Periodic organization of stress response genes on the E. coli chromosome.

Steff Horemans, Brian C. Jester, Laurent Janniere, Francois Kepes

Proceedings of the Évry Spring School
on

advances in
Systems and
Synthetic
Biology

March 19th - 23rd, 2018

Edited by

Patrick Amar, Attila Csikász-Nagy, François Képès, Vic Norris

"But technology will ultimately and usefully be better served by following the spirit of Eddington, by attempting to provide enough time and intellectual space for those who want to invest themselves in exploration of levels beyond the genome independently of any quick promises for still quicker solutions to extremely complex problems."

Strohman RC (1977) Nature Biotech 15:199

FOREWORD

Systems Biology includes the study of interaction networks and, in particular, their dynamic and spatiotemporal aspects. It typically requires the import of concepts from across the disciplines and crosstalk between theory, benchwork, modelling and simulation. The quintessence of Systems Biology is the discovery of the design principles of Life. The logical next step is to apply these principles to synthesize biological systems. This engineering of biology is the ultimate goal of Synthetic Biology: the rational conception and construction of complex systems based on, or inspired by, biology, and endowed with functions that may be absent in Nature.

This annual School started in 2002. It was the first School dedicated to Systems Biology in France, and perhaps in Europe. Since 2005, Synthetic Biology has played an increasingly important role in the School. Generally, the topics covered by the School have changed from year to year to accompany and sometimes precede a rapidly evolving scientific landscape. Its title has evolved in 2004 and again in 2012 to reflect these changes. The first School was held near Grenoble after which the School has been held in various locations. It started under the auspices of Genopole®, and has been supported by the CNRS since 2003, as well as by several other sponsors over the years.

This book gathers overviews of the talks, original articles contributed by speakers and students, tutorial material, and poster abstracts. We thank the sponsors of this conference for making it possible for all the participants to share their enthusiasm and ideas in such a constructive way.

Patrick Amar, Gilles Bernot, Marie Beurton-Aimar, Attila Csikász-Nagy, Oliver Ebenhoeh, Ivan Junier, Marcelline Kaufman, François Képès, Pascale Le Gall, Sheref Mansy, Jean-Pierre Mazat, Victor Norris, El Houssine Snoussi, Birgit Wiltschi.

ACKNOWLEDGEMENTS

We would like to thank the conference participants, who have contributed in a way or another to this book. It gathers overviews of the talks, discussions, original articles and tutorial material contributed by speakers, abstracts from attendees, short articles from students, posters and lectures proposed by the epigenesis groups to review or illustrate matters related to the scientific topic of the conference.

Of course the organisation team would like to express gratitude to all the staff of the *Ibis Evry Cathédrale* hotel for the very good conditions we have found during the conference.

Special thanks to the Epigenomics project for their assistance in preparing this book for publication. The cover photography shows a view of the *Khánh-Anh pagoda*, at Évry.

We would also like to express our thanks to the sponsors of this conference for their financial support allowing the participants to share their enthusiasm and ideas in such a constructive way.

They were:

- Genopole® Évry:
  http://www.genopole.fr
- Centre National de la Recherche Scientifique (CNRS):
  http://www.cnrs.fr
- Université d'Évry Val d'Essonnes et iSSB:
  http://www.univ-evry.fr
- Global Bioenergies
  http://www.global-bioenergies.com
- GDR CNRS 3003 Bioinformatique Moléculaire:
  http://www.gdr-bim.u-psud.fr
- Fondation Fourmentin-Guilbert:
  http://www.fourmentinguilbert.org
- Institut National de Recherche en Informatique et en Automatique (INRIA):
  http://www.inria.fr/
- LABEX Mathématique Hadamard:
  https://www.fondation-hadamard.fr
- Société Française de Biophysique:
  https://www.sfbiophys.org
- Advanced Analytical Technologies:
  https://www.aati-us.com

THE EDITORS

INVITED SPEAKERS

NED BUDISA	Technische Universität Berlin, Institut für Chemie, Berlin, D
ANNE CAMBON-THOMSEN	Épidémiologie et analyses en santé publique, INSERM / U. Toulouse, F
JÉRÔME FÉRET	Antique team, DI - École Normale Supérieure, Paris, F
MUSTAFA KHAMMASH	Department of Biosystems Science and Engineering, ETHZ, Zurich, CH
DOUGLAS MURRAY	Inst. For Advanced Biosciences, Keio University, JP
JEAN-PAUL DI RAGO	IBGC, CNRS / U. Bordeaux, F
VITOR PINHEIRO	Institute of Structural and Molecular Biology, University of London, UK
NICOLE RADDE	Institute for Systems Theory and Automatic Control, U. of Stuttgart, D
ANNE SIEGEL	Dyliss team, IRISA lab., Rennes, F
BEATRIX SÜESS	Department of Biology, Technische Universität Darmstadt, D
MATIAS ZURBRIGGEN	Inst. of Synthetic Biology, U. of Düsseldorf, D

CONTENTS

Periodic organization of stress response genes on the *E. coli* chromosome

Steff Horemans[1], Brian Jester[2], Laurent Jannière[1] and François Képès[2]

[1] iSSB, Génomique Métabolique, CEA, CNRS, Univ. Évry, Université Paris- Saclay, Genopole Campus 1, Genavenir 6, 5 rue Henri Desbruères, F-91030 Evry, France
[2] SYNOVANCE, Genopole Campus 1, Genavenir 6, 5 rue Henri Desbruères, F- 91030 Evry, France

Abstract

Escherichia coli has evolved a diverse set of transcription factors and σ factors to deal with the various stresses it encounters during its life cycle. Additionally, the very intricate structure of the *E. coli* chromosome has been found to change upon stress induction, but the importance of this for efficient stress response has been insufficiently understood. In this study, we have explored the chromosomal organization of the stress response regulons and used our Transcription Based Solenoidal framework to understand how this can implicate chromosome structure under stress and stress response control.

