayer in the sense of grouping transport, signaling, gene regulation, the regulated gene products, and either enzymes or any other product, as well as the reactions they encode, and the metabolites involved. They are preprocessed constructions based on the set of regulated genes that comprise the regulon of individual TFs. Their construction offers the highest level of integration currently available of major groups of co-regulated genes and their gene products. We have generated short summaries for several GENSOR units, describing the function of the regulated genes.

As mentioned already, elementary GENSOR units have enabled evaluations of the partial homogeneity of the functional putative response of regulons. We say “putative,” assuming all genes of a regulon are co-expressed, which we know is not the case. In that sense, complex GENSOR units define groups of genes with a much higher co-expression correlation, as mentioned before. Nonetheless, there is room for further work both experimentally as well as in the bio-curation of the regulatory modules or phrases governing complex GENSOR units to gather evidence and distinguish promoters subject to

distinct regulation under different conditions, from promoters subject to simultaneous regulation by multiple TFs.

Elementary GENSOR units were used to confirm the hypothesis that feedback is inherent to regulatory circuits devoted to address changes in the environment. As mentioned in methods, we updated the holo TF conformations, so that now they are consistently using the main name of the bound effector, enabling computing with this knowledge. These changes will be uploaded in these databases, and the list of all allosteric effectors is included in the supplementary material. As mentioned, we automatically identified feedback in 82% of the cases and manually curated the remaining cases, identifying the presence of feedback in all of those with available information. Although this analysis was limited to allosteric-mediated regulation, to the extent that all TFs are part of the sensing capability of the cell, we believe that feedback is equally expected irrespective of the signaling mechanism. A different scenario could be imagined for gene regulation in, for instance, developmental programs, where feedback to the signal might not be logically required.

The accumulated knowledge of sensing and regulated circuits has provided the basis to develop many biotechnological applications, exploiting also the rich promoter architectures, as they are based on the identification and rational manipulation of those circuits or parts of them (Zhang et al., 2012; Tao et al., 2013; Taylor et al., 2016; Graham et al., 2020; Li et al., 2020). Genome identification of whole-genome transcriptional regulatory networks may have a concomitant impact in a new generation of biotechnological applications exploiting more complex metabolic and regulatory circuits.

MATERIALS AND METHODS

GENSOR Unit Assembly

GENSOR units from the regulatory network in the format “TF-gene” available in RegulonDB v10.9 were assembled using the pipeline described in Ledezma-Tejeida, NAR, 2019, using custom Perl scripts available at (<https://github.com/PGC-CCG/Feedback-in-GUs>) (Gensor option in the Gensor-Unit pipeline), originally published in https://github.com/dledezma/gensor_units.

Curation of TF Conformations

In order to computationally identify the feedback, we revised the annotated active conformations of TFs to precisely name *holo* TF conformations using the main name of the bound effector, and not a synonym or short name as has been done before. These changes will be uploaded in these databases, and the list of all allosteric effectors is included in the supplementary material.

Identification of Feedback in GUs

The feedback loops were identified directly from the GENSOR units assembled in the previous step. We then used a custom script (feedback option of the Gensor unit pipeline available at (<https://github.com/PGC-CCG/Feedback-in-GUs>)) that was incorporated as input for the curated active and inactive TF conformations from RegulonDB v10.9 to define the metabolic effector or effectors that interact with each TF. Each GENSOR unit was then scanned for the presence of the effector or effectors as reactants and/or products in

its reactions. Feedback was assigned to a TF if one or more of its effectors were found in its regulated metabolic reactions. For some TFs, the specific stereoisomers of the effector molecules were not found in the reactions in their exact form, but a lower- or upper-class compound were. In these cases, the program suggests possible matches in feedback, as there could be a lack of knowledge of their specific spatial conformation. These cases, however, must be curated manually.

The manual curation was carried out by analyzing the compounds in non-allosteric reactions from the GENSOR unit. The reactions in which the effector is present were retrieved, and then the regulation over the genes encoding the corresponding enzymes was analyzed.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://github.com/PGC-CCG/Feedback-in-GUs>.

