Chromosomal organization is shaped by the transcription regulatory network

Ruth Hershberg[1], Esti Yeger-Lotem[1,2] and Hanah Margalit[1]

[1]Department of Molecular Genetics and Biotechnology, Faculty of Medicine, The Hebrew University, Jerusalem 91120, Israel
[2]Department of Computer Science, Technion, Haifa 32000, Israel

Transcription regulation, a key step in the control of gene expression, has been the focus of several large-scale studies; however, little attention has been given to its relationship with the chromosomal arrangement of transcription units. We studied this relationship systematically in *Escherichia coli* and *Saccharomyces cerevisiae* using network analysis methods. Our analysis reveals links between transcription regulation and chromosomal organization, suggesting that in both organisms transcription regulation has shaped the organization of transcription units on the chromosome. Differences found between the organisms reflect the inherent differences in transcription regulation between prokaryotes and eukaryotes.

Introduction

Cells invoke different cellular processes in response to dynamic changes in their environment. This is accomplished through a complex regulatory program by which distinct groups of genes are turned 'on' or 'off', depending on various environmental signals. Much of the regulation occurs at the level of transcription. It has long been known that transcription regulation is related to chromosomal gene order: prokaryotic operons, sets of genes that are arranged in tandem on the chromosome and that are co-transcribed as a single message, are the most well-known example [1,2]. In eukaryotes, non-random gene order has also been well established: it has been shown that functionally related genes, such as genes belonging to the same metabolic pathway, tend to be clustered on the chromosome [3]. In addition, these genes are more likely to be co-expressed than random pairs of genes [4–6]. Although the relationship between transcription regulation and chromosomal organization has been reported, a systematic study of this relationship has been lacking.

In this article, we conducted a comprehensive analysis of the relationship between transcription regulation and the chromosomal organization of transcription units (TUs) on two model organisms: the prokaryote *Escherichia coli* and the eukaryote *Saccharomyces cerevisiae*. We define TUs as sets of one or more genes that are transcribed as a single message. Thus, in *E. coli* a TU can include several genes if they reside in the same operon [7–9], whereas in *S. cerevisiae* each TU corresponds to a single gene [10]. Data used in this study (TUs and their regulatory interactions [7–13]) are summarized in Table 1. By representing TUs as nodes, and transcription regulation and chromosomal adjacency as two distinct types of edges between nodes, an integrated network was created for each organism. A schematic representation of the *E. coli* network is presented in Figure 1. Using the algorithm developed by Yeger-Lotem *et al.* [13], we searched these networks for integrated network motifs [12–14] – connected patterns that recur in the examined network significantly more often than expected at random and that contain the two edge types. Although the experimental data used are incomplete, it is important to note that this method is robust to errors in the data [15]. This analysis revealed several highly significant network motifs of two, three and four TUs (Table 2) that uncover the inter-relationship between chromosomal organization and transcription regulation. Two major phenomena were observed: (i) adjacent TUs are often co-regulated by the same transcription factor (TF); and (ii) TFs often regulate the TU that is adjacent to their encoding TU.

Co-regulation of neighbors

Co-regulation of adjacent TUs, which we term co-regulation of neighbors, appears in two distinct forms. In the first form, one of the neighbors itself encodes the regulating TF (Table 2b). This significant two-node motif was observed only in *E. coli* and is denoted co-regulation *in cis*. The second form, denoted co-regulation *in trans*, is a three-TU motif in both *E. coli* and *S. cerevisiae*, in which

Table 1. Number of genes, TUs and transcription-regulation relationships used in this study

Organism	Data	
Escherichia coli	Genes	4308
	Transcription units (TUs)	3405
	Transcription factor (TF)-encoding genes	111
	Regulated genes	737
	TF-encoding TUs	106
	Regulated TUs	337
	Transcriptional regulatory interactions[a]	549
Saccharomyces cerevisiae	Genes	4140
	TF-encoding genes	126
	Regulated genes	558
	Transcriptional regulatory interactions	1271
Bacillus subtilis	Genes	4225
	TUs	2229
	TF-encoding genes	90
	Regulated genes	683
	TF-encoding TUs	88
	Regulated TUs	276
	Transcriptional regulatory interactions[a]	387

[a]Number of regulatory interactions between TF-encoding TUs and regulated TUs.

