Chapter 1
Towards Synthetic Gene Circuits with Enhancers: Biology's Multi-input Integrators

Roee Amit

Abstract One of the greatest challenges facing synthetic biology is to develop a technology that allows gene regulatory circuits in microbes to integrate multiple inputs or stimuli using a small DNA sequence "foot-print", and which will generate precise and reproducible outcomes. Achieving this goal is hindered by the routine utilization of the commonplace σ^{70} promoters in gene-regulatory circuits. These promoters typically are not capable of integrating binding of more than two or three transcription factors in natural examples, which has limited the field to developing integrated circuits made of two-input biological "logic" gates. In natural examples the regulatory elements, which integrate multiple inputs are called enhancers. These regulatory elements are ubiquitous in all organisms in the tree of life, and interestingly metazoan and bacterial enhancers are significantly more similar in terms of both Transcription Factor binding site arrangement and biological function than previously thought. These similarities imply that there may be underlying enhancer design principles or grammar rules by which one can engineer novel gene regulatory circuits. However, at present our current understanding of enhancer structure-function relationship in all organisms is limited, thus preventing us from using these objects routinely in synthetic biology application. In order to alleviate this problem, in this book chapter, I will review our current view of bacterial enhancers, allowing us to first highlight the potential of enhancers to be a game-changing tool in synthetic biology application, and subsequently to draw a road-map for developing the necessary quantitative understanding to reach this goal.

Keywords Bacterial enhancers • Biological computation • Digital computation • Gene regulatory networks • Synthetic enhancer

R. Amit (✉)
Department of Biotechnology and Food Engineering,
Technion – Israel Institute of Technology, Haifa 32000, Israel
e-mail: roeeamit@technion.ac.il

X. Wang et al. (eds.), *Reprogramming Microbial Metabolic Pathways*,
Subcellular Biochemistry 64, DOI 10.1007/978-94-007-5055-5_1,

Abbreviations

NRI	nitrogen regulation I
RNAP	RNA polymerase
TF	transcription factors
TSS	transcriptional start site

1.1 Introduction

The ability of living organisms from all branches of the tree of life to convert a complex palette of variable input signals into discrete output levels that in turn trigger cell differentiation, morphogenesis, stress responses, complex metabolic reactions, or a host of other cellular phenomenon is one of the great mysteries of modern biology. One way to achieve this involves an amalgam of gene network interactions with complex regulatory regions. While the gene regulatory subnetwork structure (Ackers et al. 1982; Bintu et al. 2005a, b; Bolouri and Davidson 2002; Buchler et al. 2003; Rosenfeld et al. 2005; Stathopoulos and Levine 2005) and the input/output relationship between different genes is becoming better defined for many systems, the structure, binding-site arrangement, and the underlying mechanisms responsible for the regulatory output remains for the most part undeciphered. Consequently, our ability to engineer novel gene-regulatory circuits is severely hindered by this knowledge gap, where the process of integration of multiple regulatory inputs remains poorly understood.

1.1.1 Gene Regulation in Bacteria

In bacteria, the prevailing view of transcriptional regulation is built around the idea of regulated recruitment of RNA polymerase and the dissociable sigma factor σ^{70}. In this picture, the presence or absence of RNA polymerase at a promoter of interest is dictated by the corresponding presence or absence of batteries of transcription factors that either increase (activators) or decrease (repressors) the probability of polymerase binding. An increasingly sophisticated understanding of this kind of regulatory response has resulted in an explosion of efforts in synthetic and systems biology research efforts built using a broad palette of different activators and repressors for a range of different promoters (Belyaeva et al. 1998; Bintu et al. 2005a; Elowitz and Leibler 2000; Gardner et al. 2000; Guido et al. 2006; Joung et al. 1993, 1994; Muller et al. 1996).

Another whole set of bacterial promoters utilize an alternative sigma factor (σ^{54}) which together with RNAP form a stable closed promoter complex that, unlike its σ^{70} counterpart, is unable to initiate transcription by itself (Amit et al. 2011; Buck

et al. 2000; Bulger and Groudine 2011; Ninfa et al. 1987; Rappas et al. 2007). This effectively causes the polymerase to be poised at the gene of interest awaiting the arrival of a transcription factor partner termed the "driver", which releases the polymerase. Consequently, these promoters are regulated in a different fashion than their recruitment counterparts. The activating or transcription driving complex is typically widely separated from the promoter (100–1,000 bp (Ninfa et al. 1987)), precluding it from forming direct contact with the poised polymerase. It has been shown that DNA looping (Amit et al. 2011; Huo et al. 2006; Schulz et al. 2000; Su et al. 1990) and ATP hydrolysis (Rappas et al. 2007) are required to induce open complex formation and transcription initiation. These regulatory regions belong to a different class of regulatory elements called bacterial enhancers, whose structure and function are similar to their eukaryotic counterparts.

1.1.2 Enhancers – General Structure and Mode of Action

Enhancer elements or cis regulatory modules are ubiquitous in all genomes (Buck et al. 2000; Bulger and Groudine 2011; Rappas et al. 2007). It is hypothesized that enhancers execute their regulatory program by making direct contact with the basal promoter via DNA or chromatin looping. In general, they are made up of contiguous genomic regions that stretch from tens to thousands of base-pairs, contain several binding sites for a variety of transcription factors (TF), and often their regulatory output is independent of their location or orientation relative to the basal promoter (Amit et al. 2011; Amit and Phillips 2012; Driever and Nusslein-Volhard 1989; Driever et al. 1989; Huo et al. 2006; Ninfa et al. 1987). Furthermore, enhancers, like gene-regulatory networks themselves, can be viewed qualitatively (Amit et al. 2011) as modular entities, which in this case are made of three connected irreducible parts: the driver binding sites responsible for initiation of transcription, transcription factor binding sites responsible for the modulation of expression levels, and a basal promoter. In these systems, a basal promoter has the capability to generate little or no transcriptional output on its own (Aida et al. 2006; Boehm et al. 2003; Bulger and Groudine 2011; Gilmour 2009; Magasanik 1993; Muse et al. 2007; Ninfa et al. 1987; Rappas et al. 2007; Zeitlinger et al. 2007) but together with the rest of the enhancer it can express its full regulatory potential (Atkinson et al. 2002, 2003; Davidson 2001, 2006; Lee and Schleif 1989; Magasanik 1993; Small et al. 1992; Yuh et al. 2001). Even though many aspects associated with enhancer regulation are routinely studied in natural systems with state-of-the-art techniques in both bacteria (Amit et al. 2011; Atkinson et al. 2002, 2003; Huo et al. 2006; Ninfa et al. 1987) and higher eukaria (Bolouri and Davidson 2002; Davidson 2006; Stathopoulos and Levine 2005), the underlying mechanisms of regulatory "action-at-a-distance" responsible for integrating the various inputs in enhancers remain elusive. In order to fulfill the potential promised within synthetic gene regulatory circuits, we must rapidly close our knowledge gap between the relatively advanced understanding of dynamic phenomenon associated with gene subnetwork motifs with our meager