To this end, our in-house software GREAT:SCAN:patterns was used to perform periodicity analyses on the binding sites of the major stress response regulators in *E. coli*. Significant periodic organisations spanning the entire chromosome as well as specific regions were found for nearly all stress response regulons. These periodic organisations were distinct for different stresses, suggesting different chromosomal structures under different stresses, with one exception. A similar organisation was found in the case of the oxidative stress and iron metabolism regulation. This similarity may reflect the in vivo cofunction of these two regulatory systems. Furthermore, this work discusses the possibility how the observed chromosomal organization of the stress response genes may be compatible with exploiting known chromosomal structures such as supercoiling, microdomain and macrodomain organization.

These findings provide a new detailed picture of chromosome structure under stress conditions and may ultimately lead to new design principles for synthetic genomes and industrial strains.

1 Introduction

Bacterial chromosomes have to be compacted 500-1000 times to fit inside the bacterial cell, whilst staying accessible to the action of DNA processing proteins such as DNA polymerases, RNA polymerases and other DNA binding

proteins [1]. This is achieved by folding the chromosome into a remarkably intricate structure called the nucleoid of which the most detailed features are only now being rapidly discovered [1]. The lowest level of chromosome structure is defined by supercoiling or the twisting of the DNA helix [1]. Supercoiling is caused by the action of DNA processing enzymes like DNA and RNA polymerases, which overwind the DNA in front of them and underwind the DNA behind [2]. In order to manage this DNA supercoiling, cells have evolved dedicated enzymes called topoisomerases and gyrases, that can relax and add DNA twists, respectively. On a higher structural level, abundant nucleoid associated proteins (NAP's) organize the chromosome by locally constraining supercoiling in topologically isolated domains called microdomains. Experimental studies in *E. coli* have shown that its chromosome has 400 of these microdomains of average length between 8-12 kb [3]. Recently, new experimental studies have revealed that the *E. coli* chromosome is organized in domains of 100-200 kb long, in which DNA elements physically interact more frequently [4]. These domains represent yet another level of chromosome structural organisation. At the highest level of organization (1 Mbp), the chromosome is divided in four structured macrodomains (Ori, Left, Right, Ter) and two non- structured ones (NS-Left, NS-Right) [5]. The Ter macrodomain is organized by the MatP protein [6] and the Ori one by the maoP protein [7], but the other macrodomains seem to be defined by the replication process itself [8].

Interestingly, several features of chromosomal structural organisation have been shown to profoundly impact DNA transcription and replication while being responsive to environmental changes. For example, the degree of DNA supercoiling has been found to respond to several stresses including oxidative stress [9], osmotic stress [10], heat shock [11], cold shock [12] and nutrient starvation [13]. Conversely, supercoiling has been shown to directly affect the expression of many genes [14] and some stress responses in particular [15], making it the perfect integrator of environmental changes with gene expression responses. Moreover, constraining supercoiling to a localized domain on the chromosome is thought to cause genes within such microdomain to have correlated transcription and transcription bursting profiles, underlining the importance of these microdomains for gene expression control [16]. Macrodomains on the other hand seem to contribute to efficient replication [17,18] and limit genome plasticity [19]. To date, there is no known evidence suggesting a direct link between macrodomain organisation and gene expression.

The dependence of gene expression on its position in the chromosome sequence is a consequence of the link between chromosome structure and gene expression [20]. This implies that there is an evolutionary pressure to organize

the positioning of genes in ways that optimize their expression.

Several *in silico* analyses have shown that most genes are not randomly scattered along the chromosome sequence, but follow distinct chromosome layouts that reflect structural features of the chromosome [21,22,23]. Computational analysis of fully sequenced genomes showed that coregulated genes tended to be organized close together in regulons of 20 genes or less, whereas larger regulons tended to be organized periodically along the chromosome in many different organisms [21], most notably in *E. coli* [24] and *Saccharomyces cerevisiae* [25]. In order to explain this phenomenon, the Transcription based Solenoidal (TBS) framework of chromosome expression was elaborated by Kepes and Vaillant in 2003 [26]. According to this model, gene expression regulation is optimized by confining the necessary transcription factors, RNA polymerases and coregulated genes into a small nuclear volume called a transcription focus. Like other chemical reactions, the confining of all necessary factors inside this small volume was thought to improve the speed and robustness of gene expression and regulation [26]. These transcription foci were thought to self-assemble through mediation of bivalent transcription factors, which would recruit their target genes and then even more transcription factors to the focus in a positive feedback loop. According to the model, the spatial clustering of coregulated genes is facilitated either by organizing them close together on the chromosome sequence (1D clustering) or by organizing them periodically along the sequence (interpreted as 3D clustering). In this scenario, the model states that the 3D folding of the chromosome will bring them together in space either by folding into solenoids (eukaryotes) or plectonemes (prokaryotes) [26]. The Transcription Based Solenoidal framework establishes a clear link between chromosome organization, structure and expression. In this work, we will analyze the periodic organization of stress response genes and study its implications for chromosome structure and stress response regulation through the lens of the Transcription Based Solenoidal framework.