Corresponding author: Margalit, H. (hanah@md.huji.ac.il).
Available online 20 January 2005

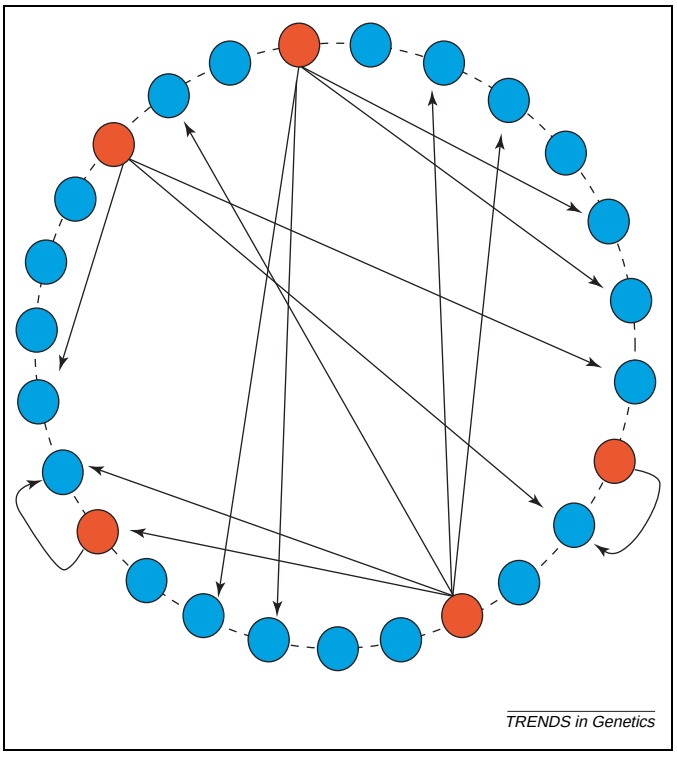


Figure 1. Schematic representation of the *Escherichia coli* integrated network. Red circles represent transcription factor (TF)-encoding transcription units (TUs); blue circles represent all other TUs; undirected edges (broken lines) connect TUs that are adjacent on the chromosome and directed edges (arrows) point from a TF-encoding TU to its target TUs.

the TF co-regulating the two neighbors is encoded by a third TU (Table 2c).

In a recent study, Teichmann and Babu [16] showed that gene duplication had a key role in the evolution of the transcriptional regulatory networks of *E. coli* and *S. cerevisiae*. To investigate whether the phenomenon of co-regulation of neighbors can be largely explained by gene duplication, we aligned the sequences of the co-regulated neighboring TUs against each other (see supplementary methods). We found that only 19% of co-regulated neighbor pairs in *S. cerevisiae* and none of the pairs in *E. coli* show significant sequence similarity and, therefore, concluded that the co-regulation of neighbors cannot be largely attributed to gene duplication.

Although co-regulation of neighbors is common to both organisms, we observed differences in the relative orientations of the co-regulated neighbors. Neighboring TUs can be divergent, convergent or unidirectional relative to one another (Figure 2). *E. coli* exhibits a clear preference for the divergent orientation: 73% of co-regulated neighbors are divergent to each other (Figure 2 and supplementary Table 1). In *S. cerevisiae,* there is less of a clear preference towards divergent orientation (Figure 2).

What underlies this orientation preference in *E. coli*? Divergent orientation enables the sharing of regulatory sequences between the adjacent TUs. We examined 16 pairs of divergent co-regulated TUs, for which the exact locations of the binding sites that mediate the regulation of each of the TUs by their common TF are known. For 14 of these TU pairs, co-regulation is mediated through at least one shared binding site. The sharing of binding sites is advantageous for transcriptional coupling: in general, coupling the transcription of co-regulated TUs is complicated by the fact that changes in the levels of certain signals can cause a particular TF to bind certain binding sites while releasing others [11]. The sharing of binding sites eliminates this level of complexity and might thus enhance the coupled transcription of co-regulated TUs.