grasp of the underlying biophysical mechanisms that are responsible for producing the enhancer regulatory output. It is the purpose of this book chapter to outline the potential benefits of utilizing enhancers routinely in synthetic biology applications, and to draw a road-map that will guide the development of the necessary knowledge base to facilitate this capability.

1.2 Structure and Function of Bacterial Enhancers

1.2.1 Enhancer Architecture and Transcriptional Kinetics

Bacterial enhancers are highly modular objects, whose binding site architecture can be grossly divided into three distinct modules (Fig. 1.1a). The driver module is typically associated with either a tandem of or three specialized binding sites that are located between 50 and 500 bp upstream of the basal promoter. The driver binding sites facilitate the cooperative assembly of a hexameric ATPase (e.g. NRI/NtrC, PspF, etc.) belonging to the AAA+family. These ATPases exist in the cytoplasm as dimers, each capable of individually binding one of the binding sites in the driver module. The assembly of the hexameric complex, apparently occurs as a result of the binding of two dimers, to which a third cooperatively binds to complete the assembly. The cooperative nature of this binding ensures that the hexameric complex is highly stable, reminiscent of other AAA+DNA bound hexamers (e.g. RuvB Amit et al. 2004) that also have an increased binding affinity as an assembled complex vs. the cytoplasmic dimers.

The second module encompasses the region in between both the promoter and driver binding sites. This region typically contains a multitude of binding sites for several (1–5) transcription factors, and its main role is to modulate the expression level that would be generated if no proteins were bound. This modulation was demonstrated recently on one natural system (Atkinson et al. 2002), and with two synthetic enhancer systems (Amit 2012; Amit et al. 2011; Huo et al. 2006), showing that the expression level can either be inhibited (repressed) or amplified (activated) depending on the type of protein that binds, the number of binding sites, the location of the binding sites with respect to the promoter within this region, and the spacing between the binding sites.

Finally, the third module is the basal promoter, which in this case binds the σ^{54}-RNAP holoenzyme complex. This module is responsible for integrating all the inputs thereby generating a particular expression pattern at some integrated rate. The integration of the inputs takes place via a sequential kinetic mechanism, whereby the σ^{54}-RNAP holoenzyme complex binds the promoter, but is unable to initiate expression, and as a result remains paused at the Transcriptional Start Site (TSS). Simultaneously, the rest of the transcription apparatus assembles at the various binding sites on the enhancer. Transcription is facilitated when the upstream assembled driver complex (e.g. NtrC – the "driver" (Amit et al. 2011)) loops and makes directs contact with the poised σ^{54}-RNAP complex. The driver

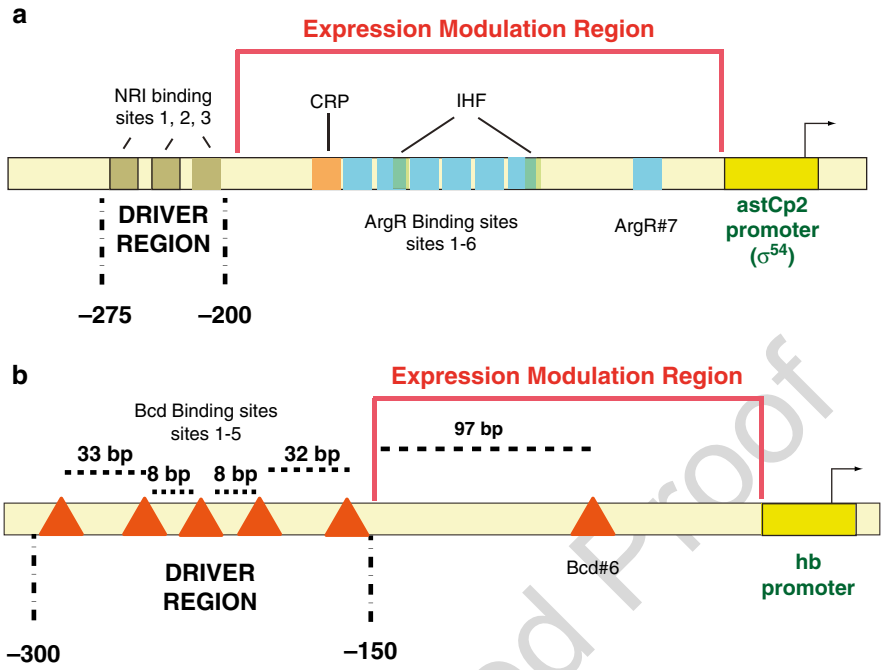

Fig. 1.1 Bacteria and Drosophila enhancers. The figure is a schematic designed to pictorially convey the similarities between a sample bacterial enhancer and a near-promoter Eukaryotic enhancer. (**a**) The astCp2 enhancer in *E. coli*, exhibiting a ~200 bp expression modulation region, at least ten binding sites for three different kinds of TFs, and three NRI~P driver binding sites (Kiupakis and Reitzer 2002). (**b**) The hb (hunchback) gene enhancer in D. melanogaster showing a very similar architecture (Driever and Nusslein-Volhard 1989; Driever et al. 1989) to bacterial enhancers in terms of binding sites, proximity to promoter, and binding site separation. Note, that in this case Bcd is also the driver in this system. This enhancer and others with similar binding-site architectures can serve as model systems for an initial Eukaryotic Rosetta stone algorithm

has a special amino acid loop termed GAFTGA (to signify the amino-acid content (Rappas et al. 2007; Zhang et al. 2002)), which effectively enters a specialized slot within the poised holoenzyme complex, similar to a "key in a hole" mechanism. Upon binding, subsequent ATP hydrolysis by the driver generates a conformational change within the holoenzyme complex, which in turn alleviates the poised state, allowing transcription to progress. The expression modulation region affects the looping rate by inducing certain structural effects that either increases the probability for a successful looping event or decreases it. Since the rates associated with ATP hydrolysis and subsequent conformational changes are fast, the rate of looping becomes the determining rate-limiting factor for transcription. Consequently, the bacterial enhancer architecture allows the promoter to modulate the expression rate based on the transcription factor content that is found upstream, thereby allowing it to function as a form of biological integrator (Amit 2012; Amit et al. 2011).