As most bacteria, *E. coli* has to cope with a rapidly changing and often stressful environment. To this end, *E. coli* has evolved mechanisms capable of sensing and coping with environmental stress. Firstly, *E. coli* has acquired a diverse set of transcription factors dedicated to sensing environmental stress and altering gene expression to increase stress resistance. These transcription factors include OxyR (H_2O_2 stress) [27], SoxR and SoxS [27] (redox cycling compound stress), Fur (iron deprivation) [28], GadE (acid stress) [29] and OmpR (osmotic stress) [30]. Additionally, *E. coli* has also evolved a set of alternative σ factors that change the specificity of its unique RNA polymerase toward stress response genes. These include σ^{24} (Extreme heat shock), σ^{28} (Flagellar synthesis/chemotaxis), σ^{32} (Heat shock), σ^{38} (General starvation) and σ^{54} (Nitrogen regulated genes) [31]. Interestingly, *E. coli* has also been

shown to drastically change the structure of its chromosome under stress conditions with global impacts on gene expression [14]. As an extreme example, the NAP Dps self-aggregates under starvation conditions while bound to DNA and thus condenses the chromosome and physically protects it from damage [32]. Despite recent advances in the elucidation of the *E. coli* chromosome structure under non-stressed conditions [33], the 3D structure of the chromosome under stress conditions, its impact on gene expression and the switching mechanisms between different chromosomal structures have all been far less understood. This knowledge is indispensable for increased understanding of the functioning of natural chromosomes as well as the elucidation of design principles for synthetic chromosomes.

The goal of this project was to study the chromosomal layout of the E. coli stress response genes and to use the TBS framework to understand its implications for gene expression control and chromosome structure under stress conditions. This was achieved by extracting the binding positions of OxyR, SoxS, Fur, GadE, OmpR, σ^{24} and σ^{28} from the literature [27,28,29,30,,34] and analyzing them with our periodicity analysis software GREAT:SCAN:patterns. This allowed us to:

- Search for global and local periods in the chromosome (section 3.1).
- Analyze their implications for chromosome structure and function (section 3.1).
- Visualize 3D clustering of the transcription factor binding sites as predicted by the TBS framework (section 3.2).

2 Method

GREAT:SCAN:Patterns

In order to investigate the interplay between periodic chromosome organization, chromosome expression and chromosome structure, our group has developed GREAT:SCAN:patterns, the main computational tool used in this work. GREAT:SCAN:patterns is a free, web-based, online tool that exhaustively analyzes all periodic patterns that can be detected from the positions of coregulated genes in the genome [35]. The input of GREAT:SCAN:patterns simply consists of a file with gene identifiers and their chromosome positions. Depending on the nature of the data that was used to create the input file, chromosome positions may mean the 5'end of the genes (transcriptome data, 3C interaction data) or the middle of a transcription factor binding site (ChIP data). Briefly, it works in three steps. The first step involves the detection and

visualization of all periods along the full length chromosome. Initially, the data is preprocessed to remove the effects of proximal gene identifiers. This is necessary, because proximal gene identifiers can artificially increase the p-value of the periods when paired with another, sufficiently distant gene identifier. This can lead to false positive periods [36]. To avoid this problem, the program replaces a set of proximal gene identifiers by its barycentre. Proximity is defined as two times the average gene distance, a parameter specified by the user (E.coli:1000 bp) [36]. Next, the p-value of each period is calculated. This p-value is defined as the probability to have a higher periodicity score when an equal number of positions is randomly drawn from a uniform distribution. This periodicity score is calculated by first performing a modulo operation on the chromosome positions for each period and then performing a cluster analysis of the resulting phase coordinates, which rewards both dense and poor regions [37]. These p-values are then corrected with a multiple testing correction that takes the period length into account. This is necessary because many periods get tested for relatively short periods, which increases the odds of finding a false positive. The periods are subsequently visualized in a periodogram. In a second step, the chromosome is scanned to determine the length of each region over which a periodicity occurs. To do this, GREAT:SCAN:patterns uses a variable size sliding window that scans the chromosome for regions in which periodic organisation is found, starting with a window size of 10 kb that geometrically increases with a constant factor and ends with a sliding window that encompasses 95 % of the chromosome. Period p-values are subsequently calculated in the same way as for the whole chromosome.

The last step involves creating a clustergram. This representation plots the phase (remainder of the modulo division of the gene coordinate on the chromosome by the period) against the gene identifier. Next, DBSCAN, a density based clustering algorithm, is used to detect clusters of "in phase" genes based on the phase coordinates.

3 Results

GREAT:SCAN:patterns reveals periodic organisations in the stress response regulons

The first goal of this work was to study the periodic organization of the binding sites of several stress response regulators on the chromosome. To this end, the exact binding positions of OxyR, SoxS, GadE, Fur, OmpR, σ^{19} σ^{24}, σ^{28}, σ^{32} and σ^{54} were extracted from Chromatin Immuno Precipitation (ChIP) experiments, remapped on the newest version of the E. coli K12 MG1655 chromosome (NC_000913.3) and analyzed with GREAT:SCAN:patterns. Sig-

nificant periods spanning the entire chromosome were detected in all datasets, except for σ^{32} and σ^{54}. Additionally, given the low amount of binding sites (7) of σ^{19}, a false positive result could not be excluded and was therefore dropped from further analysis. These results are summarized in Table 1. Additionally, the same period detection analysis was performed along stretches of the chromosome to define periodic subregions in the chromosome. These periods are summarized in Table 2. The regions associated with these periods are shown in Figure 1. To better understand the relationships between the found periods, a harmonic analysis was performed to see which periods were multiples of other periods and therefore shared common chromosomal organisations. Harmonic periods were allowed to differ 1% maximally. Periods were considered to be part of the same harmonic family if they were a harmonic of at least one period in the family. Related harmonic periods and their corresponding regulons are shown for each period in tables 1 and 2.

