Chapter 1

Structure and Evolution of Transcriptional Regulatory Networks

Guilhem Chalancon* and M. Madan Babu*
Medical Research Council
Laboratory of Molecular Biology
Hills Road, Cambridge CB2 0QH
UNITED KINGDOM
Phone: +44 (0) 1223 402208
Fax: +44 (0) 1223 213556
Email: madanm@mrc-lmb.cam.ac.uk; guilhem@mrc-lmb.cam.ac.uk

*Corresponding authors. Send proofs to:

M. Madan Babu or G. Chalancon
Medical Research Council
Laboratory of Molecular Biology
Hills Road, Cambridge CB2 0QH
United Kingdom
Email: madanm@mrc-lmb.cam.ac.uk; guilhem@mrc-lmb.cam.ac.uk

OVERVIEW
INTRODUCTION
  Concept of transcriptional regulatory networks
STRUCTURE OF TRANSCRIPTIONAL NETWORK
  Local network structure
  Global network structure
  Dynamic nature of transcriptional networks
EVOLUTION OF TRANSCRIPTIONAL NETWORKS
  Mechanisms for the evolution of gene regulatory networks: loss, gain, and rewire
  Impact of gene duplication on TRN evolution
  Horizontal gene transfer: getting connected
  Evolution of networks across organisms
OUTLOOK AND PERSPECTIVE
  Quantitative modeling of gene networks
  Natural variation and network evolution
  Noise and gene networks
  Engineering gene circuits
Acknowledgements
REFERENCES

OVERVIEW

Regulation of gene expression is primarily mediated by proteins called transcription factors (TFs), which recognize and bind specific nucleotide sequences and affect transcription of nearby genes. Over the last years, considerable information has been accumulated on regulatory interactions between the TFs and their regulated target genes (TGs) in various model prokaryotic systems such as *Escherichia coli* and *Bacillus subtilis*. This has permitted researchers to model the transcriptional regulatory system of an organism as a network, wherein TFs or TGs are represented as nodes and regulatory interactions are denoted as directed links. Representation of this information as a network has provided us with a robust conceptual framework to investigate this system, and work in the last decade has uncovered several fundamental general principles pertaining to its structure and evolution. In this chapter, we first introduce the concept of transcriptional regulatory networks. We then discuss our current understanding of the structure of transcriptional regulatory networks. Specifically, we discuss the local and global structure of such networks. We then discuss the various forces that influence network evolution such as gene duplication, horizontal gene transfer and gene loss. In particular, we discuss how the transcriptional regulatory network evolves across organisms that live in different environments. Finally, we conclude by discussing major challenges for future research and highlighting how the new understanding can have implications for biotechnology and medicine, and be exploited in applications such as microbial engineering and synthetic biology.

INTRODUCTION

The ability to co-ordinate and bring about changes in gene expression in response to environmental variation is crucial for the maintenance of cellular homeostasis. Among all the regulatory processes modulating the synthesis of a gene-product, regulation of transcription is essential, as this is the first step in a series of events that give rise to a protein. Such alterations in the expression level of particular genes eventually trigger phenotypic changes in response to the environment, thereby permitting the organism to adapt to the new environment.

Regulation of transcription is mediated through proteins called transcription factors (TFs). TFs are DNA binding proteins that bind to specific regions, the *cis*-regulatory elements, in the promoter regions of certain genes and eventually influence gene expression. In addition to a DNA binding domain (DBD) that recognizes the DNA, most TFs also contain an additional regulatory domain (e.g., a small molecule binding domain, enzymatic domain, etc) that responds to the signal (e.g., a small molecule). The affinity of the DNA-binding domain to bind a specific DNA sequence can be modulated through the state of the regulatory domain (e.g., a ligand binding to a regulatory domain). The regulatory domain itself is influenced by the presence or absence of a signal in the internal or the external environment. For example, in a simple free-living organism such as *E. coli*, studies have estimated the presence of ~320 TFs and over 80% of them have been shown to also contain a regulatory domain in addition to a DBD (Madan Babu and Teichmann 2003).

The binding of a TF to a promoter region can either result in an increased or decreased transcription of the regulated target gene. In addition to exerting their effect independently, TFs can also affect gene expression in a combinatorial manner. More specifically, TFs regulate the initiation of transcription through different strategies operating on the transcriptional machinery. In bacteria, we can roughly distinguish two classes of mechanisms for repression: the binding of TFs can block the RNA polymerase by steric hindrance, or can recruit co-repressors that decrease

the affinity of the holoenzyme ($\alpha 2\beta\beta'\omega$) for the promoter region. Similarly, activation can either be achieved through the binding of the TF which increases the local concentration of the holoenzyme at the promoter region, or by the subsequent recruitment of a co-activator. Please see Browning and Busby (Browning and Busby 2004) for a more detailed description of the other mechanisms of activation and repression of transcription in bacteria.

The affinity of transcription factor DNA-binding domains for promoters is sequence-dependent. Therefore, genes containing identical or similar DNA sequences (cis-regulatory elements) in their promoter region are susceptible to be targeted and regulated by the same TF. Moreover, the unit of the prokaryotic genome organization comprise of operons, which consist of a collection of genes that are adjacent to each other, placed under the control of a single promoter, and give rise to a poly-cistronic transcript (i.e., mRNA molecule which can have independent translation initiation sites for the generation of multiple protein products that are encoded in the same transcript) (Davies and Jacob 1968). As a consequence, genes belonging to the same operon can be regulated at once, by one single TF. As the genes contained in operons tend to have similar biological functions, this organization is considered to facilitate the coordinated regulation of gene expression (Osbourn and Field 2009).

The expression pattern of transcription factors itself is extremely dynamic and dependent on stress. In *E.coli*, a key response to stress is the general stress response, which triggers the transcription of genes required for survival during starvation. This response is induced by growth-rate reduction, which is a consequence of nutrient limitation or starvation. It can also be induced by acidic pH, rapid variations in temperature or in osmolarity (Weber et al. 2006).

Modulators of the general stress response include transcription factors and subunits of the RNA polymerase such as sigma factors. Particularly σ^{38}, also called RpoS, controls the expression of ~10% of the genome in case of starvation (Foster 2007; Weber et al. 2005). RpoS is structurally very similar to σ^{70}, which is largely expressed in rapidly growing cells, but controls the transcription of distinct set of genes, that decrease the growth rate but target DNA protection and repair. This example highlights the importance of transcriptional regulation for survival. Please refer to chapter 3 for the role of sigma factors, and to chapter 15-17 for a description of the general stress response in bacteria.

Concept of transcriptional networks

A fast, precise and global regulation of transcription is essential for cell survival in changing environments. This regulation is mostly controlled by transcription factors, which are differentially expressed or regulated depending on environmental conditions, and which specifically target promoter regions. This knowledge results from decades of detailed investigations which focused on specific cases of prokaryotic gene regulation, mostly performed in *Escherichia coli.* However, deciphering general rules governing transcription regulation at the genome-scale in bacterial organisms has become an achievable goal in recent years.

As one would imagine, not only myriads of transcriptional factors bind to promoter sequences with combinatorial effects on the transcription of downstream genes, but also those interactions are highly dynamic. This dynamics allow cells to co-ordinate elaborate responses to external and internal stimuli, but is a major challenge for understanding transcriptional regulation in its global

nature. The availability of sequenced genomes as from the late 1990s undoubtedly changed the scenario. It has now become possible to collect and analyze large amounts of information (in hundreds, then thousands) of bacterial species, allowing annotations and predictions of transcription factor binding sites. Simultaneously, the development of genome-scale high-throughput experiments detecting protein-DNA interactions became possible. For instance, chromatin-immunoprecipitation and protein-DNA microarrays played a central role in the identification of new protein-DNA interactions (Grainger et al. 2005; Grainger et al. 2009; Molle et al. 2003).

The understanding of the diverse nature of information on transcription factors and their regulated targets (See **Table 1**) was facilitated by the adoption of network theory, which permitted uncovering patterns in gene regulation on a genomic scale (Babu et al. 2004; Milo et al. 2002; Thieffry et al. 1998). The investigation of interactions between TFs and their target genes as a network provided a general framework to identify general principles that govern such complex systems. Formally, transcriptional regulatory networks (TRNs) are modelled as directed graphs, which are composed of vertices or nodes that are connected by directed edges. In this case, vertices denote both transcription factors (TF) or their target genes (TG). Directed edges, which connect a TF to its TG represents a regulatory interaction. Such an object can be studied with a set of analytic tools derived from network theory (Babu et al. 2004; Barabasi and Oltvai 2004). Consequently, during the past decade, such approaches have facilitated detailed investigations into the structure, the dynamics and the evolution of the regulation of transcription at the genome scale.