1.2.2 Biological Function

Most bacteria contain some version of the σ^{54} sigma factor. A few well-known examples include the nitrogen regulation protein C (NRI or NtrC), the nitrogen fixation protein A (NifA), the C4-dicarboxylic acidic transport protein D (DctD), the phage shock protein F (PspF), the xylene catabolism regulatory protein (XylR) and the 3,4-dimethylphenol catabolism regulatory protein (DmpR) (Xu and Hoover 2001 and references therein). A close examination of all of these examples indicates that the σ^{54} regulated genes are often activated in response to various stresses and growth inhibiting conditions (Buck et al. 2000). In such cases, bacterial cell responds to the stress by turning on a dormant metabolic pathway in order for it to cope successfully with the stress. Such a massive shifting of transcriptional resources is in many ways akin to a bacterial form of cell differentiation into a specialized cell-type designed to cope with the stressful environment.

In addition, σ^{54} promoters (Rappas et al. 2007; Xu and Hoover 2001; Zhang et al. 2002) are also over-represented in genes that play an important role in bacterial developmental-like processes. This includes the two-component nitrogen response pathway and related systems, which exhibits regulatory and signaling characteristics that are also reminiscent of a primitive developmental like process (Goldman et al. 2006; Magasanik 1993; Ninfa and Peng 2005). A more telling example is the involvement of σ^{54} promoters in the formation of *M. Xanthus* fruiting bodies (Goldman et al. 2006). In particular, a recent genomic analysis carried out on M. Xanthus genomes as compared with other bacterial genomes (Goldman et al. 2006; Jelsbak et al. 2005) revealed that the comparative number of σ54 promoters as a function of genome size is much larger as compared with other bacterial species, indicating that these promoters are likely associated with specialized biological functions in M. Xanthus fruiting body development.

Interestingly, in Eukaryotes, promoters that go through promoter proximal pausing gene activation were also found to be over-represented in developmentally-important or cell-differentiation type of processes. Recently, (Aida et al. 2006; Guenther et al. 2007; Muse et al. 2007; Rasmussen and Lis 1993; Zeitlinger et al. 2007; Zhang et al. 2007) have showed that in metazoan organisms ranging from humans to flies, "paused" genes, known to be off at particular developmental stages, tissues, or based on ambient environmental conditions were found to be occupied by an active PolII transcriptional complex (with nascent transcript) localized 20–60 nt from the Transcriptional Start Site (TSS). Release of a paused polymerase from its stalled state requires a secondary event of looping from an upstream region, which allows a specialized protein called pTEF-b (Cheng and Price 2007; Renner et al. 2001), to phosphorylate several sites on the paused PolII (Boehm et al. 2003), which in turn allows transcription to progress.

Consequently, a form of enhancer-regulated paused or poised transcription is ubiquitous to all biological kingdoms, and seems to be over-represented in genes that are known to play an important role in executing some sort of a "developmental-like" or "cell-differentiation" type program. Such programs also seem to be characterized by both precision and often synchronized initiation of transcription along a cluster of cells.

Therefore, it is tempting to speculate, that there may be some characteristic inherent to enhancer regulatory structure (Fig. 1.1b) as well as with the activation sequential kinetics, which endows the enhancer regulatory response in all organisms with precision, discrete, and possible synchronized behavior.

1.3 Engineering Gene Circuits with Synthetic Enhancers

The main premise of synthetic biology is to use "biological parts" to construct novel biological composite objects for a variety of applications from basic research to personal medicine. In order to achieve this goal and to develop this technology, a more "engineering" friendly scheme had to be adopted to describe biological function. Given the rapid development and penetration of Information Technology via the "digital computer" in recent decades, it became quaint to compare a gene being turned on/off to a process of switching a bit from 0 to 1, which as a result led to the adoption of a computational language and Boolean algebra as a generalizing mechanism (Andrianantoandro et al. 2006). This scheme was particularly commensurate with the biological function of the more commonplace bacterial σ^{70} promoters, which generate transfer functions reminiscent of sigmoidal functions that characterize digital on/off switches. Consequently, this property has made them highly attractive for designing simple computational modules from elementary biological parts (i.e. genes, promoters, etc.), and amongst other reasons has led to almost exclusive utilization of these objects in the first generation of synthetic biology works.

Unlike σ^{70} bacterial promoters, the coupled enhancer-σ^{54} promoter systems have been completely ignored by the community in the early days of synthetic biology except for one notable exception (Atkinson et al. 2003). In addition to this work, bacterial enhancers so far have been used in three additional synthetic biology works (Amit et al. 2011; Amit and Phillips 2012; Huo et al. 2006) with no real applications as of yet. In this section, I will explain in computational terms why this underutilization of enhancers is expected to change as we move into developing the next phase of synthetic biology applications.

1.3.1 Biological Computation at the Gene Regulatory Level

One of the major challenges of synthetic biology is to engineer compact, yet complex gene regulatory networks capable of carrying out complex computational operations in a precise fashion. In this case "computation" means a type of calculation process that follows a well-defined model expressed as an algorithm, protocol, network topology, or any other set of predefined rules. From a biological perspective such processes may involve sensing and processing a whole palette of chemical input signals, deciding on where and when a particular gene should be expressed, dividing into particular cell types, regulatory responses, etc. Thus far, the major workhorse

used to demonstrate novel synthetic biological computational processes have been synthetic gene regulatory circuits implemented using standard bacterial σ^{70} promoters.