AUTHOR CONTRIBUTIONS

GF developed the algorithm to identify feedback in GENSOR Units, PL and SG-C performed all manual curation, DL-T and VT created the figures, LM and CB-M provided technical assistance and recovered data from RegulonDB and Ecocyc. JC-V suggested the overall perspective and supervised the research; GF, PL, SG-C, DL-T, VT, and JC-V jointly discussed and wrote the manuscript.

FUNDING

Funding for this work came from Universidad Nacional Autónoma de México (UNAM) and by NIGMS-NIH grant numbers 5RO1GM131643 and 2R01GM077678. Funding for open access publication fees comes from NIGMS-NIH grant 5RO1GM131643. We acknowledge funding from Universidad Nacional Autónoma de México (UNAM) and by NIGMS-NIH grant numbers 5RO1GM131643 and 2R01GM077678.

ACKNOWLEDGMENTS

The authors thank Víctor Del Moral, Joel Gómez Espindola, and Alfredo Hernández Alvarez for their technical support. JC-V acknowledges previous support of a sabbatical in Barcelona by PASPA DGAPA-UNAM. GF acknowledges CONACYT. DL-T acknowledges funding through an ETH Postdoctoral Fellowship and a grant of the Botnar Child Care Foundation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2022.823240/full#supplementary-material>