In this chapter, we first discuss the main characteristics of the structure of prokaryotic transcriptional regulatory networks. In the second part, we discuss about the various forces that influence their evolution. Finally we discuss how the understanding gained is being exploited in biotechnology and medicine.

STRUCTURE OF TRANSCRIPTIONAL NETWORK

Transcriptional regulatory networks (TRNs) have a complex and hierarchical structure and can be investigated at several levels of organisation (Babu et al. 2004) (**Figure 1**). At the most basic level, the network is made up of basic units, which comprise of a transcription factor, its target gene and the *cis*-regulatory element through which it regulates the expression of the target gene (**Figure 1A**). At the local level of organisation, these basic units are arranged into recurrent wiring patterns called network motifs, which appear frequently throughout the network (**Figure 1B**). The network motifs have been shown to perform specific information processing task, and details of this is discussed below and in Chapter 2. The global level of organisation involves the set of all known regulatory interactions among the TFs and the TGs in an organism (**Figure 1C**). In particular, TRNs have been shown to be characterised by the presence of a few TFs which are referred to as global regulators as they control the expression of a large number of genes.

It should be noted that much of the work on bacterial regulatory networks has focused on *Escherichia coli* for which data are most abundant. While much of our understanding of TRNs has been obtained by investigating the *E. coli* network, work on the *B. subtilis*, *Corynebacterium* and *S. cerevisiae* network and the TRNs from other organisms have shown that the general principles of organisation are largely the same. Currently, there over 2,500 regulatory interactions in *E. coli*, which are available through the RegulonDB database (Gama-Castro et al. 2008). For a

comprehensive list of databases providing information about known and inferred transcriptional regulatory networks, please see **Table 1**.

Local network structure

At a local level, TRNs have been shown to contain small recurrent patterns of interconnections whose number of occurrence is substantially higher than what is expected by chance when compared with random networks of identical size. These structures, which were first defined by Shen-Orr *et al*. (Shen-Orr et al. 2002) are known as network motifs (Alon 2007). Please refer to Chapter 2 for more details. Milo *et al*. (Milo et al. 2002) and Lee *et al*. (Lee et al. 2002) discovered three over-represented network motifs in the *E. coli* and yeast transcriptional regulatory network (**Figure 1B**). These three motifs are referred to as (i) Feed Forward Motifs (FFM), (ii) Single Input Modules (SIM) and (iii) Multiple Input Modules (MIM). Several subsequent work have shown that each motif possess distinct kinetic properties with respect to the control of target gene expression (Alon 2007).

(i) Feed Forward Motifs: In FFMs, a top-level TF regulates a target gene and an intermediate TF, which also regulates the same target gene. One should note that since the top and the intermediate TFs can either be activators or repressors, four combinations are possible, in response to two possible input (that is activation or repression of the top-level TF) resulting in eight distinct cases. However, two particular combinations are prevalent in the *E. coli* transcriptional regulatory network (Alon 2007; Mangan and Alon 2003). In the most recurrent FFM, both TFs are activators. This pattern ensures that the TG is only transcribed when a persistent signal activates the top-level TF, as expression of the target gene relies on the activation of the two TFs. This configuration prevents fluctuating concentrations of the top-level TF from regulating the downstream target gene, thereby filtering stochastic variation or noise in the input signal. Noticeably, the second most frequent feed forward motif in the *E. coli* TRNs comprises of TFs acting in an opposing manner: the intermediate-level TF is a repressor while the top-level one is an activator. This pattern is referred to as an incoherent FFM (Mangan et al. 2006), and possesses a pulse-like dynamics in the expression of the target gene: the top-level TF activates the expression of the TG until a response threshold that activates the intermediate TF. At that point, the expression of the TG is inhibited.

(ii) Single Input Modules: In SIMs, a single TF regulates a group of target genes simultaneously, therefore allowing a coordinated regulation of those set of genes. However, the concentration of TF necessary to activate the regulated genes varies depending on their promoter strength. Therefore, a SIM can show a rather subtle behavior, as the TF concentration changes with time. Such a motif can set a temporal order in the pattern of expression of individual target genes. Such patterns have been indeed observed experimentally in several metabolic pathway genes (Zaslaver et al. 2004) and in the flagellar biogenesis pathway (Kalir et al. 2001).

(iii) Multiple Input Motifs: In this type of motif, multiple TFs regulate the expression of numerous TGs. Consequently, distinct signals can be integrated in the motif, providing distinct ways of regulating gene expression. Consistently, MIM provides a flexible regulation of their target genes in a combinatorial manner that is very likely to confer a fitness advantage under different environmental conditions.

Global structure

The global level of organization of transcriptional regulatory networks (TRNs) has been extensively studied by several groups. It has been shown that TRNs display a "scale-free" like topology (Babu et al. 2004; Madan Babu and Teichmann 2003; Thieffry et al. 1998). Such a topology is characterised by the presence of a few TFs (referred to as global regulators) that regulate a strikingly large number of target genes and a vast majority of TFs (called as fine-tuners) that regulate a small number of TGs. An analysis of the *E. coli* transcriptional network has defined global regulators as the top 20% of the TFs with the highest number of regulated target genes. An investigation of the function of the global regulators showed that they are TFs involved in carbon degradation (Mlc and Lrp), redox status sensing (ArcA, NarL and Fnr), ion transport regulation (Fur), environmental sensors (CspA and Crp) and nucleoid associated proteins (Hns, Ihf and Fis). It has been proposed that the global regulators contribute to the robustness of the gene regulatory system, where robustness is defined as the ability of the transcriptional regulatory network to remain functional while its structure is significantly perturbed (Barabasi and Albert 1999; Kitano 2004). In addition to the above mentioned topology, recent studies have also shown that the TRN of *E. coli* and that of other organisms display extensive combinatorial regulation (Balaji et al. 2007; Janga et al. 2007b) and tend to possesses a multi-layer hierarchical (i.e., a serial cascade of transcription factors) structure without feedback regulation at the transcription level (Cosentino Lagomarsino et al. 2007; Jothi et al. 2009; Ma et al. 2004; Martinez-Antonio et al. 2008; Yu and Gerstein 2006).

Dynamic nature of transcriptional networks

The maintenance of cellular homeostasis and the successful adaptation to environmental changes are challenges that microorganisms face all the time. This ability relies on the rapid integration of external and internal stimuli *via* changes in gene expression. Unsurprisingly, the capacity of the transcriptional regulatory machinery to quickly bring about changes in the gene expression pattern reflects the highly dynamic dimension of transcriptional regulatory networks. Cells must respond to change in temperature and pH, nutrient or toxins concentrations, etc. Consistently, active parts of the transcriptional regulatory network change over time. In addition to sequence specific TFs that respond to distinct signals, nucleoid-like architectural proteins have been shown to affect the local chromosome structure and influence the availability of specific sites on the DNA. Such chromosomal dynamics has been shown to influence the expression of several genes (Marr et al. 2008). In this sense, knowledge on the topological properties of regulatory network, though informative, is not sufficient to explain this fundamental function. Accordingly, a change in regulatory network topology across different conditions and the impact of architectural proteins such as Hns, Fis, etc has gained considerable attention and is a direction of current intense research (Balaji et al. 2007; Berger et al.; Dillon and Dorman; Dorman 2009a; Janga et al. 2007a; Luijsterburg et al. 2006; Luijsterburg et al. 2008; Marr et al. 2008; Martinez-Antonio et al. 2008). In addition to architectural proteins, secondary messenger molecules such as cyclic di-GMP, (p)ppGpp, riboswitches and small regulatory RNAs can affect gene expression dynamics. Their prevalence and impact on gene regulation on a genomic scale, and how they tune the transcriptional response is another intense area of research (Hengge 2009; Montange and Batey 2008; Pesavento and Hengge 2009; Schirmer and Jenal 2009; Sharma et al.; Storz et al. 2005; Waters and Storz 2009).

EVOLUTION OF TRANSCRIPTIONAL NETWORKS

The increasing availability of completely sequenced genomes and the development of high-throughput experiments have facilitated extensive investigation of gene phylogenies for all

protein families from hundreds of prokaryotic organisms. This has allowed us to gain insights into the intricate interplay of evolutionary forces that drive the evolution of transcriptional regulatory networks. In this part of the chapter, we will first provide a short overview of the major mechanisms of gene evolution and then discuss the role of these evolutionary forces in shaping the prokaryotic regulatory networks.