Promoters that belong to this family are often regulated (i.e. turned on and off) by transcription factors whose binding sites are either in the vicinity or over-lapping the RNAP binding region. Due to the transcription factors' binding sites proximity to their cognate promoters, these proteins regulate gene expression by either preventing RNAP from binding, or by recruiting RNAP and increasing the probability for transcription. As a result, the transfer functions that depict how these promoters activate gene expression as a function of intracellular transcription factor concentrations are highly reminiscent of sigmoidal switching behavior, which has been compared to a form of binary digital computation and attributed properties of buffer gates (Andrianantoandro et al. 2006; Gardner et al. 2000). This characteristic of gene expression has been one of the primary drivers for the engineering of gene-regulatory circuits that function as "noisy" biological binary-logic gates (Andrianantoandro et al. 2006 and references therein), and subsequent construction of simple circuits made of connected biological digital gates (Anderson et al. 2007). Since the gene products (proteins or RNA molecules) of these biological logic gates can be utilized with minimal effort to either feed-back on their own promoter, or participate in further down-stream regulation, such efforts have led to a plethora of implementations of composite biological circuits made of several interconnected biological gates that have been shown to be capable of executing simple computational operations akin to simple electronic circuits (Basu et al. 2005; Friedland et al. 2009; Tabor et al. 2009; Tasmir et al. 2011).

Despite the rapid progress achieved over the last 10 years with increasingly complex circuits capable of carrying out sophisticated computational algorithms, σ^{70} recruitment promoters are not capable of generating transfer functions that are sufficiently close to the digital ideal. First, the process of induction, which generates the transition between "gene-off" to "gene-on" states is typically spread over a wide-range of inducer or transcription factor concentrations. This, in turn, yields an extended range where a gradiated response is observed, which is characteristic of analog computational processes. Consequently, the sharp switching that characterizes electronic binary digital gates cannot be simply engineered with the biological versions.

Moreover, in order to execute complex computational algorithms using biological binary digital computation that often relies on "wiring" together a whole set of two-input digital gates (e.g. AND, OR, etc.) to carry out simple Boolean computations, many regulatory components are needed. Since σ^{70} promoters necessitate that the TF binding sites be present within a close proximity, individual promoters can integrate only one or two signals. This, in turn, means that in order to program cells to carry out complex computational processes, large gene regulatory network circuits with many nodes need to be designed, which translates to generating a need for engineering very large sequences, whose growth potential is limited by the biological vessel that will execute the computation. Since bacterial cells are capable of encasing ~1–10 Mbp of DNA, this suggests that very quickly a glass ceiling of computational complexity will be reached using the binary paradigm.

Finally, unlike electronic computers, which are not subject to thermal noise, biological computation is subject to large thermal noise effects. This in turns renders any biological computation operation a stochastic process, which is by definition subject to different modeling rules, then the deterministic processes that characterize conventional computational processes. In particular, the recruitment transcriptional process is particularly susceptible to molecular noise (Elowitz and Leibler 2000; Elowitz et al. 2002; Thattai and van Oudenaarden 2001), which makes this problem even more of an acute issue for these systems. Consequently, at present, any computational processes that are carried out by biological modules are not only limited in computational capability, but are also imprecise and noisy. Yet, despite the physical, energetic, and thermal limitations, natural biological computation is capable of executing tremendously complex and precise computational operations at the gene regulatory level. So the question remains how do we overcome these limitations and develop a technology that can carry out precise and reproducible molecular computation operations?

1.3.2 Biological Computation with Enhancers

Unlike σ^{70} promoters, σ^{54} are always coupled to bacterial enhancers, and in effect can be considered to be one large regulatory unit. This unit includes a multitude of binding sites for many transcription factors, which in turn can support the integration of many different input signals. Thus, enhancers provide a convenient platform for engineering Boolean digital gates with multiple inputs (n), which allows 2^2^n computational operations to be carried out at a single promoter as compared with approximately 16 for a σ^{70} promoter (e.g. an enhancer capable of integrating three or four input will support 256 AND 65,536 different computational operations respectively.) As an example for the utility and compactness of enhancer-based computation as compared with σ^{70} recruitment promoter, consider constructing a three or four input AND gate. With the latter system this will require the utilization of at least two different chemically wired two-input gate promoters, while with enhancers these operations can be carried out at a single promoter.

Another advantage of enhancers is the capability to engineer interactions between transcription factors that are bound adjacently to one another. In the case of cooperative interactions between transcription factors, enhancer output will be characterized by transfer functions (Fig. 1.2a) whose transition region occupies smaller TF concentration ranges that are much closer to the digital ideal. Alternatively, anti-cooperative or mutually exclusive interactions between bound transcription factors on the enhancer can generate transfer functions with more than two "stable" states (Fig. 1.2b, c). Having more than two stable output states supports a non-Boolean digital computation model, where instead of a 0 or 1 output, the enhancer can generate a 0,1,2 or more output. Digital computation with more than two discrete input/output states offers a much larger computational flexibility (Table 1.1), as the number of possible algebraic operations with a 2 or 3-input gates increases exponentially