Global periodicity analysis				
Regulon	Period (kb)	p-value	Cis harmonic Periods	Trans harmonic periods
SoxS (24) *	15	0,0319		
	40,5	0,0452	81/121,4/202,5	5,8 (Fur Fe starv)/ 20,3 (Fur Fe replete)
	54,4	0,0151		
	61,7	0,0254		
	81	0,0089	40,5	5,8 (Fur Fe starv)/ 20,3 (Fur Fe replete)/ 16,2 (OxyR)
	202,5	0,0234	40,5	5,8 (Fur Fe starv)/ 20,3 (Fur Fe replete)
σ^{28} (50)*	8,2	0,0094	8,2	16,5 (σ^{24})
	144,2	0,0362	288,5	
	288,5	0,0394	144,2	
	342,6	0,0222		342,5 (GadE)
σ^{24} (56)*	5,5	0,0018	11/16,5/22,1	
	11	0,024	5,5/22,1	
	16,5	0,0018	5,5	8,2 (σ^{28})
	22,1	0,0392	5,5/11	
Fur (Fe replete) (110)*	8,6	0,0222		
	227,3	0,041		
Fur (Fe starv) (59)*	148,1	0,0414		
	237,1	0,0241		
OmpR (24)*	13,2	0,0067	26,5	
	26,5	0,0273	13,2	
GadE (14)*	325,8	0,033		
OxyR (21)*	16,2	0,0114		81 (SoxS)

Table 1: Periodicity analysis of stress response regulons: Periods observed by considering binding sites along the entire chromosome. * denotes the number of binding sites in each dataset. Different background colors are added to easily distinguish the results for different datasets.

Tables 1 and 2 show significant periodic organisations in the binding positions of all the stress regulons. Short periods (5.5-25 kb), mid-range periods (25 kb-100 kb) and long range periods ($>$ 100 kb) are frequently observed across the datasets. Additionally, the local analysis in table 2 shows that most peri-

odic regions incorporate a vast majority of the total number of binding sites (table 1), which strengthens their importance. Harmonic analysis showed that harmonic periods were frequently found within datasets, but infrequently between, suggesting that the different stress response binding sites are organized independently on the chromosome. A strong exception is the harmonic link between the OxyR, SoxS and Fur (both Fe replete and Fe starv) binding sites where the major periods belong to the same harmonic family. Additionally, harmonic periods are found between $\sigma^{24/28}$, σ^{24} and OxyR, σ^{28} and OxyR and GadE and σ^{28}. Harmonic analysis also showed that many mid-range periods were harmonics of short periods, although several periods in the OxyR and SoxS datasets could not be connected to a smaller one. (29.4/68.1 kb for OxyR and 54.4/61.7 kb for SoxS).

Local periodicity analysis					
Regulon	Period	Pval	Binding sites in region	Cis harmonic periods	Trans harmonic periods
	9,3	0,004	15		
SoxS	40,5	0,003	20	81/121,4/202,5	5,8 (Fur Fe starv)/ 20,3 (Fur Fe replete)
(24) *	81	0,003	15	40,5	5,8 (Fur Fe starv)/ 20,3 (Fur Fe replete)/ 16,2 (OxyR)
	121,4	0,005	19	40,5	5,8 (Fur Fe starv)/ 20,3 (Fur Fe replete)
σ^{28}	8,2	0,001	38	8,2	16,5 (σ^{24})
(50)*	294,7	0,001	37		29,4 (OxyR)
	5,5	0,0003	49	5,5/11/16,5/22,1	
σ^{24}	15,4	0,0002	41	30,8	30,8 (OxyR)
(56)*	30,8	0,0001	41	15,4	30,8 (OxyR)
	8,9	0,0004	32		
Fur (Fe replete)	20,3	0,001	83		40,5/81/121,4/202,5 (SoxS)
(110)*	228,6	0,003	75	227,3	227,8 (Fur Fe starv)
Fur (Fe starv)	5,8	0,0006	52		40,5/81/121,4/202,5 (SoxS)
(59)*	227,8	0,004	48		227,3/228,6 (Fur Fe replete)
	10,5	0,002	9		
OmpR (24)*	13,2	0,003	22	26,5	
	15,2	0,004	15		
	5,6	0,001	11		
	14,2	0,004	10	28,5	
GadE (14)*	28,5	0,0001	10	14,2	
	149,9	0,001	11		
	342,5	0,003	12		342,6 (σ^{28})
	29,4	0,002	11		294,7 (σ^{24})
OxyR (21)*	30,8	0,003	11		15,4/30,8 (σ^{24})
	68,1	0,002	11		

Table 2: Periodicity analysis of stress response regulons: Periods observed by considering binding sites in subregions of the chromosome defined by the sliding window approach. * denotes the number of binding sites in each dataset. Different background colors are added to easily distinguish the results for different datasets.

Figure 1 shows that most of the periodic regions in table 2 cover large stretches of the chromosome. Interestingly, in the case of the dataset for each regulon, large regions of the chromosome do not reveal periodicity. Furthermore, several harmonic periods within datasets also map to the same region on the chromosome, further strengthening the idea that these periods are identical (example: GadE 14.2/28.5 kb). Also, all datasets except OxyR contain two or more local regions without periodicity that cover a different portion of the

chromosome and even in the case of OxyR, a large number of binding sites is excluded from the region with periodicity. The boundaries of these regions were therefore compared to macrodomain boundaries in order to investigate if macrodomains play a role in establishing regions with and without a certain periodic organisation. However, no clear overlap was found.

