MICROBIOLOGY AND MOLECULAR BIOLOGY REVIEWS, Sept. 2009, p. 481–509 Vol. 73, No. 3
1092-2172/09/$08.00+0 doi:10.1128/MMBR.00037-08

Mechanisms and Evolution of Control Logic in Prokaryotic Transcriptional Regulation

Sacha A. F. T. van Hijum,[1,2]†* Marnix H. Medema,[1]†‡ and Oscar P. Kuipers[1,3]*

Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands[1]; Interfacultary Centre of Functional Genomics, Ernst Moritz Arndt Universität, Greifswald, Germany[2]; and Kluyver Centre for Genomics of Industrial Fermentation, Delft, The Netherlands[3]

INTRODUCTION

Bacteria react to various environmental conditions by employing different modes of regulation, e.g., metabolic, translational, and transcriptional regulation. Their genes are organized into a hierarchical network of interconnected regulons, which is flexibly organized according to the environmental conditions that a cell faces (255). The expression of regulons is controlled by regulatory proteins (transcription factors [TFs]) with their concomitant DNA binding targets, which are known as TF binding sites (TFBSs). In some cases, the presence of cofactors is necessary for TF activity. In the end, the composition of regulons induced by a condition that the cell faces depends on the concentrations of active TFs. At gene promoters, one or more regulatory signals are integrated into one regulatory output. We term the function according to which regulatory output is determined under different conditions as the control logic of a promoter. The control logic is very important not only for the regulatory output of a promoter but also for motif stringency: how well does the TFBS fit the TFBS sequence that is optimal for binding by a given TF? A recent review by Balleza et al. focused mainly on regulatory network inference, regulatory network plasticity, chromosome structure, and how to make dynamical models of regulatory networks (11). Our review focuses on the mechanisms that determine the control logic of promoters, the relationship of motif stringency to regulatory output, and how these mechanisms are grounded in their evolutionary history. We will first briefly discuss the wide variety of basic mechanisms of regulation at bacterial promoters. We will then focus on TF target analyses, in particular on the experimental determination and in silico prediction of TFBSs and their distributions throughout the genome. Finally, the evolutionary dynamics of *cis*-regulatory regions are discussed, with a keen eye

* Corresponding author. Present address for S. A. F. T. van Hijum: NIZO Food Research, P.O. Box 20, 6710 BA Ede, The Netherlands. Phone: 0031-318-659511. Fax: 0031-318-650400. E-mail: sacha.vanhijum@nizo.nl. Mailing address for O. P. Kuipers: Molecular Genetics, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands. Phone: 0031-50-3632093. Fax: 0031-50-3632348. E-mail: o.p.kuipers@rug.nl.

† S.A.F.T.V.H. and M.H.M. contributed equally.

‡ Present address: Department of Microbial Physiology, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Kerklaan 30, 9751 NN Haren, The Netherlands.

on the evolution of regulatory networks and its relationship to TFBS motif fuzziness and stringency.

Specifically, we will address the problem that the functionality of many in silico-predicted TFBSs can often be neither confirmed nor rejected on the basis of the experimental observations described in the literature. It can be very difficult to distinguish between DNA sequences that function as a binding site for TFs (true positives) and those that do not (false positives) on the basis of a DNA motif. Such a motif is produced from an alignment of several annotated or predicted binding sites. For instance, in statistical identifications of TFBSs, the unavoidable use of a cutoff will lead to a tradeoff between false-positive and false-negative results among the sequences close to this cutoff (215). There is a genuine need to be able to distinguish true and false TFBSs within this twilight zone. Part of the problem is that often, only a limited set of true positives outside of this twilight zone is available as input data, while no ideal negative data set exists (293). However, there is also the question of whether in the end one can truly categorize every potential TFBS as being "positive" or "negative" or if one should think about TFBS functionality in a more continuous manner. In order to tackle these matters, a deeper insight into the broad mechanistic and evolutionary frameworks of the regulatory complexity present in promoter sequences is required.

The issue of operons, multiple genes that are transcribed in a single mRNA, being central in prokaryotic gene regulation and the question of which prediction methods to be used for a given organism have been reviewed recently (35) and will not be discussed further. Also, the subject of gene expression being dependent on its presence at the leading or lagging strand during DNA replication has been reviewed extensively (224, 236, 246), as has the role of protein phosphorylation on, e.g., carbohydrate metabolism regulation (70). Other mechanisms of transcriptional regulation, such as attenuation and (anti-)antitermination have been discussed in depth as well (105, 252).

While many related reviews have focused on DNA motif discovery and the computational data integration needed to reconstruct transcriptional regulatory networks (TRNs) (112, 125, 191, 257, 259), the focus here is on the biological regulatory mechanisms that combine in promoters to yield specific gene expression outputs. Central to this review are the terms control logic and motif stringency. In other words, how are signals integrated at the prokaryote promoter, and how do these signals result in a graded regulatory response? We outline that the difference between spurious and functional TFBSs largely depends on a number of factors: (i) their location, (ii) their degeneracy, and (iii) whether the corresponding TF is local or more pleiotropic. Although in a few cases we cite eukaryote research that is relevant to the topic as well, the focus is clearly on prokaryotes. Prokaryotic transcription regulation is highly complex and will leave computational biologists busy for decades to create models of it that approximate its intricate reality.

BASIC MECHANISMS OF REGULATION AT PROKARYOTIC PROMOTERS

Transcription is the process of transcribing DNA into RNA (e.g., mRNA, tRNA, rRNA, and small RNAs) and is performed primarily by RNA polymerase (RNAP). Transcription consists of five phases: (i) preinitiation, (ii) initiation, (iii) promoter clearance, (iv) elongation, and (v) termination. During preinitiation, RNAP binds to the core promoter elements (−10 and −35; positions indicate the location of each sequence with respect to the transcription start site) in the upstream region (*cis*-regulatory region) of a gene on the genome. After RNAP binding, a transcription bubble is created between positions −10 and +2 through a process termed isomerization (36). At the start of initiation, sigma (σ) factors associate with the RNAP and allow it to recognize the −35 and −10 sequences. After the first DNA base is transcribed into mRNA, the process of promoter clearance takes place. During this process, RNAP often slips from the DNA, producing incomplete transcripts (abortive initiation). RNAP no longer slips from the DNA when approximately 23-bp transcripts are formed. The elongation step involves the elongation of the mRNA transcript until transcription termination occurs. The termination of transcription is mediated either by hairpin structures in the DNA (transcriptional terminators; Rho-independent termination) or by binding of the Rho cofactor, which dissociates the mRNA from DNA (53, 123, 220).

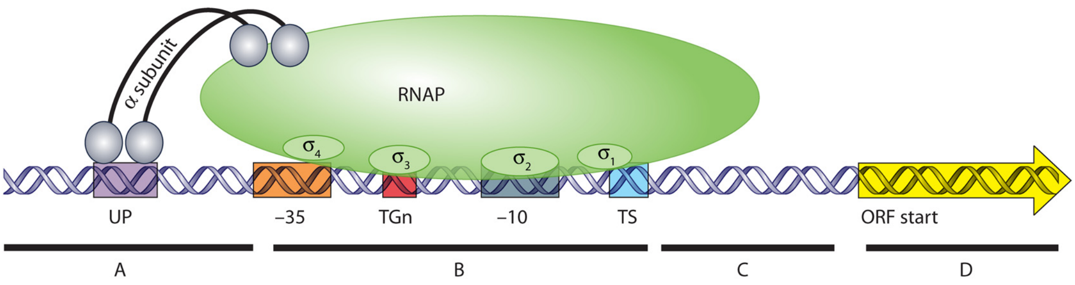
In the next paragraphs, we discuss the cofactors that are involved in RNAP binding, TFBSs, and transcriptional activation and repression.

General Promoter Architecture

Some genes are transcribed highly, while other genes are barely transcribed or even not at all. This is due in large part to the fact that transcriptional regulation takes place mainly at the initial binding of RNAP to the DNA, the isomerization process, and the earliest stages of RNAP progression along the DNA duplex (36). Because the supply of both σ-factors and free RNAP in a cell is limited, there is intense competition between promoters for the binding of the RNA holoenzyme (36, 192a).

The binding of a specific σ-subunit of RNAP plays an important role in transcriptional regulation. The three main functions of σ-factors are (i) to ensure the recognition of core promoter elements, (ii) to position the RNAP at the target promoter, and (iii) to unwind the DNA near the transcription start site (321) (Fig. 1).

One genome may encode many different σ-factors, which, in addition to specific TFs, are used to determine the transcriptional response of a bacterial cell by each one guiding the RNAP to a specific set of target genes (111). In general, bacterial housekeeping σ-factors are similar to the *Escherichia coli* σ^{70} 70-kDa σ-factor (111, 226) and regulate genes that are involved in cellular growth. Several members of the σ^{70} factor family have been described. *E. coli* K-12 has five other σ^{70} family σ-factors besides σ^{70} (231), whereas *Bacillus subtilis* has 17 known variants of σ^{70} (274). Typically, housekeeping σ^{70} σ-factors bind to the −35 and −10 DNA sequence elements in a promoter, which are relatively conserved hexanucleotide sequences with the consensus sequences TTGACA at position −35 and TATAAT at position −10 (36). The intrinsic strength of a core promoter (the level of transcription taking place from it apart from the effects of the binding of additional TFs) is determined largely by the extent to which the core promoter



Type	Mechanism	Action	TF binding			
			upstream (A)	core promoter (B)	downstream (C)	ORF (D)
repression	Steric hindrance	No RNAP binding		+		
repression	Roadblock	No transcription elongation			±	+
repression	DNA looping	No RNAP binding	+		+	±
repression	Activator modulation	Prevents activator binding	±	±		
activation	Class I	Interaction α subunit RNAP	+			
activation	Class II	Facilitates σ factor binding	+			
activation	DNA conformational change	DNA helix twist		+		
activation	Repressor modulation	Prevents repressor binding	±	±	±	±

FIG. 1. Molecular mechanism of transcription modulation. The main features of four repression and four activation types are presented. +, the TF binds at this location; ±, there are multiple places where the TF could bind. TS signifies the transcription start site, TGn signifies the extended −10 element, and UP signifies the UP element. The ORF is the gene regulated by the promoter.

elements match these consensuses (154, 157, 289). Alternative σ-factors (among which are also those of the σ^{54} family) often regulate a set of genes having a clearly defined function, but their regulons may also cover a broader set of target genes involved in diverse biological processes and overlap significantly with those of housekeeping σ-factors (306). A specific subfamily of σ-factors that directly incorporates signals from the extracellular environment in regulating transcription (ECF σ-factors) also exists (121). Excellent reviews of alternative σ-factors that discuss their diverse functionalities in detail are available (111, 121, 151). Diverse σ-factors are often regulated by anti-σ-factors, which inhibit their function under specific conditions (139).

Two other important sites are the extended −10 element and the UP element (Fig. 1). The extended −10 element is located directly upstream of the −10 element and comprises four nucleotides with the consensus sequence TRTG (304, 305), and the approximately 20-bp UP element is located upstream of the −35 element up to −80 nucleotides (84, 205). Such UP elements are easily spotted, as they are AT rich and seem to be particularly associated with strong promoters. The relative contributions of these elements to RNAP binding differ strongly between promoters. A particular combination of these elements could result in RNAP binding a promoter sequence too tightly, which would in turn prevent the RNAP from escaping the promoter. Currently, predictions of bacterial core promoter sequences can be performed using the following methods: position-weight matrix (PWM) scoring (137); comparative genomics approaches (294); classification by, e.g., support-vector machines (107, 295); and a recently developed triad algorithm that incorporates UP element detection (66) (see also Table 2 for an overview of methods that deal with promoter prediction).

In addition to these general methods that a cell uses to regulate gene expression, the cell utilizes specialized TFs that bind to specific DNA recognition sequences (TFBSs). TFBSs for a specific TF can differ in nucleotide sequence and composition, but they can be represented by a consensus DNA sequence motif, i.e., the representation of the target variability of the TF. Below, the different representations of sequence motifs are discussed.

The location and nucleotide composition of TFBSs determine in large part whether a TF represses or activates the expression of a certain gene. The length of bacterial TFBSs is usually between 12 and 30 bp, and they often appear in the form of direct repeats or palindromes, which may facilitate the dimeric binding of TFs (247). As most bacterial TFs have a helix-turn-helix domain and act as homodimers, the motifs of their TFBSs are usually structured as a “dyad” (spaced motif) with a spacing of a given number of uninformative base pairs (301). In some cases where TFBSs exist as direct repeats or palindromes, half-sites (with only one of the repeated segments or half of the palindrome) also have some functionality (168). TFBSs can be located at various positions relative to the canonical −35 and −10 promoter sequences ranging from far upstream to within and downstream of the promoter. Regulatory motifs are usually not strictly specific (as are the DNA motifs cut by restriction enzymes) but are only partially conserved and thus appear rather “fuzzy” (100, 266).

The thermodynamic state of TF proteins can be described using a three-state model (169, 283): (i) freely diffusing in three dimensions as monomers, (ii) unspecifically bound as mono- or oligomers to DNA by general electrostatic interactions and thus diffusing along the DNA backbone in one dimension, and (iii) specifically bound to a binding site at a local energy minimum through hydrogen bonds as well as hydrophobic and electrostatic interactions. Switching between the latter two states involves a conformational change of the TF protein, which is triggered by the molecular recognition of an energy minimum, most often through the binding of a protein α-helix to the major groove of the DNA (52). The combination of these three states enables the TF to find its target sites and bind to them in relatively little time (169). For a few model systems that were studied, the binding energy itself seems to be well approximated by the sum of the independent contributions of a small number of TF binding nucleotides (88, 221). The binding probability depends on the binding energy in a sigmoid way, thus generating a threshold between weak binding and strong binding that is exemplified by an insensitivity of the binding probability if the binding energy is between weak and strong binding (169).

Mechanisms of Transcriptional Repression and Activation

Some TFs function to repress transcription, while others activate transcription. Still others function as either activators or repressors, often according to the positioning of the TFBS relative to the σ-factor binding site in the target promoter (231) (see Fig. 1 for a summary of the main mechanisms). The binding and release of repressors and activators themselves are often controlled by cofactor binding. Cofactors are molecules that can range widely in size and nature, from small ions, nucleotides, covalently attached phosphate moieties, and sugars to peptides or whole proteins (2, 86, 118, 285). Although most activators function by first binding to the promoter DNA before interacting with RNAP, some activators (such as *E. coli* MarA and SoxS) also bind to free RNAP in the cytosol prior to binding their TFBSs (110, 200).

There are four main modes in which TFs have been described to mediate repression (36, 181, 247) (Fig. 1A to C): (i) repression by steric hindrance, often by binding of the repressor between or on the core promoter elements; (ii) repression by blocking of transcription elongation, often by binding at the start of the coding region (roadblock mechanism); (iii) repression by DNA looping, with binding sites often both upstream and downstream of the core promoter (in this case, an interaction between two monomers of the same TF is possible only if both TFBSs are spaced correctly); and (iv) repression by the modulation of an activator. In the latter case, a repressor binds to a TFBS that (partly) overlaps a different TFBS of an activator. The binding of the repressor to its site will then prevent the binding of the activator to its respective TFBS. An example of such an interaction is that between the CytR and CRP (for a review, see reference 36).

Similarly, four modes of activation by TFs have been described (12, 36, 181, 247, 279) (Fig. 1D to F): (i) class I activation, in which the TF binds upstream of the core promoter and interacts with the flexible α-subunit of RNAP; (ii) class II activation, in which the TF binds the DNA directly adjacent (mostly upstream) to the core promoter and promotes σ-factor binding; (iii) activation by DNA conformational change, in which the TF binds to the core promoter to enable it to be bound by a σ-factor, often by twisting the DNA helix; and (iv) activation by the modulation of a repressor, alleviating the repression effect. An example of the latter mode (also termed antirepression) was recently discovered for the *B. subtilis* competence activator ComK, a minor groove binding protein that binds adjacent to the repressors Rok and CodY at its own *comK* promoter (279). Although ComK binding to the DNA does not result in the physical displacement of Rok and CodY, it removes the repression effect and thus activates the expression of the gene (Fig. 2).

Spatial Constraints on Promoter Architecture

Although it seems obvious that spatial constraints on TFBS placement within promoters should exist, relatively few detailed experimental studies have been performed to specify these (187). Most repressor sites are located between positions −60 and +60 relative to the transcriptional start site (55, 83, 192, 212), although repressors often bind to sites much further upstream, as in the case of, e.g., DeoR repression of the *E. coli ula* operon (167). The degree of repression depends significantly on the TFBS position relative to that of the promoter (58). Activator sites are usually present upstream of or next to the −35 core promoter element (247) (Fig. 3 and Table 1). Class I activators are generally bound between positions −60 and −95, while class II activator sites are adjacent to, or overlapping with, the −35 element (12). In a recent study by Cox and coworkers (58), regulatory effects of the activators LuxR, which regulates luminescence genes in *Vibrio fischeri*, and AraC, regulating arabinose metabolism in *E. coli*, were tested in vivo using 288 artificially constructed promoters that were inserted into a plasmid with a luciferase reporter gene. The regulatory effects of activator TFBSs located downstream of the −35 core promoter element appeared to be negligible compared to the effects of upstream sites. This work clearly indicates that control logic can be inferred for a number of regulators involved in metabolism.

Other spatial constraints are formed by the fact that activation or repression often functions only if TFs bind to specific positions on the promoter DNA helix, as TF binding to a TFBS in general has to be present at the same side of the DNA duplex as RNAP binding to fulfill its function. In two independent studies, Ushida and Aiba (298) and Gaston and coworkers (95) showed that the extent to which the well-studied *E. coli* catabolite repression protein (CRP) was able to activate gene expression on *melR* and *lacZ* promoters was dependent largely on the helical face to which it bound, which had to be identical to the face to which RNAP bound (Fig. 4). Therefore, within the region between positions −60 and −95, class I activators were mostly functional only around positions −61, −71, −81, and −91 (12), the intervals which match a single helical turn (10.5 bp) of B-form DNA (Table 1). For the *Lactococcus lactis* MG1363 pleiotropic regulators CodY and CcpA, the helicity of the TFBS compared to the transcription start site was shown to be important for the regulation of target genes as well (68, 334).

In many cases when a TFBS is positioned at a relatively long distance from core promoter elements, this has a specific regulatory function. For example, the fact that in *B. subtilis*, the ComK binding site (K-box) at the promoter of the *comK* gene itself is positioned one or two helical turns further upstream than K-boxes in other promoters provides a threshold for autoactivation. This can be relieved by the adjacent binding of DegU (116, 117). Because DegU binding stimulates *comK*

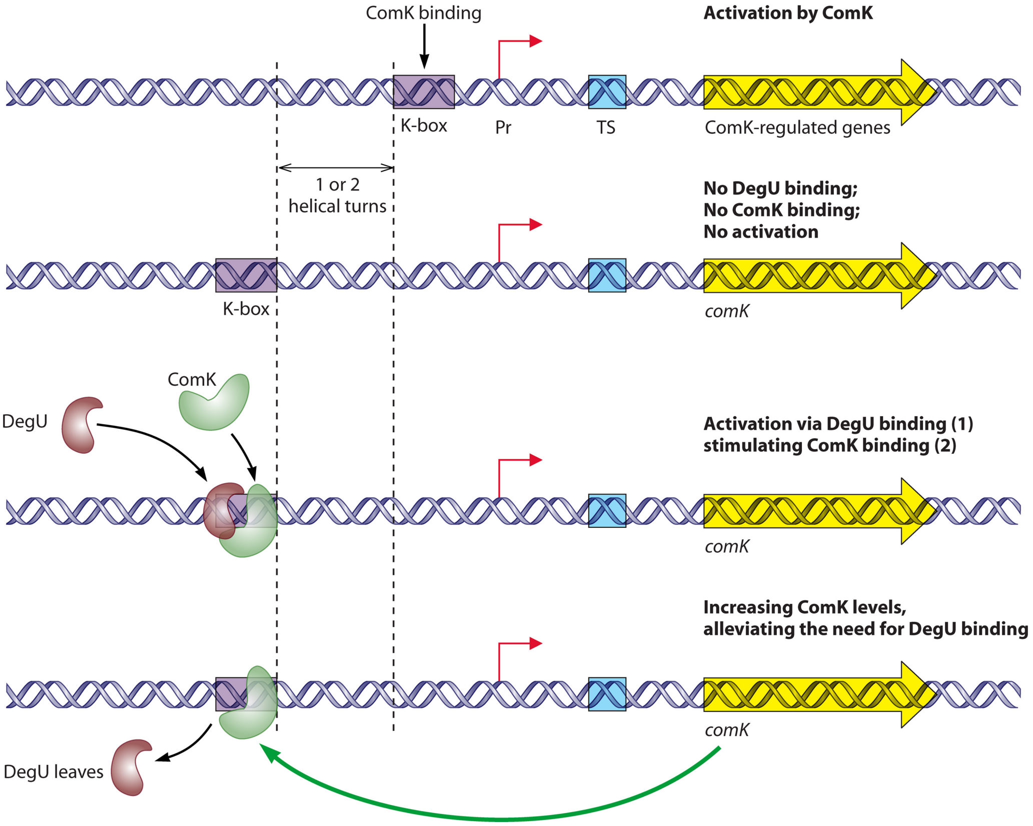


FIG. 2. ComK regulates genes of the *com* regulon by binding to K-boxes upstream of their promoters. The autoregulation of *comK* gene expression is more complex. The K-box is located either one or two helical turns further upstream relative to the locations of K-boxes in the *com* regulon. To activate this promoter, it is required that another regulator, DegU, binds first, thus recruiting ComK to bind to the K-box. Once ComK levels rise sufficiently, ComK can autoactivate its transcription without the need for DegU binding. TS signifies the transcription start site, and Pr signifies the core promoter.

transcription, the concentration of active ComK rises, and ComK can then activate the transcription of the *comK* gene without the additional help of DegU. DegU thus functions as a priming protein that can turn on an autostimulatory feedback loop (Fig. 2).

Notably, the threshold to the feedback loop has a very significant biological function, as it enables phenotypic heterogeneity of competence in *B. subtilis* populations (278, 280).

Cooperative Regulation Mechanisms and Promoter Control Logic

Based on RegulonDB, version 6.3 (94), a large percentage (about 65%) of the transcriptional units (operons or single genes) of *E. coli* K-12 that are annotated to be regulated by at least one TF are regulated by more than one TFBS for a given TF. Also, genes are often regulated by more than one different

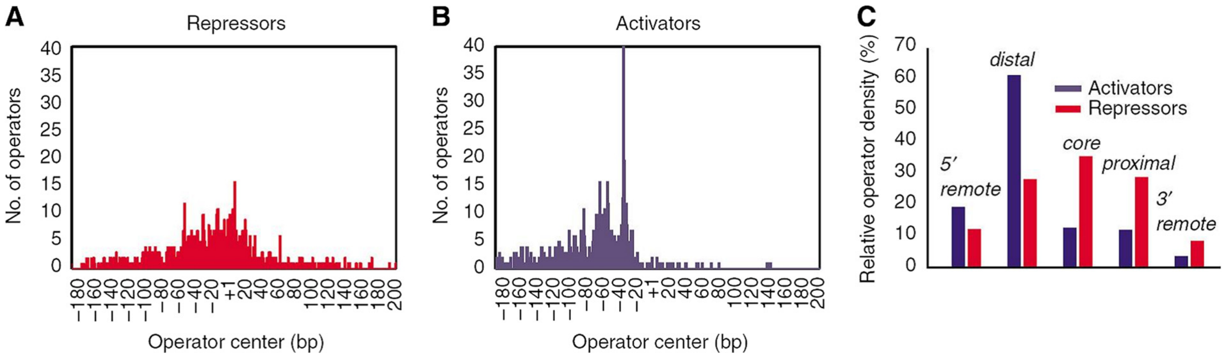


FIG. 3. Distribution of TFBS locations for activators and repressors in the RegulonDB database (94) as found in 1,102 *E. coli* promoters. (A and B) The distribution of TFBSs is shown relative to the transcription start site (+1). (C) The density of TFBSs in 554 *E. coli* σ^{70} promoters is depicted, divided into five regions: the 45-bp region upstream of the −35 box (distal), the 25-bp region between the −10 and −35 boxes (core), the 30-bp region downstream of the −10 box (proximal), and the remote 5′ and 3′ regions. (Reproduced from reference 58, which was published under a Creative Commons license.)