Mechanisms for the evolution of gene regulatory networks

Mutations in the genome of an organism contribute to the evolution of TRNs. Such mutations, which fall on a spectrum, may affect just a single or few bases (e.g., single nucleotide substitutions) or may result in the generation of a large chunk of genetic material (e.g., duplication, repeat element expansion by transposons or horizontal transfer). Accordingly, such events may have a range of outcomes; for instance, they can affect regulatory interactions either (i) at the *cis*- level, by mutating TF-binding sites or incorporate *cis*-regulatory elements upstream of genes during repeat element expansion or (ii) at the *trans*- level, through the modification or generation of new DNA-binding domains that may recognize a different DNA sequence or may respond to a different ligand. Most of these mutations are likely to either be deleterious or cause disruption of an existing regulatory interaction. Evolution of the TRNs, on the other hand, consists of addition of new nodes (TFs and TGs) and new edges (regulatory interactions). As we will see in the following sections, gain of genes is crucial for those two aspects. As illustrated in **Figure 2**, gene gain is driven in prokaryotes either by gene duplication (Brenner et al. 1995; Chothia and Gough 2009; Teichmann et al. 1998) or by horizontal gene transfer (Koonin et al. 2001; Kunin et al. 2005). While these two processes intrinsically add new nodes in TRNs, more importantly, they increase the evolvability of such network by facilitating gain and rewiring of regulatory interactions (Babu et al. 2004; Gelfand 2006; Janga and Collado-Vides 2007; McAdams et al. 2004; Perez and Groisman 2009a). This point is well illustrated by a recent work which showed that artificial incorporation of new regulatory interactions into *E. coli* is rarely a barrier for evolution and even contributes to the fitness under various selection pressures (Isalan et al. 2008). In this section, we only consider gene duplication, loss and horizontal gene transfer. We do not explicitly address evolution of new interactions through repeat element expansion, which is another mechanism that may influence network evolution (Marino-Ramirez et al. 2005).

Impact of gene duplication on TRN evolution

Evolution by gene duplication involves the generation of a second copy of the genomic segment harboring a gene, thereby resulting in the emergence of two identical copies of the same gene in a genome. Following duplication, one of the copies retains the ancestral function and the other copy may diverge under a relaxed selection pressure until it acquires a new function (neo-fuctionalization). Alternatively, the two copies may share a part of the function of the ancestral copy (sub-functionalization) or the second copy may become degenerate (Lynch and Conery 2000). In a simplistic scenario, three different cases (**Figure 2A**) must be considered: i.e., whether the duplicated segment contains either a TF or TG, or both (Madan Babu and Teichmann 2003; Teichmann and Babu 2004). As a consequence of this event, gene duplication will result in doubling the quantity of regulatory interactions in addition to the number of genes involved. In each case, the fate of those shared interactions, that is their maintenance or removal during evolution, is of crucial importance to understand the evolution of transcriptional regulatory networks.

Through a systematic analysis of the transcriptional regulatory network of the prokaryote *E. coli* and the unicellular eukaryote *S. cerevisiae*, Teichmann and Babu found that more than two-third of the interactions have evolved as a consequence of gene duplication. They also observed that over one-half of the known regulatory interactions were inherited from ancestral transcription factors or target genes after duplication with the rest of the regulatory interactions having been re-wired and gained during divergence after gene duplication (Madan Babu and Teichmann 2003; Teichmann and Babu 2004). The authors also noticed that only a small fraction of the genes and the regulatory interactions have evolved as a consequence of gene recombination or innovation (Teichmann and Babu 2004).

An obvious question that arises given the vast amount of gene duplication during the evolution of transcription networks is if this has had any significant role in the generation of the network motifs or of the global topology of the network. In the same study (Teichmann and Babu 2004), the authors investigated the individual network motifs and demonstrated that while the individual genes in the network motifs may have evolved as a consequence of gene duplication, the interactions have either been gained or have evolved as a consequence of re-wiring. Conant and Wagner (Conant and Wagner 2003) also observed the same trend by investigating the yeast and the *E. coli* network. These studies together demonstrate that network motifs have evolved independently (i.e., convergent evolution) multiple times, possibly because they contribute to fitness by tuning the expression level of genes in a way that maximizes fitness. This is supported by the observation from experimental evolution studies, where *E. coli* was found to optimize its expression level of a protein that maximizes growth rate and therefore its fitness (Dekel and Alon 2005). An investigation of the global structure of the TRN by Teichmann and Babu showed that the scale-free structure is not a direct consequence of gene duplication. While this observation is consistent with the possibility that the scale-free structure could have evolved due to selection, there are other possible mechanisms, which are non-adaptive (e.g., neutral evolution), that may also give rise to the same structure (Lynch 2007).

Taken together, these studies have shown that gene duplication has played a key role in the evolution of the network components, losses and gains of regulatory interactions. In addition, they have contributed to the growth of the TRN through the inheritance of regulatory interactions, gain and through re-wiring, thereby fuelling network evolution.

Horizontal gene transfer: getting connected

In eukaryotes, gene duplication and loss are believed to be the major source of genome diversification. However, in prokaryotes, horizontal gene transfer (HGT) of genetic material also represent a substantial source of genetic novelty (Koonin et al. 2001; Lerat et al. 2005). Interestingly, the uptake of foreign genes is often biased towards the acquisition for traits that directly contribute to fitness such as virulence, symbiosis, or resistance to toxins (Becq et al. 2007; Nakamura et al. 2004; Sorek et al. 2007). Thus while understanding the role of HGT is of particular importance in prokaryotic evolution, it also has implications for understanding how they contribute to network evolution and adaptation of organisms to new environments (Ahmed et al. 2008; Juhas et al. 2009).

HGT requires the physical incorporation of foreign DNA into the receiver organism, its integration into the host regulatory network, and eventually its selection through the bacterial population (i.e., its fixation). The incorporation of DNA during HGT is driven by three distinct

mechanisms referred to as conjugation, transduction and transformation. The molecular mechanisms of these processes have been extensively studied, and are beyond the scope of this chapter (Please see (Chen et al. 2005)). Here, we discuss the regulatory constraints and mechanisms that shape the integration of new genes in TRNs. When a segment of DNA is horizontally transferred into an individual, the immediate impact on fitness of the imported genes is indeed crucial for the adaptation and survival of the individual in a bacterial population and during changing environments. However, how the gene gets integrated into the chromosome over the long run and how it integrates into an existing regulatory network is only now being understood in detail (Dorman 2007; Dorman 2009b; Lercher and Pal 2008; Navarre et al. 2007; Stoebel et al. 2008) (**Figure 2B**).

If the transferred segment is transcriptionally active, an imported gene must be successfully translated and folded in a non-lethal protein. In such cases, its protein expression level must be adequately regulated. This implies the need for a tighter transcriptional regulation, and thus a proper recognition of its promoter region and transcription factor binding sites by the resident transcriptional network, or requires a horizontally transferred TF that came along with the segment. Therefore, the probability of integrating a transferred gene into a network is expected to generally decrease with phylogenetic distance (Sorek et al. 2007). It has been observed in *E. coli* K-12 that genes in K-loops, known to be hot-spots of HGT, are poorly translated (Taoka et al. 2004). Taoka and colleagues notably provided evidence that most of the recently acquired foreign genes in *E. coli* K-12 are generally not translated in laboratory conditions, suggesting that their expression may not be directly contribute to fitness (i.e., growth) in log-phase culture. In another study Sorek et al (Sorek et al. 2007) have shown that genes that failed to be horizontally transferred are those that are generally highly expressed. Thus, viability and successful synthesis of newly acquired genes alone are unlikely to be sufficient conditions for fixation. A balance between fitness benefits and cost in synthesis of the new gene is therefore necessary for the survival and competitiveness of the individual harboring the transferred gene in a mixed bacterial population.

How can the cell find a strategy to favor such balance? Interestingly, several recent reports have suggested that it might be important, as a first step, to silence the transferred gene. The transferred gene can then be subsequently expressed (through anti-silencing mechanisms (Stoebel et al. 2008)) when the benefit of its expression is higher than the cost of its synthesis. This is likely to tip the balance in the population, favouring the emergence of individuals who harbor the transferred gene. For example, it was observed that nucleoid-associated proteins such as Hns contribute to silencing the transcriptional activation of recently acquired genes, providing a "stealth function" minimizing the cost on fitness of their expression, thus facilitating their transmission (Doyle et al. 2007; Stoebel et al. 2008). Consistently, Navarre et al. demonstrated that in *Salmonella* Hns selectively silences horizontally acquired genes by targeting sequences with GC-content lower than the resident genome (Navarre et al. 2006). In addition to these studies, Perez and Groisman have suggested that mutations in orthologous transcription factors and in their dependent promoters in different organisms may allow bacterial transcription factors to incorporate newly acquired genes into ancestral regulatory circuits and yet retain control of the core members of a regulon (Perez and Groisman 2009b).

Taken together, these studies have begun to help us understand the role of horizontal gene transfer in network evolution and appreciate better various aspects of laterally acquired genes

which contribute to its increased likelihood to be successfully integrated into existing regulatory networks.