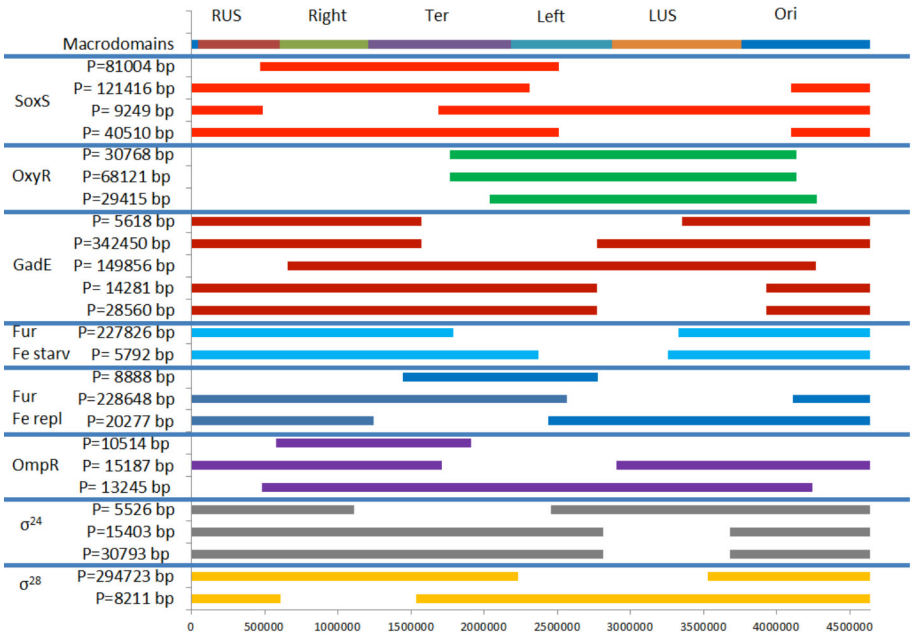

Figure 1: Periodic domain organisation shows large periodic domains. (Colored bars) Chromosomal coordinates (horizontal axis) are plotted against periods (vertical axis) detected by the sliding window approach.

Clustergrams of OxyR and SoxS reveal spatial clustering under their shared period

The second goal of this work was to study how these periodic organisations can contribute to spatial clustering of the binding sites and improved gene expression control according to the TBS framework. To this end, the phase coordinates of the binding sites were calculated and plotted against the gene that is regulated by this binding site in clustergrams. Clusters were detected using the DBSCAN algorithm. The clarity of this representation is dependent on both the size of the period and the size of the dataset. Therefore, the discussion here is limited to the clustergrams of the 16.2 kb OxyR period and the 81 kb SoxS period (Figure 2). Binding sites that don't belong to a

cluster are depicted as grey circles. In both cases, large clusters of binding sites were detected as predicted by the TBS framework. It is important to note that the clustergram representation wraps around. The OxyR period creates two clusters of binding sites. One (black triangles) contains the majority of OxyR binding sites with the binding site of the *oxyR* gene in the middle of the cluster. This cluster also contains the OxyR binding site of the *fur* gene. The other cluster (blue triangles) contains the OxyR binding sites of the genes that code for most reactive oxygen species (ROS) degrading enzymes including katG, ahpCF and sodA and ROS degrading enzyme/nucleoid binding protein dps. Incidentally, the OxyR binding site next to *soxS* is not part of any cluster. On the other hand, the SoxS period also creates two clusters of binding sites (black and blue triangles). The black cluster contains the SoxS binding sites next to the *soxS* and *fur* genes, but excludes the one of *sodA*, which is considered to be SoxS' most important target from the physiological viewpoint. The blue cluster contains the acrAB efflux pump, membrane protein ypjC, unknown ypeC and ycgZ a regulator of *ompF* which encodes a porin.

4 Discussion

The goal of this project was to study the chromosomal layout of the *E. coli* stress response genes and to use the TBS framework to understand its implications for the coordination of gene expression and chromosome structure under stress conditions. This was achieved by extracting the binding positions of OxyR, SoxS, Fur, GadE, OmpR, σ^{24} and σ^{28} from the literature [27,28,29,30, 34] and analyzing them with our periodicity analysis software GREAT:SCAN:patterns.

Both our local and our global periodicity analysis showed significant periods in three size ranges: short (5.5-25 kb), mid-range (25-100 kb) and long (gt 100 kb).

Most short periods in the stress regulons have sizes up to 16 kb, which was found by others to be one characteristic length of transcriptomic data in *E. coli* [38]. Interestingly, these authors also observed a drop in characteristic length from 7.5 kb +/-0.9 kb to 5.6 kb +/-1.3 kb periods when supercoiling was abolished by the inhibition of gyrase activity. The short period lengths also correspond remarkably well to the size of microdomains under normal conditions3. The authors therefore hypothesized that supercoiling was responsible for the establishment of the short characteristic lengths (up to 16 kb) they observed. Many of the short range periods observed in this study fall within the 7.5 (SoxS, Fur Fe replete, σ^{28}) and 5.5 kb (σ^{24}, GadE, Fur Fe starv) ranges as well as in a larger period range between 13-16 kb (OxyR, OmpR GadE,