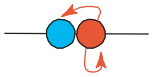
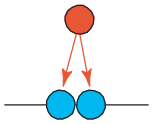
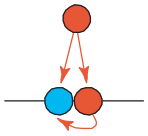
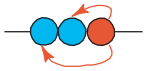
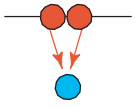
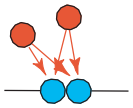
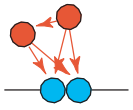
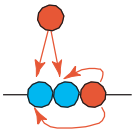
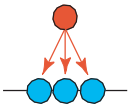
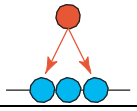
Other factors that contribute to transcriptional coupling between divergent TUs are DNA super-helical density and steric hindrance. Transcriptional coupling through changes in DNA super-helical density was observed in the divergent TUs *ilvY* and *ilvC*, where transcription of *ilvY* induces *ilvC* expression by increasing the negative supercoiling in the shared promoter region [17,18]. Steric hindrance can cause inverse coupling in the expression of adjacent divergent TUs. This has been observed for the global regulator CRP and its divergent TU *yhfA*: expression from the *yhfA* promoter prevents the binding of RNA polymerase to the *crp* promoter, thus repressing CRP expression [19,20].

These examples show that chromosomal adjacency and the orientation of TUs might affect their transcriptional coupling. To substantiate this finding, we examined the correlation between the expression profiles of all adjacent TUs in *E. coli* [21]. Pairs of adjacent TUs were divided into three groups based on their relative orientation. We found that the groups of divergent and unidirectional pairs were significantly enriched for pairs that had significant correlations in their expression profiles. No such phenomenon was found for pairs of convergent TUs (Table 3). Furthermore, we found that, in accordance with previous findings [22], co-regulation raises the probability of TU pairs being significantly correlated in their expression profiles. Remarkably, adjacency and divergence raise this probability even further (Table 3).

In a recent study, Korbel *et al.* [23] used conserved gene order for the prediction of gene function in prokaryotes. They observed that pairs of divergent genes are more often conserved in distant clades than convergent gene pairs; these conserved divergent pairs show enhanced co-expression and are often functionally associated. In accord with this, we found that among pairs of co-regulated TUs, the proportion of functionally associated pairs is greater for neighboring TUs than for non-adjacent TUs (64% and 23%, respectively; supplementary methods online).

The manifestation of co-regulation of neighbors in *S. cerevisiae* is different than in *E. coli*. Apart from the lack of co-regulation *in cis* and the weaker orientation preferences, the *S. cerevisiae* motifs show enrichment for co-regulation *in trans* of three neighboring TUs (Table 2j) and of nearly adjacent TUs (Table 2k). We also found three cases of co-regulation *in trans* of up to six adjacent TUs. This tendency can be explained by the relationship between gene expression and chromatin structure in eukaryotes [24]. Eukaryotic genes are available for transcriptional regulation by TFs only when located in regions of 'open' chromatin. Thus, it might be advantageous for a group of co-regulated genes to cluster in a chromosomal region that is 'open' under the conditions in which they are expressed. Because chromatin structure

Table 2. Integrated network motifs involving chromosomal adjacency and transcription regulation

Motif	Illustration[a]	Number of occurrences in *Escherichia coli*[b]	Number of occurrences in *Saccharomyces cerevisiae*[b]
(a) Neighbor regulation		**29**	1
(b) Co-regulation of neighbors in *cis*		**21**	0
(c) Co-regulation of neighbors *in trans*		**31**	**63**
(d) Feed-forward loop [12] in which the co-regulated TUs are neighbors		**9**	0
(e) Neighbor regulation in which the regulating TF also regulates the neighbor's neighbor		**5**	0
(f) Two adjacent TFs co-regulating a single TU		0	**5**
(g) Co-regulation of neighbors *in trans* by two separate TFs		4	**45**
(h) Two feed-forward loops [12] where the co-regulating TUs co-regulate a pair of neighbors		**5**	**6**
(i) Co-regulation of neighbors by two TFs, one of which conducts neighbor regulation		**5**	0
(j) Co-regulation *in trans* of three neighbors		2	**16**
(k) Co-regulation *in trans* of two TUs that are one TU apart		4	**13**

[a]Red circles represent TF-encoding TUs; blue circles represent TUs; red arrows represent transcription regulation interactions; the black bar represents the chromosome and adjacent TUs are drawn as adjacent circles.
[b]Numbers in bold letters designate significant motifs (patterns that appear five or more times with $P \leq 0.05$).

