Internal-sensing machinery directs the activity of the regulatory network in *Escherichia coli*

Agustino Martínez-Antonio, Sarath Chandra Janga, Heladia Salgado and Julio Collado-Vides

Programa de Genómica Computacional, Centro de Ciencias Genómicas, UNAM, Cuernavaca, Morelos, 62100, México

Individual cells need to discern and synchronize transcriptional responses according to variations in external and internal conditions. Metabolites and chemical compounds are sensed by transcription factors (TFs), which direct the corresponding specific transcriptional responses. We propose a classification of the currently known TFs of *Escherichia coli* based on whether they respond to metabolites incorporated from the exterior, to internally produced compounds, or to both. When analyzing the mutual interactions of TFs, the dominant role of internal signal sensing becomes apparent, greatly due to the role of global regulators of transcription. This work encompasses metabolite–TF interactions, bridging the gap between the metabolic and regulatory networks, thus advancing towards an integrated network model for the understanding of cellular behavior.

Introduction

Biological systems need to monitor the environment and respond to changes. Gene expression is largely controlled at the level of transcription by transcription factors (TFs). In addition to a DNA-binding domain, TFs have structural domains for covalent or allosteric modifications such as the binding of specific metabolites or chemical signaling molecules [1–5]. Specific signals determine the conformation of the TF that interacts with particular DNA-binding sites or with the rest of the transcriptional machinery [6,7]. Thus, metabolites that affect TFs constitute common signals for endogenous and exogenous changes and are the links between the genetic and metabolic networks. A specific example of this is a global study in yeast that revealed the differential patterns of expression caused by endogenous and exogenous environmental changes [8].

Here, we propose a classification of TFs based on the type of signals and direct-binding metabolites that TFs detect, specifically, whether such signals are exogenous, endogenous or a combination of both. Although it is primarily focused on *Escherichia coli*, we believe that this classification might also work for other bacteria. Approximately 70% of the known regulatory interactions in *E. coli* involve TFs using signal metabolites to modulate their activities [9,10].

As far as we know, this is the first attempt towards a large-scale classification of sensing systems and transcriptional responses at the cellular level in bacteria. The mathematical description of such biologically more realistic regulatory networks represents a novel challenge. Our proposed classification also provides the foundations for understanding the dynamics and topology of the regulatory network as it senses changes in the environment.

Classification of transcription factors

TFs are ‘two-head molecules’ (as Francis Jacob called them): one head is the DNA-binding site and the other the allosteric site to which a metabolite binds non-covalently [11]. It is convenient to think in more abstract terms of meta-allosteric or generalized allosteric sites that include both the classic non-covalent binding and covalent chemical modifications (e.g. phosphorylation) because their roles are similar for physiological and topological descriptions of the network. These meta-allosteric components connect every TF with a metabolite or element of metabolism. Thus, as concentrations of metabolites change, they trigger an element of the regulatory network and turn genes on and off. Because the products of some of these genes are enzymes, the metabolic configuration of the cell can be modified. When these interactions are appreciated, it becomes clear that the topology of the machinery of metabolic and genetic network components is inherently circular, posing a challenge for its adequate mathematical modeling.

Sensing exogenous signals

Extracellular signals could be metabolites produced by other cells, secreted or excreted by-products generated by other organisms, chemical elements such as minerals, oxygen, protons and other ions, or changes in temperature and osmotic pressure. Extracellular signals are used, for example, as energy sources, as precursor materials of cellular components or as indicators of the environment (Figure 1a; see Supplementary Material). Unfortunately, even for a model organism such as *E. coli*, some environmental signals that trigger physiological...
responses are still unknown [12,13]. This is particularly true for two-component systems because there are many without a known trigger metabolite or condition; for these, the classification as ‘external’ was based on evolutionary considerations[14]. In addition, because two-component proteins function as a unit, the signals considered were those affecting the sensor component. Some transport systems in yeast can be considered as being related to sensing [15]. In E. coli, some transport systems directly affect the activity of TFs through protein–protein interactions. For instance, in the absence of maltose, MalT (activator of maltose use) is recruited by MalY (the maltose transporter), maintaining it in a monomeric inactive state [16]. Similarly, the localization of Mlc (repressor of MalT and the phosphotransferase system, PTS) varies depending on the phosphorylation of EIICBglc (glucose-specific permease of the PTS): when EIICBglc is not phosphorylated, Mlc is sequestered in the membrane, thus avoiding the exertion of its function in the cytoplasm [17,18].

Sensing endogenous signals

In our proposed scheme, TFs are classified based on their direct signal interaction. Thus, a TF can belong to the internal-sensing class even if the signal metabolite has been induced as a consequence of a change in the environment.