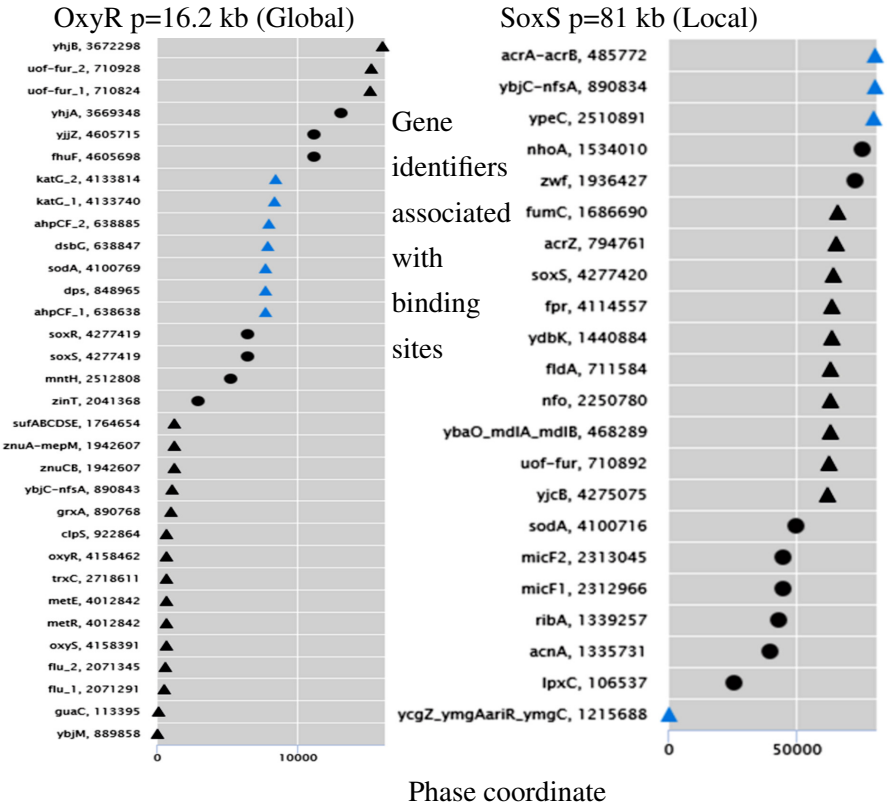

Figure 2: Clustergams of OxyR and SoxS reveal spatial clustering of the TF binding sites. (triangles) Phase coordinates (horizontal axis) of binding sites are plotted against gene identifier and chromosomal position. (Vertical axis) Binding sites that don't belong to a cluster are depicted as grey circles.

σ^{24}). Moreover, a rich body of literature shows that cellular supercoiling levels change in all the stress conditions analyzed here [9,10,11,12,13]. Thus our data suggests that the binding sites of the stress regulators are organized along the chromosome in order to take advantage of supercoiling induced structural changes. Our TBS model further suggests that these supercoiling induced structural changes will help bring together the correct stress response genes, potentially adding another indirect layer to gene expression through supercoiling.

The mid-range periods (25-100 kb) do not associate with known structural features of the *E. coli* chromosome. However, our harmonic analysis showed that many of these periods were multiples of short range periods (σ^{24} 15.4/30.8, GadE 14.2/ 28.5, SoxS 40.5 kb/81 kb). Especially the SoxS 81 kb

period is fascinating, as it is a multiple of the Fur Fe replete 20.3, the 16.2 kb OxyR period and the Fur Fe starvation 5.8 kb period. This suggests that some regulons with mid-range periods are organized to accommodate multiple chromosomal structures, which correspond to the multiple conditions under which the regulon should be active.

The long range 100 kb+ periods observed in the datasets are either multiples of the short range periods observed before, or correspond to a higher form of chromosome organization, or both. These periods have been observed before in *E. coli* and *Bacillus subtilis* [38,39]. The 121.5 kb period of SoxS (Table 2) seems to be a clear example of a 100 kb+ being a harmonic of a smaller period. Conversely, periods of 100 kb length scale may correspond to the length scale of Chromatin Interaction Domains (CID's), which have been observed experimentally in *Caulobacter crescentus* [4]. These domains are characterized by increased spatial clustering [4]. Additionally, our local analysis demonstrated regions with strong periodic patterns that covered large stretches of the chromosome, but also showed stretches of the chromosome without a certain periodic pattern, even though these regions weren't devoid of transcription factor binding sites. These regions were compared to macrodomain boundaries and only loose associations were found. One explanation may be that this is due to the small size of most datasets, which prevents GREAT:SCAN:patterns to clearly establish the boundaries of the periodic regions. Alternatively, macrodomain boundaries or other higher order structures have not been investigated under stress conditions and may be subject to change, a possibility meriting further investigation.

The harmonic analysis demonstrated that most harmonic periods were found within datasets and relatively few between, suggesting that most stress responses induce unique chromosome conformations. An interesting exception is formed by OxyR, SoxS and Fur, where the major periods of each dataset were found to be harmonics of each other. This is an interesting observation since an important part of oxidative stress defense involves the regulation of iron metabolism, due to the fact that Fe2+ can react with H_2O_2 to form highly toxic OH radicals [40]. Additionally, we found that the 81 kb period of SoxS was both a multiple of the major periods under iron replete and iron deficient conditions. These new results from the harmonic analysis suggest that the OxyR, SoxS and Fur regulons are organized along the chromosome in order to exploit a structure that permits coordinated expression of the regulons while the cell transitions from an Fe^{2+} rich state to an Fe^{2+} poor state. This transition is indeed known to involve strong implications for oxidative stress status in the cell.

To study the implications of this particular organization on chromosome structure further, the 3D clustering of the OxyR and SoxS regulons was inves-

tigated using clustergram representations. These results largely corresponded with the predictions of the TBS framework. Most binding sites were found in clusters in both the OxyR and SoxS regulon. The binding sites of the *oxyR* and *soxS* genes were both found within their clusters, an organization that decreases transcription factor search time and improves gene expression control according to the TBS framework. The OxyR and SoxS binding sites of the *fur* gene were found inside the major clusters, suggesting that the OxyR, SoxS and Fur regulons form a single spatial cluster of gene expression when all regulons are activated, rather than three separate ones. Moreover, this organisation allows each regulon as well as each combination of regulons to efficiently cluster in phase space.