spans large regions, it might similarly affect nearby genes that are not necessarily immediate neighbors regardless of their relative orientations. In support of this idea, the correlation between expression profiles of adjacent, or nearly adjacent, genes in *S. cerevisiae* tends to be significantly positive, and is more sensitive to the distance between the genes than to their relative orientation [4]. We further examined the effect of adjacency on the expression of pairs of

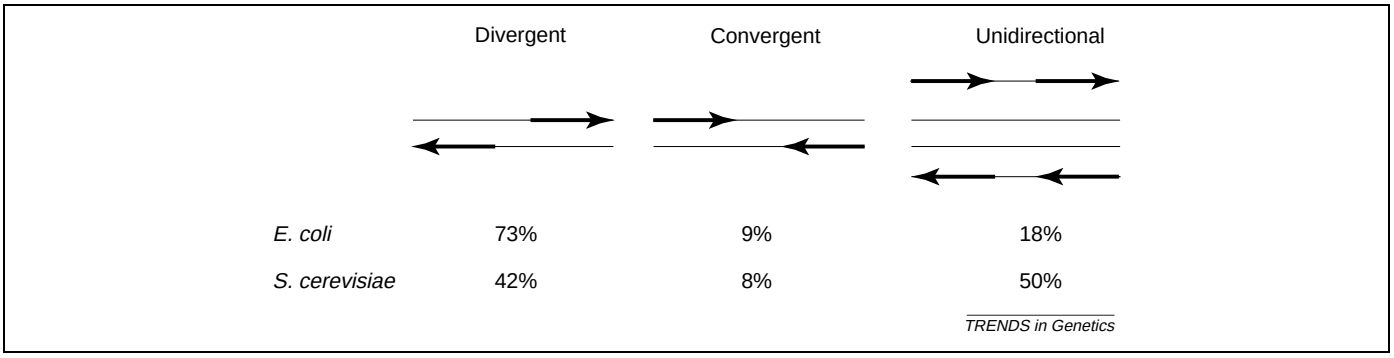


Figure 2. Distribution of orientations among adjacent co-regulated transcription units in *Escherichia coli* and *Saccharomyces cerevisiae*.

Table 3. Correlation between expression profiles of transcription unit pairs in *Escherichia coli*

Relationship between transcription units (TU)		Number of TU pairs[a]	Percentage of pairs with significantly correlated TUs[b]	*P*-value[c]
Adjacent	Divergent	481	10.8	$\ll 10^{-4}$
	Unidirectional	1632	12.6	$\ll 10^{-4}$
	Convergent	474	5.7	0.24
Co-regulated	Non-adjacent	3944	15.7	$\ll 10^{-4}$
	Adjacent	38	47.4	$\ll 10^{-4}$
	Adjacent and divergent	28	50	$\ll 10^{-4}$

[a]Number of TU pairs for which full expression data were available.
[b]Two TUs are considered to be significantly correlated in their expression profiles if $r_s \geq 0.71$ or $r_s \leq -0.38$.
[c]Probability of obtaining the observed percentage of pairs of significantly correlated TUs at random, calculated using the binomial test.

co-regulated genes using the clustering of *S. cerevisiae* genes to condition-specific 'transcription modules' by Ihmels *et al.* [25]. The fraction of gene pairs in which both pair-mates belong to a common transcription module was greater among pairs of co-regulated neighbors (47% of the 36 pairs) than among pairs of co-regulated non-adjacent TUs (21% of the 10 250 pairs). These findings indicate that adjacency indeed enhances the coupling of transcription of co-regulated genes in *S. cerevisiae*.

To conclude, in both organisms the phenomenon of co-regulation of adjacent TUs is related to transcriptional coupling. In eukaryotes, this is achieved largely through chromatin structure. In prokaryotes, coupling is enhanced either by the inclusion of genes within a single operon, where strong direct coupling is advantageous, or by adjacency and the divergent orientation of TUs, which enable a more flexible form of coupling.