TABLE 1. Overview of transcriptional regulatory mechanisms in prokaryotes[a]

Regulatory mechanism	Regulation specifically within or around promoter DNA	Main positioning relative to transcription start	Sequence-specific mechanisms
Class I activation	+	−95 to −60	+
Class II activation	+	−50 to −35	+
Activation by DNA conformational change	+	−35 to −10	+
Activation by repressor modulation	+	−60 to +60	+
Repression by steric hindrance	+	−35 to −10	+
Repression by roadblock	+	−10 to +60	+
Repression by DNA looping	+	−60 to +60	+
Repression by activator modulation	+	−95 to −10	+
Cooperative activation	+	−95 to −35	+
Cooperative repression	+	−60 to +60	+
Promoter escape regulation	+	−10 to +10 −10 and −35 elements	+
DNA methylation	+	−200 to 0	+
Riboswitches	−	5′ UTR and 3′ UTR	+
Transcriptional interference	−	−	−
Chromosome polarization	−	−	−
DNA supercoiling	−	−	±
mRNA degradation	−	−	±

[a] See also Fig. 1 for more details concerning the eight types of transcription modulation. A + signifies present or applies to, a − signifies does not apply or not present, and a ± signifies present in specific cases. UTR, untranslated region.

TF (31% of the total genes for *E. coli* K-12). For example, activators and repressors can antagonize each other at a particular promoter sequence (competitive regulation) (38, 101, 124). However, multiple different activators also frequently work together to induce transcription (cooperative regulation) (36, 38), each regulated by a different cellular or environmental signal (6). This is the case for the *B. subtilis ackA* promoter, the expression of which is governed by cellular levels of both glucose and branched-chain amino acids through activation by CcpA and CodY (210, 271). Sometimes, TFs contribute to activation independently in a combination of class I and class II interactions. In other instances, multiple activators interact with the DNA in a cooperative manner. In yet other cases, one activator functions to counter the function of a repressor while the other one performs the direct activation (27, 28, 36). Generally, two modes of cooperative binding exist (124): (i) homocooperative binding, in which more than one of the same TFs bind cooperatively to multiple instances of the same TFBS in one promoter region, and (ii) heterocooperative binding, in which different TFBSs in the same promoter are cooperatively bound by different TFs. The cooperative or competitive action of multiple TFs can result in complex regulatory events at *cis*-regulatory regions, as in the above-mentioned case of the *comK* promoter (Fig. 2).

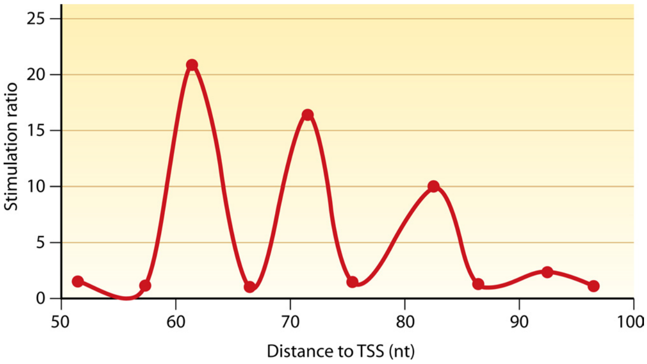
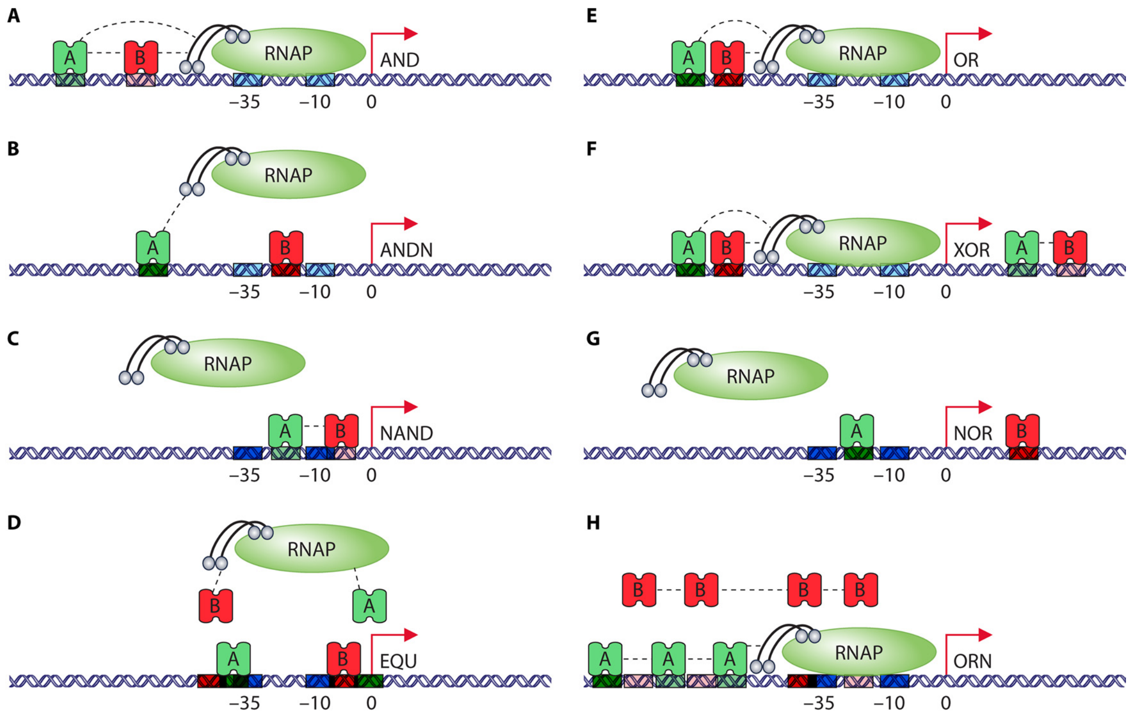


FIG. 4. Effect of the distance between the CRP site and the transcription start site (at 0) on activation by CRP. The stimulation ratio is the activity of the *lacZ* promoter relative to the activity in a *crp* mutant. TSS, transcription start site; nt, nucleotides. (Adapted from reference 298 with permission of Oxford University Press.)

Boolean logic gates such as AND, OR, and NAND (Fig. 5) can be accomplished with prokaryotic promoters by relatively simple combinations of interactions between two TFs and RNAP at a promoter (27, 28, 38, 275). For example, the AND gate, in which transcription occurs only if both of two active TFs are present at high concentrations, can be produced by two different activator TFBSs acting cooperatively. The OR gate, in which transcription occurs when either of two active TFs is present at a high concentration, can be produced by two activator TFBSs functioning independently on the same target. The NAND (not and) gate, in which transcription is repressed only when both of two active TFs is present at high concentrations, can be produced by a strong promoter regulated by two weak repressor TFBSs acting cooperatively (and requiring this cooperation to attain a significant repressive effect) (38). Thermodynamic models reported by Buchler et al. suggested that more complex Boolean logic gates (EQU and XOR) can also be attained (Fig. 5). An XOR (excluded or) gate, in which transcription occurs only when one out of two active TFs acting on a promoter is present at high concentrations, for example, can be accomplished by two different TFs acting independently as activators on two strong-affinity TFBSs while at the same time acting cooperatively as repressors on two weak-affinity TFBSs.

Finally, an EQU (equals) gate, in which transcription occurs only when the active concentrations of two TFs are approximately equal, can be produced by two different TFs acting as repressors on two strong-affinity TFBSs interfering with a strong promoter and at the same time acting as derepressors on each other's sites. Buchler et al. (38) and, later, Bintu et al. (27) also suggested options involving either multiple alterna-



TFs		Transcriptional unit activity							
C_A	C_B	AND	ANDN	NAND	EQU	OR	XOR	NOR	ORN
Low	Low	Off	Off	On	On	Off	Off	On	On
Low	High	Off	Off	On	Off	On	On	Off	Off
High	Low	Off	On	On	Off	On	On	Off	On
High	High	On	Off	Off	On	On	Off	Off	On

FIG. 5. Boolean logic of transcriptional regulation. Boolean logic gates map multiple input signals from two TFs (of concentrations c_A and c_B) to one output signal. The table specifies the status of the transcription of the transcriptional unit ("on" or "off" for each gate); for example, when a promoter functions as an ORN (or not) gate, transcription occurs when c_A is high or c_B is not high. It should be noted that because promoter outputs are generally not a binary function of regulator concentrations, a wide variety of non-Boolean logical phenotypes occur in nature. (A to F) Possible promoter configurations to attain each Boolean output. A and B are TFs. Green/red boxes signify TFBSs, and blue boxes signify core promoter elements. Darkly colored TFBSs/promoter elements have strong binding affinity for the TF/RNAP σ-factor, and lightly colored ones have weak binding affinity. Dashed lines signify cooperative interactions. (A) AND gate, in which A and B both function as class I/class II activators in a cooperative fashion. (B) ANDN gate, in which A functions as an activator and B functions as a repressor. (C) NAND gate, in which A and B cooperatively function as repressors. (D) EQU gate, in which A and B function both as repressors and derepressors on separate TFBSs. (E) OR gate, in which A and B separately function as activators. (F) XOR gate, in which A and B function both as separate activators and cooperative repressors. (G) NOR gate, in which A and B function as separate repressors. (H) ORN gate, in which an module of A activators competes for binding to the DNA with a module of B repressors. (Panels A, C, D, E, and F are based on data from reference 38; panels B and G are based on data from reference 275; panel H is based on data from reference 124.)

tive core promoters acting on a gene or repression by DNA looping similar to the mechanisms that have been described for the *E. coli lac* operon (160).

Alternatively, Hermsen et al., using similar thermodynamic models, predicted that complex Boolean logic gates (including NOR, ANDN, and ORN) (Fig. 5) can also be accomplished by the cooperative binding of TFs in complex promoters with multiple TFBSs when in the *cis*-regulatory region, two modules, both containing an array of binding sites, overlap and thus compete for cooperative binding (124). The affinity of binding of σ-factors to the core promoter and of TFs to the different TFBSs determines the precise logic function governing the conditions for transcriptional activation or repression. For example, an EQU gate requires a strong core promoter to facilitate transcription when the concentrations of both TFs are low, while two homocooperative repression modules mediate repression only when one of the two TFs is present at a sufficient concentration (124). When both TFs are present in high concentrations, this repression is countered by a heterocooperative activation module containing both TFs; this heterocooperative array of sites must then have a higher cumulative binding affinity than do the overlapping homocooperative repression modules. Problems with the predictions described Hermsen et al. appear to be that the modules which they proposed lead to an overcrowding of TFBSs within promoters that seems quite unrealistic.

The biological relevance of these theoretical studies has still to be investigated, as few experimental efforts have yet focused on identifying complex logic gates regulated by cooperative TF binding. The most extensive experimental work in this respect was done by Kaplan et al., who mapped the control logic of 19 *E. coli* sugar metabolism-related genes, which are regulated by both CRP and a specific sugar regulator, in considerable detail (148). They did this by creating a map of gene expression levels under various concentrations of cyclic AMP (the metabolite determining CRP activity) and the sugar involved in activating the specific sugar regulator. Because the conditions were chosen in such a way that the expression depended almost exclusively on the concentrations of these two input signals, they could interpret the shape of the resulting map to infer the control logic of the promoter (147, 148). Interestingly, those authors found the sugar gene promoters to contain diverse control logics, including quite complex ones such as that of *fucR*, which approximates the XOR gate, by displaying reduced expression levels when both input signals are high and when both are low (148). Another promoter region that is interesting for future study in this respect would be the *E. coli gltBDF* operon, which is involved in one of the two main pathways of ammonia assimilation in this organism. This operon was recently shown to be regulated by multiple global regulatory proteins of *E. coli* (Lrp, IHF, CRP, and ArgR) (228).

Although the complex control logic that underlies cooperative regulation has not yet been described for modeling efforts, elementary control logic represented in stoichiometry matrices was described by Klamt and coworkers, who created a modeling tool, CellNetAnalyzer (155). In the end, an understanding of the different ways in which the different types of control logic can be produced by prokaryotic promoters can both help predict the input-output relationships between factors involved in promoter regulation for purposes of transcriptional network reconstruction (78, 259) and help synthetic biology efforts in the engineering of artificial biological circuits (275).

The Wide Variety of Additional Regulation Mechanisms

Although it is not our goal to give in-depth descriptions of all other transcriptional regulatory mechanisms, it is worthwhile to give a short overview of additional regulation mechanisms that add to the complexity of transcriptional regulation. Therefore, we will shortly touch on the regulatory mechanisms of promoter escape regulation, transcriptional interference, DNA methylation, chromosome supercoiling, histones, as well as the posttranscriptional regulation mechanisms of mRNA degradation, riboswitches, and short noncoding RNAs. For details, we will refer to some excellent reviews that have recently been written on these topics.

A large part of cellular transcriptional regulation takes place at the stage of transcription initiation, in which the bound RNAP has to escape the promoter to advance to downstream regions of the DNA template (132). Besides the possibility of regulation by TFs binding upstream of the core promoter elements, RNAP promoter escape can also be regulated by specific factors which bind to the RNAP itself. Recently, it was shown for one such promoter escape-regulating factor, GreA (129, 133), that it can also be sequence specific. In a microarray study comparing cells expressing either wild-type GreA or a strain carrying an inactivated version of the same factor, Stepanova et al. identified 126 genes that were specifically transcribed in the presence of wild-type GreA (282). The mechanism by which this specificity is mediated is not yet clear.

Another way in which transcription elongation can be regulated for both σ-factors and TFs is when different transcriptional activities interfere with one another in *cis* by the collision of RNAPs bound to, or initiated from, different promoters (268). This process is called transcriptional interference and can occur in convergent promoters, tandem promoters, and overlapping promoters. Convergent promoters are promoters producing converging transcripts, the 5′ regions of which overlap at least partially; tandem promoters are promoters in which one promoter is placed upstream of the other but transcribing in the same direction, and in overlapping promoters, the RNAP binding sites are at least partially overlapping. Transcriptional interference could very well be a widespread mechanism of gene regulation. An analysis of the 4,462 *E. coli* promoters in the RegulonDB database revealed 166 tandem promoters, 54 convergent promoters, and 435 promoters that are probably overlapping (268).

Modifications to the structure of the DNA itself can also function to regulate transcription. DNA methylation is such an epigenetic regulation mechanism (48, 184, 243). The best-studied bacterial DNA methyltransferases are the *Caulobacter crescentus* CcrM methyltransferase, which methylates the N^6-adenine of GANTC (243), and the *E. coli* Dam methyltransferase, which methylates the N^6-adenine of GATC sequences (182). Methylated GATC sequences within *cis*-regulatory regions can increase, decrease, or have no effect on transcription initiation efficiency (182). Also, Dam methyltransferases regulate gene expression through the formation of DNA methylation patterns (184), which appear because regulatory proteins compete with Dam for binding to the DNA at Dam sites and prevent their methylation. DNA methylation patterns can both repress and activate gene expression by either enhancing or blocking the binding of either repressors or activators at promoters (184).

Regulation at the level of DNA structure can also take place at the level of overall chromosome organization. Such regulation provides a more global control of transcription than the control of regulators that are specifically dedicated to a relatively small set of gene promoters (199). Crucial in regulating bacterial chromosomal organization are the histone-like nucleoid proteins HU, Fis, H-NS, StpA, IHF, and Dps (208, 288). Besides their global role in regulating supercoiling and chromatin dynamics, at least some of them may also act on a local level in a gene-specific fashion. Note that Fis and IHF were also shown to bind to specific DNA recognition sequences (208) and can have different regulatory effects (activation or repression) when bound to different sites within the same promoter (37). Also, higher-level macrodomains that are related to the transcriptional response to supercoiling and correspond to the distribution of binding sites for DNA gyrase, a topoisomerase involved in creating negative supercoils, exist on bacterial chromosomes (296). However, the domains of higher levels of transcriptional activity are not caused by superhelicity only; replication polarization of the chromosome also plays a major role (1, 183).

Although outside the realm of transcriptional regulation, the regulation at the posttranscriptional level should not be neglected. Riboswitches are regulatory domains that reside in the noncoding regions of mRNAs, where they bind metabolites and control gene expression (14, 195, 308, 319). That mRNA stability can be of high importance in regulating transcript abundance is perfectly illustrated by a study by Selinger et al., who measured mRNA half-lives for 1,036 open reading frames (ORFs) in *E. coli*, which appeared to range between 1 and 2,084 min, with the majority of half-lives between 2 and 20 min, while degradation speeds differed according to the lengths of the polycistronic transcripts (264). Finally, it was also discovered that short noncoding RNAs, first thought to be important for gene expression regulation in eukaryotes only, are also prevalent in prokaryotes and function, for example, by binding specifically to certain mRNAs to repress their translation (108, 109, 203, 244).

TRANSCRIPTION FACTOR TARGET ANALYSIS

Determining target genes of transcriptional regulators is a field that has evolved quite rapidly in the past years. The reasons for this are emerging high-throughput methodologies for transcriptome analysis such as DNA microarrays (75) and mRNA sequencing (317), which allow the monitoring of thousands of transcripts simultaneously, and chromatin immunoprecipitation (ChIP) approaches, with which dozens of TF-DNA interactions can be discovered (233). Current work is in most cases focused on the association of targets (together forming a “regulon”) with their transcriptional regulator. This is done, for instance, by determining a regulon from DNA microarray targets querying a knockout of a transcriptional regulator. TRNs are reconstructed from DNA microarray data and literature data as primary data sources. Other approaches have integrated ChIP-on-chip protein-DNA interaction data, protein-protein interaction data, proteomics, metabolomics, and pathway information (162, 328, 332). Here, we discuss the techniques that are focused primarily on regulon reconstruction and how control logic is used in current approaches.

Experimental Regulon Identification

Regulons are usually identified using transcriptome comparisons between wild-type and TF knockout strains grown under one or a few conditions (329). More recently, time-series transcriptome analysis has also been performed for this purpose, e.g., the time-resolved determination of the CcpA regulons of *B. subtilis* and *L. lactis* (188, 334). From such experiments, groups of genes or operons that respond to specific environmental perturbations can be identified, which are referred to as stimulons (247). To define such stimulons, the level of gene expression of an unperturbed control is compared to that under a condition that stimulates a certain cellular response using DNA microarrays. If the mRNA is isolated under a specific condition, such experiments provide snapshot information of the regulatory role of TFs under those specific conditions (329).

In order to detect associations based on microarray data, coexpression or reciprocal expression between a TF and its target is required. The prerequisite of (anti-)correlated expression patterns is that there is an autoregulatory loop for the TF; i.e., the TF regulates its own expression. These autoregulatory loops are an important basic regulatory mechanism, especially for the negative regulatory loop, where the cell ensures that the expression of a given TF is downregulated after the TF has been produced. For *E. coli* K-12, about 50% of the TFs have negative autoregulatory loops (248). Therefore, one can conclude that for at least 50% of the TFs, a clear (anti-)correlation in expression patterns cannot be expected if other factors such as detection limits of the experimental technique are also taken into account.

In any case, additional experiments are required to distinguish between direct and indirect regulatory effects when the results of such experiments are analyzed. Some clustering algorithms that can quite effectively extract gene expression modules from perturbation data have been developed, such as the ENIGMA tool developed by Maere et al. (193). An advantage of the ENIGMA tool over to most earlier biclustering tools is that it can deal with partial coexpression between genes; i.e., genes show correlated expression only under a subset of conditions.

In order to globally identify the genomic regions that are occupied by a DNA binding TF, ChIP experiments are also used. In ChIP experiments, the chromosomal DNA is cross-linked to a tagged regulator protein, sonicated to produce small fragments, and then immunoprecipitated with an antibody against a given TF or its tag (234). In ChIP-on-chip, this enrichment of DNA binding to a certain TF is then compared to that of a control containing nonenriched chromosomal DNA with microarray analysis to reveal the binding sites of that TF on the genome (233, 329). Recently, a novel method called ChIP-Seq was also developed, in which ChIP is coupled to next-generation massively parallel sequencing technology (145, 198). Typically, only short 25- to 50-nucleotide reads (“tags”) are sequenced, and genomic regions with probable binding sites are identified by the high densities of such tags in the output (146, 299). The regulatory motif that characterizes a set of TFBSs can be detected using both gene expression (DNA microarray and mRNA sequencing) and ChIP-Seq or ChIP-on-chip data (often from genome tiling microarrays). For these methods, either (i) the *cis*-regulatory regions of genes with large differences in transcription rates between the respective TF knockout and its wild type in a microarray experiment are pooled or (ii) the *cis*-regulatory regions are precipitated with a certain TF, employing computational methods to identify overrepresented oligonucleotides in these sequences (112, 191). Tiling microarrays can also be used to obtain more precise information concerning the start of transcription, also referred to as promoter mapping (56, 242).

Other methods focus directly on identifying the DNA binding specificity of a TF and can be used to reconstruct regulons by using the resulting regulatory motifs to predict the binding site of a given TF computationally. One such methods is coined systematic evolution of ligands by exponential enrichment (SELEX). In SELEX, one starts with a random pool of oligonucleotides, after which strongly bound oligonucleotides are enriched by multiple cycles of target binding, selection, and DNA amplification (73). Although the standard SELEX method can easily be used to find the optimal consensus sequence of a TFBS motif, it fails in practice to provide a good data set for reconstructing a high-resolution motif of its DNA binding specificity because the oligonucleotide pool is too enriched for the most strongly bound sites (177). Fortunately, modifications to the protocol make it possible to obtain the needed amount of low- and medium-affinity sequences (177, 250). An even more promising approach that was recently developed is formed by protein binding microarrays, in which TF fusion proteins are bound on double-stranded DNA microarrays containing many different DNA sequence variants of a given length (23, 24, 39). Binding of the TF to spots can then be detected with fluorescently labeled antibodies against the protein to which it is fused. The method can be used in an impressively high-throughput manner to determine the DNA binding specificities of many TFs (331).

More traditional methods also reveal information on the presence of TFBSs in promoter DNA sequences. An example is DNase I footprinting. In this technique, a DNA fragment is allowed to interact with a DNA binding protein, after which the complex is partially digested with DNase I (93). The bound protein protects the region of the DNA to which it binds from DNase digestion. Subsequent electrophoresis identifies the region of protection as a gap in the background of digestion products (173). Another traditional method is the electrophoretic mobility shift assay, in which a protein-DNA mixture is separated on a gel and compared to a DNA-only control. One can then see if the protein binds the DNA: in this case, the DNA band from the protein-DNA mixture will be less mobile than that of the DNA-only control (69).