Evolution of networks across organisms

While the above studies have provided insights into how networks evolve in an organism, it is of fundamental interest to understand how transcriptional regulatory networks evolve across species. In other words, are interactions between TFs and TGs sufficiently conserved to be able to predict a regulatory interaction in an organism from a closely related one? This question is important since less information is available on the transcriptional networks of many prokaryotes, as most of the experimental studies performed over the past decades have been focused on model organisms such as *E. coli* and *B. subtilis*. Approaches used to address the problem of the inference of TRN from other prokaryotes can broadly be grouped into two categories, depending on whether we focus on orthology or on sequence similarity of transcription factor binding sites (Babu 2008; Janky et al. 2009; Venancio and Aravind 2009). The first category of methods exploits the assumption that orthologous TFs regulate orthologous TGs in distinct genomes. The latter exploits the assumption that identical binding sites upstream of two genes in closely related species imply similar regulatory interactions with orthologous TFs. Overall, these methods, in addition to methods discussed in the introduction has provides us with a deeper insight into the evolution of TRNs across organisms.

Recent studies that have investigated over 150 completely sequenced genomes have shown that TFs are less conserved across genomes than their target genes (Lozada-Chavez et al. 2006; Madan Babu et al. 2006), suggesting a greater evolvability of TFs. Noticeably, it was observed that global regulators do not differ from other TFs in terms of sequence conservation. Another study by Hershberg and Margalit showed that the mode of regulation (activation or repression) exerted by transcription factors has an effect on their evolution. Repressors were found to co-evolve tightly with their target genes. In contrast, activators were found to be lost independently of their targets. These results suggest that prokaryote organisms evolve rapidly their own set of transcriptional regulators, and are therefore able to rewire regulation interaction in a very flexible way. These observations are also supported by a study by Isalan et al (Isalan et al. 2008) which has shown that artificial incorporation of new regulatory interactions into *E. coli* is rarely a barrier for evolution and in fact contributes to the fitness under various selection pressures.

An analysis of the local structure revealed that motifs are not conserved as whole units and that individual interactions within a motif may be lost or retained. Given the functional importance of network motifs, these results may seem surprising at a first glance as one would have expected that closely related species will conserve local network structures. However, a careful analysis by Babu et al (Madan Babu et al. 2006) showed that organisms with similar lifestyle tend to conserve similar interactions and similar motifs. In fact, it was noticed that losing or gaining interactions can result in embedding orthologous genes in different motif contexts (**Figure 2C**). Thus, this result is more meaningful when one considers the environment in which an organism lives. This trend appeared to be statistically significant and the study has identified interesting examples (Madan Babu et al. 2006). For instance, in *E. coli,* it was observed that the fumarate reductase genes FrdB and FrdC are under the control of the transcription factors Fnr and NarL in a feed-forward motif. These enzymes, which convert fumarate to succinate under anaerobic conditions to derive energy, are therefore only expressed when both Fnr and NarL are active, that is only under a persistent signal for lack of oxygen. Consistently, *E. coli* faces alternations of

aerobic and anaerobic phases over long periods, which makes it important to induce fumarate reductases only when the bacteria is likely to stay in an anaerobic environment for extended periods. In contrast, *H. influenzae* is a pathogen that faces strong redox fluctuations during host infection. Interestingly, contrary to what happens in *E. coli* NarL is lost, and the expression of FrdB or FrdC only depend on Fnr. Therefore the fumarate reductases are regulated in a simpler manner (through a Single Input Motif) in this pathogen, which again seems relevant given its environmental lifestyle. Interestingly, this feed-forward motif found in *E. coli* is also conserved in distantly related organisms such as *B. pertussis* (beta-proteobacterium) and *D. hafniense* (firmicute) that have similar lifestyle.

At the level of the global structure, it was observed that global regulatory hubs are not preferentially more conserved than other TFs. It was found that the condition specific global regulatory hubs are the ones that may be lost more easily. This observation lends support to an idea that orthologous transcription factors may contribute to different fitness to organisms living in different environments and hence completely different transcription factors may emerge as global regulators. Consistent with this, an analysis of the *E. coli* and the *B. subtilis* network revealed that while the global topology was similar, very different proteins emerged as global hubs. This observation again points to the importance of the environment in shaping network structure (Madan Babu et al. 2006).

Taken together, these observations highlight an important principle which is that transcriptional regulatory networks are extremely plastic, evolve rapidly and adapt to the environment by tinkering individual interactions (Lozada-Chavez et al. 2006; Madan Babu et al. 2006; Price et al. 2007). More specifically, the specific principles can be summarized as follows (**Figure 3**): at the level of network components, TFs evolve more rapidly than their target genes, allowing organisms to organisms to evolve their own set of regulators in line with their environment. Besides, both at the basic and at the local structure level, organisms with similar lifestyle tend to possess similar regulatory interactions. Finally, at the level of the global structure, conservation of TFs is independent of their connectivity (*i.e.* the number of target genes), while the environment, again, seems to be the major force driving gain and loss of TF and regulatory interactions.

OUTLOOK AND PERSPECTIVES

In this chapter, we have introduced the concept of transcriptional regulatory networks and have discussed how representing the transcriptional regulatory system of an organism as a network could provide us with a better understanding of the complexity of gene regulation on a genomic scale. Specifically, we have discussed research in the last decade and have highlighted general principles of network structure and evolution. In this section, we discuss major challenges and important directions for future research and describe how our understanding of the structure and evolution of gene networks are already being exploited in different ways.

Quantitative modeling of gene networks

While experimental advances in sequencing are providing us with an avalanche of information about the repertoire of genes and their expression levels across different conditions from diverse microbes and microbial communities, one of the fundamental challenges for the future would be to develop conceptual and computational framework to integrate all these data to quantitatively model how individual genes are regulated within a cell in different context such as stress, during

infection, in the presence of a particular food source, etc. In this direction, computational and experimental approaches that model regulation of individual genes at high resolution (Ronen et al. 2002; Zaslaver et al. 2006) or the changes in the structure of entire regulatory network of an organism (Luscombe et al. 2004; Martinez-Antonio et al. 2008) are already being investigated. A key advance would be to investigate different biological systems such as DNA damage response, stress response, etc from diverse organisms, develop new methods for investigating network dynamics and to uncover general principles through comparative analysis.

Natural variation and network evolution

The ability to sequence different strains of the same species or different individuals from the same population is providing us with a wealth of information about natural variation in the genomic sequences of different organisms (e.g., *Mycobacterium leprae* (Monot et al. 2009), *Escherichia coli* (Ooka et al. 2009; Studier et al. 2009)). Such variation might involve single nucleotide changes (Brochet et al. 2008), or structural alterations such as insertion and deletion of sequences through transposable elements and horizontal gene transfer (Brzuszkiewicz et al. 2006). These events affect not only protein coding regions, but also inter-genic regions and hence may influence the expression of relevant genes. For example, it was recently shown that the gain of a regulatory interaction through mutations in the promoter region of *Salmonella typhimurium* strains allowed the regulation of a virulence gene. This feature conferred a fitness advantage to those strains and permitted them to adapt better to the host environment (Osborne et al. 2009). Given the fluid nature of bacterial genomes, another important future direction would be to understand natural variation in gene circuits within distinct populations of the same species. Such an understanding can provide fundamental insights into the emergence of pathogens (Brzuszkiewicz et al. 2006) and has implications for human health and disease (Ahmed et al. 2008).

Noise and gene networks

Non-genetic cell-to-cell variation in gene experssion (i.e., noise) has been another exciting area that has gained attention recently (Losick and Desplan 2008; Raj and van Oudenaarden 2008) (see Chapter 22). Such stochastic variation in a cell population can be beneficial where phenotypic diversity is advantageous but detrimental if homogeneity and fidelity in cellular behaviour is required. Recent work in this direction has shown that different circuits have the potential to either amplify or buffer noise (Losick and Desplan 2008; Raj and van Oudenaarden 2008). For instance, it was recently shown that while seemingly different alternative circuits can provide similar patterns of outputs in gene expression, the impact of fluctuations in protein levels was shown to be an important determinant of why some circuits were selected in evolution (Cagatay et al. 2009). An important challenge in this direction would be to understand the interplay between network structure and the noise level of individual genes in such networks. In this direction, a recent study by Jothi et al (Jothi et al. 2009) has shown that TFs which are in the top of the hierarchy generally tend to show higher cell-to-cell variation in their expression level. Based on this and other observations, it was proposed that the interplay between network organization and TF dynamics could permit differential utilization of the same underlying network by distinct members of a clonal cell population. Gaining a better understanding of how gene circuits could influence stochasticity in gene expression will have a significant impact in understanding phenomenon such as (i) bacterial persistence or adaptive resistance (e.g., (Balaban et al. 2004; Jayaraman 2008)), (ii) differential cell-fate outcome in response to the same uniform stimulus (e.g., (Maamar et al. 2007)), (iii) phenotypic variability in fluctuating environments

(e.g., (Acar et al. 2008)), and (iv) cellular differentiation and development (e.g., (Suel et al. 2006; Suel et al. 2007)).