REFERENCES

- Ali Azam, T., Iwata, A., Nishimura, A., Ueda, S., and Ishihama, A. (1999). Growth Phase-dependent Variation in Protein Composition of the *Escherichia coli* Nucleoid. *J. Bacteriol.* 181, 6361–6370. doi:10.1128/jb.181.20.6361-6370.1999
- Ansaldi, M., Théraulaz, L., and Méjean, V. (2004). TorI, a Response Regulator Inhibitor of Phage Origin in *Escherichia coli*. *Proc. Natl. Acad. Sci.* 101, 9423–9428. doi:10.1073/pnas.0401927101
- Anzai, T., Imamura, S., Ishihama, A., and Shimada, T. (2020). Expanded Roles of Pyruvate-Sensing PdhR in Transcription Regulation of the *Escherichia coli* K-12 Genome: Fatty Acid Catabolism and Cell Motility. *Microb. Genom.* 6, mgen000442. doi:10.1099/mgen.0.000442
- Aquino, P., Honda, B., Jaini, S., Lyubetskaya, A., Hosur, K., Chiu, J. G., et al. (2017). Coordinated Regulation of Acid Resistance in *Escherichia coli*. *BMC Syst. Biol.* 11 (1), 1. doi:10.1186/s12918-016-0376-y
- Azam, T. A., and Ishihama, A. (1999). Twelve Species of the Nucleoid-Associated Protein from *Escherichia coli*. *J. Biol. Chem.* 274, 33105–33113. doi:10.1074/jbc.274.46.33105
- Balderas-Martínez, Y. I., Savageau, M., Salgado, H., Pérez-Rueda, E., Morett, E., and Collado-Vides, J. (2013). Transcription Factors in *Escherichia coli* Prefer the Holo Conformation. *PLoS One* 8, e65723. doi:10.1371/journal.pone.0065723
- Behr, S., Fried, L., and Jung, K. (2014). Identification of a Novel Nutrient-Sensing Histidine Kinase/response Regulator Network in *Escherichia coli*. *J. Bacteriol.* 196, 2023–2029. doi:10.1128/jb.01554-14
- Bintu, L., Buchler, N. E., García, H. G., Gerland, U., Hwa, T., Kondev, J., et al. (2005). Transcriptional Regulation by the Numbers: Applications. *Curr. Opin. Genet. Develop.* 15, 125–135. doi:10.1016/j.gde.2005.02.006
- Bintu, L., Buchler, N. E., García, H. G., Gerland, U., Hwa, T., Kondev, J., et al. (2005). Transcriptional Regulation by the Numbers: Models. *Curr. Opin. Genet. Develop.* 15, 116–124. doi:10.1016/j.gde.2005.02.007
- Buchler, N. E., Gerland, U., and Hwa, T. (2003). On Schemes of Combinatorial Transcription Logic. *Proc. Natl. Acad. Sci.* 100, 5136–5141. doi:10.1073/pnas.0930314100
- Changela, A., Chen, K., Xue, Y., Holschen, J., Outten, C. E., O'Halloran, T. V., et al. (2003). Molecular Basis of Metal-Ion Selectivity and Zeptomolar Sensitivity by CueR. *Science* 301, 1383–1387. doi:10.1126/science.1085950
- Chen, S., and Calvo, J. M. (2002). Leucine-induced Dissociation of *Escherichia coli* Lrp Hexadecamers to Octamers. *J. Mol. Biol.* 318, 1031–1042. doi:10.1016/S0022-2836(02)00187-0
- Chen, S., Hao, Z., Bieniek, E., and Calvo, J. M. (2001). Modulation of Lrp Action in *Escherichia coli* by Leucine: Effects on Non-specific Binding of Lrp to DNA. *J. Mol. Biol.* 314, 1067–1075. doi:10.1006/jmbi.2000.5209
- Collado-Vides, J. (1992). Grammatical Model of the Regulation of Gene Expression. *Proc. Natl. Acad. Sci.* 89, 9405–9409. doi:10.1073/pnas.89.20.9405
- Fernandez-López, R., Ruiz, R., de la Cruz, F., and Moncalián, G. (2015). Transcription Factor-Based Biosensors Enlightened by the Analyte. *Front. Microbiol.* 6, 648. doi:10.3389/fmicb.2015.00648
- Gama-Castro, S., Salgado, H., Santos-Zavaleta, A., Ledezma-Tejeda, D., Muñoz-Rascado, L., García-Sotelo, J. S., et al. (2016). RegulonDB Version 9.0: High-Level Integration of Gene Regulation, Coexpression, Motif Clustering and beyond. *Nucleic Acids Res.* 44, D133–D143. doi:10.1093/nar/gkv1156
- Gao, Y., Lim, H. G., Verkler, H., Szubin, R., Quach, D., Rodionova, I., et al. (2021). Unraveling the Functions of Uncharacterized Transcription Factors in *Escherichia coli* Using ChIP-Exo. *Nucleic Acids Res.* 49, 9696–9710. doi:10.1093/nar/gkab735
- 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. doi:10.1038/nrmicro1932
- Graham, G., Csicsery, N., Stasiowski, E., Thouvenin, G., Mather, W. H., Ferry, M., et al. (2020). Genome-scale Transcriptional Dynamics and Environmental Biosensing. *Proc. Natl. Acad. Sci. USA* 117, 3301–3306. doi:10.1073/pnas.1913003117
- Groisman, E. A. (2016). Feedback Control of Two-Component Regulatory Systems. *Annu. Rev. Microbiol.* 70, 103–124. doi:10.1146/annurev-micro-102215-095331
- Hart, B. R., and Blumenthal, R. M. (2011). Unexpected Coregulator Range for the Global Regulator Lrp of *Escherichia coli* and *Proteus mirabilis*. *J. Bacteriol.* 193, 1054–1064. doi:10.1128/jb.01183-10