However, some observations on the clustergram remain puzzling. For instance, under the OxyR period a separate cluster of genes was observed for important ROS degrading enzymes. This cluster, however, is exactly half a period away from the major one, which means it would fuse into a single cluster if the physiological period is 8.1 kb instead of 16.2 kb. This is an attractive possibility, given the closeness of 8.1 kb to microdomain lengths. Another remarkable observation is the fact that *sodA* is in a cluster under the OxyR period, but not under the SoxS period. This is remarkable, because *sodA* codes for a superoxide dismutase [41] and is therefore of paramount importance for defense against superoxides and redox cycling compounds. A possible explanation relies on the fact the TBS framework requires transcription factors that can form multimers to establish transcription foci. OxyR is known to form tetramers [42], but SoxS and Fur are not, meaning it may be more important to be in phase with OxyR than SoxS or Fur. Last, SoxS is also not in phase with any of the OxyR clusters. This may point to the formation of spatially close, but separate SoxS, Fur and OxyR clusters. Alternatively, the TBS framework may be too strict a representation of chromosome architecture and microdomain size may be more flexible than assumed by the model. This view is supported by the fact that microdomain sizes are not equal across the chromosome, but exhibit a certain distribution3.

In summary, the results presented here suggest that the binding sites of major stress response regulators are not randomly scattered across the chromosome, but form clear periodic layouts. In most cases, our results show distinct periodic layouts for the different stress response regulators, suggesting distinct associated chromosome structures. An important exception is formed by the oxidative stress regulon and iron metabolism regulon, which share a single family of periods and have to work together to fend off oxidative stress. Furthermore, our data suggest important roles for supercoiling, microdomains and maybe higher order structures in stress response regulation. Ultimately, the further elucidation and experimental verification of the organizational princi-

ples of the chromosome presented here will lead to a dynamic understanding of the regulation of chromosome structure and open new paths to rational engineering of synthetic chromosomes and industrial strains.

References

[1] Dorman, C. J. Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nat Rev Microbiol* **11**, 349–355 (2013).

[2] Liu, L. F. & Wang, J. C. Supercoiling of the DNA template during transcription. *Proc. Natl. Acad. Sci.* **84**, 7024–7027 (1987).

[3] Postow, L., Hardy, C. D., Arsuaga, J. & Cozzarelli, N. R. Topological domain structure of the Escherichia coli chromosome. *Genes Dev.* **18**, 1766–1779 (2004).

[4] Le, T. B. K., Imakaev, M. V., Mirny, L. A. & Laub, M. T. High-Resolution Mapping of the Spatial Organization of a Bacterial Chromosome. *Science* **342**, 731–734 (2013).

[5] Valens, M., Penaud, S., Rossignol, M., Cornet, F. & Boccard, F. Macrodomain organization of the Escherichia coli chromosome. *EMBO J.* **23**, 4330–4341 (2004).

[6] Thiel, A., Valens, M., Vallet-Gely, I., Espéli, O. & Boccard, F. Long-range chromosome organization in E. coli: A site-specific system isolates the ter macrodomain. *PLoS Genet.* **8**, 112 (2012).

[7] Valens, M., Thiel, A. & Boccard, F. The MaoP/maoS Site-Specific System Organizes the Ori Region of the E. coli Chromosome into a Macrodomain. *PLoS Genet.* **12**, 123 (2016).

[8] Duigou, S. & Boccard, F. Long range chromosome organization in Escherichia coli: The position of the replication origin defines the non-structured regions and the Right and Left macrodomains. *PLoS Genet.* **13**, (2017).

[9] Horiuchi, H., Takagi, M. & Yano, K. Relaxation of supercoiled plasmid DNA by oxidative stresses in Escherichia coli. *J. Bacteriol.* **160**, 1017–1021 (1984).

[10] McClellan, J. A., Boublíková, P., Palecek, E. & Lilley, D. M. Superhelical torsion in cellular DNA responds directly to environmental and genetic factors. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 8373–8377 (1990).

[11] Mizushima, T., Natori, S. & Sekimizu, K. Relaxation of supercoiled DNA associated with induction of heat shock proteins in Escherichia coli. *Mol Gen Genet* **238**, 15 (1993).

[12] Mizushima, T., Kataoka, K., Ogata, Y., Inoue, R. & Sekimizu, K. Increase in negative supercoiling of plasmid DNA in Escherichia coli exposed to cold shock. *Mol. Microbiol.* **23**, 381–386 (1997).

[13] Balke, V. L. & Gralla, J. D. Changes in the linking number of supercoiled DNA accompany growth transitions in Escherichia coli. *J. Bacteriol.* **169**, 4499–4506 (1987).

[14] Peter, B. J. et al. Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. *Genome Biol.* **5**, R87. 1–15 (2004).

[15] Schneider, R., Travers, A. & Muskhelishvili, G. The expression of the Escherichia coli fis gene is strongly dependent on the superhelical density of DNA. *Mol. Microbiol.* **38**, 167–175 (2000).

[16] Bohrer, C. H. & Roberts, E. A biophysical model of supercoiling dependent transcription predicts a structural aspect to gene regulation. *BMC Biophys.* **9**, 2 (2015).

[17] Lesterlin, C., Mercier, R., Boccard, F., Barre, F. X. & Cornet, F. Roles for replichores and macrodomains in segregation of the Escherichia coli chromosome. *EMBO Rep.* **6**, 557–562 (2005).