TFs binding to endogenous signal metabolites sense the intracellular status of cell machinery. Such metabolites can be involved in replication, or can function as energy metabolites and cofactors for enzyme catalysis or as internal alarmones (molecules that alert the cell to the onset of metabolic stress for regulation of gene expression and enzyme function; Figure 1b). In some cases, these same metabolites regulate, by allosterism, enzymes in the same or different pathways (e.g. pyruvate, phosphoenolpyruvate and fructose 1,6-bisphosphate). Most metabolites are synthesized by one or two enzymes, whereas pyruvate – biochemically the most connected metabolite [19,20] – is synthesized by 27 enzymes that participate in several biochemical pathways, which reflects its importance as a central molecule in metabolism. In some metabolic states, such as microaerobiosis or anaerobiosis, reducing power accumulates and could be lethal to the cell. This condition generates signals for TFs that sense the oxide-reduction state of the cell, such as FNR, SoxR and OxyR.

Sensing hybrid signals

Depending on growth conditions, some metabolites can be synthesized endogenously by the cell or transported and incorporated from the external environment. TFs that sense this type of metabolite are classed as ‘hybrid sensing’ (Figure 1c; see Supplementary Material). Of the 33 TFs in this class, 50% sense amino acids and the remainder sense sugars, modified sugars and carbohydrates from intermediate metabolism.

Influences of the sensing classes on the regulatory network

To understand the role of the different sensing classes from a topological perspective, it is necessary to determine how gene distribution is regulated by the different classes
Approximately 50% of the genes in *E. coli* are regulated by more than one TF [21]. Figure 2a shows the overlap of regulated operons as governed by internal (I; 29 TFs), external (E; 58 TFs) and hybrid (H; 33 TFs) classes of sensing. The numbers of operons regulated by TFs of each class are shown; numbers in parentheses are the corresponding percentages of regulated operons. **Figure 2a** Fraction of genes regulated by each sub-class of TFs that are also co-regulated by TFs of other sub-classes. E-TC, red bars; E-TM, green; I-DB, turquoise; I-SM, dark blue; H, yellow. **Figure 2b** Fraction of genes regulated by each class of TFs and co-regulated by a given number of other TFs. External-sensing, orange bars; internal-sensing, green; hybrid-sensing, blue. **Figure 2c** Transcriptional regulation of the genes encoding TFs in each class. The numbers represent the percent of genes in each set regulated by: auto-regulation (unbroken curved arrow) and regulation by TFs of the same class (broken curved arrow). Arrows between two classes represent the fraction of TFs (where the arrow ends) regulated by TFs of the other class (where the arrow originates). The percent of transcriptional regulation in each class is shown.

**Co-regulation between subclasses of TFs**

The external-sensing TFs of the two-component systems (E-TC) have a slightly higher tendency to co-regulate with internal-sensing TFs for DNA-binding (I-DB) (Figure 2b). This interesting observation suggests that, to regulate gene expression, the cell integrates the status of signals that are not transportable with the topology of the DNA. Similarly, external-sensing TFs using transported metabolites (E-TM) work together with TFs that sense metabolites generated internally (I-SM). Indeed, it is logical that regulation of substrate metabolites (by E-TM) might depend on the availability of their derived metabolites in the cell (sensed by I-SM). Interestingly, TFs of the hybrid class almost exclude co-regulation with TFs of the external class (E-TC and E-TM). In this respect, hybrid systems (mainly amino acid regulation) are modeled by the cell as the external system, with both systems requiring coordination with the internal-sensing sub-component. Figure 2c describes co-regulatory interactions in terms of the number of TFs affecting a particular gene in each sensing class. Close to 40% of the genes in the internal-sensing category are regulated by two or more TFs. Approximately 27% of the genes in external-sensing and a mere 18% of the genes in the hybrid class fall within this range; when the threshold is raised to three or more regulators of a gene, the percentages drop to 10%, 8% and 2%, respectively. These results indicate that, compared with external-sensing or hybrid classes, a larger fraction of genes that are subject to internal-sensing regulation is co-regulated by additional TFs. This again indicates the
The dense co-regulation of the internal class of TFs might be a consequence of the regulatory contribution of five of the seven global TFs defined in *E. coli* [21] being internal-sensing TFs. CRP (cAMP receptor protein), which senses levels of cAMP, and FNR (fumarate- and nitrate-reductase regulatory protein), which senses the redox state of the cell by binding to [4Fe–4S]^{2+} or [2Fe–2S]^{2+}, belong to the I-SM sub-class. FIS (factor for inversion stimulation), IHF (integration host factor) and Hns (histone-like protein or nucleoid-associated protein) are in the I-DB sub-class. ArcA (aerobic-respiration regulatory protein) belongs to the external-sensing (E-TC sub-class) and Lrp (leucine-responsive regulatory protein) belongs to the hybrid class. The small fraction of internal global TFs has a strong influence on the network; they affect 80% of the genes regulated exclusively by the internal class, and 95% of all co-regulation provided by the internal system is accomplished through these five global regulators. By contrast, ArcA, the external global regulator, only affects 15% of the genes that are regulated exclusively by the external class, 27% of the genes that are regulated by TFs from the overlap of internal and external classes, and 75% of those that are regulated by TFs of the hybrid and external class overlap. Lrp, the global TF of the hybrid class, affects 40% of the genes only regulated in this class, 21% of the genes regulated by the three classes, and 15% of genes co-regulated with the external class. Thus, the dominant role of these five global internal-sensing TFs is clear in their connections with the hybrid- and external-sensing sub-components.