- Hu, L. I., Chi, B. K., Kuhn, M. L., Filippova, E. V., Walker-Peddakotla, A. J., Basell, K., et al. (2013). Acetylation of the Response Regulator RcsB Controls Transcription from a Small RNA Promoter. *J. Bacteriol.* 195, 4174–4186. doi:10.1128/jb.00383-13
- Huang, Z., Zhu, L., Cao, Y., Wu, G., Liu, X., Chen, Y., et al. (2011). ASD: a Comprehensive Database of Allosteric Proteins and Modulators. *Nucleic Acids Res.* 39, D663–D669. doi:10.1093/nar/gkq1022
- Huerta, A., Salgado, H., Thieffry, D., and Collado-Vides, J. (1998). RegulonDB: a Database on Transcriptional Regulation in *Escherichia coli*. *Nucleic Acids Res.* 26, 55–59. doi:10.1093/nar/26.1.55
- Jeon, Y., Lee, Y. S., Han, J. S., Kim, J. B., and Hwang, D. S. (2001). Multimerization of Phosphorylated and Non-phosphorylated ArcA Is Necessary for the Response Regulator Function of the Arc Two-Component Signal Transduction System. *J. Biol. Chem.* 276, 40873–40879. doi:10.1074/jbc.m104855200
- Joly, N., Danot, O., Schlegel, A., Boos, W., and Richet, E. (2002). The Aes Protein Directly Controls the Activity of MalT, the Central Transcriptional Activator of the *Escherichia coli* Maltose Regulon. *J. Biol. Chem.* 277, 16606–16613. doi:10.1074/jbc.m200991200
- Ju, X., Li, D., and Liu, S. (2019). Full-length RNA Profiling Reveals Pervasive Bidirectional Transcription Terminators in Bacteria. *Nat. Microbiol.* 4, 1907–1918. doi:10.1038/s41564-019-0500-z
- Kaczmarek, J. A., and Prather, K. L. J. (2021). Effective Use of Biosensors for High-Throughput Library Screening for Metabolite Production. *J. Ind. Microbiol. Biotechnol.* 48, kuab049. doi:10.1093/jimb/kuab049
- Keseler, I. M., Gama-Castro, S., Mackie, A., Billington, R., Bonavides-Martínez, C., Caspi, R., et al. (2021). The EcoCyc Database in 2021. *Front. Microbiol.* 12, 711077. doi:10.3389/fmicb.2021.711077
- Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., et al. (2017). The EcoCyc Database: Reflecting New Knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* 45, D543–d550. doi:10.1093/nar/gkw1003
- Khoroshilova, N., Popescu, C., Münck, E., Beinert, H., and Kiley, P. J. (1997). Iron-sulfur Cluster Disassembly in the FNR Protein of *Escherichia coli* by O₂: [4Fe-4S] to [2Fe-2S] Conversion with Loss of Biological Activity. *Proc. Natl. Acad. Sci.* 94, 6087–6092. doi:10.1073/pnas.94.12.6087
- Kroner, G. M., Wolfe, M. B., and Freddolino, P. L. (2019). *Escherichia coli* Lrp Regulates One-Third of the Genome via Direct, Cooperative, and Indirect Routes. *J. Bacteriol.* 201, e00411. doi:10.1128/JB.00411-18
- Ledezma-Tejeda, D., Altamirano-Pacheco, L., Fajardo, V., and Collado-Vides, J. (2019). Limits to a Classic Paradigm: Most Transcription Factors in *E. coli* Regulate Genes Involved in Multiple Biological Processes. *Nucleic Acids Res.* 47, 6656–6667. doi:10.1093/nar/gkz525
- Ledezma-Tejeda, D., Ishida, C., and Collado-Vides, J. (2017). Genome-Wide Mapping of Transcriptional Regulation and Metabolism Describes Information-Processing Units in *Escherichia coli*. *Front. Microbiol.* 8, 1466. doi:10.3389/fmicb.2017.01466
- Lee, S.-J., Boos, W., Bouché, J. P., and Plumbridge, J. (2000). Signal Transduction between a Membrane-Bound Transporter, PtsG, and a Soluble Transcription Factor, Mlc, of *Escherichia coli*. *Embo j* 19, 5353–5361. doi:10.1093/emboj/19.20.5353
- Leitch, S., Bradley, M. J., Rowe, J. L., Chivers, P. T., and Maroney, M. J. (2007). Nickel-specific Response in the Transcriptional Regulator, *Escherichia coli* NikR. *J. Am. Chem. Soc.* 129, 5085–5095. doi:10.1021/ja068505y
- Lempp, M., Farke, N., Kuntz, M., Freibert, S. A., Lill, R., and Link, H. (2019). Systematic Identification of Metabolites Controlling Gene Expression in *E. coli*. *Nat. Commun.* 10, 4463. doi:10.1038/s41467-019-12474-1
- Li, J. W., Zhang, X. Y., Wu, H., and Bai, Y. P. (2020). Transcription Factor Engineering for High-Throughput Strain Evolution and Organic Acid Bioproduction: A Review. *Front. Bioeng. Biotechnol.* 8 (98), 98. doi:10.3389/fbioe.2020.00098
- Lorca, G. L., Ezersky, A., Lunin, V. V., Walker, J. R., Altamentova, S., Evdokimova, E., et al. (2007). Glyoxylate and Pyruvate Are Antagonistic Effectors of the *Escherichia coli* IclR Transcriptional Regulator. *J. Biol. Chem.* 282, 16476–16491. doi:10.1074/jbc.m610838200