TFBS Motif Representation

A major issue in TFBS motif discovery is the way in which motifs are represented (Fig. 6). Many different representations exist, and the choice is often determined by the level of accuracy, simplicity, interpretability, representational power, or computational convenience (191) (see Table 2 for an overview of methods involved in TFBS discovery and visualization). Probably the simplest way of motif representation is the use of a consensus sequence of preferred nucleotides (A, C, G, and T). Either such a consensus sequence can be represented in a strict manner, in which case it represents only the optimal sequence, or degeneracy can be built in (e.g., R is purine, Y is pyrimidine, S is strong, W is weak, K is keto, M is amino, and N is any nucleotide, according to IUPAC nomenclature [http://www.chem.qmul.ac.uk/iubmb/misc/naseq.html]) (63), in which case a limited amount of information can be represented on the proportions of nucleotides at the given positions. PWMs are currently the most common model for identifying TFBS motifs and are more precise than consensus-based representations (284, 293). In PWMs, the nucleotide observed at a position is assumed to be independent of the nucleotides at other positions. Motifs are visualized conveniently by sequence logos consisting of an ordered stack of letters in which the letter's height indicates the amount of information that the motif contains at that position (59, 262).

As a final critical note, a consensus sequence, PWM, or sequence logo does not necessarily convey all biologically relevant features of a DNA sequence that enables it to be bound by a TF. It is merely the result of determining the overrepresentation of nucleotides in a number of *cis*-regulatory regions of coregulated genes.

In Silico Prediction of TFBSs

The identification of the sequence motifs that constitute the range of sequences functioning as TFBSs for a certain TF in a particular genome remains a challenge in computational biology, and a large array of options have been exploited to predict such motifs in silico (61, 71, 112, 191, 218, 257, 303) (see Table 2 for an overview of methods involved in TFBS discovery and visualization). In general, one starts out with a set of DNA sequences that are a priori believed to be coregulated and therefore likely to be bound by one or more regulatory proteins (329). This list of genes can be determined, e.g., based on candidates that are differentially expressed in a DNA microarray experiment querying a perturbation or by determining coexpression over a compendium of microarray data (see above). Computational algorithms are then used to identify the motifs that could be responsible for this binding of TFs (191). Finally, the motifs that are found to be overrepresented in the DNA sequences of the coregulated genes can be used to search the genome for other additional putative TFBSs that match the motif. Recently, in a study reported by Westholm and coworkers (314a), where a meta-analysis of predicted TFBS distributions across the *Saccharomyces cerevisiae* genome was performed, it was demonstrated that there are significant numbers of TFBS motifs for which (a combination of) location and orientation are important for functionality. They also provided a Web tool (ContextFinder) that will allow researchers to perform such an analysis on a regular basis.

The basic algorithmic approaches that have been used thus far to identify DNA motifs can be grouped into two main categories: enumerative or word-based methods and probabilistic methods (61, 71, 112, 191, 218). Generally speaking, probabilistic methods are more appropriate for finding motifs in prokaryotes, as they are better suited to identifying longer sequence motifs in terms of computational cost (61, 323). However, in contrast to enumerative methods, they do not always find the global optimum in their search space.

Enumerative methods exhaustively catalogue DNA oligonucleotide words, which are then scored by statistical significance on a set of reference sequences to identify the most significantly overrepresented motif strings of a certain length (191). Multiples of these string-based motifs can then be merged into one approximate motif, if necessary. van Helden et al. developed the oligonucleotide analysis motif-finding algorithm based on this approach (300). Later, they adapted their method for the analysis of bipartite dyad motifs with a low-information-content linker region (301), which are characteristic for dimeric TFs and require algorithm alterations for detection (see also references 25, 50, and 315). While their method is exhaustive, its detection range is relatively limited, identifying patterns mainly with one or more highly conserved cores. An advantage of the oligonucleotide analysis and dyad analysis tools is that they are integrated into a wide collection of modular tools (RSAT) by which, for example, the string-based motifs that they produce can also be converted to PWMs (291, 297). A general limitation of enumerative methods is that searching for long sequence motifs is computationally expensive, and exhaustive searches become impractical for motif lengths longer than 10 nucleotides (232). One method to solve the problem is the use of suffix trees, as introduced by Sagot

A. CodY-binding sites from *L. lactis* MC1363 promoter

ctrA	A	A	T	T	G	T	C	T	G	A	C	A	A	T	T
dppA	A	T	T	T	T	T	C	T	G	A	C	A	A	T	T
gltA	T	A	T	T	T	T	C	T	G	A	A	A	A	T	T
dppP	T	A	T	T	G	T	C	A	G	A	A	A	A	T	T
serC	A	A	T	T	A	T	C	A	G	A	A	A	A	T	T
oppD	A	A	T	G	T	T	C	A	G	A	A	A	A	T	T
dppA	A	A	T	A	T	T	C	T	G	A	A	A	A	T	T
ilvD	A	A	T	G	T	T	C	T	G	A	C	A	A	A	T
asnB	A	A	T	T	T	C	C	A	G	A	C	A	A	T	T
dppP	T	G	T	T	T	T	C	T	G	A	A	A	A	T	T
gltB	T	A	T	A	T	T	C	T	G	A	T	A	A	T	T
gltA	A	A	T	T	T	T	C	G	G	A	A	T	A	A	A
ctrA	A	T	T	C	G	T	C	A	G	T	A	A	A	T	T
pepC	A	A	T	T	A	T	C	A	A	A	A	A	A	A	T
ctrA	T	T	T	T	T	T	C	A	A	A	A	A	A	A	T
prtP	A	A	T	T	T	A	C	A	G	A	T	A	A	A	A
gltA	T	A	T	T	T	T	C	T	A	A	A	A	A	A	A
ilvD	A	T	T	T	A	T	C	G	G	A	A	T	A	A	T
Consensus	A	A	T	T	T	T	C	W	G	A	A	A	A	T	T

B. Position frequency matrix (PFM)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
A	12	13	0	2	3	1	0	8	3	17	12	16	18	7	3
C	0	0	0	2	0	1	18	0	0	0	4	0	0	0	0
G	0	1	0	2	3	0	0	2	15	0	0	0	0	0	0
T	6	4	18	12	12	16	0	8	0	1	2	2	0	11	15

C. Position weight matrix (PWM)

A	0, 90	1, 01	−2, 39	−1.09	−0, 71	−1, 60	−2, 39	0, 39	−0, 71	1, 36	0, 90	1, 28	1, 44	0, 23	−0, 71
C	−2, 39	−2, 39	−2, 39	−0, 53	−2, 39	−1, 18	2, 24	−2, 39	−2, 39	−2, 39	0, 26	−2, 39	−2, 39	−2, 39	−2, 39
G	−2, 39	−1, 18	−2, 39	−0, 53	−0, 08	−2, 39	−2, 39	−0, 53	1, 99	−2, 39	−2, 39	−2, 39	−2, 39	−2, 39	−2, 39
T	0, 01	−0, 41	1, 44	0, 90	0, 90	1, 28	−2, 39	0, 39	−2, 39	−1, 60	−1, 09	−1, 09	−2, 39	0, 79	1, 19

D. Sequence logo

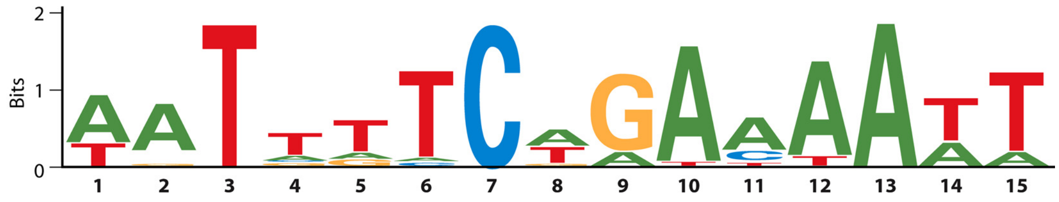


FIG. 6. Different representations of the CodY consensus sequence. (A) Data collection for a group of TFBSs for a specific TF (predicted CodY binding sites in promoters of genes from the *L. lactis* CodY regulon [68]). A consensus sequence can be calculated from such a list, in which degeneracy is taken into account according to IUPAC nomenclature. (B) Position frequency matrix taken from the data in A. (C) PWM in which the frequencies are weighed by the total sites assessed and the percent GC content of the genome. (D) Sequence logo (59, 262). Heights of the letters represent the information contents at each position. It should be noted that sequence logos are only visual representations, which are not used as an input for computational algorithms.

(254), which effectively reduce the size of the search space, making search time exponential with respect not to motif length but to the number of mismatches allowed in the motif. The well-known motif-finding algorithms Weeder and MITRA are also equipped with such suffix trees (81, 229). Recently, hybrid algorithms have also been proposed, in which probabilistic models are incorporated in dictionary-based methods related to enumerative algorithms (41, 253, 253, 309).

Probabilistic methods mostly first develop a probabilistic model (mostly a PWM) of the sequence data and then optimize it to find motifs common to multiple input sequences. Two algorithms frequently used for optimization are the expectation-maximization (EM) algorithm (43, 67, 172) and Gibbs sampling (98, 171). EM algorithms start off with a guess PWM as an initial motif model, consisting of a single oligonucleotide subsequence (n-mer) and background oligonucleotide frequencies. For each n-mer in the target sequence, the probability that it was generated by the motif instead of chance effects in the background sequence is calculated. Subsequently, the algorithm iterates between calculating a new motif model based on the old model plus the added motif sequences and calculating the probabilities of n-mers in the target sequence given this model (43, 67, 172) until a convergence criterion is reached. A disadvantage of the EM algorithm is that it is a

TABLE 2. Bioinformatics tools and databases for analysis, visualization, or discovery of TFs, TFBSs, and gene regulatory networks

Tool or database	Description	URL
TFBS/motif discovery tools		
MEME	Motif discovery algorithm	http://meme.sdsc.edu/
Weeder	Motif discovery algorithm	http://159.149.109.9:8080/weederweb2006
AlignACE	Motif discovery algorithm	http://atlas.med.harvard.edu/
MDScan/BioProspector	Motif discovery algorithm	http://seqmotifs.stanford.edu/
Consensus	Motif discovery algorithm	ftp://www.genetics.wustl.edu/pub/stormo/Consensus/
PhyloCon	Motif discovery algorithm	http://ural.wustl.edu/~twang/PhyloCon/
PhyloGibbs	Motif discovery algorithm	http://www.phylogibbs.unibas.ch
RSAT	Large series of regulatory analysis tools, containing oligonucleotide and dyad analysis	http://rsat.ulb.ac.be/rsat/
SCOPE	Ensemble motif discovery tool	http://genie.dartmouth.edu/scope/
MotifVoter	Ensemble motif discovery tool	http://www.comp.nus.edu.sg/~bioinfo/MotifVoter/
rVista	Phylogenetic footprinting tool	http://rvista.dcode.org/
Footprinter	Phylogenetic footprinting tool	http://genome.cs.mcgill.ca/cgi-bin/FootPrinter3.0/
Consite	Phylogenetic footprinting tool	http://consite.genereg.net/
Promoter prediction		
PPP	Promoter prediction tool	http://bioinformatics.biol.rug.nl/websoftware/ppp/
PromEC	Database of *E. coli* promoters	http://bioinfo.md.huji.ac.il/marg/promec
SAK	σ^{70} promoter analysis	http://nostradamus.cs.rhul.ac.uk/~leo/sak_demo/
Beagle	σ^{70} promoter analysis	http://eresearch.fit.qut.edu.au/Beagle/
Databases		
RegulonDB	Database of *E. coli* transcriptional regulation	http://regulondb.ccg.unam.mx/
DBTBS	Database of *B. subtilis* transcriptional regulation	http://dbtbs.hgc.jp/
BacTregulators	Database of prokaryotic TFs	http://www.bactregulators.org/
MicrobesOnline	Expression data, evolutionary relationships, operon/regulon predictions	http://www.microbesonline.org/
DBD	Database of predicted TFs	http://transcriptionfactor.org
RegTransBase	Database of prokaryotic TFBSs and regulatory interactions	http://regtransbase.lbl.gov
Prodoric	Database of prokaryotic gene regulation	http://prodoric.tu-bs.de
MtbRegList	Database of transcriptional regulation in *Mycobacterium tuberculosis*	www.usherbrooke.ca/vers/MtbRegList
cTFbase	Database of cyanobacterial TFs	http://bioinformatics.zj.cn/cTFbase/
Coryneregnet	Database of corynebacterial TFs and gene regulatory networks	http://www.coryneregnet.de/
TractorDB	Database of gammaproteobacterial regulatory networks	http://www.tractor.lncc.br/
ArchaeaTF	Database of archaeal TFs	http://bioinformatics.zj.cn/archaeatf/
Visualization		
Genome2D	Motif detection in genomic context	http://molgen.biol.rug.nl/molgen/research/molgensoftware.php
MOTIFATOR	Motif detection and visualization in gene context	http://www.motifator.nl
Weblogo	Motif visualization	http://weblogo.berkeley.edu/
Motif Distribution Viewer	Motif distribution visualization	http://h-invitational.jp/mdv/
Network analysis		
Cytoscape	Network visualization and analysis	http://www.cytoscape.org/
Visant	Network visualization and analysis	http://visant.bu.edu/
Osprey	Network visualization and analysis	http://biodata.mshri.on.ca/osprey/index.html
Pajek	Network visualization and analysis	http://pajek.imfm.si/
Vanted	Network visualization and analysis	http://vanted.ipk-gatersleben.de/
Biotapestry	Network visualization and analysis	http://www.biotapestry.org/
TYNA/Topnet	Network analysis	http://tyna.gersteinlab.org/tyna/
Bioconductor	Network visualization and analysis	http://www.bioconductor.org/

local optimization method that is sensitive to the initialization point. The well-known motif discovery tool MEME, which is based on the EM algorithm, largely avoids local maxima by performing a single iteration for each *n*-mer in the target sequences and iterating the best motif from this set to convergence (10). Gibbs sampling can be considered a stochastic variant of EM (98, 171). In Gibbs sampling, the algorithm starts off with a number of *n*-mers randomly sampled from the input sequences. It then probabilistically decides for each iteration whether to remove an old site from and/or add a new site to the motif model. The probability is weighted by the binding probability for those sites based on the old model (180). Well-known motif discovery tools based on Gibbs sampling are AlignACE (249), MotifSampler (290), and BioProspector (179). Like the EM algorithm, Gibbs sampling can suffer from the problem of the presence of local optima. GibbsST is a promising algorithm that circumvents this problem in a new way, by a thermodynamic method called simulated tempering (269).

Because so many different tools are available for DNA motif discovery, balanced comparisons are of major importance. Although some efforts in this have been attempted (134, 276, 293), it remains a major challenge to the work field to find objective standards for algorithm evaluation. The main reason for this is that the various tools score differently depending on the data sets, and absolute benchmarks are lacking (256, 293). Tompa et al., who created eukaryotic benchmark data sets with which they tested 13 commonly used algorithms, found no single program to be superior across all performance measures and data sets (although Weeder outperformed the other tools in most cases) (293). Hu et al. performed a similar analysis with prokaryotic benchmark data sets for five motif discovery tools, although their analysis differed in that they allowed minimal parameter tuning during performance evaluation (134). Both studies found that the absolute measures of correctness of all programs were quite low, although Hu et al. found that the algorithms which they tested were capable of predicting at least one binding site accurately more than 90% of the time (134). Because of the limitations inherent in any single motif discovery tool, users are advised to use multiple algorithms, to run probabilistic algorithms multiple times, to pursue the top few motifs instead of the single most significant one, to combine similar motifs, and to evaluate the resulting motifs in terms of group specificity, set specificity, and positional bias.

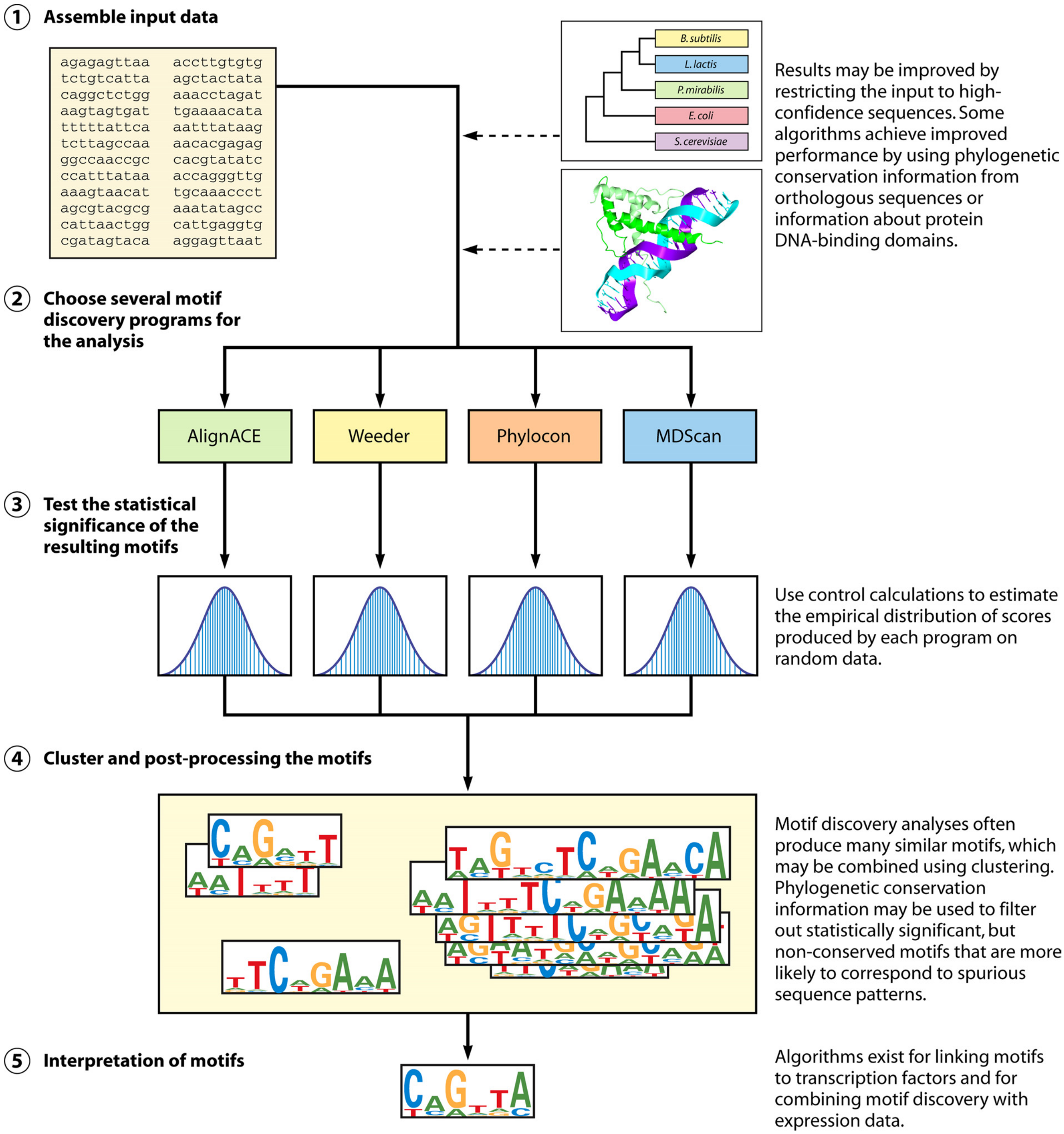
A consensus is now emerging that because no single program is superior for all data sets, several programs (preferably based on different methodologies) should be combined to achieve optimal results (119, 134, 293) (Fig. 7). Hu et al. found that an ensemble method that combined outcomes of the tools that they tested increased both sensitivity and specificity considerably (134). They later extended the method in their EMD algorithm (135). Recently, two additional applications (SCOPE and MotifVoter) that combine the results of different motif search algorithms for prokaryote data have become available (44, 316). The application MOTIFATOR is focused on prokaryote data analysis and uses the SCOPE algorithm to search for overrepresented DNA motifs in upstream regions of DNA microarray targets (31). The resulting motifs are presented in combination with functional enrichment and a visualization of the putative TFBSs in relation to the ORF to allow the user to prioritize results. While SCOPE merges the scores of three complementary algorithms (BEAM [45] for nondegenerate motifs, PRISM [46] for degenerate motifs, and SPACER [50] for bipartite motifs), MotifVoter extracts its motifs by clustering the results of up to 10 well-known motif discovery tools such as Weeder, MEME, and AlignACE. Notably, MotifVoter significantly outperformed earlier ensemble algorithms on the benchmark data set reported by Tompa et al. as well as on a bacterial (*E. coli*) benchmark data set (316).

Comparative genomic approaches can also be used to detect TFBSs or to filter results from enumerative and alignment methods by using the assumption that nucleotides in a binding site motif are generally better conserved than the nucleotides in the vicinity of the binding site. With these so-called phylogenetic footprinting approaches, conserved regions that point to the presence of important functionality, i.e., TFBSs (and also RNAP/ribosome binding sites), are identified (30, 131). The most basic methodology is to construct a global multiple sequence alignment of the orthologous promoter sequences using an alignment tool such as ClustalW (292) and then to manually identify conserved regions within this alignment (Fig. 8). Genomes of three species having the optimal phylogenetic distance toward each other could be sufficient for the detection of such conservation (206). However, such an approach to phylogenetic footprinting does not always work because it may be difficult to obtain an accurate alignment, or an obtained alignment may be uninformative. Therefore, several motif-finding algorithms have been adapted to detect phylogenetic footprints in promoters of orthologous genes in tools such as OrthoMEME (235), Footprinter/MicroFootprinter (29, 217), PhyloCon (310, 311), PhyME (277), and PhyloGibbs (273). Some methodologies that avoid the use of alignments altogether have even been proposed (79, 106). Recently, an approach in which predicted motifs throughout different taxonomic levels can be compared has also been developed, which enables one to detect not only motif conservation but also motif divergence (144). Finally, the conservation of the genomic context of TFs can be used to detect genes regulated by a TF, after which motifs of such a TF can be obtained through the footprinting of all orthologues that share this identical genomic context (89, 313). Although the degeneration or turnover of a TFBS in one or more specific phylogenetic lineages is a potential hazard to the phylogenetic footprinting approach (136), a computational approach (CSMET) that takes into account such a lineage-specific evolution of TFBSs has recently been developed (241).

Finally, prediction approaches that make use of structural information about the TF (4a, 149, 159, 178, 213), either from crystallographic structures or from homology models, have recently been applied. Although the use of such models for ab initio predictions of TFBSs is still limited, Morozov and Siggia have shown that it can be used successfully to compute a PWM for a certain TFBS motif using the combination of structural information and a single strict consensus sequence (213). The method proceeds from the assumption that the conservation of a base pair in the binding site is correlated with the number of atomic contacts between that base pair and the TF, which functions as a reliable proxy of TF-TFBS binding affinity.