Engineering gene circuits

Another major challenge would be to exploit the knowledge gained about regulatory networks to engineer gene circuits with defined properties (e.g., tunable circuits (An and Chin 2009)) for different applications. In this context, several groups have made important contributions and synthetic gene circuits are already being exploited in medicine (e.g., engineering interactions between bacterial and human cells (Anderson et al. 2006; Steidler et al. 2000); see Chapter 23), bioenergy (e.g., production of fatty-acid derived fuels (Steen et al.); see Chapter 31), bioremediation (e.g., to harness the concentration gradient of metals (Xu and Lavan 2008); see Chapter 32), laboratory applications (e.g., creation of bacterial strains resistant to specific antibiotics for selection experiments (Dantas et al. 2008; Martinez 2008); see Chapter30) and in biotechnology (e.g., for the production of proteins(Alper et al. 2005)). For a more detailed and current account of synthetic biology and engineering of gene circuits, the reader is recommended to the following reviews by Chin JW (Chin 2006), Kiel et al (Kiel et al.) and Lu et al (Lu et al. 2009).

In conclusion, this is truly an exciting time for experimental and computational biologists who aim to understand gene regulatory networks. Especially, with the advances in computing and genomic technologies, we foresee the availability of more extensive and detailed maps of transcriptional regulation and other mechanisms of regulation (e.g., riboswitches and small RNAs; see Chapter 5) in a number of microorganisms. The availability of such information will fuel research that addresses fundamental questions linking different types of regulation (Leonard et al. 2008; Purnick and Weiss 2009). All these advancements collectively have the potential to transform our understanding of gene regulation in the near future.

Acknowledgements

The authors would like to thank the Medical Research Council, UK for funding their research.
GC thanks the ENS Cachan for financial support.

REFERENCES

Acar, M., J.T. Mettetal, and A. van Oudenaarden. 2008. Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet* **40:** 471-475.

Ahmed, N., U. Dobrindt, J. Hacker, and S.E. Hasnain. 2008. Genomic fluidity and pathogenic bacteria: applications in diagnostics, epidemiology and intervention. *Nat Rev Microbiol* **6:** 387-394.

Alon, U. 2007. Network motifs: theory and experimental approaches. *Nat Rev Genet* **8:** 450-461.

Alper, H., C. Fischer, E. Nevoigt, and G. Stephanopoulos. 2005. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A* **102:** 12678-12683.

An, W. and J.W. Chin. 2009. Synthesis of orthogonal transcription-translation networks. *Proc Natl Acad Sci U S A* **106:** 8477-8482.

Anderson, J.C., E.J. Clarke, A.P. Arkin, and C.A. Voigt. 2006. Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol* **355:** 619-627.

Babu, M.M. 2008. Computational approaches to study transcriptional regulation. *Biochem Soc Trans* **36:** 758-765.

Babu, M.M., N.M. Luscombe, L. Aravind, M. Gerstein, and S.A. Teichmann. 2004. Structure and evolution of transcriptional regulatory networks. *Curr Opin Struct Biol* **14:** 283-291.

Balaban, N.Q., J. Merrin, R. Chait, L. Kowalik, and S. Leibler. 2004. Bacterial persistence as a phenotypic switch. *Science* **305:** 1622-1625.

Balaji, S., M.M. Babu, and L. Aravind. 2007. Interplay between network structures, regulatory modes and sensing mechanisms of transcription factors in the transcriptional regulatory network of E. coli. *J Mol Biol* **372:** 1108-1122.

Barabasi, A.L. and R. Albert. 1999. Emergence of scaling in random networks. *Science* **286:** 509-512.

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

Becq, J., M.C. Gutierrez, V. Rosas-Magallanes, J. Rauzier, B. Gicquel, O. Neyrolles, and P. Deschavanne. 2007. Contribution of horizontally acquired genomic islands to the evolution of the tubercle bacilli. *Mol Biol Evol* **24:** 1861-1871.

Berger, M., A. Farcas, M. Geertz, P. Zhelyazkova, K. Brix, A. Travers, and G. Muskhelishvili. Coordination of genomic structure and transcription by the main bacterial nucleoid-associated protein HU. *EMBO Rep* **11:** 59-64.

Brenner, S.E., T. Hubbard, A. Murzin, and C. Chothia. 1995. Gene duplications in H. influenzae. *Nature* **378:** 140.

Brochet, M., C. Rusniok, E. Couve, S. Dramsi, C. Poyart, P. Trieu-Cuot, F. Kunst, and P. Glaser. 2008. Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of Streptococcus agalactiae. *Proc Natl Acad Sci U S A* **105:** 15961-15966.

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

Brzuszkiewicz, E., H. Bruggemann, H. Liesegang, M. Emmerth, T. Olschlager, G. Nagy, K. Albermann, C. Wagner, C. Buchrieser, L. Emody, G. Gottschalk, J. Hacker, and U. Dobrindt. 2006. How to become a uropathogen: comparative genomic analysis of extraintestinal pathogenic Escherichia coli strains. *Proc Natl Acad Sci U S A* **103:** 12879-12884.

Cagatay, T., M. Turcotte, M.B. Elowitz, J. Garcia-Ojalvo, and G.M. Suel. 2009. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell* **139:** 512-522.

Chen, I., P.J. Christie, and D. Dubnau. 2005. The ins and outs of DNA transfer in bacteria. *Science* **310:** 1456-1460.

Chin, J.W. 2006. Modular approaches to expanding the functions of living matter. *Nat Chem Biol* **2:** 304-311.

Chothia, C. and J. Gough. 2009. Genomic and structural aspects of protein evolution. *Biochem J* **419:** 15-28.

Conant, G.C. and A. Wagner. 2003. Convergent evolution of gene circuits. *Nat Genet* **34:** 264-266.

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

Dantas, G., M.O. Sommer, R.D. Oluwasegun, and G.M. Church. 2008. Bacteria subsisting on antibiotics. *Science* **320:** 100-103.

Davies, J. and F. Jacob. 1968. Genetic mapping of the regulator and operator genes of the lac operon. *J Mol Biol* **36:** 413-417.

Dekel, E. and U. Alon. 2005. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436:** 588-592.

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

Dorman, C.J. 2007. H-NS, the genome sentinel. *Nat Rev Microbiol* **5:** 157-161.

Dorman, C.J. 2009a. Nucleoid-associated proteins and bacterial physiology. *Adv Appl Microbiol* **67:** 47-64.

Dorman, C.J. 2009b. Regulatory integration of horizontally-transferred genes in bacteria. *Front Biosci* **14:** 4103-4112.

Doyle, M., M. Fookes, A. Ivens, M.W. Mangan, J. Wain, and C.J. Dorman. 2007. An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. *Science* **315:** 251-252.

Foster, P.L. 2007. Stress-induced mutagenesis in bacteria. *Crit Rev Biochem Mol Biol* **42:** 373-397.

Gama-Castro, S., V. Jimenez-Jacinto, M. Peralta-Gil, A. Santos-Zavaleta, M.I. Penaloza-Spinola, B. Contreras-Moreira, J. Segura-Salazar, L. Muniz-Rascado, I. Martinez-Flores, H. Salgado, C. Bonavides-Martinez, C. Abreu-Goodger, C. Rodriguez-Penagos, J. Miranda-Rios, E. Morett, E. Merino, A.M. Huerta, L. Trevino-Quintanilla, and J. Collado-Vides. 2008. RegulonDB (version 6.0): gene regulation model of Escherichia coli K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. *Nucleic Acids Res* **36:** D120-124.

Gelfand, M.S. 2006. Evolution of transcriptional regulatory networks in microbial genomes. *Curr Opin Struct Biol* **16:** 420-429.

Grainger, D.C., D. Hurd, M. Harrison, J. Holdstock, and S.J. Busby. 2005. Studies of the distribution of Escherichia coli cAMP-receptor protein and RNA polymerase along the E. coli chromosome. *Proc Natl Acad Sci U S A* **102:** 17693-17698.

Grainger, D.C., D.J. Lee, and S.J. Busby. 2009. Direct methods for studying transcription regulatory proteins and RNA polymerase in bacteria. *Curr Opin Microbiol* **12:** 531-535.

Hengge, R. 2009. Principles of c-di-GMP signalling in bacteria. *Nat Rev Microbiol* **7:** 263-273.

Isalan, M., C. Lemerle, K. Michalodimitrakis, C. Horn, P. Beltrao, E. Raineri, M. Garriga-Canut, and L. Serrano. 2008. Evolvability and hierarchy in rewired bacterial gene networks. *Nature* **452:** 840-845.