Neighbor regulation

The second major phenomenon revealed by the motif analysis is that in *E. coli* a striking number of TFs (44% of the TFs in our data) regulate a TU adjacent to their encoding TU (Table 2a,b, supplementary Table 2). We call this phenomenon neighbor regulation. Neighbor regulation, although prominent in *E. coli*, does not seem to have a role in the yeast *S. cerevisiae*. To examine whether neighbor regulation can be considered a prokaryotic phenomenon, we examined an additional, evolutionarily distant prokaryote, *Bacillus subtilis*. Examination of available data of operon structure and transcription regulation in *B. subtilis* [26,27] (Table 1) showed that, similar to *E. coli*, 42% of *B. subtilis* TFs conduct neighbor regulation. The widespread occurrence of neighbor regulation in the two prokaryotes that we studied, and its absence in yeast, can be explained by the 'selfish operon' model [1,2]. According to this model, genes that are functionally related tend to cluster on the chromosome to increase their chances of being co-transferred during horizontal-transfer events. By mutual transfer the genes can confer a desirable function onto the recipient organism, thus increasing their chances of being retained in that genome. Although in a recent study, Pal *et al.* [28] challenged the selfish operon model as a full explanation of gene clustering, it might still partly explain the high incidence of neighbor regulation in prokaryotes. It is compelling to hypothesize that neighbor regulation would be even more advantageous for TFs that regulate single TUs (not including auto-regulation), because in such cases a complete regulon can be co-transferred. Intriguingly, and consistent with this hypothesis, we found that a remarkable 81% of the 42 *E. coli* TFs regulating a single target conduct neighbor regulation, whereas only 26% of the 57 TFs regulating more than one target do so. The selfish operon model can also account for the lack of neighbor regulation in yeast, because horizontal-transfer events are less frequent in eukaryotes.

An additional rationale for neighbor regulation, which might also account for its lack in eukaryotes, is mechanistic. In prokaryotes, the processes of transcription and translation are coupled; therefore, it is possible that the concentration of a protein will be greater in the vicinity of the segment of DNA encoding its gene. It was thus suggested that some DNA-binding proteins regulate *cis* targets that are encoded close to their own genes more efficiently [29].

Concluding remarks

By applying network analysis methods, we managed to identify significant inter-relationships between chromosomal adjacency and transcription regulation automatically. We showed that in both *E. coli* and *S. cerevisiae* chromosomal adjacency affects the transcriptional regulatory program. This effect implies that transcription regulation considerations have been an important factor in shaping the order and orientation of TUs on the chromosome. Accordingly, the differences observed between the integrated networks of the two organisms reflect the differences between the transcriptional regulatory mechanisms of prokaryotes and eukaryotes.

Our findings might also have a predictive value: the high incidence of neighbor regulation in *E. coli* and *B. subtillis* suggests that in prokaryotes the cellular function of a TF-encoding TU can serve to predict the cellular function of its neighboring TU or vice versa. In *S. cerevisiae*, we found that clusters of co-regulated adjacent TUs are over-represented. Thus, finding that a TF binds to the promoters of adjacent TUs in genome-wide chromatin immunoprecipitation experiments might increase the reliability of these experimental results.

The findings presented here are important for a better understanding of both gene-expression regulation and evolution of chromosomal organization, and might also have implications in the genetic engineering of synthetic regulatory circuits.

Update

While this article was in press, a statistical analysis of the spatial distribution of operons in the transcriptional regulation network of *E. coli* was described by Warren

and ten Wolde [30]. The results of this study are consistent with our results in *E. coli*.

Supplementary material

Supplementary material and a detailed description of the methods used are available at: http://margalit.huji.ac.il/chromosomal_org

Acknowledgements

This study was supported by the Israeli Science Foundation, administered by the Israeli Academy of Sciences and Humanities. R.H. and E.Y.-L. are supported by the Yeshaya Horowitz Association through The Center for Complexity Science.