Transcriptional Regulatory Network Analysis and Reconstruction

Gene regulatory networks or TRNs have become an important tool in studying global transcriptional regulation in prokaryotes (5, 11, 17, 125, 150, 259, 260). Figure 9 shows an example of the visualization of the *E. coli* K-12 TRN. In this figure, the nodes (boxes) correspond to genes, and the edges (lines) are the interactions between the genes. An interaction between the TF and its target is denoted as an edge between the TF node and its target node. The network is built by interconnecting the TF nodes to form larger network structures. Within a TRN, smaller network modules can be distinguished (for a review, see reference 3). These network modules are (i) positive and negative autoregulation (a TF regulates its own expression); (ii) feed-forward loops, where regulator A regulates the expression of regulator B and target C. (regulator B additionally regulates the expression of C; there are eight



1 Assemble input data
agagagttaa accttgtgtg
tctgtcatta agctactata
caggctctgg aaacctagat
aagtagtgat tgaaaacata
tttttattca aatttataag
tcttagccaa aacacgagag
ggccaaccgc cacgtatatc
ccatttataa accagggttg
aaagtaacat tgcaaaccct
agcgtacgcg aaatatagcc
cattaactgg cattgaggtg
cgatagtaca aggagttaat
B. subtilis
L. lactis
P. mirabilis
E. coli
S. cerevisiae
Results may be improved by restricting the input to high-confidence sequences. Some algorithms achieve improved performance by using phylogenetic conservation information from orthologous sequences or information about protein DNA-binding domains.
2 Choose several motif discovery programs for the analysis
AlignACE
Weeder
Phylocon
MDScan
3 Test the statistical significance of the resulting motifs
Use control calculations to estimate the empirical distribution of scores produced by each program on random data.
4 Cluster and post-processing the motifs
Motif discovery analyses often produce many similar motifs, which may be combined using clustering. Phylogenetic conservation information may be used to filter out statistically significant, but non-conserved motifs that are more likely to correspond to spurious sequence patterns.
5 Interpretation of motifs
Algorithms exist for linking motifs to transcription factors and for combining motif discovery with expression data.

FIG. 7. Motif discovery workflow. (Adapted from reference 191, which was published under a Creative Commons license.)

different regulatory combinations possible depending on the Boolean logic) (Fig. 5); and (iii) dense overlapping regulons, where gene expression is driven by a combination of TFBSs for different TFs.

These networks allow the study of the signal integration occurring at the promoters of genes (which are represented as nodes in the network) in a wider context. Additionally, predictions of the functioning of larger regulatory structures in the cell can follow from studying TRNs (265). Another example of analysis of TRNs is given by Carrera and coworkers, who described a method that allows predictions of the response of a TRN following perturbations (e.g., knockout of a TF) (47). A combination of analysis and reconstruction was given by Barrett and Palsson, who described an algorithm that allows the reconstruction of a TRN of a given organism by the iteration of a prediction of the most informative perturbation, performing that perturbation in the laboratory, and reconstructing the TRN including the new information (13). (see Table 2 for an overview of methods involved with [gene regulatory] network analysis and visualization). Below, we nonexhaustively describe some approaches to gene network reconstruction, i.e., computationally determining interactions between genes.

The most common approaches are the modeling of Boolean logic networks (33, 163, 201, 272) or the use of Bayesian mod-

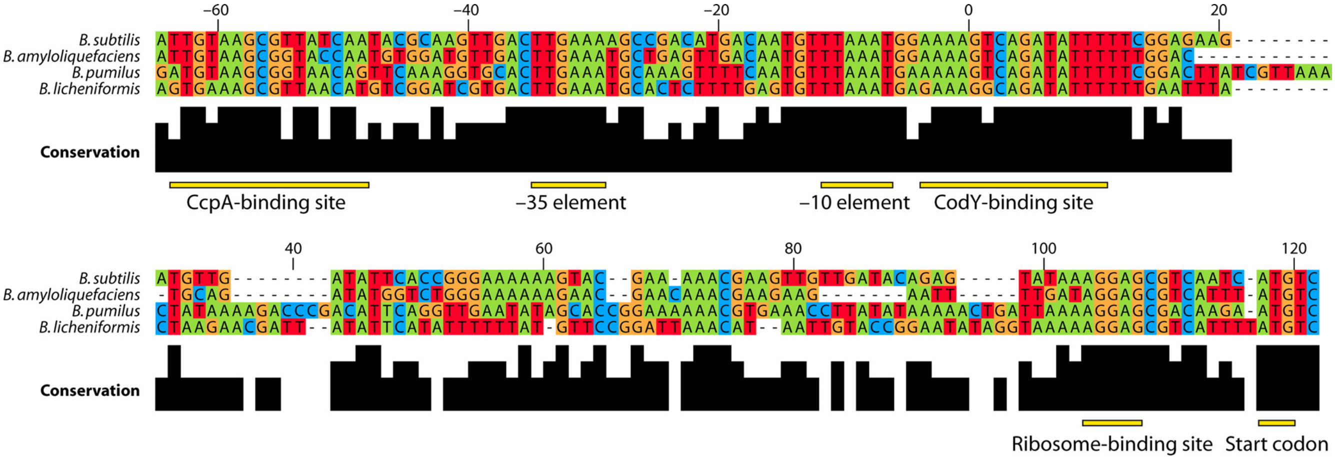

−60
−40
−20
0
20
B. subtilis ATTGTAAGCGTTATCAATACGCAAGTTGACTTGAAAAGCCGACATGACAATGTTTAAATGGAAAAGTCAGATATTTTTCGGAGAAG---------
B. amyloliquefaciens ATTGTAAGCGGTACCAATGTGGATGTTGACTTGAAATGCTGAGTTGACAATGTTTAAATGGAAAAGTCAGATATTTTTTGGAC----------
B. pumilus GATGTAAGCGGTAACAGTTCAAAGGTGCACTTGAAATGCAAAGTTTTCAATGTTTAAATGAAAAAAGTCAGATATTTTTCGGACTTATCGTTAAA
B. licheniformis AGTGAAAGCGTTAACATGTCGGATCGTGACTTGAAATGCACTCTTTTGAGTGTTTAAATGAGAAAGGCAGATATTTTTTTGAATTTA---------
Conservation
CcpA-binding site
−35 element
−10 element
CodY-binding site
40
60
80
100
120
B. subtilis ATGTTG---------ATATTCACCGGGAAAAAAGTAC--GAA-AAACGAAGTTGTTGATACAGAG-----TATAAAGGAGCGTCAATC-ATGTC
B. amyloliquefaciens -TGCAG---------ATATGGTCTGGGAAAAAAGAAC--GAACAAACGAAGAAG-------AATT-----TTGATAGGAGCGTCATTT-ATGTC
B. pumilus CTATAAAAGACCCGACATTCAGGTTGAATATAGCACCGGAAAAAAACGTGAAACCTTATATAAAAACTGATTAAAAGGAGCGACAAGA-ATGTC
B. licheniformis CTAAGAACGATT--ATATTCATATTTTTTAT-GTTCCGGATTAAACAT--AATTGTACCGGAATATAGGTAAAAAGGAGCGTCATTTTATGTC
Conservation
Ribosome-binding site
Start codon

FIG. 8. Example of phylogenetic footprinting in which promoter sequences from different genomes are aligned to find stretches of nucleotides that are evolutionarily conserved and are thus probable to have regulatory functions. This example shows phylogenetic footprinting of the *ackA* promoter in four *Bacillus* species.

els or coexpression measures to create probabilistic networks (90, 91, 161, 258). More complex network models have also been introduced, such as continuous (rather than logical) models (57, 77, 216, 227) and single-molecule-level models (103, 263, 333). Reference or template-based network reconstruction is a methodology that uses reference networks to predict edges between genes for a given organism (18). CoryneRegNet is a database that contains data for regulatory interactions for a number of organisms, including *E. coli* K-12, that can be used for this purpose (15). Each of these methodologies has its own advantages: logical models allow relatively easy and flexible fitting to large-scale biological phenomena, continuous models allow an understanding of more confined processes that rely on finer timing and exact molecular concentrations, and single-molecule-level models allow study of the stochastic aspects of gene regulation. Template-based methods allow one to use knowledge on TRNs generated for different organisms. Although TRNs can be quite well compared between some related organisms (16), it remains to be established whether this assumption generally holds for other species and more special-

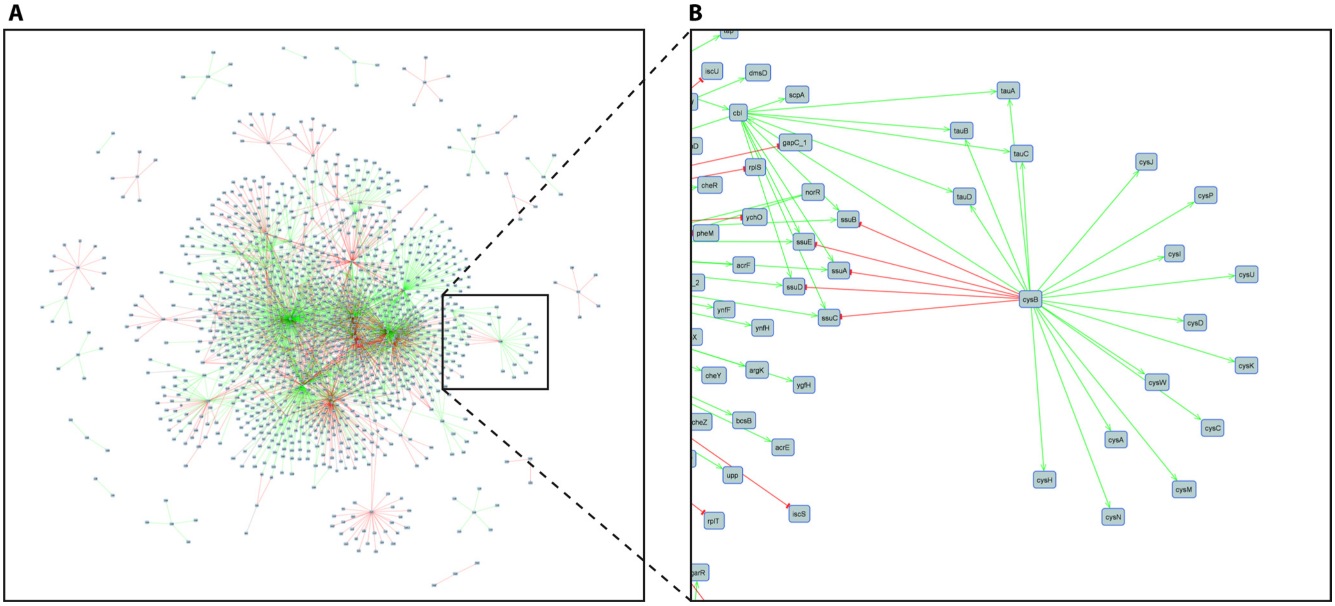


FIG. 9. The *E. coli* K-12 transcriptional regulatory network. The *E. coli* K-12 network was obtained from RegulonDB, version 5 (94), and visualized using GeneVis software (314) with a spring layout. (A) The entire gene regulatory network consists of interconnected TFs (hubs in the network). (B) Detail of the network consisting of the *cysB* regulator and its targets. The nodes (boxes) correspond to the genes, and the edges (lines) denote interactions between the genes (nodes). The direction of an interaction is indicated from the base of the arrow (regulator or TF) to the arrowhead (target gene). An interaction is between either the TF and its target (e.g., between CysB and TauC) or two TFs (e.g., between CysB and Cbl). The green edge (with arrow) indicates the activation of target expression by the TF; e.g., CysB activates the expression of the TF Cbl. The red edge indicates a repression of the target expression by the TF; e.g., CysB inhibits the expression of the SsuABCDE operon. The gray edge indicates an interaction of a TF, and its target is unknown.

ized gene regulatory modules. The major data source for the above-mentioned approaches is gene expression data obtained from microarray experiments. Based on benchmarks of reconstruction using different algorithms and synthetic data, the reconstruction of TRNs, and conceivably determining regulon structure (see above), has been shown to be most effective when small time series of genetic perturbations are used, as opposed to larger-time-series microarray data (97).

For all these approaches, the process of reconciling laboratory data (gene expression data and ChIP-on-chip) with bioinformatic regulon predictions is of major importance (126, 302). This integration step is necessary to be able to reliably analyze genome-scale models of TRNs to predict the effects of the application of different stimuli to an organism (19). Schlitt and Brazma proposed subdividing regulatory network models into four categories: (i) part lists (systematized lists of network elements in a particular organism or system), (ii) topology models (the parts including their interconnections), (iii) control logic models (the description of the combinatorial effects of regulatory signals), and (iv) dynamic models (the simulation of the network in time) (259, 260). Currently, there are large gaps between part lists that, for example, constitute a regulon and topology models, in which the part lists are integrated to yield a network topology (259). A further level of complexity is added with the control logic of networks, which has been described in a number of studies (62, 96, 162, 163, 255).

There are still a number of categories of inconsistency between the models and experimental observations. For example, not all physical interactions reported by, e.g., ChIP-on-chip between TFs and *cis*-regulatory regions result in significant functional regulatory effects that are detectable in gene expression data (259). Moreover, many transcripts remain below detection limits of the techniques used in high-throughput gene expression studies (32). Also, many inconsistencies exist between TF-DNA interactions predicted by computational approaches (e.g., PWM-based methods) and ChIP-on-chip data (170, 211). Even in large collections of gene expression data collected under many different conditions, sometimes no transcriptional effects are discovered for certain TFBSs (127). Last but not least, only a small complement of an organism's genes is active under the laboratory conditions (single-species growth in liquid culture) in which they are commonly grown, so available microarray data query only a limited part of the regulatory space (281).

TFBS DISTRIBUTION THROUGHOUT THE GENOME

Currently, reconstruction of regulons or networks of regulons is done primarily by using DNA microarray data in conjunction with literature knowledge and in some cases is supplemented with data for protein-DNA interactions. This involves searching for overrepresented DNA motifs in the upstream regions of target genes (see above). Current algorithms that were developed for searching overrepresented DNA motifs create a background model of the genome. These background models are based mostly on (oligo)nucleotide distributions across genomic regions. In the following section, the genomic distribution of TFBSs is discussed. This information can be used to further improve detection of TFBSs and to reduce the number of false-positive and false-negative results.

Distribution of Spurious and Functional Sites

A main obstacle for TFs to locate their functional binding sites across the chromosome are spurious TFBSs, sites with relatively high binding affinity (and relatively close to the TFBS consensus sequence) that have arisen nonadaptively throughout the genome without having been selected for a particular biological function (169). The fact that TFs do not have strict sequence specificity means that through simple mutations, spurious binding sites can quite easily appear by chance at positions where they do not significantly affect the transcription of nearby genes (174). Such spurious binding sites will lower the effective TF concentration within a cell.

Initial investigations into the distribution and dynamics of spurious TFBSs have been made by Huerta and coworkers (137, 138), who focused on the distribution of RNAP σ-factor binding core promoter elements throughout eubacterial and archaeal genomes. Their statistical investigation, in which they counted the number of RNAP binding motifs throughout different regions of 44 genomes, has shown that σ^{70} binding to −35 and −10 core promoter elements is overrepresented in regulatory regions (generally upstream regions) compared to nonregulatory regions of genomes (138). In two other studies, dinucleotide and/or trinucleotide frequencies of different genome regions were incorporated into the analysis to show that RNAP binding sites are also present below expectations (the number of motifs expected to arise by chance given certain oligonucleotide frequencies) in both coding and noncoding regions of bacterial genomes, which implies that natural selection acts to counter the appearance of spurious sites (92, 114). RNAP binding to −10 sites appeared to be overrepresented within regulatory regions relative to nonregulatory regions, even when the sites at the −10 position itself are not taken into account (92). Multiple −10 sites throughout promoter regions could perhaps function to maintain local RNAP abundance in these regions, or a *cis*-regulatory region may contain two promoters in tandem. In a study by Radonjic and coworkers, RNAP was reported to be present in the upstream regions of genes to ensure a fast response when the eukaryote *Saccharomyces cerevisiae* exits the stationary growth phase (238). RNAP binding sites downstream of the core promoter can also have important functions, such as the −10 site-resembling element at the transcriptional start site of the *E. coli lac* promoter, which mediates a transcription pause (34, 219). This mechanism possibly functions as a negative regulator of transcription in which the rescue of the stalled RNAP complex is dependent on one or more other TFs.

A similar, although not as extensive, study has been done by Hamoen et al. on a specific TF, the competence factor ComK, which was previously mentioned (115). Those researchers found that while both of the ComK binding sites (K-boxes) without any mismatches and 18 out of 25 of the K-boxes with one mismatch from the strict consensus were positioned in intergenic regions, only 56 of the 171 K-boxes with two mismatches were positioned in intergenic regions. Also, of the K-boxes with three mismatches, only 280 of the 864 were present in intergenic regions. Yet still, K-boxes with three mismatches were overrepresented in these intergenic regions, as they cover only 12% of the genome, and the difference in the percent GC content between genic and intergenic regions also could not account for the 32% of the triple-mismatched K-boxes found there. The only drawback of this study is that it did not take into account oligonucleotide frequencies, which in genic regions, for example, may be influenced by codon biases (4).

In general, it is also important that it is difficult to pinpoint which TFBSs are spurious and which are not. In a recent study, Shimada et al. found that 14 out of the 20 targets of the *E. coli* RutR TF found by ChIP-on-chip analysis were located in coding regions of the DNA and had little or no effect on transcription levels when tested (270). However, the computational prediction that other bacteria containing RutR homologues also have RutR TFBSs that are overrepresented in coding regions makes it tempting to suggest that RutR has some unknown function within these regions (270). Nonetheless, these results can just as well be explained as being an evolutionary relic.

One source of biological information that may help to distinguish between spurious and functional binding sites is that most local (nonpleiotropic) TFs tend to be encoded in close chromosomal proximity with one of their target genes, as was shown for *E. coli* by Janga et al. (142). Multiple biophysical models have shown that this makes sense because it allows the TFs to quickly reach their targets after translation, even at low concentrations, by sliding along the adjacent DNA (20, 158, 322). This implies that quite probably, an important part of the information determining the biological relevance of a TFBS is not present in its sequence but rather is present in its position on the chromosome (143).

Natural Selection and TFBS Motif-Like DNA Sequences

Studying the effect of natural selection on the abundance of TFBS motif-like sequences in different genomic regions may reveal much about their functionality (113). Given the fact that in large bacterial populations, natural selection is by far the major determinant of genomic sequence (189), selection can be quantified quite easily by comparing the abundance of TFBS motif-like sequences with the abundance expected from chance alone. The genomic abundance of every short DNA motif sequence expected by chance can be calculated from oligonucleotide frequencies. These comparisons between observed and expected abundances can be performed with different regions of prokaryotic genomes (e.g., coding and noncoding). For many TFs, such analyses can reveal in which regions there is selection either for or against the presence of their TFBSs.

Besides distinguishing between general sequence categories such as coding or regulatory regions, different genomic regions can be specified for this analysis. For example, the abundance of certain TFBSs in different parts of *cis*-regulatory regions (e.g., −30, −50, and −100 nucleotides relative to the transcriptional start site) could be assessed separately to specifically identify the regions within promoters to which particular TFs generally bind. For example, the observed/expected abundance ratios of certain TFBS motif-like sequences in the first 50 to 100 nucleotides of coding regions could be compared with the observed/expected abundance ratios of these sites in coding regions. This might give insight in the natural selection leading to a large abundance of TFBS motif-like sequences in the 5′ part of coding regions. This, in turn, would then point to a possible roadblock function of the corresponding TF.

Methodology for Studying Natural Selection on TFBS Motifs

Both the hidden Markov model analysis used to calculate expectations of TFBS abundance from genomic oligonucleotide frequencies (60, 130) and the sliding-window approach to count the number of DNA motifs with a certain number of mismatches from the strict consensus (8) are straightforward. A more elaborate algorithm for detecting the positional overrepresentation of TFBSs that uses PWMs of spatially conserved motifs based on comparative genomics techniques was developed by Defrance and Touzet (65). Therefore, these analyses have the potential to become standard tools for the study of transcription as an addition to the most commonly used PWM-based tools. Information from such methods would enhance standard positional statistics of TF distribution throughout genomic regions, as was used by Huerta and coworkers, for example, who created simple motif density maps showing the location distribution of core promoter-like elements throughout regulatory regions (138). It should be kept in mind, however, that the regions selected as input for the analyses should be sufficiently large so that the motifs under study do not significantly affect the di- or trinucleotide frequencies themselves.

Finally, when TFBS sequences of a TF are sorted based on the distance from the degenerate consensus sequence, the strictness of the motif-TF interaction could perhaps be monitored by observing the effect of natural selection (assuming this to be the major mechanism shaping genomic sequences of bacteria) on the abundance of these sequences in relation to their distance to the degenerate consensus. Because consensus methods probably do not offer sufficient accuracy for such an analysis, PWMs could be used, by calculating PWM scores for all TFBS-like DNA words and observing the effect of natural selection on the abundance of groups of DNA words with different PWM scores. An alternative method is to approach this problem from the perspective of the actual biological effects of the TFBS-like sequences by performing a two-dimensional clustering with the PWM scores as a function of gene expression. This would also provide a means of visualization, and currently, several groups are following this approach (54, 64).

EVOLUTIONARY DYNAMICS OF *cis*-REGULATORY REGIONS

In order to realistically predict which DNA motifs in a genome are functional TFBSs and which are not, it is important to understand how TFBSs evolve. After all, the sequence of any TFBS is shaped by its evolutionary heritage. In the following section, we review the evolution of the information content of the nucleotides making up TFBSs. This information is a highly important yet a complex piece of the transcriptional regulation puzzle.

Evolution of Regulatory Networks

An important aspect of TRNs is that they can evolve rapidly (6, 7, 99, 141, 186). Therefore, transspecies extrapolation of information from TRNs is possible in only a very limited taxonomic range (16). The regulatory effects of a TF often vary already significantly between different strains of the same bacterial species (122). The structure and sequence of *cis*-regulatory elements may change even when gene expression patterns are conserved because there is a significant turnover of binding site sequences (74, 136, 187). In such a turnover event, a new TFBS bound by the same TF evolves next to the original TFBS, after which the original TFBS degenerates (100). Furthermore, TFs for which strong phylogenetic evidence exist that they are evolutionary orthologues rarely regulate orthologous genes (237). Amazingly, a recent study even shows that a TF (Lrp) from *Proteus mirabilis* that was heterologously expressed in the closely related bacterium *E. coli* regulated only 51% of the genes that were regulated by its highly similar (98% sequence identity) *E. coli* Lrp orthologue under the same conditions (176). In another study, *B. subtilis* ComK, which is normally a transcriptional activator, appeared to function mainly as a repressor when it was heterologously expressed in *L. lactis* (286). Also, by studying PhoP orthologues from *Salmonella enterica* and *Yersinia pestis* (79% identical), Perez and Groisman found that they acted differently on promoters (one able and one unable to induce transcription) in the two species even in a case where both orthologues bound the PhoP binding site in the promoter effectively (230). Apparently, the evolution of TF-TFBS interactions involves a complex interplay of both minor modifications to the sequences of TFs and functional changes in the architecture of promoters.

In the long run, TFs seem to evolve quite independently of their target genes through the rapid genome-wide tinkering of transcriptional interactions (7). Genes coding for repressors coevolve more tightly with their targets than do genes encoding activators. An activator can be lost when its targets remain in the genome. In contrast, a repressor usually can be lost by a genome only after either its target genes have also been lost or the TRNs have rewired significantly to diminish the regulatory role of the repressor (128). Therefore, the information content of repressor TFBSs is expected to be more conserved across related organisms.