Janga, S.C. and J. Collado-Vides. 2007. Structure and evolution of gene regulatory networks in microbial genomes. *Res Microbiol* **158:** 787-794.

Janga, S.C., H. Salgado, J. Collado-Vides, and A. Martinez-Antonio. 2007a. Internal versus external effector and transcription factor gene pairs differ in their relative chromosomal position in Escherichia coli. *J Mol Biol* **368:** 263-272.

Janga, S.C., H. Salgado, A. Martinez-Antonio, and J. Collado-Vides. 2007b. Coordination logic of the sensing machinery in the transcriptional regulatory network of Escherichia coli. *Nucleic Acids Res* **35:** 6963-6972.

Janky, R., J. Helden, and M.M. Babu. 2009. Investigating transcriptional regulation: from analysis of complex networks to discovery of cis-regulatory elements. *Methods* **48:** 277-286.

Jayaraman, R. 2008. Bacterial persistence: some new insights into an old phenomenon. *J Biosci* **33:** 795-805.

Jothi, R., S. Balaji, A. Wuster, J.A. Grochow, J. Gsponer, T.M. Przytycka, L. Aravind, and M.M. Babu. 2009. Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. *Mol Syst Biol* **5:** 294.

Juhas, M., J.R. van der Meer, M. Gaillard, R.M. Harding, D.W. Hood, and D.W. Crook. 2009. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol Rev* **33:** 376-393.

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

Kiel, C., E. Yus, and L. Serrano. Engineering signal transduction pathways. *Cell* **140:** 33-47.

Kitano, H. 2004. Biological robustness. *Nat Rev Genet* **5:** 826-837.

Koonin, E.V., K.S. Makarova, and L. Aravind. 2001. Horizontal gene transfer in prokaryotes: quantification and classification. *Annu Rev Microbiol* **55:** 709-742.

Kunin, V., L. Goldovsky, N. Darzentas, and C.A. Ouzounis. 2005. The net of life: reconstructing the microbial phylogenetic network. *Genome Res* **15:** 954-959.

Lee, T.I., N.J. Rinaldi, F. Robert, D.T. Odom, Z. Bar-Joseph, G.K. Gerber, N.M. Hannett, C.T. Harbison, C.M. Thompson, I. Simon, J. Zeitlinger, E.G. Jennings, H.L. Murray, D.B. Gordon, B. Ren, J.J. Wyrick, J.B. Tagne, T.L. Volkert, E. Fraenkel, D.K. Gifford, and R.A. Young. 2002. Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science* **298:** 799-804.

Leonard, E., D. Nielsen, K. Solomon, and K.J. Prather. 2008. Engineering microbes with synthetic biology frameworks. *Trends Biotechnol* **26:** 674-681.

Lerat, E., V. Daubin, H. Ochman, and N.A. Moran. 2005. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol* **3:** e130.

Lercher, M.J. and C. Pal. 2008. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol Biol Evol* **25:** 559-567.

Losick, R. and C. Desplan. 2008. Stochasticity and cell fate. *Science* **320:** 65-68.

Lozada-Chavez, I., S.C. Janga, and J. Collado-Vides. 2006. Bacterial regulatory networks are extremely flexible in evolution. *Nucleic Acids Res* **34:** 3434-3445.

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

Luijsterburg, M.S., M.C. Noom, G.J. Wuite, and R.T. Dame. 2006. The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: a molecular perspective. *J Struct Biol* **156:** 262-272.

Luijsterburg, M.S., M.F. White, R. van Driel, and R.T. Dame. 2008. The major architects of chromatin: architectural proteins in bacteria, archaea and eukaryotes. *Crit Rev Biochem Mol Biol* **43:** 393-418.

Luscombe, N.M., M.M. Babu, H. Yu, M. Snyder, S.A. Teichmann, and M. Gerstein. 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* **431:** 308-312.

Lynch, M. 2007. The evolution of genetic networks by non-adaptive processes. *Nat Rev Genet* **8:** 803-813.

Lynch, M. and J.S. Conery. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290:** 1151-1155.

Ma, H.W., J. Buer, and A.P. Zeng. 2004. Hierarchical structure and modules in the Escherichia coli transcriptional regulatory network revealed by a new top-down approach. *BMC Bioinformatics* **5:** 199.

Maamar, H., A. Raj, and D. Dubnau. 2007. Noise in gene expression determines cell fate in Bacillus subtilis. *Science* **317:** 526-529.

Madan Babu, M. and S.A. Teichmann. 2003. Evolution of transcription factors and the gene regulatory network in Escherichia coli. *Nucleic Acids Res* **31:** 1234-1244.

Madan Babu, M., S.A. Teichmann, and L. Aravind. 2006. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. *J Mol Biol* **358:** 614-633.

Mangan, S. and U. Alon. 2003. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci U S A* **100:** 11980-11985.

Mangan, S., S. Itzkovitz, A. Zaslaver, and U. Alon. 2006. The incoherent feed-forward loop accelerates the response-time of the gal system of Escherichia coli. *J Mol Biol* **356:** 1073-1081.

Marino-Ramirez, L., K.C. Lewis, D. Landsman, and I.K. Jordan. 2005. Transposable elements donate lineage-specific regulatory sequences to host genomes. *Cytogenet Genome Res* **110:** 333-341.

Marr, C., M. Geertz, M.T. Hutt, and G. Muskhelishvili. 2008. Dissecting the logical types of network control in gene expression profiles. *BMC Syst Biol* **2:** 18.

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

Martinez, J.L. 2008. Antibiotics and antibiotic resistance genes in natural environments. *Science* **321:** 365-367.

McAdams, H.H., B. Srinivasan, and A.P. Arkin. 2004. The evolution of genetic regulatory systems in bacteria. *Nat Rev Genet* **5:** 169-178.

Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. Network motifs: simple building blocks of complex networks. *Science* **298:** 824-827.

Molle, V., Y. Nakaura, R.P. Shivers, H. Yamaguchi, R. Losick, Y. Fujita, and A.L. Sonenshein. 2003. Additional targets of the Bacillus subtilis global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. *J Bacteriol* **185:** 1911-1922.

Monot, M., N. Honore, T. Garnier, N. Zidane, D. Sherafi, A. Paniz-Mondolfi, M. Matsuoka, G.M. Taylor, H.D. Donoghue, A. Bouwman, S. Mays, C. Watson, D. Lockwood, A. Khamispour, Y. Dowlati, S. Jianping, T.H. Rea, L. Vera-Cabrera, M.M. Stefani, S. Banu,

M. Macdonald, B.R. Sapkota, J.S. Spencer, J. Thomas, K. Harshman, P. Singh, P. Busso, A. Gattiker, J. Rougemont, P.J. Brennan, and S.T. Cole. 2009. Comparative genomic and phylogeographic analysis of Mycobacterium leprae. *Nat Genet* **41:** 1282-1289.

Montange, R.K. and R.T. Batey. 2008. Riboswitches: emerging themes in RNA structure and function. *Annu Rev Biophys* **37:** 117-133.

Nakamura, Y., T. Itoh, H. Matsuda, and T. Gojobori. 2004. Biased biological functions of horizontally transferred genes in prokaryotic genomes. *Nat Genet* **36:** 760-766.

Navarre, W.W., M. McClelland, S.J. Libby, and F.C. Fang. 2007. Silencing of xenogeneic DNA by H-NS-facilitation of lateral gene transfer in bacteria by a defense system that recognizes foreign DNA. *Genes Dev* **21:** 1456-1471.

Navarre, W.W., S. Porwollik, Y. Wang, M. McClelland, H. Rosen, S.J. Libby, and F.C. Fang. 2006. Selective silencing of foreign DNA with low GC content by the H-NS protein in Salmonella. *Science* **313:** 236-238.

Ooka, T., Y. Ogura, M. Asadulghani, M. Ohnishi, K. Nakayama, J. Terajima, H. Watanabe, and T. Hayashi. 2009. Inference of the impact of insertion sequence (IS) elements on bacterial genome diversification through analysis of small-size structural polymorphisms in Escherichia coli O157 genomes. *Genome Res* **19:** 1809-1816.

Osborne, S.E., D. Walthers, A.M. Tomljenovic, D.T. Mulder, U. Silphaduang, N. Duong, M.J. Lowden, M.E. Wickham, R.F. Waller, L.J. Kenney, and B.K. Coombes. 2009. Pathogenic adaptation of intracellular bacteria by rewiring a cis-regulatory input function. *Proc Natl Acad Sci U S A* **106:** 3982-3987.

Osbourn, A.E. and B. Field. 2009. Operons. *Cell Mol Life Sci* **66:** 3755-3775.

Perez, J.C. and E.A. Groisman. 2009a. Evolution of transcriptional regulatory circuits in bacteria. *Cell* **138:** 233-244.