**Functional analysis of gene products regulated by external- and internal-sensing TFs**

Our proposed classification of TFs raises several questions: what do the different classes (namely external, internal and hybrid) comprise in terms of cell function? How do these classes compare in terms of their activities at the cellular level?

To understand the distribution of functional categories in these classes, the database of clusters of orthologous groups (COG) was used [22] to describe the functions of gene products in the cell. Figure 3 shows the distribution of COG categories for products of genes regulated by each TF class and sub-class. The first seven functional categories in Figure 3 comprise genes that are regulated by all classes of signal sensing. However, there are some biases in the regulation of these functional categories: energy production and conversion (COG category 1) is regulated mainly by TFs of the E-TC system; genes for carbohydrate transport and metabolism are co-regulated by I-SM and E-TM systems; inorganic ion transport and metabolism is widely regulated by all the systems but mostly by TFs of the external class. Amino acid transport and metabolism is a category more frequently regulated by hybrid systems, but signal-transduction mechanisms are only regulated by E-TC. Functions of defense mechanisms and intracellular trafficking, secretion, and vesicular transport are regulated only by E-TM. Transcription, replication, recombination and repair, and translation, ribosomal structure and biogenesis are regulated by the internal
class and, to a minor degree, by the hybrid system. Cell motility is characteristically regulated only by I-DB and the hybrid systems, and cell-cycle control, cell division and chromosome partitioning is regulated only by the I-DB system. In summary, this analysis suggests that metabolism, transport of compounds, defense and cell communication are functions that are widely co-regulated by internal and external classes of sensing, whereas the integrity of the basal machinery, cell structure and genetic information is mostly regulated by the internal class. This shows clear compartmentalization and integration of the regulatory network of the cell to sense and respond to external and internal conditions. In addition, systems for extracellular sensing regulate fewer COG categories (ten for E-TC and 12 for E-TM), whereas systems for internal sensing regulate a wider variety of categories (15 for I-SM and 16 for I-DB).

This functional analysis reinforces the notion that the class of internal-sensing TFs is the master set for regulating cellular function.

Concluding remarks
Regulatory networks have not been analyzed in conjunction with their meta-allosteric interactions that connect them with metabolism. The analysis of the E. coli network based on external, internal and hybrid TF sensing is a first step in this direction. We conclude that the global transcriptional response is mainly directed by the internal class of sensing based on the following observations: (i) TFs responding to the internal class of signals control most of the regulons; (ii) TFs of external and hybrid classes are mostly regulated by TFs from the internal class; (iii) the internal class has the highest incidence of co-regulation; and (iv) most of the global regulators belong to the internal class of sensing. Our classification of sensing in E. coli is validated by the functional analysis of the regulated genes for each sub-class. Clearly, TFs that sense internal conditions are the dominant regulators – mostly global ones – affecting all three classes of sensing TFs in E. coli. We believe that this will be universally valid in all bacteria, but whether or not this is the case is currently unknown.

The distribution of internal-, external- and hybrid-sensing TFs within the internal network components [21, 23] of modules [24,25] and motifs [26–29] will help to provide a more detailed understanding of how the cell responds to signals. A regulated transition might start with a change of metabolite concentration which induces a signal that interacts with a TF. This TF might then generate – or repress production of – an enzyme, in turn affecting the concentration of certain metabolites. When considering these meta-allosteric interactions, the circularity in the flux of actions supporting a cellular transition becomes apparent. The mathematical modeling of such network behavior is certainly a challenge for systems biology.

Acknowledgements
We thank G. Moreno-Hagelsieb and E. Pérez-Rueda for their critical review of the manuscript, and V. Jimenez, S. Gama-Castro, M. Peralta-Gil, A. Santos, E. Díaz and F. Sánchez for their technical support. This work was partially supported by NIH grant RO1 GM 071962 and PAPIIT-UNAM IN214905. We also acknowledge three anonymous referees and the Editor for their helpful suggestions.

Supplementary data
Supplementary Material associated with this article can be found at doi:10.1016/j.tim.2005.11.002

References
16 Schlegel, A. et al. (2002) The N terminus of the Escherichia coli transcription activator MalT is the domain of interaction with MalY. J. Bacteriol. 184, 3069–3077