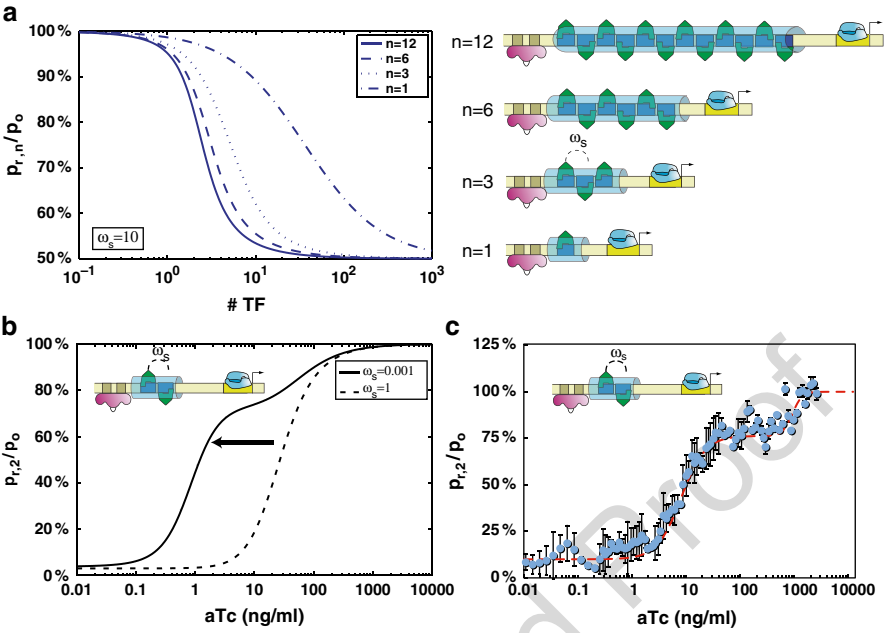

Fig. 1.2 Enhancer Transfer Function. (**a**) Transfer functions for bacterial enhancers characterized by two input signals: driver bound to a tandem of binding sites upstream of the poised RNAP, and a variable number of binding sites (1,3,6,12) in the expression modulation region for some Transcription Factor (TF). The TF is assumed to rigidify the DNA when bound to the DNA leading to repression (see SI of (Amit et al. 2011) for definition of repression and the values on the y-axis), and to bind the enhancer cooperatively (quantified by a protein-interaction parameter ωs >1). The model shows that given these assumptions, it is possible to generate sharper transfer functions by simply increasing the number of binding sites. (**b**) Alternatively, one can generate a step-like response in the model using the exact same binding architecture by setting the protein-interaction parameter to some value $\omega s \ll 1$. Therefore, a wide-array of possible transfer functions may be possible depending on a handful of characteristics such as type of protein bound, number of binding sites, spacing between binding sites, etc. (**c**) Data published previously (Amit et al. 2011) showing that by varying the number of active TetR proteins inside cells via the inducer anhydrous-tetracycline (aTc) and using an enhancer structure containing two TetR binding sites with 16 bp spacing between sites, a step-like response is generated. For further detail of model and experimental data see (Amit 2012; Amit et al. 2011; Amit and Phillips 2012)

with the number of possible output states. Consequently, the enhancer's capability to both integrate multiple inputs and to generate multiple stable state transfer functions endows them with a tremendous computational flexibility and complexity, which can only be produced by σ^{70} based gene circuits that are composed of multiple promoters and require a significantly larger sequence signature.

Finally, unlike electronic computers, where thermal noise plays a minor role, in biological systems thermal noise plays a critical role in regulation. The sources of noise of have been enumerated and quantified in several recent publications (Elowitz et al. 2002; Friedman et al. 2006; Golding et al. 2005; Ozbudak et al. 2002; Paulsson

[AU1]

Table 1.1 Computational complexity of different digital computational systems

# inputs (n)	Boolean-logic[a,b]	Combined-logic[c,d]	3-logic[e]
1	4	8	9
2	16	512	3^9
3	256	2^{27}	3^{27}
N	2^2^n	2^3^n	3^3^n

[a] The number of possible operations with conventional "Boolean-logic" defined as 2^2^n operations

[b] The left "2" in 2^2^n corresponds to the number of values each input channel can accept, the middle "2" to the number of output values that are possible, and n is the number of input channels

[c] The number of possible operations with "Combined-logic" defined as 2^3^n operations

[d] 2^3^n operations can be executed by a gate that can accept only two values (i.e. protein bound/unbound), but can output three discrete values (e.g. enhancer)

[e] The number of possible operations with "3-logic" defined as 3^3^n operations. "3-logic" is the "three" version of Boolean logic, which can also be implemented with enhancers

2004; Pedraza and Paulsson 2008; Raser and O'Shea 2004; Sanchez et al. 2011; Thattai and van Oudenaarden 2001), and are mostly due to a small finite number of interacting objects, the kinetics of binding and unbinding, and variation of different molecular species from cell to cell (i.e. some cells may have more RNAP molecules available than others, etc.). A major challenge of synthetic biology is to not only construct synthetic circuits capable of carrying out complex computation, but to do so with minimal noise effects. Since noise is an additive quantity (Pedraza and Paulsson 2008; Sanchez et al. 2011), circuits with multiple promoters and components are inherently susceptible to thermal noise, and as such must constantly add elements that simultaneously mitigate the deleterious noise affects (Andrianantoandro et al. 2006). Enhancers, on the other hand, which presumably can carry out complex calculations at a single promoter, double almost by default as a noise-minimizing mechanism due to the compactness of the molecular design. Consequently, enhancers have the potential to not only allow us to code complex biological algorithm, but to do so in a noise-minimal fashion as well.

1.3.3 Putting It Altogether – Constructing Circuits

Despite their potential, to date only three synthetic biology works have utilized enhancers to generate novel regulatory effects. These works have either altered bacterial enhancers to generate novel regulatory effects (Amit et al. 2011; Huo et al. 2006) or wired two enhancers together to generate a damped oscillator characterized by a periodicity that was an order of magnitude or so larger than the standard time-scale associated with a bacterial cell-cycle (Atkinson et al. 2003).

In the former works, Huo et al. (2006)) showed that by careful positioning of a binding site for IHF at different locations along the enhancer, the regulatory effects

can either be sharply repressive or highly activating, with a periodicity that is commensurate with the DNA helical pitch. In a recent work, we carried out a systematic analysis of many synthetic enhancers, which showed that altering the enhancer's ability to loop using bound transcription factors affects regulation. We (Amit et al. 2011) were able to show transfer functions (Fig. 1.2c) that are characterized by multiple output levels with sharp transitions between states that pointed to a combined cooperative and anti-cooperative effect (Amit 2012) in the binding of TetR proteins. The next stage will be to utilize these libraries of characterized synthetic enhancers to engineer gene-circuits capable of carrying out complex computation in a noise minimizing and compact genomic architecture. My lab is advancing towards this goal with our current research.

Based on these early achievements, it seems that utilizing synthetic enhancers to construct synthetic gene circuits promises to generate some very interesting applications in the very near future. Complex circuits that can induce bacterial cell differentiation in response to stimuli, convert continuous input signals into some discrete output, and function as intra-cellular detectors are all possible. While it may be possible to develop such applications using the standard gene regulatory network coupled to the recruitment promoter tool kits, it will likely take up a larger space of sequence, and be composed of many more components. Finally, one can imagine constructing complex multi-enhancer synthetic circuits, adding another level of complexity, which can push us closer to the dream of building biological integrated circuits. Therefore, coupling a library of synthetic enhancers with characterized transfer functions to known circuit architectures can lead to a great advance in biological circuit capabilities.