Detailed computer simulations have shown that *cis*-regulatory regions can evolve in relatively little time through local point mutations, although the details of these models were based on eukaryotic genomes (22, 76). Also, local duplications caused by DNA strand slippage during replication, promoter rearrangements, and transposition of *cis*-regulatory regions between promoters can quickly generate novel TFBSs, although most of these processes have been studied in detail only for eukaryotes (156, 207). Furthermore, gene duplications that include *cis*-regulatory regions complicate the picture, because *cis*-regulatory regions of duplicate genes are known to be able to diverge rapidly (175, 223).

Interdependency of TFBS Nucleotides

Although, as noted above, the energy of the binding of a TF to one of its TFBSs is often quite well approximated by the sum of the independent contributions of several important nucleotides, which has been referred to as the "additivity hypothesis" (21), the correlation between binding affinity and distance to the strict consensus sequence or the PWM score of a TFBS may not be quite perfect. Two studies have shown that the additivity hypothesis cannot fully account for the binding energies in the sequence space of TFBS motifs. In the first study, binding affinities of the Mnt repressor of *Salmonella* phage P22 were determined for its binding sites, in which positions 16 and 17 of the 21-bp operator had been varied to account for all 16 possible dinucleotide combinations (194). The two nucleotides appeared to be clearly interdependent: if position 17 was not a C, the preference of position 16 changed from A to C. In the second study, Bulyk and coworkers used protein binding microarrays to assess the binding affinities of a TFBS of the mouse zinc finger protein Zif268 for all 64 combinations of three nucleotides (40). Their analysis showed that a dinucleotide model (in which the effect of every nucleotide is dependent on the adjacent nucleotides) fitted their data better than a mononucleotide model (in which every nucleotide is scored independently) (40).

A reanalysis of these studies showed that the interdependency of nucleotides in a TFBS differed between different TFs, with the information increasing 2 to 15% when shifting from a mononucleotide to a dinucleotide model (21). Although those authors concluded that additive models are still accurate enough to be of use, it is still probable that—especially for TFs with a lower affinity for their binding sites (21)—search models based upon the assumption of additivity (which constitute the large majority of models used) will produce more false-positive and false-negative results than models in which nucleotide interdependencies are incorporated (330). When not taking into account the interdependencies of nucleotides for TF binding, interpretation problems might arise, especially when assessing large motifs, for example, when an asymmetric high level of conservation of a large part of a motif boosts the PWM score, while another part of a motif governing an essential structural DNA-protein interaction has degenerated. In a recent study, this has been shown to be the case for ComK binding sites in *B. subtilis*, in which transcription activation was almost completely abolished when the second thymine of the K-box was mutated into a guanine, even though the rest of the motif stayed intact (287). Similar results were also reported by Michal et al. for the Ndt80 motif in the eukaryote *S. cerevisiae* (209) and by Francke et al. for the LacI family, where the central CG nucleotides in the motif are essential for TF binding (89).

Furthermore, the surrounding sequence could have a significant effect on the effective binding affinity of a site (204) because the binding affinity of the surrounding sequence affects the time required for a TF to find its target through one-dimensional diffusion along the DNA. It may also affect the half-life of TF-TFBS binding, because if the surrounding sequence has a relatively high affinity for the TF, it will diffuse away more easily. An actual example of the influence of the surrounding sequence composition on TFBS functionality is the TFBSs of *B. subtilis* CcpA (*cre* boxes), which are more active when positioned in an AT-rich nucleotide context than when positioned in a GC-rich context (325).

Information Content of TFBSs

The functionality or nonfunctionality of TFBSs is governed by evolutionary forces (215), which act mainly on the information content of DNA sequences. TFBSs are hard to identify because of the evolutionary tolerance (due to insufficient selection) of nonfunctional-site-resembling oligonucleotides and because of the array of evolutionary processes (the balance between selection, drift, and mutation) allowing fuzziness in functional binding sites.

If the size of the genome and the number of functional TFBSs is known, the amount of information needed for a TF to identify the site ($R_{\mathrm{frequency}}$) can be computed from the size of this genome and the number of sites (152). The information content of a TFBS (R_{sequence}) depends on motif length, motif stringency, and the genomic frequency of the nucleotides present in the motif (152). Evolutionary simulations have shown that the information content R_{sequence} of TFBSs will evolve to a value close to $R_{\mathrm{frequency}}$ (152, 261), and a clear inverse correlation between TF binding specificity and pleiotropy (defined by the number of functional target sites to which a TF binds) has been found for the genomes of *E. coli* and *B. subtilis* (185, 266). The possibility that regulon size could therefore be estimated from the information content of TFBSs is intriguing. Francke and coworkers described CcpA and LacI operator motifs (TFBSs) for *Lactobacillus plantarum* (89). Those authors indeed reported that the CcpA operator *cis*-acting replication element site is quite degenerate, which reflects the global role that CcpA has in the control of cellular metabolism (89). It should be noted that part of the observed degeneracy could also be due to the higher number of sequences on which the motif representation is based. However, in a broader study of this in *E. coli* and *B. subtilis*, Lozada-Chavez et al. found a clear general negative correlation between the DNA binding specificity and pleiotropy of TFs as well (185). Notably, it could also be predicted that certain classes of TFs are structurally fit to function as nonpleiotropic regulators provided that their three-dimensional structure permits binding to TFBSs with larger motif lengths that can contain more information. In the end, the information content of a motif is a tradeoff between motif length and motif stringency (89).

The sequences of pleiotropic regulator TFBSs tend to be more conserved during evolution because of higher functional constraints (240, 266). This points to an interesting paradox, where the motif stringency does not have to correlate with motif sequence conservation, as is the case for sequence motifs for nonpleiotropic regulators, which are more stringent but not more conserved at the sequence level. Therefore, TFBSs of regulators that bind only at a single promoter may be very hard to trace because no overrepresentation of them can be found within the genome itself and because the rapid coevolution of the TFBSs with the gene of its TF may make phylogenetic footprinting impossible. However, the TFBSs of regulators that bind to very few targets could be determined by combining conserved gene context with phylogenetic footprinting within a limited phylogenetic range (89).

TFBS Motif Fuzziness

The evolution of TFBSs has generally produced nonrandom fuzziness of TFBS sequence motifs relative to their strict consensus sequence. The variation that occurs for particular nucleotides is different for every position in a certain TFBS motif, which reflects the importance of each nucleotide in establishing the specific binding of a TFBS by its TF. The position-specific variation that can be found within one genome is generally conserved throughout other relatively closely related genomes (214). A common problem in identifying TBFSs is that the number of regulated genes should be sufficient to determine a degenerate consensus sequence. Phylogenetic footprinting is a powerful tool to increase the number of TFBSs that can be used for this. Furthermore, comparing the position-specific stringency of TFBS nucleotides within a genome together with their conservation degrees across genomes can give an indication of selective pressures that have acted on certain TFBS nucleotides.

Two main evolutionary scenarios have been proposed to explain TFBS motif fuzziness from an evolutionary perspective (100). One scenario is that the binding affinity of each site is optimized evolutionarily to maximize the functionality of the site. Because the functionality of the site may demand a low binding affinity, fuzziness is a logical evolutionary consequence. This has been observed, for instance, for LacI, where the perfect palindrome has a higher affinity than the actual motif (89). In a second scenario, the fuzziness of TFBSs is attained automatically as a consequence of the balance between mutation and selection, because the function of a TFBS would be insensitive to its precise TF binding affinity as long as it is above some threshold. It should be noted that the two scenarios are not in contradiction and may both account for a part of the observed fuzziness.

In *cis*-regulatory regions that contain a single TFBS, the first scenario would play out if the expression of the gene has a graded response to the TF concentration, while the second scenario would play out if it responds in a binary or sigmoid fashion (104). A gene regulation model which incorporates the rate of transcription in combination with motif stringency and TF concentration would be more accurate compared to on/off models. In such a model, the threshold of TF abundance should also be modeled. Protein binding microarrays (9) could be used to determine the in vitro threshold concentration that results in the binding of a TF to its TFBS as a function of the TFBS sequence. Interestingly, Bilu and Barkai conducted a genome-wide survey of TFBSs in the yeast *Saccharomyces cerevisiae* in which they found that binding sites tend to be shorter and fuzzier if they are situated in more complex promoters containing more than one TFBS (26). Because promoters of essential genes tend to be bound by fewer TFs (26), one possible explanation for this fuzziness is that promoters can evolve to a larger complexity when they are under low selective pressure.

Stabilizing selection on a promoter sequence is weak when variation in the transcription rate of a gene is not likely to result in a deletion of the gene (190, 239). In such a situation, the emergence of a novel TFBS is also less likely to have deleterious effects, and there is more opportunity for evolutionary processes to incorporate such novel sites in a manner that is advantageous to the organism while still allowing for fuzzy TFBSs (26, 190). Indeed, data from comparative genomic analyses suggest that new TFBSs tend to appear in promoters that already contain multiple sites (26). However, it could also be argued that this is merely because in a promoter that already contains multiple TFBSs, a new TFBS confers a smaller change in the transcription rate, while the selective pressure on this transcription rate may be just as high as that for other genes.

Cooperative and Competitive DNA Binding and Motif Stringency

The above-mentioned explanation of fuzziness may not be the whole story, as has been shown with examples of promoters with multiple TFBSs involved in cooperative or competitive DNA binding. Using a biophysical model of transcriptional regulation in which *cis*-regulatory regions with either homocooperative or heterocooperative sites were studied, Hermsen et al. found that TFBSs for which their TFs have weak binding affinity (i.e., fuzzy sites) probably have specific functions in cooperative transcription activation and repression (124).

In homocooperative activation, auxiliary TFBSs, which do not interact directly with the RNAP, need to be bound by their TF with higher affinity than does the primary site that interacts directly with the RNAP, in order to maximize the steepness of the response to the TF concentration (which is the primary function of homocooperativity) (38). On the other hand, in homocooperative repression, the auxiliary sites should be bound by their TFs with much weaker affinity than the primary site to establish a steep response (124). Thus, in each of these cases, the binding affinity of a TF for the auxiliary site is adapted in order to reach an optimal TF concentration dependence of the response. Experimental support for these results comes from *E. coli cis*-regulatory regions containing homocooperative LysR family activator binding sites and others containing homocooperative Fur repressor binding sites, which have both been studied in some detail and confirm the role of strong and weak binding sites proposed by the model (80, 165, 318). Therefore, in the case that multiple TFBSs for a single TF are present in a promoter, the secondary sites are expected to be more conserved than the primary TFBS in the case of activators and the other way around in the case of repressors.

Whereas most simple combinations of TFBSs lead to Boolean NOR or ANDN gates (153) (Fig. 5), the model constructed by Hermsen and coworkers also predicts that heterocooperativity or heterocompetitivity may facilitate more complex transcriptional responses, such as the Boolean AND or OR gates (Fig. 5). This finding is in accordance with earlier results reported by Buchler and coworkers (38). In promoters functioning as an AND gate, the core promoter (−35 and −10 RNAP binding sites) is weak, so there is no transcription without specific activation, and two TFBSs (binding TF1 and TF2) are present, which are both too weak to function by themselves and induce activation only cooperatively. Additional sites binding either TF1 or TF2 may be present in the promoter to steepen the response to the TF concentration (124). In such promoters, the fuzziness of the TFBSs is selective, since because of the lower binding affinity of the sites, both TFs are required to be present at sufficient concentrations to activate transcription.

The biological importance of cooperative regulation also has consequences for the prediction of TFBSs. Currently, TFBSs are determined case by case. The determination of TFBSs would benefit from the integration of searches on different TFBSs on the same *cis*-regulatory regions. In cases where multiple TFBSs are identified in a given *cis*-regulatory region, the motif detection stringency should be decreased in order to account for the more complex promoters in which cooperative or competitive regulation takes place. One reason is that TFBSs probably need less motif stringency to be functional if they are positioned next to another TFBS with a high binding affinity for the same TF because this will cause the local TF concentration in this promoter to be higher than normal. A second reason is that the biological usefulness of cooperative and competitive regulation mechanisms can be expected to have increased the frequency of TFBSs in promoters during evolution beyond the level that would be expected on the basis of selection acting on single TFBSs.

We conclude that in order to understand the structure and response of a TRN, the fuzziness of a TFBS should be considered in context with (the nature of) other TFBSs in the same promoter region. A fuzzy TFBS could still have an equally important role as a “perfect” TFBS in the case of homo- or heterocooperativity.

Fuzziness Due to Insufficient Selective Force on Stringency

The scenario in which TFBS fuzziness is a result of mutational entropy has been developed in two theoretical studies using the assumption that the fitness of a TFBS depends solely on its binding affinity for its TF. Gerland and Hwa reported that the fuzziness of motifs arises naturally from the balance between selection and mutation: mutations that slightly lower the affinity of TF binding to the TFBS are not rapidly removed by selection compared to the event of a new mutation (100). A study reported by Sengupta and coworkers emphasized this point, while they also found that TFBSs of TFs governing large regulons were more fuzzy than those of TFs targeting only a few specific sites (266). This may be both because the amount of mutational variation in the TFBSs of a certain TF increases with the number of TFBSs (higher mutational forces) and because the information required to identify a TFBS is less specific if more TFBSs are present in the genome (lower selective force [see above]).

It seems that the two scenarios (the selective scenario and the mutation-selection balance scenario) explaining TFBS motif fuzziness are in reality probably intertwined and that both processes play important roles in TFBS evolution. The contribution of each process probably differs according to both the complexity of promoters and the selective pressures acting upon them. Finally, from a broader perspective, another reason for TFBS fuzziness could be that less strict binding of TFs to DNA motifs (unlike restriction endonucleases) both creates robustness to deleterious mutations and enhances the evolvability of new TFBSs (307). Interestingly, in two studies of the *E. coli lac* operon, it was predicted that this operon can easily evolve from its intermediate form to a pure AND gate or a pure OR gate, because the fact that fuzziness is allowed during evolution facilitates the discovery of nearby sequence space (204, 267). In the context of the ever-varying evolutionary challenges that bacteria must face, it is the inherent evolutionary versatility and adaptability of the relationship between TFs and TFBSs that make these systems so successful.

Evolution of TFBS Multiplets by Binding Site Turnover

Although homocooperative interactions may explain the appearance of TFBS multiplets (multiple adjacent occurrences of the same type of TFBS) in quite a number of promoters, it probably does not account for all multiplets present in *cis*-regulatory regions. For example, probably not all TFs oligomerize on the DNA because they may bind to different faces of the DNA helix or may not have three-dimensional domains that strongly interact (320). In some promoters, multiple TF proteins act simultaneously with the RNAP to either repress or activate transcription in a synergetic manner without oligomerization, as was observed for the CRP TF in *E. coli* (166). An evolutionary model has confirmed that under high selective pressures, more than one TFBS can indeed be maintained in promoters when the binding sites contribute independently to transcriptional activation (100, 112). This process could also be a driving force in the apparently frequent process of binding site turnover because new sites can evolve under selection, while after relief of selection, the old site may degenerate instead of the new site (74, 187). For phylogenetic footprinting approaches to TFBS discovery, such binding site turnover forms a serious hazard.

CONCLUSIONS AND FUTURE PERSPECTIVES

In order to fully understand bacterial TRNs and to integrate experimental and computational information, an appreciation of the biological mechanistic intricacies of gene expression regulation is needed. As can be seen in Table 1 and Fig. 1, there is a large variety of biological mechanisms by which transcription is regulated in prokaryotes. So far, only a few of them have been taken into account in regulon reconstruction and TRN reconstruction efforts. Spatial positioning of TFBSs, motif stringency, and combinatorial regulation mechanisms should especially be taken into account.

Barrett and Palsson as well as Covert and colleagues predicted that through an iterative model-building strategy in which iterations of high-throughput experiments and in silico modeling are performed subsequently, regulatory network elucidation for the model organism *E. coli* could be completed within years (13, 57). Such iterative approaches are indeed promising, because in this way, future experimental research will be streamlined effectively to yield the most information-dense results. However, if complex regulatory mechanisms such as those discussed in this review play a major role in prokaryotes, the outlook given by Barrett and Palsson as well as Covert and coworkers is probably too optimistic. More complex models may be needed to arrive at a TRN with a minimum number of inconsistencies. Moreover, there are more general issues in network reconstruction. In many cases, DNA microarray data are still used as a primary data source. Large compendia of microarray data obtained under different conditions are required to distinguish between direct and indirect regulatory effects (87, 326). Even when large data sets are available, a limitation of this approach is that only those networks which are (differentially) expressed under the conditions in which the transcriptome analysis was performed can be reconstructed. Furthermore, current efforts are focused on the association of targets with their transcriptional regulator. This involves the assumption that the transcriptional regulator should be coexpressed with its targets. The problem with this assumption is that this is the case only for TFs with autoregulation; i.e., the TF regulates its own expression. For *E. coli* K-12, the numbers of TFs that negatively regulate themselves are most common and have been estimated to be about 50% (248). Another scenario could be that the transcriptional regulator is expressed earlier than its targets, which would require aligning and phasing gene expression patterns of the regulator and its targets (324). Still another approach could be to (i) determine stimulons (120), i.e., genes for which the expression is changed when applying a stimulus; (ii) determine the different regulons that are part of a stimulon; and (iii) determine the causal (TF-target) relationships within the regulons.

In order to be able to reliably reconstruct TRNs with the correct interactions between their nodes, it seems absolutely vital that more functionality of a promoter can be predicted and used as input than just the presence or absence of a certain TFBS in it. For one, the positions of TFBSs for activators and repressors should be taken into account. Recently, such information has been successfully implemented to increase motif search accuracies by searching for sequence motifs that are nonhomogeneously distributed within promoters (49). Another, more specific, possibility to use positional information is to weight the predicted transcriptional effect of class II activators by the helical face at which they are positioned relative to the core promoter elements (115). Furthermore, some classes of activators or repressors function only in a specific positional range relative to the transcription start site, so putative TFBSs for these TFs outside these regions could be discarded (although not if experimental information points to functionality). However, in the end, one would like to predict the steepness and control logic of the response of a promoter to the concentrations of numbers of active TFs in the cell. In order to accomplish this, a "grammar" should be constructed that can predict the promoter function from the positioning and combinations of multiple TFBSs in a promoter.

Synthetic biology is the field of research where biological building blocks conferring a certain functionality are identified by a combination of molecular biology, bioinformatics, and engineering. These building blocks can subsequently be transferred to a different organism to add biological functionality. Synthetic biology approaches seem both the solution and an additional application for this: a solution because synthetic approaches will allow the construction of large synthetic promoter libraries with which such a grammar can be constructed (58, 102, 153) and an application because such grammar definitions will allow the de novo design of synthetic promoters with any control logic of choice to function in a synthetic regulatory module (197). Once such a detailed grammar has been constructed, it can be employed to take into account combinatorial regulation in reconstructing TRNs by using and improving on software systems such as the newly developed RENCO (251). The possible role of homocooperative binding could be integrated by boosting TFBS motif scores if they are positioned next to TFBSs for the same TF, especially because it has been shown that weak-affinity TFBSs can function as strong-affinity TFBSs if they are positioned next to additional strong-affinity TFBSs (102). In cooperative binding at multiple TFBSs, but also when homocooperative binding takes place at a single TFBS (binding of a homodimer), it should be taken into account that at such TFBSs, regulation will probably be more sigmoid because of a larger concentration effect. This can be integrated into Boolean logic functions as previously demonstrated (57) (Fig. 5).

To be able to predict gene expression regulation more accurately, it is also vital that many less-well-studied regulation mechanisms are understood. One example of regulatory sequences that have been neglected is formed by the core promoter elements themselves as well as the associated UP element and the extended -10 element. How the affinity of RNAP binding to core promoters is determined by their sequence and how this affects the transcription rate of a downstream gene should be studied in more detail. This would then result in the possibility of giving genes specific coefficients that signify the strength of the core promoter without other regulatory interactions. Also, promoters that could be regulated by transcriptional interference (42, 268) should probably be assessed separately, as the mechanisms are quite complex and have not yet been studied in detail. A kind of "dominance factor" could be used to indicate the activity of one promoter at the expense of the activity of another. More global information such as the chromosomal positioning (1, 183) of genes is probably easy to integrate into in silico regulatory networks, as all genes can be given a constant that lowers the predicted level of expression of genes if they are closer to the terminus of replication. Similarly, the distance of putative TFBSs from the gene encoding their TF (corresponding to the search speed of the TF toward it) can be taken into account. Finally, non-TF sequence-specific DNA binding proteins such as Dam methyltransferases and some nucleoid proteins should be added to promoter and network analyses. Perhaps transcript cleavage factors such as GreA (133, 282) also have some sequence specificity, and this could be investigated further experimentally.

When attempting to validate in silico-predicted TRNs, it should at all times be noted that expression data actually do not represent gene transcription rates but represent merely mRNA abundance rates. More global assessments of mRNA stability such as that performed by Selinger and coworkers (264) and the more recent mRNA sequencing techniques by, e.g., Illumina (http://www.illumina.com) are probably adequate and necessary to quantify the role of selective mRNA degradation, which may be the cause of many inconsistencies between high-throughput expression data and computational predictions. Also, at the mRNA level, the role of riboswitches should not be underestimated (14, 195, 308, 319).

Recent bioinformatic applications are increasingly appreciating the biophysical reality of protein-DNA binding. For example, Manke et al. recently demonstrated that it is possible to accurately predict regulatory interactions using a continuous model of TFBS binding affinities instead of discrete descriptions of the absence or presence of TFBSs (196). Importantly, this model also takes into account the binding affinity of a TF for the background sequence. Also, nucleotide interdependencies have started to be modeled into motif discovery algorithms (222, 327, 330). In complex promoters where homocooperative regulation takes place, there is a good chance that fuzzy motifs that function specifically to bind TFs with weak affinity are not incorporated in in silico predictions because of a lack of statistical significance. The role of information content in the fuzziness of motifs can also be integrated into models, as the pleiotropy of a TF could be used to determine how distant from the degenerate consensus motif a TFBS sequence is allowed to be in order to attain an optimal balance between false-positive and false-negative results. However, the problem with using the degree of TF pleiotropy for this is that one attempts to use the output of a model (the regulon size) as input before actually obtaining this output. However, integration with experimental data and iterative modeling should be able to solve this.

On the experimental side, new technical possibilities are opening up as well. Protein binding microarrays are further complementing the ChIP-on-chip approach in identifying DNA sequences that bind to specific TFs. They also have potential in discovering the functional sequence space of TFBSs if, for example, many different degenerate versions of the consensus sequence are put on an array. Maybe even more promising are ChIP-Seq approaches (198, 245). Initial methodological tests reveal that especially a combination of traditional ChIP-on-chip and ChIP sequencing yields a more comprehensive list of functional TFBSs throughout genomes (85). Also, combining high-resolution ChIP-on-chip or ChIP-Seq data with gene expression data appears to be promising for the network reconstruction of specific regulons (51). The integration of transcriptomics and metabolomics will finally also reveal more insights into the role of small-molecule concentrations (for example, as small TF binding ligands or in riboswitch regulation) in regulating gene transcription on a global scale (202), which is especially important when integrating experimental data for organisms grown in different media.