Perez, J.C. and E.A. Groisman. 2009b. Transcription factor function and promoter architecture govern the evolution of bacterial regulons. *Proc Natl Acad Sci U S A* **106:** 4319-4324.

Pesavento, C. and R. Hengge. 2009. Bacterial nucleotide-based second messengers. *Curr Opin Microbiol* **12:** 170-176.

Price, M.N., P.S. Dehal, and A.P. Arkin. 2007. Orthologous transcription factors in bacteria have different functions and regulate different genes. *PLoS Comput Biol* **3:** 1739-1750.

Purnick, P.E. and R. Weiss. 2009. The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol* **10:** 410-422.

Raj, A. and A. van Oudenaarden. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* **135:** 216-226.

Ronen, M., R. Rosenberg, B.I. Shraiman, and U. Alon. 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.

Schirmer, T. and U. Jenal. 2009. Structural and mechanistic determinants of c-di-GMP signalling. *Nat Rev Microbiol* **7:** 724-735.

Sharma, C.M., S. Hoffmann, F. Darfeuille, J. Reignier, S. Findeiss, A. Sittka, S. Chabas, K. Reiche, J. Hackermuller, R. Reinhardt, P.F. Stadler, and J. Vogel. The primary transcriptome of the major human pathogen Helicobacter pylori. *Nature* **464:** 250-255.

Shen-Orr, S.S., R. Milo, S. Mangan, and U. Alon. 2002. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat Genet* **31:** 64-68.

Sorek, R., Y. Zhu, C.J. Creevey, M.P. Francino, P. Bork, and E.M. Rubin. 2007. Genome-wide experimental determination of barriers to horizontal gene transfer. *Science* **318:** 1449-1452.

Steen, E.J., Y. Kang, G. Bokinsky, Z. Hu, A. Schirmer, A. McClure, S.B. Del Cardayre, and J.D. Keasling. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463:** 559-562.

Steidler, L., W. Hans, L. Schotte, S. Neirynck, F. Obermeier, W. Falk, W. Fiers, and E. Remaut. 2000. Treatment of murine colitis by Lactococcus lactis secreting interleukin-10. *Science* **289:** 1352-1355.

Stoebel, D.M., A. Free, and C.J. Dorman. 2008. Anti-silencing: overcoming H-NS-mediated repression of transcription in Gram-negative enteric bacteria. *Microbiology* **154:** 2533-2545.

Storz, G., S. Altuvia, and K.M. Wassarman. 2005. An abundance of RNA regulators. *Annu Rev Biochem* **74:** 199-217.

Studier, F.W., P. Daegelen, R.E. Lenski, S. Maslov, and J.F. Kim. 2009. Understanding the differences between genome sequences of Escherichia coli B strains REL606 and BL21(DE3) and comparison of the E. coli B and K-12 genomes. *J Mol Biol* **394:** 653-680.

Suel, G.M., J. Garcia-Ojalvo, L.M. Liberman, and M.B. Elowitz. 2006. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440:** 545-550.

Suel, G.M., R.P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo, and M.B. Elowitz. 2007. Tunability and noise dependence in differentiation dynamics. *Science* **315:** 1716-1719.

Taoka, M., Y. Yamauchi, T. Shinkawa, H. Kaji, W. Motohashi, H. Nakayama, N. Takahashi, and T. Isobe. 2004. Only a small subset of the horizontally transferred chromosomal genes in Escherichia coli are translated into proteins. *Mol Cell Proteomics* **3:** 780-787.

Teichmann, S.A. and M.M. Babu. 2004. Gene regulatory network growth by duplication. *Nat Genet* **36:** 492-496.

Teichmann, S.A., J. Park, and C. Chothia. 1998. Structural assignments to the Mycoplasma genitalium proteins show extensive gene duplications and domain rearrangements. *Proc Natl Acad Sci U S A* **95:** 14658-14663.

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

Venancio, T.M. and L. Aravind. 2009. Reconstructing prokaryotic transcriptional regulatory networks: lessons from actinobacteria. *J Biol* **8:** 29.

Waters, L.S. and G. Storz. 2009. Regulatory RNAs in bacteria. *Cell* **136:** 615-628.

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

Weber, H., T. Polen, J. Heuveling, V.F. Wendisch, and R. Hengge. 2005. Genome-wide analysis of the general stress response network in Escherichia coli: sigmaS-dependent genes, promoters, and sigma factor selectivity. *J Bacteriol* **187:** 1591-1603.

Xu, J. and D.A. Lavan. 2008. Designing artificial cells to harness the biological ion concentration gradient. *Nat Nanotechnol* **3:** 666-670.

Yu, H. and M. Gerstein. 2006. Genomic analysis of the hierarchical structure of regulatory networks. *Proc Natl Acad Sci U S A* **103:** 14724-14731.

Zaslaver, A., A. Bren, M. Ronen, S. Itzkovitz, I. Kikoin, S. Shavit, W. Liebermeister, M.G. Surette, and U. Alon. 2006. A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nat Methods* **3:** 623-628.

Zaslaver, A., A.E. Mayo, R. Rosenberg, P. Bashkin, H. Sberro, M. Tsalyuk, M.G. Surette, and U. Alon. 2004. Just-in-time transcription program in metabolic pathways. *Nat Genet* **36:** 486-491.

FIGURE LEGENDS

Figure 1: Structure of transcriptional regulatory network (A) The basic unit consists of a transcription factor (TF) which recognised specific regulatory sequence upstream of its target gene (TG) (B) At the local level, the basic units assemble to form network motifs: the Feed-Forward Motif (FFM), Single Input Motif (SIM) and Multiple Input Motif (MIM). (C) At the global level, transcriptional regulatory networks display a scale-free topology, which is characterised by the presence of a few TFs (hubs or global regulators) that regulate many genes and many TFs that regulate a few genes.

Figure 2: The major evolutionary forces that drive transcriptional regulatory network evolution.

Figure 3: General principles of evolution at three distinct levels of network organisation.

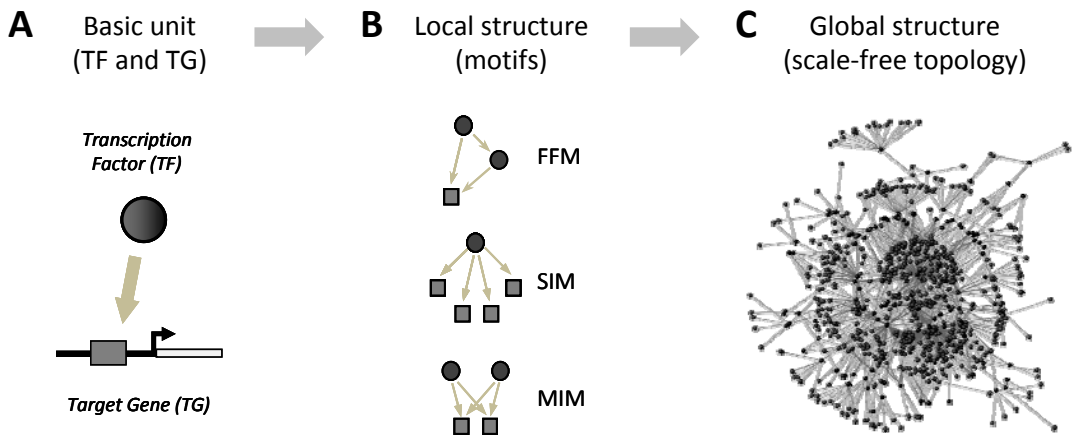

Figure 1: Structure of transcriptional regulatory network. (A) The basic unit consists of a transcription factor (TF) which recognised specific regulatory sequence upstream of its target gene (TG) (B) At the local level, the basic units assemble to form network motifs: the Feed-Forward Motif (FFM), Single Input Motif (SIM) and Multiple Input Motif (MIM). (C) At the global level, transcriptional regulatory networks display a scale-free topology, which is characterised by the presence of a few TFs (hubs or global regulators) that regulate many genes and many TFs that regulate a few genes.