1.4 Synthetic Enhancers as a Basic Research Tool for a Biological Rosetta Stone Algorithm

In order to reach this goal and to be able to engineer routinely gene circuits with synthetic or natural bacterial enhancers as regulatory code, we must first decipher the regulatory code encoded within natural enhancers so that they can provide a credible starting point. However, due to their modular architecture and large binding site content, enhancers are notoriously difficult to dissect, often requiring large and labor-intensive collaborative efforts. To understand the scope of the problem, consider the following example in eukaryotes: the regulatory region (Davidson 2006) of the gene *endo16* in the sea urchin *S. purpuratus.* This is a "run-of-the-mill" gene that participates in the endo-mesoderm formation in early sea-urchin development. It has a regulatory region that spans ~2.3 kbp, with purportedly seven cis-regulatory modules that play a role in defining the time and place of *endo16* expression. Of those modules only two modules have been quantitatively characterized using a "knock down and rescue" type of approach, which necessitated many years' worth of man-work. While the *endo16* analysis and similar works (Atkinson et al. 2002; Davidson 2006; Driever and Nusslein-Volhard 1989; Driever et al. 1989; Small

et al. 1992) have led to provocative data that spawned a vibrant research field, the labor-intensive nature of the research has generated slow progress, which resulted in only a handful of enhancers (bacterial or Eukaryotic) that have been quantitatively characterized to this day. Consequently, one of the greatest challenges facing modern day biological research is to develop a high-throughput methodology for the decipherment of the regulatory programs encoded within enhancers.

Interestingly, as a result of the handful of examples dissected thus far (Atkinson et al. 2002; Davidson 2006; Driever and Nusslein-Volhard 1989; Driever et al. 1989; Ninfa et al. 1987; Small et al. 1992; Yuh et al. 2001) an interesting pathway for a more rapid decipherment of the regulatory programs encoded within enhancers may have emerged. A close examination of the data indicates that there may be a regulatory code characterized by "grammar" (Datta and Small 2011) or design rules encoded into both metazoan and bacterial enhancers. These grammar rules, once deciphered, can in principle allow us to predict the regulatory output of an enhancer based on sequence information alone. If there is a regulatory code encoded into enhancer sequences, what is the best strategy to go about developing a decoding algorithm? One possible method is to work in "reverse": namely, try encoding "words" or "sentences" and testing the decoding algorithm to see if its output recovered the original information. For a biological application, this approach implies developing a synthetic biology strategy for the decipherment of the regulatory output encoded into enhancers. In effect, to engineer using synthetic biology a Biological Rosetta stone algorithm for the regulatory code.

Unlike the archaeological Rosetta Stone, which contained panels of identical messages written in three different scripts and two languages, the biological Rosetta Stone is still missing two of the three panels (Fig. 1.3). In order to develop a draft for this algorithm, we need to construct the two additional panels to complement the sequence panel that we want to decipher. One possible way to do this is by engineering a "collection" or library of simple synthetic enhancers from the ground up, which will be coupled to a high-throughput analysis platform. Results from this analysis can then be used as "training" tool for candidate Rosetta Stone algorithms. Unlike the traditional approach of "knock-out and rescue", building enhancers from the ground up allows one to systematically increase the complexity of the enhancer and enhancer circuit designs in a controlled fashion, which, in turn, provides the opportunity to reconstruct regulatory behavior revealed by quantitative analysis of natural enhancers in an insulated fashion. Therefore, the synthetic approach allows one to dissect quantitatively a multitude of enhancers in substantially less time and manpower.

While it may be difficult to develop such a strategy in metazoans, bacteria are perfectly suited for an initial development of this approach. Recently, we took the first step towards this goal (Amit 2012; Amit et al. 2011; Amit and Phillips 2012) by constructing a "rough sketch" of a bacterial Rosetta Stone algorithm (Fig. 1.3), which in turn had enabled us to formulate qualitative predictions for the expression level outputs of heretofore unexplored bacterial enhancers based purely on sequence analysis. If this strategy proves to be successful in bacteria, a similar strategy for Eukaryotic enhancers can be developed as the next step, while simultaneously

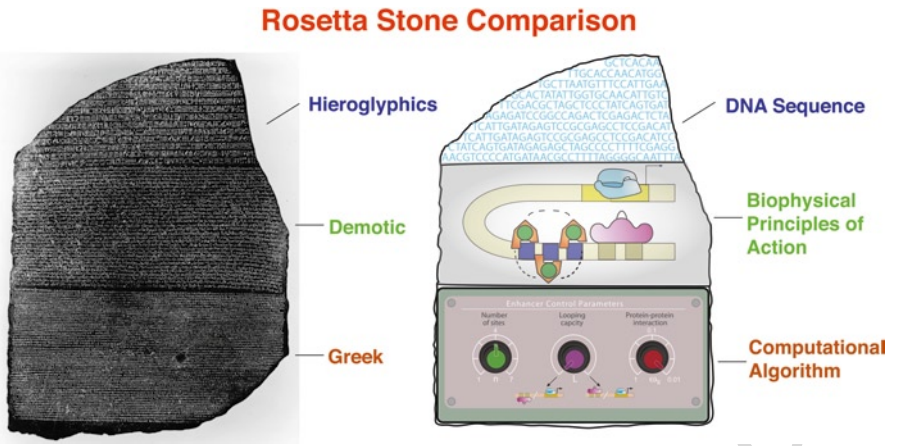

Fig. 1.3 The Biological Rosetta Stone. The Rosetta stone is an archeological artifact that contains three identically written segments in three different scripts and two languages. The Greek (*bottom*) and Demotic (*middle*) segments allowed researchers to interpret the Hieroglyphics script (*top*), which in turn provided archaeologists with a decoding "algorithm" that allowed them to read many previously undecipherable ancient Egyptian texts. The Biological Rosetta Stone strategy's is based on producing the regulatory code equivalent of the real Rosetta stone, where the *top panel* in this case is the DNA sequence. The *middle panel* is biophysical principles or "machine-code" deciphered via the synthetic enhancer experiments. The *bottom panel* is the computational algorithm executed by the enhancer, which is encoded within the sequence depicted by the *top panel*. Such a tool can then be used as a decoding algorithm to predict regulatory output from the sequence of naturally occurring enhancers

allowing us to progress in implementing this technology in bacterial applications. Since current methodologies for the decipherment of the regulatory code are dependent on the arduous and labor-intensive "knock-down and rescue" approaches, I am certain that complementing the standard dissection or reductionist approach with this synthetic methodology will substantially accelerate our ability to decipher the regulatory code, and as such will impact this field to a large extent. Consequently, the ability to construct synthetic enhancer gene regulated circuits in microbial organisms has the potential to not only spawn a quantum leap for a whole host of synthetic biology applications in therapeutics, environmental challenges, biofuel production, etc., but also to bring us a step closer to deciphering the significantly more complex Eukaryotic regulatory code. Successful implementation of such modalities will bring the field closer to fulfilling its great technological potential that had so far proven to be somewhat elusive.