Conceivably, it will not be possible to integrate all biological mechanisms mentioned in this review into computational methods for regulatory network reconstruction. Due to the complexity of the matter, early attempts to integrate these mechanisms in general models will probably yield results that are not more accurate than the results of highly optimized simplistic models. Of course, in order to design methodologies that we can use within a reasonable time, we need to get closer to the complex biological reality without overfitting the data on too-complex and noisy models that involve too many parameters that fail to be informative (72, 150). Only a first level of biological complexity is handled by current computational approaches. New models based on the features that are described here may significantly enhance our grip on the TRNs as they actually are. Such modeling will point out (i) which features do and which do not contribute to the successful separation of genes as being part of certain regulons or not and (ii) which genes cannot be correctly classified based on the current features and thus contain features or "biology" missing in the model. As Sandve and colleagues recently mentioned, "another mathematical reformulation of existing approaches will certainly not change the status of the field" (257). However, if the integration of biological mechanisms into computational models goes hand-in-hand with advances in algorithm development and the increasing use of high-throughput experimental data to validate network reconstructions, significant advances in grasping the regulatory complexity residing inside bacterial cells can surely be expected.

ACKNOWLEDGMENTS

Work of S.A.F.T.V.H. was in part supported by the BMBF (grant number 0313978A) within the framework of the transnational SysMO initiative in the project BaCell-SysMO.

We thank Christof Francke for critical reading of the manuscript and Siger Holsappel for graphically enhancing Fig. 2. We thank the anonymous reviewers for their constructive comments.

REFERENCES

1. **Abby, S., and V. Daubin.** 2007. Comparative genomics and the evolution of prokaryotes. Trends Microbiol. **15:**135–141.
2. **Aki, T., H. E. Choy, and S. Adhya.** 1996. Histone-like protein HU as a specific transcriptional regulator: co-factor role in repression of gal transcription by GAL repressor. Genes Cells **1:**179–188.
3. **Alon, U.** 2007. Network motifs: theory and experimental approaches. Nat. Rev. Genet. **8:**450–461.
4. **Andersson, S. G., and C. G. Kurland.** 1990. Codon preferences in free-living microorganisms. Microbiol. Rev. **54:**198–210.

4a. **Angarica, V. E., A. G. Pérez, A. T. Vasconcelos, J. Collado-Vides, and B. Contreras-Moreira.** 2008. Prediction of TF target sites based on atomistic models of protein-DNA complexes. BMC Bioinformatics **9:**436.

5. **Babu, M. M., B. Lang, and L. Aravind.** 2009. Methods to reconstruct and compare transcriptional regulatory networks. Methods Mol. Biol. **541:**1–18.
6. **Babu, M. M., N. M. Luscombe, L. Aravind, M. B. Gerstein, and S. A. Teichmann.** 2004. Structure and evolution of transcriptional regulatory networks. Curr. Opin. Struct. Biol. **14:**283–291.
7. **Babu, M. M., S. A. Teichmann, and L. Aravind.** 2006. Evolutionary dynamics of prokaryotic transcriptional regulatory networks. J. Mol. Biol. **358:** 614–633.
8. **Baerends, R. J. S., W. K. Smits, A. de Jong, L. W. Hamoen, J. Kok, and O. P. Kuipers.** 2004. Genome2D: a visualization tool for the rapid analysis of bacterial transcriptome data. Genome Biol. **5:**R37.
9. **Bai, Y., Q. Ge, J. Wang, T. Li, Q. Liu, and Z. Lu.** 2005. Investigation of DNA-protein sequence-specific interactions with a ds-DNA array. Molecules **10:**417–426.
10. **Bailey, T. L., N. Williams, C. Misleh, and W. W. Li.** 2006. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res. **34:**W369–W373.
11. **Balleza, E., L. N. Lopez-Bojorquez, A. Martinez-Antonio, O. Resendis-Antonio, I. Lozada-Chavez, Y. I. Balderas-Martinez, S. Encarnacion, and J. Collado-Vides.** 2009. Regulation by transcription factors in bacteria: beyond description. FEMS Microbiol. Rev. **33:**133–151.
12. **Barnard, A., A. Wolfe, and S. J. W. Busby.** 2004. Regulation at complex bacterial promoters: how bacteria use different promoter organizations to produce different regulatory outcomes. Curr. Opin. Microbiol. **7:**102–108.
13. **Barrett, C. L., and B. O. Palsson.** 2006. Iterative reconstruction of transcriptional regulatory networks: an algorithmic approach. PLoS Comput. Biol. **2:**e52.
14. **Barrick, J. E., K. A. Corbino, W. C. Winkler, A. Nahvi, M. Mandal, J. Collins, M. Lee, A. Roth, N. Sudarsan, I. Jona, J. K. Wickiser, and R. R. Breaker.** 2004. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. Proc. Natl. Acad. Sci. USA **101:**6421–6426.
15. **Baumbach, J.** 2007. CoryneRegNet 4.0—a reference database for corynebacterial gene regulatory networks. BMC Bioinformatics **8:**429.
16. **Baumbach, J., S. Rahmann, and A. Tauch.** 2009. Reliable transfer of transcriptional gene regulatory networks between taxonomically related organisms. BMC Syst. Biol. **3:**8.
17. **Baumbach, J., A. Tauch, and S. Rahmann.** 2009. Towards the integrated analysis, visualization and reconstruction of microbial gene regulatory networks. Brief. Bioinform. **10:**75–83.
18. **Baumbach, J., T. Wittkop, C. K. Kleindt, and A. Tauch.** 2009. Integrated analysis and reconstruction of microbial transcriptional gene regulatory networks using CoryneRegNet. Nat. Protoc. **4:**992–1005.
19. **Beer, M. A., and S. Tavazoie.** 2004. Predicting gene expression from sequence. Cell **117:**185–198.
20. **Benichou, O., C. Loverdo, and R. Voituriez.** 2008. How gene colocalization can be optimized by tuning the diffusion constant of transcription factors. Europhys. Lett. **84:**38003.
21. **Benos, P. V., M. L. Bulyk, and G. D. Stormo.** 2002. Additivity in protein-DNA interactions: how good an approximation is it? Nucleic Acids Res. **30:**4442–4451.
22. **Berg, J., S. Willmann, and M. Lässig.** 2004. Adaptive evolution of transcription factor binding sites. BMC Evol. Biol. **4:**42.
23. **Berger, M. F., and M. L. Bulyk.** 2009. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nat. Protoc. **4:**393–411.
24. **Berger, M. F., A. A. Philippakis, A. M. Qureshi, F. S. He, P. W. Estep III, and M. L. Bulyk.** 2006. Compact, universal DNA microarrays to comprehensively determine transcription-factor binding site specificities. Nat. Biotechnol. **24:**1429–1435.
25. **Bi, C., and P. K. Rogan.** 2006. BIPAD: a Web server for modeling bipartite sequence elements. BMC Bioinformatics **7:**76.
26. **Bilu, Y., and N. Barkai.** 2005. The design of transcription-factor binding sites is affected by combinatorial regulation. Genome Biol. **6:**R103.
27. **Bintu, L., N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, T. Kuhlman, and R. Phillips.** 2005. Transcriptional regulation by the numbers: applications. Curr. Opin. Genet. Dev. **15:**125–135.
28. **Bintu, L., N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, and R. Phillips.** 2005. Transcriptional regulation by the numbers: models. Curr. Opin. Genet. Dev. **15:**116–124.
29. **Blanchette, M., and M. Tompa.** 2003. FootPrinter: a program designed for phylogenetic footprinting. Nucleic Acids Res. **31:**3840–3842.
30. **Blanchette, M., and M. Tompa.** 2002. Discovery of regulatory elements by a computational method for phylogenetic footprinting. Genome Res. **12:** 739–748.
31. **Blom, E. J., J. B. Roerdink, O. P. Kuipers, and S. A. F. T. van Hijum.** 2009. MOTIFATOR: detection and characterization of regulatory motifs using prokaryote transcriptome data. Bioinformatics **25:**550–551.
32. **Bon, M., S. J. McGowan, and P. R. Cook.** 2006. Many expressed genes in bacteria and yeast are transcribed only once per cell cycle. FASEB J. **20:**1721–1723.
33. **Bornholdt, S.** 2008. Boolean network models of cellular regulation: prospects and limitations. J. R. Soc. Interface **5**(Suppl. 1)**:**S85–S94.
34. **Brodolin, K., N. Zenkin, A. Mustaev, D. Mamaeda, and H. Heumann.** 2004. The sigma 70 subunit of RNA polymerase induces *lacUV5* promoter-proximal pausing of transcription. Nat. Struct. Mol. Biol. **11:**551–557.
35. **Brouwer, R. W. W., O. P. Kuipers, and S. A. F. T. van Hijum.** 2008. The relative value of operon predictions. Brief. Bioinform. **9:**367–375.
36. **Browning, D. F., and S. J. W. Busby.** 2004. The regulation of bacterial transcription initiation. Nat. Rev. Microbiol. **2:**57–65.
37. **Browning, D. F., J. A. Cole, and S. J. Busby.** 2008. Regulation by nucleoid-associated proteins at the *Escherichia coli nir* operon promoter. J. Bacteriol. **190:**7258–7267.
38. **Buchler, N. E., U. Gerland, and T. Hwa.** 2003. On schemes of combinatorial transcription logic. Proc. Natl. Acad. Sci. USA **100:**5136–5141.
39. **Bulyk, M. L.** 2006. Analysis of sequence specificities of DNA-binding proteins with protein binding microarrays. Methods Enzymol. **410:**279–299.
40. **Bulyk, M. L., P. L. F. Johnson, and G. M. Church.** 2002. Nucleotides of transcription factor binding sites exert interdependent effects on the binding affinities of transcription factors. Nucleic Acids Res. **30:**1255–1261.
41. **Bussemaker, H. J., H. Li, and E. D. Siggia.** 2000. Building a dictionary for genomes: identification of presumptive regulatory sites by statistical analysis. Proc. Natl. Acad. Sci. USA **97:**10096–10100.
42. **Callen, B. P., K. E. Shearwin, and J. B. Egan.** 2004. Transcriptional interference between convergent promoters caused by elongation over the promoter. Mol. Cell **14:**647–656.
43. **Cardon, L. R., and G. D. Stormo.** 1992. Expectation maximization algorithm for identifying protein-binding sites with variable lengths from unaligned DNA fragments. J. Mol. Biol. **223:**159–170.
44. **Carlson, J. M., A. Chakravarty, C. E. DeZiel, and R. H. Gross.** 2007. SCOPE: a Web server for practical de novo motif discovery. Nucleic Acids Res. **35:**W259–W264.
45. **Carlson, J. M., A. Chakravarty, and R. H. Gross.** 2006. BEAM: a beam search algorithm for the identification of cis-regulatory elements in groups of genes. J. Comput. Biol. **13:**686–701.
46. **Carlson, J. M., A. Chakravarty, R. S. Khetani, and R. H. Gross.** 2006. Bounded search for de novo identification of degenerate *cis*-regulatory elements. BMC Bioinformatics **7:**254.
47. **Carrera, J., G. Rodrigo, and A. Jaramillo.** 2009. Model-based redesign of global transcription regulation. Nucleic Acids Res. **37:**e38.
48. **Casadesus, J., and D. Low.** 2006. Epigenetic gene regulation in the bacterial world. Microbiol. Mol. Biol. Rev. **70:**830–856.
49. **Casimiro, A. C., S. Vinga, A. T. Freitas, and A. L. Oliveira.** 2008. An analysis of the positional distribution of DNA motifs in promoter regions and its biological relevance. BMC Bioinformatics **9:**89.
50. **Chakravarty, A., J. M. Carlson, R. S. Khetani, C. E. DeZiel, and R. H. Gross.** 2007. SPACER: identification of cis-regulatory elements with non-contiguous critical residues. Bioinformatics **23:**1029–1031.
51. **Cho, B. K., C. L. Barrett, E. M. Knight, Y. S. Park, and B. O. Palsson.** 2008. Genome-scale reconstruction of the Lrp regulatory network in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **105:**19462–19467.
52. **Choo, Y., and A. Klug.** 1997. Physical basis of a protein-DNA recognition code. Curr. Opin. Struct. Biol. **7:**117–125.
53. **Ciampi, M. S.** 2006. Rho-dependent terminators and transcription termination. Microbiology **152:**2515–2528.
54. **Clements, M., E. P. van Someren, T. A. Knijnenburg, and M. J. Reinders.** 2007. Integration of known transcription factor binding site information and gene expression data to advance from co-expression to co-regulation. Genomics Proteomics Bioinform. **5:**86–101.
55. **Collado-Vides, J., B. Magasanik, and J. D. Gralla.** 1991. Control site loca-

tion and transcriptional regulation in *Escherichia coli*. Microbiol. Mol. Biol. Rev. **55:**371–394.

56. **Cordero, F., M. Botta, and R. A. Calogero.** 2007. Microarray data analysis and mining approaches. Brief. Funct. Genomics Proteomics **6:**265–281.
57. **Covert, M. W., E. M. Knight, J. L. Reed, M. J. Herrgard, and B. O. Palsson.** 2004. Integrating high-throughput and computational data elucidates bacterial networks. Nature **429:**92–96.
58. **Cox, R. S., III, M. G. Surette, and M. B. Elowitz.** 2007. Programming gene expression with combinatorial promoters. Mol. Syst. Biol. **3:**145.
59. **Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner.** 2004. WebLogo: a sequence logo generator. Genome Res. **14:**1188–1190.
60. **Cuticchia, A. J., R. Ivarie, and J. Arnold.** 1992. The application of Markov chain analysis to oligonucleotide frequency prediction and physical mapping of *Drosophila melanogaster*. Nucleic Acids Res. **20:**3651–3657.
61. **Das, M. K., and H. K. Dai.** 2007. A survey of DNA motif finding algorithms. BMC Bioinformatics **8**(Suppl. 7)**:**S21.
62. **Davidich, M. I., and S. Bornholdt.** 2008. Boolean network model predicts cell cycle sequence of fission yeast. PLoS ONE **3:**e1672.
63. **Day, W. H., and F. R. McMorris.** 1992. Critical comparison of consensus methods for molecular sequences. Nucleic Acids Res. **20:**1093–1099.
64. **de Bleser, P., B. Hooghe, D. Vlieghe, and F. van Roy.** 2007. A distance difference matrix approach to identifying transcription factors that regulate differential gene expression. Genome Biol. **8:**R83.
65. **Defrance, M., and H. Touzet.** 2006. Predicting transcription factor binding sites using local over-representation and comparative genomics. BMC Bioinformatics **7:**396.
66. **Dekhtyar, M., A. Morin, and V. Sakanyan.** 2008. Triad pattern algorithm for predicting strong promoter candidates in bacterial genomes. BMC Bioinformatics **9:**233.
67. **Dempster, A., N. Laird, and D. Rubin.** 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B **39:**1–38.
68. **den Hengst, C. D., S. A. F. T. van Hijum, J. M. Geurts, A. Nauta, J. Kok, and O. P. Kuipers.** 2005. The *Lactococcus lactis* CodY regulon: identification of a conserved *cis*-regulatory element. J. Biol. Chem. **280:**34332–34342.
69. **Dent, C. L., and D. S. Latchman.** 1993. The DNA mobility shift assay, p. 1–26. *In* D. S. Latchman (ed.), Transcription factors: a practical approach. Oxford University Press, Oxford, United Kingdom.
70. **Deutscher, J., C. Francke, and P. W. Postma.** 2006. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. Microbiol. Mol. Biol. Rev. **70:**939–1031.
71. **D'haeseleer, P.** 2006. How does DNA sequence motif discovery work? Nat. Biotechnol. **24:**959–961.
72. **Dietterich, T.** 1995. Overfitting and undercomputing in machine learning. ACM Comp. Surv. **27:**326–327.
73. **Djordjevic, M.** 2007. SELEX experiments: new prospects, applications and data analysis in inferring regulatory pathways. Biomol. Eng. **24:**179–189.
74. **Doniger, S. W., and J. C. Fay.** 2007. Frequent gain and loss of functional transcription factor binding sites. PLoS Comput. Biol. **3:**932–942.
75. **Dufva, M.** 2009. Introduction to microarray technology. Methods Mol. Biol. **529:**1–22.
76. **Durrett, R., and D. Schmidt.** 2007. Waiting for regulatory sequences to appear. Ann. Appl. Prob. **17:**1–32.
77. **Edwards, J. S., M. Covert, and B. Palsson.** 2002. Metabolic modelling of microbes: the flux-balance approach. Environ. Microbiol. **4:**133–140.
78. **Elati, M., P. Neuvial, M. Bolotin-Fukuhara, E. Barillot, F. Radvanyi, and C. Rouveirol.** 2007. LICORN: learning cooperative regulation networks from gene expression data. Bioinformatics **23:**2407–2414.
79. **Elemento, O., and S. Tavazoie.** 2005. Fast and systematic genome-wide discovery of conserved regulatory elements using a non-alignment based approach. Genome Biol. **6:**R18.
80. **Escolar, L., J. Perez-Martin, and V. de Lorenzo.** 2000. Evidence for an unusually long operator for the Fur repressor in the aerobactin promoter of *Escherichia coli*. J. Biol. Chem. **275:**24709–24714.
81. **Eskin, E., and P. A. Pevzner.** 2002. Finding composite regulatory patterns in DNA sequences. Bioinformatics **18**(Suppl. 1)**:**S354–S363.
82. Reference deleted.
83. **Espinosa, V., A. D. Gonzalez, A. T. Vasconcelos, A. M. Huerta, and J. Collado-Vides.** 2005. Comparative studies of transcriptional regulation mechanisms in a group of eight gamma-proteobacterial genomes. J. Mol. Biol. **354:**184–199.
84. **Estrem, S. T., W. Ross, T. Gaal, Z. W. S. Chen, W. Niu, R. H. Ebright, and R. L. Gourse.** 1999. Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase alpha subunit. Genes Dev. **13:**2134–2147.
85. **Euskirchen, G. M., J. S. Rozowsky, C. L. Wei, W. H. Lee, Z. D. Zhang, S. Hartman, O. Emanuelsson, V. Stolc, S. Weissman, M. B. Gerstein, Y. Ruan, and M. Snyder.** 2007. Mapping of transcription factor binding regions in mammalian cells by ChIP: comparison of array- and sequencing-based technologies. Genome Res. **17:**898–909.
86. **Fabret, C., V. A. Feher, and J. A. Hoch.** 1999. Two-component signal transduction in *Bacillus subtilis*: how one organism sees its world. J. Bacteriol. **181:**1975–1983.
87. **Faith, J. J., B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. J. Collins, and T. S. Gardner.** 2007. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. PLoS Biol. **5:**e8.
88. **Fields, D. S., Y. He, A. Al-Uzri, and G. D. Stormo.** 1997. Quantitative specificity of the *mnt* repression. J. Mol. Biol. **271:**194.
89. **Francke, C., R. Kerkhoven, M. Wels, and R. J. Siezen.** 2008. A generic approach to identify transcription factor-specific operator motifs; inferences for LacI-family mediated regulation in *Lactobacillus plantarum* WCFS1. BMC Genomics **9:**145.
90. **Friedman, N.** 2004. Inferring cellular networks using probabilistic graphical models. Science **303:**799–805.
91. **Friedman, N., M. Linial, I. Nachman, and D. Pe'er.** 2000. Using Bayesian networks to analyze expression data. J. Comput. Biol. **7:**601–620.
92. **Froula, J. L., and M. P. Francino.** 2007. Selection against spurious promoter motifs correlates with translational efficiency across Bacteria. PLoS ONE **8:**e745.
93. **Galas, D. J., and A. Schmitz.** 1978. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res. **5:**3157–3170.
94. **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. Breu-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–D124.
95. **Gaston, K., A. Bell, A. Kolb, H. Buc, and S. J. W. Busby.** 1990. Stringent spacing requirements for transcription activation by CRP. Cell **62:**733–743.
96. **Gat-Viks, I., A. Tanay, D. Raijman, and R. Shamir.** 2006. A probabilistic methodology for integrating knowledge and experiments on biological networks. J. Comput. Biol. **13:**165–181.
97. **Geier, F., J. Timmer, and C. Fleck.** 2007. Reconstructing gene-regulatory networks from time series, knock-out data, and prior knowledge. BMC Syst. Biol. **1:**11.
98. **Gelfand, A. E., and A. F. M. Smith.** 1990. Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. **85:**398–409.
99. **Gelfand, M. S.** 2006. Evolution of transcriptional regulatory networks in microbial genomes. Curr. Opin. Struct. Biol. **16:**420–429.
100. **Gerland, U., and T. Hwa.** 2002. On the selection and evolution of regulatory DNA motifs. J. Mol. Evol. **55:**386–400.
101. **Gertz, J., and B. A. Cohen.** 2009. Environment-specific combinatorial cis-regulation in synthetic promoters. Mol. Syst. Biol. **5:**244.
102. **Gertz, J., E. D. Siggia, and B. A. Cohen.** 2009. Analysis of combinatorial cis-regulation in synthetic and genomic promoters. Nature **457:**215–218.
103. **Gillespie, D. T.** 2007. Stochastic simulation of chemical kinetics. Annu. Rev. Phys. Chem. **58:**35–55.
104. **Gjuvsland, A. B., E. Plahte, and S. W. Omholt.** 2007. Threshold-dominated regulation hides genetic variation in gene expression networks. BMC Syst. Biol. **1:**57.
105. **Gollnick, P., and P. Babitzke.** 2002. Transcription attenuation. Biochim. Biophys. Acta **1577:**240–250.
106. **Gordan, R., L. Narlikar, and A. J. Hartemink.** 2008. A fast, alignment-free, conservation-based method for transcription factor binding site discovery. Lect. Notes Comp. Sci. **4955:**98–111.
107. **Gordon, J. J., M. W. Towsey, J. M. Hogan, S. A. Mathews, and P. Timms.** 2006. Improved prediction of bacterial transcription start sites. Bioinformatics **22:**142–148.
108. **Gottesman, S.** 2005. Micros for microbes: non-coding regulatory RNAs in bacteria. Trends Genet. **21:**399–404.
109. **Gottesman, S., G. Storz, C. Rosenow, N. Majdalani, F. Repoila, and K. M. Wassarman.** 2001. Small RNA regulators of translation: mechanisms of action and approaches for identifying new small RNAs. Cold Spring Harb. Symp. Quant. Biol. **66:**353–362.
110. **Griffith, K. L., I. M. Shah, T. E. Myers, M. C. O'Neill, and R. E. Wolf.** 2002. Evidence for 'pre-recruitment' as a new mechanism of transcription activation in *Escherichia coli*: the large excess of *soxS* binding sites per cell relative to the number of *soxS* molecular per cell. Biochem. Biophys. Res. Commun. **291:**979–986.
111. **Gruber, T. M., and C. A. Gross.** 2003. Multiple sigma subunits and the partitioning of bacterial transcription space. Annu. Rev. Microbiol. **57:**441–466.
112. **GuhaThakurta, D.** 2006. Computational identification of transcriptional regulatory elements in DNA sequence. Nucleic Acids Res. **34:**3585–3598.
113. **Hahn, M. W.** 2007. Detecting natural selection on *cis*-regulatory DNA. Genetica **129:**7–18.
114. **Hahn, M. W., J. E. Stajich, and G. A. Wray.** 2003. The effects of selection against spurious transcription factor binding sites. Mol. Biol. Evol. **20:**901–906.
115. **Hamoen, L. W., W. K. Smits, A. de Jong, S. Holsappel, and O. P. Kuipers.** 2002. Improving the predictive value of the competence transcription factor

(ComK) binding site in *Bacillus subtilis* using a genomic approach. Nucleic Acids Res. **30:**5517–5528.