- Madan Babu, M., and Teichmann, S. A. (2003). Evolution of Transcription Factors and the Gene Regulatory Network in *Escherichia coli*. *Nucleic Acids Res.* 31, 1234–1244. doi:10.1093/nar/gkg210
- Magasanik, B. (2000). Global Regulation of Gene Expression. *Proc. Natl. Acad. Sci.* 97, 14044–14045. doi:10.1073/pnas.97.26.14044
- Mandrich, L., Caputo, E., Martin, B. M., Rossi, M., and Manco, G. (2002). The Aes Protein and the Monomeric α -Galactosidase from *Escherichia coli* Form a Non-covalent Complex. *J. Biol. Chem.* 277, 48241–48247. doi:10.1074/jbc.m207398200
- Martínez-Antonio, A., Janga, S. C., Salgado, H., and Collado-Vides, J. (2006). Internal-sensing Machinery Directs the Activity of the Regulatory Network in *Escherichia coli*. *Trends Microbiol.* 14, 22–27. doi:10.1016/j.tim.2005.11.002
- Mejía-Almonte, C., Busby, S. J. W., Wade, J. T., van Helden, J., Arkin, A. P., Stormo, G. D., et al. (2020). Redefining Fundamental Concepts of Transcription Initiation in Bacteria. *Nat. Rev. Genet.* 21, 699–714. doi:10.1038/s41576-020-0254-8
- Mermod, M., Magnani, D., Solioz, M., and Stoyanov, J. V. (2012). The Copper-Inducible ComR (YcfQ) Repressor Regulates Expression of ComC (YcfR), Which Affects Copper Permeability of the Outer Membrane of *Escherichia coli*. *Biomaterials* 25, 33–43. doi:10.1007/s10534-011-9510-x
- Milanesio, P., Arce-Rodríguez, A., Muñoz, A., Calles, B., and de Lorenzo, V. (2011). Regulatory Exaptation of the Catabolite Repression Protein (Crp)-cAMP System in *Pseudomonas putida*. *Environ. Microbiol.* 13, 324–339. doi:10.1111/j.1462-2920.2010.02331.x
- Nam, T.-W., Cho, S. H., Shin, D., Kim, J. H., Jeong, J. Y., Lee, J. H., et al. (2001). The *Escherichia coli* Glucose Transporter Enzyme IICB_{glc} Recruits the Global Repressor Mlc. *Embo j* 20, 491–498. doi:10.1093/emboj/20.3.491
- Pardee, A. B., Jacob, F., and Monod, J. (1958). The Role of the Inducible Alleles and the Constitutive Alleles in the Synthesis of Beta-Galactosidase in Zygotes of *Escherichia coli*. *C. R. Hebd Seances Acad. Sci.* 246, 3125–3128.
- Pérez-Rueda, E., and Collado-Vides, J. (2000). The Repertoire of DNA-Binding Transcriptional Regulators in *Escherichia coli* K-12. *Nucleic Acids Res.* 28, 1838–1847. doi:10.1093/nar/28.8.1838
- Piazza, I., Kochanowski, K., Cappelletti, V., Fuhrer, T., Noor, E., Sauer, U., et al. (2018). A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* 172, 358–372. doi:10.1016/j.cell.2017.12.006
- Santos-Zavaleta, A., Sánchez-Pérez, M., Salgado, H., Velázquez-Ramírez, D. A., Gama-Castro, S., Tierrafria, V. H., et al. (2018). A Unified Resource for Transcriptional Regulation in *Escherichia coli* K-12 Incorporating High-Throughput-Generated Binding Data into RegulonDB Version 10.0. *BMC Biol.* 16 (91), 91. doi:10.1186/s12915-018-0555-y
- Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-Romero, L., Ledezma-Tejeda, D., et al. (2019). RegulonDB V 10.5: Tackling Challenges to Unify Classic and High Throughput Knowledge of Gene Regulation in *E. coli* K-12. *Nucleic Acids Res.* 47, D212–d220. doi:10.1093/nar/gky1077
- Schlegel, A., Danot, O., Richet, E., Ferenci, T., and Boos, W. (2002). The N Terminus of the *Escherichia coli* Transcription Activator MalT Is the Domain of Interaction with MalY. *J. Bacteriol.* 184, 3069–3077. doi:10.1128/jb.184.11.3069-3077.2002