References

1 Lawrence, J.G. (2002) Shared strategies in gene organization among prokaryotes and eukaryotes. *Cell* 110, 407–413
2 Lawrence, J.G. (2003) Gene organization: selection, selfishness, and serendipity. *Annu. Rev. Microbiol.* 57, 419–440
3 Hurst, L.D. *et al*. (2004) The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* 5, 299–310
4 Cohen, B.A. *et al*. (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat. Genet.* 26, 183–186
5 Spellman, P.T. and Rubin, G.M. (2002) Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol.* 1, 5
6 Williams, E.J. and Bowles, D.J. (2004) Coexpression of neighboring genes in the genome of *Arabidopsis thaliana*. *Genome Res.* 14, 1060–1067
7 Rudd, K.E. (2000) EcoGene: a genome sequence database for *Escherichia coli* K-12. *Nucleic Acids Res.* 28, 60–64
8 Tjaden, B. *et al*. (2002) Transcriptome analysis of *Escherichia coli* using high-density oligonucleotide probe arrays. *Nucleic Acids Res.* 30, 3732–3738
9 Salgado, H.*et al.* (2004) RegulonDB (version 4.0): transcriptional regulation, operon organization and growth conditions in *Escherichia coli* K-12. *Nucleic Acids Res.* 32, Database issue: D303–D306
10 Christie, K.R. *et al*. (2004) *Saccharomyces* Genome Database (SGD) provides tools to identify and analyze sequences from *Saccharomyces cerevisiae* and related sequences from other organisms. *Nucleic Acids Res.* 32, D311–D314
11 Wall, M.E. *et al*. (2004) Design of gene circuits: lessons from bacteria. *Nat. Rev. Genet.* 5, 34–42
12 Shen-Orr, S.S. *et al*. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* 31, 64–68
13 Yeger-Lotem, E. *et al*. (2004) Network motifs in integrated cellular networks of transcription-regulation and protein–protein interaction. *Proc. Natl. Acad. Sci. U. S. A.* 101, 5934–5939
14 Lee, T.I. *et al*. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804
15 Milo, R. *et al*. (2002) Network motifs: simple building blocks of complex networks. *Science* 298, 824–827
16 Teichmann, S.A. and Babu, M.M. (2004) Gene regulatory network growth by duplication. *Nat. Genet.* 36, 492–496
17 Opel, M.L. *et al*. (2001) The effects of DNA supercoiling on the expression of operons of the *ilv* regulon of *Escherichia coli* suggest a physiological rationale for divergently transcribed operons. *Mol. Microbiol.* 39, 1109–1115
18 Hatfield, G.W. and Benham, C.J. (2002) DNA topology-mediated control of global gene expression in *Escherichia coli*. *Annu. Rev. Genet.* 36, 175–203
19 Hanamura, A. and Aiba, H. (1992) A new aspect of transcriptional control of the *Escherichia coli crp* gene: positive autoregulation. *Mol. Microbiol.* 6, 2489–2497
20 Ishizuka, H. *et al*. (1994) Mechanism of the down-regulation of cAMP receptor protein by glucose in *Escherichia coli*: role of autoregulation of the crp gene. *EMBO J.* 13, 3077–3082
21 Allen, T.E. *et al*. (2003) Genome-scale analysis of the uses of the *Escherichia coli* genome: model-driven analysis of heterogeneous data sets. *J. Bacteriol.* 185, 6392–6399
22 Yu, H. *et al*. (2003) Genomic analysis of gene expression relationships in transcriptional regulatory networks. *Trends Genet.* 19, 422–427
23 Korbel, J.O. *et al*. (2004) Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nat. Biotechnol.* 22, 911–917
24 Khorasanizadeh, S. (2004) The nucleosome: from genomic organization to genomic regulation. *Cell* 116, 259–272
25 Ihmels, J. *et al*. (2002) Revealing modular organization in the yeast transcriptional network. *Nat. Genet.* 31, 370–377
26 Makita, Y. *et al.* (2004) DBTBS: database of transcriptional regulation in *Bacillus subtilis* and its contribution to comparative genomics. *Nucleic Acids Res.* 32, Database issue: D75–D77
27 De Hoon, M.J. *et al*. (2004) Predicting the operon structure of *Bacillus subtilis* using operon length, intergene distance, and gene expression information. *Pac Symp Biocomput* 2004, 276–287
28 Pal, C. and Hurst, L.D. (2004) Evidence against the selfish operon theory. *Trends Genet.* 20, 232–234
29 McFall, E. (1986) *cis*-acting proteins. *J. Bacteriol.* 167, 429–432
30 Warren, B.P. and ten Wolde, P.R. (2004) Statistical analysis of the spatial distribution of operons in the transcriptional regulation network of *Escherichia coli*. *J. Mol. Biol.* 342, 1379–1390

0168-9525/$ - see front matter
doi:10.1016/j.tig.2005.01.003