116. **Hamoen, L. W., A. F. van Werkhoven, J. J. Bijlsma, D. Dubnau, and G. Venema.** 1998. The competence transcription factor of *Bacillus subtilis* recognizes short A/T-rich sequences arranged in a unique, flexible pattern along the DNA helix. Genes Dev. **12:**1539–1550.
117. **Hamoen, L. W., A. F. van Werkhoven, G. Venema, and D. Dubnau.** 2000. The pleiotropic response regulator DegU functions as a priming protein in competence development in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA **97:**9246–9251.
118. **Hantke, K.** 2001. Iron and metal regulation in bacteria. Curr. Opin. Microbiol. **4:**172–177.
119. **Harbison, C. T., D. B. Gordon, T. I. Lee, N. J. Rinaldi, K. D. MacIsaac, T. W. Danford, N. M. Hannett, J. B. Tagne, D. B. Reynolds, J. Yoo, E. G. Jennings, J. Zeitlinger, D. K. Pokholok, M. Kellis, P. A. Rolfe, K. T. Takusagawa, E. S. Lander, D. K. Gifford, E. Fraenkel, and R. A. Young.** 2004. Transcriptional regulatory code of a eukaryotic genome. Nature **431:** 99–104.
120. **Hecker, M., and U. Volker.** 2004. Towards a comprehensive understanding of *Bacillus subtilis* cell physiology by physiological proteomics. Proteomics **4:**3727–3750.
121. **Helmann, J. D.** 2002. The extracytoplasmic function (ECF) sigma factors. Adv. Microb. Physiol. **46:**47–110.
122. **Hendriksen, W. T., N. Silva, H. J. Bootsma, C. E. Blue, G. K. Paterson, A. R. Kerr, A. de Jong, O. P. Kuipers, P. W. Hermans, and T. J. Mitchell.** 2007. Regulation of gene expression in *Streptococcus pneumoniae* by response regulator 09 is strain dependent. J. Bacteriol. **189:**1382–1389.
123. **Henkin, T. M.** 2000. Transcription termination control in bacteria. Curr. Opin. Microbiol. **3:**149–153.
124. **Hermsen, R., S. Tans, and P. R. ten Wolde.** 2006. Transcriptional regulation by competing transcription factor modules. PLoS Comput. Biol. **2:**1552–1560.
125. **Herrgard, M. J., M. W. Covert, and B. O. Palsson.** 2004. Reconstruction of microbial transcriptional regulatory networks. Curr. Opin. Biotechnol. **15:** 70–77.
126. **Herrgard, M. J., M. W. Covert, and B. O. Palsson.** 2003. Reconciling gene expression data with known genome-scale regulatory network structures. Genome Res. **13:**2423–2434.
127. **Herring, C. D., M. Raffaelle, T. E. Allen, E. I. Kanin, R. Landick, A. Z. Ansari, and B. O. Palsson.** 2005. Immobilization of *Escherichia coli* RNA polymerase and location of binding sites by use of chromatin immunoprecipitation and microarrays. J. Bacteriol. **187:**6166–6174.
128. **Hershberg, R., and H. Margalit.** 2006. Co-evolution of transcription factors and their targets depends on mode of regulation. Genome Biol. **7:**R62.
129. **Hochschild, A.** 2007. Gene-specific regulation by a transcript cleavage factor: facilitating promoter escape. J. Bacteriol. **189:**8769–8771.
130. **Hong, J.** 1989. Prediction of oligonucleotide frequencies based upon dinucleotide frequencies obtained from the nearest neighbor analysis. Nucleic Acids Res. **18:**1625–1628.
131. **Hooghe, B., P. Hulpiau, F. van Roy, and P. de Bleser.** 2008. ConTra: a promoter alignment analysis tool for identification of transcription factor binding sites across species. Nucleic Acids Res. **36:**W128–W132.
132. **Hsu, L. M.** 2002. Promoter clearance and escape in prokaryotes. Biochim. Biophys. Acta **1577:**191–207.
133. **Hsu, L. M., N. V. Vo, and M. J. Chamberlin.** 1995. *Escherichia coli* transcript cleavage factors GreA and GreB stimulate promoter escape and gene expression in vivo and in vitro. Proc. Natl. Acad. Sci. USA **92:**11588–11592.
134. **Hu, J., B. Li, and D. Kihara.** 2005. Limitations and potentials of current motif discovery algorithms. Nucleic Acids Res. **33:**4899–4913.
135. **Hu, J., Y. D. Yang, and D. Kihara.** 2006. EMD: an ensemble algorithm for discovering regulatory motifs in DNA sequences. BMC Bioinformatics **7:**342.
136. **Huang, W., J. R. Nevins, and U. Ohler.** 2007. Phylogenetic simulation of promoter evolution: estimation and modeling of binding site turnover events and assessment of their impact on alignment tools. Genome Biol. **8:**R225.
137. **Huerta, A. M., and J. Collado-Vides.** 2003. Sigma70 promoters in *Escherichia coli*: specific transcription in dense regions of overlapping promoter-like signals. J. Mol. Biol. **333:**261–278.
138. **Huerta, A. M., J. Collado-Vides, and M. P. Francino.** 2006. Selection for unequal densities of sigma70 promoter-like signals in different regions of large bacterial genomes. PLoS Genet. **2:**e185.
139. **Hughes, K. T., and K. Mathee.** 1998. The anti-sigma factors. Annu. Rev. Microbiol. **52:**231–286.
140. Reference deleted.
141. **Janga, S. C., and J. Collado-Vides.** 2007. Structure and evolution of gene regulatory networks in microbial genomes. Res. Microbiol. **158:**787–794.
142. **Janga, S. C., H. Salgado, J. Collado-Vides, and A. Martinez-Antonio.** 2007. Internal versus external effector and transcription factor gene pairs differ in their relative chromosomal position in Escherichia coli. J. Mol. Biol. **368:** 263–272.
143. **Janga, S. C., H. Salgado, and A. Martinez-Antonio.** 2009. Transcriptional regulation shapes the organization of genes on bacterial chromosomes. Nucleic Acids Res. **37:**3680–3688.
144. **Janky, R., and J. van Helden.** 2008. Evaluation of phylogenetic footprint discovery for predicting bacterial *cis*-regulatory elements and revealing their evolution. BMC Bioinformatics **9:**37.
145. **Johnson, D. S., A. Mortazavi, R. M. Myers, and B. Wold.** 2007. Genome-wide mapping of in vivo protein-DNA interactions. Science **316:**1497–1502.
146. **Jothi, R., S. Cuddapah, A. Barski, K. Cui, and K. Zhao.** 2008. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res. **36:**5221–5231.
147. **Kaplan, S., A. Bren, E. Dekel, and U. Alon.** 2008. The incoherent feed-forward loop can generate non-monotonic input functions for genes. Mol. Syst. Biol. **4:**203.
148. **Kaplan, S., A. Bren, A. Zaslaver, E. Dekel, and U. Alon.** 2008. Diverse two-dimensional input functions control bacterial sugar genes. Mol. Cell **29:**786–792.
149. **Kaplan, T., N. Friedman, and H. Margalit.** 2005. Ab initio prediction of transcription factor targets using structural knowledge. PLoS Comput. Biol. **1:**e1.
150. **Karlebach, G., and R. Shamir.** 2008. Modelling and analysis of gene regulatory networks. Nat. Rev. Mol. Cell Biol. **9:**770–780.
151. **Kazmierczak, M. J., M. Wiedmann, and K. J. Boor.** 2005. Alternative sigma factors and their roles in bacterial virulence. Microbiol. Mol. Biol. Rev. **69:**527–543.
152. **Kim, J. T., T. Martinetz, and D. Polani.** 2003. Bioinformatic principles underlying the information content of transcription factor binding sites. J. Theor. Biol. **220:**529–544.
153. **Kinkhabwala, A., and C. C. Guet.** 2008. Uncovering cis regulatory codes using synthetic promoter shuffling. PLoS ONE **3:**e2030.
154. **Kiryu, H., T. Oshima, and K. Asai.** 2005. Extracting relations between promoter sequences and their strengths from microarray data. Bioinformatics **21:**1062–1068.
155. **Klamt, S., J. Saez-Rodriguez, and E. D. Gilles.** 2007. Structural and functional analysis of cellular networks with CellNetAnalyzer. BMC Syst. Biol. **1:**2.
156. **Kloeckener-Gruissem, B., and M. Freeling.** 1995. Transposon-induced promoter scrambling: a mechanism for the evolution of new alleles. Proc. Natl. Acad. Sci. USA **92:**1836–1840.
157. **Kobayashi, M., K. Nagata, and A. Ishihama.** 1990. Promoter selectivity of Escherichia coli RNA polymerase: effect of base substitutions in the promoter −35 region on promoter strength. Nucleic Acids Res. **18:**7367–7372.
158. **Kolesov, G., Z. Wunderlich, O. N. Laikova, M. S. Gelfand, and L. A. Mirny.** 2007. How gene order is influenced by the biophysics of transcription regulation. Proc. Natl. Acad. Sci. USA **104:**13948–13953.
159. **Kono, H., and A. Sarai.** 1999. Structure-based prediction of DNA target sites by regulatory proteins. Proteins **35:**114–131.
160. **Kuhlman, T., Z. G. Zhang, M. H. Saier, and T. Hwa.** 2007. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **104:**6043–6048.
161. **Lahdesmaki, H., S. Hautaniemi, I. Shmulevich, and O. Yli-Harja.** 2006. Relationships between probabilistic Boolean networks and dynamic Bayesian networks as models of gene regulatory networks. Signal Processing **86:**814–834.
162. **Lahdesmaki, H., A. G. Rust, and I. Shmulevich.** 2008. Probabilistic inference of transcription factor binding from multiple data sources. PLoS ONE **3:**e1820.
163. **Lähdesmäki, H., I. Shmulevich, and O. Yli-Harja.** 2003. On learning gene regulatory networks under the Boolean network model. Machine Learning **52:**147–167.
164. Reference deleted.
165. **Lamblin, A., and J. Fuchs.** 1994. Functional analysis of the *Escherichia coli* K-12 *cyn* operon transcriptional regulation. J. Bacteriol. **176:**6613–6622.
166. **Langdon, R. C., and A. Hochschild.** 1999. A genetic method for dissecting the mechanism of transcriptional activator synergy by identical activators. Proc. Natl. Acad. Sci. USA **96:**12673–12678.
167. **Lanzer, M., and H. Bujard.** 1988. Promoters largely determine the efficiency of repressor action. Proc. Natl. Acad. Sci. USA **85:**8973–8977.
168. **Larsen, R., J. Kok, and O. P. Kuipers.** 2005. Interaction between ArgR and AhrC controls regulation of arginine metabolism in *Lactococcus lactis*. J. Biol. Chem. **280:**19319–19330.
169. **Lässig, M.** 2007. From biophysics to evolutionary genetics: statistical aspects of gene regulation. BMC Bioinformatics **8**(Suppl. 6)**:**S7.
170. **Laub, M. T., S. L. Chen, L. Shapiro, and H. H. McAdams.** 2002. Genes directly controlled by CtrA, a master regulator of the Caulobacter cell cycle. Proc. Natl. Acad. Sci. USA **99:**4632–4637.
171. **Lawrence, C. E., S. F. Altschul, M. S. Boguski, J. S. Liu, A. F. Neuwald, and J. C. Wootton.** 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. Science **262:**208–214.
172. **Lawrence, C. E., and A. A. Reilly.** 1990. An expectation maximization (EM) algorithm for the identification and characterization of common sites in unaligned biopolymer sequences. Proteins **7:**41–51.

173. **Leblanc, B., and T. Moss.** 2001. DNase I footprinting. Methods Mol. Biol. **148:**31–38.

174. **Li, Q. M., and S. A. Johnston.** 2001. Are all DNA binding and transcription regulation by an activator physiologically relevant? Mol. Cell. Biol. **21:** 2467–2474.

175. **Li, W. H., J. Yang, and X. Gu.** 2005. Expression divergence between duplicate genes. Trends Genet. **21:**602–607.

176. **Lintner, R. E., P. K. Mishra, P. Srivastava, B. M. Martinez-Vaz, A. B. Khodursky, and R. M. Blumenthal.** 2008. Limited functional conservation of a global regulator among related bacterial genera: Lrp in *Escherichia*, *Proteus* and *Vibrio*. BMC Microbiol. **8:**60.

177. **Liu, J., and G. D. Stormo.** 2005. Combining SELEX with quantitative assays to rapidly obtain accurate models of protein-DNA interactions. Nucleic Acids Res. **33:**e141.

178. **Liu, L. A., and J. S. Bader.** 2009. Structure-based ab initio prediction of transcription factor-binding sites. Methods Mol. Biol. **541:**1–19.

179. **Liu, X., D. L. Brutlag, and J. S. Liu.** 2001. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. Pac. Symp. Biocomput. 127–138.

180. **Liu, Y., L. Wei, S. Batzoglou, D. L. Brutlag, J. S. Liu, and X. S. Liu.** 2004. A suite of Web-based programs to search for transcriptional regulatory motifs. Nucleic Acids Res. **32:**W204–W207.

181. **Lloyd, G., P. Landini, and S. Busby.** 2001. Activation and repression of transcription initiation in bacteria. Essays Biochem. **37:**17–31.

182. **Lobner-Olesen, A., O. Skovgaard, and M. G. Marinus.** 2005. Dam methylation: coordinating cellular processes. Curr. Opin. Microbiol. **8:**154–160.

183. **Lobry, J. R., and J. M. Louarn.** 2003. Polarisation of prokaryotic chromosomes. Curr. Opin. Microbiol. **6:**101–108.

184. **Low, D. A., N. J. Weyand, and M. J. Mahan.** 2001. Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. Infect. Immun. **69:**7197–7204.

185. **Lozada-Chavez, I., V. E. Angarica, J. Collado-Vides, and B. Contreras-Moreira.** 2008. The role of DNA-binding specificity in the evolution of bacterial regulatory networks. J. Mol. Biol. **379:**627–643.

186. **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.

187. **Ludwig, M. Z.** 2002. Functional evolution of noncoding DNA. Curr. Opin. Genet. Dev. **12:**634–639.

188. **Lulko, A. T., G. Buist, J. Kok, and O. P. Kuipers.** 2007. Transcriptome analysis of temporal regulation of carbon metabolism by CcpA in *Bacillus subtilis* reveals additional target genes. J. Mol. Microbiol. Biotechnol. **12:** 82–95.

189. **Lynch, M.** 2007. The origins of genome architecture. Sinauer Associates, New York, NY.

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

191. **MacIsaac, K. D., and E. Fraenkel.** 2006. Practical strategies for discovering regulatory DNA sequence motifs. PLoS Comput. Biol. **2:**201–210.

192. **Madan, B. M., and S. A. Teichmann.** 2003. Functional determinants of transcription factors in *Escherichia coli*: protein families and binding sites. Trends Genet. **19:**75–79.

192a. **Maeda, H., N. Fujita, and A. Ishihama.** 2000. Competition between seven *Escherichia coli* σ subunits: relative binding affinities to the core RNA polymerase. Nucleic Acids Res. **28:**3497–3503.

193. **Maere, S., P. van Dijck, and M. Kuiper.** 2008. Extracting expression modules from perturbational gene expression compendia. BMC Syst. Biol. **2:**33.

194. **Man, T. K., and G. D. Stormo.** 2001. Non-independence of Mnt repressor-operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay. Nucleic Acids Res. **29:**2471–2478.

195. **Mandal, M., B. Boese, J. E. Barrick, W. C. Winkler, and R. R. Breaker.** 2003. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. Cell **113:**577–586.

196. **Manke, T., H. G. Roider, and M. Vingron.** 2008. Statistical modeling of transcription factor binding affinities predicts regulatory interactions. PLoS Comput. Biol. **4:**e1000039.

197. **Marchisio, M. A., and J. Stelling.** 2008. Computational design of synthetic gene circuits with composable parts. Bioinformatics **24:**1903–1910.

198. **Mardis, E. R.** 2007. ChIP-seq: welcome to the new frontier. Nat. Methods **4:**613–614.

199. **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.

200. **Martin, R. G., W. K. Gilette, N. I. Martin, and J. L. Rosner.** 2002. Complex formation between activator and RNA polymerase as the basis for transcriptional activation by *marA* and *soxS* in *Escherichia coli*. Mol. Microbiol. **43:**355–370.

201. **Martin, S., Z. Zhang, A. Martino, and J. L. Faulon.** 2007. Boolean dynamics of genetic regulatory networks inferred from microarray time series data. Bioinformatics **23:**866–874.

202. **Mashego, M. R., K. Rumbold, M. De Mey, E. Vandamme, W. Soetaert, and J. J. Heijnen.** 2007. Microbial metabolomics: past, present and future methodologies. Biotechnol. Lett. **29:**1–16.

203. **Masse, E., N. Majdalani, and S. Gottesman.** 2003. Regulatory roles for small RNAs in bacteria. Curr. Opin. Microbiol. **6:**120–124.

204. **Mayo, A. E., Y. Setty, S. Shavit, A. Zaslaver, and U. Alon.** 2006. Plasticity of the cis-regulatory input function of a gene. PLoS Biol. **4:**555–561.

205. **McCracken, A., M. S. Turner, P. Giffard, L. M. Hafner, and P. Timms.** 2000. Analysis of promoter sequences from *Lactobacillus* and *Lactococcus* and their activity in several *Lactobacillus* species. Arch. Microbiol. **173:**383–389.

206. **McCue, L. A., W. Thompson, C. S. Carmack, and C. E. Lawrence.** 2002. Factors influencing the identification of transcription factor binding sites by cross-species comparison. Genome Res. **12:**1523–1532.

207. **McGregor, A., P. Shaw, J. Hancock, D. Bopp, M. Hediger, and N. Wratten.** 2001. Rapid restructuring of bicoid-dependent hunchback promoters within and between dipteran species: implications for molecular evolution. Evol. Dev. **3:**397–407.

208. **McLeod, S. M., and R. C. Johnson.** 2001. Control of transcription by nucleoid proteins. Curr. Opin. Microbiol. **4:**152–159.

209. **Michal, L., O. Mizrahi-Man, and Y. Pilpel.** 2008. Functional characterization of variations on regulatory motifs. PLoS Genet. **4:**e1000018.

210. **Moir-Blais, T. R., F. J. Grundy, and T. M. Henkin.** 2001. Transcriptional activation of the *Bacillus subtilis ackA* promoter requires sequences upstream of the CcpA binding site. J. Bacteriol. **183:**2389–2393.

211. **Molle, V., M. Fujita, S. T. Jensen, P. Eichenberger, J. E. Gonzalez-Pastor, J. S. Liu, and R. Losick.** 2003. The Spo0A regulon of *Bacillus subtilis*. Mol. Microbiol. **50:**1683–1701.

212. **Moreno-Campuzano, S., S. C. Janga, and E. Pérez-Rueda.** 2006. Identification and analysis of DNA-binding transcription factors in *Bacillus subtilis* and other Firmicutes—a genomic approach. BMC Genomics **7:**147.

213. **Morozov, A. V., and E. D. Siggia.** 2007. Connecting protein structure with predictions of regulatory sites. Proc. Natl. Acad. Sci. USA **104:**7068–7073.

214. **Moses, A. M., D. Y. Chiang, M. Kellis, E. S. Lander, and M. B. Eisen.** 2003. Position specific variation in the rate of evolution in transcription factor binding sites. BMC Evol. Biol. **3:**19.

215. **Mustonen, V., and M. Lassig.** 2005. Evolutionary population genetics of promoters: predicting binding sites and functional phylogenies. Proc. Natl. Acad. Sci. USA **102:**15936–15941.

216. **Nachman, I., A. Regev, and N. Friedman.** 2004. Inferring quantitative models of regulatory networks from expression data. Bioinformatics **20**(Suppl. 1)**:**i248–i256.

217. **Neph, S., and M. Tompa.** 2006. MicroFootPrinter: a tool for phylogenetic footprinting in prokaryotic genomes. Nucleic Acids Res. **34:**W366–W368.

218. **Nguyen, T. T., and I. P. Androulakis.** 2009. Recent advances in the computational discovery of transcription factor binding sites. Algorithms **2:**582–605.

219. **Nickels, B. E., J. Mukhopadhyay, S. J. Garrity, R. H. Ebright, and A. Hochschild.** 2004. The sigma(70) subunit of RNA polymerase mediates a promoter-proximal pause at the lac promoter. Nat. Struct. Mol. Biol. **11:** 544–550.

220. **Nudler, E., and M. E. Gottesman.** 2002. Transcription termination and anti-termination in E. coli. Genes Cells **7:**755–768.

221. **Oda, M., K. Furukawa, K. Ogata, A. Sarai, and H. Nakamura.** 1998. Thermodynamics of specific and non-specific DNA binding by the c-Myb DNA-binding domain. J. Mol. Biol. **276:**571–590.

222. **O'Flanagan, R. A., G. Paillard, R. Lavery, and A. M. Sengupta.** 2005. Non-additivity in protein-DNA binding. Bioinformatics **21:**2254–2263.

223. **Ohta, T.** 2003. Evolution by gene duplication revisited: differentiation of regulatory elements versus proteins. Genetica **118:**209–216.

224. **Omont, N., and F. Kepes.** 2004. Transcription/replication collisions cause bacterial transcription units to be longer on the leading strand of replication. Bioinformatics **20:**2719–2725.

225. Reference deleted.

226. **Paget, M. S., and J. D. Helmann.** 2003. The sigma70 family of sigma factors. Genome Biol. **4:**203.

227. **Pan, Y., T. Durfee, J. Bockhorst, and M. Craven.** 2007. Connecting quantitative regulatory-network models to the genome. Bioinformatics **23:**i367–i376.

228. **Paul, L., P. K. Mishra, R. M. Blumenthal, and R. G. Matthews.** 2007. Integration of regulatory signals through involvement of multiple global regulators: control of the *Escherichia coli gltBDF* operon by Lrp, IHF, Crp, and ArgR. BMC Microbiol. **7:**2.

229. **Pavesi, G., P. Mereghetti, G. Mauri, and G. Pesole.** 2004. Weeder Web: discovery of transcription factor binding sites in a set of sequences from co-regulated genes. Nucleic Acids Res. **32:**W199–W203.

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

231. **Pérez-Rueda, E., and J. Collado-Vides.** 2000. The repertoire of DNA-binding transcriptional regulators in *Escherichia coli* K-12. Nucleic Acids Res. **28:**1838–1847.

232. **Pevzner, P. A., and S. H. Sze.** 2000. Combinatorial approaches to finding

subtle signals in DNA sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. **8:**269–278.