- Schleif, R. (2010). AraC Protein, Regulation of the L-Arabinose Operon in *Escherichia coli*, and the Light Switch Mechanism of AraC Action. *FEMS Microbiol. Rev.* 34, 779–796. doi:10.1111/j.1574-6976.2010.00226.x
- Schreiber, V., Steegborn, C., Clausen, T., Boos, W., and Richet, E. (2000). A New Mechanism for the Control of a Prokaryotic Transcriptional Regulator: Antagonistic Binding of Positive and Negative Effectors. *Mol. Microbiol.* 35, 765–776. doi:10.1046/j.1365-2958.2000.01747.x
- Seo, S. W., Kim, D., Latif, H., O'Brien, E. J., Szubin, R., and Palsson, B. O. (2014). Deciphering Fur Transcriptional Regulatory Network Highlights its Complex Role beyond Iron Metabolism in *Escherichia coli*. *Nat. Commun.* 5, 4910. doi:10.1038/ncomms5910
- Sirithanakorn, C., and Cronan, J. E. (2021). Biotin, a Universal and Essential Cofactor: Synthesis, Ligation and Regulation. *FEMS Microbiol. Rev.* 45, fuab003. doi:10.1093/femsre/fuab003
- Su'tsugu, M., Shimuta, T.-r., Ishida, T., Kawakami, H., and Katayama, T. (2005). Protein Associations in DnaA-ATP Hydrolysis Mediated by the Hda-Replicase Clamp Complex. *J. Biol. Chem.* 280, 6528–6536. doi:10.1074/jbc.m412060200
- Takahashi, K., Kawazoe, Y., Sakumi, K., Nakabeppu, Y., and Sekiguchi, M. (1988). Activation of Ada Protein as a Transcriptional Regulator by Direct Alkylation with Methylating Agents. *J. Biol. Chem.* 263, 13490–13492. doi:10.1016/s0021-9258(18)68265-1
- Tao, H.-C., Peng, Z.-W., Li, P.-S., Yu, T.-A., and Su, J. (2013). Optimizing Cadmium and Mercury Specificity of CadR-Based *E. coli* Biosensors by Redesign of CadR. *Biotechnol. Lett.* 35, 1253–1258. doi:10.1007/s10529-013-1216-4
- Taylor, N. D., Garruss, A. S., Moretti, R., Chan, S., Arbing, M. A., Cascio, D., et al. (2016). Engineering an Allosteric Transcription Factor to Respond to New Ligands. *Nat. Methods* 13, 177–183. doi:10.1038/nmeth.3696
- van Heeswijk, W. C., Westerhoff, H. V., and Booger, F. C. (2013). Nitrogen Assimilation in *Escherichia coli*: Putting Molecular Data into a Systems Perspective. *Microbiol. Mol. Biol. Rev.* 77, 628–695. doi:10.1128/mmr.00025-13
- Wall, M. E., Hlavacek, W. S., and Savageau, M. A. (2004). Design of Gene Circuits: Lessons from Bacteria. *Nat. Rev. Genet.* 5, 34–42. doi:10.1038/nrg1244
- Yan, B., Boitano, M., Clark, T. A., and Ettwiller, L. (2018). SMRT-Cappable-seq Reveals Complex Operon Variants in Bacteria. *Nat. Commun.* 9, 3676. doi:10.1038/s41467-018-05997-6
- Zhang, F., Carothers, J. M., and Keasling, J. D. (2012). Design of a Dynamic Sensor-Regulator System for Production of Chemicals and Fuels Derived from Fatty Acids. *Nat. Biotechnol.* 30, 354–359. doi:10.1038/nbt.2149
- Zimmer, D. P., Soupene, E., Lee, H. L., Wendisch, V. F., Khodursky, A. B., Peter, B. J., et al. (2000). Nitrogen Regulatory Protein C-Controlled Genes of *Escherichia coli*: Scavenging as a Defense against Nitrogen Limitation. *Proc. Natl. Acad. Sci.* 97, 14674–14679. doi:10.1073/pnas.97.26.14674

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

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Femerling, Gama-Castro, Lara, Ledezma-Tejeda, Tierrafria, Muñoz-Rascado, Bonavides-Martínez and Collado-Vides. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.