[18] Espéli, O. et al. A MatP-divisome interaction coordinates chromosome segregation with cell division in E. coli. *EMBO J.* **31**, 3198–3211 (2012).

[19] Esnault, E., Valens, M., Espéli, O. & Boccard, F. Chromosome Structuring Limits Genome Plasticity in Escherichia coli. *PLoS One* **9**, 18 (2007).

[20] Bryant, J. A., Sellars, L. E., Busby, S. J. W. & Lee, D. J. Chromosome position effects on gene expression in Escherichia coli K-12. *Nucleic Acids Res.* **42**, 11383–11392 (2014).

[21] Junier, I., Hérisson, J. & Képès, F. Genomic organization of evolutionarily correlated genes in bacteria: Limits and strategies. *J. Mol. Biol.* **419**, 369–386 (2012).

[22] Sobetzko, P., Travers, A. & Muskhelishvili, G. Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc. Natl. Acad. Sci.* **109**, E42-E50 (2012).

[23] Sobetzko, P., Glinkowska, M., Travers, A. & Muskhelishvili, G. DNA thermodynamic stability and supercoil dynamics determine the gene expression program during the bacterial growth cycle. *Mol. Biosyst.* **9**, 1643 (2013).

[24] Képès, F. Periodic transcriptional organization of the E. coli genome. *J. Mol. Biol.* **340**, 957–964 (2004).

[25] Képès, F. Periodic epi-organization of the yeast genome revealed by the distribution of promoter sites. *J. Mol. Biol.* **329**, 859–865 (2003).

[26] Képès, F. & Vaillant, C. Transcription-Based Solenoidal Model of Chromosomes. *Complexus* **1**, 171–180 (2004).

[27] Seo, S. W., Kim, D., Szubin, R. & Palsson, B. O. Genome-wide Reconstruction of OxyR and SoxRS Transcriptional Regulatory Networks under Oxidative Stress in Escherichia coli K-12 MG1655. *Cell Rep.* **12**, 1289–1299 (2015).

[28] Seo, S. W. et al. Deciphering fur transcriptional regulatory network highlights its complex role beyond iron metabolism in Escherichia coli. *Nat. Commun.* **5**, 110 (2014).

[29] Seo, S. W., Kim, D., O'Brien, E. J., Szubin, R. & Palsson, B. O. Decoding genome-wide GadEWX-transcriptional regulatory networks reveals multifaceted cellular responses to acid stress in Escherichia coli. *Nat. Commun.* **6**, 18 (2015).

[30] Seo, S. W. et al. Revealing genome-scale transcriptional regulatory landscape of OmpR highlights its expanded regulatory roles under osmotic stress in Escherichia coli K-12 MG1655. *Sci. Rep.* **7**, 110 (2017).

[31] Chung, H. J., Bang, W. & Drake, M. a. Comprehensive Stress Response of Reviews Escherichia coli in Food Science and Food Safety. *Compr. Rev. inFood Sci. Food Saf.* **5**, 52–64 (2006).

[32] Ceci, P. et al. DNA condensation and self-aggregation of Escherichia coli Dps are coupled phenomena related to the properties of the N-terminus. *Nucleic Acids Res.* **32**, 5935–5944 (2004).

[33] Hacker, W. C., Li, S. & Elcock, A. H. Features of genomic organization in a nucleotide-resolution molecular model of the Escherichia coli chromosome. *Nucleic Acids Res.* **45**, 7541–7554 (2017).

[34] Cho, B.-K., Kim, D., Knight, E. M., Zengler, K. & Palsson, B. O. Genome- scale reconstruction of the sigma factor network in Escherichia coli: topology and functional states. *BMC Biol.* **12**, 4 (2014).

[35] Bouyioukos, C., Elati, M. & Képès, F. Analysis tools for the interplay between genome layout and regulation. *BMC Bioinformatics* **17**, 191 (2016).

[36] Bouyioukos, C., Elati, M. & Képès, F. Protocols for probing genome architecture of regulatory networks in hydrocarbon and lipid microorganisms. *Hydrocarb. Lipid Microbiol. Protoc. - Springer Protoc. Handbooks* 1–29 (2015). doi:10.1007/8623

[37] Junier, I., Hérisson, J. & Képès, F. Periodic pattern detection in sparse boolean sequences. *Algorithms Mol. Biol.* **5**, 31 (2010).

[38] Jeong, K. S., Ahn, J. & Khodursky, A. B. Spatial patterns of transcriptional activity in the chromosome of Escherichia coli. *Genome Biol.* **5**, R86 (2004).

[39] Carpentier, a, Torrésani, B., Grossmann, a & Hénaut, a. Decoding the nucleoid organisation of Bacillus subtilis and Escherichia coli through gene expression data. *BMC Genomics* **6**, 84 (2005).

[40] Imlay, J. A. The molecular mechanisms and physiological consequences of oxidative stress: Lessons from a model bacterium. *Nat. Rev. Microbiol.* **11**, 443–454 (2013).

[41] Britton, L. & Fridovich, I. Intracellular localization of the superoxide dismutases of Escherichia coli: a Intracellular Localization of the Superoxide Dismutases of Escherichia coli: a Reevaluation. **131**, 815–820 (1977).

[42] Kullik, I., Toledano, M. B., Tartaglia, L. A. & Storz, G. Mutational Analysis of the Redox-Sensitive Transcriptional Regulator OxyR: Regions Important for DNA Binding and Multimerization. *J. Bacteriol.* **177**, 1285–1291 (1995).