233. **Pillai, S., and S. P. Chellappan.** 2009. ChIP on chip assays: genome-wide analysis of transcription factor binding and histone modifications. Methods Mol. Biol. **523:**341–366.
234. **Pillai, S., P. Dasgupta, and S. P. Chellappan.** 2009. Chromatin immunoprecipitation assays: analyzing transcription factor binding and histone modifications in vivo. Methods Mol. Biol. **523:**323–339.
235. **Prakash, A., M. Blanchette, S. Sinha, and M. Tompa.** 2004. Motif discovery in heterogeneous sequence data. Pac. Symp. Biocomput. 348–359.
236. **Price, M. N., E. J. Alm, and A. P. Arkin.** 2005. Interruptions in gene expression drive highly expressed operons to the leading strand of DNA replication. Nucleic Acids Res. **33:**3224–3234.
237. **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.
238. **Radonjic, M., J. C. Andrau, P. Lijnzaad, P. Kemmeren, T. T. Kockelkorn, D. van Leenen, N. L. van Berkum, and F. C. Holstege.** 2005. Genome-wide analyses reveal RNA polymerase II located upstream of genes poised for rapid response upon *S. cerevisiae* stationary phase exit. Mol. Cell **18:**171–183.
239. **Raijman, D., R. Shamir, and A. Tanay.** 2008. Evolution and selection in yeast promoters: analyzing the combined effect of diverse transcription factor binding sites. PLoS Comput. Biol. **4:**e7.
240. **Rajewsky, N., N. D. Socci, M. Zapotocky, and E. D. Siggia.** 2002. The evolution of DNA regulatory regions for proteo-gamma bacteria by interspecies comparisons. Genome Res. **12:**298–308.
241. **Ray, P., S. Shringarpure, M. Kolar, and E. P. Xing.** 2008. CSMET: comparative genomic motif detection via multi-resolution phylogenetic shadowing. PLoS Comput. Biol. **4:**e1000090.
242. **Reed, B. D., A. E. Charos, A. M. Szekely, S. M. Weissman, and M. Snyder.** 2008. Genome-wide occupancy of SREBP1 and its partners NFY and SP1 reveals novel functional roles and combinatorial regulation of distinct classes of genes. PLoS Genet. **4:**e1000133.
243. **Reisenauer, A., L. S. Kahng, S. McCollum, and L. Shapiro.** 1999. Bacterial DNA methylation: a cell cycle regulator? J. Bacteriol. **181:**5135–5139.
244. **Repoila, F., N. Majdalani, and S. Gottesman.** 2003. Small non-coding RNAs, co-ordinators of adaptation processes in *Escherichia coli*: the RpoS paradigm. Mol. Microbiol. **48:**855–861.
245. **Robertson, G., M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. M. Euskirchen, B. Bernier, R. Varhol, A. Delaney, N. Thiessen, O. L. Griffith, A. He, M. Marra, M. Snyder, and S. Jones.** 2007. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat. Methods **4:**651–657.
246. **Rocha, E. P. C., and A. Danchin.** 2003. Essentiality, not expressiveness, drives gene-strand bias in bacteria. Nat. Genet. **34:**377–378.
247. **Rodionov, D. A.** 2007. Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. Chem. Rev. **107:**3467–3497.
248. **Rosenfeld, N., M. B. Elowitz, and U. Alon.** 2002. Negative autoregulation speeds the response times of transcription networks. J. Mol. Biol. **323:**785–793.
249. **Roth, F. P., J. D. Hughes, P. W. Estep, and G. M. Church.** 1998. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. Nat. Biotechnol. **16:**939–945.
250. **Roulet, E., S. Busso, A. A. Camargo, A. J. Simpson, N. Mermod, and P. Bucher.** 2002. High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites. Nat. Biotechnol. **20:**831–835.
251. **Roy, S., M. Werner-Washburne, and T. Lane.** 2008. A system for generating transcription regulatory networks with combinatorial control of transcription. Bioinformatics **24:**1318–1320.
252. **Rutberg, B.** 1997. Antitermination of transcription of catabolic operons. Mol. Microbiol. **23:**413–421.
253. **Sabatti, C., L. Rohlin, K. Lange, and J. C. Liao.** 2005. Vocabulon: a dictionary model approach for reconstruction and localization of transcription factor binding sites. Bioinformatics **21:**922–931.
254. **Sagot, M.** 1998. Spelling approximate repeated or common motifs using a suffix tree. Lect. Notes Comp. Sci. **1380:**111–127.
255. **Samal, A., and S. Jain.** 2008. The regulatory network of *E. coli* metabolism as a Boolean dynamical system exhibits both homeostasis and flexibility of response. BMC Syst. Biol. **2:**21.
256. **Sandve, G. K., O. Abul, V. Walseng, and F. Drablos.** 2007. Improved benchmarks for computational motif discovery. BMC Bioinformatics **8:**193.
257. **Sandve, G. K., and F. Drablos.** 2006. A survey of motif discovery methods in an integrated framework. Biol. Direct **1:**11.
258. **Schafer, J., and K. Strimmer.** 2005. An empirical Bayes approach to inferring large-scale gene association networks. Bioinformatics **21:**754–764.
259. **Schlitt, T., and A. Brazma.** 2007. Current approaches to gene regulatory network modelling. BMC Bioinformatics **8**(Suppl. 6)**:**S9.
260. **Schlitt, T., and A. Brazma.** 2005. Modelling gene networks at different organisational levels. FEBS Lett. **579:**1859–1866.
261. **Schneider, T. D.** 2000. Evolution of biological information. Nucleic Acids Res. **28:**2794–2799.
262. **Schneider, T. D., and R. M. Stephens.** 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. **18:**6097–6100.
263. **Schultz, D., J. E. Ben, J. N. Onuchic, and P. G. Wolynes.** 2007. Molecular level stochastic model for competence cycles in Bacillus subtilis. Proc. Natl. Acad. Sci. USA **104:**17582–17587.
264. **Selinger, D. W., R. M. Saxena, K. J. Cheung, G. M. Church, and C. Rosenow.** 2003. Global RNA half-life analysis in *Escherichia coli* reveals positional patterns of transcript degradation. Genome Res. **13:**216–223.
265. **Semsey, S., S. Krishna, K. Sneppen, and S. Adhya.** 2007. Signal integration in the galactose network of *Escherichia coli*. Mol. Microbiol. **65:**465–476.
266. **Sengupta, A. M., M. Djordjevic, and B. I. Shraiman.** 2002. Specificity and robustness in transcription control networks. Proc. Natl. Acad. Sci. USA **99:**2072–2077.
267. **Setty, Y., A. E. Mayo, M. G. Surette, and U. Alon.** 2003. Detailed map of a *cis*-regulatory input function. Proc. Natl. Acad. Sci. USA **100:**7702–7707.
268. **Shearwin, K. E., B. P. Callen, and J. B. Egan.** 2005. Transcriptional interference—a crash course. Trends Genet. **21:**339–345.
269. **Shida, K.** 2006. GibbsST: a Gibbs sampling method for motif discovery with enhanced resistance to local optima. BMC Bioinformatics **7:**486.
270. **Shimada, T., A. Ishihama, S. J. Busby, and D. C. Grainger.** 2008. The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. Nucleic Acids Res. **36:**3950–3955.
271. **Shivers, R. P., S. S. Dineen, and A. L. Sonenshein.** 2006. Positive regulation of Bacillus subtilis *ackA* by CodY and CcpA: establishing a potential hierarchy in carbon flow. Mol. Microbiol. **62:**811–822.
272. **Shmulevich, I., E. R. Dougherty, S. Kim, and W. Zhang.** 2002. Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. Bioinformatics **18:**261–274.
273. **Siddharthan, R., E. D. Siggia, and E. van Nimwegen.** 2005. PhyloGibbs: a Gibbs sampling motif finder that incorporates phylogeny. PLoS Comput. Biol. **1:**e67.
274. **Sierro, N., Y. Makita, M. de Hoon, and K. Nakai.** 2008. DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. Nucleic Acids Res. **36:**D93–D96.
275. **Silva-Rocha, R., and V. de Lorenzo.** 2008. Mining logic gates in prokaryotic transcriptional regulation networks. FEBS Lett. **582:**1237–1244.
276. **Singh, C. P., F. Khan, B. N. Mishra, and D. S. Chauhan.** 2008. Performance evaluation of DNA motif discovery programs. Bioinformation **3:**205–212.
277. **Sinha, S., M. Blanchette, and M. Tompa.** 2004. PhyME: a probabilistic algorithm for finding motifs in sets of orthologous sequences. BMC Bioinformatics **5:**170.
278. **Smits, W. K., C. C. Eschevins, K. A. Susanna, S. Bron, O. P. Kuipers, and L. W. Hamoen.** 2005. Stripping *Bacillus*: ComK auto-stimulation is responsible for the bistable response in competence development. Mol. Microbiol. **56:**604–614.
279. **Smits, W. K., T. T. Hoa, L. W. Hamoen, O. P. Kuipers, and D. Dubnau.** 2007. Antirepression as a second mechanism of transcriptional activation by a minor groove binding protein. Mol. Microbiol. **64:**368–381.
280. **Smits, W. K., O. P. Kuipers, and J. W. Veening.** 2006. Phenotypic variation in bacteria: the role of feedback regulation. Nat. Rev. Microbiol. **4:**259–271.
281. **Snyder, M., and M. B. Gerstein.** 2003. Genomics. Defining genes in the genomics era. Science **300:**258–260.
282. **Stepanova, E., J. Lee, M. Ozerova, E. Semenova, K. Datsenko, B. L. Wanner, K. Severinov, and S. Borukhov.** 2007. Analysis of promoter targets for *Escherichia coli* transcription elongation factor GreA in vivo and in vitro. J. Bacteriol. **189:**8772–8785.
283. **Stormo, G. D., and D. S. Fields.** 1998. Specificity, free energy and information content in protein-DNA interactions. Trends Biochem. Sci. **23:**109–113.
284. **Stormo, G. D., T. D. Schneider, L. Gold, and A. Ehrenfeucht.** 1982. Use of the 'Perceptron' algorithm to distinguish translational initiation sites in *E. coli*. Nucleic Acids Res. **10:**2997–3011.
285. **Stulke, J., and W. Hillen.** 1999. Carbon catabolite repression in bacteria. Curr. Opin. Microbiol. **2:**195–201.
286. **Susanna, K. A., C. D. den Hengst, L. W. Hamoen, and O. P. Kuipers.** 2006. Expression of transcription activator ComK of *Bacillus subtilis* in the heterologous host *Lactococcus lactis* leads to a genome-wide repression pattern: a case study of horizontal gene transfer. Appl. Environ. Microbiol. **72:**404–411.
287. **Susanna, K. A., A. M. Mironczuk, W. K. Smits, L. W. Hamoen, and O. P. Kuipers.** 2007. A single, specific thymine mutation in the ComK-binding site severely decreases binding and transcription activation by the competence transcription factor ComK of *Bacillus subtilis*. J. Bacteriol. **189:**4718–4728.
288. **Swinger, K. K., and P. A. Rice.** 2004. IHF and HU: flexible architects of bent DNA. Curr. Opin. Struct. Biol. **14:**28–35.
289. **Szoke, P. A., T. L. Allen, and P. L. deHaseth.** 1987. Promoter recognition by *Escherichia coli* RNA polymerase: effects of base substitutions in the −10 and −35 regions. Biochemistry **26:**6188–6194.
290. **Thijs, G., M. Lescot, K. Marchal, S. Rombauts, B. de Moor, P. Rouze, and Y. Moreau.** 2001. A higher-order background model improves the detection

of promoter regulatory elements by Gibbs sampling. Bioinformatics **17:** 1113–1122.

291. **Thomas-Chollier, M., O. Sand, J. V. Turatsinze, R. Janky, M. Defrance, E. Vervisch, S. Brohee, and J. van Helden.** 2008. RSAT: regulatory sequence analysis tools. Nucleic Acids Res. **36:**W119–W127.
292. **Thompson, J. D., T. J. Gibson, and D. G. Higgins.** 2002. Multiple sequence alignment using ClustalW and ClustalX. Curr. Protoc. Bioinform. **2:**3.2.
293. **Tompa, M., N. Li, T. L. Bailey, G. M. Church, B. de Moor, E. Eskin, A. V. Favorov, M. C. Frith, Y. T. Fu, W. J. Kent, V. J. Makeev, A. A. Mironov, W. S. Noble, G. Pavesi, G. Pesole, M. Regnier, N. Simonis, S. Sinha, G. Thijs, J. van Helden, M. Vandenbogaert, Z. P. Weng, C. Workman, C. Ye, and Z. Zhu.** 2005. Assessing computational tools for the discovery of transcription factor binding sites. Nat. Biotechnol. **23:**137–144.
294. **Touzain, F., S. Schbath, I. Bled-Rennesson, B. Aigle, G. Kucherov, and P. Leblond.** 2008. SIGffRid: a tool to search for sigma factor binding sites in bacterial genomes using comparative approach and biologically driven statistics. BMC Bioinformatics **9:**73.
295. **Towsey, M. W., J. J. Gordon, and J. M. Hogan.** 2006. The prediction of bacterial transcription start sites using SVMs. Int. J. Neural Syst. **16:**363–370.
296. **Travers, A., and G. Muskhelishvili.** 2005. Bacterial chromatin. Curr. Opin. Genet. Dev. **15:**507–514.
297. **Turatsinze, J. V., M. Thomas-Chollier, M. Defrance, and J. van Helden.** 2008. Using RSAT to scan genome sequences for transcription factor binding sites and cis-regulatory modules. Nat. Protoc. **3:**1578–1588.
298. **Ushida, C., and H. Aiba.** 1990. Helical phase dependent action of CRP: effect of the distance between the CRP site and the −35 region on promoter activity. Nucleic Acids Res. **18:**6325–6330.
299. **Valouev, A., D. S. Johnson, A. Sundquist, C. Medina, E. Anton, S. Batzoglou, R. M. Myers, and A. Sidow.** 2008. Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat. Methods **5:**829–834.
300. **van Helden, J., B. Andre, and J. Collado-Vides.** 1998. Extracting regulatory sites from the upstream region of yeast genes by computational analysis of oligonucleotide frequencies. J. Mol. Biol. **281:**827–842.
301. **van Helden, J., A. F. Rios, and J. Collado-Vides.** 2000. Discovering regulatory elements in non-coding sequences by analysis of spaced dyads. Nucleic Acids Res. **28:**1808–1818.
302. **Veber, P., C. Guziolowski, B. M. Le, O. Radulescu, and A. Siegel.** 2008. Inferring the role of transcription factors in regulatory networks. BMC Bioinformatics **9:**228.
303. **Vingron, M., A. Brazma, R. Coulson, J. van Helden, T. Manke, K. Palin, O. Sand, and E. Ukkonen.** 2009. Integrating sequence, evolution and functional genomics in regulatory genomics. Genome Biol. **10:**202.
304. **Voskuil, M. I., and G. H. Chambliss.** 1998. The −16 region of *Bacillus subtilis* and other gram-positive bacterial promoters. Nucleic Acids Res. **26:**3584–3590.
305. **Voskuil, M. I., and G. H. Chambliss.** 2002. The TRTGn motif stabilizes the transcription initiation open complex. J. Mol. Biol. **322:**521–532.
306. **Wade, J. T., D. C. Roa, D. C. Grainger, D. Hurd, S. J. Busby, K. Struhl, and E. Nudler.** 2006. Extensive functional overlap between sigma factors in *Escherichia coli*. Nat. Struct. Mol. Biol. **13:**806–814.
307. **Wagner, A.** 2005. Robustness and evolvability in living systems. Princeton University Press, Princeton, NJ.
308. **Wakeman, C. A., W. C. Winkler, and C. E. Dann III.** 2007. Structural features of metabolite-sensing riboswitches. Trends Biochem. Sci. **32:**415–424.
309. **Wang, G., T. Yu, and W. Zhang.** 2005. WordSpy: identifying transcription factor binding motifs by building a dictionary and learning a grammar. Nucleic Acids Res. **33:**W412–W416.
310. **Wang, T.** 2007. Using PhyloCon to identify conserved regulatory motifs. Curr. Protoc. Bioinform. **2:**2.12.
311. **Wang, T., and G. D. Stormo.** 2003. Combining phylogenetic data with co-regulated genes to identify regulatory motifs. Bioinformatics **19:**2369–2380.
312. Reference deleted.
313. **Wels, M., C. Francke, R. Kerkhoven, M. Kleerebezem, and R. J. Siezen.** 2006. Predicting cis-acting elements of Lactobacillus plantarum by comparative genomics with different taxonomic subgroups. Nucleic Acids Res. **34:**1947–1958.
314. **Westenberg, M. A., S. A. F. T. van Hijum, O. P. Kuipers, and J. B. T. M. Roerdink.** 2008. Visualizing genome expression and regulatory network dynamics in genomic and metabolic context. Comput. Graph. Forum **27:** 887–894.

314a.**Westholm, J. O., F. Xu, H. Ronne, and J. Komorowski.** 2008. Genome-scale study of the importance of binding site context for transcription factor binding and gene regulation. BMC Bioinformatics **9:**484.

315. **Wijaya, E., K. Rajaraman, S. M. Yiu, and W. K. Sung.** 2007. Detection of generic spaced motifs using submotif pattern mining. Bioinformatics **23:** 1476–1485.
316. **Wijaya, E., S. M. Yiu, N. T. Son, R. Kanagasabai, and W. K. Sung.** 2008. MotifVoter: a novel ensemble method for fine-grained integration of generic motif finders. Bioinformatics **24:**2288–2295.
317. **Wilhelm, B. T., and J. R. Landry.** 2009. RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. Methods **48:**249–257.
318. **Wilson, R., M. Urbanowski, and G. Stauffer.** 1995. DNA binding sites of the LysR-type regulator GcvA in the *gcv* and *gcvA* control regions of *Escherichia coli*. J. Bacteriol. **177:**4940–4946.
319. **Winkler, W. C., and R. R. Breaker.** 2005. Regulation of bacterial gene expression by riboswitches. Annu. Rev. Microbiol. **59:**487–517.
320. **Wolberger, C.** 1999. Multiprotein-DNA complexes in transcriptional regulation. Annu. Rev. Biophys. Biomol. Struct. **28:**29–56.
321. **Wösten, M. M. S. M.** 1998. Eubacterial sigma-factors. FEMS Microbiol. Rev. **22:**127–150.
322. **Wunderlich, Z., and L. A. Mirny.** 2008. Spatial effects on the speed and reliability of protein-DNA search. Nucleic Acids Res. **36:**3570–3578.
323. **Yanover, C., M. Singh, and E. Zaslavsky.** 2009. M are better than one: an ensemble-based motif finder and its application to regulatory element prediction. Bioinformatics **25:**868–874.
324. **Yu, H., N. M. Luscombe, J. Qian, and M. Gerstein.** 2003. Genomic analysis of gene expression relationships in transcriptional regulatory networks. Trends Genet. **19:**422–427.
325. **Zalieckas, J. M., L. V. Wray, Jr., and S. H. Fisher.** 1998. Expression of the *Bacillus subtilis acsA* gene: position and sequence context affect *cre*-mediated carbon catabolite repression. J. Bacteriol. **180:**6649–6654.
326. **Zare, H., D. Sangurdekar, P. Srivastava, M. Kaveh, and A. B. Khodursky.** 2009. Reconstruction of Escherichia coli transcriptional regulatory networks via regulon-based associations. BMC Syst. Biol. **3:**39.
327. **Zare-Mirakabad, F., H. Ahrabian, M. Sadeghi, A. Nowzari-Dalini, and B. Goliaei.** 2009. New scoring schema for finding motifs in DNA sequences. BMC Bioinformatics **10:**93.
328. **Zhang, Y., J. Xuan, B. G. de Los Reyes, R. Clarke, and H. W. Ressom.** 2008. Network motif-based identification of transcription factor-target gene relationships by integrating multi-source biological data. BMC Bioinformatics **9:**203.
329. **Zhou, D., and R. Yang.** 2006. Global analysis of gene transcription regulation in prokaryotes. Cell. Mol. Life Sci. **63:**2260–2290.
330. **Zhou, Q., and J. S. Liu.** 2004. Modeling within-motif dependence for transcription factor binding site predictions. Bioinformatics **20:**909–916.
331. **Zhu, C., K. J. Byers, R. P. McCord, Z. Shi, M. F. Berger, D. E. Newburger, K. Saulrieta, Z. Smith, M. V. Shah, M. Radhakrishnan, A. A. Philippakis, Y. Hu, F. de Masi, M. Pacek, A. Rolfs, T. Murthy, J. Labaer, and M. L. Bulyk.** 2009. High-resolution DNA-binding specificity analysis of yeast transcription factors. Genome Res. **19:**556–566.
332. **Zhu, J., B. Zhang, E. N. Smith, B. Drees, R. B. Brem, L. Kruglyak, R. E. Bumgarner, and E. E. Schadt.** 2008. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. Nat. Genet. **40:**854–861.
333. **Zhu, R., A. S. Ribeiro, D. Salahub, and S. A. Kauffman.** 2007. Studying genetic regulatory networks at the molecular level: delayed reaction stochastic models. J. Theor. Biol. **246:**725–745.
334. **Zomer, A. L., G. Buist, R. Larsen, J. Kok, and O. P. Kuipers.** 2007. Time-resolved determination of the CcpA regulon of *Lactococcus lactis* subsp. *cremoris* MG1363. J. Bacteriol. **189:**1366–1381.

Continued next page

Sacha van Hijum was born in Bedum, The Netherlands, in 1972. He studied bacterial molecular biology (University of Groningen, The Netherlands) and obtained his Ph.D. at the Microbial Physiology Department (University of Groningen). He did three postdoctorals at the Molecular Genetics Department (University of Groningen) and one at the Interfacultary Centre of Functional Genomics (University of Greifswald, Germany). Presently, he is working at NIZO Food Research (Ede, The Netherlands) and the CMBI Bacterial Genomics Group (Radboud University, Nijmegen, The Netherlands). For the past years, research focus was on studying gene regulatory interactions in prokaryotes using computational biology techniques. Currently, the focus has broadened to data analysis and mining of high-throughput technologies such as DNA microarrays, proteomics, metabolomics, and next-generation sequencing. Bioinformatics is used to integrate these complex and multivariate data sources in order to understand the complex interactions occurring at various regulatory levels (e.g., transcriptional and metabolic networks) underlying an organism's response to its changing environment.

Marnix Medema was born in Vaassen, The Netherlands, in 1986. He obtained his B.Sc. in biology at the Radboud University of Nijmegen and then finished the top master program Biomolecular Sciences at the University of Groningen, from which he graduated in 2008. He is now starting his Ph.D. research at the department of Microbial Physiology in Groningen, on genomics and systems biology of the actinomycete bacteria.

Oscar Kuipers was born in Rotterdam, The Netherlands. He studied Biology at Utrecht University and received his master's degree in Molecular Biology, Biochemistry, and Informatics in 1986. He obtained his Ph.D. in protein engineering of porcine pancreatic phospholipase A_2 in 1990, after which he was appointed as postdoctoral and later project leader and group leader of genetics at NIZO Food Research in Ede, The Netherlands. In 1999, he was appointed Full Professor in Molecular Genetics at the University of Groningen, The Netherlands. His current research interests include functional genomics and physiology studies of low-GC gram-positive bacteria. Currently, he is studying gene regulatory networks in these organisms as well as the phenomenon of phenotypic bistability occurring in, e.g., competence for genetic transformation and sporulation. Moreover, he has a keen interest in biosynthesis, regulation, immunity, and mode of action of a number of different antimicrobial peptides, in the role of metal ions in virulence and pathogenesis, and in general stress responses in bacteria.