1.5 Conclusions

Enhancers are a class of ubiquitous regulatory objects that potentially can alter the way by which we construct gene circuits. They are capable of executing complex molecular computational operations via a promiscuous architecture capable of integrating multiple

binding sites for several transcription factors. This compact architecture can potentially be used to engineer gene circuits capable of executing complex computational operations that are currently untenable with standard approaches. This ability to integrate multiple signals can be used a fine tuned spatio-temporal control of gene expression, a capability which may be crucial for most future synthetic biology applications from biofuels to smart drug designs.

Even though enhancers are more commonly associated with Eukaryotic regulation, their prevalence and utilization in similar biological function in bacteria points to a largely untapped potential in utilization for synthetic biology applications. However, at present the difficulty in using these components is not rooted in our ability to produce large sequences of DNA, but rather in our ignorance as to the basic operating principles that underlie many of the regulatory effects that are generated by these modules. Hence, before progressing to actively constructing circuits from these objects, we must first develop a better understanding of the basic design rules that guide enhancer regulatory function. One such approach is to develop a biological Rosetta stone via synthetic enhancers to try to distill the rules, which can then be applied on natural bacterial enhancers to test the level of our new understanding. Once we have this tool in place, the engineering of gene circuits with enhancers as basic regulatory modules can finally tap the nearly unlimited computational potential provided by these modules.

References

Ackers GK, Johnson AD, Shea MA (1982) Quantitative model for gene regulation by lambda phage repressor. Proc Natl Acad Sci U S A 79:1129–1133

Aida M, Chen Y, Nakajima K, Yamaguchi Y, Wada T, Handa H (2006) Transcriptional pausing caused by NELF plays a dual role in regulating immediate-early expression of the junB gene. Mol Cell Biol 24:2444–2454

Amit R (2012) TetR anti-cooperative and cooperative interactions in synthetic enhancers. J Comp Biol 19:115–125

[AU2] Amit R, Phillips R (2012) Design rules for synthetic enhancers. Mol Syst Biol in review

[AU3] Amit R, Gileadi O, Stavans J (2004) Direct observation of RuvAB-catalyzed branch migration of single Holliday junctions. Proc Natl Acad Sci U S A 101:11605–11610

Amit R, Garcia HG, Phillips R, Fraser SE (2011) Building enhancers from the ground up: a synthetic biology approach. Cell 146:105–118

Anderson JC, Voigt CA, Arkin AP (2007) Environmental signal integration by a modular AND gate. Mol Syst Biol 3:133

Andrianantoandro E, Basu S, Karig DK, Weiss R (2006) Synthetic biology: new engineering rules for an emerging discipline. Mol Syst Biol 2:1–14

Atkinson MR, Pattaramanon N, Ninfa AJ (2002) Governor of the glnAp2 promoter of *Escherichia coli*. Mol Microbiol 46:1247–1257

Atkinson MR, Savageau MA, Myers JT, Ninfa AJ (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. Cell 113:597–607

Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R (2005) A synthetic multicellular system for programmed pattern formation. Nature 434:1130–1134

Belyaeva TA, Rhodius VA, Webster CL, Busby SJ (1998) Transcription activation at promoters carrying tandem DNA sites for the *Escherichia coli* cyclic AMP receptor protein: organisation of the RNA polymerase alpha subunits. J Mol Biol 277:789–804

Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Kuhlman T, Phillips R (2005a) Transcriptional regulation by the numbers: applications. Curr Opin Genet Dev 15:125–135

Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Phillips R (2005b) Transcriptional regulation by the numbers: models. Curr Opin Genet Dev 15:116–124

Boehm AK, Saunders A, Werner J, Lis JT (2003) Transcription factor and polymerase recruitment, modification, movement on dhsp70 in vivo in the minutes following heat shock. Mol Cell Biol 23:7628–7637

Bolouri H, Davidson EH (2002) Modeling DNA sequence-based cis-regulatory gene networks. Dev Biol 246:2–13

Buchler NE, Gerland U, Hwa T (2003) On schemes of combinatorial transcription logic. Proc Natl Acad Sci U S A 100:5136–5141

Buck M, Gallegos MT, Studholme DJ, Guo Y, Gralla JD (2000) The bacterial enhancer-dependent sigma(54) (sigma(N)) transcription factor. J Bacteriol 182:4129–4136

Bulger M, Groudine M (2011) Functional and mechanistic diversity of distal transcription enhancers. Cell 144:327–339

Cheng B, Price DH (2007) Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. J Biol Chem 282:21901–21912

Datta RR, Small S (2011) Gene regulation: piecing together the puzzle of enhancer evolution. Curr Biol 21:R542–R543

Davidson EH (2001) Genomic regulatory systems: development and evolution. Academic, New York

Davidson EH (2006) The regulatory genome. Elsevier, Burlington

Driever W, Nusslein-Volhard C (1989) The bicoid protein is a positive regulator of hunchback transcription in the early drosophila embryo. Nature 337:138–143

Driever W, Thoma G, Nusslein-Volhard C (1989) Determination of spatial domains of zygotic gene expression in the drosophila embryo by the affinity of binding sites for the bicoid morphogen. Nature 340:363–367

Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. Nature 403:335–338

Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science 297:1183–1186

Friedland AE, Lu TK, Wang X, Shi D, Church G, Collins JJ (2009) Synthetic gene networks that count. Science 324:1199–1202

Friedman N, Cai L, Xie XS (2006) Linking stochastic dynamics to population distribution: an analytical framework of gene expression. Phys Rev Lett 97:168302

Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. Nature 403:339–342

Gilmour DS (2009) Promoter proximal pausing on genes in metazoans. Chromosoma 118:1–10

Golding I, Paulsson J, Zawilski SM, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. Cell 123:1025–1036

Goldman BS, Nierman WC, Kaiser D, Slater SC, Durkin AS, Eisen JA, Ronning CM, Barbazuk WB, Blanchard M, Field C, Halling C, Hinkle G, Iartchuk O, Kim HS, Mackenzie C, Madupu R, Miller N, Shvartsbeyn A, Sullivan SA, Vaudin M, Wiegand R, Kaplan HB (2006) Evolution of sensory complexity recorded in a myxobacterial genome. Proc Natl Acad Sci USA 103:15200–15205

Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA (2007) A chromatin landmark and transcription initiation at most promoters in human cells. Cell 130:77–88

Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, Elston TC, Collins JJ (2006) A bottom-up approach to gene regulation. Nature 439:856–860

Huo YX, Tian ZX, Rappas M, Wen J, Chen YC, You CH, Zhang X, Buck M, Wang YP, Kolb A (2006) Protein-induced DNA bending clarifies the architectural organization of the sigma54-dependent glnAp2 promoter. Mol Microbiol 59:168–180

Jelsbak L, Givskov M, Kaiser D (2005) Enhancer-binding proteins with a forkhead- associated domain and the σ54 regulon in myxococcus Xanthus fruiting body development. Proc Natl Acad Sci U S A 102:3010–3015

Joung JK, Le LU, Hochschild A (1993) Synergistic activation of transcription by *Escherichia coli* cAMP receptor protein. Proc Natl Acad Sci U S A 90:3083–3087
Joung JK, Koepp DM, Hochschild A (1994) Synergistic activation of transcription by bacteriophage lambda cI protein and *E. coli* cAMP receptor protein. Science 265:1863–1866
Kiupakis AK, Reitzer LJ (2002) ArgR-independent induction and ArgR-dependent superinduction of the astCADBE operon in *Escherichia coli*. J Bacteriol 184:2940–2950
Lee DH, Schleif RF (1989) In vivo DNA loops in araCBAD: size limits and helical repeat. Proc Natl Acad Sci U S A 86:476–480
Magasanik B (1993) The regulation of nitrogen utilization in enteric bacteria. J Cell Biochem 51:34–40
Muller J, Oehler S, Muller-Hill B (1996) Repression of lac promoter as a function of distance, phase and quality of an auxiliary lac operator. J Mol Biol 257:21–29
Muse GW, Gilchrist DA, Nechaev S, Shah R, Parker JS, Grissom SF, Zeitlinger J, Adelman K (2007) RNA polymerase is poised for activation across the genome. Nat Genet 39:1507–1511
Ninfa AJ, Peng J (2005) PII signal transduction proteins:sensors of α-ketoglutarate that regulate nitrogen metabolism. Curr Opin Microbiol 8:168–173
Ninfa AJ, Reitzer LJ, Magasanik B (1987) Initiation of transcription at the bacterial glnAp2 promoter by purified *E. coli* components is facilitated by enhancers. Cell 50:1039–1046
Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. Nat Genet 31:69–73
Paulsson J (2004) Summing up the noise in gene networks. Nature 427:415–418
Pedraza JM, Paulsson J (2008) Effects of moelcular memory and bursting on fluctuations in gene expression. Science 319:339–343
Rappas M, Bose D, Zhang X (2007) Bacterial enhancer-binding proteins: unlocking sigma-54 dependent gene transcription. Curr Opin Struct Biol 17:110–116
Raser JM, O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. Science 304:1811–1814
Rasmussen EB, Lis JT (1993) In vivo transcriptional pausing and cap formation on three drosophila heat shock genes. Proc Natl Acad Sci U S A 90:7923–7927
Renner DB, Yamaguchi Y, Wada T, Handa H, Price DH (2001) A highly purified RNA polymerase II elongation control system. J Biol Chem 276:42601–42609
Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. Science 307:1962–1965
Sanchez A, Garcia HG, Jones D, Phillips R, Kondev J (2011) Effect of promoter architecture on the cell-to-cell variability in gene expression. PLOS Comput Biol 7:e1001100
Schulz A, Langowski J, Rippe K (2000) The effect of the DNA conformation on the rate of NtrC activated transcription of *Escherichia coli* RNA polymerase sigma(54) holoenzyme. J Mol Biol 300:709–725
Small S, Blair A, Levine M (1992) Regulation of even-skipped stripe 2 in the drosophilla embryo. EMBO J 11:4047–4057
Stathopoulos A, Levine M (2005) Genomic regulatory networks and animal development. Dev Cell 9:449–462
Su W, Porter S, Kustu S, Echols H (1990) DNA-looping and enhacer activity: association between DNA-bound NtrC activator and RNA polymerase at the bacterial glnA promoter. Proc Natl Acad Sci U S A 87:5504–5508
Tabor JJ, Salis HM, Simpson ZB, Chevalier AA, Levskaya A, Marcotte EM, Voigt CA, Ellington AD (2009) A synthetic genetic edge detection program. Cell 137:1272–1281
Tasmir A, Tabor JJ, Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. Nature 469:212–215
Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. Proc Natl Acad Sci U S A 98:8614–8619
Xu R, Hoover TR (2001) Transcriptional regulation at a distance in bacteria. Curr Opin Microbiol 4:138–144
Yuh CH, Bolouri H, Davidson EH (2001) Cis-regulatory logic in the endo16 gene: switching from specification to a differentiation mode of control. Development 128:617–629

618 Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M, Young RA (2007)
619 RNA polymerase stalling at developmental control genes in the drosophila melanogaster
620 embryo. Nat Genet 39:1512–1516
621 Zhang X, Chaney M, Wigneshweraraj SR, Schumacher J, Bordes P, Cannon W, Buck M (2002)
622 Mechanochemical ATPases and transcriptional activation. Mol Microbiol 45:895–893
623 Zhang Z, Klatt A, Gilmour DS, Henderson AJ (2007) Negative elongation factor NELF represses
624 human immunodeficiency virus transcription by pausing the RNA polymerase II complex.
625 J Biol Chem 282:16981–16988