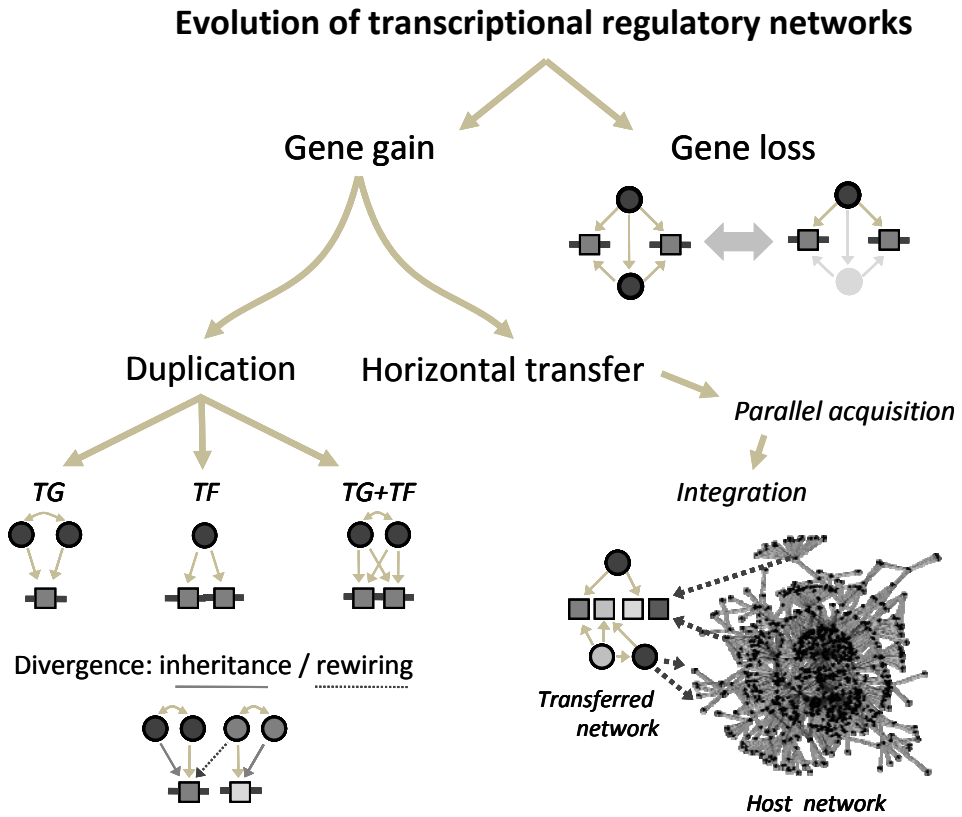

Figure 2: The major evolutionary forces that drive transcriptional regulatory network evolution.

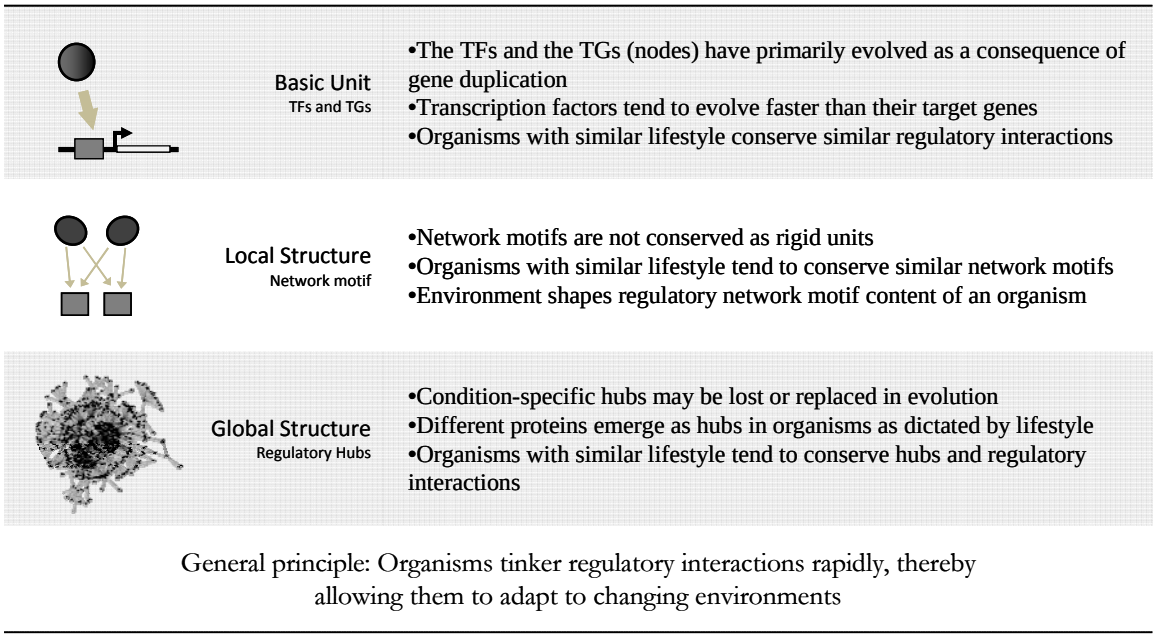

Figure 3: General principles of evolution at three distinct levels of network organisation.

Table 1: Databases and computer programs for investigating transcriptional regulatory networks. Adapted from Babu MM (Babu 2008) and Janky *et al*. (Janky et al. 2009).

Databases containing regulatory information	Comment	Website
RegTransBase	TF-binding sites and regulatory interactions	http://regtransbase.lbl.gov/cgi-bin/regtransbase?page=main
ORegAnno	An open access database for gene regulatory element and polymorphism annotation	http://www.oreganno.org/
STRING	Genome context and SMART (simple modular architecture research tool), domain assignment	http://smart.embl-heidelberg.de/
RegulonDB	Database of TFs and binding sites for *E. coli*	http://regulondb.ccg.unam.mx/
DBTBS	Database of TFs and binding sites for *B. subtilis*	http://dbtbs.hgc.jp/
Coryneregnet	Database of regulatory network for several microbes	http://www.coryneregnet.de/
Prodoric	Prokaryotic database of gene regulation	http://www.prodoric.de/
TractorDB	Predicted TF-binding sites in gamma proteobacterial genomes	http://www.tractor.lncc.br/
Microbes Online	Domain assignment, expression data, evolutionary relationships and operon structure	http://www.microbesonline.org/
BacTregulators	Database of transcription factors in bacteria and archaea	http://www.bactregulators.org/
DBD	Database of predicted transcription factors of over 700 completely sequenced genomes based on SCOP DNA binding domains	http://dbd.mrc-lmb.cam.ac.uk/DBD/index.cgi?Home
RegPrecise	Database of curated genomic inference of regulons in prokaryotic genomes	http://regprecise.lbl.gov/RegPrecise/
Transfac	Transcription factor database	http://www.biobase-international.com/pages/index.php?id=transfac
ArchaeaTF	Archaeal transcription factor database	http://bioinformatics.zj.cn/archaeatf/Homepage.php
Tools for analysis of transcription regulation	**Comment**	**Website**
Vista	Tools for comparative analysis of genomic sequences	http://genome.lbl.gov/vista/index.shtml
RSAT	A very powerful platform for regulatory sequence analysis	http://rsat.ulb.ac.be/rsat/
Webmotifs	motif discovery, scoring, analysis, and visualization using different programs	http://fraenkel.mit.edu/webmotifs/finalout.html

seqVISTA	Platform for binding site discovery	http://zlab.bu.edu/SeqVISTA/index.htm
Weblogo	Visualizing binding site information	http://weblogo.berkeley.edu/
Enologos	Logo visualization	http://biodev.hgen.pitt.edu/cgi-bin/enologos/enologos.cgi
Network visualization	**Comment**	**Website**
Biolayout	Visualization	http://cgg.ebi.ac.uk/services/biolayout/
Cytoscape	Visualization and analysis	http://www.cytoscape.org/
GraphViz	Visualization	http://www.graphviz.org/
H3Viewer	Visualization	http://graphics.stanford.edu/~munzner/h3/
Neat	Visualization and analysis	http://rsat.ulb.ac.be/rsat/index_neat.html
Netminer	Visualization and analysis (Commercial)	http://www.netminer.com/
Osprey	Visualization and analysis	http://biodata.mshri.on.ca/osprey/index.html
Pajek	Visualization and analysis	http://vlado.fmf.uni-lj.si/pub/networks/pajek/
Visant	Visualization and analysis	http://visant.bu.edu/
Yed	Visualization and analysis	http://www.yworks.com/
Network analysis	**Comment**	**Website**
Mfinder	Network motif finder	http://www.weizmann.ac.il/mcb/UriAlon/groupNetworkMotifSW.html
FanMod	Network motif finder	http://www.minet.uni-jena.de/~wernicke/motifs/
Clique finder	Identification of cliques	http://topnet.gersteinlab.org/clique/
MCode	Identification of densely connected sub-network	http://baderlab.org/Software/MCODE
Cytoscape	Several plugins in cytoscape allows advanced analysis of network topology	http://www.cytoscape.org/
Vanted	Analysis of network with experimental data	http://vanted.ipk-gatersleben.de/
Biotapestry	Drawing, analysis and visualization	http://www.biotapestry.org/
TYNA / Topnet	Network analysis	http://tyna.gersteinlab.org/tyna/
NCT	Network comparison toolkit	http://chianti.ucsd.edu/nct/
Bioconductor	Network analysis and visualization	http://www.bioconductor.org/